

File ID	uvapub:175958
Filename	Proefschrift
Version	final

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	PhD thesis
Title	Beoordelen van informatievaardigheden in het hoger onderwijs
Author(s)	A.A.J. van Helvoort
Faculty	FGw
Year	2016

FULL BIBLIOGRAPHIC DETAILS:

<http://hdl.handle.net/11245/1.539501>

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).

BEOORDELEN VAN INFORMATIEVAARDIGHEDEN IN HET HOGER ONDERWIJS

Jos van Helvoort



**BEOORDELEN VAN
INFORMATIEVAARDIGHEDEN
IN HET HOGER ONDERWIJS**

JOS VAN HELVOORT

© Jos van Helvoort, Delft, 2016

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Beeld Omslag: Putti in een bibliotheek, atelier van Bernard Picart, 1712,
Collectie Rijksmuseum Amsterdam, © Rijksmuseum Amsterdam

Vormgeving: Mustafa Özbek

Druk: CPI Koninklijke Wöhrmann, Zutphen

ISBN 978-90-73077-79-9

NUR 840

BEOORDELEN VAN INFORMATIEVAARDIGHEDEN IN HET HOGER ONDERWIJS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 15 september 2016, te 14.00 uur
door Adrianus Antonius Josephus van Helvoort
geboren te Eindhoven.

Promotiecommissie:

Promotoren: Prof. Dr. F.J.M. Huysmans, Universiteit van Amsterdam
Prof. Dr. F.L.J.M. Brand-Gruwel, Open Universiteit

Copromotor: Dr. E. Sjoer, De Haagse Hogeschool

Overige leden: Prof. Dr. T.H.P.M. Thomassen, Universiteit van Amsterdam
Prof. Dr. J.S. Mackenzie Owen, Universiteit van Amsterdam
Prof. Dr. M.L.L. Volman, Universiteit van Amsterdam
Dr. D. Joosten-ten Brinke, Open Universiteit
Prof. Dr. P. Nieuwenhuysen, Vrije Universiteit Brussel

Faculteit: Geesteswetenschappen

Het hier beschreven onderzoek werd mede mogelijk gemaakt door steun van

De Stichting GO Fonds

Stichting **GO**FONDS
ontwikkeling van de informatiedienstverlening

De Haagse Hogeschool

DE HAAGSE
HOGESCHOOL

Voor mijn ouders

Riet van Helvoort - van Hulten † 6 februari 2015

Cees van Helvoort † 9 februari 2015

Inhoudsopgave

Voorwoord	1
1 Probleemstelling en context van het onderzoek	5
1.1 Relevantie van Informatievaardigheden	7
1.1.1 Informatiesamenleving	7
1.1.2 Web 2.0	8
1.1.3 Een leven lang leren	10
1.1.4 Ontwikkelingen in het onderwijs	11
1.1.5 Rol van onderzoek in het hoger onderwijs	11
1.2 Probleembeschrijving en doelstelling	13
1.3 Eisen aan het beoordelingsinstrument	14
1.3.1 Generieke toepasbaarheid	14
1.3.2 Aansluiting bij actuele ontwikkelingen op het gebied van toetsing en onderwijs	15
1.4 Onderzoeksvragen en onderzoeksmethodes	16
1.5 Opbouw van het proefschrift	18
Deel I Theoretische uitgangspunten en constructie van het beoordelingsmodel	21
2 Wetenschappelijk onderzoek naar informatievaardigheden	23
2.1 Verschillende onderzoeksbenaderingen	25
2.1.1 Uitkomstgerichte gedragsbenadering ('behavioural approach')	25
2.1.2 Procesgerichte constructivistische benadering	26
2.1.3 Relationele benadering	28
2.2 Relevantie van de drie onderzoeksbenaderingen voor de beantwoording van de centrale onderzoeksvraag	29
2.3 Verschillende modellen	31
2.4 Een passend IPS-model voor het onderhavige onderzoek	39
2.5 Constructie van (metacognitieve) kennis	42
2.6 Conclusies en discussie	45
3 Meten van informatievaardigheden	47
3.1 Vragenlijsten	50
3.1.1 Tests	50
3.1.2 Enquêtes	53
3.2 Performance assessments	55
3.2.1 Studieopdrachten	56
3.2.2 Portfolio's	58
3.2.3 Directe observatie	59
3.2.4 Scoringsrubrieken	60
3.3 Conclusies en discussie	62

4 Constructie van de scoringsrubriek	65
4.1 Ontwerpprocedure	66
4.2 Verkenning	68
4.3 Uitwerking	70
4.4 Testen	73
4.4.1 Methode	73
4.4.2 Resultaten	73
4.4.3 Revisie van de eerste concept versie	74
4.5 Prototype na revisie	75
4.6 Conclusies en discussie	76
Deel II Testen en evalueren van het prototype van de scoringsrubriek	79
5 Evalueren van de bruikbaarheid bij Saxion; Testen van betrouwbaarheid en validiteit bij de Nederlandse Defensie Academie en Open Universiteit	83
5.1 Evaluatie bij Saxion	84
5.1.1 Methode	84
5.1.2 Resultaten	85
5.1.3 Conclusies	86
5.2 Test bij de Nederlandse Defensie Academie	87
5.2.1 Methode	88
5.2.2 Resultaten	89
5.2.3 Conclusies	91
5.3 Test bij de Open Universiteit	93
5.3.1 Methode	93
5.3.2 Resultaten	95
5.3.3 Conclusies	97
5.4 Algehele conclusies en discussie	97
5.4.1 Praktische bruikbaarheid van de scoringsrubriek	98
5.4.2 Interbeoordelaarsbetrouwbaarheid	98
5.4.3 Validiteit	99
6 Testen van betrouwbaarheid en validiteit bij de opleiding Mediastudies van de Universiteit van Amsterdam	107
6.1 Inhoud en vorm van de gebruikte studieopdracht	109
6.2 Vergelijking van de scoringsrubriek en het UvA-beoordelingsformulier	110
6.3 Methode	115
6.4 Kwantitatieve resultaten	116
6.5 Kwalitatieve resultaten	122
6.6 Conclusies	124
6.7 Discussie	126
7 Evalueren met studenten uit de deeltijdopleiding idm van De Haagse Hogeschool	135
7.1 Literatuur	136

7.2	Enquête	139
7.3	Focusgroepsgesprek	140
7.3.1	Daadwerkelijk gebruik van de scoringsrubriek	143
7.3.2	Potentieel gebruik van de scoringsrubriek	145
7.4	Conclusies en discussie	147
8	Evalueren met docenten van De Haagse Hogeschool	151
8.1	Literatuur	152
8.2	Methode	155
8.3	Resultaten	156
8.4	Conclusie en discussie	158
9	Algemene conclusies en discussie	163
9.1	Resultaten en conclusies van het onderzoek	164
9.2	Reflectie op de onderzoeksaanpak	170
9.3	Reikwijdte van het onderzoek	171
9.3.1	Verhouding tussen het IPS-model en de Scoringsrubriek voor Informatievaardigheden	171
9.3.2	Beoordeling van het proces van ‘information seeking’	172
9.3.3	De Scoringsrubriek voor Informatievaardigheden en het Persoonlijk Kennissysteem	173
9.4	Additionele inzichten	173
9.4.1	Transfer van informatievaardigheden	173
9.4.2	Afstemming van de scoringsrubriek op specifieke disciplines	174
9.4.3	Gebruik van de scoringsrubriek voor onderwijsontwikkeling	176
9.5	Aanbevelingen voor het onderwijs	177
9.6	Eindwaardering en suggesties voor nader onderzoek	179
9.6.1	Onderzoek naar het beoordelen van het ‘Information Problem Solving’- proces	179
9.6.2	Informatievaardigheden van docenten	180
9.6.3	Digitale versie van de scoringsrubriek	181
	Referenties	185
	Samenvatting	199
	Summary	209
	Lijst van bijlagen	218
	Lijst van figuren	219
	Lijst van tabellen	220

Voorwoord

Dit voorwoord wil ik graag beginnen met een metafoor: 'Een proefschrift schrijven, is fietsen op de Mont Ventoux'. Dat deze vergelijking om minstens twee verschillende redenen opgaat, zal ik verderop toelichten. Het beeld van de fietstocht op de Mont Ventoux is echter ook passend omdat daarmee bovendien wordt verwezen naar een dichtbundel uit de jaren '70 waarvan de auteur Jan Kal in het dagelijkse leven werkzaam was als bibliotheekmedewerker. En het onderwerp van mijn proefschrift kan juist vanuit deze beroepsgroep rekenen op een warme belangstelling.

De eerste reden waarom de vergelijking tussen het schrijven van een proefschrift en fietsen op de Mont Ventoux opgaat, wordt door Jan Kal in zijn titelgedicht expliciet genoemd. In de derde regel van het sextet staat: 'bezint eer gij begint'. Ik heb zelf in beide situaties aan den lijve ondervonden hoe moeilijk het is om vooraf in te schatten welke hindernissen je zult tegenkomen. En zoals het beklimmen van de Mont Ventoux toch heus wat anders is dan 'drie keer de Cauberg', zo was de ervaring die ik had opgedaan met onderzoek en publiceren nog geen goede indicatie voor het werk dat gemoeid is met een promotieonderzoek. Mijn eerste poging tot het schrijven van een proefschrift moest ik zelf dan ook – precies als mijn eerste beklimming van de Mont Ventoux – voortijdig staken.

Maar gelukkig doet zich altijd wel een tweede gelegenheid voor en geldt zeker anno 2016 dat je er niet alleen voor staat als je begint aan de klim van de Mont Ventoux. En daarmee kom ik op de tweede reden waarom de metafoor zo passend is. Op deze plek wil ik allen heel hartelijk danken die mij gedurende het onderzoeksproject hebben geïnspireerd, gemotiveerd en met wie ik heb mogen samenwerken. Een paar mensen wil ik uitdrukkelijk noemen.

Als eerste geldt hier zeker een woord van dank aan Albert Boekhorst. Zonder hem zou ik aan dit onderzoek nooit zijn begonnen. Hij was in Nederland de eerste die aandacht voor het onderwerp Informatievaardigheden vroeg en toen ik meer dan tien jaar geleden bij hem aanklopte met de vraag of hij iets zag in een onderzoek naar het beoordelen van informatievaardigheden, stond hij direct klaar om me op weg te helpen. Jammer genoeg was hij niet in de gelegenheid om mijn onderzoek persoonlijk te blijven begeleiden, maar zonder Albert had ik misschien wel een proefschrift geschreven doch zeker niet dit proefschrift. Mijn oprechte dank gaat uit naar hem voor de ideeën die hij met mij deelde en zijn oprechte belangstelling als we elkaar weer eens op een internationale conferentie troffen.

Hij was het ook die mij voorstelde om het eerste resultaat van mijn onderzoek (een scoringsrubriek voor informatievaardigheden en het verhaal over de constructie daarvan) in te dienen als

research paper bij het Journal of Information Literacy. Voordat het zover was, moest er nog wel veel onderzoek worden gedaan naar het functioneren van het beoordelingsinstrument. Juist toen bleek hoe groot het belang is van samenwerking bij het doen van wetenschappelijk onderzoek. Ik bedank alle collega's van De Haagse Hogeschool die tijdens het ontwerpproces inhoudelijk feedback hebben gegeven op de verschillende concepten. Zij waren niet de enigen maar de bijdragen van Jos van Dijk en Loes Tromp droegen zeker bij aan het succes van het eindproduct. Caroline Timmers en haar collega's van Saxion, en Maarten van Veen en Peter Jongejan van de Nederlandse Defensie Academie worden geprezen om hun collegiale inzet bij de evaluatie van het beoordelingsinstrument. Ook zonder de hulp van Ron Oostdam, toen nog werkzaam bij het SCO-Kohnstamm Instituut van de Universiteit van Amsterdam, was de publicatie nooit tot stand gekomen.

Het is niet gebleven bij dit artikel in het Journal of Information Literacy. Geïnspireerd door de internationale contacten die ik opdeed en vooral door de welgemeende interesse die ik ondervond van Serap Kurbanoglu van Hacettepe Universiteit (Ankara) en Sonja Špiranec van de Universiteit van Zagreb, bouwde ik erop voort. Studenten van de opleiding IDM van De Haagse Hogeschool en collega's van verschillende andere faculteiten waren een vruchtbare informatiebron. De resultaten van mijn onderzoek kon ik vooral goed kwijt op de conferenties van ECIL die vanaf 2013 werden georganiseerd. Helemaal aan het einde van het traject waren Mashya Boon en Halbe Kuipers de ideale waterdragers die mij de laatste, onontbeerlijke data hebben aangereikt. Duizend maal dank Mashya en Halbe voor het werk dat jullie hebben verzet.

Dat ik er plezier in had om mee te draaien in het internationale onderzoek op het gebied van informatiewetenschappen, bleef niet onopgemerkt door Frank Huysmans. Hij vroeg mij, in zijn functie als bijzonder hoogleraar Bibliotheekwetenschap aan de Universiteit van Amsterdam, of ik er niet voor voelde om onder zijn begeleiding een tweede poging te doen om een proefschrift te schrijven. Na de eerste verkenningen werden we het eens over een planning en werd het promotietraject in gang gezet. Het begeleidingsteam werd uitgebreid met professor Saskia Brand-Gruwel van de Open Universiteit en een copromotor vanuit mijn werkgever De Haagse Hogeschool, lector Ellen Sjoer. Met veel genoegen kijk ik terug op de bijeenkomsten die wij met vieren hadden.

Saskia wil ik uitdrukkelijk bedanken om haar onderwijskundige adviezen, haar welgemeende e-mails die zeker als aanmoediging hebben gefunctioneerd en de plezierige manier waarop zij mij ook betrok bij activiteiten van haar onderzoeksgroep in Heerlen, zoals het congres van de International Association of School Librarianship in juni en juli 2015 in Maastricht. Zonder Ellen had het proefschrift zeker minder diepgang gekregen. Zij heeft mij aangezet tot een extra empirisch onderzoek met resultaten waar ik bijzonder trots op ben. Dank Ellen dat je al die tijd in mijn werk bent blijven geloven en het beste bij mij naar boven haalde.

Frank ten slotte bleek de meest aimabele begeleider te zijn die ik me had kunnen voorstellen. Hij heeft veel geduld met me gehad, wist volgens mij precies waar mijn zwakke plekken lagen maar wist me, zonder me voor het hoofd te stoten, ook steeds de juiste weg op te sturen. Ik ben erg blij Frank dat ik het geluk heb gehad om met jou te kunnen samenwerken en ik hoop van harte dat dat met de verdediging van dit proefschrift niet afgelopen is.

Behalve door de mensen die mij input hebben gegeven, hen die mij hebben gecoacht en de mannen en vrouwen met wie ik bij wijze van spreken ben opgefietst – zoals alle leden van de kenniskring van het lectoraat Duurzame Talentontwikkeling – ben ik gedurende vele jaren gesteund door sponsors en volgers, zoals dat in welertermen heet. Ik dank Gert de Ruiter, directeur van de faculteit IT & Design en Willy van der Kwaak, tot voor kort teamleider van de opleiding IDM, voor de kansen die ze me hebben geboden en het vertrouwen dat ze al die tijd in me hebben gehad. Ook zonder de steun van Ineke van der Meule van het Centrum voor Lectoraten en Onderzoek van De Haagse Hogeschool en die van De Stichting GO Fonds was dit onderzoek nooit tot een goed einde gekomen. Tot mijn belangrijkste volgers behoorden de vele collega's van De Haagse Hogeschool van wie ik er twee in het bijzonder noem: Peter Becker en de eerder genoemde Jos van Dijk. Met twee zulke informatieprofessionals als paranimf naast me kan het bij de verdediging van het proefschrift niet misgaan.

Aan het einde van de rit wijd ik nog een gedachte aan het thuisfront. Het lijkt een gemeenplaats te zijn als je bij het voltooien van een proefschrift opmerkt dat je partner er veel voor over heeft gehad om jou deze gelegenheid te bieden. Toch is dat ook in mijn geval niet onwaar. Daar staat ter geruststelling tegenover dat Wilma en ik uit het zelfde hout gesneden lijken te zijn en dat we het juist in elkaar waarderen dat we zo gek zijn op het doen van onderzoek. Ze gaf me, met andere woorden, niet alleen de gelegenheid maar heeft ook steeds als voorbeeld gediend. Een betere steun en toeverlaat had ik me niet kunnen wensen.

1

Probleemstelling en context van het onderzoek

Abstract: Informatievaardigheden worden gedefinieerd als een set van deelvaardigheden die nodig zijn bij het gebruik van informatie. Ze zijn belangrijk voor het functioneren in een samenleving waarin kennis en informatie een steeds belangrijker economische rol spelen. In het hoger onderwijs wordt de laatste jaren bovendien meer en meer gebruik gemaakt van werkvormen waarbij óók een beroep op informatievaardigheden wordt gedaan. Uit eerder onderzoek blijkt echter dat jongeren niet vanuit zichzelf al over dergelijke vaardigheden beschikken.

Bij het opleiden van toekomstige professionals en betrokken burgers dient de verwerving van informatievaardigheden aandacht te krijgen. Een belangrijk principe bij het ontwerpen van onderwijs, is dat het beoordelingsproces in overeenstemming wordt gebracht met de gewenste leeruitkomsten en leeractiviteiten. In dit proefschrift is de vraag aan de orde hoe de beoordeling van informatievaardigheden in het hoger onderwijs eruit zou kunnen zien.

Inleiding

Informatievaardigheden (in het Engels meestal aangeduid als “Information Literacy”)¹ staan internationaal in de belangstelling sinds eind jaren ‘80 (Bruce, 2000) en in het Nederlandse taalgebied sinds circa 2000. De meeste auteurs zijn het erover eens dat het begrip betrekking heeft op iemands bekwaamheid in het gebruik van informatie, waarbij de volgende deelvaardigheden worden onderscheiden:

- onderkennen van een informatiebehoefte
- formuleren van een zoekvraag
- selecteren en raadplegen van geschikte informatiebronnen (zoals het internet, een digitaal bestand of een bibliotheekcollectie)
- gebruik van informatie- en communicatietechnologie bij het raadplegen van de informatiebronnen
- selecteren en beoordelen van de informatie die is gevonden
- verwerken (analyseren en synthetiseren), hergebruiken en verspreiden van de gevonden informatie

(Boekhorst, 2000, p. 103; Webber & Johnston, 2000, p. 382). Ook een expertmeeting in Praag in september 2003 met vertegenwoordigers uit 23 landen en georganiseerd met steun van de Unesco, spreekt in het kader van Information Literacy over deze zes deelvaardigheden (Unesco, 2003).

Definities gaan vaak over een nadere invulling van bovenstaande activiteiten waarbij opvalt dat ze zich niet zozeer richten op een definitie van de vaardigheid zelf maar op de definiëring van een ‘informatievaardige persoon’ (zie bijvoorbeeld Boekhorst, 2000, p. 103: “Iemand is informatievaardig indien deze in staat is...”). Informatievaardigheid wordt met andere woorden beschouwd als een kwaliteit van een persoon, wat overigens niet wil zeggen dat die kwaliteit een vaststaand gegeven is.

Het concept informatievaardigheden wordt vaak gehanteerd binnen bibliotheken die opereren in de context van onderwijs en onderzoek maar het reikt verder dan bibliotheek- of ict-vaardigheden, vooral omdat ook het informatiegebruik er deel van uitmaakt (Johnston & Webber, 2003, pp. 335-336)². Wie zijn oor te luisteren legt in het onderwijs zal van docenten te horen krijgen dat

1 In dit onderzoek wordt aangesloten bij de opvatting van Boekhorst om voor het Amerikaanse begrip ‘information literacy’ in het Nederlands de term ‘informatievaardigheden’ te hanteren (Boekhorst, 2000, p. 97). Ondanks het feit dat ‘information literacy’ meer dan alleen maar een paar vaardigheden lijkt te behelzen (in hoofdstuk 2 zal worden stilgestaan bij het feit dat het gaat om een set vaardigheden die geïntegreerd en gecoördineerd moet worden ingezet), is de term informatievaardigheden sindsdien de meest gangbare aanduiding in het Nederlands voor het betreffende fenomeen. Overigens heeft de zelfde discussie over de aanduiding van het verschijnsel ook in het Verenigd Koninkrijk gespeeld. Ook daar is door de taskforce van de Society of College, National and University Libraries (SCONUL) gekozen voor de enigszins beperkte aanduiding ‘information skills’ (Johnson, 2003, p. 48).

2 Het onderscheid met mediawijsheid of ‘media literacy’ is volgens Lee en So (2014, p. 143) dat het laatste begrip vooral betrekking heeft op het kritische gebruik en de analyse van massamedia. De Nederlandse netwerkorganisatie Mediawijzer.net rekent ook het begrip en het gebruik van nieuwe media (internettoepassingen, sms) tot het terrein van mediawijsheid (Mediawijzer.net). In beide gevallen is het onderscheid met informatievaardigheden dat die laatste zich ook (maar niet uitsluitend) richten op de wetenschappelijke en professionele communicatiekanalen.

scholieren en studenten juist op dát gebied vaak oppervlakkig te werk gaan. Ook in het pre-internet tijdperk gold wellicht dat studenten informatie uit boeken en tijdschriften wat al te gemakkelijk overnamen in hun eigen stukken, de alomtegenwoordigheid van het internet heeft ervoor gezorgd dat dit op een veel grotere schaal gebeurt. Veel werkstukken bestaan voor een groot deel uit een compilatie van gekopieerde stukken van het internet en bronverwijzingen zijn slordig (als ze al aanwezig zijn). Ook blijkt dat het voor veel studenten erg lastig is om op basis van de opgespoorde informatie een vraagstuk vanuit een zinvol en eigen perspectief te behandelen. De manier waarop studenten informatie uit internetbronnen gebruiken, heeft dan ook nogal eens het karakter van ‘copy-paste’ terwijl de cognitieve verwerking van de gevonden gegevens ontbreekt (Roth, 1999; Decoo, 2003; Tweede Fase Adviespunt, 2005, p. 90). En juist die cognitieve verwerking is het doel waarvoor studieopdrachten of ‘assignments’ in het onderwijs worden ingezet.

1.1 Relevantie van Informatievaardigheden

1.1.1 Informatiesamenleving

De groeiende aandacht voor informatievaardigheden aan het einde van de jaren ‘80 is nauw verweven met het ontstaan van een post-industriële samenleving waarin de neerslag van wetenschappelijke informatie (“codifications of theoretical knowledge”, Bell, 1976b, p. 46) een steeds belangrijker economische rol is gaan spelen. Dit in tegenstelling tot de daaraan voorafgaande industriële samenleving waarin de productie van tastbare, fysieke goederen het economische primaat had.

Bell wees al vroeg op het ontstaan van zo’n post-industriële samenleving maar merkte ook op dat deze de industriële samenleving niet volledig zou vervangen. Voor de productie van fysieke goederen zouden wel steeds minder arbeidskrachten nodig zijn (Bell, 1976a, p. 578). Mensen verdienen inmiddels inderdaad steeds vaker hun inkomen met activiteiten in de sfeer van zakelijke dienstverlening, persoonlijke dienstverlening - zoals onderwijs en de gezondheidszorg - en in onderzoek en advisering (zie bijvoorbeeld Centraal Bureau voor de Statistiek, 2012). Volgens Webster is die verschuiving van beroepsactiviteiten één van de criteria die in de vakliteratuur wordt gebruikt voor het onderscheiden van een nieuw soort samenleving (2014, p. 10 en pp. 17-19). Andere criteria die worden gebruikt zijn volgens hem:

- Het voortschrijdend gebruik van nieuwe technologieën, zowel in privé huishoudens als in bedrijven en andere professionele organisaties. Onder nieuwe technologieën rekent hij dan zowel apparaten als softwaretoepassingen en services op het internet.
- Macro-economische kengetallen zoals de bijdrage van een bedrijfstak (bijvoorbeeld zakelijke dienstverlening) aan het bruto binnenlands product.
- Het gebruik van netwerken die het mogelijk maken dat ruimtelijke afstanden worden overbrugd.
- Het toenemende *aanbod aan informatie* via verschillende media en over verschillende kanalen. Als voorbeelden noemt Webster de grote verscheidenheid aan radio- en televisiezenders maar ook de alomtegenwoordigheid van reclame-uitingen en de boodschappen die mensen zelf uitzenden door bijvoorbeeld de kleding die ze dragen.

Hoewel men al deze criteria in de literatuur tegenkomt, is er volgens Webster bij geen van deze kwantitatieve criteria aanleiding om te spreken van een werkelijke breuk met de voorgaande maatschappijvorm, de industriële samenleving. Hij benadrukt dan ook dat er sprake is van een geleidelijke overgang in plaats van een radicale verandering (Webster, 2014, pp. 340-357). Wel herkent hij zich in de observatie van Bell dat het belang van theoretische, abstracte informatie vanaf de tweede helft van de twintigste eeuw is toegenomen. Het toenemende belang van theoretische informatie geldt volgens hem echter niet alleen voor beroepsactiviteiten maar ook voor het leven dat burgers privé leiden. De maatschappelijke omgeving verwacht steeds vaker dat individuen goed geïnformeerd zijn en zelfstandig beslissingen nemen (Webster, 2014, pp. 34-37 en pp. 144-145). Hij heeft het dan ook niet over een post-industriële samenleving maar, net als Bell in een later stadium, over een *informatiesamenleving*. Anderen spreken echter liever over een kennismaatschappij (Drucker, 2001b) of juist over een netwerkmaatschappij (Castells, 1996; Van Dijk, 2001). Met dat laatste wordt dan vooral benadrukt dat individuen steeds meer met elkaar in verbinding staan, onder andere via internet (Van Dijk, 2011).

Drucker publiceerde in *The Economist* in november 2001 een serie artikelen over de impact die de dominantie van kenniswerk heeft op de organisatie van het werk zelf en die van de werkplekken (Drucker 2001a; Drucker 2001b). Net als Bell wijst Drucker op het feit dat de nieuwe professionals beter presteren als ze de gelegenheid hebben om zelfstandig beslissingen te nemen en minder dan hun industriële voorgangers in een keurslijf worden gedrongen.

Eveneens in de lijn van het eerdere betoog van Bell, wijst Boekhorst (refererend aan de socioloog Castells) erop dat informatie ook al in eerdere maatschappijtypen (de industriële samenleving en de pre-industriële agrarische samenleving) een rol speelde maar dat pas in de informatiesamenleving de productie, de verwerking en de verspreiding van informatie de overheersende factoren in de economie zijn geworden (Boekhorst, 2000, p. 64; maar zie ook Hamelink, 1989, p. 16). Informatie is de laatste dertig jaar dus, in economische zin, een steeds belangrijker productiefactor geworden maar speelt ook een belangrijke rol bij het functioneren van individuen als volwaardige burgers. De kwantitatieve groei van informatiegebruik betekent echter niet dat ook de kwaliteit van het gebruik is toegenomen. Er zijn aanwijzingen dat juist het tegenovergestelde aan de orde is (Webster, 2014, pp. 32-34) terwijl er wel steeds hogere eisen aan werknemers en burgers worden gesteld (Castells, 2010, pp. 240-241). Van het hoger onderwijs wordt gevraagd om daar op in te spelen (Allen & Van der Velden, 2011).

De laatste jaren worden de kennis en de vaardigheden die zijn vereist in de moderne economie ook wel aangeduid als “21st century skills” (Van der Velden, 2011). In veel van de modellen die daar betrekking op hebben, worden informatievaardigheden uitdrukkelijk genoemd als onderdeel van de “applied ICT literacy” die weer deel uitmaken van de 21st century skills zelf (Voogt en Pareja Roblin, 2010, p. 19 ff.).

1.1.2 Web 2.0

Digitale technologieën en vooral ook het internet en de sociale media hebben er voor gezorgd dat zowel de toegang tot informatie als de productie en distributie ervan sterk zijn gedemocratiseerd

(Anderson, 2006, p. 58 ff. en p. 179; De Haan & Huysmans, 2006, p. 93 en pp. 95-96). Het is vooral de ontwikkeling van het world wide web tot web 2.0, met als kenmerken een intensieve gebruikersparticipatie en een open karakter, dat daaraan heeft bijgedragen. Of zoals Shirky het formuleert: de technologie die tot dan toe werd gebruikt voor persoonlijke doeleinden (en dan vooral voor consumptie) kreeg een sociaal karakter en maakte het bovendien ook voor willekeurige burgers mogelijk om informatie te produceren en met elkaar te delen (Shirky, 2010, p. 14 ff.). Voor velen van hen is het daarbij niet van belang of ze er voor betaald worden: ze dragen vooral bij aan een platform omdat ze daar plezier aan beleven en om bij de betreffende gemeenschap te horen (Shirky, 2010, pp. 56-60).

Toegepast op het hoger onderwijs, kan men stellen dat deze ontwikkeling van het sociale internet er voor gezorgd heeft dat de informatiebronnen die studenten ter beschikking staan om de vraagstukken die ze voorgelegd krijgen op te lossen, inderdaad niet langer alleen bestaan uit de traditionele handboeken en de wetenschappelijke literatuur maar ook uit gratis internetbronnen afkomstig uit het meer informele circuit zoals Wikipedia, blogposts (vaak ook van serieuze experts), forums en frequently asked questions. Ook biedt het web 2.0 studenten de mogelijkheid om dergelijke opdrachten in gezamenlijke projecten te onderzoeken en de resultaten daarvan in een handomdraai te publiceren (Alexander, 2006).

De invloed van het sociale karakter van het world wide web reikt echter verder. Steeds vaker komt het voor dat ook bedrijven en professionals hun kennis en informatieproducten in hun netwerken 'delen' in plaats van dat zij deze zorgvuldig 'bewaken'. Succesvolle bedrijven plaatsen hun diensten op het openbare web en binden op die manier klanten aan zich die vervolgens zorgen voor inkomsten door de afname van aanvullende diensten of advertentie-inkomsten. Popmuzikanten publiceren clips met hun muziek op een videokanaal als Youtube met het doel op die manier extra publiek naar hun concerten te trekken. Als laatste voorbeeld kunnen onderzoekers worden genoemd die gratis downloads van hun tijdschriftartikelen, conferentiepapers en analyses in repositories plaatsen, óók met het doel hun impact en reputatie te vergroten (Anderson, 2006, p. 74).

Kelly sprak in dat verband in 1998 al over "New rules for the new economy" (Kelly, 1998). Deze nieuwe economische wetten zorgen voor een exponentiële groei van het informatieaanbod, die zich zowel manifesteert in een toename van het aantal informatieproducten en -diensten zelf (bijvoorbeeld door het aanbieden van verschillende varianten van één en het zelfde basisproduct) als in de verscheidenheid aan distributiekkanalen die worden gebruikt voor de verspreiding van dat aanbod (Anderson, 2006, pp. 219-220). Daarbij geldt tevens dat traditionele poortwachters zoals uitgeverij, boekhandel en bibliotheek niet langer het alleenrecht op die distributie kunnen claimen. Van der Weel spreekt in dat verband over het verschijnsel van "disintermediatie" (2012, p. 31). Auteurs publiceren rechtstreeks op het internet of via sites als Lulu.com en xlibris.com, verkopen hun boeken zelf via internetboekhandels en attenderen elkaar op nieuwe informatiebronnen via Twitter, Facebook en andere online media.

Al met al kan de conclusie zijn dat men een steeds grotere verscheidenheid aan media-uitingen tegenkomt, zowel van professionals als van niet-professionals / amateurs. Om in die overvloed van het informatieaanbod de weg te vinden en onderscheid te maken tussen zinvolle en zinloze infor-

matie, dienen mensen te beschikken over informatievaardigheden en dan vooral over het vermogen om de relevantie en betrouwbaarheid van informatie te beoordelen (zie bijvoorbeeld Walraven, Brand-Gruwel & Boshuizen, 2009). Meer recent wordt benadrukt dat ook het vormgeven van een eigen informatieomgeving daarbij van belang wordt (Sieverts, 2013). Te denken valt daarbij aan het gebruik van een Personal Library, deelname aan een 'Virtueel Kennis Centrum' en het gebruik van Personal Information Management (Liefsoens, 2013).

Betrouwbaarheid van informatie speelt ook nog op een andere manier een rol, namelijk in de betekenis van de 'veiligheid' van informatie en informatiebronnen. De ruime beschikbaarheid van internet heeft er ook voor gezorgd dat gebruikers geconfronteerd worden met verschillende vormen van fraude - zoals het ontfutselen van wachtwoorden en het inbreken op iemands bankrekening - en met andere internetgebruikers die iemands privacy schenden op een van de sociale netwerken zoals Facebook of Pinterest. De Europese commissie ondersteunt ook in Nederland programma's die tot doel hebben om kinderen en jongeren te beschermen tegen dergelijke vormen van oplichting en pesterijen (European Commission, 2015). Daarbij kan het zowel gaan om het aanbieden van een veilige internetomgeving als om het bevorderen van kennis over dit soort zaken. Voorlichting en lesprogramma's over veilig internetten worden over het algemeen echter niet gerekend tot het thema 'informatievaardigheden' maar tot 'mediawijsheid' (Mediawijzer.net).

1.1.3 Een leven lang leren

Een derde ontwikkeling die de relevantie van informatievaardigheden onderstreept, is het gegeven dat internet en ICT er ook voor hebben gezorgd dat er op de werkplek steeds meer belang wordt gehecht aan *generieke* vaardigheden. Een van de belangrijkste daarvan is de vaardigheid om te 'leren' omdat veranderingen in het eigen vak zo snel plaatsvinden dat beroepskennis die is opgedaan op school snel veroudert (Drucker, 2001a; De Haan & Huysmans, 2006, p. 104). Het belang van het 'leren leren' in het onderwijs werd recent nog eens benadrukt in een rapport van de Wetenschappelijke Raad voor het Regeringsbeleid (2013, p. 272) en in de eerder genoemde discussies rond "21st century skills".

Eveneens binnen de Nederlandse context spreekt Thijssen (2001) in dat verband over het toenemende belang van "leercompetenties" en "exploratieve informatie-verwerking" binnen arbeidssituaties. Met dat laatste doelt hij op het "adequaat gebruik maken van informatie-aanbod in (schriftelijke of elektronische) informatiebronnen om kennis te exploreren" (Thijssen, 2001, p. 24). Het is duidelijk dat het door hem gehanteerde begrip exploratieve informatieverwerking veel overlap vertoont met wat hierboven "informatievaardigheden" zijn genoemd.

Lifelong Learning was ook de naam van een programma van de Europese Commissie waarmee onder andere het innovatieve gebruik van ict, voor bijvoorbeeld e-learning en mobiele apparaten in het onderwijs, werd gestimuleerd (European Commission, 2014; Drigas & Tsolaki, 2015). Aan openbare bibliotheken wordt een belangrijke rol toegedicht bij het toegankelijk houden van actuele ict-toepassingen en het bieden van ondersteuning bij het gebruik daarvan ten behoeve van 'informeel leren' (Drigas & Tsolaki, 2015). Ook met betrekking tot dit thema gaat de bemoei-

enis van de Europese Commissie echter meer in de richting van mediawijsheid dan in die van informatievaardigheid.

1.1.4 Ontwikkelingen in het onderwijs

Informatievaardigheden zijn niet alleen relevant vanwege maatschappelijke en technologische ontwikkelingen, ook veranderingen in het onderwijs zelf hebben ervoor gezorgd dat het belang ervan is toegenomen. Ongeveer in dezelfde periode dat West-Europa transformeerde van een industriële samenleving naar een informatiesamenleving vond in het hoger onderwijs (en zeker ook in Nederland) een transitie plaats van een leeromgeving die vooral gebaseerd was op instructie naar een omgeving die de constructie van kennis door de lerenden zelf centraal stelt (zie bijvoorbeeld Biggs, 1996, p. 348).³ Kenmerkend voor de constructivistische benadering van leren is onder andere dat er gebruik wordt gemaakt van informatierijke leeromgevingen waarin studenten actief werken aan authentieke problemen of vraagstukken met een complex en samengesteld karakter. Waar bij het instructiemodel het accent ligt op het leren oplossen van discrete, los van elkaar staande problemen, legt de constructivistische onderwijsbenadering de nadruk op een authentieke context die ervoor zorgt dat studenten zichzelf trainen in het oplossen van complexe problemen waarbij zij de aandacht op verschillende aspecten tegelijk dienen te richten (Elshout-Mohr, Oostdam & Overmaat, 2002, p. 372).

Deze didactische vernieuwingen staan bekend onder namen als probleemgestuurd onderwijs, projectonderwijs en competentiegericht onderwijs (Van Merriënboer, Van der Klink & Hendriks, 2002, p. 24) en worden onder andere gekenmerkt door het feit dat van studenten verwacht wordt dat zij zelf op zoek gaan naar informatie om de voorgelegde problemen op te lossen en deze informatie te combineren met reeds aanwezige kennis (Brand-Gruwel, Wopereis & Vermetten, 2005). De vaardigheid in het vergaren van informatie maar vooral ook die in het definiëren van het informatieprobleem en het selecteren, beoordelen en verwerken van de informatie, kunnen binnen deze onderwijsconcepten niet gemist worden. Dat studenten deze vaardigheden uit zichzelf wel zullen verwerven en wellicht al verworven hebben, is een vaak voorkomend misverstand (Kolowich, 2011; Thompson, 2011; Kirschner & Van den Berg, 2012). Het aanleren van informatievaardigheden op een conceptueel niveau waarbij de nadruk op de informatie zelf ligt in plaats van op ict, is cruciaal om succesvol te studeren binnen het hoger onderwijs. Informatievaardigheid is voor studenten dan ook niet alleen van belang om binnen hun toekomstige arbeidscontext te kunnen blijven leren (het argument van Thijssen dat genoemd werd in paragraaf 1.1.3), het is tevens een opleidingscompetentie, een vermogen waarover studenten dienen te beschikken om hun initiële studie überhaupt te kunnen afronden (Elshout-Mohr & Oostdam, 2001, p. 2).

1.1.5 Rol van onderzoek in het hoger onderwijs

Aandacht voor informatievaardigheden speelt ook een rol in de discussie over onderzoek in het hoger onderwijs - en meer specifiek in het hbo - die sinds het einde van de 20^e eeuw in Nederland wordt gevoerd. In tegenstelling tot veel andere Europese landen kent het Nederlandse hoger onderwijs een

3 Elshout-Mohr, Oostdam en Overmaat (2002, p. 369) dateren de introductie van het concept constructivisme in het onderwijs in de jaren '80.

binaire structuur waarbij onderscheid wordt gemaakt tussen Universities of Applied Sciences (hogescholen) en Research Universities (UASnet, z.d.). De kerntaak van de hogescholen is weliswaar het opleiden van professionals voor de beroepspraktijk (waarmee zij zich onderscheiden van de Research Universities) maar de wet op het hoger onderwijs en wetenschappelijk onderzoek uit 1993 gaf hun ook de mogelijkheid tot het verrichten van onderzoeksactiviteiten voor zover die verband houden met het onderwijs aan die betreffende instelling (Borgdorff, Van Staa & Van der Vos, 2007).

Na de ondertekening van het Bologna Verdrag (1999) en de invoering van de bachelor master-structuur in het hoger onderwijs in september 2002 (UASnet, z.d.) wordt van die onderzoeksfunctie in het hoger beroepsonderwijs uitdrukkelijk meer werk gemaakt. Sinds 2001 bevordert de Nederlandse overheid praktijkgericht onderzoek in het hbo, onder andere door de mogelijkheid om aan de hogescholen lectoren te benoemen (UASnet, z.d.). HBO-onderzoek wordt vooral geacht bij te dragen aan verbetering en vernieuwing van de beroepspraktijk door middel van “ontwerp en ontwikkeling” en kenmerkt zich daardoor als “vraaggericht” (Adviesraad voor het Wetenschaps- en Technologiebeleid, 2005). De kerntaak van het hoger beroepsonderwijs is volgens de Adviesraad weliswaar het opleiden van goed geschoold personeel maar daar hoort (volgens die zelfde Adviesraad) ook bij dat dit hoog opgeleide personeel in staat is om zelf onderzoek uit te voeren. De beroepspraktijk wordt steeds complexer en de AWT meent dat afgestudeerden aan het hbo meer dan in het verleden in de praktijk taken uitvoeren die men kan aanduiden als onderzoek: “situaties doorlichten, analyseren in het licht van bekende kennis, afwegingen maken over te volgen aanpak en dergelijke” (p. 14). Onderzoek door studenten in het hbo dient volgens de AWT dan ook minder de functie te hebben van het oplossen van de praktijkproblemen zelf (daarvoor bestaan ook andere organisaties zoals ingenieursbureaus en adviesbureaus) maar zich vooral te richten op het uiteindelijke opleidingsdoel van hogescholen, het aanleren van de benodigde vaardigheden bij studenten. De vaardigheid in het doen van onderzoek maakt daar deel van uit. Feitelijke oplossingen die worden ontwikkeld zouden volgens de AWT dan vooral moeten worden gezien als praktische “bijvangst”. In termen van Jenkins en Healy is er dan sprake van “research-oriented teaching” (2005, pp. 21-22) .

De opvatting van de AWT met betrekking tot de onderzoekstaak van het hbo is een zeer beperkte en wordt zeker niet door alle andere partijen onderschreven. Zo is in de in paragraaf 1.1.4 besproken vormen van constructivistisch leren een veel grotere didactische rol weggelegd voor het doen van (informatie-)onderzoek. Jenkins en Healy spreken in dat geval over “research-based” of “inquiry-based learning” (2005, p. 22).

Maar ook in de beperkte opvatting van de AWT geldt dat informatievaardigheid een van de kwaliteiten is die iemand nodig heeft voor het doen van onderzoek, naast andere vaardigheden met betrekking tot het verzamelen, analyseren, interpreteren en verwerken van data. Dat geldt eens te meer als tot het begrip ‘onderzoeksvaardigheid’ ook het vermogen van professionals wordt gerekend om hun kennis blijvend te vernieuwen, zoals Borgdorff, Van Staa en Van der Vos (2007) doen.

Een andere onderzoeksvaardigheid die studenten dienen te verwerven, is die in het *lezen* van onderzoek. Deze wordt vaak onderscheiden van de vaardigheid in het doen van onderzoek zelf (zie bijvoorbeeld Keshav, 2007). Feitelijk gaat het om het efficiënt kunnen selecteren en beoordelen van

wetenschappelijke output (op basis van titel, abstract, figuren, tabellen en de referentielijst) en het ontwikkelen van een kritische houding met betrekking tot de inhoud ervan, bijvoorbeeld door op zoek te gaan naar verborgen aannames en methodologische keuzes. Deze vaardigheid in het lezen van onderzoeksliteratuur komt grotendeels overeen met één van de belangrijkste deelactiviteiten die binnen het complexe begrip Informatievaardigheid zijn onderscheiden, te weten de vaardigheid in het selecteren en beoordelen van informatie. Hoewel deze deelvaardigheid zich strikt genomen niet beperkt tot onderzoeksliteratuur, is het duidelijk dat er grote overlap is met wat elders de vaardigheid in het lezen van onderzoeksverslagen en -artikelen wordt genoemd.

1.2 Probleembeschrijving en doelstelling

In het voorgaande is betoogd dat informatievaardigheid een kwaliteit is die door professionals in hun arbeidsomgeving en door studenten gedurende hun studie niet kan worden gemist. Op basis van literatuur werd bovendien geconcludeerd dat het een illusie is te denken dat studenten deze vaardigheid uit zichzelf wel zullen verwerven. Daaruit volgt dat er behoefte is aan onderwijs dat studenten helpt om informatievaardig te worden.

Voor het Nederlandse taalgebied is intussen een aantal methodes voor dergelijk onderwijs in informatievaardigheden beschikbaar, zoals het studieboek van Boekhorst, Kwast en Wevers (2004), dat van Van Veen en Westerkamp (2010) en dat van Brand-Gruwel en Wopereis (2014). De wijze waarop informatievaardigheden kunnen worden *beoordeeld* heeft in die onderwijsmethodes echter weinig aandacht gekregen. In dit onderzoek waarvan deelstudies werden uitgevoerd vanaf 2009, is de focus uitdrukkelijk gericht op dat beoordelingsproces en de wijze waarop ook dát het leren kan bevorderen.

De geringe aandacht in het Nederlandse taalgebied voor het beoordelen van informatievaardigheden is opmerkelijk omdat onderwijsontwikkelaars over het algemeen bepleiten om vanaf het begin van het ontwikkelproces rekening te houden met de wijze waarop de te behalen doelstellingen zullen worden getoetst. Hoewel men bij toetsen in het kader van onderwijs vaak denkt aan de afsluiting van een onderwijseenheid is het bij het ontwerpen van onderwijs een algemeen aanvaard principe dat reeds in de beginfase van het ontwerpproces wordt nagedacht over de leerdoelen die men wil bereiken en de soort beoordelingsinstrumenten waarmee men het bereiken van die leerdoelen wil vaststellen (Straetmans, 2006, p. 7). De gewenste leeruitkomsten, de leeractiviteiten en toetsvormen worden daarvoor op elkaar afgestemd, het idee van “constructive alignment” (Biggs, 2003). Bovendien is uit diverse onderzoeken bekend dat de wijze van beoordelen in sterke mate sturing geeft aan het studiegedrag van studenten (Frederiksen, 1984; Driessen & Van der Vleuten, 2000; Dochy, Segers & De Rijdt, 2002, pp. 11-12).

In dit proefschrift ga ik dan ook op zoek naar een of meer instrumenten die geschikt zijn voor het *beoordelen* van informatievaardigheden in de context van het Nederlandse hoger onderwijs. Uiteraard worden aan zo’n beoordelingsinstrument de gangbare eisen gesteld op het gebied van validiteit, betrouwbaarheid en praktische bruikbaarheid (Van de Watering en Dierick, 2002, p. 63).

De uiteindelijke doelstelling van dit onderzoek (de reden waarom het onderzoek wordt uitgevoerd; Verschuren, 2005, pp. 34-37) is echter breder dan het beoordelen op zich en richt zich op een bijdrage aan de ontwikkeling van de informatievaardigheden van studenten in het hoger onderwijs.

1.3 Eisen aan het beoordelingsinstrument

Zoals hiervoor opgemerkt worden aan beoordelingsinstrumenten van oudsher eisen gesteld op het gebied van validiteit, betrouwbaarheid en praktische bruikbaarheid. In dit onderzoek geldt echter een aantal aanvullende eisen die in de volgende twee paragrafen worden benoemd en toegelicht.

1.3.1 Generieke toepasbaarheid

Het onderwerp informatievaardigheden heeft in de afgelopen vijftien jaren vooral veel aandacht gekregen vanuit de universiteits- en hogeschoolbibliotheken. Daarbij hebben vooral de universiteitsbibliotheken aanvankelijk veel tijd gestoken in het zelf verzorgen van trainingen en instructies maar steeds vaker is daarbij nu ook sprake van inbedding in de verschillende curricula (Dirkx et al., 2013). Deze auteurs geven aan dat dit belangrijk is “omdat de motivatie van studenten aanzienlijk groter is wanneer zij het geleerde direct kunnen toepassen in opdrachten” (p. 4). Brand-Gruwel en Wopereis (2006) concluderen niet alleen dat ingebedde instructie motiverend werkt maar ook dat de leereffecten ervan groter zijn dan bij speciale cursussen die losstaan van de vakinhoud. Zij pleiten ervoor om het oplossen van informatieproblemen aan te leren binnen de context van authentieke opdrachten waar studenten aan werken tijdens hun studie.

Ook informatiespecialisten die werkzaam zijn in hogeschool- en universiteitsbibliotheken lijken behoorlijk wat tijd te besteden aan het geven van instructies (Van Helvoort, 2012a). Er zijn echter geen gegevens bekend over de vraag in hoeverre daarbij sprake is van integratie in een onderwijseenheid of juist van losstaande cursussen die verzorgd worden door de bibliotheek. Om eerder genoemde redenen - motivatie van studenten, een groter leereffect - wordt in deze studie gepleit voor de integratie van informatievaardighedeninstructie in de onderwijseenheden die door de opleidingen zelf worden verzorgd.

Voor het hier te ontwikkelen beoordelingsinstrument betekent dit dat het toegepast zou moeten kunnen worden binnen verschillende vakgebieden. Uit de internationale vakliteratuur over het toetsen van informatievaardigheden waren al een paar cursusgerelateerde beoordelingsinstrumenten bekend maar een generiek beoordelingsinstrument ontbrak. Een pluspunt van een algemeen toepasbaar model zou zijn dat zo'n instrument het mogelijk maakt om groepen studenten en / of opleidingen met elkaar te vergelijken. Bijkomend voordeel is dat opleidingen die nog aan het begin van de ontwikkeling van onderwijs in informatievaardigheden staan, daarmee een instrument wordt aangeboden dat zij na kleine aanpassingen snel kunnen implementeren in hun eigen curriculum. Ook in die zin sluit een algemeen toepasbaar instrument goed aan bij de hierboven geformuleerde doelstelling van dit onderzoek, het leveren van een bijdrage aan de ontwikkeling van de informatievaardigheden van studenten in het hoger onderwijs.

1.3.2 Aansluiting bij actuele ontwikkelingen op het gebied van toetsing en onderwijs

In ieder geval bij de constructivistische onderwijsvormen die in paragraaf 1.1.4 zijn onderscheiden (probleemgestuurd onderwijs, projectonderwijs, competentiegericht onderwijs) geldt dat toetsing niet alleen aan het einde van een onderwijseenheid een rol speelt maar ook tussentijds. De feedback die deze tussentijdse beoordelingen opleveren, wordt geacht de studenten aan te zetten tot dieper leren. Toetsing heeft in de constructivistische leeropvatting dus niet alleen een certificerende (“summatieve”) functie als sluitstuk van het onderwijs maar gedurende het leerproces ook een informerende en motiverende rol. Die laatste vorm van toetsen, die uitdrukkelijk is bedoeld om het leren van de student te bevorderen, wordt in de literatuur ook wel aangeduid als “formatief toetsen” (Nitko & Brookhart, 2007, p. 8 en pp. 107-109). In het geval dat zulke tussentijdse beoordelingen niet alleen een cijfer opleveren (of überhaupt geen cijfer) maar gepaard gaan met uitgebreide inhoudelijke feedback, wordt ook wel gesproken over een “assessment FOR learning” (Stiggins, 2005).

In het geval dat toetsen en beoordelen betrekking hebben op het uitvoeren van een complete (beroeps) taak of een functie - die zelf vaak weer zijn samengesteld uit verschillende deelactiviteiten - wordt in de onderwijskundige literatuur ook wel gesproken over “assessments” (Elshout-Mohr & Oostdam, 2001, p. 3; Dochy & Nickmans 2005, p. 55) of over “performance assessments” (Nitko & Brookhart, 2007, p. 244). Bij informatievaardigheden gaat het eveneens over de uitvoering van zo’n ‘hele taak’. Om informatievaardig te zijn is het immers niet voldoende om te beschikken over afzonderlijke kenniselementen maar moet de student die ook geïntegreerd en gecoördineerd kunnen toepassen. In het vervolg zal daarom ook wel de uitdrukking “assessment van informatievaardigheden” worden gehanteerd. Omdat het begrip ‘assessment’ in de internationale literatuur echter meer betekenissen heeft dan in het Nederlandse taalgebruik, dient dat eerst nader afgebakend te worden.

Radcliff, Jensen, Salem, Burhanna & Gedeon (2007, pp. 11-14) onderscheiden drie niveaus waarop assessments plaatsvinden:

- Het niveau van een cursus of onderwijseenheid (“classroom assessments”). Het object van assessment is in dit geval de prestatie van individuele studenten. Zoals hiervoor al opgemerkt kunnen classroom assessments formatief zijn (informerend; vooral gericht op de voortgang van het leren en op feedback) of summatief (selecterend; voor een cijfer).
- Curriculum. Het assessment is formatief van aard en richt zich op de prestaties van een groep studenten. Doel is te achterhalen op welke aspecten studenten als cohort goed presteren of juist tekort schieten. De resultaten van het assessment kunnen worden gebruikt om na te gaan welke onderdelen van een curriculum in aanmerking komen voor verbetering.
- Instituut. Object van assessment is een of meer cohorten van studenten, verdeeld over verschillende studierichtingen of faculteiten/academies. Ook institutionele assessments zijn formatief van aard en de resultaten kunnen bijvoorbeeld gebruikt worden om de verschillende studierichtingen of faculteiten/academies onderling met elkaar te vergelijken (benchmarking).

Het onderhavige onderzoek is in eerste instantie gericht op de ontwikkeling van een instrument voor “classroom assessments” op het gebied van informatievaardigheden. Assessments op het niveau van een curriculum of een instituut worden niet bij het onderzoek betrokken omdat de scope beperkt blijft tot de bevordering van informatievaardigheden bij individuele studenten. Het

is uiteraard wel zo dat de resultaten van de individuele assessments geaggregeerd kunnen worden tot het niveau van een cohort of instituut en dat die vervolgens gebruikt kunnen worden voor bijvoorbeeld benchmarking. Voorwaarde is dan wel dat het instrument generiek genoeg is om toegepast te kunnen worden op verschillende soorten opdrachten uit verschillende vakgebieden.

Samenvattend wordt gesteld dat het te ontwikkelen toetsinstrument geschikt dient te zijn voor

- het beoordelen van individuele “performances” op het gebied van informatievaardigheden
- ingezet moet kunnen worden binnen verschillende vakgebieden
- en zowel gebruikt kan worden bij summatieve toetsing als bij formatieve toetsing.

De aandacht in het onderzoek voor zowel summatieve als voor formatieve toetsing, heeft ook een onderzoeksmethodologisch aspect. Continue analyse van relevante literatuur over assessments en assessmentinstrumenten in de beginfase van het onderzoek én de reflectie daarop, hebben duidelijk gemaakt dat beoordelen en leren in elkaar grijpen en niet onafhankelijk van elkaar bestudeerd kunnen worden. De formulering van de eisen waaraan het beoordelingsinstrument zou moeten voldoen, is daarmee een duidelijk voorbeeld geweest van het cyclische proces zoals dat volgens Wester en Peters (2004, pp. 83-86) in kwalitatief onderzoek wordt doorlopen. Een ander voorbeeld van zo’n cyclisch proces wordt beschreven in hoofdstuk 4 waarin het ontwerpproces van de scoringsrubriek wordt besproken.

1.4 Onderzoeksvragen en onderzoeksmethodes

Gegeven de probleembeschrijving uit paragraaf 1.2 en de eisen uit paragraaf 1.3 komt de centrale onderzoeksvraag als volgt te luiden:

Wat is een betrouwbaar, valide en praktisch bruikbaar instrument dat generiek toepasbaar is voor het beoordelen én het bevorderen van de informatievaardigheden van individuele studenten in het hoger onderwijs?

Voor beantwoording van deze centrale onderzoeksvraag dient onderzoek te worden gedaan naar verschillende deelvragen. Deze deelvragen zijn de onderzoeksvragen van het proefschrift.

Onderzoeksvraag 1: Wat is de meest geschikte assessmentmethode voor het beoordelen en bevorderen van informatievaardigheden?

In deze eerste onderzoeksvraag komt expliciet naar voren dat een te kiezen methode voor het toetsen en beoordelen van informatievaardigheden ook een functie dient te vervullen bij het bevorderen van het leerproces van studenten. De vraag zal worden beantwoord door middel van literatuurstudie. De in paragraaf 1.3 omschreven eisen die aan het instrument worden gesteld, zullen worden gebruikt om het meest passende antwoord te formuleren.

Het antwoord op onderzoeksvraag 1 zal nog steeds een betrekkelijk algemeen karakter hebben, in die zin dat de uitkomst niet een instrument zelf is maar een beschrijving van een bepaald *type*. In onderzoeksvraag 2 wordt antwoord gegeven op de vraag hoe zo'n instrument (in dit geval een 'scoringsrubriek') er dan concreet uit ziet.

Onderzoeksvraag 2: Hoe ziet een scoringsrubriek eruit die geschikt is voor het beoordelen en bevorderen van informatievaardigheden bij studenten in het hoger onderwijs?

De vraag zal worden beantwoord door het ontwerpen van een concreet beoordelingsinstrument, een scoringsrubriek voor informatievaardigheden. Het te ontwikkelen prototype van de scoringsrubriek wordt vervolgens getest en geëvalueerd aan de hand van twee vervolgvragen.

Onderzoeksvraag 3: Hoe betrouwbaar en valide is de scoringsrubriek?

Voor het bepalen van de betrouwbaarheid en de validiteit van de scoringsrubriek wordt deze getest in drie verschillende situaties. Het gaat daarbij om drie verschillende typen opleidingen van drie verschillende onderwijsinstellingen. In alle drie de gevallen is gebruik gemaakt van bestaande studieopdrachten. De resultaten van de beoordelingen worden steeds geanalyseerd met het pakket SPSS.

Onderzoeksvraag 4: Hoe wordt de scoringsrubriek door studenten en docenten in het hoger onderwijs gebruikt en hoe ervaren zij het gebruik ervan?

Onderzoeksvraag 4 wordt beantwoord aan de hand van verschillende onderzoeken. Gezien de verschillende groepen respondenten en de gebruikte onderzoeksmethodes worden er drie verschillende varianten van de onderzoeksvraag gebruikt.

Onderzoeksvraag 4a: Hoe ervaren docenten in hun rol van beoordelaar het gebruik van de scoringsrubriek?

De onderzoeksmethodes bij deze onderzoeksvraag zijn open interviews en een vragenlijst.

Onderzoeksvraag 4b: Hoe wordt de scoringsrubriek door studenten in het hoger onderwijs gebruikt en hoe ervaren zij het gebruik ervan?

De onderzoeksmethode bij deze onderzoeksvraag is een focusgroepsprek.

Onderzoeksvraag 4c: Hoe wordt de scoringsrubriek door docenten in het hoger onderwijs gebruikt en hoe ervaren zij in het algemeen het gebruik ervan?

Voor de beantwoording van deze vraag worden semi gestructureerde interviews afgenomen.

1.5 Opbouw van het proefschrift

Het vervolg van dit proefschrift is opgesplitst in twee delen. In deel I worden de theoretische uitgangspunten beschreven en wordt toegelicht hoe het ontwerpproces van de scoringsrubriek is verlopen. De theoretische uitgangspunten hebben weer betrekking op enerzijds het conceptueel kader dat behoort bij het begrip ‘informatievaardigheden’ (hoofdstuk 2) en anderzijds op de onderwijskundige achtergronden die behoren bij toetsen en beoordelen (hoofdstuk 3). De beschrijving van de constructie van een scoringsrubriek in hoofdstuk 4 eindigt met de presentatie van een prototype. Dat tussenproduct fungeert tevens als afsluiting van deel I.

In deel II van het proefschrift wordt het prototype getest en geëvalueerd. Hoofdstuk 5 beschrijft drie kleinere onderzoeken bij achtereenvolgens Saxion, de Nederlandse Defensie Academie en de Open Universiteit. Hoofdstuk 6 beschrijft een grootschaliger onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam. Bij al deze onderzoeken ligt de nadruk op het verzamelen van kwantitatieve gegevens. Met uitzondering van het onderzoek bij Saxion hebben deze onderzoeken vooral het karakter van een ‘test’.

De onderzoeken in hoofdstuk 7 en 8 zijn meer kwalitatief van aard en gaan over ervaringen van studenten en docenten in het hoger onderwijs met het gebruik van de scoringsrubriek in de praktijk. Anders dan bij de onderzoeken in hoofdstuk 5 en 6 lag bij de onderzoeken in hoofdstuk 7 en 8 de nadruk dan ook meer op een evaluatie dan op een test.

Na deel II volgt nog hoofdstuk 9 waarin het proefschrift wordt afgesloten met algemene conclusies en een discussie waarin wordt gereflecteerd op de onderzoeksaanpak, de reikwijdte ervan, additionele inzichten, aanbevelingen voor het onderwijs en suggesties voor verder onderzoek in de toekomst. In een schema ziet het hoofdstukoverzicht eruit zoals in tabel 1.1 op de volgende pagina.

Over een aantal van de deelonderzoeken die zijn uitgevoerd, is eerder gerapporteerd in (peer-reviewed) tijdschriftartikelen en conferentiebijdragen. Deze studies vormden de basis van hoofdstuk 2 (Van Helvoort, 2014), hoofdstuk 3 tot en met 5 (Van Helvoort, 2010), hoofdstuk 7 (Van Helvoort, 2012b) en hoofdstuk 8 (Van Helvoort, 2013). Daar waar gebruik is gemaakt van de gegevens uit eerdere publicaties wordt dat aan het begin van het betreffende hoofdstuk of paragraaf in een voetnoot aangegeven.

Dit proefschrift is meer dan een optelsom van de eerdere publicaties in wetenschappelijke tijdschriften en conferentiebundels. Niet alleen werden een inleiding op het onderwerp informatievaardigheden, een aanvullende empirische studie en een afsluitend discussiehoofdstuk toegevoegd, de eerdere publicaties werden ook herschreven om ze beter met elkaar te integreren. Om een heel praktische reden, het plezier dat de onderzoeker heeft in het gebruik van zijn moedertaal, is ervoor gekozen om de publicatie te schrijven in het Nederlands. Dat was met name ook een optie omdat een bijdrage aan de internationale wetenschappelijke discussie over informatievaardigheden ook al geleverd was met de publicatie van de afzonderlijke artikelen en conferentiebijdragen.

Tabel 1.1

Hoofdstukindeling en opbouw van het proefschrift

	Hoofdstuk	Onderwerp	Onderzoeksmethode(n)
DEEL 1	1	Probleembeschrijving, doelstelling en onderzoeksvragen van de studie	• Literatuuranalyse
	2	Conceptueel kader van het construct informatievaardigheden	• Literatuuranalyse
	3	Theoretische achtergronden bij beoordelen in het hoger onderwijs. Onderzoeksvraag 1: Wat is de meest geschikte assessmentmethode voor het beoordelen en bevorderen van informatievaardigheden?	• Literatuuronderzoek en -analyse
	4	Constructie van een scoringsrubriek. Onderzoeksvraag 2: Hoe ziet een scoringsrubriek eruit die geschikt is voor het beoordelen en bevorderen van informatievaardigheden bij studenten in het hoger onderwijs?	• Ontwerpen “Step-by-Step Procedure” voor het ontwerpen van scoringsrubrieken van Mertler (2001)
DEEL 2	5	Testen van de scoringsrubriek. Onderzoeksvraag 3: Hoe betrouwbaar en valide is de scoringsrubriek?	• Betrouwbaarheidsanalyses • onderzoek naar de interne consistentie • onderzoek naar de concurrentvaliditeit
		Evalueren van de scoringsrubriek Onderzoeksvraag 4a: Hoe ervaren docenten in hun rol van beoordelaar het gebruik van de scoringsrubriek?	• Vragenlijst • open interviews
	6	Testen van de scoringsrubriek. Onderzoeksvraag 3: Hoe betrouwbaar en valide is de scoringsrubriek?	• Betrouwbaarheidsanalyse • exploratieve factoranalyse • onderzoek naar de interne consistentie • onderzoek naar de concurrentvaliditeit
		Evalueren van de scoringsrubriek Onderzoeksvraag 4a: Hoe ervaren docenten in hun rol van beoordelaar het gebruik van de scoringsrubriek?	• Open interviews
	7	Evalueren met studenten. Onderzoeksvraag 4b: Hoe wordt de scoringsrubriek door studenten in het hoger onderwijs gebruikt en hoe ervaren zij het gebruik ervan?	• Focusgroepsgesprek
	8	Evalueren met docenten. Onderzoeksvraag 4c: Hoe wordt de scoringsrubriek door docenten in het hoger onderwijs gebruikt en hoe ervaren zij in het algemeen het gebruik ervan?	Individuele semi gestructureerde interviews
	9	Welke conclusies met betrekking tot het onderwijs in informatievaardigheden in het hoger onderwijs kunnen, op basis van de verkregen resultaten, worden getrokken? Welke onderwerpen vergen nader onderzoek?	Reflectie

Deel I

**Theoretische uitgangspunten en
constructie van het beoordelingsmodel**

De onderzoeksvragen van dit proefschrift zijn typisch vragen die behoren bij een 'ontwerpgericht wetenschappelijk onderzoek'. In paragraaf 1.5 werden ze als volgt geformuleerd:

- Onderzoeksvraag 1: Wat is de meest geschikte assessmentmethode voor het beoordelen en bevorderen van informatievaardigheden?*
- Onderzoeksvraag 2: Hoe ziet een scoringsrubriek eruit die geschikt is voor het beoordelen en bevorderen van informatievaardigheden bij studenten in het hoger onderwijs?*
- Onderzoeksvraag 3: Hoe betrouwbaar en valide is de scoringsrubriek?*
- Onderzoeksvraag 4: Hoe wordt de scoringsrubriek door studenten en docenten in het hoger onderwijs gebruikt en hoe ervaren zij het gebruik ervan?*

Volgens Van Aken en Andriessen (2011, p. 16) is kenmerkend voor ontwerpgericht wetenschappelijk onderzoek dat het zich niet beperkt tot het beschrijven en verklaren van problemen uit de omringende werkelijkheid maar dat het zich ook richt op het ontwikkelen en testen van oplossingen daarvoor. Het probleem waarvoor in dit onderzoek een oplossing wordt gezocht is dat er in het hoger onderwijs nog geen algemeen toepasbare methode bestaat voor het beoordelen van informatievaardigheden. In hoofdstuk 1 is met argumenten onderbouwd dat dat daadwerkelijk een probleem is.

In deel I van het proefschrift staat de ontwikkeling van een oplossing voor bovengenoemd probleem centraal. In de hoofdstukken 2 en 3 wordt daarvoor eerst onderzoek gedaan naar twee kernbegrippen uit het onderzoek: 'informatievaardigheden' en 'beoordelen'. Op basis van verschillende modellen en benaderingen die in de literatuur zijn gevonden, wordt in hoofdstuk 2 onder andere een Information Problem Solving-model ontwikkeld dat voor de rest van het onderzoek fungeert als conceptueel kader. Het onderzoek in hoofdstuk 3 is meer beschrijvend en analytisch van aard. Het richt zich op een beschrijving en vergelijking van verschillende benaderingen die in de literatuur zijn gevonden voor toetsen en beoordelen in het hoger onderwijs. De uitkomst van die literatuuranalyse is wel bepalend voor het soort instrument waarvoor uiteindelijk wordt gekozen, een zogenaamde scoringsrubriek. Hoofdstuk 4 beschrijft de ontwikkeling van een prototype van zo'n scoringsrubriek voor informatievaardigheden.

De tweede fase van het ontwerpgerichte onderzoek – het testen en evalueren van de ontwikkelde oplossing – wordt beschreven in vier hoofdstukken in deel II van dit proefschrift.

2

Wetenschappelijk onderzoek naar informatievaardigheden

Abstract: In het wetenschappelijk onderzoek naar informatievaardigheden worden drie benaderingen onderscheiden waarvan er twee relevant zijn voor dit onderzoek: de uitkomstgerichte gedragsbenadering en de procesgerichte constructivistische benadering. Op basis van andere onderzoeken uit de procesgerichte constructivistische benadering wordt een model van het Information Problem Solving-proces geschetst dat voor de rest van het onderzoek fungeert als conceptueel kader.

Het oplossen van informatieproblemen is niet alleen een doel op zich zelf maar is ook bedoeld om nieuwe kennis op te bouwen. Daarvoor wordt het idee van een Persoonlijk Kennissysteem geïntroduceerd. Wetenschappelijke theorieën over leren zijn gebruikt om het concept van een Persoonlijk Kennissysteem nader te operationaliseren.

Inleiding

Belangstelling van wetenschappers en opiniemakers voor informatievaardigheden is nauw verbonden met het ontstaan van de informatiemaatschappij. Aan het einde van de jaren '80 van de vorige eeuw kwam men met name in de Verenigde Staten, en iets later ook in Australië, tot het besef dat vaardigheid in het gebruik van informatie verder gaat dan behendigheid in het gebruik van computers. Informatievaardigheid - zowel die van individuele werknemers als die van een organisatie als geheel - werd gezien als cruciaal voor het bedrijfsleven (Drucker, 1992). Pleitbezorgers van initiatieven op het gebied van informatievaardigheid richtten zich dan ook niet alleen tot het onderwijs maar ook tot arbeidsorganisaties en de overheid. Aan de laatste werd onder andere de taak toegedicht ervoor te zorgen dat minder kansrijke groepen niet worden buitengesloten (ALA, 1989). Toegang tot informatie, vaardigheid in het opsporen en kritisch gebruik ervan, werden bovendien gezien als cruciaal voor de ontwikkeling van democratisch burgerschap (ALA, 1989).

In de eerste helft van de jaren '90 ontstond er schoorvoetend wetenschappelijke belangstelling voor onderzoek naar het fenomeen informatievaardigheid waarbij er een nadruk lag op definitiestudies en op operationalisering van de 'information skills' die noodzakelijk werden geacht (Bruce, 2000). Veel van het onderzoek werd uitgevoerd binnen het (hoger) onderwijs, enkele onderzoekers (Cheuk, 1998; Kuhlthau & Tama, 2001) richtten hun vizier (ook) op werkplekken binnen bedrijven. De meeste onderzoekers hebben een achtergrond in bibliotheek- en informatiewetenschappen. Onderwerpen die in het wetenschappelijk onderzoek aan bod komen zijn normen voor informatievaardigheid (ACRL, 2000; Bundy, 2004), de toepassing ervan binnen specifieke disciplines (bijvoorbeeld Cooney & Hiris, 2003), ervaringen van professionals met betrekking tot informatievaardigheid (Bruce, 1997) en studies naar het effect van instructie in informatievaardigheid (Knight, 2006). Vanaf circa 2005 wordt ook de ict-vaardigheid van jongeren die opgroeien in het digitale tijdperk, een belangrijk onderwerp van discussie (Kvavik, 2005; Windham, 2006; Kirschner & Van den Berg, 2012).

In dit hoofdstuk wordt stilgestaan bij de wetenschappelijke en in het bijzonder de onderwijskundige context van het internationale onderzoek naar informatievaardigheden. De te presenteren inzichten bieden later in het onderzoek hulp bij de beslissingen die genomen worden in het kader van het ontwerp van het beoordelingsinstrument. De volgende vier onderzoeksvragen komen aan de orde:

- a. Welke wetenschappelijke benaderingen worden in het onderzoek naar informatievaardigheden in het hoger onderwijs, onderscheiden?
- b. Welke van deze benaderingen zijn het meest passend voor de beantwoording van de centrale onderzoeksvraag?
- c. Welke modellen worden in het wetenschappelijk onderzoek voor het fenomeen informatievaardigheden gehanteerd?
- d. Hoe ziet een model eruit dat passend is voor de beantwoording van de centrale onderzoeksvraag?

Onderzoeksvraag a zal in paragraaf 2.1 worden beantwoord op basis van een analyse van relevante onderzoeksliteratuur. De uitkomsten daarvan worden bediscussieerd in paragraaf 2.2 hetgeen zal

leiden tot beantwoording van onderzoeksvraag b. Criteria die worden gebruikt bij de bespreking van de verschillende wetenschappelijke benaderingen worden afgeleid van de eisen die in hoofdstuk 1 voor het te ontwikkelen beoordelingsinstrument werden genoemd. Dit zijn achtereenvolgens:

- Het te ontwikkelen instrument dient zowel geschikt te zijn voor summatieve beoordeling van informatievaardigheden als ingezet kunnen worden als formatief beoordelingsinstrument dat leidt tot bevordering van informatievaardigheden.
- Het instrument dient zich te richten op de uitvoering van complete informatietaken en niet op afzonderlijke kenniselementen.
- Het instrument dient generiek toepasbaar te zijn.

Van de beoordelingsinstrumenten zelf wordt uiteraard ook geëist dat ze valide en betrouwbare beoordelingen opleveren. Deze laatste eisen die worden gesteld aan de beoordelingsinstrumenten spelen echter geen rol bij de bespreking van de wetenschappelijke benaderingen.

Onderzoeksvraag c die wordt behandeld in paragraaf 2.3 zal opnieuw worden beantwoord aan de hand van een literatuurstudie. Bij de beantwoording van de onderzoeksvraag zal ook worden ingegaan op de verschillen en overeenkomsten tussen de verschillende modellen. In paragraaf 2.4 wordt op basis van de onderscheiden modellen een voorstel gedaan voor een eigen model dat passend is voor het onderhavige onderzoek. Daarbij worden de eerder genoemde eisen die aan het beoordelingsinstrument worden gesteld, gehanteerd als richtinggevend.

In paragraaf 2.5 wordt stilgestaan bij de vraag hoe informatievaardigheden volgens het gekozen model bijdragen aan de constructie van nieuwe kennis. Het hoofdstuk wordt afgesloten met een samenvatting en discussie over het belang van de verworven inzichten (paragraaf 2.6).

2.1 Verschillende onderzoeksbenaderingen

Het onderzoek richt zich op het beoordelen en bevorderen van informatievaardigheid in het *hoger onderwijs*. In de internationale onderzoeksliteratuur naar informatievaardigheid in het hoger onderwijs worden door de meeste onderzoekers - zoals bijvoorbeeld Webber en Johnston (2000) - drie onderzoeksbenaderingen onderscheiden.

2.1.1 Uitkomstgerichte gedragsbenadering ('behavioural approach')

Kenmerkend voor deze benadering is dat informatievaardigheid wordt gedefinieerd als een set van 'skills' (vaardigheden) of 'attributes' (eigenschappen) waarover een individu beschikt (Partridge, 2007, p. 2-15 ff.). Het onderscheid tussen 'skills' en 'attributes' is daarbij gelegen in het feit dat de laatste waarneembaar gedrag beschrijven (bijvoorbeeld: "an information literate person ... develops successful search strategies"; Doyle, 1992, p. 2) en als aanwijzing gelden voor het feit dat de betreffende persoon over de bijbehorende 'skills' beschikt (Doyle 1992, p. 2). De 'attributes' die op die manier zijn geformuleerd zijn prescriptief van karakter (zie ook Sundin, 2008, p. 32 en p. 34) en kunnen in het onderwijs worden gehanteerd om de mate van informatievaardigheid van een student op objectieve wijze vast te stellen. Ze zijn daarmee vergelijkbaar met wat in de onderwijsliteratuur 'learning outcomes' (leeruitkomsten of eindtermen) worden genoemd (zie ook Nuffic, 2012).

De beschrijvingen van ‘attributes’ of ‘outcomes’ in de vorm van waarneembaar gedrag, fungeren steeds als normen voor het meten van informatievaardigheid. In het vervolg van dit onderzoeksverslag zal de gedragsbenadering ook wel worden aangeduid als de “uitkomstgerichte benadering van informatievaardigheden”.

In deze benadering proberen onderzoekers tot consensus te komen over de vraag wat onder een informatievaardige persoon moet worden verstaan (Bruce, 2000). Een goed voorbeeld van zo’n studie is die van Doyle (1992) waarin op basis van een Delphi-studie zo’n poging tot definitie en formulering van gewenste ‘outcomes’ werd gedaan. Wetenschapsfilosofisch kan de gedragsbenadering worden gekenmerkt als positivistisch omdat ervan wordt uitgegaan dat de beheersing van informatievaardigheid kan worden gemeten en objectief kan worden vastgesteld.

Veel wetenschappelijk onderzoek naar informatievaardigheden uit de jaren ’90 van de vorige eeuw is te kenschetsen als uitkomstgericht. Het belangrijkste voorbeeld vormen de ACRL-standaarden (2000) die wereldwijd hebben bijgedragen aan de totstandkoming van onderzoek naar informatievaardigheid en een praktische vertaling daarvan naar de praktijk van onderwijs en bibliotheekinstructie. Ook in latere jaren vormen de ACRL-standaarden vaak nog de basis van surveys naar informatievaardigheid (bijvoorbeeld voor die van Timmers & Glas, 2010). Dergelijke tests en surveys kunnen vanuit het perspectief van toetsing worden gekenmerkt als psychometrisch. Hiermee wordt bedoeld dat er wordt uitgegaan van de klassieke testtheorie uit de psychologie waarbij gebruik wordt gemaakt van gestandaardiseerde toetsen die als doel hebben optimaal te discrimineren tussen de testpersonen (Van de Watering & Dierick, 2002, p. 63; Dochy & McDowell, 1997, p. 282).

Onder andere Webber en Johnston wijzen op het gevaar dat deze uitkomstgerichte benadering de complexe vaardigheid Information Literacy opdeelt in kleine eenheden, hetgeen bij de beoordeling ervan uiteindelijk leidt tot afvinklijsten. Een student zou zijn of haar informatievaardigheid dan bewijzen als hij/zij aantoonde de afzonderlijke ‘losse kunstjes’ te beheersen, wat geen recht doet aan het complexe karakter ervan (Webber & Johnston, 2000, p. 384). In dat verband is het inderdaad tekenend dat de vijf ACRL-standaarden leiden tot 22 gedragsindicatoren en maar liefst 87 ‘student outcomes’ of leeruitkomsten (ACRL, 2000). Ook zou de uitkomstgerichte gedragsbenadering volgens Webber en Johnston bevorderen dat studenten slechts oppervlakkig leren en de verworven kennis en vaardigheden niet transfereren naar een andere context.

2.1.2 Procesgerichte constructivistische benadering

Bij deze benadering wordt het proces bestudeerd dat een student doorloopt bij het oplossen van een specifiek informatieprobleem. De nadruk ligt dus meer op de aanpak en de stappen die worden gezet dan op het uiteindelijke resultaat. Studies die door een procesgerichte benadering worden gekenmerkt, spreken in plaats van over ‘Information Literacy’ dan ook wel over het proces van ‘Information Problem Solving’ (Brand-Gruwel, Wopereis & Vermetten, 2005) of het ‘Information Search Process’ (Kuhlthau, 1993). De meeste onderzoekers die het oplossen van informatieproblemen bestuderen, maken onderscheid tussen verschillende cognitieve deelvaardigheden of -activiteiten die nodig zijn om het proces tot een goed einde te brengen (Irving, 1985; Berkowitz & Eisenberg, 1990; Boekhorst, 2000; Brand-Gruwel et al., 2005). Het werk van Kuhlthau is daarop

een uitzondering omdat zij juist veel aandacht besteedt aan de affectieve aspecten (de ‘feelings’) gedurende verschillende fases van het informatiezoekproces (Kuhlthau, 1993).

Hoewel veel onderzoek binnen de procesgerichte benadering zich beperkt tot de bestudering van het oplossen van specifieke informatieproblemen, rechtvaardigen de meeste auteurs de aandacht voor Information Problem Solving met het argument dat een goede beheersing van dit proces ook bijdraagt aan het leerproces van studenten in het algemeen. Zo benadrukt Boekhorst dat het niet alleen gaat om het oplossen van concrete informatieproblemen maar ook om de verwerving van metacognitieve kennis (Boekhorst, 2000, pp. 102-103) en spreekt Bruce (2000) in navolging van Kuhlthau over “information literacy as a way of learning”. Het oplossen van informatieproblemen leidt in deze opvatting dus tevens tot de constructie van nieuwe kennis door de student (Webber & Johnston, 2000, p. 385) en de benadering is behalve als procesgericht daarom tevens te kenmerken als ‘constructivistisch’. Met dat laatste wordt dan vooral bedoeld op het doel van het Information Problem Solving-proces - de constructie van nieuwe kennis - en minder op het constructivisme in de kennis theoretische betekenis van het woord. Alle onderzoekers die de procesgerichte benadering van informatievaardigheden hanteren gaan er namelijk wel vanuit dat het fenomeen informatievaardigheid in een model te vatten is dat verwijst naar een in de werkelijkheid bestaand verschijnsel. Constructivisme in de epistemologische betekenis van het woord speelt wel een rol bij de in paragraaf 2.1.3 te bespreken *relationele* benadering.

De bijdrage die het oplossen van specifieke informatieproblemen levert aan kennisconstructie door de lerende, leidt er toe dat informatievaardigheid in de procesgerichte opvatting ook wordt beschouwd als een algemene leervaardigheid. De veronderstelling is dat studenten door regelmatige toepassing en oefening, de algemene vaardigheid in het oplossen van informatieproblemen verwerven die zij zowel in het vervolg van hun studie als in hun (latere) beroepspraktijk nodig hebben. Anders dan bij de uitkomstgerichte gedragsbenadering ligt de nadruk hier dus op de transfer naar nieuwe situaties (zie bijvoorbeeld ook Walraven, Brand-Gruwel & Boshuizen, 2008). In onderwijskundige termen gesproken is informatievaardigheid in deze opvatting onder andere te kenschetsen als een van de *opleidingscompetenties* die studenten nodig hebben om hun studie succesvol af te ronden (zie ook paragraaf 1.1.4).

Als we de uitkomstgerichte benadering van informatievaardigheden karakteriseren als voortkomend uit de testcultuur waarbij de nadruk ligt op standaard vragenlijsten en standaard tests, dan kan van de procesgerichte benadering worden gezegd dat zij zich kenmerkt door een focus op het leerproces. De procesgerichte benadering van informatievaardigheden heeft niet voor niets haar wortels in de constructivistische leertheorieën (Webber & Johnston, 2000, pp. 384-385). Wanneer in die constructivistische benadering van onderwijs wordt ingegaan op het onderwerp toetsing, dan wordt ook dat besproken met de vraag in het achterhoofd hoe het kan bijdragen aan het leren van de student. Het gaat daarbij dan dus niet alleen om beoordelen en meten (‘assessment of learning’) maar ook om de bevordering van het leerproces van de student (‘assessment for learning’; Stiggins, 2002).

Een overeenkomst tussen de uitkomstgerichte gedragswetenschappelijke benadering van informatievaardigheden en de procesgerichte constructivistische benadering, is dat voor beide geldt dat zij

uitgaan van het idee dat het fenomeen informatievaardigheid op een objectieve manier kan worden beschreven en dat het normatief van aard is. Ook in de modellen die op basis van de constructivistische benadering worden gehanteerd, gaan de onderzoekers uit van het ideaal van een typisch informatievaardige persoon.

2.1.3 Relationale benadering

De relationele benadering staat ook wel bekend als fenomenografisch. Dergelijk onderzoek richt zich op de verschillende manieren waarop een verschijnsel door verschillende individuen wordt ervaren (Bruce, 1999, pp. 34-35). Centraal staat dus niet een fenomeen of concept zelf maar de contextafhankelijke relatie van individuele personen met het betreffende fenomeen. De relationele benadering van informatievaardigheden is daarmee verwant aan het constructivistisch wetenschappelijke perspectief dat als uitgangspunt gebruikt dat de werkelijkheid niet te kennen is zonder dat ook het perspectief van de handelende persoon er bij wordt betrokken (Murphy, 1997).

Een belangrijk onderscheid met de uitkomstgerichte en de procesgerichte benadering is dat informatievaardigheid in de relationele benadering veel minder een normatief begrip is en dat de laatste probeert een beschrijving te geven van de werkelijkheid in al haar verschillende verschijningsvormen. Onderzoek naar informatievaardigheden is in de relationele benadering met andere woorden meer descriptief dan prescriptief.

Het proefschrift van Christine Bruce (1997) wordt algemeen beschouwd als hét voorbeeld van fenomenografisch onderzoek naar informatievaardigheden. Voor een weergave van haar onderzoek en de resultaten daarvan, is gebruik gemaakt van haar publicatie in het tijdschrift *Higher Education Research & Development* (Bruce, 1998).

Op basis van 16 interviews en schriftelijke antwoorden van 44 respondenten, onderscheidt Bruce 7 'conceptions' (niet te verwarren met 'concepten') van informatievaardigheid, zoals die door professionals in het hoger onderwijs worden ervaren.

- a. gebruik van informatie- en communicatietechnologie ("Information Technology Conception")
- b. kennis van informatiebronnen en daarmee kunnen werken ("Information Sources Conception")
- c. verwerken van informatie in het geval iemand voor een nieuw vraagstuk komt te staan en een tekort aan kennis ervaart ("Information Process Conception")
- d. managen van (informatie)bronnen ("Information Control Conception")
- e. construeren van kennis rond een nieuw onderwerp ("Knowledge Construction Conception")
- f. verwerven van nieuwe inzichten op basis van ervaringen ("Knowledge Extension Conception")
- g. informatie in verschillende contexten op zo'n manier gebruiken dat ook anderen ervan profiteren ("Wisdom Conception").

Vervolgens benadrukt Bruce dat een consequentie van de uitkomsten van haar onderzoek is dat onderwijs in informatievaardigheden studenten zou moeten helpen om met zoveel mogelijk verschillende opvattingen van het betreffende fenomeen kennis te maken en hun ervaringen ermee

te verdiepen (Bruce, 1998, p. 41; maar ook Bruce, 1999, p. 35). Hoewel de hier geciteerde publicaties van Bruce over het algemeen worden beschouwd als sleutelpublicaties in het onderzoek naar informatievaardigheden (zie onder andere Webber & Johnston, 2000, p. 385) kan daarbij wel de kanttekening worden geplaatst dat in haar onderzoek de ervaringen van *professionals* in het hoger onderwijs werden bestudeerd (docenten, informatiespecialisten, stafmedewerkers en studiebegeleiders) en niet die van studenten.

Een ander veel besproken fenomenografisch onderzoek dat zich wel op studenten richtte, is dat van Limberg (1999). Zij bestudeerde de opvattingen van 25 Zweedse 'high school seniors' over het informatiezoekproces en informatiegebruik, en vond een sterk verband tussen de wijze waarop het zoeken naar informatie werd ervaren - als 'fact finding' of juist meer als middel om uitgebreid geïnformeerd te raken - en de inhoudelijke kwaliteit van het werk bij een gegeven studieopdracht. De uitwerkingen van de oppervlakkige informatiezoekers waren ook inhoudelijk van mindere kwaliteit dan die van de studenten die er een degelijker opvatting over informatie zoeken op na hielden. Limberg maakt echter ook een andere opmerking die in het kader van dit hoofdstuk interessant is. Zij wijst erop dat haar onderzoek bovendien inzicht geeft in de feitelijke beleving van het informatiezoekproces door studenten, ook als dat niet volgens de ideale modellen van de onderzoekers wordt uitgevoerd. Daarin wijkt haar eigen onderzoek volgens haar zelf af van de meer traditionele onderzoeken naar informatievaardigheid: "Researchers have shown less interest in describing how users, who do not follow the models, actually think or act". Daarmee geeft ze een goede karakteristiek van het verschil tussen fenomenografisch opgezet onderzoek naar informatievaardigheden en onderzoek dat op een uitkomstgerichte of procesgerichte leest is geschoeid.

Een derde document dat in het kader van de relationele benadering van informatievaardigheden genoemd dient te worden is het Framework for Information Literacy for Higher Education waarvan de tweede draft in juni 2014 door de ACRL werd gepubliceerd (ACRL, 2015). Dit framework dat aanvankelijk bedoeld was als de opvolger van de ACRL-standaarden, is georganiseerd rond kernbegrippen (in het document "threshold concepts" genoemd) die door de ontwikkelaars van het informatievaardighedenonderwijs nog nader moeten worden uitgewerkt tot te hanteren eindtermen of 'learning outcomes' (Oakleaf, 2014). Gebaseerd op onder andere het eerder genoemde onderzoek van Bruce en Limberg wordt in het Framework benadrukt dat het begrip information literacy door zowel groepen als individuen heel verschillend wordt opgevat, afhankelijk van bijvoorbeeld context en discipline waarin het wordt gebruikt. Daarom wordt er van uitgegaan dat de uiteindelijk te gebruiken eindtermen door de onderwijsontwikkelaars zelf worden geformuleerd. In het Framework wordt dan ook afstand genomen van de prescriptieve benadering die bij de standaarden uit 2000 werd gehanteerd.

2.2 Relevantie van de drie onderzoeksbenaderingen voor de beantwoording van de centrale onderzoeksvraag

Het onderzoek heeft betrekking op assessmentinstrumenten die (onder andere) toepasbaar zijn voor het toetsen van informatievaardigheden van individuele studenten in het hoger onderwijs.

Zoals hiervoor aangegeven is de relationele of fenomenografische benadering meer beschrijvend dan normatief van aard. Zonder de waarde van fenomenografisch onderzoek naar informatievaardigheden in het hoger onderwijs in twijfel te trekken, kan worden vastgesteld dat deze benadering minder geschikt is als we instrumenten willen ontwikkelen voor het *meten* van informatievaardigheid. Voor dat laatste is het namelijk wel gewenst dat duidelijk is welke leeruitkomsten van de studenten worden verwacht of dat daar in ieder geval consensus over is tussen deskundigen. In dit onderzoek naar het beoordelen van informatievaardigheden speelt de relationele benadering dan ook geen rol van betekenis. De zeven 'conceptions' van informatievaardigheid die Bruce (1998) onder professionals in het hoger onderwijs onderscheidt (zie paragraaf 2.1.3) zullen in het vervolg van dit onderzoek echter wel een rol spelen.

Van de overige twee besproken onderzoeksbenaderingen sluit de uitkomstgerichte gedragsbenadering het best aan bij het onderwerp toetsing. In onderzoeken die deze benadering hanteren, staat de operationalisering van 'vaardig gedrag' tijdens het oplossen van informatieproblemen centraal. Die operationalisering kan zich dan overigens nog steeds richten op een product dat wordt geproduceerd of op de processtappen die daarbij worden uitgevoerd (Nitko & Brookhart, 2007, pp. 263-264).

De operationalisering van informatievaardig gedrag is een noodzakelijke stap om te komen tot een meetinstrument dat informatievaardigheid op een valide en betrouwbare wijze beoordeelt. Bij het onderhavige onderzoek zullen de validiteit en de betrouwbaarheid van de te ontwikkelen instrumenten (en die van de daarbij te gebruiken variabelen als operationalisering van informatievaardig gedrag) als criteria worden gebruikt om uitspraken te doen over de geschiktheid van het instrument.

Bij de keuze voor een of meer geschikte meetinstrumenten dient echter ook rekening te worden gehouden met de maatschappelijke en onderwijskundige context waarbinnen de instrumenten gebruikt kunnen worden. De aandacht voor informatievaardigheden is in hoofdstuk 1 al gemotiveerd met de constatering dat het hoger onderwijs, onder invloed van ontwikkelingen naar een kenniseconomie, behoefte heeft aan onderwijsvormen die de zelfwerkzaamheid en de onderzoeksvaardigheid van aankomende professionals stimuleren. Daar werd ook opgemerkt dat in het hedendaagse hoger onderwijs bovendien het idee heeft post gevat dat het voldoende recht dient te doen aan het complexe en samengestelde karakter van de beroepstaken waarmee de aankomende professionals steeds vaker worden geconfronteerd. Vanuit dat perspectief loopt de uitkomstgerichte benadering van informatievaardigheden het risico dat de focus op operationalisering van informatievaardig gedrag, leidt tot een te ver doorgeslagen aandacht voor deelfacetten van het complexe fenomeen informatievaardigheid waarbij er geen aandacht meer is voor de wijze waarop de verschillende facetten ervan elkaar beïnvloeden.

De procesgerichte benadering van informatievaardigheden sluit wat dat betreft veel beter aan bij de hedendaagse denkwijze over hoger onderwijs. De diverse modellen die worden gehanteerd voor het proces van Information Problem Solving onderscheiden weliswaar verschillende deelvaardigheden maar deze worden wel als een samengesteld geheel beschouwd. Daar staat weer tegenover dat in het procesgerichte onderzoek naar informatievaardigheden (zoals de naam al suggereert)

de focus vooral is gericht op het leerproces van de student en dat in de meeste onderzoeken geen aandacht wordt besteed aan de wijze waarop de leeruitkomsten daarvan kunnen worden gemeten. Zoals aangegeven in de centrale onderzoeksvraag zal in het onderhavige onderzoek juist de vraag aan de orde komen wat voor soort instrument zowel geschikt is voor het beoordelen van informatievaardigheden als voor het bevorderen daarvan. In de onderzoeken die in de volgende hoofdstukken worden beschreven wordt dan ook een voorstel gedaan om een juiste mix te vinden van de uitkomstgerichte en de procesgerichte benadering. Een van de eisen die in paragraaf 1.3 aan de instrumenten werd gesteld, is dat ze voldoende recht doen aan het gegeven dat ook het oplossen van een informatieprobleem een samengestelde taak is met een holistisch karakter. Om zicht te krijgen op dat laatste zal in de volgende paragraaf worden stilgestaan bij verschillende modellen die binnen de procesgerichte benadering van informatievaardigheden worden gehanteerd. Deze modellen schetsen op verschillende manieren de context waarbinnen het te ontwikkelen beoordelingsinstrument dient te kunnen functioneren. Met betrekking tot toetsing zelf kan worden aangesloten bij de constructivistische *leertheorieën* waar de procesgerichte benadering van informatievaardigheden door is geïnspireerd en die in het huidige hoger onderwijs een prominente plaats innemen. Dat betekent, zoals opgemerkt in paragraaf 2.1.2, dat de te kiezen toetsinstrumenten ook ingezet zouden moeten kunnen worden als een ‘tool for learning’.

2.3 Verschillende modellen

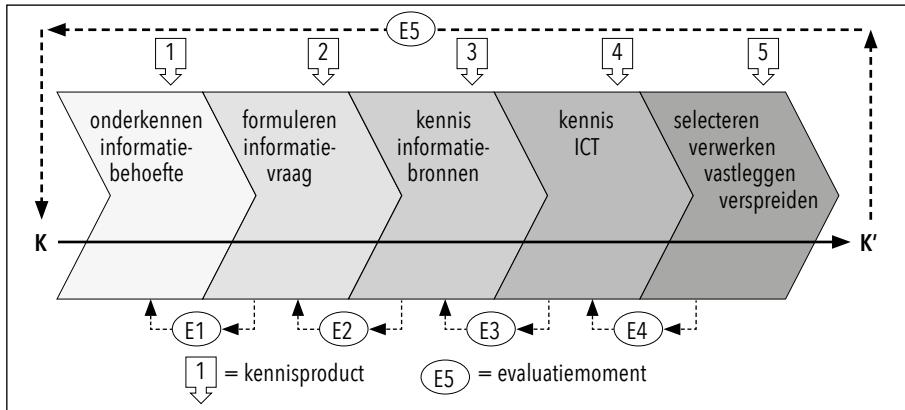
Informatievaardigheden worden beschouwd als de vaardigheden die nodig zijn bij het oplossen van informatieproblemen (Brand-Gruwel et al., 2005). In de internationale vakliteratuur worden diverse modellen van de complexe vaardigheid Information Problem Solving beschreven. In deze paragraaf wordt de bespreking beperkt tot vier van deze modellen. Deze worden met elkaar vergeleken en op basis van de onderwijskundige context waarbinnen het eigen onderzoek wordt uitgevoerd, zal in paragraaf 2.4 een voorstel worden gedaan voor een te hanteren model waarin aspecten uit verschillende modellen terugkomen.

Het eerste model dat wordt besproken is dat van Boekhorst (1999) omdat dit richtinggevend is geweest voor veel theorievorming en de praktijk van het onderwijs in informatievaardigheden in Nederland. Vervolgens komt het model voor Information Problem Solving van Brand-Gruwel et al. aan de orde (Brand-Gruwel et al., 2005). Daarvoor is gekozen omdat het goed aansluit bij ontwikkelingen in het Nederlandse hoger onderwijs die zijn beschreven in hoofdstuk 1 en paragraaf 2.2. Als tegenhanger van deze twee modellen wordt stilgestaan bij Kuhlthaus Information Search Process-model (Kuhlthau, 1993) omdat het, anders dan het model van Boekhorst en dat van Brand-Gruwel et al., sterk de nadruk legt op het gebruikersperspectief van het informatiezoekproces. Het vierde model dat de revue passeert is dat van Irving (1985) dat een voorbeeld is van een model dat specifiek aandacht besteedt aan het proces van kennis*constructie*.

Bij de te bespreken modellen ontbreekt het Big Six Skills model van Berkowitz en Eisenberg (1990). Hoewel dit internationaal veel aandacht heeft getrokken zal het hier niet besproken worden omdat het model van Brand-Gruwel et al. sterk op het Big Six Skills-model is geënt. Brand-Gruwel et al.

voegden echter een regulatiecomponent aan het model van Berkowitz en Eisenberg toe (Brand-Gruwel et al., 2005, p. 489). Het belang daarvan zal duidelijk worden als het model van Brand-Gruwel et al. wordt besproken.

Boekhorst



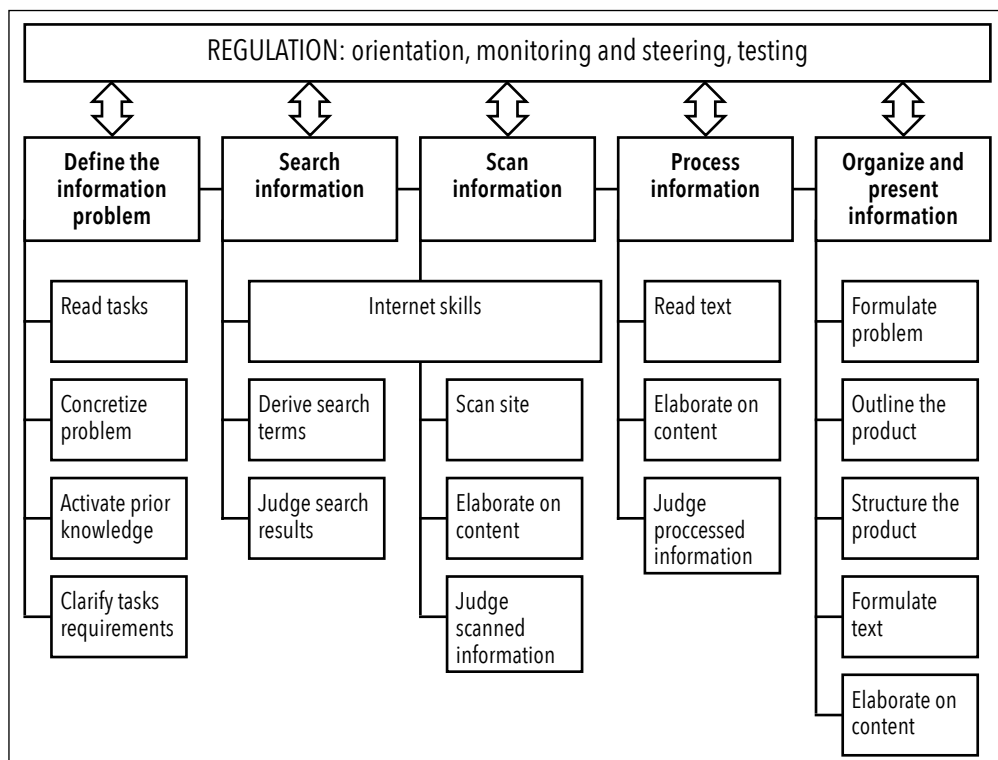
Figuur 2.1

Model Informatievaardigheden van Boekhorst
(Boekhorst, Koers & Kwast, 1999)

In het “Model Informatievaardigheden” van Boekhorst (figuur 2.1) herkent men de zes deelvaardigheden die in hoofdstuk 1 voor het concept Informatievaardigheden zijn onderscheiden. Dat is niet toevallig omdat het onderscheid van de zes deelvaardigheden onder andere was gebaseerd op de definitie die Boekhorst zelf gaf in zijn proefschrift uit het jaar 2000.

De deelvaardigheid “selecteren en beoordelen van de informatie die is gevonden” maakt in zijn model onderdeel uit van stap 5 maar tevens van de verschillende evaluatiemomenten (E1 tot en met E5). Anders dan de modellen van Brand-Gruwel en Kuhlthau die hierna besproken zullen worden, is het model van Boekhorst niet gericht op het uitvoeren van specifieke studie- of onderzoeksopdrachten maar op het oplossen van informatieproblemen in het algemeen. Dat komt onder andere tot uitdrukking in het feit dat de stappen 3 en 4 (anders dan 1, 2 en 5) niet geformuleerd zijn als een activiteit maar als “kennis”. Ook volgens Boekhorst zelf is zijn model dan ook “geen inhoudelijk onderwijskundig of didactisch model”, hoewel er wel uit afgeleid kan worden aan welke aspecten in het onderwijs expliciet aandacht besteed dient te worden (2000, pp. 102-103). Als kader voor het ontwikkelen van een assessmentinstrumentarium is het echter niet direct bruikbaar wat inderdaad wordt veroorzaakt door het feit dat zijn model niet refereert naar de uitvoering van afzonderlijke studie-opdrachten en dat er verschillende grootheden (‘activiteiten’ en ‘kennisgebieden’) naast elkaar worden gebruikt.

Een belangrijk element in het model van Boekhorst dat wel goed bruikbaar is en dat aansluit bij de huidige onderwijskundige ideeën, is het iteratieve karakter ervan dat wordt uitgedrukt door de terugkerende evaluatiemomenten, zowel aan het einde van iedere afzonderlijke stap als aan het einde van het totale informatieverwerkingsproces.



Figuur 2.2
Information Problem Solving-model van Brand-Gruwel et al. (2005)

Het model van Brand-Gruwel et al. hanteert de theoretische uitgangspunten van het instructieontwerp voor het aanleren van complexe vaardigheden (Brand-Gruwel et al., 2005). Zo'n complexe vaardigheid wordt daarbij opgevat als een samengesteld geheel van deelvaardigheden die gecoördineerd en geïntegreerd uitgevoerd dienen te worden. Juist bij het idee van 'performance assessments' en de beoordeling van 'hele beroepstaken' dat werd besproken in paragraaf 1.3.2 van dit proefschrift, spelen dergelijke complexe vaardigheden volgens Brand-Gruwel et al. een belangrijke rol.

De onderzoekers hebben het model opgesteld op basis van de uitkomsten van een experimenteel onderzoek naar internetgedrag bij het oplossen van informatieproblemen in het kader van studietaken.⁴ Een belangrijke aanpassing in vergelijking met het oorspronkelijke Big Six skills-model van Berkowitz en Eisenberg, is de vervanging van de evaluatiefase (stap 6 bij Berkowitz en Eisenberg)

⁴ Het model is onder andere gebaseerd op observaties die de onderzoekers deden bij studenten tijdens de uitvoering van een opdracht om informatie op te sporen op het internet en deze te gebruiken bij het schrijven van een kort essay (maximaal 400 woorden). Voor de uitvoering van de opdracht was in zijn geheel 90 minuten beschikbaar.

door een component met regulatieactiviteiten die het model van Brand-Gruwel et al. een iteratief karakter geeft (Brand-Gruwel et al., 2005). Ook dit iteratieve karakter is volgens deze auteurs kenmerkend voor de ‘hele taak-benadering’.

Het onderscheid tussen een taakgericht model en een meer algemeen model als dat van Boekhorst blijkt al direct bij de eerste stap die Brand-Gruwel et al. formuleren (“Define the information problem”). In het model van Brand-Gruwel et al. wordt uitgegaan van een bestaand informatieprobleem terwijl er in het model van Boekhorst van wordt uitgegaan dat die informatiebehoefte eerst onderkend dient te worden. In het model van Brand-Gruwel et al. heeft die eerste stap zowel betrekking op de inhoud van de studietaak als op het formuleren van een onderzoeksvraag en de bijbehorende subvragen. Dit wordt uitgedrukt door de deelvvaardigheid “Concretize problem”.

In de toelichting op het model wordt duidelijk dat de stap “Search information” ook betrekking heeft op de selectie van geschikte informatiebronnen (in dit geval dus bronnen die beschikbaar zijn op het internet) en het gebruik van het internet zelf. In de studie van Brand-Gruwel et al. wordt op de benodigde internetvaardigheden niet nader ingegaan, al wordt wel erkend dat ze een rol spelen in het zoekproces. Van belang is op te merken dat in het model wordt onderkend dat het selecteren van geschikte zoektermen ook in deze fase (en dus niet alleen in de oriëntatiefase vooraf) plaatsvindt. Het verschijnsel dat zoektermen steeds vaker tijdens het zoekproces (‘in action’) worden geformuleerd is inderdaad kenmerkend voor het zoekgedrag van mensen in een tijdperk waarin zij continu in een online verbinding met informatiesystemen staan.

Het onderscheid tussen “scannen” van informatie en “process information” is gelegen in het feit dat het “scannen” beperkt blijft tot het oppervlakkig analyseren van de geboden content en dat de informatiegebruiker op basis van die analyse besluit om van de betreffende sites wel of niet diepgaander kennis te gaan nemen (Brand-Gruwel et al., 2005)⁵. In de laatste stap van het model vindt de synthese plaats die leidt tot een nieuw ‘eigen’ kennisproduct, bijvoorbeeld in de vorm van een paper of een presentatie. De regulatieactiviteiten “steering” (plannen) en “monitoring” hebben vooral betrekking op de uit te voeren deelactiviteiten, de “orientation” is meer gericht op de inhoud van de volledige studietaak. “Testing” tenslotte komt vooral overeen met wat in andere modellen wordt aangeduid als “evaluatie” (Brand-Gruwel et al., 2005, p. 491). In de onderwijskundige literatuur worden dergelijke regulatieactiviteiten ook wel aangeduid als “metacognitieve vaardigheden” (Vermunt & Verloop, 1999, pp. 262-263).

In hun studie vergelijken Brand-Gruwel et al. het oplossen van informatieproblemen door experts (PhD studenten) met de manier waarop eerstejaars (“freshmen”) van een universitaire opleiding dat aanpakken. Op basis van waargenomen verschillen in tijdbesteding doen zij de suggestie om

5 Irving, wiens model we verderop bespreken, maakt nog een onderscheid tussen “scanning” en “skimming” van informatie. Bij “scannen” gaat het erom dat de informatiegebruiker op zoek gaat naar bekende woorden (bijvoorbeeld door gebruik van een register en vervolgens de betreffende pagina in vogelvlucht door te nemen), bij “skimmen” probeert de informatiegebruiker snel een overzicht van de inhoud te krijgen (bijvoorbeeld door te “bladeren”) (Irving, 1985, p. 81). Waar Brand-Gruwel et al. in hun model spreken over “scannen” wordt (ook) “skimmen” bedoeld.

voor de instructie met betrekking tot Information Problem Solving of informatievaardigheden de nadruk te leggen op:

definiëren van het informatieprobleem (stap 1)

verwerken van informatie (stap 4)

beoordelen van informatie (subvaardigheid van stap 2, 3 en 4)

en de component regulatieactiviteiten.

De ontwikkeling van deze laatste metacognitieve vaardigheden komt volgens de auteurs het beste tot stand door een combinatie van observatie, samenwerkend oefenen en begeleiding (zie ook Lazonder, 2005).

Kuhlthau

Het Information Search Process (ISP)-model van Kuhlthau wordt gekenmerkt door een expliciete aandacht voor de affectieve aspecten van het zoekproces, de manier waarop het zoekproces door de gebruiker wordt *beleefd* (de “feelings” in figuur 2.3). Daarnaast spelen cognitieve (“thoughts”) en uitvoerende activiteiten (“actions”) een rol. De laatste hebben gedurende alle fases van het proces betrekking op het zoeken naar informatie. In de fases die voorafgaan aan de “Focus Formulation” gaat het daarbij vooral om bronnenexploratie met als doel in het onderwerp thuis te raken. Gedurende de voortgang van het proces ontwikkelt het zoeken naar *relevante informatie* zich volgens Kuhlthau tot het opsporen van *informatie met pertinentie*, wat inhoudt dat de informatie niet alleen relevant is maar bovendien iets toevoegt aan de kennis die al beschikbaar is (Kuhlthau, 1993, p. 39). Pertinentie is daardoor meer subjectief van karakter dan relevantie.

Een cruciale fase van het zoekproces is volgens het model van Kuhlthau “stage” 4, de “focus formulation”. Dat is het punt waarop onzekerheid bij de informatiezoeker omslaat in vertrouwen (Kuhlthau, 1993, p. 46).

Stages	Task Initiation	Topic Selection	Prefocus Exploration	Focus Formulation	Information Collection	Search Closure	Starting Writing
Feelings	uncertainty	optimism	confusion, frustration, and doubt	clarity	sense of direction/ confidence	relief	satisfaction or dissatisfaction
Thoughts		ambiguity	-----> specificity				
				----->			
				increased interest			
Actions	seeking relevant information -----> seeking pertinent information						

Figuur 2.3

Model van het Information Search Process volgens Kuhlthau (1993, p. 43)

De toegevoegde waarde van Kuhlthaus ISP-model is het belang dat daarin wordt toegekend aan de fases die voorafgaan aan het formuleren van de focus. Vooral de exploratie van het te onderzoeken onderwerp door het opsporen en bestuderen van achtergrondliteratuur vóórdat men begint aan het oplossen van de concrete informatievraag, is volgens Kuhlthau cruciaal om het zoekproces tot

een goed einde te brengen. In termen van de eerder genoemde deelvaardigheden uit hoofdstuk 1 gaat het hierbij om het “onderkennen van een informatiebehoefte”, een strategische informatievaardigheid die studenten niet zomaar uit zichzelf zullen verwerven. De oriëntatie op het onderwerp door je eerst goed ‘in te lezen’ is inderdaad strategisch van aard omdat het hier gaat om een activiteit die aanvankelijk weinig direct resultaat oplevert maar die zich op termijn uitbetaalt.

In vergelijking met de eerder besproken modellen van Boekhorst en Brand-Gruwel et al. is het ISP-model van Kuhlthau beperkter in zijn reikwijdte omdat de uiteindelijke creatie van een nieuw kennisproduct er geen deel van uitmaakt. Dit kan worden opgevat als een tekortkoming van het model van Kuhlthau omdat de ervaringen van een informatiegebruiker tijdens het werken aan een nieuw kennisproduct immers ook aanleiding kunnen zijn tot nieuwe informatiebehoefes en bijstelling van het informatiezoekproces.

Irving

Het ‘model’ van Irving is gebaseerd op 9 stappen die een student doorloopt bij het werken aan studietaken (Irving, 1985). In haar boek geeft de auteur geen visuele weergave daarvan maar het geheel is niettemin te beschouwen als een ‘model’ dat beschrijft over welke vaardigheden een student dient te beschikken om een studieopdracht tot een goed einde te brengen. Figuur 2.4 geeft een overzicht van de afzonderlijke vaardigheden die Irving onderscheidt, aangevuld met concrete voorbeelden van activiteiten of te gebruiken kennis. Anders dan andere auteurs die haar model bespreken (zoals bijvoorbeeld Eisenberg, 1996) wordt daarbij gebruik gemaakt van de terminologie die Irving zelf in hoofdstuk 3 van haar studie (p. 33 ff.) heeft gehanteerd.

Hoewel ook het model van Irving uitgaat van de vaardigheden die nodig zijn om afzonderlijke studietaken uit te voeren, legt zij in de nadere toelichting op de verschillende activiteiten de nadruk op de wijze waarop die bijdragen aan de constructie van mentale modellen en cognitieve strategieën. Haar model besteedt dan ook meer dan de andere besproken modellen aandacht aan de cognitieve verwerking van informatie bij de creatie van een nieuw kennisproduct (“Effective use”, “Making records”, “Making sense”). Ook de aandacht voor ‘concept mapping’ in de fases “Defining tasks” en “Finding resources” geeft aan dat haar model uiteindelijk vooral gericht is op de constructie van persoonlijke kennis.

Het bedenken van mogelijke zoektermen maakt bij Irving zowel deel uit van stap 1 als van stap 3. De uitwerking in haar boek van de stappen 2 en 3 (“Considering sources” en “Finding resources”) dragen wel de sporen van de tijd en deze is inmiddels dan ook enigszins achterhaald.

1. Defining tasks

Voorbeelden van activiteiten: formuleren van de informatiebehoefte; in kaart brengen van bestaande kennis; concept mapping; bedenken zoektermen

2. Considering sources

Voorbeelden van mogelijke informatiebronnen: bibliotheken en hun netwerk; informatiecentra; experts; online bestanden

3. Finding resources

Voorbeelden van activiteiten: selecteren van zoektermen en zoekingen; concept mapping

4. Making selections

Voorbeelden van activiteiten: kiezen van te gebruiken primaire bronnen; verwerpen van bronnen die niet geschikt gevonden worden

5. Effective use

Voorbeelden van activiteiten: gebruik van focusgerichte leesstrategieën zoals "scannen" en "skimmen"; interpreteren van tekst en beeld

6. Making records

Voorbeelden van activiteiten: aantekeningen maken; onderscheiden van hoofd- en bijzaken

7. Making sense

Voorbeelden van activiteiten: reflecteren, analyseren en synthetiseren tot een persoonlijk 'construct'

8. Presenting work

Voorbeelden van activiteiten: structureren; schrijven van een onderzoeksverslag of case studie; maken van een audiovisuele presentatie

9. Assessing progress

Voorbeelden van activiteiten: zelfevaluatie; peer-assessment; assessment door docenten

Figuur 2.4

Information skills-model van Irving (1985)

Vergelijking van de vier IPS-modellen

De vier modellen worden in onderstaande figuur 2.5 vergeleken met elkaar en met de zes deelvaardigheden die in hoofdstuk 1 zijn onderscheiden. In het vervolg van deze paragraaf wordt, behalve op de verschillen in processtappen bij de vier modellen, ingegaan op andere karakteristieken die kenmerkend zijn voor de afzonderlijke modellen.

6 deeltaalvaardigheden	Boekhorst	Brand-Gruwel	Kuhlthau	Irving			
1. Onderkennen informatiebehoefte	1. Onderkennen informatie-behoefte	1. Define information problem	1.Task initiation		1. Defining task		
	Evalueren		2. Topic selection				
2. Formuleren zoekvraag	2. Formuleren informatievraag		3. Prefocus exploration	4. Focus formulation		3. Finding resources	
	Evalueren						Regulation
3. Selecteren informatiebronnen	3. Kennis informatie-bronnen	2. Search information			2. Considering sources		
	Evalueren						
4. Gebruik van ICT	4. Kennis ICT	Regulation	5. Information collection		3. Finding resources		
	Evalueren						
5. Selecteren en beoordelen van informatie die is gevonden	5. Selecteren, verwerken vastleggen en verspreiden	3. Scan information					4. Making selections
		Regulation					
6. Verwerken, hergebruiken en verspreiden van de informatie		4. Process information	6. Search closure		5. Effective use		
		Regulation			6. Making records		
		5. Organize and present information	7. Starting writing		7. Making sense		
		Evalueren	Regulation			8. Presenting work	
					9. Assessing progress		

Figuur 2.5

Vergelijking van 4 Information Problem Solving-modellen

De modellen van Brand-Gruwel et al. en van Kuhlthau zijn beide uitdrukkelijk gericht op het oplossen van een specifiek informatieprobleem en sluiten daarmee aan bij de procesgerichte opvatting van informatievaardigheden die in paragraaf 2.1 is onderscheiden. Het model van Kuhlthau beperkt zich echter tot het informatie *zoek*proces en de fase van informatieverwerking wordt daarin niet meegenomen. Dat maakt het model van Kuhlthau minder bruikbaar voor het onderhavige onderzoek. Bovendien biedt het weinig zicht op de verschillende zoekactiviteiten die informatiezoekers uitvoeren (zoeken met keywords, grasduinen in een gestructureerde verzameling, surfen via verwijzingen etcetera).

Het model van Kuhlthau voegt wel het waardevolle inzicht toe dat een goede oriëntatie op het probleem of het vraagstuk door middel van bronnenexploratie essentieel is voor het oplossen van een informatieprobleem. Ook het model van Brand-Gruwel et al. onderkent het belang van een goede definitie van het informatieprobleem (en de auteurs geven aan dat het zoekproces van experts zich onder andere op dit punt onderscheidt van dat van beginnelingen) maar de expliciete aandacht voor de rol van bronnenexploratie daarbij ontbreekt in hun model.

Andere bezwaren tegen het model van Brand-Gruwel et al. zijn de beperking tot een studieopdracht die in 90 minuten uitgevoerd dient te worden en de beperking van het gebruik van informatiebronnen tot het internet. Verder wordt in het model eveneens geen onderscheid gemaakt tussen verschillende zoekmethodes die gebruikt kunnen worden maar in de toelichtende studie wordt bij de internet skills wel onderscheid gemaakt tussen browsen, meanderen (wat te beschouwen is als de internetvariant van 'backward' en 'forward chaining') en het gebruik van een search engine. Ook sluit het goed aan bij de eerder besproken trend om het hoger onderwijs meer te richten op complexe (samengestelde) beroepstaken: de rol van de component "Regulation" in het model maakt duidelijk dat het oplossen van informatieproblemen zelf een complexe taak is die is samengesteld uit verschillende deeltaken die geïntegreerd en gecoördineerd worden uitgevoerd. Van de negen "information skills" die Irving onderscheidt, hebben de laatste vijf betrekking op de deelvaardigheid "Verwerken, hergebruiken en verspreiden van de informatie"⁶. De nadruk bij Irvings model op de cognitieve verwerking van informatie sluit goed aan bij de opvatting dat het oplossen van specifieke informatieproblemen tevens bijdraagt aan de constructie van (metacognitieve) kennis. Als model voor het proces van Information Problem Solving schiet het echter tekort omdat het iteratieve karakter van dat proces in haar model niet tot uitdrukking komt.

Ook het model van Boekhorst is in het kader van de onderzoeksvraag minder bruikbaar. Het toont overeenkomsten met het model van Brand-Gruwel et al. en met dat van Irving maar is als model voor het ontwerpen van onderwijs en toetsing minder consequent uitgewerkt. Wel kunnen er de cognitieve elementen uit afgeleid worden die kunnen dienen als leerinhoud voor onderwijs in informatievaardigheden.

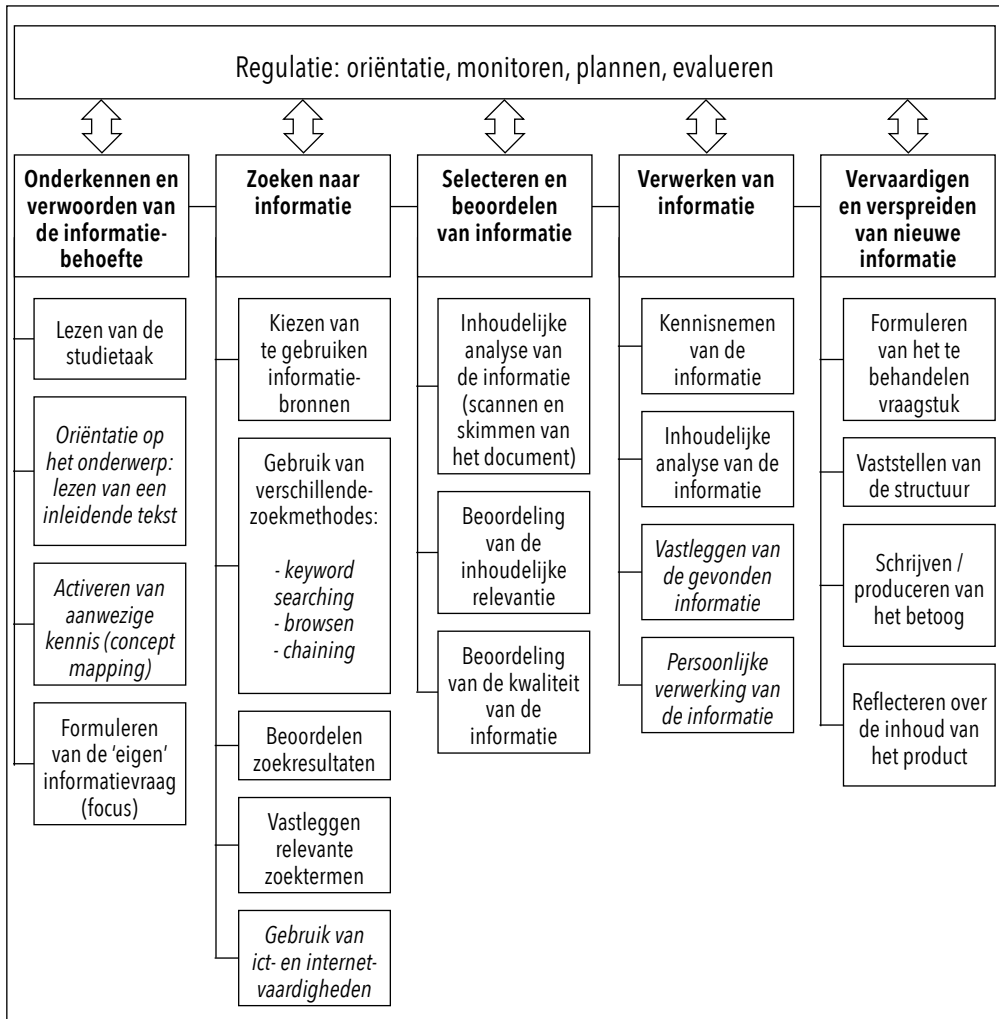
2.4 Een passend IPS-model voor het onderhavige onderzoek⁷

Het Information Problem Solving-model van Brand-Gruwel et al. sluit het beste aan bij de onderwijskundige opvattingen die momenteel in het Nederlandse hoger onderwijs worden gehanteerd. Vooral de uitwerking en de rol van regulatieactiviteiten in dat model doen recht aan de benadering van informatievaardigheden als een complexe beroepstaak. Bezwaarlijk in hun model is echter de beperking tot het gebruik van het internet en het feit dat het is gebaseerd op waarnemingen tijdens het oplossen van informatieproblemen in het kader van zeer strak (in tijd) afgebakende studieopdrachten.

In figuur 2.6 wordt daarom een variant van hun model voorgesteld die tegemoet komt aan de bovengenoemde bezwaren. Het model wordt gevolgd door een toelichting op de verschillende deelactiviteiten en de afwijkingen ten opzichte van het uitgangsmodel van Brand-Gruwel et al.

6 Als de laatste stap van haar model ("Assessing progress") niet als een aparte informatieactiviteit wordt gezien maar als een vorm van evaluatie die te vergelijken is met de regulatieactiviteiten van Brand-Gruwel et al., dan gaat het nog altijd om vier van de acht onderscheiden stappen.

7 De inhoud van paragraaf 2.4 en 2.5 is gedeeltelijk eerder gepubliceerd in: The personal knowledge base conception of information literacy. *Communications in Computer and Information Science*, 492, 31-36. (Van Helvoort, 2014).



Figuur 2.6

Het model voor het Information Problem Solving-proces (IPS-model) zoals voorgesteld in het kader van het onderhavige onderzoek

De eerste stap in het model van Brand-Gruwel is omschreven als “Define the information problem”. De auteurs geven aan dat dit is te beschouwen als een operationalisering van het “onderkennen van een informatiebehoefte”. Bij de activiteiten die behoren bij deze eerste stap ontbreekt echter aandacht voor de rol van een goede oriëntatie op het opgegeven onderwerp of vraagstuk, de stap van de “prefocus exploration” in het model van Kuhlthau. Deze fase van bronnenexploratie is bij uitstek een onderkenning van het feit dat er behoefte is aan extra (achtergrond)informatie voordat men zelf een goede (gefocuste) informatievraag kan formuleren. In het model dat hier wordt voorgesteld is de eerste stap daarom gedefinieerd als “Onderkennen en verwoorden van de informatiebehoefte” en is de oriëntatie op het onderwerp door het lezen van een inleidende tekst als

deelvaardigheid toegevoegd. De deelvaardigheid “Concretize problem” is in termen van Kuhlthau geformuleerd als “het formuleren van een eigen (gefocuste) informatievraag”. De laatste deelactiviteit (“Clarify task requirements”) is vervallen omdat die ook beschouwd kan worden als onderdeel van de eerste (“Read task”).

Bij stap 2 (“zoeken naar informatie”) is het kiezen van de te gebruiken informatiebronnen (anders dan bij Brand-Gruwel et al. die zich in hun experiment hebben beperkt tot het gebruik van internet) wel als aparte deelactiviteit opgenomen.

Als tweede deelactiviteit werd “gebruik van verschillende zoekmethodes” toegevoegd. Behalve om het formuleren van een zoekvraag in het kader van “keyword searching” kan het daarbij ook gaan om browsen of het volgen van verwijzingen (sneeuwballen of citaatzoeken; “chaining”). Aan welke van deze zoekmethodes een student de voorkeur geeft kan per student en per situatie verschillen. Dat het gebruik van de verschillende methodes optioneel is, wordt uitgedrukt doordat ze cursief staan vermeld. Dat een student de zoekmethodes in voorkomende gevallen alle drie goed kan toepassen, is wel een indicatie voor zijn of haar mate van informatievaardigheid.

Op basis van de zoekresultaten maakt de student een eerste selectie van relevante documenten of ‘hits’. Een goede student zorgt ervoor dat hij / zij tegelijkertijd direct de zoektermen vastlegt waarmee de relevante documenten teruggevonden kunnen worden. Dit hoeven niet per se de zoektermen te zijn waarmee de student de betreffende documenten heeft gevonden. Het gaat er in deze fase vooral om dat de student relevante zoektermen herkent om ervoor te zorgen dat de relevante documenten (en andere relevante informatie) in een later stadium teruggevonden kunnen worden. Met name voor het gebruik van de verschillende zoekmethodes dient de student te beschikken over ict- en internetvaardigheden. Het gaat hierbij vooral om operationele vaardigheden in het gebruik van digitale hulpmiddelen en om navigatievaardigheden waarvan wordt aangenomen dat de meeste jongeren die reeds redelijk goed ontwikkeld hebben (Van Deursen & Van Dijk, 2008, pp. 5-6).

Bij de deelactiviteit “Scan information” gaat het volgens Brand-Gruwel et al. om een beoordeling van de gevonden informatie en selectie op basis van relevantie en kwaliteit. In het eigen model is er voor gekozen om dit als zodanig in de naamgeving tot uitdrukking te laten komen. De beoordeling van de relevantie van de gevonden informatie in het kader van het op te lossen informatieprobleem, is te beschouwen als een andere deelvaardigheid dan de beoordeling en selectie op basis van kwaliteitscriteria zoals betrouwbaarheid, actualiteit en volledigheid. Bij het beoordelen van de relevantie vindt eerst een globale inhoudelijke analyse van de informatie plaats.

Verwerking van de informatie gebeurt door er eerst kennis van te nemen (in de meeste gevallen door te lezen) en deze inhoudelijk te analyseren. De derde deelactiviteit die in het kader van informatieverwerking van belang is (het vastleggen van de gevonden informatie; “Making records” in het model van Irving) ontbreekt in het model van Brand-Gruwel et al. De reden daarvan kan zijn dat hun experiment betrekking had op een studietaak binnen een betrekkelijk korte tijdsperiode. Voor de meeste taken waaraan studenten in het kader van hun studie werken is het echter van belang dat ze zichzelf aanleren om gevonden informatie vast te leggen. Daarbij gaat het dan zowel om de formele kenmerken die ervoor zorgen dat de informatie opnieuw kan worden teruggevonden (auteur, titel,

jaar, url) als om inhoudelijke gegevens in de vorm van een korte samenvatting. Dat stelt hen in staat om de informatie snel te gebruiken tijdens de laatste fase van het informatiegebruik, het produceren van een eigen kennisproduct (Walhout, Brand-Gruwel & Martens, 2011). Het totale proces van informatieverwerking leidt als het goed is tot een diepgaandere beoordeling en integratie van de gevonden informatie dan in de vorige fase. Hoewel Brand-Gruwel et al. in deze fase spreken over “Judge processed information” is er in het model de voorkeur aan gegeven om het aan te duiden als “Persoonlijke verwerking van informatie” (“Making sense” in het model van Irving).

In de laatste fase vindt de uiteindelijke vervaardiging en de verspreiding van het nieuwe kennisproduct plaats. Na de definitieve formulering van het te beantwoorden vraagstuk wordt de structuur van het kennisproduct vastgesteld, in de vorm van bijvoorbeeld een ‘bouwplan’ of een ‘storyboard’. In het eigen model is de door Brand-Gruwel et al. voorgestelde “Outline” daarbij inbegrepen. De structuur wordt ingevuld in de schrijf- of productiefase. De student reflecteert tenslotte over de inhoud van het concept product wat aanleiding kan zijn om nieuwe informatievragen te stellen of aanvullende informatie op te zoeken.

De verschillende activiteiten worden in het model horizontaal en lineair gepresenteerd. De component regulatieactiviteiten leidt er echter toe dat de student op basis van monitoring en evaluatie na iedere activiteit een of meer stappen terug kan doen om een reeds uitgevoerde activiteit in aangepaste vorm te herhalen.

Ook de opeenvolging van de verschillende deelactiviteiten (in het model weergegeven in verticale kolommen) is niet in alle gevallen lineair. Zo is het bij de eerste activiteit (“Onderkennen en verwoorden van de informatiebehoefte”) arbitrair of het activeren van aanwezige kennis voorafgaat aan de oriëntatie door middel van het lezen van een inleidend document of andersom, en is het mogelijk dat bij de verwerking van informatie (activiteit 4) de persoonlijke verwerking en de vastlegging van de informatie tegelijkertijd plaatsvinden. Om die reden zijn in het model de betreffende vier deelactiviteiten (net als de drie te onderscheiden zoekmethodes bij activiteit 2) cursief weergegeven. Daarmee wordt aangegeven dat de volgorde waarin de deelactiviteiten worden uitgevoerd willekeurig is. Het gebruik van ict- en internetvaardigheden is bij “Zoeken naar informatie” cursief geplaatst omdat dit bij verschillende deelactiviteiten een rol kan spelen.

2.5 Constructie van (metacognitieve) kennis

Information Problem Solving is in de procesgerichte benadering van informatievaardigheden niet alleen een doel op zichzelf maar draagt ook bij aan de constructie van kennis door de lerende. In het kader van de Knowledge Construction Conception, Knowledge Extension Conception en de Wisdom Conception spreekt Bruce (wiens onderzoek in paragraaf 2.1.3 werd besproken) dan over een “Personal Knowledge Base”. In het vervolg van het eigen onderzoek zal daarvoor de term “Persoonlijk Kennissysteem” worden gebruikt. Verschillende auteurs binnen de procesgerichte benadering van informatievaardigheden gebruiken dit idee van kennisconstructie. De activiteiten die in het model van Irving worden benoemd hebben bijvoorbeeld vaak betrekking op de mentale verwerking van informatie maar ook in haar model wordt niet concreet gemaakt wat voor soort

kennis dat oplevert. De onderwijskundige literatuur op het gebied van instructie-ontwerp geeft daarvan een duidelijker beeld.

Van Merriënboer en anderen spreken over de constructie van cognitieve schema's die behoren bij een bepaald vakgebied (Van Merriënboer, Clark & De Croock, 2002; Janssen-Noordman & Van Merriënboer, 2002, p. 67; Kirschner & Van Merriënboer, 2013, p. 33 en p. 140). Deze cognitieve schema's kunnen op hun beurt weer bestaan uit mentale modellen (Wat is het? Hoe is het georganiseerd? Hoe werkt het?) en cognitieve strategieën (Hoe pak je het aan?). Tijdens het werken aan de leertaken raken studenten vertrouwd met dit soort heuristische informatie die bij een vakgebied behoort (met name door er zelf naar op zoek te gaan) maar zij construeren ook hun eigen 'body of knowledge' die vervolgens gebruikt kan worden bij de uitvoering van andere, vergelijkbare taken ('transfer'). In die zin maken de mentale modellen en cognitieve strategieën onderdeel uit van het persoonlijke kennissysteem dat de student opbouwt en dat hierboven benoemd is als een van de resultaten van het Information Problem Solving-proces.

Schaap, De Bruijn, Van der Schaaf en Kirschner (2009) spreken in het kader van beroepsonderwijs over de constructie van Personal Professional Theories (PPT's), een begrip dat veel verwantschap heeft met wat hierboven een Persoonlijk Kennissysteem is genoemd. PPT's behoren bij een bepaald vakgebied en worden opgeslagen in het lange termijn geheugen van de lerende (Schaap et al., 2009, p. 484). Ze bestaan behalve uit professionele kennis (zowel declaratieve kennis als procedurele) uit een stelsel van gemeenschappelijke normen, waarden en overtuigingen die een rol spelen in het betreffende vakgebied (Schaap et al., 2009, p. 482). Voor zover mentale modellen en cognitieve strategieën bestaan uit formele kennis, kunnen ze worden opgedaan uit studieboeken, vakartikelen en instructies (Schaap et al., 2009, p. 485). Het is vooral dit type formele kennis dat tijdens het Information Problem Solving-proces door de lerenden wordt verworven. Ook de normen en waarden die bij een vakgebied horen, kunnen als expliciete kennis zijn vastgelegd in publicaties en instructies en kunnen dus tijdens het Information Problem Solving-proces worden opgedaan, maar bij verwerving dáárvan speelt ook de interactie met (andere) professionals een grote rol. Professionele normen, waarden en overtuigingen worden dan ook meer dan declaratieve en procedurele kennis verworven tijdens het leren op de werkplek zoals tijdens stages of in een werkend leren-traject (Schaap et al., 2009, p. 482).

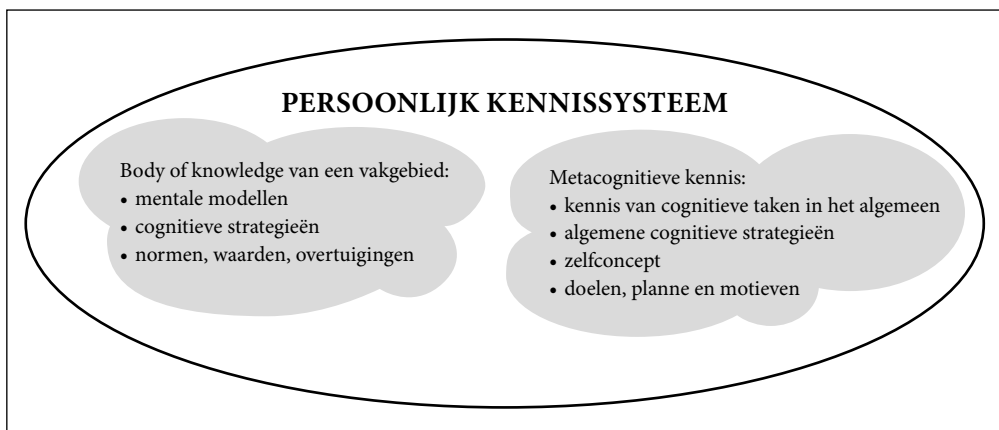
Behalve het onderscheid tussen drie soorten vakgerelateerde kennis maken Schaap et al. nog een tweede onderscheid, namelijk dat tussen de hierboven bedoelde formele kennis, werkproces gerelateerde kennis en praktische kennis. Werkproces gerelateerde kennis wordt volgens hen opgedaan tijdens het uitvoeren van professionele taken en komt tot stand door een symbiose van theoretische, formele kennis en ervaringskennis die wordt opgedaan tijdens de taakuitvoering (Schaap et al., 2009, p. 485). Het Information Problem Solving-model verwijst ook naar deze vorm van kennisverwerving, omdat het oplossen van informatieproblemen immers wordt beschouwd als een onderdeel van het uitvoeren van professionele taken. Praktische kennis daarentegen is de meest subjectieve en impliciete vorm van kennis die vooral wordt opgedaan in interactie met anderen. Volgens Schaap et al. heeft praktische kennis een situationeel karakter en maakt ze geen deel uit van iemands PPT (2009, p. 486).

De verschillende soorten kennis hebben alle betrekking op het vakgebied waarvoor de lerenden worden opgeleid. Van de auteurs van wie de modellen in paragraaf 2.3 zijn besproken, is het vooral Boekhorst die benadrukt dat het oplossen van informatieproblemen ook bijdraagt aan de verwerking van meer algemene kennis met betrekking tot het leren zelf (Boekhorst, 2000, pp. 102-103). Zijn suggestie dat oefening in het gebruik en hergebruik van informatie (evalueren, analyseren, parafraseren, synthetiseren) het 'leren leren' van de student bevordert, wordt door de onderwijspsycholoog Pintrich beaamd (2002, pp. 220-221). De laatste onderscheidde 3 soorten van zulke metacognitieve kennis (Pintrich, 2002):

- kennis van algemene cognitieve strategieën die niet bij een bepaald vakgebied horen (verschillende manieren om studietaken aan te pakken)
- kennis van (soorten) cognitieve taken in het algemeen
- kennis van eigen sterke en zwakke punten op het gebied van leren ('zelfconcept'; kennis van de eigen leerstijl).

De zelfde drie soorten metacognitieve kennis worden ook onderscheiden in de literatuurstudie van Alexander, Schallert en Hare (1991, pp. 328-330). Zij menen echter dat er nog een vierde component een rol speelt bij metacognitieve kennis, namelijk iemands plannen en doelen. In hun toelichting geven ze aan dat het daarbij meer om een affectieve dimensie gaat waarbij de persoonlijke motieven en overwegingen van de lerende een belangrijke rol spelen (p. 329). Het onderscheid met het Zelfconcept is dat bij iemands plannen, doelen en motieven heel andere overwegingen een rol kunnen spelen dan alleen de vraag hoe sterk iemand is in het uitvoeren van een bepaalde taak of activiteit (zie bijvoorbeeld ook Petty & Cacioppo, 1981). Dat deze affectieve component van kennis ook bij het Information Problem Solving-proces een belangrijke rol speelt wordt onder andere aangegeven door Walraven 2008 (p. 36) en in de eerder aangehaalde studie van Kuhlthau (1993).

Op basis van de hierboven besproken literatuur met betrekking tot kennisconstructie wordt in figuur 2.7 een schematische weergave gegeven van de elementen die deel uitmaken van een Persoonlijk Kennissysteem zoals dat door studenten wordt opgebouwd.



Figuur 2.7
Elementen waaruit een Persoonlijk Kennissysteem bestaat

De combinatie van het Information Problem Solving-model in figuur 2.6 en het Persoonlijk Kennissysteem in figuur 2.7 levert een compleet beeld op van activiteiten die in het kader van informatietaken worden uitgevoerd en de (nieuwe) kennis die daarmee wordt opgebouwd.

Het IPS-model uit figuur 2.6 functioneert in de volgende hoofdstukken als een conceptueel kader voor het onderzoek naar de wijze waarop de informatievaardigheid van studenten kan worden gemeten en de vraag hoe die meetinstrumenten kunnen bijdragen aan het leren zelf. Zoals al eerder bleek is de inhoud van het Persoonlijk Kennissysteem meer object van onderzoek in de onderwijs- en leerwetenschappen dan in de informatiewetenschappen. Daarmee valt het buiten de reikwijdte van het eigen onderzoek. Om een volledig beeld te schetsen van het verschijnsel informatievaardigheid en de impact die dat kan hebben op het leren in het hoger onderwijs is het Persoonlijk Kennissysteem echter wel onderdeel geweest van het theoretische onderzoek dat in dit hoofdstuk is besproken.

2.6 Conclusies en discussie

Wetenschappelijk onderzoek naar informatievaardigheden kent diverse benaderingen waarvan er hier drie zijn besproken. Van deze drie is de uitkomstgerichte gedragsbenadering in ieder geval passend voor de beantwoording van de centrale onderzoeksvraag uit hoofdstuk 1 omdat deze benadering goed aansluit bij het streven om een beoordelingsinstrument te ontwikkelen. De kern van de uitkomstgerichte gedragsbenadering ligt immers in het formuleren van leeruitkomsten, een belangrijke stap bij het ontwikkelen van toetsen en beoordelingsinstrumenten. De tweede benadering, de constructivistische procesgerichte benadering, bevat echter elementen die beter dan de uitkomstgerichte gedragsbenadering aansluiten bij de andere eisen die we in hoofdstuk 1 hebben gesteld aan het te ontwikkelen beoordelingsinstrument. Daarmee wordt onder andere gedoeld op het uitgangspunt om het volledige proces van het oplossen van informatieproblemen centraal te stellen en de aandacht niet te richten op afzonderlijke kenniselementen of deelvaardigheden. Ook sluit de constructivistische procesgerichte benadering goed aan bij het streven om een beoordelingsinstrument te ontwikkelen dat ook ingezet kan worden als een 'tool for learning' en dus tevens een functie heeft om de *verwerving* van informatievaardigheden te bevorderen. Voor het onderhavige onderzoek zal daarom een benadering worden gekozen die haar wortels heeft in de uitkomstgerichte gedragsbenadering maar waarin de aansprekende elementen van de constructivistische procesgerichte benadering toch voldoende tot hun recht komen. Deze keuze zal met name een rol spelen in hoofdstuk 3 waarin verschillende methodes en instrumenten voor het meten van informatievaardigheden de revue zullen passeren en in hoofdstuk 4 waarin wordt beschreven hoe een concreet instrument wordt ontwikkeld.

De derde benadering die in paragraaf 2.1 is besproken, de relationele benadering die in fenomenografisch onderzoek wordt gebruikt, is als aanpak voor het onderhavige onderzoek verworpen omdat die te beschrijvend is en onvoldoende aansluit bij het centrale onderwerp van het proefschrift, de ontwikkeling van een *beoordelingsinstrument* voor informatievaardigheden.

Een tweede onderwerp in dit hoofdstuk was de vergelijking van verschillende ‘modellen’ die worden gehanteerd om het proces van het oplossen van informatieproblemen (“Information Problem Solving”) te visualiseren. In het kader van het onderhavige onderzoek is op basis van vier van deze IPS-modellen een eigen model ontwikkeld. Dit model functioneert in de vervolghoofdstukken als een raamwerk voor het nemen van ontwerpbeslissingen en de analyse van onderzoeksgegevens. Het informatievaardige gedrag dat in het kader van het te ontwikkelen beoordelingsinstrument wordt geformuleerd, dient aan te sluiten bij de deelactiviteiten die in het model zijn benoemd.

Het proces van Information Problem Solving is in het onderwijs echter niet alleen een doel op zich. Verschillende auteurs beroepen zich erop dat het oplossen van informatieproblemen ook helpt bij het construeren van nieuwe kennis. Het argument dat Information Problem Solving helpt bij het opbouwen van een ‘Persoonlijk Kennissysteem’ wordt door verschillende auteurs gehanteerd maar vooralsnog ontbreekt in de informatiewetenschappelijke literatuur een heldere operationalisering daarvan. Op basis van literatuur uit de onderwijs- en leerwetenschappen is er een voorstel voor gedaan. Deze operationalisering van het Persoonlijk Kennissysteem voegt aan de bestaande literatuur toe dat een vollediger beeld wordt geschetst van de relevantie en de ‘scope’ van informatievaardigheden.

Meten van informatievaardigheden⁸

Abstract: In de wetenschappelijke literatuur over methodes en instrumenten voor het beoordelen van informatievaardigheden, wordt onderscheid gemaakt tussen vragenlijsten (tests en enquêtes) en diverse vormen van performance assessment (studieopdrachten, portfolio's en directe observatie). Als beoordelingsinstrument wordt bij studieopdrachten en portfolio's vaak gebruik gemaakt van zogenaamde scoringsrubrieken.

Op basis van de gevonden literatuur wordt geconcludeerd dat voor het beoordelen van het individuele vermogen om informatieproblemen op te lossen, vragenlijsten te veel gericht zijn op kennis en begrip en te weinig op de authentieke taakuitvoering. Het gebruik van studieopdrachten in combinatie met het gebruik van een scoringsrubriek als beoordelingsinstrument, heeft voor dergelijke performance assessments de voorkeur.

⁸ De inhoud van dit hoofdstuk is eerder gepubliceerd in: A scoring rubric for performance assessment of information literacy in Dutch Higher Education. *Journal of Information Literacy*, 4 (1), 22-39 (Van Helvoort, 2010).

Onderzoeksvraag en –aanpak voor hoofdstuk 3

De onderzoeksvraag die in dit hoofdstuk wordt behandeld luidt: Wat is de meest geschikte assessmentmethode voor het beoordelen en bevorderen van informatievaardigheden? Om deze onderzoeksvraag te beantwoorden is een literatuurstudie uitgevoerd waarbij eerst de vraag wordt beantwoord welke assessmentmethodes er überhaupt bestaan voor het beoordelen van informatievaardigheden van studenten in het hoger onderwijs. Voor de keuze van het meest geschikte beoordelingsinstrument zijn de criteria gebruikt die ook in hoofdstuk 2 werden genoemd bij de bespreking van de wetenschappelijke benaderingen en modellen:

- Het te ontwikkelen beoordelingsinstrument dient betrouwbare en valide beoordelingen op te leveren.
- Het instrument dient, conform de procesgerichte constructivistische benadering, zowel geschikt te zijn voor summatieve beoordeling van informatievaardigheden als ingezet kunnen worden als formatief beoordelingsinstrument dat leidt tot bevordering van informatievaardigheden.
- Het instrument dient zich te richten op de uitvoering van complete informatietaken zoals weergegeven in het IPS-model en niet op afzonderlijke kenniselementen.
- Het instrument dient generiek toepasbaar te zijn.

Uitgangspunt bij het literatuuronderzoek is een sleutelpublicatie geweest van Gratch-Lindauer (2003). De door haar genoemde literatuur is opgespoord, bestudeerd en verwijzingen vanuit die publicaties werden nagetrokken. De resultaten werden aangevuld met die uit keyword searches in LISA, LISTA, ERIC en Web of Science. Deze keyword searches werden afgesloten in februari 2010 en hebben daarom betrekking op publicaties uit het jaar 2009 of eerder. In januari 2013 is een aantal titels opnieuw opgespoord om de bibliografische gegevens te controleren.

De gebruikte query voor de aanvullende keyword searches bestond uit drie componenten die kunnen worden samengevat als “het beoordelen van informatievaardigheden in het hoger onderwijs”. Voor het begrip Informatievaardigheden is de term “Information Literacy” gebruikt. Deze term is in alle vier de bibliografische zoeksystemen gebruikt als ‘Subject’ (LISTA), ‘Topic’ (Web of Science) of als descriptor uit een thesaurus (LISA en ERIC). Voor de component ‘beoordelen’ is de zoekterm “evaluation” gebruikt, wat de meest algemene aanduiding is voor het beoordelen van studenten en hun werk. Binnen LISA, Web of Science en ERIC kon met de term “evaluation” als descriptor uit een thesaurus respectievelijk als ‘Topic’ worden gezocht, binnen LISTA is met deze term in alle velden gezocht wat uiteraard leidde tot een aantal gevonden documenten dat na lezing van het abstract minder relevant bleek te zijn. Voor de beperking tot het hoger onderwijs is in alle vier de zoeksystemen gebruik gemaakt van de zoekfrase “higher education”. In Eric kon deze frase gebruikt worden als descriptor uit de thesaurus, in Web of Science als ‘Topic’ en in LISA en LISTA werd de frase toegepast op ‘alle velden’. De drie componenten zijn in alle gevallen verbonden met Booleaanse AND operatoren. De query werd bovendien beperkt tot peer-reviewed publicaties uit de periode 2000-2009.

In totaal zijn 37 relevante publicaties gevonden. Tabel 3.1 geeft verwijzingen naar alle publicaties die gebruikt zijn voor de samenstelling van onderstaand overzicht van beoordelingsinstrumenten.

De volledige bibliografische gegevens zijn terug te vinden in de referentielijst. Niet opgenomen zijn verwijzingen naar publicaties over scoringsrubrieken en onderwijskundige onderwerpen in het algemeen. Die komen, indien van toepassing, wel voor in de afsluitende referentielijst.

Tabel 3.1

Literatuur over het beoordelen van informatievaardigheden in het hoger onderwijs, periode 2000-2009

Brown en Gratch-Lindauer, 2004	Gross en Latham, 2007	Oakleaf, 2008 en idem 2009
Buchanan, Luck en Jones, 2002	Hignite, Margavio en Margavio, 2009	Radcliff, Jensen, Salem Jr., Burhanna en Gedeon, 2007
Burke, Germain en Xu, 2005	Hook, Bracke, Greenfield en Mills, 2003	Rockman, 2002
Burkhardt, Kinnie en Cournoyer, 2008	Ivanitskaya, O'Boyle en Casey, 2006	Scharf, Elliot, Huey, Briller en Joshi, 2007
Cameron, Wise en Lottridge, 2007	Knight, 2006	Sonley, Turner, Myer en Cotton, 2007
Cherry, Fiegen en Watson, 2002	Kurbanoglu, Akkoyunlu en Umay, 2006	Stoffberg en Blignaut, 2008
Cochrane, 2006	Lazonder, 2005	Streatfield en Markless, 2008
Diller en Phelps, 2008	Mittermeyer en Quirion, 2003	Thirion en Pochet, 2009
Dunn, 2002	Monoi, O'Hanlon en Diaz, 2005	Timmers en Glas, 2010
Emmett en Emde, 2007	Neely, 2006	Walsh, 2009
Fourie en Van Niekerk, 1999 en idem 2001	Novotny en Cahoy, 2006	Webber en Johnston, 2000
Gratch-Lindauer, 2003	Nutefall, 2005	

De sleutelpublicatie van Gratch-Lindauer (2003) onderscheidt verschillende soorten instrumenten om de leeruitkomsten van studenten op het gebied van informatievaardigheid te meten: tests, quizzen, studieopdrachten, rechtstreekse observaties, enquêtes, logboeken en portfolio's. Een tweede literatuuroverzicht dat pas in latere instantie via de keyword search werd opgespoord (Walsh, 2009) komt met een vergelijkbare indeling zij het dat daarin voor de studieopdrachten nog onderscheid wordt gemaakt tussen "analyse van bibliografieën" en "essays". Tevens onderscheidt hij nog "simulaties" en "final grades" maar van deze beide laatste merkt hij op dat er maar weinig studies zijn die rapporteren over het gebruik van deze methodes voor het meten van informatievaardigheden (Walsh, 2009, p. 25).

Gratch-Lindauer verdeelt de verschillende meetinstrumenten en -methodes die zij onderscheidt over drie leerdomeinen die teruggrijpen op de taxonomie van Bloom (1956):

- het cognitieve domein: wat de studenten weten of begrijpen.
- het uitvoeringsdomein ("performance based"): wat de studenten daadwerkelijk doen. Dit is het leerdomein dat door Bloom wordt aangeduid als het "psychomotor domain" (Bloom, 1956, p. 7).
- het affectieve domein: hoe de studenten hun eigen bekwaamheid ervaren.

Elk van de meetinstrumenten en –methodes wordt volgens Gratch-Lindauer vooral toegepast binnen één specifiek leerdomein (tests bijvoorbeeld in het cognitieve domein, studieopdrachten en observaties in het uitvoeringsdomein en enquêtes en logboeken in het affectieve domein) maar zij benadrukt dat de meeste ook toegepast kunnen worden binnen méér dan één van de leerdomeinen. Als voorbeeld noemt ze tests die ook items kunnen bevatten die vragen naar het daadwerkelijke gedrag van studenten (pp. 27-28).

In de volgende paragrafen volgt een inhoudelijke bespreking van diverse soorten instrumenten en beoordelingsmethodes zoals de auteurs die in de literatuur die is opgespoord, behandelen. Voor de beoordelingsmethodes en meetinstrumenten die door Gratch-Lindauer worden aangeduid als behorende tot het ‘uitvoeringsdomein’ wordt in het overzicht de aanduiding ‘performance assessment’ gebruikt omdat deze benaming het beste aansluit bij de internationale terminologie die wordt gehanteerd in het onderwijswetenschappelijke onderzoek.

3.1 Vragenlijsten

In navolging van Gratch-Lindauer (2003) wordt onderscheid gemaakt tussen tests en quizen aan de ene kant en enquêtes aan de andere kant. Deze drie assessmentvormen worden wel in één hoofdpagina behandeld omdat ze alle drie gebruik maken van schriftelijke of elektronische vragenlijsten.

Het onderscheid tussen tests en quizen is gelegen in het feit dat er bij het eerste type assessment een beoordeling wordt uitgesproken over de (meestal) cognitieve vaardigheden van een individuele student (hetzij summatief, hetzij formatief of diagnostisch) terwijl bij quizen het accent ligt op een spelelement dat de student zou moeten aanzetten tot leren. Uiteraard kan de student bij een quiz ook individuele feedback ontvangen maar in dat geval worden de individuele prestaties op de quiz niet gebruikt door de docent voor het uitspreken van een beoordeling of het inrichten van (vervolg)onderwijs. Omdat het onderhavige onderzoek zich uitdrukkelijk wel richt op een assessmentmethode die ook gebruikt kan worden voor individuele beoordelingen zal aan quizen in het vervolg geen aandacht besteed worden. Wel is het zo dat tests over het algemeen ook als een quiz in het onderwijs kunnen worden ingezet.

Bij enquêtes wordt vooral gevraagd naar het gedrag van de student maar dit wordt niet (zoals bij de nog te bespreken vormen van performance assessment) op een directe manier waargenomen. Terwijl tests betrekking hebben op het cognitieve domein, worden enquêtes in de terminologie van Gratch-Lindauer gerekend tot het affectieve domein. Ze geven inzicht in de manier waarop de studenten hun eigen informatievaardige gedrag ervaren. Het gevaar van sociaal wenselijke antwoorden is bij enquêtes echter niet denkbeeldig, wat deze vorm, net als quizen, ongeschikt maakt voor het vaststellen van individuele beoordelingen.

3.1.1 Tests

De schriftelijke test is het instrument dat in het kader van het meten van informatievaardigheden waarschijnlijk het vaakst wordt gebruikt. Neely (2006) noemt in een bijlage circa 70 voorbeel-

den van assessmentinstrumenten waarvan de meeste de vorm van zo'n schriftelijke test hebben. Overigens zijn ze in de meeste gevallen niet beschreven in gepubliceerde vakliteratuur maar worden ze (vaak tegen betaling) beschikbaar gesteld op het World Wide Web (Neely 2006, p. 158).

De vorm van de vragen is in de meeste gevallen die van meer keuze-vragen gecombineerd met waar/onwaar-vragen en combinatievragen ('matching items'). De items uit een test liggen vrijwel altijd op het cognitieve gebied: het gaat om wat studenten weten of kunnen beredeneren bij zogenaamde toepassingsvragen. Het onderscheid met enquêtes is dat bij de laatste vorm meestal wordt gevraagd naar opinies, houdingen of het gangbare gedrag. Voor het scoren daarvan wordt vaak een Likertschaal gebruikt.

Tests kunnen zowel worden gebruikt voor het beoordelen van individuele studenten als voor het beoordelen van groepen studenten ('cohorts'). In het geval van individuele beoordelingen kunnen uiteraard totaalscores voor het betreffende cohort worden berekend en geanalyseerd, bijvoorbeeld door de spreiding en gemiddeldes te bepalen.

De populariteit van tests - en wellicht ook die van enquêtes - heeft te maken met het feit dat de antwoorden eenvoudig te verwerken zijn (Walsh, 2009, p. 21). Andere voordelen van tests zijn (Oakleaf, 2008, pp. 235-236):

- de resultaten van afzonderlijke groepen studenten kunnen gemakkelijk met elkaar worden vergeleken
- door middel van een pre- en post-test kan de voortgang van studenten op het gebied van informativaardigheden worden vastgesteld
- mits de lijst met items goed is samengesteld kan er een hoge mate van betrouwbaarheid bereikt worden.

Amelie Brown en Bonnie Gratch-Lindauer wijzen er echter op dat juist dat laatste veel inspanning kost en dat daarvoor een behoorlijke kennis op het gebied van toetstheorie is vereist (Brown & Gratch-Lindauer 2004).

Een alternatief voor zelf ontwikkelde tests is het gebruik van standaard tests. Het bekendste voorbeeld daarvan voor het hoger onderwijs ('universities' en 'colleges') zijn de standaard tests die binnen het project SAILS van de Amerikaanse Kent State University zijn ontwikkeld. Universiteiten kunnen twee soorten tests inkopen, individuele en cohort tests. De kosten zijn afhankelijk van het aantal deelnemende studenten. Op de website van het project SAILS is een kleine set met 'sample questions' beschikbaar (SAILS, 2012). Project SAILS is uitdrukkelijk gericht op de Amerikaanse markt maar men heeft Engelstalige versies in voorbereiding die zich richten op een internationale markt. In de internationale versies zullen geen items worden opgenomen die verwijzen naar een situatie in een specifiek land, bijvoorbeeld op het gebied van auteursrecht (persoonlijke correspondentie per mail d.d. 26 november 2012).

In de gevonden literatuur had één vakpublicatie betrekking op het *ontwikkelp*roces van zo'n 'information literacy test' (Cameron et al., 2007). In het desbetreffende artikel wordt vooral ingegaan op de wijze waarop de betrouwbaarheid en de validiteit van de test als geheel en die van de afzonderlijke items zijn vastgesteld. De betreffende test (ILT) werd aanvankelijk voor lokaal gebruik

ontwikkeld maar wordt inmiddels ook als standaard test op commerciële basis geëxploiteerd door Madison Assessment, een verzelfstandigd onderdeel van James Madison University.

Zoals hierboven vermeld geeft Gratch-Lindaur (2003) aan dat tests vooral worden toegepast op het cognitieve domein. Voor zover dat kan worden nagegaan geldt dat ook voor zowel de SAILS test als de Information Literacy Test van Madison Assessment. Als voorbeeld volgt hieronder een afbeelding van een van de vragen uit de 'sample questions' die SAILS op zijn webpagina beschikbaar stelt. Het betreft een typisch voorbeeld van een reproductievraag.

Sample Question Three

What is a term used to describe what a book or journal article is about?

CHOOSE ONE ANSWER

- ☐ Bibliography
- ☐ Keyword
- ☐ Library catalog
- ☐ Research database
- ☒ Subject heading

Figuur 3.1
Voorbeeld uit de 'sample questions' van SAILS (2012)

Tests op een specifiek vakgebied

Ivanitskaya, O'Boyle en Casey ontwikkelden een online standaard test ("Research Readiness Self-Assessment" - RRSA) voor het meten van de informatievaardigheden van "health information consumers" (Ivanitskaya et al., 2006). Hun testitems komen alle uit het cognitieve domein maar hebben behalve op 'weten' en 'begrijpen' ook betrekking op 'toepassing' en analytische vaardigheden op cognitief gebied. Een voorbeeld van een item dat betrekking heeft op toepassing en analyse wordt gegeven in afbeelding 3.2. De RRSA is weliswaar een gestandaardiseerde test maar heeft wel betrekking op het specifieke vakgebied "health information".

16 - You have found a good reference for your research paper. It is a book edited by Williams and others. The book title is "The epidemiology of diabetes mellitus: An international perspective." Find it by searching the library catalog: <http://catalog.lib.cmich.edu/> What is the CALL NUMBER of this book?

- ☐ OA 645.D5 E654
- ☐ PA 645.D5 E654
- ☐ RA 645.D5 E654
- ☐ SA 645.D5 E654
- ☐ TA 645.D5 E654

Figuur 3.2
Testitem nummer 16 uit het "Research Readiness Self-Assessment" (Ivanitskaya et al., 2006)

Een alternatief voor standaard tests is de ontwikkeling van een test die nauw gelieerd is aan één bepaalde cursus. Cherry et al. (2002) beschrijven de ontwikkeling van zo'n test voor een Business Curriculum. Het type vragen dat gebruikt wordt is ook in hun geval sterk gericht op het cognitieve domein maar de testitems hebben wel een toepassingsgericht karakter en lijken goed te passen bij het vakgebied waarvoor de betreffende cursus opleidt. De constructie van een dergelijke test vereist echter een innige samenwerking van informatieprofessionals en docenten. Figuur 3.3 geeft een voorbeeld uit de door hen ontwikkelde test.

5. Books, articles, newspapers, statistics as well as free and fee-based Web sites on the Internet all have a place in business research. From the list of topics below identify what you think is the best format or media type (e.g. books, newspaper articles, etc.) for locating information on that topic. Explain where you would go to get this type of information and why you would go there. Where would you find information on ...

Example:

Recent layoff announcements from across the country

Possible answer:

Popular press magazines (like *Time*, *Newsweek*, and *Business Week*)

- a. Current examples of industry-specific practices

Answer:

- b. Company/Organization press releases

Answer:

- c. Employment laws and regulations

Answer:

Figuur 3.3

Voorbeeld van een testitem uit de lokaal ontwikkelde test van Cherry et al. (2002)

3.1.2 Enquêtes

Enquêtes hebben vaak de vorm van een vragenlijst waarin bij ieder item een bepaalde vorm van gedrag wordt beschreven en waarbij de respondenten vervolgens op een Likertschaal aangeven hoe vaak zij dit gedrag vertonen. In de vakliteratuur worden ze vaak aangeduid als “scales”. Kurbanoglu, Akkoyunlu en Umay (2006) beschrijven de constructie van een “information literacy self-efficacy scale”. De geconstrueerde ‘scale items’ werden door hen vervolgens getest op betrouwbaarheid en validiteit. Self-efficacy is te beschouwen als het vertrouwen dat mensen hebben in hun eigen vermogen om een bepaalde taak te kunnen uitvoeren (Kurbanoglu et al., 2006, p. 731). De ‘self-efficacy scale’ van Kurbanoglu et al. heeft duidelijk betrekking op wat Gratch-Lindauer het affectieve domein noemt. Het type vragen dat wordt gesteld leent zich niet om studenten individueel te beoordelen. Daarvoor is de kans te groot dat studenten de in hun ogen door de docent gewenste antwoorden gaan geven. Een tweede aanwijzing voor het feit dat de ‘scale’ van Kurbanoglu et al.

niet de pretentie heeft om studenten te beoordelen, is het feit dat de groep proefpersonen waarmee de ‘scale’ werd getest niet bestond uit studenten maar uit onderwijsgevendenden op diverse niveaus van het onderwijssysteem in Turkije.

Van het instrument dat Kurbanoglu et al. ontwikkelden is zowel een concept versie als de uiteindelijke versie opgenomen als bijlage bij hun publicatie. Hun ‘scale’ is dus beschikbaar voor gebruik in het onderwijs.

Een studie die qua opzet vergelijkbaar is met die van Kurbanoglu et al. is die van Timmers en Glas (2010). Zij ontwikkelden een enquête naar het informatie zoekgedrag van studenten aan de Saxion Hogeschool en de Universiteit Twente. Ook in hun geval is een Engelstalige versie van de definitieve vragenlijst beschikbaar als bijlage bij de publicatie. De oorspronkelijke Nederlandstalige vragenlijst bestaat uit 46 items met vier antwoordmogelijkheden op een Likertschaal (“altijd, vaak, soms, bijna nooit”), aangevuld met een vijfde optie om aan te geven dat de respondent niet bekend is met het onderwerp in het betreffende item (“Wat is dat?”). De items hebben allemaal betrekking op het vertoonde gedrag (bijvoorbeeld: “Als ik informatie zoek voor een studie-opdracht... maak ik gebruik van Wikipedia”) en konden anoniem worden beantwoord. Een voorbeeldvraag is weergegeven in figuur 3.4. Het item in het voorbeeld is ontleend aan de betreffende tijdschriftpublicatie en is ten behoeve daarvan vertaald naar het Engels.

Gezien de aard van de items en het risico op sociaal wenselijke antwoorden als deze niet anoniem worden beantwoord, geldt ook voor de enquête van Timmers en Glas dat deze niet gebruikt kan worden voor het beoordelen van individuele studenten.

When I search information for study assignments ...	always	often	sometimes	seldom or never	what is that?
I use study materials handed by my teacher	☐	☐	☐	☐	☐
I go to the library	☐	☐	☐	☐	☐
I consult librarians	☐	☐	☐	☐	☐
I consult the library catalogue	☐	☐	☐	☐	☐
I consult databases	☐	☐	☐	☐	☐

Figuur 3.4

Voorbeeld van een item uit de enquête van Timmers en Glas (2010)

Een voorbeeld van een enquête waarbij niet gevraagd werd naar ‘gedrag’ maar naar opinies, wordt gegeven door Cochrane (2006, pp. 106-112). Voor verschillende “information literacy skills” werd studenten gevraagd om een score te geven op een schaal van 1 tot en met 5 voor zowel de moeilijkheidsgraad, als voor het ervaren ‘nut’ (“usefulness”) en de mate waarin de student zichzelf als ‘competent’ ervaart. Bij deze opzet lijkt nog veel meer dan bij enquêtes die vragen naar gedrag, het gevaar aanwezig dat de uitkomsten worden beïnvloed door hetgeen de respondenten ervaren als ‘sociaal gewenst’ (zie ook Cochrane, 2006, p. 116). Het is bovendien een algemeen bekend verschijnsel dat

studenten hun eigen vaardigheden op het gebied van informatiegebruik, en met name die op het gebied van online zoeken, verkeerd inschatten (zie bijvoorbeeld Monoi et al., 2005, p. 102).

Thirion en Pochet (2009) gebruikten een test van 20 items voor het meten van de informatievaardigheid van eerstejaars studenten in Franstalig België. Zij maakten daarvoor gebruik van een instrument dat door Mittermeyer en Quirion (2003) was ontwikkeld voor Franstalig Canada (Québec). Hoewel de antwoorden op de test van Thirion en Pochet wel degelijk ‘goed’ of ‘fout’ werden gerekend kan hun vragenlijst (en dus ook die van Mittermeyer en Quirion) toch eerder als een enquête dan als een echte test worden beschouwd. De vragenlijst werd anoniem ingevuld en de resultaten werden gebruikt voor een onderzoek naar het daadwerkelijke niveau van de studenten die instromen in het Franstalige hoger onderwijs in België. Precies zoals bij de eerder besproken enquêtes geldt dat de waarde van de gegevens die zij verzamelden, afhankelijk is van de mate waarin de respondenten op een eerlijke en open manier antwoord hebben gegeven en de mate waarin ze gemotiveerd waren om mee te doen aan het onderzoek (zie ook Gross & Latham, 2007, p. 349).

3.2 Performance assessments

Een veel voorkomend vraagstuk in de opgespoorde literatuur heeft betrekking op de keuze tussen klassieke tests of enquêtes zoals die zijn besproken in paragraaf 3.1 en verschillende vormen van performance assessments. In een van de oudste artikelen die over assessment van informatievaardigheden werd gevonden (Rockman, 2002) wordt al geclaimd dat tests en surveys “[...] do not demonstrate how well a student has actually learned to navigate through a search strategy process to find, evaluate, use and apply information to meet a specific need” (p. 193). Rockmans bezwaren tegen het gebruik van “objective tests” voor het beoordelen van de informatievaardigheden van studenten komen overeen met de conclusies van veel onderwijshervormers die opmerkten dat het toenemende gebruik van gemakkelijk te verwerken standaard tests en andere vormen van toetsen met behulp van computerprogramma’s, hebben geleid tot een verminderde aandacht voor het aanleren en beoordelen van complexe samengestelde vaardigheden (zie bijvoorbeeld Frederiksen, 1984, p. 195).

In een meer recent artikel over het beoordelen van informatievaardigheden bevestigt Megan Oakleaf de bovengenoemde bezwaren tegen wat zij noemt “Fixed-Choice Tests”. Oakleaf gaat daarbij in op de theoretische achtergrond van de klassieke testcultuur (het “scientific measurement”) waarbij complexe leertaken worden opgedeeld in verschillende deeltaken die afzonderlijk worden onderwezen, geoefend en getoetst (Oakleaf, 2008, p. 234). Zoals in paragraaf 3.1 al is aangegeven onderkent Oakleaf dat het gebruik van dergelijke tests grote voordelen heeft maar ze merkt op dat het ook veel beperkingen kent. Als belangrijkste nadelen van het gebruik van dergelijke tests voor het meten van informatievaardigheid noemt ze (Oakleaf, 2008, pp. 237-238):

- Ze zijn gericht op afzonderlijke onderdelen van het concept informatievaardigheid en niet op het volledige complexe ‘construct’. Doordat het begrip informatievaardigheid verwijst naar zo’n complex samengesteld construct is er een reëel gevaar dat de test niet het volledige begrip representeert.

- Tests creëren een onnatuurlijke situatie waardoor niet wordt gemeten hoe de lerende (de testpersoon) zich ‘in het echt’ zou gedragen. De testsituatie creëert met andere woorden een gebrek aan authenticiteit. Tests hebben volgens Oakleaf het gevaar dat ze te veel de nadruk leggen op het ‘knowing what’ en te weinig op het ‘knowing how’. Daardoor ontstaat het risico dat studenten ‘voor de test gaan leren’.

Hoewel Oakleaf zich in het betreffende artikel met haar kritische bespreking beperkt tot beoordelingen met behulp van tests, mag worden aangenomen dat haar opmerkingen volgens haar ook opgaan voor andere min of meer gestandaardiseerde meetinstrumenten zoals enquêtes. Ook in de besproken voorbeelden van enquêtes zijn de vragen immers meer gericht op afzonderlijke onderdelen van informatievaardigheid dan op het totale construct. Bovendien hoeven de antwoorden van respondenten bij een enquête niet overeen te komen met hun daadwerkelijke gedrag.

Een derde auteur die uitdrukkelijk bezwaar maakt tegen het gebruik van tests of vragenlijsten voor het beoordelen van een complexe vaardigheid is Lorrie Knight. In een artikel uit 2006 merkt zij op dat er een trend gaande is “toward authentic assessment, a process that measures how students apply their knowledge to real-time tasks. [...] Authentic assessment is a promising method for the evaluation of information literacy learning outcomes, as it measures not only what students learn through library instruction, but also how the learning is subsequently incorporated in their academic work” (Knight, 2006, p. 45). Het door Knight gehanteerde concept van “authentic assessment” heeft betrekking op de zelfde assessmentvormen als de “performance based” instrumenten die worden onderscheiden door Gratch-Lindauer (2003, pp. 29-30) en Oakleaf (2008, p. 240 ff.) zoals studieopdrachten, schrijfopdrachten en de verzorging van presentaties. Ze hebben stuk voor stuk het karakter van een ‘complexe taak’. Oakleaf (2008, p. 240) licht toe dat dergelijke assessments een achtergrond hebben in de constructivistische leertheorieën die ervan uitgaan dat een assessment niet alleen bedoeld is voor het beoordelen van leeruitkomsten maar ook een functie hebben als een ‘tool for learning’: “students should learn by completing an assessment” (Oakleaf 2008, p. 241). Performance assessments zijn in deze visie - meer dan tests, enquêtes of quizen - geschikt om aan te zetten tot leren. Bovendien zouden ze, zoals eerder aangegeven, beter geschikt zijn voor het beoordelen van complexe en abstracte vaardigheden (‘higher order skills’) zoals informatievaardigheid (Oakleaf, 2008, p. 242-243; Scharf et al., 2007, p. 462). Een bijkomend voordeel van dergelijke ‘authentic’ of ‘performance based’ assessmentvormen zou zijn dat ze zijn ingebed in een realistische context, wat bevordert dat de verworven vaardigheden kunnen worden getransfereerd naar andere studie- of werksituaties (Oakleaf, 2008, p. 243). In de paragrafen 3.2.1 tot en met 3.2.4 wordt stilgestaan bij concrete vormen die zo’n ‘performance assessment’ kan aannemen.

3.2.1 Studieopdrachten

Studieopdrachten of ‘assignments’ die zich goed lenen om de informatievaardigheid van studenten te beoordelen zijn presentaties, essays, onderzoeksverslagen (schriftelijk of in een mondelinge presentatie), eindschrijvingen en (wellicht in iets mindere mate) stageverslagen (Gratch-Lindauer, 2003, p. 30; Walsh, 2009, pp. 22-25). Deze opdrachten hebben gemeenschappelijk dat er in veel gevallen externe informatie (van het internet of uit een bibliotheek) dient te worden verzameld, beoordeeld en te worden verwerkt in een eigen kennisproduct. In de vakliteratuur zijn enkele voorbeelden

gevonden van casussen waarbij de beoordeling van informatievaardigheid op dit type opdrachten werd toegepast. Scharf et al. (2007) beschrijven een project waarbij de beoordeling van een portfolio met schrijfopdrachten werd uitgebreid met beoordelingscriteria op het gebied van informatievaardigheid. Zij maakten daarvoor gebruik van een zelf ontwikkelde scoringsrubriek (zie paragraaf 3.2.4) waarvan bovendien de interbeoordelaarsbetrouwbaarheid en de interne consistentie werden getest. Hun artikel is daarmee een van de meest aansprekende voorbeelden van performance assessment van informatievaardigheden.

Knight beschrijft een project waarbij één specifiek assignment, de samenstelling van een geannoteerde bibliografie als onderdeel van een uitgebreidere 'course', werd beoordeeld met een speciaal voor dat doel ontwikkelde scoringsrubriek. In dit geval kunnen vraagtekens geplaatst worden bij de authenticiteit van de opdracht aangezien niet geheel duidelijk is of het betreffende assignment voldoende was ingebed in het grotere geheel van de betreffende cursus.

De twee voorbeelden van 'performance assessment' die hiervoor zijn beschreven, hebben allebei betrekking op de beoordeling van eindproducten. Jennifer Nutefall (2005) beschrijft een casus waarbij de nadruk ligt op de beoordeling van het onderzoeksproces. In haar voorbeeld kregen deelnemers aan een "Oral Communication and Information Literacy Class" onder andere de opdracht om een presentatie voor te bereiden én een opdracht om het onderzoeksproces tijdens de voorbereiding van die presentatie vast te leggen in een "Paper Trail", een beschrijving van het doorlopen zoekproces waarin bovendien uitdrukkelijk aandacht besteed diende te worden aan de reflectie daarop (Nutefall, 2005, p. 93). De instructie aan de studenten voor de 'Paper Trail' en de beoordelingscriteria die werden gehanteerd zijn als een bijlage opgenomen in haar artikel. Ook Nutefalls publicatie is daarmee een duidelijk voorbeeld van toepassing van performance assessment bij een studieopdracht maar in haar geval wordt helaas geen aandacht besteed aan de betrouwbaarheid van de beoordelingen, een van de vaak genoemde risico's bij deze vorm van assessment.

Webber en Johnston (2000) gaan nog een stap verder op het pad van reflectie. Zij beschrijven de opzet en uitvoering van een speciale "information literacy class" die werd aangeboden aan studenten van de Business School van de University of Strathclyde, waarbij studenten expliciete opdrachten kregen om te reflecteren over de aard en het belang van het begrip "information literacy" zelf. In hun geval beperkt zich dat niet tot wat we in hoofdstuk 2 het Information Problem Solving-proces hebben genoemd maar wordt uitdrukkelijk geprobeerd om de studenten door middel van de reflectie-opdrachten op het niveau te krijgen dat ze informatievaardigheid gaan ervaren als het vermogen om zelfstandig te leren en kritisch te denken (Webber en Johnston, 2000, pp. 390-391). Patrick Tompkins (2007) gebruikt een zelfde soort opzet met veel aandacht voor een "metacognitive essay", zij het dat het "information-literacy curriculum" in zijn geval onderdeel uitmaakt van een "College Composition class" waarin studenten gezamenlijk onderzoek doen en individueel een aantal "writing assignments" uitvoeren. Behalve het eerder genoemde "metacognitive essay" zijn dat

- het bijhouden van een project logboek,
- het schrijven van een aantal abstracts
- en een uitgebreide samenvatting (met bibliografische verwijzingen) van een of meer informatiebronnen die in het kader van het onderzoeksproject werden opgespoord.

Van de vijf hierboven besproken casussen zijn er drie die betrekking hebben op projecten waarbij docenten en informatiespecialisten nauw samenwerkten, namelijk die van Scharf et al. (2007), Knight (2006) en Nutefall (2007). Die samenwerking tussen onderwijsstaf en informatiespecialisten is ook in paragraaf 3.1.1 genoemd voor de tests die lokaal en voor specifieke cursussen of opleidingen werden ontwikkeld. De ervaring is dat informatiespecialisten wel veel extra training nodig hebben in het opstellen van geschikte studieopdrachten (Hook et al., 2003) en het gebruik van beoordelingscriteria (Oakleaf, 2009, p. 981).

3.2.2 Portfolio's

Portfolio's kennen hun oorsprong in het kunstonderwijs en hebben daar de betekenis van een verzameling van gemaakte werkstukken. In het hoger onderwijs bestaat veel aandacht voor het portfolio als instrument waarmee de student

- a) het bereiken van de gewenste eindtermen of competenties aantoont en
- b) zijn / haar ontwikkeling vastlegt en aantoonbaar maakt.

Bij de bespreking van het onderwerp 'studieopdrachten' in paragraaf 3.2.1 was al een aantal malen sprake van een 'portfolio'. In de terminologie van Smith en Tillema (2003, pp. 627-628) is er in dergelijke gevallen sprake van een "training portfolio" dat betrekking heeft op een specifieke cursus of een specifiek onderwijsonderdeel. Behalve eindproducten wordt in het portfolio aanvullend documentatiemateriaal verzameld, zoals beoordelingen door derden en reflecties over het eigen functioneren (Smith & Tillema, 2003, p. 627; Radcliff et al. 2007, pp. 132-133). Een vroeg voorbeeld van zo'n portfolio op het gebied van informatievaardigheden is het werk van Fourie en Van Niekerk (1999 en 2001). Ook Sonley et al. beschrijven het gebruik van een trainingsportfolio (Sonley et al., 2007). In hun geval gaat het om een verslag dat behoort bij een cursus waarin studenten de opdracht krijgen om een geannoteerde bibliografie samen te stellen. Aan de hand van diverse templates wordt de studenten gevraagd om in te gaan op de gekozen zoektermen en evaluaties van gevonden bronnen.

Portfolio's kunnen echter ook worden gebruikt om over een langere periode te worden bijgehouden. Diller en Phelps beschrijven het gebruik van zo'n ontwikkelingsgericht portfolio. Hun voorbeeld van een ePortfolio is een integraal onderdeel van het curriculum waarmee studenten hun voortgang op het gebied van informatievaardigheid zelf bijhouden (Diller & Phelps, 2008, pp. 78-79). Juist in deze opzet lijkt de kracht van een portfolio als 'tool for learning' goed tot zijn recht te komen omdat studenten nog meer dan bij een specifieke opdracht zelf het bewijsmateriaal verzamelen voor het bereiken van de leerdoelen en daar vervolgens op reflecteren.

Dagboeken en interviews

De reflectiecomponent in een portfolio kan ook de vorm hebben van een dagboek of een weblog hoewel daar het gevaar aan kleeft dat de student zich te veel richt op een chronologische beschrijving van het proces ('en toen... en toen...'). Een dagboek of weblog kan echter wel goed gebruikt worden om de meer feitelijke gang van zaken te registreren en de daarin verzamelde gegevens te gebruiken voor het schrijven van het reflectieverslag (Streatfield & Markless, 2008, p. 105).

De verschillende vormen van portfolio assessment worden vaak gecombineerd met een 'criteriumgericht interview' (Dochy & Nickmans, 2005, pp. 81-83) dat dan onder andere bedoeld is om de authenticiteit van het portfolio vast te stellen. In de gevonden literatuur over het beoordelen van informatievaardigheden is daar echter geen voorbeeld van gevonden. Wel schrijven Buchanan, Luck en Jones over het gebruik van een online forum waar de docenten vragen stellen en studenten reflecteren over (bijvoorbeeld) ethische en juridische aspecten van informatiegebruik door met elkaar over zo'n vraag in discussie te gaan (Buchanan et al., 2002, p. 155). De bijdragen aan het discussieforum worden in hun geval uitdrukkelijk beoordeeld en de resultaten daarvan tellen mee in de toekenning van studiepunten.

3.2.3 Directe observatie

Bij portfolio's die worden gebruikt om behalve het studentenproduct ook het onderzoeksproces te beoordelen, is altijd sprake van een indirecte vorm van beoordeling. Het proces zelf (datgene wat de student daadwerkelijk doet) onttrekt zich immers aan de waarneming. Beoordeeld worden slechts de weergave van dat proces door de student en zijn of haar reflectie daarop. Kathleen Dunn (2002) beschrijft een methode om het gevolgde onderzoeksproces wel rechtstreeks te observeren door middel van "screen captures" (Dunn, 2002, pp. 31-32). Met behulp van de screen capture-software werd zowel vastgelegd wat de studenten zagen op het scherm als al hun activiteiten aan de computer zoals de woorden die werden ingetypt, de muisbewegingen, de sites die werden bezocht enzovoorts. Voor een beperkte groep van de studenten werd deze methode gecombineerd met een etnografische onderzoeksmethode waarbij een van de onderzoekers een student observeerde en vragen stelde over gemaakte keuzes gedurende de studietaak. Aan de studenten was vooraf wel gezegd dat antwoorden niet 'goed' of 'fout' konden zijn en de observanten waren geïnstrueerd om de studenten niet te assisteren of te beïnvloeden tijdens het werk aan hun studietaak (Dunn, 2002, p. 32). Het doel van het door Dunn beschreven onderzoek was dan ook niet de beoordeling van individuele studenten maar het verkrijgen van inzicht in het feitelijke zoekgedrag van studenten (Dunn, 2002, p. 31 en p. 34).

Een ander voorbeeld van directe observatie met behulp van screen capture-software (in dit geval in combinatie met een 'hardop denken'-protocol) wordt beschreven door Novotny en Cahoy (2006). Ook dit assessment had echter niet als doel om studenten zelf te beoordelen. In het beschreven onderzoek dat gebaseerd was op gegevens van een zeer kleine groepen studenten, wordt nagegaan wat het effect is van een gerichte instructie op de vaardigheid van de betreffende studenten met het gebruik van de online bibliotheekcatalogus. Dat laatste is bovendien maar een zeer klein facet van het complexe construct informatievaardigheid zodat deze studie in het kader van het onderhavige onderzoek niet relevant te noemen is.

Een derde voorbeeld van directe observatie waarbij gebruik werd gemaakt van schermopnamen is dat van Lazonder (2005). Hij liet studenten zoekacties uitvoeren en legde onder andere vast wat de kwaliteit was van hun regulatieactiviteiten gedurende het zoekproces (Lazonder, 2005, p. 469). Ook in dit geval werden de uitkomsten echter niet gebruikt voor het beoordelen van de taakuitvoering of voor het geven van feedback. De zoekacties werden uitgevoerd in een experimentele onderzoekssituatie waarbij het doel was om inzicht te krijgen in de vraag of samenwerking bij het zoeken op het web leidt tot betere zoekresultaten dan individueel zoeken.

Emmett en Emde beschrijven een beoordelingspraktijk waarbij studenten rechtstreeks werden geobserveerd tijdens het uitvoeren van een informatietaak (een 'proeve van bekwaamheid'). Ze maakten daarvoor gebruik van "information-need scenarios" (Emmett & Ende, 2007, Appendix 1). Tekenend is echter dat deze vorm van assessment in hun praktijk na twee jaar werd vervangen door een gestandaardiseerde test omdat de observatiemethode uiteindelijk toch als te tijdrovend werd ervaren. Ook Hignite et al. (2009) maakten in hun "Information and Communications Technology (ICT) exam" gebruik van 'taken' om de cognitieve en kritische vaardigheden van studenten te meten. In hun geval is echter niet duidelijk hoe de taken er precies uit zagen.

3.2.4 Scoringsrubrieken

Voor 'performance assessment' van informatievaardigheden - zij het door middel van studieopdrachten, zij het met behulp van portfolio's of een combinatie van die twee - is een checklist of een 'scoringsrubriek' voor het vaststellen van de beoordeling een onmisbaar hulpmiddel (Fourie & Van Niekerk, 1999, p. 342; Gratch-Lindauer, 2003, p. 31; Knight, 2006, p. 45 ff.; Oakleaf, 2008, p. 244 ff.). Jonsson en Svingby definiëren zo'n scoringsrubriek als "a scoring tool for qualitative rating of authentic or complex student work" (2007, p. 131). Het bevat zowel de beoordelingscriteria voor de belangrijkste dimensies van de uit te voeren taak als de normen (de "standards of attainment") die worden gehanteerd bij het gebruik van die criteria. Volgens Knight wordt in het hoger onderwijs steeds vaker onderkend dat scoringsrubrieken een geschikt hulpmiddel zijn om de uitvoering van authentieke beroepstaken op een betrouwbare manier te beoordelen (Knight, 2006, p. 45). In Nederland worden ze vaak toegepast voor het beoordelen van communicatieve vaardigheden bij schrijfopdrachten en mondelinge presentaties (persoonlijke communicatie Ron Oostdam, Universiteit van Amsterdam, SCO-Kohnstamm Instituut). Ter illustratie geeft figuur 3.5 een fragment van een scoringsrubriek die door Oakleaf werd ontwikkeld voor het beoordelen van een opdracht aan studenten om een website te evalueren.

Evaluation Criteria	Beginning	Developing	Exemplary
Articulates Criteria	0 - Student does not address authority issues. 	1 - Student addresses authority issues, but does not use criteria terminology. 	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship. 
Cites Indicators of Criteria	0 - Student does not address authority indicators. 	1 - Student refers vaguely or broadly to authority indicators, but does not cite specific indicators. 	2 - Student cites specific authority indicators such as: domain, server/publisher/host, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials. 
Links	0 - Student does not	1 - Student refers vaguely or	2 - Student cites specific

Figuur 3.5

Fragment uit de scoringsrubriek van Oakleaf (2009).

De studieopdracht had betrekking op de evaluatie van een website (Oakleaf, 2009, p. 974).

In de onderwijswetenschappen worden scoringsrubrieken beschouwd als een van de “standard oriented” beoordelingsinstrumenten die vooral hun nut bewijzen bij het beoordelen van ‘algemene vaardigheden’ (‘general skills’). In het kader van toetsen en beoordelen onderscheiden Elshout-Mohr, Oostdam en Overmaat dergelijke algemene vaardigheden van de ‘discrete kennis’ die bij een vakgebied behoort en competenties die bij één specifiek beroep horen (2002, p. 374). Voor de algemene vaardigheden die niet bij een specifiek beroep horen geldt volgens hen dat ze vooral een rol spelen bij ‘open studietaken’ die studenten de ruimte bieden om ze, binnen zekere grenzen, op hun eigen manier aan te pakken (Elshout-Mohr et al., 2002, pp. 374-375; Jonsson & Svingby, 2007, p. 131). Zoals ook in hoofdstuk 2 al werd betoogd is Information Problem Solving een voorbeeld van zo’n taak met een open karakter. Studenten kiezen immers vaak een eigen onderwerp voor hun onderzoek of dienen op zijn minst hun eigen focus te formuleren op het vraagstuk dat door hun docent is voorgelegd.

Ook scoringsrubrieken zijn er natuurlijk in verschillende soorten. Mertler (2001) onderscheidt holistische van analytische scoringsrubrieken. Een holistische scoringsrubriek wordt gebruikt om een algemene beoordeling uit te spreken over het studieproduct of het proces, zonder dat de verschillende dimensies of ‘traits’ afzonderlijk beoordeeld worden. Dit in tegenstelling tot analytische ‘rubrics’ waarbij “the teacher scores separate individual parts of the product or performance first, then sums the individual scores to obtain a total score” (Mertler, 2001). Het gebruik van een holistische rubriek is voor beoordelaars uiteraard minder tijdrovend maar analytische scoringsrubrieken hebben voor studenten het voordeel dat ze veel uitgebreidere feedback verschaffen. In praktijk komt het overigens ook voor dat van een analytische scoringsrubriek één of meer specifieke dimensies of ‘traits’ worden gebruikt bij een bepaalde opdracht (Nitko & Brookhart 2007, p. 269). Het gebruik van de verschillende dimensies kan dan gedurende het curriculum worden opgebouwd zodat er sprake is van een begeleide ontwikkeling van de complexe beroepstaak.

In de literatuur op het gebied van information literacy worden de volgende voordelen van het gebruik van scoringsrubrieken bij beoordelingen genoemd:

- De subjectiviteit bij het beoordelen van studieopdrachten wordt verminderd doordat er duidelijke omschrijvingen bestaan van de verschillende niveaus waarop deze uitgewerkt kunnen zijn (Oakleaf, 2008, pp. 245-246).
- De scoringsrubriek kan worden gebruikt om uitgebreid feedback te geven (Oakleaf, 2008, p. 245).
- Tijdens het ontwikkelen van een scoringsrubriek worden de docenten en / of beoordelaars gestimuleerd om een gemeenschappelijk beeld te vormen van de inhoud van het construct Informatievaardigheid c.q. Information Problem Solving (Knight, 2006, p. 52; Oakleaf, 2008, p. 246). Scharf et al. (2007, p. 469) wijzen erop dat de betrouwbaarheid van een scoringsrubriek als beoordelingsinstrument juist ook tot stand kan komen als resultaat van het overleg dat de beoordelaars met elkaar gehad hebben tijdens het ontwikkelproces. Dit geldt uiteraard vooral als er geen gebruik wordt gemaakt van een standaard model maar de betreffende rubriek door het team zelf wordt geconstrueerd.
- Als de scoringsrubriek vooraf wordt verspreid en met de studenten wordt besproken, helpt dat hen om te begrijpen wat er in het kader van de studieopdracht van hen wordt verwacht (Scharf et al., 2007, p. 471; Oakleaf, 2008, p. 245).

- Studenten kunnen de scoringsrubriek gebruiken om zichzelf of elkaar tussentijds te beoordelen (Oakleaf, 2008, p. 246).
- Net als standaard tests kunnen gestandaardiseerde scoringsrubrieken worden gebruikt om verschillende cohorten studenten (uit verschillende jaren of van verschillende studieprogramma's) met elkaar te vergelijken (Oakleaf, 2008, pp. 245-246).

Uiteraard zijn er ook nadelen verbonden aan het gebruik van scoringsrubrieken. Het in de literatuur meest genoemde probleem is de tijd die is gemoeid met de constructie ervan of, in het geval dat een bestaande scoringsrubriek wordt gebruikt, de moeite die het kost om de beoordelaars er mee te leren werken (Knight, 2006, p. 52; Diller & Phelps, 2008, p. 82). Dat laatste dient niet te worden onderschat: Oakleaf beschrijft het belang van uitgebreide trainingssessies met beoordelaars die nodig zijn om tot een aanvaardbare mate van interbeoordelaarsbetrouwbaarheid te komen (2009, pp. 975-976). Aan de andere kant dragen dergelijke sessies natuurlijk wel bij aan de vorming van een gemeenschappelijk beeld van het ingewikkelde construct Informatievaardigheid.

3.3 Conclusies en discussie

In dit hoofdstuk zijn verschillende methodes en instrumenten besproken die gebruikt kunnen worden om informatievaardigheden van individuele studenten te beoordelen. Uitgangspunt was dat een geschikt instrument aan de volgende eisen dient te voldoen:

- het levert betrouwbare en valide beoordelingen op
- het beperkt zich niet tot het beoordelen van informatievaardigheden maar kan ook worden gebruikt om deze te bevorderen
- het richt zich op het volledige construct informatievaardigheid en
- het is generiek toepasbaar.

De gevonden methodes voor het beoordelen van informatievaardigheid vallen in eerste instantie uiteen in vragenlijsten en studieopdrachten⁹. Van de vragenlijsten zijn enquêtes die naar opinies of gedrag vragen, niet geschikt voor het beoordelen van individuele studenten. Schriftelijke tests kunnen daarvoor wel worden gebruikt maar hebben vrijwel altijd slechts betrekking op het cognitieve domein. Behalve reproductievragen kunnen ook begrips- en toepassingsvragen worden gesteld maar er zijn geen voorbeelden van tests gevonden die betrekking hebben op de uitvoering van de volledige taak Information Problem Solving. In praktijk hebben ze vrijwel altijd betrekking op bronnenkennis, zoektechnieken en bronbeoordelingen (zie ook Stoffberg & Blignaut, 2008, p. 7) en juist niet op ACRL standaard 4 die betrekking heeft op het *gebruik* van informatie (Cameron et al., 2007, p. 231; Burkhardt et al., 2008, p. 383). Voor dat laatste zijn 'performance assessments' in de vorm van studieopdrachten beter geschikt (Cameron et al., 2007, p. 231).

⁹ Het in paragraaf 3.2.2 besproken beoordelingsinstrument 'Portfolio' en de 'Scoringsrubrieken' uit paragraaf 3.2.4 worden niet als een zelfstandig te onderscheiden methode gezien omdat het in beide gevallen gaat om een hulpmiddel dat wordt ingezet bij de beoordeling van studieopdrachten.

Voor de beoordeling van het Information Problem Solving-proces tijdens de uitvoering van dergelijke studieopdrachten zijn echter wel extra beoordelingsinstrumenten nodig. In de gevonden literatuur kwam naar voren dat een analytische scoringsrubriek een geschikt instrument is. Mits goed geconstrueerd, komt zo'n 'rubric' tegemoet aan verschillende eisen die aan het beoordelingsinstrument zijn gesteld.

- Door een goede operationalisering van de verwachte leeruitkomsten en beschrijvingen van gedrag op verschillende niveaus, bevordert ze de betrouwbaarheid van beoordelingen.
- Ze kan zowel worden ingezet voor summatieve beoordelingen als gedurende het onderwijsproces als een 'tool for learning', bijvoorbeeld als formatieve toets. Een scoringsrubriek sluit dus goed aan bij de procesgerichte constructivistische benadering van informatievaardigheden waarvan in hoofdstuk 2 is opgemerkt dat die de context vormt waarin het beoordelingsinstrument ook toegepast dient te kunnen worden.
- Omdat een scoringsrubriek kan worden opgebouwd uit verschillende dimensies kan het instrument als geheel betrekking hebben op een complete samengestelde taak. Toegepast op de vraag die in dit onderzoek aan de orde is, kan een scoringsrubriek voor informatievaardigheden dus gebruikt worden voor een beoordeling van het volledige Information Problem Solving-proces.

Over de andere eisen die aan het te ontwikkelen instrument worden gesteld - validiteit en generieke toepasbaarheid - valt in zijn algemeenheid niets te zeggen. Die kenmerken worden bepaald door de eigenschappen van het instrument dat daadwerkelijk wordt ontwikkeld.

Een alternatieve methode voor het beoordelen van het Information Problem Solving-proces van studenten, de zogenaamde directe observatie, bleek vooral toegepast te worden in het kader van *onderzoek* naar informatievaardigheden. De methodes die daarbij gebruikt worden zijn te bewerkelijk om in de reguliere onderwijspraktijk te worden ingezet.

In de literatuur zijn een paar voorbeelden van scoringsrubrieken voor informatievaardigheden gevonden maar deze hadden alle betrekking op een specifieke cursus. In het onderhavige onderzoek wordt er de voorkeur aan gegeven om een generiek toepasbaar instrument te ontwikkelen zodat het ook een rol kan spelen bij de implementatie van onderwijs in informatievaardigheden in opleidingen of instituten waar dit tot nu toe nog niet van de grond is gekomen. In dit proefschrift wordt er daarom voor gekozen om een scoringsrubriek voor informatievaardigheden te ontwikkelen die - met eventueel kleine aanpassingen - gebruikt kan worden in verschillende opleidingen in het hoger onderwijs. De constructie van zo'n algemeen toepasbare scoringsrubriek voor informatievaardigheden is onderwerp van het volgende hoofdstuk. In deel II van dit proefschrift zal moeten blijken of met zo'n generiek beoordelingsinstrument betrouwbare en valide beoordelingen kunnen worden opge maakt.

Constructie van de scoringsrubriek¹⁰

Abstract: Beschreven wordt hoe een analytische scoringsrubriek is ontworpen waarmee het daadwerkelijke informatiegebruik van studenten tijdens het oplossen van informatieproblemen kan worden beoordeeld en die tevens kan functioneren als een feedback-instrument. Tijdens het ontwerp (waarbij gebruik werd gemaakt van een variant van de 7 stappen-procedure van Mertler, 2001) is een iteratief proces doorlopen waarbij gebruik werd gemaakt van vakliteratuur op het gebied van informatievaardigheden, internationale standaarden op dat gebied, observatie van studenten tijdens practica en collegiale toetsing van concept versies door docenten die werkzaam zijn in het hbo. De eerste concept versie van de scoringsrubriek is getest aan de hand van een studieopdracht tot het schrijven van een landenprofiel en is op basis van de daarbij opgedane ervaringen gereviseerd. Het uiteindelijke prototype van de scoringsrubriek bestaat uit zeven afzonderlijke beoordelingscriteria met omschrijvingen van professioneel respectievelijk inadequaar gedrag, aangevuld met een 6 punts Likertschaal.

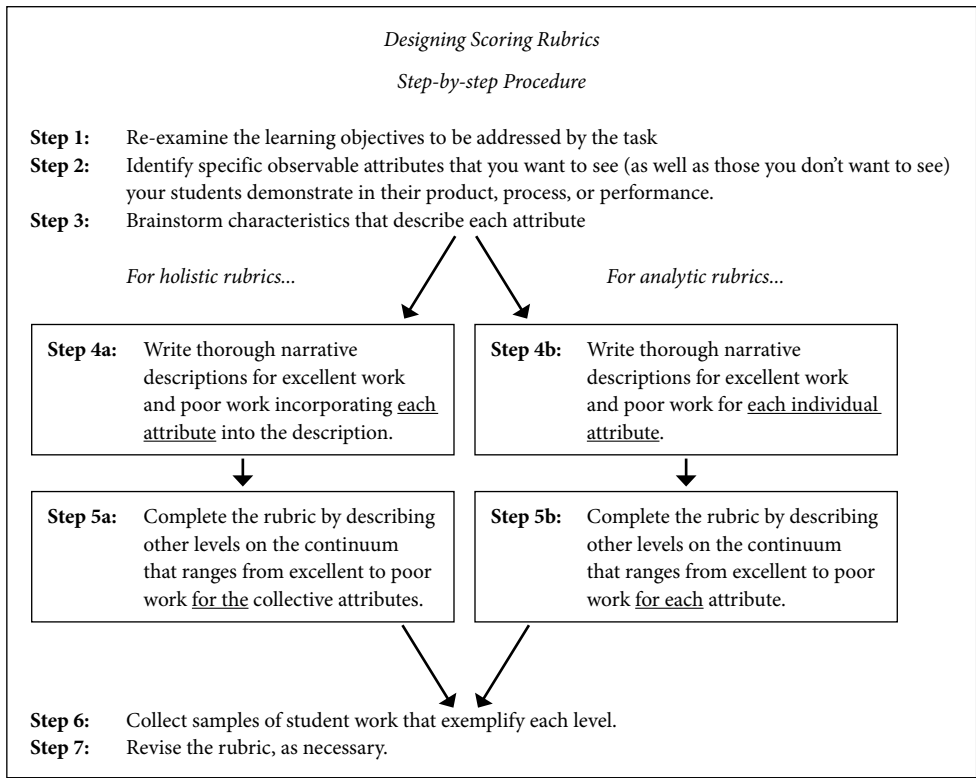
¹⁰ De inhoud van dit hoofdstuk is eerder gepubliceerd in: A scoring rubric for performance assessment of information literacy in Dutch Higher Education. *Journal of Information Literacy*, 4 (1), 22-39 (Van Helvoort, 2010).

Inleiding

In dit hoofdstuk wordt beschreven hoe een beoordelingsinstrument is geconstrueerd waarmee het daadwerkelijke informatiegebruik van studenten in het hoger onderwijs kan worden beoordeeld en dat tevens gebruikt kan worden als feedback-instrument om het leren van studenten op dat gebied te stimuleren. In hoofdstuk 3 is aangegeven waarom uiteindelijk is gekozen voor de beoordeling van informatievaardigheden aan de hand van studieopdrachten en het gebruik daarbij van een *scoringsrubriek*. Omdat naast de beoordelingsfunctie ook veel waarde wordt gehecht aan de feedbackfunctie van het instrument is gekozen voor een *analytische* scoringsrubriek. In paragraaf 3.2.4 is uiteen gezet dat juist die vorm van een scoringsrubriek de voorkeur verdient als het beoordelingsinstrument ook wordt gebruikt als een ‘tool for learning’.

4.1 Ontwerpprocedure

Voor het ontwerpen van een scoringsrubriek heeft Mertler (2001) op basis van verschillende informatiebronnen een 7 stappen-procedure voorgesteld (figuur 4.1).



Figuur 4.1

"Step-by-Step Procedure" voor het ontwerpen van scoringsrubrieken (Mertler, 2001, Figure 2)

Bij het ontwerp van de Scoringsrubriek voor Informatievaardigheden is een aangepaste versie van Mertlers procedure voor het ontwikkelen van een analytische scoringsrubriek toegepast. Zijn procedure kon niet stap voor stap worden gevolgd omdat de scoringsrubriek die in het eigen onderzoek werd ontwikkeld, niet bedoeld is voor één concrete studie-opdracht in het kader van een specifieke cursus of les. De scoringsrubriek voor informatievaardigheden zou breder toepasbaar moeten zijn, namelijk voor de beoordeling van verschillende studieopdrachten die studenten in het kader van verschillende cursussen en onderwijsseenheden uitvoeren. Om die reden kon stap 1 uit Mertlers procedure (“Re-examine the learning objectives to be addressed by the task”) niet worden uitgevoerd. In plaats daarvan startte het eigen onderzoek met het vaststellen en de nadere beschrijving van de te gebruiken beoordelingscriteria (Mertlers stappen 2 en 3). Vervolgens werd voor ieder beoordelingscriterium zowel het ‘professionele’ als het ‘inadequate’ gedrag zo concreet mogelijk uitgeschreven (Mertlers stap 4b). Met het ‘inadequate gedrag’ werd in dit geval bedoeld op fouten en missers die vaak in het werk van studenten worden aangetroffen. Het verzamelen van ideeën over de te gebruiken beoordelingscriteria en de voorbeelden van professioneel en inadequaar gedrag was in deze eerste fases van het ontwerpproces niet aan strenge regels gebonden. Input van verschillende kanten was welkom, met name omdat deze fase van het ontwerpproces nog kon worden beschouwd als een verkenningsfase.

In de zeven-stappen-procedure van Mertler volgt na de formulering van het professionele en het inadequate gedrag een nadere beschrijving van het gedrag op de verschillende tussenliggende niveaus (Mertlers stap 5b). Voor de eigen scoringsrubriek diende daaraan voorafgaand nog wel het aantal gewenste niveaus te worden vastgesteld. Nadat de eerste versie van de scoringsrubriek op alle niveaus is uitgeschreven volgt in Mertlers procedure een test aan de hand van bestaande uitwerkingen van de opdracht, bij voorkeur op zoveel mogelijk verschillende niveaus. Omdat de scoringsrubriek voor informatievaardigheden geen betrekking heeft op een bestaande studieopdracht gold opnieuw dat hier enigszins afgeweken moest worden van de zeven-stappen-procedure. In het eigen onderzoek werd voor het testen van de eerste versie gebruik gemaakt van een opdracht bij een recent ontwikkelde cursus. Daarvoor werd een bijdrage gevraagd van twee collega-docenten die de uitwerkingen van een opdracht voor een essay beoordeelden met behulp van de concept versie. Met hun feedback kon de scoringsrubriek worden aangepast. Figuur 4.2 geeft een samenvatting van het ontwerpproces. In de paragrafen 4.2 tot en met 4.4 worden de afzonderlijke stappen zoals ze daadwerkelijk zijn uitgevoerd nader beschreven. Het hoofdstuk besluit met een beschrijving van het prototype dat het resultaat is van het ontwerpproces (paragraaf 4.5) en een paragraaf met conclusies en discussie (4.6).

a. Vaststellen van de gewenste criteria b. Beschrijving van het ‘professionele gedrag’ voor ieder criterium c. Beschrijving van het ‘inadequaar gedrag’ voor ieder criterium	Verkennen
d. Keuze voor het aantal niveaus dat gebruikt gaat worden in de scoringsrubriek e. Beschrijving van het gedrag op de tussenliggende niveaus voor ieder criterium	Uitwerken
f. Testen van de eerste concept versie met uitwerkingen van studenten op verschillende niveaus g. Aanpassing van de scoringsrubriek als dat gewenst is	Testen

Figuur 4.2

Ontwerpproces van de Scoringsrubriek voor Informatievaardigheden

4.2 Verkenning

Om de juiste beoordelingscriteria op het spoor te komen en zowel het professionele gedrag als het inadequate gedrag te operationaliseren, zijn drie onderzoeksmethodes gebruikt:

- Om te beginnen zijn de ACRL-standaarden (2000) en de daar geformuleerde gedragsindicatoren bestudeerd.
- Gedurende twee periodes van 10 weken waarin de onderzoeker ook een training Deskresearch verzorgde voor een groep studenten van verschillende opleidingen van De Haagse Hogeschool, is het zoekgedrag van de studenten gemonitord en zijn er aantekeningen gemaakt van veel voorkomende fouten.
- Concept versies van de criteria en de gedragsbeschrijvingen zijn in drie afzonderlijke review bijeenkomsten besproken met collega's van de opleiding informatiedienstverlening en -management én met collega's van andere opleidingen (hoger europees beroeps onderwijs, bedrijfseconomie en hbo verpleegkunde).

Bij de bestudering van de ACRL-standaarden en de bijbehorende gedragsindicatoren is gebruik gemaakt van Neely (2006). Gebruik van dat boek had voordelen boven het gebruik van de tekst van de standaarden zelf omdat Neely behalve de letterlijke tekst van de standaarden, de gedragsindicatoren en de leeruitkomsten tevens verschillende voorbeelden geeft van 'good practices' uit het onderwijs zelf. Haar boek was dan ook een goede bron bij het formuleren van de verschillende criteria en het 'professionele' gedrag daarbij. Een tweede bron die in deze fase van het onderzoek veel ideeën heeft opgeleverd is die van Scharf, Elliot, Huey, Briller en Joshi (2007). Van het ACRL Framework for Information Literacy for Higher Education waarvan de eerste concept versies in 2014 verschenen, is ook in een later stadium geen gebruik gemaakt omdat de daar beschreven 'frames', anders dan de standaarden uit 2000, onvoldoende houvast boden in de vorm van duidelijk omschreven leeruitkomsten (zie ook paragraaf 2.1.3).

De eerste fase leverde de volgende beoordelingscriteria op:

1. Oriëntatie op het onderwerp
2. Gebruik van deskresearch als onderzoeksmethode
3. Aanwezigheid van een literatuurlijst en de kwaliteit van de beschrijvingen
4. Kwaliteit van de gebruikte primaire informatiebronnen (boeken, artikelen, websites)
5. Bronverwijzingen in de tekst
6. Verwerking van relevante informatie tot nieuwe kennis
7. Gebruik van zoektermen
8. Gebruik van diverse secundaire informatiebronnen.

De eerste zes beoordelingscriteria kunnen worden toegepast op het kennisproduct dat de student oplevert (het essay, verslag, de presentatie), de laatste twee kunnen alleen worden gebruikt als van de student ook wordt gevraagd dat deze een zoekstrategie inlevert. Bij de keuze voor de te gebruiken beoordelingscriteria lag de nadruk dus op eigenschappen van het eindproduct. De motivatie daarvoor is dat Information Problem Solving kan worden beschouwd als een betrekkelijk generieke vaardigheid waarbij het redelijk is om studenten de gelegenheid te bieden om, binnen zekere

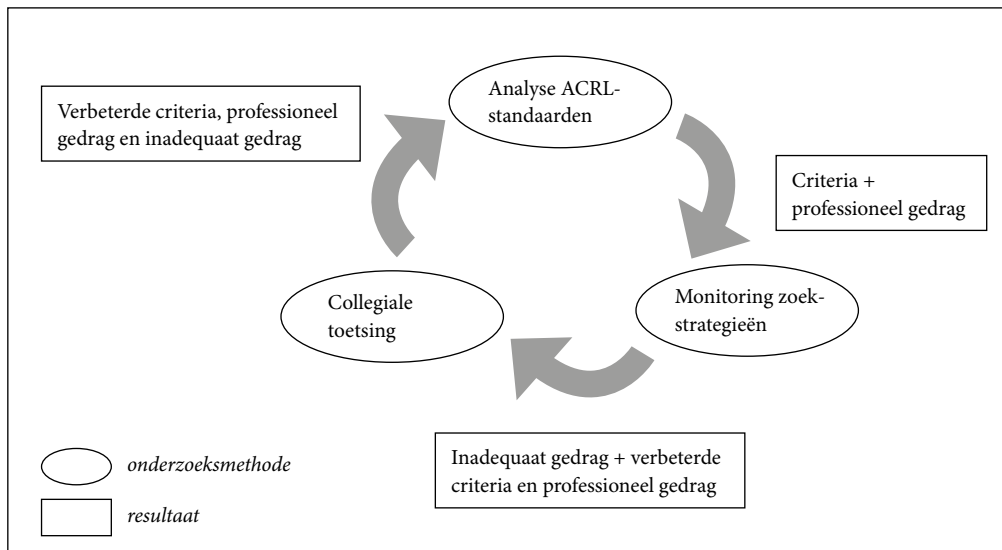
grenzen, zelf te kiezen hoe ze de opdracht aanpakken. Een voorbeeld is het opsporen van informatie door het ‘volgen’ van hyperlinks naast het gebruik van keyword searches. Sommige studenten zullen zo’n methode van meanderen ervaren als een efficiënte manier om snel relevante informatie te vinden, anderen raken daarbij al snel het spoor bijster. Het lijkt weinig zin te hebben om studenten te verplichten om (ook) op die manier te zoeken of hen dat juist te verbieden. Aan de hand van de kwaliteiten van het eindproduct kan echter wel worden vastgesteld hoe goed de informatie was die de gevolgde methode heeft opgeleverd.

Nadat de beoordelingscriteria waren vastgesteld, werden de beschrijvingen van professioneel en inadequaat gedrag opgesteld. Beschrijvingen van inadequaat gedrag werden aanvankelijk vooral ontleend aan voorbeelden die de onderzoeker zelf tegenkwam tijdens de begeleiding van practica ‘retrieval’ in het kader van het 10 weken durend onderwijsblok Deskresearch dat door de opleiding idm van De Haagse Hogeschool wordt verzorgd. Tijdens de cursus leggen de studenten een zoekstrategie vast die ze iedere week tijdens het practicum met de begeleidend docent bespreken. Vooraf krijgen de studenten toegelicht dat het bij het onderdeel ‘zoekstrategie’ niet zozeer gaat om een beschrijving van alle stappen die zijn gezet maar wel om de stappen waarvan ze zelf ervaren hebben dat ze het meest hebben opgeleverd. De zoekstrategie kan dus wekelijks worden aangepast en de uiteindelijke zoekstrategie is als het goed is een beschrijving van de bronnen, zoektermen en zoekmethodes waarmee het onderzoek herhaald kan worden.

Het monitoren van deze zoekstrategieën gedurende twee cursussen in het schooljaar 2007-2008 leverde tal van voorbeelden op van veel voorkomende fouten en misverstanden. Deze werden in de scoringsrubriek verwerkt in de beschrijvingen van het inadequate gedrag.

Het overzicht van beoordelingscriteria met zowel beschrijvingen op professioneel niveau als die op inadequaat niveau is besproken in drie review sessies, één bijeenkomst met collega’s van de opleiding idm en twee bijeenkomsten met collega’s van andere opleidingen. De feedback op het concept van de scoringsrubriek tijdens deze ‘collegiale toetsing’ heeft een belangrijke rol gespeeld bij de uiteindelijke formulering van de criteria. De organisatie van deze sessies met collega’s had vooral als doel om te komen tot omschrijvingen van beoordelingscriteria die worden herkend in de onderwijspraktijk van ervaren docenten en beoordelaars, de ‘inhoudsvaliditeit’ van de te gebruiken beoordelingscriteria (Cronbach & Meehl, 1955; Tigelaar, Dolmans, Wolfhagen & Van der Vleuten, 2005, p. 597).

De gebruikte onderzoeksmethodes (analyse van de literatuur / ACRL standaarden, het monitoren van de zoekstrategieën en de collegiale toetsing) zijn hierboven weliswaar in een lineaire volgorde gepresenteerd maar zijn feitelijk niet in die strikte volgorde uitgevoerd. De ontwerpactiviteiten startten met de analyse van de ACRL-standaarden maar het volledige proces is drie maal volledig doorlopen. Deze iteratie van het ontwerpproces heeft er voor gezorgd dat het concept van de scoringsrubriek steeds kon worden verbeterd. Zo leidden de reviews met de collega’s niet alleen tot scherpere formuleringen van de beoordelingscriteria en het ‘professionele’ gedrag maar ook tot duidelijker voorbeelden van zwakke zoekstrategieën en betere beschrijvingen van ‘inadequaat’ gedrag. Het iteratieve karakter van het ontwerpproces in deze verkennende fase wordt uitgedrukt in figuur 4.3.

**Figuur 4.3**

Vaststellen van de beoordelingscriteria en de beschrijving van professioneel en inadequaat gedrag tijdens de verkennende fase van het ontwerp van de scoringsrubriek

4.3 Uitwerking

In deze fase van het ontwerpproces wordt vastgesteld hoeveel tussenliggende niveaus in de scoringsrubriek gebruikt gaan worden en wordt voor ieder van die niveaus het kenmerkende gedrag uitgeschreven. In het geval van de scoringsrubriek voor informatievaardigheden was het de bedoeling dat deze gebruikt zou gaan worden door docenten en studenten uit verschillende vakgebieden in het hoger onderwijs. Om de scoringsrubriek niet te ingewikkeld te maken voor deze docenten en studenten die zelf geen informatieprofessional of -specialist zijn, werd er voor gekozen om de rubriek te beperken tot drie niveaus en dus slechts één tussenniveau te formuleren. Uitgaande van het beschreven 'professionele' en 'inadequate' gedrag werd het gedrag omschreven dat 'voor verbetering vatbaar' was. Deze versie van de scoringsrubriek met gedragsbeschrijvingen op drie niveaus is opgenomen als tabel 4.1.

Tabel 4.1:*Eerste concept versie van de Scoringsrubriek voor Informatievaardigheden***Onderdeel: studentenproduct**

<i> criterium</i>	<i> Professioneel</i>	<i> Voor verbetering vatbaar</i>	<i> Inadequaat</i>
1 Oriëntatie op het onderwerp	☐ Uit het product blijkt dat de student zich goed op de onderzoeksvraag of het onderwerp heeft georiënteerd en op basis daarvan een eigen invalshoek (focus) heeft geformuleerd.	☐ Uit het product blijkt dat de student externe bronnen heeft gebruikt om zich te oriënteren op de onderzoeksvraag of het onderwerp maar hij/zij heeft geen eigen invalshoek (focus) geformuleerd.	☐ Uit het product blijkt dat de student is uitgegaan van de oorspronkelijke vraag of de oorspronkelijke formulering van het onderwerp zonder dat hij/zij die verder heeft uitgediept.
2 Gebruik van deskresearch als onderzoeksmethode	☐ Uit het product blijkt dat de student deskresearch heeft toegepast als onderzoeksmethode voor de vraagstukken die daarvoor in aanmerking komen. Voorbeeld: de student heeft zowel • kernbegrippen gedefinieerd op basis van relevante literatuur als • voor een bedrijfsanalyse gebruik gemaakt van externe bronnen.	☐ Uit het product blijkt dat de student gebruik heeft gemaakt van deskresearch maar niet voor alle vragen die daarvoor in aanmerking komen. Voorbeeld: de student • heeft kernbegrippen gedefinieerd op basis van relevante literatuur maar • heeft voor een bedrijfsanalyse slechts gebruik gemaakt van informatie uit de betreffende organisatie zelf.	☐ Uit het product blijkt dat de student veel te weinig gebruik heeft gemaakt van deskresearch terwijl er wel vragen zijn die daarvoor in aanmerking komen. Voorbeeld: de student • heeft geprobeerd om kernbegrippen zelf te definiëren en heeft • bij een bedrijfsanalyse bovendien slechts gebruik gemaakt van informatie uit de betreffende organisatie zelf.
3 Bronnen-overzicht / literatuurlijst	☐ Het product bevat een bronnenoverzicht dat volledig is en waarbij een standaard citation style correct is toegepast. Aan de hand van het bronnenoverzicht zijn de gebruikte bronnen eenduidig te achterhalen.	Het product bevat een bronnenoverzicht met de belangrijkste gegevens om de genoemde bronnen te achterhalen maar ☐ het format voor de titelbeschrijvingen is niet juist of niet consequent toegepast Voorbeeld: • bij een monografie worden auteur, titel en jaar van uitgave vermeld maar niet de naam van de uitgever en / óf ☐ sommige bronnen die in de tekst worden genoemd (hooguit één of twee) ontbreken in het bronnenoverzicht.	☐ In het product ontbreekt een bronnenoverzicht en / óf ☐ het bronnenoverzicht is onvolledig (van verschillende documenten waar in de tekst naar wordt verwezen ontbreken de gegevens) en / óf ☐ cruciale bibliografische gegevens (titel, auteur, publicatiejaar) ontbreken. Voorbeeld: • alleen de url is vastgelegd.
4 Kwaliteit van de gebruikte externe informatie-bronnen	☐ Uit het bronnenoverzicht dat bij het product behoort blijkt dat de student gebruik heeft gemaakt van relevante, betrouwbare (bij voorkeur zo authentiek mogelijke) en actuele informatiebronnen die het onderwerp bovendien vanuit verschillende gezichtspunten behandelen.	☐ Sommige van de gebruikte bronnen zijn niet echt relevant, te licht of verouderd en / óf ☐ de gebruikte literatuur is eenzijdig.	☐ Het merendeel van de gebruikte externe bronnen is te licht, verouderd of onvoldoende relevant.

5	Bronverwijzingen in de tekst	☐ In de tekst van het product zelf heeft de student eenduidig verantwoord welke informatiebronnen hij of zij heeft gebruikt.	☐ In de tekst van het product zelf heeft de student verantwoord welke informatiebronnen hij/zij heeft gebruikt maar de manier van verwijzen is niet geheel correct waardoor onduidelijkheid ontstaat over de vraag wat de eigen bijdrage is en wat is ontleend.	☐ De student heeft werk van anderen (stukken tekst, afbeeldingen, geluidsfragmenten) verwerkt in zijn/haar eigen product zonder dat hij/zij de gebruikte bron heeft vermeld. Hoewel vaak niet opzettelijk, is er strikt gesproken sprake van 'plagiaat'.
---	------------------------------	---	--	--

Criterium		Professioneel	Voor verbetering vatbaar	Inadequaar
6	Verwerking van relevante informatie tot nieuwe kennis en/of inzichten	☐ Uit het product blijkt dat de student informatie uit verschillende bronnen heeft geanalyseerd en op basis daarvan nieuwe toepassingen, inzichten of hypothesen heeft geformuleerd.	☐ Uit het product blijkt dat de student informatie uit verschillende bronnen heeft geanalyseerd maar hij/zij heeft zelf geen nieuwe toepassingen, inzichten of hypothesen geformuleerd.	In het studentenproduct heeft de student ☐ de inhoud van de gebruikte informatiebronnen niet correct of niet helder weergegeven, en / óf ☐ in het geheel geen aandacht besteed aan analyse van de opgespoorde informatiebronnen, en / óf ☐ zich gebaseerd op slechts één informatiebron en hij/zij heeft de daar gevonden informatie niet ter discussie gesteld terwijl daar wel aanleiding toe is.

Scoringsrubriek voor Informatievaardigheden

Onderdeel: procesverslag / zoekstrategie

	<i>Criterium</i>	<i>Professioneel</i>	<i>Voor verbetering vatbaar</i>	<i>Inadequaat</i>
7	Zoektermen	☐ De student heeft de zoektermen gebruikt die relevant zijn voor het te onderzoeken onderwerp / vraagstuk. Hij/zij heeft daarbij rekening gehouden met het vakjargon, de synoniemen en (zoek)termen in een andere taal (meestal Engels).	☐ De student heeft (zoek)termen gebruikt die relevant zijn voor het te onderzoeken onderwerp / vraagstuk maar het zijn er te weinig of essentiële ontbreken.	☐ De zoektermen die de student heeft gebruikt zijn te algemeen of komen uit het algemeen spraakgebruik.. en / óf ☐ de student heeft geen rekening gehouden met synoniemen, verwante begrippen en (zoek) termen in een andere taal (meestal Engels).
8	Gebruik van diverse secundaire informatiebronnen	☐ De student heeft gebruik gemaakt van diverse typen secundaire bronnen (search engines, boeken, wetenschappelijke tijdschriften, databases, sociale netwerken). Indien van toepassing heeft hij/zij bovendien gebruik gemaakt van mogelijkheden buiten de eigen lokatie (IBL, documentleverantie).	☐ De student heeft gebruik gemaakt van professionele secundaire bronnen (naast search engines ook bibliografische databases en wetenschappelijke tijdschriften) maar hij/zij heeft een of meer relevante bronnen over het hoofd gezien (bijvoorbeeld een belangrijk vaktijdschrift).	☐ De student heeft alleen informatiebronnen gebruikt die gemakkelijk toegankelijk zijn. Voorbeeld: de student heeft alleen gebruik gemaakt van • het internet via het 'eenvoudige zoekscherm' van een grote algemene zoekmachine en • lesmateriaal dat door docenten is verstrekt.

4.4 Testen

Bij de test van het eerste concept werd de volgende onderzoeksvraag gebruikt: Wat is volgens beoordelaars die de concept versie van de scoringsrubriek gebruiken bij het beoordelen van studentenproducten de noodzaak om de scoringsrubriek als geheel of afzonderlijke onderdelen daarvan aan te passen?

4.4.1 Methode

De concept versie werd getest tijdens het eerste deel van een ‘internationaal semester’ van de Academie voor ICT & Media van De Haagse Hogeschool. Dit betrof een algemene cursus van 10 weken waarin studenten werden voorbereid op deelname aan een internationale stage of een afstudeertraject in het buitenland. Onderdeel van de cursus was een workshop ‘Efficient information-searching’ met aandacht voor naslagwerken, het gebruik van bibliografische databases, beoordeling van zoekresultaten en verslaglegging van het zoekproces in de vorm van een zoekstrategie. Aan het einde van de workshop kregen de studenten de opdracht om, met behulp van het geleerde, een landenprofiel samen te stellen van het land waar de betreffende student stage wilde gaan lopen. De landenprofielen werden met behulp van de concept scoringsrubriek beoordeeld op de manier waarop de studenten waren omgegaan met het vergaren en verwerken van informatie. Bij een voldoende werden er studiepunten toegekend.

De beoordeling vond plaats door twee docenten van de opleiding bedrijfskundige informatica, de studenten zelf waren afkomstig van verschillende opleidingen binnen het domein ict & media (informatica, technische informatica, bedrijfskundige informatica, communicatie en multimedia design, idm). Uiteindelijk hebben 15 studenten een landenprofiel ingeleverd dat gebruikt kon worden voor de evaluatie van de eerste concept versie van de scoringsrubriek.

De landenprofielen en de bijbehorende procesverslagen (waarin onder andere aandacht besteed moest worden aan de gevolgde zoekstrategie) werden met behulp van de concept scoringsrubriek beoordeeld door de docent die de student ook had begeleid tijdens het schrijfproces. De ene docent beoordeelde acht werkstukken, de andere zeven. Daarnaast werden alle vijftien werkstukken ook beoordeeld door de onderzoeker.

4.4.2 Resultaten

Het aantal beoordeelde uitwerkingen van de studieopdracht (7 respectievelijk 8 voor de docenten en 15 voor de onderzoeker) was niet groot genoeg om statistisch verantwoorde uitspraken te doen over de mate waarin de beoordelingen van de docenten overeenstemden met die van de onderzoeker. De gevonden gegevens met betrekking tot de consensus over de beoordelingen en de informatie uit de evaluatiegesprekken met de twee docenten konden, conform de vraagstelling van het onderzoek, wel worden gebruikt als een indicatie van de noodzaak om de formuleringen van de beoordelingscriteria en de omschrijvingen van het gedrag op drie verschillende niveaus aan te passen.

4.4.3 Revisie van de eerste concept versie

De belangrijkste conclusies die konden worden afgeleid uit de test en uit de bespreking met de betrokken collega's waren:

- Het onderscheid tussen criterium nummer 1 uit tabel 4.1 ('Oriëntatie op het onderwerp') en nummer 2 uit tabel 4.1 ('Gebruik van deskresearch als onderzoeksmethode') was voor de beoordelaars niet duidelijk. Ook de onderzoeker zelf merkte tijdens het beoordelingsproces verschillende malen dat hij moeite had de twee criteria te onderscheiden. Dat heeft er toe geleid dat besloten werd criterium 2 uit tabel 4.1 te laten vervallen. In de uiteindelijke formulering van het professionele gedrag bij criterium 1 werd nog toegevoegd dat een goede oriëntatie op het onderwerp onder andere kan blijken uit duidelijk geformuleerde onderzoeksvragen.
- Bij sommige beoordelingscriteria waren de gedragsomschrijvingen voor de docenten die als beoordelaar deelnamen aan de test niet duidelijk. Een voorbeeld is criterium 4 uit tabel 4.1. De kwalificaties 'te licht' en 'eenzijdig' bij het inadequate gedrag en het gedrag dat voor verbetering vatbaar is, bleken verschillend geïnterpreteerd te worden. In de uiteindelijke versie van het prototype is de betreffende gedragsomschrijving duidelijker gedefinieerd door toevoeging van voorbeelden.

Een ander criterium waarvoor de gedragsbeschrijving in de definitieve versie van het prototype is verduidelijkt, is criterium 6 uit tabel 4.1 dat correspondeert met criterium 5 uit het definitieve prototype. Bij dat criterium werd juist aan de omschrijving van het professionele gedrag een toelichting toegevoegd met betrekking tot het onderscheid tussen de analyse van informatie en het formuleren van een nieuwe toepassing of een nieuw inzicht.

- De beoordelaars hadden soms moeite om te kiezen tussen de drie niveaus. Ze zouden het plezieriger gevonden hebben als er meer te kiezen was geweest, bijvoorbeeld tussen vijf niveaus in plaats van drie. Aanvankelijk is geprobeerd om een of twee niveaus toe te voegen aan de scoringsrubriek. Het was echter niet altijd duidelijk welk gedrag dan lager of juist hoger gewaardeerd moest worden. Een voorbeeld daarvan was het gedrag bij criterium 3 uit tabel 4.1 (Bronnenoverzicht / literatuurlijst). Onduidelijk is daar welk gedrag aanleiding is voor een lagere beoordeling: het ontbreken van titels in de referentielijst of het ontbreken van een correct format voor de titelbeschrijvingen. Na nadere bestudering van de literatuur (onder andere Scharf et al., 2007) werd een alternatief gevonden door de omschrijvingen te beperken tot het 'professionele' en het 'inadequate' gedrag maar vervolgens wel een Likertschaal toe te voegen ten behoeve van het daadwerkelijke 'scoren'. Om te voorkomen dat beoordelaars het 'midden' van de schaal opzoeken is daarbij gekozen voor een even aantal punten, in dit geval 6. Afbeelding 4.4 geeft een voorbeeld uit het prototype, in dit geval voor wat uiteindelijk criterium 2 werd, de 'Bronnenlijst'.

Criterium		Professioneel gedrag			Inadequaaf gedrag			Score 1-10=
2	Bronnen- overzicht / literatuurlijst	☐ Het product bevat een bronnenoverzicht dat volledig is en waarbij een standaard citation style correct is toegepast. Aan de hand van het bronnenoverzicht zijn de gebruikte bronnen eenduidig te achterhalen. Toelichting: het laatste is belangrijker dan het format van de titelbeschrijvingen (de 'citation style'). Voor de score 'uitstekend' dient ook de citation style echter correct te zijn toegepast.			☐ In het product ontbreekt een bronnenoverzicht en / óf ☐ het bronnenoverzicht is verre van volledig (van verschillende documenten waar in de tekst naar wordt verwezen ontbreken de gegevens) en / óf ☐ cruciale bibliografische gegevens (titel, auteur, publicatiejaar) ontbreken. Voorbeeld dat vaak voorkomt in de praktijk: van internetbronnen zijn alleen de url's vastgelegd.			
Beoordeling:		0 uitstekend	0 goed	0 voldoende	0 zwak	0 slecht	0 zeer slecht	

Figuur 4.4

Onderdeel uit het prototype van de Scoringsrubriek voor Informatievaardigheden

4.5 Prototype na revisie

De volledige versie van het prototype van de scoringsrubriek is opgenomen als bijlage 4.1. Voor ieder beoordelingscriterium (kolommen 1 en 2) wordt daarin het professionele en het inadequaat gedrag beschreven. Deze omschrijvingen hebben als doel om aan de beoordelaars duidelijk te maken waar ze bij het betreffende criterium op dienen te letten. Ze zijn voorzien van 'check boxes' die kunnen worden gebruikt om aan te geven welke specifieke vormen van gedrag van toepassing zijn in de betreffende uitwerking van een studieopdracht. In de uiteindelijke vorm van de scoringsrubriek kan 'inadequaaf gedrag' immers op verschillende manieren blijken. Criterium 2 in figuur 4.4 is daar een duidelijk voorbeeld van.

Behalve door te werken met de check boxes kunnen beoordelaars relevante elementen uit de gedragsoomschrijvingen overigens ook markeren, onderstrepen of omcirkelen.

Nadat een beoordelaar een studentenproduct heeft gecontroleerd op het voorkomen van professioneel of inadequaaf gedrag voor een bepaald criterium, wordt hij/zij geacht een beoordeling uit te spreken op de 6 punts Likertschaal die loopt van 'uitstekend' tot 'zeer slecht'. Tenslotte kan de beoordelaar voor ieder afzonderlijk criterium een cijfer toekennen op een schaal van 1-10 of een schaal van 1-20 voor de beoordelingscriteria die verwijzen naar hogere cognitieve vaardigheden, de criteria 1, 3 en 5.

De beoordelingen op de Likertschaal kunnen als onderstaand worden vertaald naar het in Nederland gebruikte cijfersysteem van 1-10 respectievelijk van 1-20:

Uitstekend = 10 / 20

Goed = 8 / 16

Voldoende = 6 / 12

Zwak = 5 / 10

Slecht = 3 / 6

Zeër slecht = 1 / 2.

Het gebruik van de laatste kolom voor het toekennen van een cijfer is overigens niet noodzakelijk. Het nut van de scoringsrubriek is vooral gelegen in de onderbouwing van de beoordeling en een beoordelaar kan zich dus beperken tot de Likertschaal, zolang duidelijk is dat de beoordeling is gebaseerd op het waargenomen professionele en/of inadequate gedrag.

4.6 Conclusies en discussie

Het beoordelingsinstrument dat in deze deelstudie is ontwikkeld, is een analytische scoringsrubriek. De onderzoeksliteratuur die in hoofdstuk 3 werd besproken suggereert dat dit een geschikt instrument is voor de beoordeling van het daadwerkelijke informatiegedrag van studenten bij het oplossen van informatieproblemen in het kader van hun studietaken.

Het vaststellen van de te gebruiken criteria en de beschrijving van het gedrag op elk van die criteria op twee verschillende niveaus (professioneel en inadequaar), vond plaats gedurende een iteratief proces waarbij verschillende onderzoeksactiviteiten drie maal achtereenvolgens afwisselend werden toegepast. Het ontwerp van de scoringsrubriek was daardoor ook een duidelijk voorbeeld van een cyclisch proces dat volgens Wester en Peters zo kenmerkend is voor kwalitatief onderzoek (Wester & Peters, 2004, pp. 83-86).

Onderdeel van het constructieproces was bovendien een test waarvan de resultaten aanleiding gaven tot een revisie van het eerste concept van de scoringsrubriek. Belangrijk onderdeel van de revisie was de toevoeging van een 6 punts Likertschaal voor het scoren op de afzonderlijke criteria. Omdat op de Likertschaal oordelen worden uitgesproken in de vorm van tekst en niet in de vorm van cijfers, is juist dit onderdeel van de scoringsrubriek geschikt voor formatieve beoordelingen die voor de student vooral een informerend karakter hebben.

Bijlage 4.1:

Prototype van de Scoringsrubriek voor Informatievaardigheden

Onderdeel: studentenproduct

Criterium		Professioneel gedrag			Inadequaaf gedrag			
1	Oriëntatie op het onderwerp	☐ Uit het product blijkt dat de student zich goed op het vraagstuk of het onderwerp heeft georiënteerd en op basis daarvan een eigen invalshoek (focus) heeft geformuleerd. Dat blijkt ook uit het feit dat de student duidelijk heeft geformuleerd welke (onderzoeks)vragen hij/zij wil beantwoorden.			☐ Uit het product blijkt dat de student is uitgegaan van de oorspronkelijke vraag of de oorspronkelijke formulering van het onderwerp (zoals die zijn gegeven in het assignment) zonder dat hij/zij die verder heeft uitgediept. Dat blijkt bijvoorbeeld uit het feit dat de betekenis van kernbegrippen bekend wordt verondersteld en deze niet nader worden gedefinieerd.			Score 1-20=
Beoordeling:		0 uitstekend	0 goed	0 voldoende	0 zwak	0 slecht	0 zeer slecht	

Criterium		Professioneel gedrag			Inadequaaf gedrag			Score 1-10=
2	Bronnen- overzicht / literatuurlijst	☐ Het product bevat een bronnenoverzicht dat volledig is en waarbij een standaard citation style correct is toegepast. Aan de hand van het bronnenoverzicht zijn de gebruikte bronnen eenduidig te achterhalen. Toelichting; het laatste is belangrijker dan het format van de titelbeschrijvingen (de 'citation style'). Voor de score 'uitstekend' dient ook de citation style echter correct te zijn toegepast.			☐ In het product ontbreekt een bronnenoverzicht en / óf ☐ het bronnenoverzicht is verre van volledig (van verschillende documenten waar in de tekst naar wordt verwezen ontbreken de gegevens) en / óf ☐ cruciale bibliografische gegevens (titel, auteur, publicatiejaar) ontbreken. Voorbeeld dat vaak voorkomt in de praktijk: van internetbronnen zijn alleen de url's vastgelegd.			
Beoordeling:		0 uitstekend	0 goed	0 voldoende	0 zwak	0 slecht	0 zeer slecht	

Criterium		Professioneel gedrag			Inadequaaf gedrag			
3	Kwaliteit van de gebruikte primaire informatie-bronnen (boeken, artikelen, websites etc.)	☐ Uit het bronnenoverzicht dat bij het product behoort, blijkt dat de student gebruik heeft gemaakt van relevante, betrouwbare (bij voorkeur zo authentiek mogelijke) en actuele informatiebronnen die het onderwerp bovendien vanuit verschillende gezichtspunten behandelen.			☐ De gebruikte externe bronnen zijn te licht, verouderd of onvoldoende relevant. Een voorbeeld van 'te licht' is als de student alleen internetsites als informatiebron heeft gebruikt. En / óf ... ☐ De gebruikte literatuur is eenzijdig. Er is bijvoorbeeld alleen gebruik gemaakt van overheidsinformatie (.gov-sites) of van publicaties van één bepaalde auteur.			Score 1-20=
Beoordeling:		0 uitstekend	0 goed	0 voldoende	0 zwak	0 slecht	0 zeer slecht	

Criterium		Professioneel gedrag	Inadequaat gedrag	
4	Bronverwijzingen in de tekst	☐ In de tekst van het product zelf heeft de student eenduidig verantwoord welke informatiebronnen hij of zij heeft gebruikt. Dat geldt ook voor afbeeldingen of voor audiovisuele illustraties in het geval van een digitaal product.	☐ De student heeft werk van anderen (stukken tekst, afbeeldingen, geluidsfragmenten) verwerkt in zijn/haar eigen product zonder dat hij/zij de gebruikte bron heeft vermeld. Hoewel vaak niet opzettelijk, is er strikt gesproken sprake van 'plagiaat'.	Score 1-10=
Beoordeling:		0 uitstekend 0 goed 0 voldoende	0 zwak 0 slecht 0 zeer slecht	

Criterium		Professioneel gedrag	Inadequaat gedrag	
5	Verwerking van relevante informatie tot nieuwe kennis en/of inzichten	☐ Uit het product blijkt dat de student informatie uit verschillende bronnen heeft geanalyseerd en op basis daarvan nieuwe toepassingen, inzichten of hypotheses heeft geformuleerd. Toelichting: in praktijk blijkt dat veel studenten wel toekomen aan de analyse en vergelijking van verschillende informatiebronnen, maar niet in staat zijn om de gevonden gegevens te synthetiseren tot een nieuw inzicht of nieuwe toepassing. Dit criterium wordt in dat geval beoordeeld als 'voldoende' of 'zwak'.	In het studentenproduct heeft de student ☐ de inhoud van de gebruikte informatiebronnen niet correct of niet helder weergegeven, en / óf ☐ in het geheel geen aandacht besteed aan analyse van de opgespoorde informatiebronnen, en / óf ☐ zich gebaseerd op slechts één informatiebron en hij/zij heeft de daar gevonden informatie niet ter discussie gesteld terwijl daar wel aanleiding toe is.	Score 1-20=
Beoordeling:		0 uitstekend 0 goed 0 voldoende	0 zwak 0 slecht 0 zeer slecht	

Scoringsrubriek voor Informatievaardigheden

Onderdeel: procesverslag / zoekstrategie

Criterium		Professioneel gedrag	Inadequaat gedrag	
6	Zoektermen	☐ De student heeft de zoektermen gebruikt die relevant zijn voor het te onderzoeken onderwerp / vraagstuk. Hij/zij heeft daarbij rekening gehouden met het vakjargon, de synoniemen en (zoek)termen in een andere taal (meestal Engels).	☐ De zoektermen die de student heeft gebruikt zijn te algemeen of komen uit het algemeen spraakgebruik.. en / óf ☐ de student heeft geen rekening gehouden met synoniemen, verwante begrippen en (zoek) termen in een andere taal (meestal Engels).	Score 1-10=
Beoordeling:		0 uitstekend 0 goed 0 voldoende	0 zwak 0 slecht 0 zeer slecht	

Criterium		Professioneel	Inadequaat gedrag	
7	Gebruik van diverse secundaire informatiebronnen	☐ De student heeft gebruik gemaakt van diverse typen secundaire bronnen (search engines, boeken, wetenschappelijke tijdschriften, databases, sociale netwerken). Indien van toepassing heeft hij/zij bovendien gebruik gemaakt van mogelijkheden buiten de eigen locatie (IBL, documentleverantie).	☐ De student heeft vooral informatiebronnen gebruikt die gemakkelijk toegankelijk zijn. Voorbeeld: de student heeft alleen gebruik gemaakt van • het internet via het 'eenvoudige zoek-scherm' van een grote algemene zoekmachine en / of • lesmateriaal dat door docenten is verstrekt.	Score 1-10=
Beoordeling:		0 uitstekend 0 goed 0 voldoende	0 zwak 0 slecht 0 zeer slecht	

Totaalscore (maximaal 100) =

Deel II

**Testen en evalueren van het prototype
van de scoringsrubriek**

Om het prototype van de scoringsrubriek (bijlage 4.1) te testen en te evalueren, is er op verschillende onderwijsinstellingen onderzoek gedaan naar de kwaliteiten van het beoordelingsinstrument en de ervaringen die studenten en docenten er mee hebben opgedaan. Voor de tests in hoofdstuk 5 en 6 is vooral gebruik gemaakt van kwantitatief onderzoek. De nadruk lag daarbij op wat Van de Watering en Dierick “klassieke” psychometrische kenmerken van een beoordelingsinstrument noemen: betrouwbaarheid en validiteit (Van de Watering & Dierick, 2002, p. 63). Tijdens deze onderzoeken zijn echter ook wel meer kwalitatieve opmerkingen gemaakt door de docenten en beoordelaars die bij de tests waren betrokken. Deze onbedoelde kwalitatieve uitkomsten van de kwantitatieve onderzoeken zijn uiteraard wel vastgelegd en worden bij de resultaten kort vermeld. Eén van de onderzoeken uit hoofdstuk 5, dat bij Saxion, had wel een meer kwalitatief karakter en richtte zich dan ook specifiek op de bruikbaarheid van de scoringsrubriek en de efficiency in het gebruik.

Voor een meer volledig zicht op de manier waarop de scoringsrubriek kan worden gebruikt en de ervaringen die studenten en docenten er mee hebben opgedaan, was een strikt kwalitatief ingericht onderzoek beter geschikt. Er zijn twee van dergelijke kwalitatieve onderzoeken uitgevoerd die in hoofdstuk 7 en 8 worden beschreven. De opzet van deze onderzoeken was fundamenteel verschillend van die van de kwantitatieve onderzoeken. De vraagstelling was van een andere aard en de resultaten van het kwalitatieve onderzoek en die van het kwantitatieve onderzoek zijn aanvullend op elkaar. Het gebruik van de twee verschillende soorten onderzoeksmethodes kan dan ook worden getypeerd als ‘complementair’ (De Boer, 2006). In het vervolg van deze inleiding tot de hoofdstukken 5 tot en met 8 wordt nu eerst specifiek ingegaan op de onderzochte kenmerken van het prototype zelf. Aan het slot van de inleiding wordt nog stil gestaan bij de achtergronden van het onderzoek naar de manier waarop het door docenten en studenten wordt gebruikt.

Kwantitatief onderzoek

De kwantitatieve onderzoeken hadden vooral betrekking op de betrouwbaarheid en validiteit van het prototype. Betrouwbaarheid wordt in de onderzoeken in hoofdstuk 5 en 6 opgevat als interbeoordelaarsbetrouwbaarheid. Dit begrip verwijst naar de mate van overeenstemming tussen twee of meer beoordelaars als ieder van hen dezelfde set uitwerkingen op het zelfde moment beoordeelt (Stemler, 2004). In de te beschrijven studies worden verschillende indicatoren gebruikt, de meeste gebaseerd op een eenvoudig te bepalen percentage overeenstemming. In een van de onderzoeken, dat bij de opleiding Mediastudies van de UvA dat in hoofdstuk 6 wordt besproken, wordt behalve van zo’n percentage overeenstemming gebruik gemaakt van de Intraclass Correlation Coefficient (ICC). Dit laatste kengetal houdt rekening met de kans dat beoordelaars met elkaar overeenstemmen en is daardoor als statistische maat een betere indicator dan het simpele ‘percentage overeenstemming’. De aantallen beoordeelde werkstukken in de onderzoeken van hoofdstuk 5 waren echter te laag om de ICC ook daar als indicator voor interbeoordelaarsbetrouwbaarheid te gebruiken.

Met betrekking tot de validiteit van het prototype wordt onderscheid gemaakt tussen constructvaliditeit en criteriumvaliditeit. Moskal en Leydens (2000) geven aan dat constructvaliditeit onder andere onderzocht kan worden door na te gaan of alle criteria van een beoordelingsinstrument wel relevant zijn voor het meten van het betreffende onderliggende ‘construct’.

In de onderzoeken in hoofdstuk 5 wordt dit onderzocht door met SPSS de onderlinge consistentie van de verschillende beoordelingscriteria vast te stellen. In het onderzoek dat wordt beschreven in hoofdstuk 6 wordt behalve dit onderzoek naar interne consistentie een factoranalyse uitgevoerd om na te gaan in welke mate de beoordelingscriteria van de scoringsrubriek verwijzen naar één en het zelfde construct.

Onder criteriumvaliditeit wordt verstaan dat de beoordelingen die worden opgemaakt met het ene instrument positief correleren met de beoordelingen die worden opgemaakt met een ander instrument dat beoogt het zelfde construct te meten. Als er sprake is van min of meer gelijktijdige beoordeling met de twee verschillende instrumenten, wordt ook wel gesproken over concurrentvaliditeit (Cronbach & Meehl, 1955). In het onderzoek bij de Open Universiteit en dat bij de opleiding Mediastudies van de UvA kon ook deze criteriumvaliditeit worden onderzocht omdat er naast de scoringsrubriek tevens een beoordelingsmodel beschikbaar was dat de betreffende opleiding zelf al gebruikte.

De derde onderzochte eigenschap van het prototype, de praktische bruikbaarheid, is in de kwantitatieve onderzoeken van hoofdstuk 5 en 6 minder systematisch onderzocht. Zoals ook hierboven al werd opgemerkt, hebben de docenten / beoordelaars die bij het onderzoek waren betrokken wel opmerkingen gemaakt over de gebruiksvriendelijkheid van de scoringsrubriek.

Kwalitatief onderzoek

De ervaringen tijdens het gebruik van de scoringsrubriek zijn expliciet object van onderzoek geweest in de kwalitatieve studies die worden beschreven in hoofdstuk 7 en 8. De onderzoeksvraag heeft in beide hoofdstukken betrekking op het daadwerkelijke gebruik van de scoringsrubriek. In hoofdstuk 7 gaat het om het gebruik door studenten, in hoofdstuk 8 over de manier waarop de scoringsrubriek door docenten wordt gebruikt. In allebei de onderzoeken zijn uitspraken over het gebruik van de scoringsrubriek verzameld door het voeren van gesprekken die werden vooraf gegaan door gericht literatuuronderzoek naar voor- en nadelen van het gebruik van scoringsrubrieken in het algemeen.

Daar waar de betrouwbaarheid en de validiteit worden beschouwd als de “psychometrische eigenschappen” van een beoordelingsinstrument die in een meer traditioneel onderzoek worden onderzocht, worden de kenmerken van het gebruik ervan ook wel opgevat als “edumetrische eigenschappen” die meer thuis horen bij een ‘vernieuwde’ opvatting over toetsen en beoordelen (Dierick, Van de Watering & Muijtjens, 2002, pp. 95-96). Een verschil met de meer klassieke onderzoeken is dat de nadruk niet alleen ligt op het selecterende of certificerende karakter van het beoordelingsinstrument maar ook op de wijze waarop het instrument het onderwijs- en leerproces stimuleert. In dergelijk onderzoek wordt ook de vraag gesteld welke gevolgen het gebruik van het instrument heeft voor het onderwijs en het leerproces van de studenten. Dierick, Van de Watering en Muijtjens spreken daarom ook wel over de “consequentieële validiteit” van het beoordelingsinstrument (p. 117 ff.). De invalshoek van de onderzoeken in hoofdstuk 7 en 8 sluit dan ook goed aan bij de eerder in hoofdstuk 1 geconstateerde tendens om toetsen en beoordelen te beschouwen als een proces dat ook dient om het leerproces van studenten te faciliteren (“assessment for learning”).

5

Evalueren van de bruikbaarheid bij Saxion; Testen van betrouwbaarheid en validiteit bij de Nederlandse Defensie Academie en Open Universiteit

Abstract: Voor het vaststellen van de kwaliteiten van het prototype van de scoringsrubriek uit hoofdstuk 4, is gebruik gemaakt van drie verschillende contexten. Onder acht docenten van de opleiding commerciële economie van Saxion is een schriftelijke vragenlijst uitgezet met als doel te achterhalen of de scoringsrubriek voor hen een bruikbaar beoordelingsinstrument was. Bij de Nederlandse Defensie Academie werd aan de hand van beoordelingen die twee docenten met behulp van de scoringsrubriek hadden opgemaakt, een analyse uitgevoerd van de interbeoordelaarsbetrouwbaarheid en de interne consistentie van de beoordelingen. Een soortgelijke analyse van interbeoordelaarsbetrouwbaarheid en interne consistentie werd uitgevoerd bij de faculteit Psychologie en Onderwijswetenschappen van de Open Universiteit. Bij deze laatste instelling werd ook een onderzoek uitgevoerd naar de criteriumvaliditeit van de scoringsrubriek door de correlatie vast te stellen tussen beoordelingen die werden opgemaakt met de scoringsrubriek en de beoordelingen die werden opgemaakt met behulp van een beoordelingsmodel van de Open Universiteit zelf.

De zeven criteria van het prototype uit hoofdstuk 4 werden door de docenten van Saxion en die van de Nederlandse Defensie Academie ervaren als een valide afspiegeling van het construct informatievaardigheid. De uitwerkingen op het niveau van professioneel en inadequaat gedrag werden bovendien ervaren als een goed hulpmiddel om tijdens het beoordelingsproces goed gefocust te blijven. Zowel bij de Nederlandse Defensie Academie als bij de Open Universiteit werd een zeer behoorlijke aanpalende overeenstemming tussen twee verschillende beoordelaars bereikt. De interne consistentie was wisselend voor de vier beoordelaars maar voor sommigen van hen juist zeer hoog. Deze beoordelaars waren dus in staat om de scoringsrubriek consequent en strikt toe te passen. De correlatie tussen de beoordelingen met de scoringsrubriek en die met het OU beoordelingsmodel was significant positief.

Inleiding

Aan een beoordelingsinstrument worden voor gebruik in het onderwijs traditioneel eisen gesteld op het gebied van betrouwbaarheid, validiteit en gebruiksgemak (Van de Watering en Dierick, 2002, p. 63). Om de praktische bruikbaarheid, de interbeoordelaarsbetrouwbaarheid en de validiteit van het prototype van de scoringsrubriek vast te stellen, is het instrument uit bijlage 4.1 in drie verschillende contexten onderzocht:

- bij de opleiding Commerciële Economie van Saxion te Enschede (School of Marketing & International Management)
- bij de opleiding Militaire Bedrijfs Wetenschappen van de Nederlandse Defensie Academie
- en bij de cursus Informatievaardigheden voor sociale wetenschappers van de Open Universiteit (faculteit Psychologie en onderwijswetenschappen).

Binnen ieder van deze contexten werd onderzoek gedaan naar het beoordelen van authentiek studentenwerk met behulp van het prototype van de scoringsrubriek. Voor de drie verschillende settings was het aantal beschikbare beoordelaars en / of het aantal voorbeelden van studentenwerk steeds anders. De behandelde onderzoeksvragen en de gebruikte onderzoeksmethodes waren dan ook voor ieder van de drie situaties verschillend. Bij Saxion lag het accent van het onderzoek op de praktische bruikbaarheid en de efficiëntie van de scoringsrubriek. De onderzoeksvragen bij de Nederlandse Defensie Academie en de Open Universiteit Nederland richtten zich op de interbeoordelaarsbetrouwbaarheid en de interne consistentie van de beoordelingen. Bij de Open Universiteit werd bovendien onderzoek gedaan naar de criteriumvaliditeit van de scoringsrubriek. De drie onderzoeken worden hieronder in verschillende paragrafen besproken. In paragraaf 5.4 worden op basis van alle drie de onderzoeken conclusies getrokken met betrekking tot de kwaliteiten van de scoringsrubriek.

5.1 Evaluatie bij Saxion¹¹

In het onderzoek bij Saxion werd antwoord gezocht op de volgende drie onderzoeksvragen:

- a. In welke mate geeft de scoringsrubriek een valide beeld van de informatievaardigheden van een student?
- b. Hoe bruikbaar is de scoringsrubriek in de onderwijspraktijk?
- c. Hoe efficiënt is de scoringsrubriek als beoordelingsinstrument?

Dat laatste werd geïnterpreteerd als de tijd die beoordelaars nodig hebben om de beoordeling op te maken.

5.1.1 Methode

Acht docenten van de bachelor opleiding Commerciële Economie, Academie Marketing & International Management, waren bij het onderzoek betrokken. Allen hadden in het betreffende

¹¹ De inhoud van paragraaf 5.1 en 5.2 is eerder gepubliceerd in: A scoring rubric for performance assessment of information literacy in Dutch Higher Education. *Journal of Information Literacy*, 4 (1), 22-39 (Van Helvoort, 2010).

studieproject de rol van ‘tutor’ maar geen van hen had vanuit zichzelf een specifieke betrokkenheid bij het onderwerp informatievaardigheid.

De gebruikte studieopdracht was onderdeel van een reguliere onderwijsseenheid uit het eerste jaar van het major onderwijs. De opdracht hield in dat er een externe analyse voor een bedrijf of not-for-profitorganisatie werd uitgevoerd waarbij gebruik werd gemaakt van deskresearch. De eerste vijf criteria van de scoringsrubriek werden door de docenten gebruikt bij een summatieve beoordeling van de in te leveren “individuele procesverslagen”. Criteria 6 en 7 van de scoringsrubriek werden door hen niet gebruikt omdat de studenten geen zoekstrategie hoefden op te leveren.

Aan het begin van het project volgden alle acht docenten een instructie van een onderwijskundig onderzoeker die op dat moment zelf in het kader van een ander project onderzoek deed naar het onderwerp assessment van informatievaardigheid. De docenten werd de betekenis en de reikwijdte van ieder beoordelingscriterium uitgelegd om ervoor te zorgen dat de verschillende beoordelaars zich een gemeenschappelijk beeld zouden vormen van het oplossen van informatieproblemen zoals besproken in hoofdstuk 2. Ieder van de docenten beoordeelde uiteindelijk vijf procesverslagen van studenten met het prototype van de scoringsrubriek. Vervolgens beantwoordden ze een korte survey met stellingen en vragen. De survey zelf is opgenomen als bijlage 5.1 aan het einde van dit hoofdstuk.

5.1.2 Resultaten

Tabel 5.1 geeft een schematisch overzicht van de antwoorden die de docenten van Saxion hebben gegeven.

Tabel 5.1

Antwoorden op de stellingen en vragen die zijn voorgelegd aan de docenten van de opleiding Commerciële Economie van Saxion

Deelnemer	001	002	003	004	005	006	007	008
Criteria zijn alle 5 noodzakelijk	eens	eens	oneens	eens	eens	oneens	eens	oneens
Criteria zijn voldoende	eens	eens	oneens	eens	eens	eens	eens	eens
Formulering gedrag is duidelijk	oneens	oneens	oneens	eens	eens	eens	eens	eens
Rubriek zorgde ervoor dat ik wist waar ik op moest letten	eens	eens	eens	oneens	eens	eens	eens	eens
Ik ga de scoringsrubriek gebruiken voor summatieve toetsing	eens	oneens	oneens	eens	oneens	weet niet	oneens	oneens
Ik ga de scoringsrubriek gebruiken voor instructie of feedback	eens	oneens	oneens	eens	eens	eens	eens	eens
Benodigde tijd in minuten voor het invullen van de scoringsrubriek	<5	<5	<5	<5	<5	5-10	<5	5-10

Uit het overzicht en de toelichtingen die de respondenten bij hun antwoorden gaven, werden de volgende resultaten afgeleid:

- Vijf van de acht beoordelaars waren van mening dat alle vijf criteria nodig waren voor het beoordelen van de informatievaardigheid van studenten. Van de drie die dat niet vonden, vond één dat criterium 5 ('verwerking van relevante informatie tot nieuwe kennis') te hoog gegrepen was voor eerstejaars studenten en een ander dat criterium 3 ('kwaliteit van de gebruikte primaire informatiebronnen') kon worden ondergebracht bij een van de andere criteria.
- Zeven van de acht beoordelaars waren van mening dat er geen beoordelingscriterium werd gemist. De ene beoordelaar die dat niet vond, heeft niet aangegeven wat hij/zij miste.
- Vijf beoordelaars vonden de formulering van het professionele en het inadequate gedrag in alle gevallen duidelijk. Van de drie die dat niet vonden lichtte één toe dat hij/zij dat moeilijk te beoordelen vond. Een ander antwoordde dat hij/zij het plezierig had gevonden als ook het gedrag op de tussenliggende niveaus was omschreven. Op dit punt wordt teruggekomen in paragraaf 5.4 (Algehele conclusies en discussie).
- Zeven van de acht beoordelaars waren van mening dat de scoringsrubriek hen had geholpen om goed gefocust te blijven op de in het kader van informatievaardigheid te beoordelen aspecten van het betreffende studentenwerk.
- Zes van de acht beoordelaars gaven aan dat ze minder dan 5 minuten per procesverslag nodig hadden gehad om de beoordeling met behulp van de scoringsrubriek op te maken. De andere twee respondenten gaven aan daarvoor 5 tot 10 minuten nodig gehad te hebben. In de vragenlijst waren ook de opties '10-15 minuten' en '> 15 minuten' aanwezig maar geen van de respondenten heeft deze als antwoord gebruikt.

De vraag aan de docenten van de opleiding Commerciële Economie bij Saxion of zij de scoringsrubriek ook in een later stadium nog zouden gebruiken voor instructie of het formatief (diagnostisch) beoordelen van studentenwerk, werd door zes van de acht positief beantwoord. De meesten van hen waren er echter niet zeker van of ze de scoringsrubriek ook zouden gebruiken voor een summatieve beoordeling ('voor een cijfer'). Geen van de respondenten heeft dit echter gemotiveerd en omdat de enquête anoniem werd afgenomen kon hier geen navraag naar worden gedaan.

5.1.3 Conclusies

Drie van de acht geënuquêteerde docenten vonden dat de vijf gebruikte beoordelingscriteria niet allemaal echt noodzakelijk waren voor het beoordelen van de informatievaardigheid van studenten. Uit de toelichtingen van twee van hen blijkt dat echter iets genuanceerder te liggen dan het in eerste instantie lijkt. Deze twee vonden de betreffende criteria (criterium 3 'kwaliteit van de gebruikte primaire informatiebronnen' en criterium 5 'verwerking van relevante informatie tot nieuwe kennis') wel relevant maar ze vonden ook dat het betreffende criterium

- óf nog niet van toepassing is in het eerste jaar van het bachelor onderwijs (één van hen over criterium 5)
- óf onvoldoende onderscheidend was om als apart criterium te worden opgenomen (de ander over criterium 3).

Deze twee respondenten gaven in de survey dus aan dat zij het niet eens waren met de stelling dat alle beoordelingscriteria nodig waren, maar trokken de validiteit van het betreffende criterium voor het beoordelen van de informatievaardigheid van studenten niet in twijfel. Dat geldt wel voor de derde respondent die op deze stelling negatief reageerde. Deze respondent vond de formuleringen “onnodig lang”.

Met betrekking tot de praktische bruikbaarheid gaf één van de respondenten aan dat hij/zij het plezierig had gevonden als er meer dan twee niveaus van gedragsbeschrijvingen waren gebruikt. Deze opmerking raakt direct aan een van de eerder genomen ontwerpbeslissingen uit hoofdstuk 4. Om die reden wordt de discussie over dit onderwerp uitgesteld tot paragraaf 5.4 waar de meer algemene discussie over de resultaten uit alle drie de deelonderzoeken van dit hoofdstuk zal worden gevoerd.

5.2 Test bij de Nederlandse Defensie Academie

Bij dit deelonderzoek stond de vraag naar de interbeoordelaarsbetrouwbaarheid centraal. Dit begrip verwijst naar de mate van overeenstemming tussen twee of meer beoordelaars als ieder van hen de zelfde set uitwerkingen op het zelfde moment beoordeelt (Stemler, 2004). Behalve op de absolute overeenstemming kan daarbij ook worden gelet op de ‘aanpalende overeenstemming’, wat betekent dat de beoordelaars niet meer dan één punt van elkaar verschillen op een beoordelingschaal (‘adjacent agreement’).

Volgens Stemler (2004) is het heel gebruikelijk om de ‘aanpalende overeenstemming’ te gebruiken als maatstaf voor het uitdrukken van consensus over beoordelingen tussen twee of meer beoordelaars. Dat is volgens hem met name zo omdat het trainen en normeren van beoordelaars te veel tijd zou kosten als men daarmee door zou willen gaan tot het moment dat absolute overeenstemming wordt bereikt.

Een derde maatstaf om interbeoordelaarsbetrouwbaarheid vast te stellen (naast ‘absolute’ en ‘aanpalende overeenstemming’), is volgens Stemler (2004) de onderlinge consistentie tussen de beoordelingen. Met dat laatste begrip wordt bedoeld dat de relatieve rangorde van de beoordeelde producten, als ze van hoog naar laag beoordeeld worden gerangschikt, voor beide beoordelaars met elkaar correleren. Studentenproducten die door beoordelaar nummer 1 hoog worden beoordeeld, worden met andere woorden ook door beoordelaar 2 hoog beoordeeld, ook al hoeven de twee beoordelaars geen exacte overeenkomst in het toegekende cijfer te hebben.

De mate van overeenstemming werd niet alleen berekend voor de daadwerkelijke scores van de beoordelaars maar tevens voor de behaalde voldoende respectievelijk onvoldoendes. Deze berekening van de consistentie in het geven van voldoende respectievelijk onvoldoendes door beoordelaars werd ingegeven door Nitko en Brookhart die het gebruik van deze “pass / fail decisions” bepleiten (Nitko & Brookhart, 2007, p. 79). De ‘decision consistency’ is uiteraard vooral van belang bij summatieve beoordelingen waar de beslissing om een voldoende of een onvoldoende te geven grote consequenties kan hebben voor de voortgang van de studie van de betreffende studenten.

Een tweede onderzoeksvraag die tijdens het onderzoek bij de Nederlandse Defensie Academie werd beantwoord was die naar de interne consistentie van de beoordelingen die de afzonderlijke beoordelaars hadden opgemaakt. Daarmee wordt bedoeld op de vraag of de beoordelingscriteria gezamenlijk één goede schaal vormen. Interne consistentie kan worden bepaald door middel van statistische analyse en dan met name door het vaststellen van de 'item-totaal correlaties'. Een hoge interne consistentie impliceert de constructvaliditeit van het meetinstrument, in dit geval de scoringsrubriek (Moskal & Leydens, 2000 ; Gormally, Brickman & Lutz, 2012, pp. 365-366).

Samenvattend kan worden gesteld dat in het onderzoek bij de Nederlandse Defensie Academie antwoord werd gezocht op de volgende twee onderzoeksvragen:

- a. Wat is de interbeoordelaarsbetrouwbaarheid van de scoringsrubriek als twee verschillende beoordelaars de zelfde werkstukken beoordelen?
- b. Wat is de interne consistentie van de beoordelingen die zijn opgemaakt met de scoringsrubriek?

5.2.1 Methode

De scoringsrubriek werd gebruikt door twee docenten Informatiekunde van de opleiding militaire bedrijfswetenschappen, een wetenschappelijke studie op bachelorniveau. De docenten beoordeelde ieder de zelfde set van 27 studentenproducten die waren gemaakt in het kader van de opdracht om een literatuuroverzicht samen te stellen over een zelf te kiezen onderwerp met betrekking tot ict en de samenleving. Het verzamelen van informatie door middel van deskresearch en het schrijven van de literatuurstudie behoorden beide tot de leerdoelen van de betreffende onderwijseenheid (Informatiekunde, cursushandleiding 2008-2009, p. 6). Bij de beoordeling met behulp van de scoringsrubriek maakten de beoordelaars zowel gebruik van de 6 punts Likertschaal als van de laatste kolom uit het prototype om te komen tot een eindbeoordeling op een schaal van 1 tot 10.

Beide docenten hadden zelf een achtergrond in de informatiewetenschap en waren behoorlijk vertrouwd met het begrip informatievaardigheid. Er werd dan ook besloten om voor hen geen trainingssessie te organiseren waarin de betekenis en de relevantie van de afzonderlijke beoordelingscriteria nog eens werden uitgelegd. Wel werd een instructie samengesteld waarin werd toegelicht hoe ze de scoringsrubriek geacht werden te gebruiken. Het werk aan de studieopdrachten door de studenten werd gedurende een periode van zeventien weken gecoacht door de zelfde twee docenten die aan het einde van het project de scoringsrubriek gebruikten voor de eindbeoordeling.

Omdat de twee beoordelaars van de Nederlandse Defensie Academie een gemeenschappelijke achtergrond als informatiewetenschapper hadden en bekend waren met het verschijnsel informatievaardigheid, was de veronderstelling vooraf dat zij tot een behoorlijk hoge consensus en consistentie in de beoordelingen zouden geraken.

De door hen ingevulde beoordelingsformulieren werden gekopieerd en opgestuurd naar de onderzoeker / samensteller van de scoringsrubriek die ze vervolgens verwerkte met Microsoft Excel en SPSS. De data werden op interbeoordelaarsbetrouwbaarheid geanalyseerd door het genereren van overzichten in de vorm van kruistabellen, zowel voor de scores op de afzonderlijke beoordelingscriteria als voor de totaalscores. Voor het onderzoek naar de interne consistentie van de scorings-

rubriek werd met behulp van SPSS Cronbach's Alpha vastgesteld en vervolgens 'Cronbach's Alpha indien Item Verwijderd' en de 'Gecorrigeerde Item-Totaal Correlaties'. Deze twee indicatoren kunnen worden opgeroepen bij de procedure Betrouwbaarheidsanalyse binnen SPSS.

De resultaten van het onderzoek zijn gerapporteerd aan de twee betrokken respondenten en vervolgens zijn ze in een fysieke bijeenkomst met hen bediscussieerd.

5.2.2 Resultaten

Interbeoordelaarsbetrouwbaarheid

De eerste onderzoeksvraag had betrekking op de mate van overeenstemming tussen de twee beoordelaars. Tabel 5.2 geeft een overzicht van de verschillende soorten consensus die werden bereikt voor de afzonderlijke criteria en voor de eindbeoordelingen.

Tabel 5.2

Verschillende vormen van consensus tussen beoordelaar 1 en beoordelaar 2 bij de Nederlandse Defensie Academie

Criterium	N =	Absolute overeenstemming	Aanpalende overeenstemming	Overeenstemming over de beslissing voldoende - onvoldoende ("Decision consistency")
1 Oriëntatie	27	11 (41%)	24 (89%)	23 (85%)
2 Literatuurlijst	27	7 (26%)	23 (85%)	18 (67%)
3 Primaire bronnen	27	11 (41%)	23 (85%)	27 (100%)
4 Bronverwijzingen	27	12 (44%)	23 (85%)	20 (74%)
5 Informatieverwerking	27	5 (19%)	23 (85%)	19 (70%)
6 Zoektermen ^a	25	14 (56%)	24 (96%)	23 (92%)
7 Secundaire bronnen ^a	25	4 (16%)	17 (68%)	23 (92%)
Eindbeoordeling (1-10)	27	3 (11%)	22 (81%)	25 (93%)

^a Criteria 6 en 7 konden voor slechts 25 studenten worden beoordeeld omdat twee studenten geen zoekstrategie hadden opgeleverd.

Kolom 3 van de tabel geeft aan dat de twee beoordelaars slechts in een minderheid van de gevallen het helemaal met elkaar eens waren. De aanpalende overeenstemming was echter zowel voor de eindbeoordeling (>80% op de schaal van 1-10) als voor vrijwel alle afzonderlijke beoordelingscriteria (85% of meer op de 6 punts Likertschaal) acceptabel. Ballator, Farnum en Kaplan (1999, p. 38) geven aan dat op een 6 punts schaal een "adjacent agreement" van 80% over het algemeen als acceptabel wordt beschouwd.

Wat niet wordt weergegeven in tabel 5.2 maar wel het vermelden waard is, is dat ook de onderlinge consistentie tussen de twee eindbeoordelingen met een Cronbach's Alpha van 0,78 goed te noemen was. Alleen voor criterium 7 (het gebruik van secundaire bronnen) gold dat het percentage aanpalende overeenstemming veel lager was dan gewenst. In paragraaf 5.2.3 (de paragraaf met 'conclusies') wordt daarvoor een verklaring gegeven.

De overeenstemming over de beoordeling voldoende / onvoldoende was bij de criteria 2, 4 en 5 lager dan 85 %. Ook dat wordt als problematisch ervaren en is aanleiding om er in de conclusieparagraaf op terug te komen.

Interne consistentie

Voor ieder van de beoordelaars afzonderlijk werd vastgesteld in welke mate de beoordelingen op de afzonderlijke criteria consistent waren met de totale beoordeling door middel van een betrouwbaarheidsanalyse. Cronbach's Alpha voor docent 1 kwam uit op 0,877, die voor docent 2 op 0.613. Voor de twee beoordelaars gezamenlijk kwam het gemiddelde daarmee uit op 0.745.

Daarnaast werden voor ieder van de beoordelaars de zogenaamde 'Item-Totaal Statistieken' berekend (tabel 5.3 en 5.4). De gerapporteerde 'Gecorrigeerde Item-Totaal Correlatie' geeft antwoord op de vraag in welke mate de scores op de afzonderlijke criteria correleren met de totaalscore. De uitkomsten per criterium zijn onder andere een indicatie voor de mate waarin het betreffende criterium heeft bijgedragen aan de beoordeling van het volledige construct en zijn daarmee tevens een maat voor de 'constructvaliditeit' van het beoordelingsinstrument.

Tabel 5.3

Item-Totaal Statistieken voor NDA beoordelaar 1

Criterium	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd^a
1 Oriëntatie	,664	,859
2 Literatuurlijst	,714	,853
3 Primaire bronnen	,601	,869
4 Bronverwijzingen	,488	,884
5 Informatieverwerking	,683	,857
6 Zoektermen	,763	,845
7 Secundaire bronnen	,766	,845

^a Cronbach's Alpha: 0,877

Tabel 5.4

Item-Totaal Statistieken voor NDA beoordelaar 2

Criterium	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd^a
1 Oriëntatie	,325	,581
2 Literatuurlijst	,268	,614
3 Primaire bronnen	,283	,594
4 Bronverwijzingen	,387	,562
5 Informatieverwerking	,488	,518
6 Zoektermen	,474	,537
7 Secundaire bronnen	,185	,614

^a Cronbach's Alpha: 0,613

De gemiddelde alpha coëfficiënt voor de twee beoordelaars samen (0.745) mag volgens de geraadpleegde literatuur in ieder geval als acceptabel beschouwd worden (Jonsson & Svingby 2007, p. 133-134). Daarbij moet wel opgemerkt worden dat voor beoordelaar 2 de coëfficiënt is berekend voor 25 in plaats van 27 studentenproducten omdat bij 2 studenten geen gegevens ingevuld waren voor de criteria 6 en 7 (zie ook de opmerking bij tabel 5.2).

Beoordelaar 1 blijkt in staat om de scoringsrubriek heel consistent te gebruiken. Hij behaalt met zijn beoordelingen niet alleen een hoge Cronbach's Alpha (0,877), ook de 'Gecorrigeerde Item-Totaal Correlaties' zijn bijzonder goed wat erop wijst dat alle criteria correleren met de totaalscore. De 'Gecorrigeerde Item-Totaal Correlaties' in tabel 5.4 laten zien dat beoordelaar 2 vooral onder de maat presteert bij criterium 7. De waarde 0,185 ligt onder 0,20 wat wordt beschouwd als het minimum om nog acceptabel te zijn (Everitt, 2002, p. 196). De beoordelingen van beoordelaar 2 op het betreffende criterium ('Gebruik van diverse secundaire informatiebronnen') zijn dus niet consistent met die op de andere criteria.

Na afloop van de test op de Nederlandse Defensie Academie zijn de resultaten uit het onderzoek besproken met de twee betrokken docenten. De belangrijkste conclusies uit deze evaluatie waren dat de zeven beoordelingscriteria goed gekozen waren en dat ze gezamenlijk ook voldoende waren om de informatievaardigheid van studenten te meten. De beoordelaars misten echter criteria voor de meer formele aspecten van het literatuuroverzicht (vormgeving, spelling en grammatica), onderdelen van een studieopdracht die inderdaad niet zijn opgenomen in het concept van informatievaardigheid dat in dit onderzoek is gehanteerd. De gedragsomschrijvingen op twee niveaus in de scoringsrubriek waren volgens de betreffende beoordelaars weer wel bijzonder zinvol en hielpen hen bij het gebruik van het scoringsschema in de praktijk.

5.2.3 Conclusies

De aanpalende overeenstemming van de beoordelaars was in het onderzoek bij de Nederlandse Defensie Academie voldoende voor zes van de zeven beoordelingscriteria én voor de eindbeoordeling. Alleen voor criterium 7 (gebruik van secundaire bronnen) geldt dat niet. Voor dat criterium was de aanpalende overeenstemming van 68% onder de maat. Tijdens de nabespreking werd als verklaring daarvoor genoemd dat het gebruik van bepaalde (met naam genoemde) wetenschappelijke databases één van de eisen was die uitdrukkelijk in de opdracht vermeld had gestaan. Dat leidde er toe dat beoordelaar 2 steeds betrekkelijk hoge scores toekende voor het betreffende criterium terwijl de beoordelaar 1 zijn aandacht bij dit criterium juist had gericht op de vraag of de student ook andere wetenschappelijk bronnen had gebruikt dan die welke in de opdracht als verplicht te gebruiken vermeld hadden gestaan. Dat verschil in benadering komt ook tot uitdrukking in de item-totaal statistieken voor beoordelaar 2. Het is juist bij dit criterium dat beoordelaar 2 scores toekent die afwijken van die op de andere criteria.

Bij de beoordeling van de literatuuroverzichten met behulp van de scoringsrubriek legden de docenten van de Nederlandse Defensie Academie de cesuur, zoals ook gesuggereerd door de onderzoeker in de instructie die zij vooraf hadden ontvangen, op het midden van de Likertschaal (de beslissing tussen 'voldoende' of 'zwak'). Hier moet worden opgemerkt dat deze cesuur kan

verschuiven afhankelijk van de studieopdracht die wordt gebruikt en de fase van het curriculum waarin de scoringsrubriek wordt gebruikt. Zo kan bijvoorbeeld voor eerstejaars studenten worden afgesproken dat de score 'zwak' op één bepaald criterium nog goed genoeg is om de studiepunten voor het betreffende onderdeel toegekend te krijgen.

In het geval van de casus bij de Nederlandse Defensie Academie bleek de overeenstemming over de beslissing voldoende-onvoldoende (kolom 5 van tabel 5.2) bij drie beoordelingscriteria (criterium 2, 4 en 5) op minder dan 85% te liggen, wat wordt ervaren als problematisch. Bij de evaluatie van de onderzoeksresultaten met de twee beoordelaars gaven zij aan dat ze tijdens het beoordelingsproces inderdaad fouten en tekortkomingen over het hoofd hadden gezien, iets wat de beoordelaars makkelijk kan overkomen bij de beoordelingscriteria 2 en 4 (het bronnenoverzicht en de bronverwijzingen in de tekst). Voor de betrekkelijk geringe mate van overeenstemming over de beslissing voldoende-onvoldoende bij beoordelingscriterium 5 (de verwerking van relevante informatie tot nieuwe kennis) is een andere verklaring meer voor de hand liggend. Dit criterium is van zichzelf meer vatbaar voor subjectieve interpretatie door de beoordelaars. Deze uitkomsten van dit deelonderzoek leidden tot de belangrijke conclusie dat het aan te raden is om de scoringsrubriek voor informatievaardigheden in haar geheel te gebruiken op het moment dat zij wordt gebruikt voor summatieve beoordelingen waar studiepunten aan verbonden zijn. Bij gebruik van afzonderlijke criteria als een 'primary trait rubric' (zie paragraaf 3.2.4 voor dat begrip) loopt de beoordeling blijkbaar het risico minder betrouwbaar te worden.

In het geval van de besproken casus blijken eindbeoordelingen die zijn gebaseerd op alle zeven beoordelingscriteria tezamen over het algemeen dus betrouwbaarder te zijn dan beoordelingen die zijn gebaseerd op afzonderlijke criteria. Dit wordt nog eens bevestigd door het feit dat ook de consistentie in beoordeling (de correlatie tussen de relatieve rangordes) beter was voor de eindbeoordelingen dan voor de meeste afzonderlijke beoordelingscriteria. De onderlinge consistentie van de beoordelingen kan worden uitgedrukt door middel van het kengetal Cronbach's Alpha (Stemler, 2004). In het onderhavige geval werd voor de eindbeoordelingen een waarde van 0,78 (N=25) bereikt wat kan worden beschouwd als betrekkelijk goed. De relatief hoge consistentie bij de eindbeoordelingen wordt zonder twijfel bereikt door het feit dat mogelijke beoordelingsfouten nog kunnen worden gecompenseerd in het geval dat alle zeven criteria gezamenlijk worden gebruikt.

Over de interne consistentie is opgemerkt dat die voor beoordelaar 2 veel lager is dan die voor beoordelaar 1. Hierboven is al een verklaring gegeven voor het feit dat criterium 7 slecht correleert met de scores op het gehele instrument maar ook voor de andere criteria geldt dat de resultaten op de item-totaal correlaties voor beoordelaar 2 minder goed zijn dan die van beoordelaar 1. Nadere analyse van de scores zelf maakte duidelijk dat beoordelaar 2 niet alleen minder consistent was in het gebruik van de verschillende criteria maar dat hij bovendien de neiging had om hogere scores toe te kennen dan beoordelaar 1.

5.3 Test bij de Open Universiteit

De onderzoeksvragen die bij de Open Universiteit Nederland werden onderzocht waren vergelijkbaar met die bij de Nederlandse Defensie Academie (interbeoordelaarsbetrouwbaarheid en interne consistentie). Omdat in het geval van de Open Universiteit ook de beoordeling met een concurrerend beoordelingsinstrument beschikbaar was, kon voor elke beoordelaar afzonderlijk echter tevens de correlatie worden vastgesteld tussen de beoordeling met het OU-model en die met de scoringsrubriek. In het kader van het onderhavige onderzoek was dat relevant omdat die correlatie wordt beschouwd als een maatstaf voor de criteriumvaliditeit (Cronbach & Meehl, 1955).

De onderzoeksvragen die bij het onderzoek aan de Open Universiteit werden gebruikt, waren:

- Wat is de interbeoordelaarsbetrouwbaarheid van de scoringsrubriek als twee verschillende beoordelaars de zelfde werkstukken beoordelen?
- Wat is de interne consistentie van de beoordelingen die zijn opgemaakt met de scoringsrubriek?
- In welke mate geven de beoordelingen die worden opgemaakt met de scoringsrubriek een valide beeld van de beoordelingen die worden opgemaakt met het beoordelingsmodel van de opleiding zelf?

5.3.1 Methode

Van de twee beoordelaars bij dit deelonderzoek was slechts één tevens docent in de betreffende cursus. De tweede beoordelaar was de onderzoeker / ontwerper van de scoringsrubriek zelf.

Het onderzoek werd uitgevoerd aan de hand van studiedossiers die waren samengesteld in het kader van een cursus Informatievaardigheden voor sociale wetenschappers. Er waren 21 dossiers beschikbaar die betrekking hadden op een literatuuronderzoek en de daarop volgende opdracht voor het schrijven van een essay met betrekking tot het onderwerp Self Regulated Learning. Over het literatuuronderzoek dienden de studenten te rapporteren in een 'Proceswerkblad', het essay had de vorm van een blog.

Een deel van de dossiers was door de docent van de OU reeds eerder beoordeeld met een beoordelingsmodel dat speciaal voor de cursus was ontwikkeld (bijlage 5.2 aan het einde van dit hoofdstuk). Samenwerking tussen de onderzoeker / ontwerper van de scoringsrubriek en de beoordelaar van de OU-opdrachten creëerde de mogelijkheid om de studentendossiers ook met de scoringsrubriek te beoordelen en de scores met de twee beoordelingsinstrumenten vervolgens met elkaar te vergelijken.

Om deze vergelijking uit te voeren moesten de criteria uit de twee beoordelingsmodellen zelf vooraf nog wel met elkaar vergeleken worden. Die vergelijking leverde op dat twee criteria uit het OU-model niet in het onderzoek naar de criteriumvaliditeit werden meegenomen omdat vooraf duidelijk was dat die geen onderdeel uitmaakten van de scoringsrubriek (schriftelijke taalvaardigheid en kwaliteit van de reflectie). Een overzicht van de vergelijking is opgenomen als bijlage 5.3 aan het einde van dit hoofdstuk.

Tijdens het beoordelingsproces zelf bleek dat criterium 7 van de scoringsrubriek ook niet bruikbaar was omdat de keuze van de secundaire bronnen in dit geval bij de opdracht al meegegeven was:

student hadden de opdracht gekregen om te werken met het bronnenapparaat dat bij Ebscohost beschikbaar is. Deze problematiek met betrekking tot het gebruik van criterium 7 is vergelijkbaar met de kwestie die speelde bij de Nederlandse Defensie Academie. In de casus van de OU is criterium 7 verder buiten de analyse gehouden.

Tijdens een fysieke bijeenkomst beoordeelden de onderzoeker / ontwikkelaar van de scoringsrubriek (in het vervolg te benoemen als beoordelaar 1) en de betrokken docent van de OU (in het vervolg te benoemen als beoordelaar 2) ieder het zelfde dossier met de scoringsrubriek. Vervolgens werden de beoordelingen met elkaar vergeleken met het doel interpretatieverschillen van het beoordelingsmodel op het spoor te komen. Hetzelfde gebeurde met het OU-beoordelingsmodel, zij het dat daar een ander dossier voor gebruikt werd en dat de beoordeling met het eigen model door de OU-docent reeds beschikbaar was. Door de beoordelingen met elkaar te vergelijken, hoopten de beoordelaars te bereiken dat ze met een gemeenschappelijke blik naar de overige uitwerkingen zouden kijken.

De overige 19 dossiers werden verdeeld in twee groepen. Negen van die 19 dossiers waren in het verleden al beoordeeld door beoordelaar 2 met het OU-model. Beoordelaar 1 begon met deze dossiers ook met het OU-model te beoordelen. De overige 10 dossiers werden door hem eerst met de scoringsrubriek beoordeeld. Zeven weken later werden de dossiers nog eens beoordeeld maar toen met het instrument dat in de eerste ronde niet gebruikt was. Deze procedure van randomisering werd toegepast om het ontstaan van toevallige meetfouten, en dan met name fouten die veroorzaakt zouden kunnen worden door het tijdstip van beoordelen, zoveel als mogelijk te verspreiden over de te beoordelen dossiers.

Beoordelaar 2 beoordeelde in de eerste ronde alle dossiers met de scoringsrubriek en 6 weken later de dossiers die zij in een eerder stadium nog niet had beoordeeld met het OU-model.

Precies als in het onderzoek bij de Nederlandse Defensie Academie werden de data uit het onderzoek bij de Open Universiteit verwerkt in kruistabellen en werden voor de scoringsrubriek zowel de percentages voor de absolute overeenstemming als die voor de aanpalende overeenstemming vastgesteld. Anders dan bij de casus van de Nederlandse Defensie Academie was het niet zinvol om een uitspraak te doen over de overeenstemming met betrekking tot de beslissing voldoende / onvoldoende, omdat voor de dossiers van de Open Universiteit gold dat ze alle al eens beoordeeld waren met een 'voldoende'. Wel werd ook voor de beide beoordelaars van het werk van de Open Universiteit de interne consistentie vastgesteld aan de hand van Cronbach's Alpha, 'Cronbach's Alpha indien Item Verwijderd' en de 'Gecorrigeerde Item-Totaal Correlaties'. Om de criteriumvaliditeit van de scoringsrubriek vast te stellen werd voor iedere beoordelaar afzonderlijk bovendien de Pearson correlatie vastgesteld tussen de beoordeling met de scoringsrubriek en die met het OU-model.

Van de 21 beschikbare dossiers zijn er uiteindelijk 19 met de scoringsrubriek beoordeeld. Eén dossier was gebruikt voor de normering vooraf en is in het eigenlijke onderzoek niet meegenomen. Een tweede dossier was om technische redenen door beoordelaar 1 niet leesbaar te maken.

5.3.2 Resultaten

Interbeoordelaarsbetrouwbaarheid

Tabel 5.5 geeft een overzicht van de absolute en van de aanpalende overeenstemming tussen de twee beoordelaars bij het gebruik van de scoringsrubriek, zowel voor de criteria afzonderlijk als voor de eindbeoordelingen.

Tabel 5.5

Absolute en aanpalende overeenstemming tussen beoordelaar 1 en beoordelaar 2 bij de Open Universiteit Nederland

Criterium	N =	Absolute overeenstemming	Aanpalende overeenstemming
1 Oriëntatie	19	6 (32%)	17 (90%)
2 Literatuurlijst	19	3 (16%)	13 (68%)
3 Primaire bronnen	19	10 (53%)	18 (95%)
4 Bronverwijzingen	19	5 (26%)	16 (84%)
5 Informatieverwerking	19	8 (42%)	17 (90%)
6 Zoektermen	19	6 (32%)	17 (90%)
Eindbeoordeling (1-10)	19	7 (37%)	11 (58%)

Een vergelijking van de scores op de eerste zes criteria afzonderlijk geeft een beeld dat vergelijkbaar is met de resultaten die werden gescoord bij de Nederlandse Defensie Academie, zij het dat de absolute overeenstemming in het geval van OU iets lager ligt. Behalve voor criterium 2 geldt voor de afzonderlijke criteria dat de aanpalende overeenstemming op de 6 punts Likertschaal tussen de twee beoordelaars hoger dan 80% is. Criterium 2 (volledigheid van de literatuurlijst en van de vermelde bibliografische gegevens) geeft blijkbaar aanleiding om fouten over het hoofd te zien of om van mening te verschillen over de mate waarin fouten moet worden aangerekend.

Nadere beschouwing van de scores voor criterium 2 geeft aan dat beoordelaar 1 over het algemeen veel lagere scores toekende dan beoordelaar 2. Dat geldt ook voor de eindbeoordelingen waarvoor de aanpalende overeenstemming eveneens veel lager is dan bij de casus van de Nederlandse Defensie Academie. Cronbach's Alpha voor de twee reeksen eindcijfers (0,65, N=19) bevestigt dat de twee beoordelaars in de setting van de Open Universiteit minder overeenstemming bereikten dan de twee beoordelaars van de Nederlandse Defensie Academie (0,78, N=25).

Interne consistentie

Ook voor de beoordelaars in de setting van de Open Universiteit werd vastgesteld in welke mate de beoordelingen op de afzonderlijke criteria consistent waren met hun totale beoordeling. De betrouwbaarheidsanalyse met SPSS leverde voor beoordelaar 1 een Cronbach's Alpha op van 0,746, voor beoordelaar 2 een van 0,857. Voor de twee beoordelaars samen was de gemiddelde Cronbach's Alpha in dit geval 0,802. De tabellen 5.6 en 5.7 geven de Item-Totaal Statistieken voor de afzonderlijke OU-beoordelaars.

Tabel 5.6

Item-Totaal Statistieken voor OU beoordelaar 1

Criterium	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd^a
1 Oriëntatie	,254	,761
2 Literatuurlijst	,614	,658
3 Primaire bronnen	,515	,705
4 Bronverwijzingen	,821	,623
5 Informatieverwerking	,672	,654
6 Zoektermen	,136	,783

^a Cronbach's Alpha: 0,746**Tabel 5.7**

Item-Totaal Statistieken voor OU beoordelaar 2

Criterium	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd^a
1 Oriëntatie	,685	,827
2 Literatuurlijst	,579	,849
3 Primaire bronnen	,583	,846
4 Bronverwijzingen	,568	,846
5 Informatieverwerking	,864	,798
6 Zoektermen	,671	,828

^a Cronbach's Alpha: 0,857

Ook in de setting die gebruikt werd bij de Open Universiteit Nederland bleek dat één van de beoordelaars in staat was om de scoringsrubriek heel consistent te gebruiken. Bij beoordelaar 2 komt de Gecorrigeerde Item-Totaal Correlatie voor geen van de criteria onder de grens van 0,500. Bovendien geldt voor geen enkele van de criteria dat Cronbach's Alpha uitstijgt boven de vastgestelde waarde van 0.857 als het betreffende criterium zou worden verwijderd. Voor beoordelaar 1 blijkt vooral het gebruik van criterium 6 ('zoektermen') problematisch. De Gecorrigeerde Item-Totaal Correlatie blijft daar onder de minimaal gewenste waarde van 0,200.

Correlatie tussen beoordelingen met scoringsrubriek en die met het OU-model

Zoals al aangegeven in de inleiding tot deze paragraaf 5.3 bood de OU-casus tevens de mogelijkheid om de criteriumvaliditeit van de scoringsrubriek vast te stellen door de beoordelingen van de afzonderlijke beoordelaars te vergelijken met de beoordelingen die zij hadden opgemaakt met het OU-model. Het aantal beschikbare beoordelingen voor deze analyse (N) was 18 aangezien er zowel voor de scoringsrubriek als voor het OU-model een dossier was gebruikt voor de normering én omdat één dossier om technische redenen niet leesbaar was te maken. Voor beoordelaar 1 werd een Pearson correlatie vastgesteld van 0.68 ($p < 0.01$) en voor

beoordelaar 2 van 0.55 ($p < 0.05$).¹² Uit deze resultaten kan geconcludeerd worden dat er een significante positieve correlatie bestond tussen de beoordelingen met de scoringsrubriek en die met het OU-model. Een kanttekening die bij deze kengetallen mag worden geplaatst is dat de beoordelingen betrekkelijk weinig spreiding kenden, onder andere omdat alle dossiers in een eerder stadium al beoordeeld waren als ‘voldoende’.

5.3.3 Conclusies

Hoewel de interbeoordelaarsbetrouwbaarheid bij de Open Universiteit voor de afzonderlijke criteria heel behoorlijk was (met uitzondering van die voor criterium 2, het bronnenoverzicht/de literatuurlijst), was deze voor de eindbeoordelingen lager dan gewenst. Dat blijkt ook uit Cronbach's Alpha voor de twee reeksen eindcijfers (0,65, $N=19$).

Net als voor criterium 2 geldt voor de eindbeoordelingen dat beoordelaar 1 met de scoringsrubriek steeds relatief lage beoordelingen heeft toegekend. Concluderend kan worden gesteld dat, vooral ook gezien de onbekendheid van beoordelaar 1 met de betreffende studieopdrachten, het wenselijk geweest was als de beoordelaars vooraf meer tijd hadden gestoken in de normering van het beoordelingsproces. Daarnaast kan een rol gespeeld hebben dat beoordelaar 1 überhaupt minder bekend was met het vakgebied waar de studieopdrachten betrekking op hadden. Dit kan gevolgen hebben gehad voor de interpretatie van een of meer onderdelen van de uitgewerkte dossiers.

Dat beoordelaar 1 meer moeite had om de scoringsrubriek goed op de betreffende dossiers toe te passen blijkt ook uit Item-Totaal Statistieken (tabel 5.6). Vooral criterium 6 (kwaliteit van de gebruikte zoektermen) lijkt beoordelaar 1 moeilijk te kunnen toepassen. Onbekendheid met het vakgebied waarop de studieopdracht betrekking had, kan daarbij een rol gespeeld hebben. Beoordelaar 2 blijkt juist wel heel consistent in de beoordelingen te zijn (tabel 5.7).

De correlaties tussen de beoordelingen met de scoringsrubriek en die met het OU-model, zijn voor beide beoordelaars significant. Die voor beoordelaar 1 ($r=0.68$, $p < 0.01$) geeft aanleiding om te veronderstellen dat de moeite die deze beoordelaar heeft met het opmaken van consistente beoordelingen eerder wordt veroorzaakt door zijn onbekendheid met het vakgebied dan met de hantering van de beoordelingsinstrumenten zelf.

5.4 Algehele conclusies en discussie

Een analytische scoringsrubriek is een geschikt instrument voor de beoordeling van het daadwerkelijke informatiegedrag van studenten bij het oplossen van informatieproblemen in het kader van hun studietaken. Hoofdstuk 4 beschreef de constructie van zo'n scoringsrubriek. In dit hoofdstuk 5 is aandacht besteed aan drie verschillende onderzoeksvragen: die naar de praktische bruikbaarheid van de Scoringsrubriek voor Informatievaardigheden, die naar de interbeoordelaarsbetrouwbaarheid en die naar de validiteit. In deze paragraaf 5.4 zal in aparte subparagrafen worden stilgestaan bij de antwoorden op ieder van deze deelvragen.

¹² Als criterium 7 wordt meegewogen zijn de correlaties iets hoger: 0,71 en 0,56

5.4.1 Praktische bruikbaarheid van de scoringsrubriek

Zowel docenten van de opleiding Commerciële Economie van Saxion als de docenten Informatiekunde van de Nederlandse Defensie Academie bevestigden dat het scoringsschema hen behulpzaam was geweest om bij de beoordeling van studentenproducten goed gefocust te blijven op de aspecten die relevant zijn in het kader van informatievaardigheid. Er zijn geen aanwijzingen gevonden dat het gebruik van de scoringsrubriek buitensporig veel tijd vergt van de beoordelaars.

Wel gaf één van de acht respondenten van Saxion bij de evaluatie als opmerking mee dat hij of zij het plezierig had gevonden als ook het gedrag op de tussenliggende niveaus beschreven was geweest. Zoals aangegeven in paragraaf 4.4.3 werden gedragsbeschrijvingen in de definitieve versie van het prototype beperkt tot het hoogste en het laagste niveau en werd er een Likertschaal toegevoegd om de beoordeling zelf uit te drukken. De reden daarvoor was dat 'onvoldoende gedrag' zich op veel verschillende manieren kan voordoen. De oorzaak daarvan zou kunnen zijn dat de gekozen beoordelingscriteria nog altijd betrekkelijk generiek zijn en dat er in de scoringsrubriek voor informatievaardigheden verschillende operationalisaties van onvoldoende gedrag, bijvoorbeeld in de vorm van de check boxes, worden gebruikt. Indien het gedrag ook op tussenliggende niveaus beschreven zou worden, zou ieder van die operationalisaties moeten worden uitgewerkt tot een apart beoordelingscriterium. Uitsplitsing van de zeven beoordelingscriteria in meer specifieke criteria zou de scoringsrubriek echter veel uitgebreider maken. Beoordelingscriterium 3 bijvoorbeeld ('Kwaliteit van de gebruikte primaire informatiebronnen') zou dan moeten worden uitgewerkt tot minstens vier afzonderlijke criteria voor achtereenvolgens herkomst, authenticiteit, actualiteit en relevantie van de informatiebronnen. Zo'n uitgebreid scoringsschema zou docenten kunnen afschrikken om het te gaan gebruiken.

Het onderzoek onder docenten die de scoringsrubriek hebben gebruikt gaf inderdaad aan dat juist de compactheid van het schema bijdraagt aan de aantrekkelijkheid ervan. Het moet echter worden toegegeven dat de praktische bruikbaarheid van de scoringsrubriek voor informatievaardigheden op dit punt enigszins ten koste gaat van de meer theoretisch volledigheid als beoordelingsmodel.

In de hoofdstukken 7 en 8 wordt uitgebreider stilgestaan bij de praktische bruikbaarheid van de scoringsrubriek voor studenten en docenten.

5.4.2 Interbeoordelaarsbetrouwbaarheid

Tijdens twee afzonderlijke onderzoeken naar het gebruik van het prototype bleek de betrouwbaarheid van de scoringsrubriek afhankelijk te zijn van de gehanteerde opvatting van het begrip interbeoordelaarsbetrouwbaarheid.

De twee beoordelaars van de Nederlandse Defensie Academie beoordeelden ieder de zelfde set van 27 uitwerkingen maar zij kwamen bij hun beoordelingen op de 6 punts Likertschaal niet tot een hoge mate van absolute overeenstemming. De mate van 'aanpalende overeenstemming' was echter zowel voor de meeste afzonderlijke criteria als voor de eindbeoordeling aanzienlijk, vooral als in aanmerking wordt genomen dat deze werd bereikt zonder dat de beoordelaars vooraf getraind waren in het gebruik van de scoringsrubriek.

Overeenstemming over de beslissing voldoende / onvoldoende was acceptabel voor de eindbeoordelingen maar bleek voor drie van de afzonderlijke criteria problematisch. Daaruit is geconcludeerd dat het bij summatieve beoordelingen de voorkeur verdient om de complete scoringsrubriek te gebruiken omdat de consistentie van de beoordelingen hoger was voor de eindbeoordelingen dan voor die op de afzonderlijke criteria. Uiteraard kan het in praktijk voorkomen dat een van de criteria uit de scoringsrubriek niet wordt gebruikt omdat het corresponderende element niet voorkomt in de te beoordelen opdracht maar tegelijkertijd dient men zich te realiseren dat men niet meer het complete ‘construct’ meet als men te weinig criteria overhoudt.

Het onderzoek waarbij gebruik werd gemaakt van studentendossiers die waren aangelegd in het kader van de OU-cursus ‘Informatievaardigheden voor sociale wetenschappers’, gaf een vergelijkbaar beeld van de interbeoordelaarsbetrouwbaarheid. De twee beoordelaars die in dat geval waren betrokken bij de test bereikten eveneens een hoge aanpalende overeenstemming over de beoordelingen op de afzonderlijke criteria. Met betrekking tot de eindbeoordelingen verschilden zij echter meer van mening. Een verklaring die daarvoor werd gevonden is dat een van de beoordelaars niet betrokken is geweest bij de betreffende cursus en überhaupt geen deskundige was op het vakgebied waar de opdracht en de dossiers betrekking op hadden. Meer in het algemeen kan daaruit worden geconcludeerd dat de scoringsrubriek niet helemaal los kan worden gezien van het kennisdomein waarin ze wordt toegepast.

Uit de onderzoeken bij de Nederlandse Defensie Academie en de Open Universiteit Nederland blijkt met andere woorden dat de scoringsrubriek als beoordelingsinstrument betrouwbaar kan worden gebruikt maar dat een van de voorwaarden wel is dat de beoordelaars ook thuis zijn in het vakgebied waar de te beoordelen opdracht betrekking op heeft.

5.4.3 Validiteit

Tijdens de test bij de Nederlandse Defensie Academie bleek dat alle zeven beoordelingscriteria van de scoringsrubriek een bijdrage leverden aan de betrouwbaarheid van de beoordeling in haar geheel. Dat kan worden beschouwd als een aanwijzing voor het feit dat ze ook werkelijk de inhoud beoordelen die ermee beoogd wordt en dus ‘valide’ zijn.

Het feit dat een van de beoordelaars voor criterium 7 (‘gebruik van secundaire bronnen’) tot scores kwam die de interne consistentie geweld aan deden, bleek te worden veroorzaakt door een misinterpretatie die door onderling overleg tussen de beoordelaars makkelijk voorkomen had kunnen worden. Ook voor de casus van de Open Universiteit bleek criterium 7 niet goed bruikbaar omdat het gebruik van Ebscohost in dat geval al onderdeel van de instructie bij de opdracht was.

Zonder te zeggen dat criterium 7 in de scoringsrubriek gemist zou kunnen worden, is deze ervaring met twee deelonderzoeken wel aanleiding om er bij gebruik van dit criterium alert op te zijn dat het niet voor iedere opdracht van toepassing zal zijn. Zoals ook al opgemerkt in paragraaf 5.4.2 dient daar echter wel de kanttekening bij te worden geplaatst dat de scoringsrubriek minder betrouwbaar wordt als er te weinig criteria overblijven.

Dat een van de beoordelaars van de OU-dossiers ook op criterium 6 behoorlijk afwijkend scoorde, kan verklaard worden door onbekendheid van de betreffende persoon met het vakgebied. Dat toont opnieuw aan dat de scoringsrubriek niet inhoudsneutraal is en is een argument om, zoals

eerder ook al bepleit, het gebruik van de scoringsrubriek goed te integreren in het onderwijs op het betreffende vakgebied.

Voor de dossiers die waren opgemaakt in het kader van de OU-cursus, waren behalve de beoordelingen met de scoringsrubriek tevens beoordelingen beschikbaar die waren opgemaakt met het oorspronkelijk OU-beoordelingsmodel. Een vergelijking van de beoordelingen met de twee verschillende modellen gaf aan dat er voor ieder van de beoordelaars sprake was van een significante statistische relatie tussen de beoordelingen met de scoringsrubriek en die met het OU-model, zij het dat die niet hoog was. Daarbij dient wel opgemerkt te worden dat dit onderzoek naar de criteriumvaliditeit slechts betrekking had op 18 werkstukken.

Slotopmerkingen

Voor al in het geval van Saxion was het aantal participanten laag. Dit betekent dat de resultaten uit dat gedeelte van het onderzoek niet kunnen worden gegeneraliseerd naar willekeurige andere onderwijssituaties. Het aantal studentenproducten voor de casussen van de Nederlandse Defensie Academie en de Open Universiteit waren juist groot genoeg om op een verantwoorde wijze statistische waarden vast te stellen voor de overeenstemming tussen de twee beoordelaars, de consistentie van de beoordelingen en de correlaties met beoordelingen die waren opgemaakt met een ander, speciaal daarvoor ontwikkeld, beoordelingsmodel.

Beperkingen bij het onderzoek aan de Open Universiteit waren de onbekendheid van één van de beoordelaars met het vakgebied en het gegeven dat slechts werkstukken werden gebruikt die eerder al waren beoordeeld als 'voldoende'.

De evaluatie van de scoringsrubriek was niet alleen beperkt qua aantal respondenten, ook het toepassingsgebied was steeds strak afgebakend. Het onderzoek maakte weliswaar gebruik van drie verschillende contexten maar deze hadden alle drie wel betrekking op een in de tijd afgebakende studieopdracht. Toepassing van de rubriek in een ontwikkelingsgericht portfolio waardoor de groei van informatievaardigheden bij een student kan worden gevolgd (Diller & Phelps, 2008, pp. 78-79), ligt zeer voor de hand. In dat geval zullen docenten echter wel goede afspraken met elkaar moeten maken over de vraag hoe de scoringsrubriek door het curriculum heen wordt gehanteerd.

Tijdens de bespreking van de interbeoordelaarsbetrouwbaarheid en de validiteit van de scoringsrubriek is geconstateerd dat het voorkomt dat niet alle zeven beoordelingscriteria van toepassing zijn op de betreffende opdracht. Daarover is opgemerkt dat de scoringsrubriek minder betrouwbaar zal worden naarmate er minder criteria worden gebruikt. Tevens geldt in dat geval dat het beoordelingsinstrument minder valide wordt omdat men dan niet het volledige construct meer meet. Andersom geldt echter dat de scoringsrubriek voor informatievaardigheden ook kan helpen om goede opdrachten te formuleren waarbij het construct informatievaardigheden zo volledig mogelijk gemeten wordt. Op die manier is de scoringsrubriek dus tevens een hulpmiddel bij het ontwikkelen van onderwijs in plaats van alleen een toetsinstrument.

De tests en evaluaties van het prototype van de scoringsrubriek bleven in deze fase van het onderzoek beperkt tot min of meer cijfermatige studies. Ze concentreerden zich op de toepasbaarheid in de onderwijspraktijk, efficiëntie in gebruik, de interne consistentie, de interbeoordelaarsbetrouwbaarheid en de correlatie met een reeds bestaand beoordelingsinstrument. Al deze onderzoeken hebben niet geleid tot inhoudelijke aanpassingen van het prototype uit bijlage 4.1. Wel is er later een wijziging in vormgeving aangebracht ten behoeve van een publicatie in het vaktijdschrift *Informatie Professional* (Van Helvoort, 2009). Door het gebruik van kleuren en professionele typografie kon het schema in die versie worden teruggebracht tot één pagina A4. Om deze heel praktische reden is die versie het vaakst geciteerd en gebruikt.

Bijlage 5.1: *Survey die werd gebruikt bij de opleiding Commerciële Economie van Saxion*

1. De beoordelingscriteria van de scoringsrubriek zijn volgens mij *alle 5* noodzakelijk voor het beoordelen van de competentie 'informatievaardigheden' aan de hand van een concreet studentenproduct.

☐ eens Toelichting (indien van toepassing):
☐ oneens
2. De 5 beoordelingscriteria van de scoringsrubriek zijn volgens mij voldoende voor het beoordelen van de competentie 'informatievaardigheden' aan de hand van een concreet studentenproduct.

☐ eens Toelichting (indien van toepassing):
☐ oneens
3. De formulering van het 'professionele' en het 'inadequate gedrag' voor de diverse beoordelingscriteria was voor mij *in alle gevallen* duidelijk.

☐ eens Toelichting (indien van toepassing):
☐ oneens
4. De scoringsrubriek zorgde ervoor dat ik goed wist waar ik op moest letten bij het beoordelen van de informatievaardigheden van de betreffende student.

☐ eens Toelichting (indien van toepassing):
☐ oneens
5. Ik ben van plan om de scoringsrubriek in de toekomst te gaan gebruiken voor het summatief beoordelen van andere studentenproducten (bijvoorbeeld afstudeerverslagen en research papers).

☐ eens Toelichting (indien van toepassing):
☐ oneens
6. Ik denk dat ik de scoringsrubriek in de toekomst goed kan gebruiken *in andere situaties* dan een summatieve beoordeling, bijvoorbeeld voor het geven van instructie of het geven van feedback.

☐ eens Toelichting (indien van toepassing):
☐ oneens

Tenslotte volgen hieronder nog 5 vragen naar meer feitelijke gegevens

7. Hoeveel studentenproducten heb je met de scoringsrubriek beoordeeld voordat je deze vragenlijst hebt ingevuld?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ meer dan 15

8. Hoeveel tijd had je gemiddeld per studentproduct nodig om de scoringsrubriek in te vullen?

- ☐ <5 minuten
- ☐ 5-10 minuten
- ☐ 10-15 minuten
- ☐ 15 minuten of meer

9. Wat is je geslacht?

- ☐ man
- ☐ vrouw

10. Bij welke hogeschool of universiteit heb je de scoringsrubriek gebruikt?

.....

11. Bij welke opleiding, afdeling of faculteit heb je de scoringsrubriek gebruikt?

.....

Bijlage 5.2: *OU-model voor het beoordelen van het literatuuronderzoek en het essay met betrekking tot Self Regulated Learning*

Proceswerkblad

1. De onderzoeksvraag is helder geformuleerd
2. De zoekstrategie die is gekozen volstaat
3. De gekozen zoektermen zijn goed gekozen en de systematiek is helder
4. De juiste bronnen zijn aan de hand van criteria geselecteerd
5. De bronnen zijn in een tabel goed beschreven
6. De bronnen zijn op een adequate wijze verwerkt
7. De reflectie op het proces bevat een omschrijving van de goede punten en de leermomenten

Verslag / Blog

1. De probleemstelling en vraagstelling zijn helder beschreven
2. In de tekst wordt een oplossingsrichting gegeven van het probleem gebruik makend van de gelezen literatuur
3. In het verslag worden de APA-richtlijnen gehanteerd
4. De tekst is goed gestructureerd en taalkundig correct geschreven.

Elk criterium wordt beoordeeld op een schaal van 1-5 punten. Bij een maximum van 55 punten ligt, bij een cesuur van 55%, het minimum aantal punten op 31 punten.

Bijlage 5.3: *Vergelijking van de beoordelingscriteria uit de Scoringsrubriek voor Informatievaardigheden met de beoordelingscriteria die worden gebruikt door de Open Universiteit in de cursus 'Informatievaardigheden voor sociale wetenschappers'*

<i>Scoringsrubriek voor Informatievaardigheden (studentenproduct)</i>	<i>Beoordelingsmodel OU</i>
<i>Oriëntatie op het onderwerp</i>	<i>De onderzoeksvraag is helder geformuleerd (Proceswerkblad)</i> <i>De probleemstelling en vraagstelling zijn helder beschreven (Blog)</i>
<i>Bronnenoverzicht</i>	<i>In het verslag worden de APA-richtlijnen gehanteerd (Blog)</i>
<i>Kwaliteit van de gebruikte primaire informatiebronnen</i>	<i>De juiste bronnen zijn aan de hand van criteria geselecteerd (Proceswerkblad)</i>
Expliciete aandacht voor een zelfstandig bronnensysteem (een literatuurmatrix) ontbreekt.	<i>De bronnen zijn in een tabel goed beschreven (Proceswerkblad)</i>
Expliciete aandacht voor een zelfstandig bronnensysteem (een literatuurmatrix) ontbreekt. Bij <i>Verwerking van relevante informatie tot nieuwe kennis en/of inzichten</i> is er wel aandacht voor de inhoudelijke analyse van gevonden bronnen.	<i>De bronnen zijn op een adequate wijze verwerkt (Proceswerkblad)</i>
Een criterium met betrekking tot reflectie ontbreekt.	<i>De reflectie op het proces bevat een omschrijving van de goede punten en de leermomenten (Proceswerkblad)</i>
<i>Bronverwijzingen in de tekst</i>	<i>In het verslag worden de APA-richtlijnen gehanteerd (Blog)</i>
<i>Verwerking van relevante informatie tot nieuwe kennis en/of inzichten</i>	<i>In de tekst wordt een oplossingsrichting gegeven van het probleem gebruik makend van de gelezen literatuur (Blog).</i> Aandacht voor de inhoudelijke analyse is er vooral bij het criterium <i>De bronnen zijn op een adequate wijze verwerkt (Proceswerkblad)</i> .
<i>Zoektermen</i>	<i>De gekozen zoektermen zijn goed gekozen en de systematiek is helder (Proceswerkblad).</i>
<i>Gebruik van diverse secundaire informatiebronnen.</i> Accent ligt op het gegeven dat de student diverse secundaire informatiebronnen heeft gebruikt en zich niet heeft beperkt tot de makkelijk toegankelijke bronnen.	<i>De zoekstrategie die is gekozen volstaat (Proceswerkblad).</i> Uit de uitwerking blijkt dat het gaat om een combinatie van 'methode' (citaatzoeken, blokkendoosmethode, parelgroeiemethode) en te gebruiken informatiebron (zoekmachine, database etc.)
Een criterium met betrekking tot schriftelijke taalvaardigheid ontbreekt	<i>De tekst is goed gestructureerd en taalkundig correct geschreven (Blog).</i>

Noot: De tekst die cursief en vet staat komt letterlijk uit het betreffende beoordelingsschema; de tekst in normale typografie is toelichting.

6

Testen van betrouwbaarheid en validiteit bij de opleiding Mediastudies van de Universiteit van Amsterdam

Abstract: Doel van het onderzoek dat in dit hoofdstuk wordt beschreven was het vaststellen van de betrouwbaarheid en validiteit van de scoringsrubriek voor informatievaardigheden.

Daarvoor zijn 80 werkstukken van de opleiding Mediastudies van de Universiteit van Amsterdam door twee beoordelaars beoordeeld met twee verschillende beoordelingsinstrumenten: de Scoringsrubriek voor Informatievaardigheden en het beoordelingsinstrument dat de opleiding al gebruikte. Op basis van de beoordelingen is vastgesteld dat beide beoordelingsinstrumenten 'goed' scoren op het gebied van interbeoordelaarsbetrouwbaarheid en interne consistentie. Omdat de eindcijfers die met de twee beoordelingsmodellen werden gegeven sterk met elkaar correleerden, is bovendien geconcludeerd dat het zeer aannemelijk is dat de informatievaardigheden van studenten met de scoringsrubriek op een valide manier worden beoordeeld.

Inleiding

In hoofdstuk 5 is het prototype van de scoringsrubriek in drie verschillende contexten getest en geëvalueerd. In dit hoofdstuk wordt aandacht besteed aan een vierde onderzoek waarbij gebruik gemaakt werd van 80 uitwerkingen van een studieopdracht, een aantal dat vele malen groter was dan de aantallen die bij de eerdere onderzoeken gebruikt werden. Bovendien was er bij dit vierde onderzoek tevens sprake van een ander type opleiding, te weten de bachelor opleiding Mediastudies van de Faculteit der Geesteswetenschappen aan de Universiteit van Amsterdam. Dat zorgde er voor dat de eigenschappen van het prototype van de scoringsrubriek uiteindelijk op vier verschillende vakgebieden en zowel in het hoger beroepsonderwijs als in het wetenschappelijk onderwijs zijn onderzocht.

De eigenschappen waar onderzoek naar werd gedaan waren opnieuw de betrouwbaarheid, de validiteit en de praktische bruikbaarheid van het prototype. De betrouwbaarheid van de scoringsrubriek werd, net zoals in hoofdstuk 5 bij de Nederlandse Defensie Academie en de Open Universiteit, opgevat als interbeoordelaarsbetrouwbaarheid. Omdat er voor het onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam 80 uitwerkingen beschikbaar waren, kon er echter, behalve van het percentage overeenstemming, ook gebruik gemaakt worden van de Intraclass Correlation Coefficient om de overeenstemming tussen twee beoordelaars van het zelfde werk vast te stellen.

Met betrekking tot de validiteit van het prototype werd onderscheid gemaakt tussen constructvaliditeit en criteriumvaliditeit. Moskal en Leydens (2000) geven aan dat constructvaliditeit onder andere onderzocht kan worden door na te gaan of alle criteria van een beoordelingsinstrument wel relevant zijn voor het meten van het betreffende onderliggende 'construct'. In dit onderzoek werd dit vastgesteld door met SPSS zowel de onderlinge consistentie van de verschillende beoordelingscriteria vast te stellen als een factoranalyse uit te voeren.

Onder criteriumvaliditeit wordt verstaan dat de beoordelingen die worden opgemaakt met het ene instrument positief correleren met de beoordelingen die worden opgemaakt met een ander instrument. Als er sprake is van min of meer gelijktijdige beoordeling met de twee verschillende instrumenten, wordt ook wel gesproken over concurrentvaliditeit. In het onderzoek bij de opleiding Mediastudies van de UvA kon ook deze criteriumvaliditeit worden onderzocht omdat er naast de scoringsrubriek tevens een beoordelingsmodel beschikbaar was dat de opleiding zelf al gebruikte. De derde onderzochte eigenschap van het prototype, de praktische bruikbaarheid, is minder systematisch onderzocht. De beoordelingen hebben wel plaats gevonden in vier verschillende 'rondes' en na afloop van elke ronde is met de twee beoordelaars steeds een korte nabespreking georganiseerd. Tijdens deze gesprekken hebben de beoordelaars uitdrukkelijk opmerkingen gemaakt over de gebruiksvriendelijkheid van de twee verschillende beoordelingsinstrumenten. Uitgebreidere onderzoeken naar de praktische bruikbaarheid van de scoringsrubriek worden beschreven in de hoofdstukken 7 en 8.

De onderzoeksvragen die bij het onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam werden gehanteerd luiden:

- a. Wat is de interbeoordelaarsbetrouwbaarheid van de scoringsrubriek als de 80 werkstukken door twee verschillende beoordelaars worden beoordeeld?
- b. Wat is de interne consistentie van de beoordelingen die zijn opgemaakt met de scoringsrubriek?
- c. Wat is de homogeniteit van de beoordelingen die zijn opgemaakt met de scoringsrubriek?
- d. In welke mate geven de beoordelingen die worden opgemaakt met de scoringsrubriek een valide beeld van de beoordelingen die worden opgemaakt met het beoordelingsmodel van de opleiding zelf?

Deze vragen komen overeen met die uit de onderzoeken bij de Nederlandse Defensie Academie en de Open Universiteit die in hoofdstuk 5 zijn beschreven, met uitzondering van de vraag naar de homogeniteit van de beoordelingen met de scoringsrubriek. Deze laatste werd toegevoegd omdat het aantal van 80 werkstukken groot genoeg was om een factoranalyse uit te voeren die aangeeft in hoeverre er binnen de scoringsrubriek sprake is van één enkele dimensie.

6.1 Inhoud en vorm van de gebruikte studieopdracht

De studieopdracht die werd gebruikt was onderdeel van het eerstejaars vak ‘Mediageschiedenis : werkgroep’. De opdracht werd op de Blackboard Course van de opleiding als volgt omschreven: “De basiskennis van de mediageschiedenis die in het vak Inleiding Mediacultuur is verworven, wordt in het vak Mediageschiedenis: werkgroep gebruikt voor historisch georiënteerd mediaonderzoek. Studenten maken kennis met de basisbeginselen van wetenschappelijk onderzoek, zoals het ontwikkelen van een onderzoeksvraag, het maken van een onderzoeksvoorstel, het selecteren en beschrijven van bronnen etc. Deze vaardigheden worden geoefend aan de hand van een thema in historisch onderzoek naar films, televisieprogramma’s of digitale media. Studenten werken aan een concrete casus binnen een thema. De casus wordt individueel maar in overleg met de docent gekozen en afgebakend.”

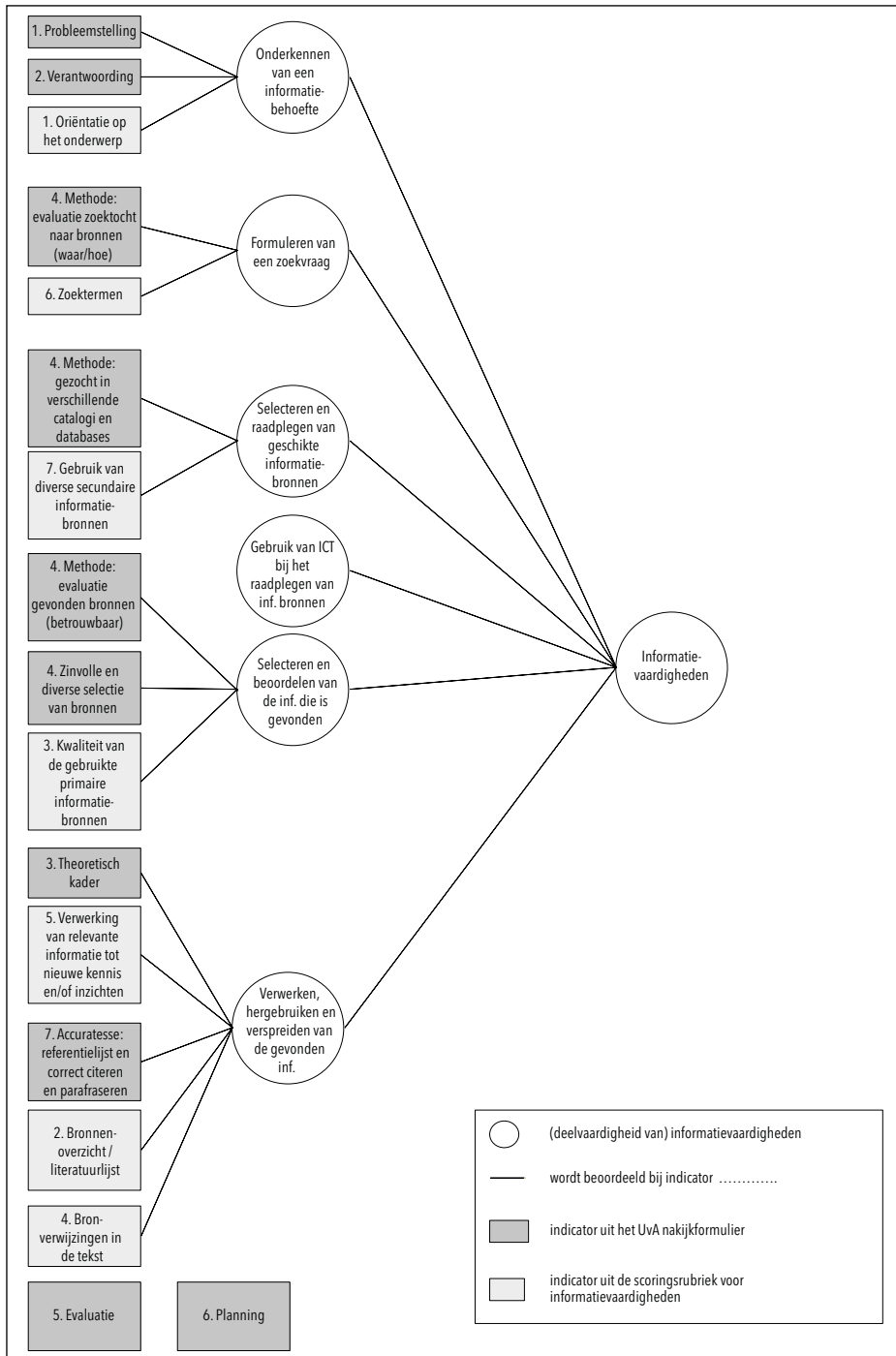
Het vak ‘Mediageschiedenis: werkgroep’ kent drie toetsen: de onderzoeksvraag (I), het onderzoeksvoorstel (II) en het onderzoeksverslag (III). In het empirisch onderzoek naar de validiteit en de betrouwbaarheid van de scoringsrubriek is gebruik gemaakt van toets II, het onderzoeksvoorstel. De aanwijzing in de studieopdracht gaf aan dat de uitwerkingen uit circa 2500 woorden diende te bestaan (exclusief referentielijst en notenapparaat). Deze omvang maakte het praktisch beter uitvoerbaar dan bij de grotere opdracht III (het ‘onderzoeksverslag’) om in een beperkt tijdsbestek een substantieel aantal beoordelingen te laten opmaken. Bovendien bleek bij een vergelijking van de scoringsrubriek met het beoordelingsmodel van de opleiding zelf dat de toets ‘onderzoeksvoorstel’ tevens bijna alle onderdelen van het construct informativaardigheden omvatte. Op deze vergelijking wordt in de volgende paragraaf ingegaan. Een vergelijkbare analyse is overigens gedaan voor het onderzoeksverslag (toets III). Uit die vergelijking bleek dat het beoordelingsmodel van het onderzoeksverslag veel minder aspecten van het construct informativaardigheden dekte.

6.2 Vergelijking van de scoringsrubriek en het UvA-beoordelingsformulier

De opleiding Mediastudies gebruikt voor het beoordelen van de onderzoeksvoorstellen een eigen beoordelingsformulier dat door hen zelf “Nakijkformulier Toets II” wordt genoemd. Het volledige formulier is hier opgenomen als bijlage 6.1. Dit beoordelingsformulier dat in het vervolg zal worden aangeduid als “het UvA-nakijkformulier”, heeft (anders dan de scoringsrubriek) niet de uitdrukkelijke intentie om informatievaardigheden te beoordelen. Om het te kunnen gebruiken als onderzoeksmateriaal in het kader van de laatste onderzoeksvraag (de vraag in welke mate de scoringsrubriek correleert met een concurrerend beoordelingsinstrument) diende de inhoud van het formulier daarom eerst vergeleken te worden met de inhoud van het construct informatievaardigheden. Daarbij viel op dat de verschillende hoofd- en subcriteria in het UvA-nakijkformulier een wat onevenwichtige indruk maken. Met name bij het onderdeel Methode waarvoor (zonder nadere specificatie) 30 punten zijn te behalen, lijken de subcriteria van verschillende aard. De ene keer gaat het om een evaluatie van het zoekproces, een andere keer om het evalueren van de gevonden bronnen zelf en een derde keer om een indeling in inhoudelijke clusters.

Dat komt ook tot uitdrukking in figuur 6.1. In deze afbeelding worden de deelvaardigheden die zijn onderscheiden in hoofdstuk 1 weergegeven als latente variabelen van het samengestelde construct informatievaardigheden. Deze latente variabelen worden vervolgens gerelateerd aan de indicatoren uit de scoringsrubriek en aan die uit het UvA-nakijkformulier. In werkelijkheid worden de indicatoren of beoordelingscriteria gebruikt om het construct daadwerkelijk te meten. Benadrukt wordt dat het daarbij altijd zal gaan om meer of minder imperfecte metingen. Het onderzoek naar de constructvaliditeit (onderzoeksvragen b en c) is bedoeld om vast te stellen hoe goed die metingen uiteindelijk blijken te zijn.

Een vergelijking van de twee groepen indicatoren maakt ook duidelijk in welke mate de scoringsrubriek en het UvA-nakijkformulier met elkaar én met de inhoud van het begrip informatievaardigheden overeenkomen.



Figuur 6.1

Vergelijking van de deelvaardigheden behorende bij het construct informatievaardigheden met de indicatoren van het UvA-nakijkformulier en die van de Scoringsrubriek voor Informatievaardigheden

Uit de analyse in figuur 6.1 blijkt dat

- de deelvaardigheid ‘gebruik van ict bij het raadplegen van informatiebronnen’ in geen van beide beoordelingsmodellen als criterium wordt gebruikt en
- dat het UvA-nakijkformulier twee beoordelingscriteria kent die geen deel uitmaken van het theoretische construct informatievaardigheid dat in hoofdstuk 1 is onderscheiden.

Dat het gebruik van ict niet is opgenomen als beoordelingscriterium in de scoringsrubriek, komt voort uit het eerder gekozen uitgangspunt om bij de ontwikkeling van de scoringsrubriek de nadruk te leggen op het gebruik van informatie op een conceptueel niveau en niet op ict-vaardigheden sec. Ook de ontwikkelaars van het nakijkformulier van de UvA hebben blijkbaar geen aanleiding gezien om ICT-gebruik als apart beoordelingscriterium op te nemen.

De twee beoordelingscriteria van het UvA-nakijkformulier die niet in de scoringsrubriek zijn gebruikt - ‘planning’ en ‘evaluatie’ - zijn nauw gerelateerd aan de manier waarop het onderwijs in dit onderdeel van de opleiding Mediastudies is ingericht. In de invulling van het construct informatievaardigheden die in hoofdstuk 1 werd gehanteerd en in de scoringsrubriek ontbreken ze als deelvaardigheid en beoordelingscriterium omdat die modellen immers zijn bedoeld om meer generiek te worden toegepast.

Het volledige nakijkformulier van de UvA waar de analyse uit figuur 6.1 op is gebaseerd, is opgenomen als bijlage 6.1. In het onderzoek dat in dit hoofdstuk wordt beschreven is van het UvA-nakijkformulier alleen gebruik gemaakt van de criteria die een corresponderend criterium hadden in de scoringsrubriek. Bijlage 6.2 geeft het UvA-nakijkformulier zoals het uiteindelijk in het onderzoek is gebruikt. De criteria die niet zijn gebruikt, zijn in die bijlage grijs afgedrukt.

Tenslotte bleek al vroeg tijdens het onderzoek dat het beoordelingscriterium ‘zoektermen’ uit de scoringsrubriek niet goed gebruikt kon worden omdat in de opgave niet expliciet om vermelding van de gebruikte zoektermen werd gevraagd. Sommige studenten gaven uit zich zelf wel een overzicht van de gebruikte zoekwoorden, anderen beperkten zich in de weergave van hun ‘methode’ echter tot bijvoorbeeld een paar sleutelpublicaties en het gebruik van de ‘sneeuwbalmethode’. De onduidelijkheid hierover leidde tot het besluit om criterium 6 van de scoringsrubriek niet bij het onderzoek te betrekken.

Weging van de verschillende indicatoren

Behalve dat de indicatoren uit het nakijkmodel van de UvA inhoudelijk dienden te corresponderen met die van de scoringsrubriek, was het ook van belang dat de weging van de afzonderlijke indicatoren min of meer vergelijkbaar was. Het gewicht van de afzonderlijke indicatoren van de twee beoordelingsinstrumenten, wordt uitgedrukt in tabel 6.1. Voor de groepering van de indicatoren is uitgegaan van de volgorde die wordt gehanteerd in het nakijkformulier van de UvA. Drie keer komt het voor dat een cel bij een van de twee beoordelingsmodellen twee of meer indicatoren bevat. Dat wordt dan veroorzaakt door het feit dat het andere model voor het corresponderende onderdeel slechts één indicator kent.

Tabel 6.1

Weging van de afzonderlijke indicatoren van zowel het UvA-nakijkformulier als van de Scoringsrubriek voor Informatievaardigheden

UvA-nakijkformulier		Scoringsrubriek voor Informatievaardigheden	
Probleemstelling	10	Oriëntatie op het onderwerp	20
Verantwoording	5		
Theoretisch kader	20	Verwerking van relevante informatie tot nieuwe kennis en/of inzichten	20
Methode	30		
• evaluatie zoektocht naar bronnen (waar/hoe)		Zoektermen	10*
• evaluatie gevonden bronnen (betrouwbaar)		Kwaliteit van de gebruikte primaire informatiebronnen	20
• zinvolle en diverse selectie van bronnen		Gebruik van diverse secundaire informatiebronnen	10
• gezocht in verschillende catalogi en databases			
Evaluatie	10*		
Planning	5*		
Accuratesse	20	Bronnenoverzicht/literatuurlijst	10
• referentielijsten		Bronverwijzingen in de tekst	10
• correct citeren en parafraseren			

* Deze criteria zijn in het onderzoek niet gebruikt omdat ze geen corresponderend criterium hadden in de scoringsrubriek respectievelijk het UvA-nakijkformulier

Omdat bij het onderzoek naar de validiteit van de scoringsrubriek twee indicatoren van het UvA-nakijkformulier (Evaluatie en Planning) niet gebruikt zijn, was de in het onderzoek maximaal te behalen score met het UvA-nakijkformulier 85 punten. Voor de scoringsrubriek was het maximum aantal te behalen punten 90, aangezien daar het criterium zoektermen niet in het onderzoek werd betrokken. Tabel 6.2 geeft de weging van de indicatoren die wel gebruikt werden, maar dan uitgedrukt in percentages van het maximaal te behalen aantal punten. Uit de tabel blijkt dat inhoudelijk vergelijkbare indicatoren qua weging wel van elkaar verschillen maar dat die verschillen betrekkelijk gering zijn. Het grootste verschil bevindt zich bij de eerste groep waar de probleemstelling en verantwoording bij het UvA-nakijkformulier samen voor 17,7% meetellen terwijl het corresponderende criterium in de scoringsrubriek – oriëntatie op het onderwerp – voor 22% deel uitmaakt van de beoordeling.

Tabel 6.2

Wegingspercentages van de in het onderzoek bij Mediastudies daadwerkelijk gebruikte indicatoren

UvA-nakijkformulier		Scoringsrubriek voor Informatievaardigheden	
Probleemstelling	11,8%	Oriëntatie op het onderwerp	22%
Verantwoording	5,9%		
Theoretisch kader	23,5%	Verwerking van relevante informatie tot nieuwe kennis en/of inzichten	22%
Methode	35,3%		
• evaluatie zoektocht naar bronnen (waar/hoe)		Kwaliteit van de gebruikte primaire informatiebronnen	22%
• evaluatie gevonden bronnen (betrouwbaar)		Gebruik van diverse secundaire informatiebronnen	11%
• zinvolle en diverse selectie van bronnen			
• gezocht in verschillende catalogi en databases			
Accuratesse	23,5%	Bronnenoverzicht/literatuurlijst	11%
• referentielijsten		Bronverwijzingen in de tekst	11%
• correct citeren en parafraseren			

Verskil tussen de twee beoordelingsformulieren

Een vergelijking van de twee volledige beoordelingsmodellen zoals die zijn opgenomen als bijlage 6.2 en bijlage 6.3 maakt duidelijk dat er niet alleen sprake is van verschillen in de criteria die worden gebruikt maar ook van verschillen in de aanpak van het beoordelingsproces. De scoringsrubriek maakt gebruik van een 6 punts Likertschaal en een uitgebreidere toelichting in zinsverband op de criteria terwijl het beoordelingsmodel van de opleiding zelf slechts korte aanduidingen van de beoordelingscriteria kent, die vervolgens wel op een schaal met een grotere spreiding worden beoordeeld. De voor- en nadelen van deze verschillen in aanpak zijn gedurende het onderzoek expliciet onderwerp van gesprek geweest tijdens de nabesprekingen met de beoordelaars.

Verskil tussen het prototype van de scoringsrubriek uit bijlage 4.1 en de versie die bij de opleiding Mediastudies werd gebruikt

Na de eerste trainingsronde werd op verzoek van de beoordelaars besloten om de scoringsrubriek op twee punten aan te passen:

- Voor de duidelijkheid werd bij ieder punt op de Likertschaal het aantal toe te kennen punten voor de score in de laatste kolom toegevoegd
- Omwille van de symmetrie werd de omschrijving van de hoogste categorie op de Likertschaal gewijzigd van “uitstekend” in “zeer goed”.

Bijlage 6.3 geeft de scoringsrubriek zoals die in het onderzoek bij Mediastudies werd gebruikt.

6.3 Methode

Participanten

Het onderzoek is uitgevoerd aan de hand van de uitwerkingen van 80 eerstejaars studenten. Het aantal te beoordelen werkstukken is vooraf om budgettaire redenen tot dat getal beperkt. De gebruikte werkstukken waren random gekozen uit een sample van 119 die weer onderdeel was van een set van circa 300 werkstukken. De sample van 119 was gebaseerd op een groep docenten / begeleiders die op uitnodiging van de coördinerende docent toestemming hadden gegeven om de werkstukken te gebruiken voor het onderzoek. Nadat ook aan alle betrokken eerstejaars studenten toestemming was gevraagd om van hun uitwerking gebruik te mogen maken, bleef de sample van 119 werkstukken over.

Procedure

Voor de beoordelingen van de werkstukken werden twee recent afgestudeerde master studenten gevraagd in de rol van student-assistent. De betreffende oud studenten waren tevens betrokken geweest bij de begeleiding van studenten tijdens het werk aan de opdracht. Ze hadden in die periode echter geen beoordelingen opgemaakt en waren dan ook niet bekend met de eerder toegekende beoordelingen.

De te beoordelen werkstukken zijn vooraf anoniem gemaakt en genummerd. Om counterbalancing te kunnen toepassen met het doel storende effecten van de taakvolgorde uit te schakelen, werden de 119 beschikbare uitwerkingen random ingedeeld in drie groepen: 20 werkstukken in een set die gebruikt werd om de beoordelaars vooraf te trainen (groep T), 40 werkstukken in groep A en 40 werkstukken in groep B. Daarnaast was er een restgroep van 19 werkstukken die niet werden gebruikt.

De 80 werkstukken uit groep A en B werden door ieder van de twee beoordelaars twee maal beoordeeld: een maal met de scoringsrubriek voor informatievaardigheden en een maal met het beoordelingsmodel van de opleiding Mediastudies zelf. Voorafgaand aan het beoordelingsproces zijn twee trainingsmiddagen georganiseerd waarin de beoordelaars aan de hand van uitwerkingen uit groep T oefenden met de scoringsrubriek respectievelijk het UvA-nakijkformulier. Aan het einde van de trainingen bereikten de beoordelaars een hoge mate van overeenstemming over de beoordeling van identieke werkstukken. Voor de scoringsrubriek gold dat de beoordelaars aan het einde van de training overeenstemming hadden over alle criteria op één na, waarvoor ze aanpalende overeenstemming bereikten.

Tijdens het eigenlijke beoordelingsproces werkten de beoordelaars niet tegelijkertijd aan de zelfde uitwerkingen. De verdeling van het werk vond plaats volgens onderstaand schema:

Tabel 6.3

Tijdschema voor het beoordelen van de werkstukken bij de opleiding Mediastudies van de UvA

Beoordelaar 1	Beoordelaar 2	Periode
Groep A Scoringsrubriek	Groep B Scoringsrubriek	Week 1-3
Groep B UvA-nakijkformulier	Groep A UvA-nakijkformulier	Week 4-6
Groep B Scoringsrubriek	Groep A Scoringsrubriek	Week 7-9
Groep A UvA-nakijkformulier	Groep B UvA-nakijkformulier	Week 10-11

De beoordelaars noteerden hun bevindingen op papieren beoordelingsformulieren en legden hun scores vast in een Excel formulier. Bovendien werd na afloop van iedere beoordelingsronde een bijeenkomst georganiseerd waarin de ervaringen werden uitgewisseld en nabesproken.

Data-analyse

De interbeoordelaarsbetrouwbaarheid werd bepaald door zowel voor de afzonderlijke beoordelingscriteria van de scoringsrubriek als voor de eindcijfers vast te stellen in welke mate er sprake was:

- van absolute overeenstemming,
- aanpalende overeenstemming en
- overeenstemming over de beslissing voldoende-onvoldoende.

Nadeel van het gebruik van dergelijke percentages overeenstemming is echter dat er geen rekening wordt gehouden met de kans dat beoordelaars bij toeval overeenstemmen. In het geval dat er voldoende data beschikbaar zijn, is het dan ook verstandig om gebruik te maken van een geavanceerder kengetal. Bij intervaldata zoals die op de Likertschaal van de scoringsrubriek wordt de Intraclass Correlation Coefficient aanbevolen (Hallgren, 2012, p. 29). Van deze coëfficiënt bestaan verschillende types. In dit onderzoek was het type Two-Way Mixed (ICC(3)) het meest passend omdat er sprake was van twee beoordelaars die ieder de zelfde set van werkstukken beoordeelden (Landers, 2011). Bij deze ICC(3) kan vervolgens nog een keuze worden gemaakt tussen een variant waarbij absolute overeenstemming het uitgangspunt is (type A) en een variant waarbij onderlinge consistentie het uitgangspunt is (type C). In dit onderzoek worden beide types vastgesteld en gerapporteerd. Als norm werd aangehouden dat een waarde van 0,60 of hoger een indicatie is van een goede interbeoordelaarsbetrouwbaarheid; een waarde vanaf 0,75 werd beschouwd als excellent (Cicchetti, 1994, p. 286).

Tevens werden met behulp van de procedures factoranalyse (extractiemethode Principal Axis Factoring, rotatiemethode Oblimin met Kaiser Normalisatie) en betrouwbaarheidsanalyse in SPSS de homogeniteit en de interne consistentie van de scoringsrubriek vastgesteld.

De derde onderzoeksvraag tenslotte, die naar de criteriumvaliditeit van de scoringsrubriek, werd beantwoord door de Pearson Correlatie Coëfficiënt (r) vast te stellen tussen de eindcijfers die de afzonderlijke beoordelaars gaven met de scoringsrubriek en de eindcijfers die ze gaven met het beoordelingsmodel van de opleiding zelf.

De nabesprekingen met de twee beoordelaars die na afloop van iedere beoordelingsronde zijn georganiseerd, werden opgenomen met een memorecorder en samengevat in een schriftelijk verslag. De onderwerpen die tijdens deze nabesprekingen aan de orde kwamen, worden behandeld in een aparte paragraaf met kwalitatieve resultaten.

6.4 Kwantitatieve resultaten

Onderzoeksvraag a: Wat is de interbeoordelaarsbetrouwbaarheid van de scoringsrubriek als de 80 werkstukken door twee verschillende beoordelaars worden beoordeeld?

Tabel 6.4 geeft de absolute aantallen en de percentages voor de verschillende vormen van overeenstemming bij het gebruik van de scoringsrubriek. De aanpalende overeenstemming is steeds 80% of hoger, met uitzondering voor die van de eindbeoordelingen met een cijfer op de schaal van 1 tot 10. De overeenstemming over de beslissing voldoende-onvoldoende is wat teleurstellend en bij de eindcijfers duidelijk minder dan die bij de casus van de Nederlandse Defensie Academie in paragraaf 5.2.

Tabel 6.4

Verskillende vormen van consensus tussen beoordelaar 1 en beoordelaar 2 bij het gebruik van de scoringsrubriek bij de opleiding Mediastudies van de UvA

Criterion	N =	Absolute overeen- stemming	Aanpalende overeen- stemming	Overeenstemming over de beslissing voldoende - onvoldoende ("Decision consistency")
1 Oriëntatie (1-6)	80	30 (38%)	64 (80%)	54 (68%)
2 Literatuurlijst (1-6)	80	31 (39%)	66 (83%)	71 (89%)
3 Primaire bronnen (1-6)	80	34 (43%)	67 (84%)	65 (81%)
4 Bronverwijzingen (1-6)	80	23 (29%)	71 (89%)	58 (73%)
5 Informatieverwerking (1-6)	80	29 (36%)	67 (84%)	57 (71%)
7 Secundaire bronnen (1-6)	80	31 (39%)	65 (81%)	64 (80%)
Cijfer (1-10)	80	25 (31%)	61 (76%)	65 (81%)

De Intraclass Correlation Coefficient is zowel voor de afzonderlijke criteria vastgesteld als voor de totaalscores op de schaal van 9-90 en de uiteindelijk toegekende cijfers op een schaal van 1-10. Uit tabel 6.5 blijkt dat de resultaten, met uitzondering van die voor het eerste criterium 'oriëntatie', voor alle criteria hoger zijn dan 0.60. Voor de eindscores overschrijdt de uitkomst zelfs 0,75. Voor de uiteindelijke cijfers op een schaal van 1-10 wordt deze excellente waarde net niet bereikt.

Tabel 6.5

Intraclass Correlation Coefficient type Two-Way Mixed (ICC(3)) voor de beoordelingen met de scoringsrubriek bij de opleiding Mediastudies van de UvA

Criterion	ICC(3) type A (absolute overeenstemming)	ICC(3) type C (onderlinge consistentie)
1 Oriëntatie (1-6)	,566	,578
2 Literatuurlijst (1-6)	,618	,701
3 Primaire bronnen (1-6)	,708	,706
4 Bronverwijzingen (1-6)	,749	,751
5 Informatieverwerking (1-6)	,635	,632
7 Secundaire bronnen (1-6)	,639	,659
Eindscore (9 -90)	,763	,763
Cijfer (1-10)	,736	,736

De Intraclass Correlation Coefficient is ook vastgesteld voor de voldoende-onvoldoendes (tabel 6.6). Zoals te verwachten op basis van de cijfers uit tabel 6.4 valt ICC(3) voor de voldoende-onvoldoendes lager uit dan die voor de 6 punts schalen en de eindbeoordelingen.

Tabel 6.6

Intraclass Correlation Coefficient type Two-Way Mixed (ICC(3)) voor de beslissing voldoende-onvoldoende met de scoringsrubriek bij de opleiding Mediastudies van de UvA

Criterion	ICC(3) type A (absolute overeenstemming)	ICC(3) type C (onderlinge consistentie)
1 Oriëntatie (O-V)	,359	,363
2 Literatuurlijst (O-V)	,524	,539
3 Primaire bronnen (O-V)	,602	,599
4 Bronverwijzingen (O-V)	,575	,592
5 Informatieverwerking (O-V)	,575	,575
7 Secundaire bronnen (O-V)	,603	,603
Cijfer (O-V)	,628	,630

Onderzoeksvraag b: Wat is de interne consistentie van de beoordelingen die zijn opgemaakt met de scoringsrubriek?

Onderzoeksvraag c: Wat is de homogeniteit van de beoordelingen die zijn opgemaakt met de scoringsrubriek?

Met behulp van de procedure factoranalyse in SPSS (extractiemethode Principal Axis Factoring) werd voor beide beoordelaars vastgesteld dat de zes criteria samen één homogene factor vormen: van de zes potentiële factoren heeft er bij beoordelaar 1 slechts één een Eigenwaarde groter dan 1. Bij beoordelaar 2 is er een tweede factor met een Eigenwaarde groter dan 1 (namelijk 1,005) maar in het geval dat bij deze beoordelaar twee factoren worden geëxtraheerd waarbij de rotatiemethode Oblimin met Kaiser Normalisatie wordt gebruikt, laden zowel het criterium Oriëntatie als het criterium Secundaire bronnen hoger dan 0,300 op beide factoren. Bovendien laat de Scree Plot een duidelijke ‘knik’ zien na factor nummer 1. Beide constatering zijn een aanwijzing dat handhaving van één factor is aan te bevelen.

Uit tabel 6.7 blijkt dat voor beide beoordelaars geldt dat als alle zes beoordelingscriteria op die ene factor worden geladen, ieder criterium een hogere waarde heeft dan 0,4. Voor beoordelaar 1 komen de ladingen zelfs niet onder 0,6. Dat alles maakt voldoende aannemelijk dat de criteria gezamenlijk verwijzen naar één onderliggend construct. De item-totaal statistieken in tabel 6.8 en 6.9 bevestigen dat beeld. De waarden van Cronbach’s Alpha zijn voor beide beoordelaars zonder meer goed te noemen, waarbij die van beoordelaar 1 zelfs de 0,9 benadert. Bovendien heeft geen van de item-totaal correlaties een waarde onder 0,4 en zou alleen bij beoordelaar 2 verwijdering van een criterium leiden tot een lichte verbetering van Cronbach’s Alpha. Die verbeteringen zijn echter zo gering (0,005 voor criterium 2 ‘Literatuurlijst’ en 0,017 voor criterium 4 ‘Bronverwijzingen’) dat daar geen betekenis aan toegekend hoeft te worden.

Tabel 6.7

Factor Matrices voor de scoringsrubriek bij extractie van één factor

Criterium	Beoordelaar 1	Beoordelaar 2
1 Oriëntatie	,772	,817
2 Literatuurlijst	,717	,521
3 Primaire bronnen	,940	,881
4 Bronverwijzingen	,723	,452
5 Informatieverwerking	,822	,765
7 Secundaire bronnen	,658	,805

Extractiemethode Principal Axis Factoring

Tabel 6.8

Item-Totaal Statistieken voor de scoringsrubriek (UvA-beoordelaar 1)

Criterium	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd^a
1 Oriëntatie	,707	,881
2 Literatuurlijst	,686	,883
3 Primaire bronnen	,869	,853
4 Bronverwijzingen	,689	,883
5 Informatieverwerking	,766	,872
7 Secundaire bronnen	,629	,895

^a Cronbach's Alpha: 0,896**Tabel 6.9**

Item-Totaal Statistieken voor de scoringsrubriek (UvA-beoordelaar 2)

Criterium	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd^a
1 Oriëntatie	,744	,807
2 Literatuurlijst	,460	,857
3 Primaire bronnen	,782	,797
4 Bronverwijzingen	,422	,869
5 Informatieverwerking	,723	,810
7 Secundaire bronnen	,740	,812

^a Cronbach's Alpha: 0,852

De item-totaal statistieken in tabel 6.8 en tabel 6.9 geven geen aanleiding om te denken dat een of meer van de beoordelingscriteria zou moeten worden verwijderd. De resultaten van beoordelaar 2 geven wel aan dat de twee meer formele criteria (criterium 2 'literatuurlijst' en criterium 4 'bronverwijzingen') ieder wat minder bijdragen aan de interne consistentie van de beoordelingen. Ook bij de factoranalyse laadden juist deze twee criteria minder hoog op de ene factor die werd onderscheiden.

Onderzoeksvraag d: In welke mate geven de beoordelingen die worden opgemaakt met de scoringsrubriek een valide beeld van de beoordelingen die worden opgemaakt met het beoordelingsmodel van de opleiding zelf?

Voordat wordt ingegaan op de mate waarin de beoordelingen met de scoringsrubriek overeenstemmen met de beoordelingen die zijn opgemaakt met het 'nakijkformulier' van de opleiding Mediastudies zelf, volgen hieronder eerst enkele gegevens over ervaringen met dat laatste beoordelingsinstrument.

Zoals ook voor de scoringsrubriek is gebeurd, is voor het UvA-nakijkformulier per beoordelaar de interne consistentie vastgesteld. In tabel 6.10 en tabel 6.11 worden per beoordelaar Cronbach's Alpha en de Item-Totaal Statistieken gegeven.

Tabel 6.10

Item-Totaal Statistieken voor het UvA-nakijkformulier (UvA-beoordelaar 1)

Criterion	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd ^a
1 Probleemstelling	,759	,837
2 Verantwoording	,655	,859
3 Theoretisch kader	,809	,764
4 Methode	,836	,772
5 Accuratesse	,776	,774

^a Cronbach's Alpha: 0,843

Tabel 6.11

Item-Totaal Statistieken voor het UvA-nakijkformulier (UvA-beoordelaar 2)

Criterion	Gecorrigeerde Item-Totaal Correlatie	Cronbach's Alpha indien Item Verwijderd ^a
1 Probleemstelling	,526	,767
2 Verantwoording	,513	,794
3 Theoretisch kader	,817	,633
4 Methode	,753	,711
5 Accuratesse	,603	,721

^a Cronbach's Alpha: 0,778

Hoewel voor beide beoordelaars geldt dat Cronbach's Alpha iets lager uitvalt dan bij de scoringsrubriek, maken de gegevens uit de tabellen 6.10 en 6.11 duidelijk dat de twee beoordelaars ook met het UvA-nakijkformulier heel consistente beoordelingen hebben opgemaakt.

Ook de verschillende soorten overeenstemming over het eindcijfer zijn vergelijkbaar met die bij het gebruik van de scoringsrubriek: 27 x absolute overeenstemming en 58 x aanpalende overeenstemming op een schaal van 1-10; 64 x overeenstemming over de beslissing voldoende-onvoldoende. De

Intraclass Correlation Coefficienten (type Two-Way Mixed ICC(3)) voor de eindcijfers, zijn bij het UvA-nakijkformulier iets hoger: ,787 voor absolute overeenstemming en ,784 voor onderlinge consistentie tegenover ,736 voor beide in het geval van de scoringsrubriek. De ICC(3) voor de voldoende-onvoldoendes waren bij het UvA-nakijkformulier weer wat lager: ,603 zowel voor absolute overeenstemming als voor consistentie terwijl die voor de scoringsrubriek ,628 respectievelijk ,630 waren.

Al met al geven deze kengetallen voldoende aanleiding om het UvA-nakijkformulier te gebruiken als alternatief beoordelingsinstrument om de scoringsrubriek ter validatie mee te vergelijken. Voor allebei de beoordelaars werd een Pearson correlatiematrix opgesteld voor de eindcijfers die werden gegeven met de scoringsrubriek en die met het UvA-nakijkformulier. Voor beoordelaar 2 werd een correlatie vastgesteld van 0,76 ($p < 0,01$) en voor beoordelaar 1 zelfs van 0,93 ($p < 0,01$) (tabel 6.12). Indien wordt uitgegaan van de toegekende totaalscores zijn deze waarden 0,77 respectievelijk 0,97 (tabel 6.13). Al deze waarden worden beschouwd als hoog respectievelijk zeer hoog.

Tabel 6.12

Pearson correlatiematrix tussen de eindbeoordelingen met de scoringsrubriek en die met het UvA-nakijkformulier voor beoordelaar 1 en beoordelaar 2

	Sc Beo1	Sc Beo2	UvA Beo1	UvA Beo2
Sc Beo1	1	0,587	0,927	0,626
Sc Beo2		1	0,626	0,758
UvA Beo1			1	0,657
UvA Beo2				1

Tabel 6.13

Pearson correlatiematrix tussen de totaalscores met de scoringsrubriek (maximaal 90 punten) en die met het UvA-nakijkformulier (maximaal 85 punten) voor beoordelaar 1 en beoordelaar 2

	Sc Beo1	Sc Beo2	UvA Beo1	UvA Beo2
Sc Beo1	1	0,622	0,971	0,629
Sc Beo2		1	0,660	0,768
UvA Beo1			1	0,665
UvA Beo2				1

Wel is het zo dat de eindbeoordelingen met de scoringsrubriek voor beide beoordelaars iets hoger uitvielen (gemiddeld 6,5 respectievelijk 6,6) dan die met het nakijkformulier van de opleiding (gemiddeld 6,3 voor allebei de beoordelaars).

Tenslotte gaan we in op een verschijnsel dat enigszins verborgen in de tabellen 6.12 en 6.13 is opgenomen. Ook de correlaties waar niet naar gezocht werd - zoals die tussen de scores van beoordelaar 1 met de scoringsrubriek en beoordelaar 2 met het UvA-nakijkformulier en andersom - blijken nog hoger te zijn dan 0,60, wat heel behoorlijk is. Deze correlatie is echter wel minder hoog dan in de situatie dat de twee formulieren door één en de zelfde beoordelaar worden gebruikt.

6.5 Kwalitatieve resultaten

Tijdens de nabesprekingen van de verschillende beoordelingsrondes hebben de beoordelaars opmerkingen gemaakt over de manier waarop zij het gebruik van de scoringsrubriek hebben ervaren. In deze paragraaf worden deze opmerkingen rond vijf thema's geclusterd en kort besproken.

Likertschalen

De scoringsrubriek werd ervaren als gebruikersvriendelijk, zeker in vergelijking met het nakijkformulier van de opleiding zelf. De beoordelaars merkten op dat het gebruik van de Likertschalen in de scoringsrubriek er voor zorgde dat je geen onderdelen van de beoordeling vergat in te vullen en dat je betrekkelijk makkelijk en snel tot een score kwam. Ze hadden dan ook het gevoel dat het opmaken van een beoordeling met het UvA-nakijkformulier meer tijd kostte dan het opmaken van een beoordeling met behulp van de scoringsrubriek.

Daar stond tegenover dat hun ervaring in de derde beoordelingsronde was dat ze met het formulier van de UvA over een aantal aspecten een genuanceerder oordeel konden uitspreken dan met de scoringsrubriek. Dat speelde met name bij criteria waar in het nakijkformulier van Mediastudies 10 punten of meer te verdelen waren terwijl je bij het vergelijkbare criterium uit de scoringsrubriek slechts kon kiezen uit de 6 alternatieven van de Likertschaal. Deze ervaring kwam vooral in de derde beoordelingsronde aan het licht toen met de scoringsrubriek werkstukken werden beoordeeld die een paar weken eerder al waren beoordeeld met het UvA-nakijkformulier.

Beoordelen van inhoudelijke kennis en competenties

De beoordelaars kwamen regelmatig terug op de vraag of de beoordelingsinstrumenten wel voldoende aandacht besteden aan kennis en competenties die specifiek betrekking hebben op het vak Mediastudies. Daarbij maakten zij onderscheid tussen meer formele criteria - zoals correct gebruik van een citation style in de beschrijvingen van referentielijst en de bronverwijzingen in de tekst - en meer inhoudelijke criteria. Met dat laatste werd dan bijvoorbeeld gedoeld op het beoordelen van de relevantie van een bron voor het beantwoorden van de onderzoeksvraag en de bespreking van begrippen en theorieën, wat in het UvA-nakijkformulier wordt geschaard onder het criterium 'Theoretisch kader'.

In de scoringsrubriek zoals die door de beoordelaars in dit project werd gebruikt, hadden slechts drie van de zes criteria wat meer betrekking op vakinhoud (oriëntatie, kwaliteit van de primaire bronnen en verwerking van relevante informatie tot nieuwe kennis). Daarbij dient opgemerkt te worden dat deze drie criteria wel extra zwaar mee telden bij het vaststellen van de eindscore, zodat ze samen 6/9 (67%) van het eindcijfer uit maakten. Een van de beoordelaars wees echter op het feit dat dat het gevoel van onvoldoende aandacht voor de inhoud niet altijd weg neemt. Hij verwoordde het als volgt: "Ik denk dat het idee om er meer gewicht aan te geven misschien nog niet voldoende is. Het is namelijk ook een kwantitatief iets, zo van.. hoeveel criteria heb je ten opzichte van de criteria die over de inhoud gaan? Als je twee criteria over inhoud hebt en je hebt er vier over de andere dingen, dan voelt het voor jou als nakijker altijd heel erg dat je niet genoeg met de inhoud bezig bent."

Omdat de beoordelaars met deze ervaring nauw aansloten bij andere geluiden in de academische wereld over de aandacht voor 'algemene vaardigheden' (zie bijvoorbeeld Godwin, 2005, p. 70), zal in paragraaf 6.7 ('Discussie') nader worden ingegaan op hun gevoel en de wijze waarop dat strookt met de andere gegevens uit dit onderzoek.

Interbeoordelaarsbetrouwbaarheid voor criterium 1, 'Oriëntatie'

Waarom ICC(3) juist voor het eerste criterium (oriëntatie) onder het gewenste niveau was, is tijdens de nabespreking met de beoordelaars kort aan de orde geweest. Beoordelaar 2 gaf als mogelijke verklaring dat hij juist bij dit criterium extra streng is geweest. Hoewel dat geen gevolgen hoeft te hebben voor de onderlinge *consistentie* op dat criterium, is zijn opmerking wel aanleiding geweest om na te gaan wat voor de afzonderlijke beoordelaars de gemiddelde score en de standaarddeviatie per criterium was. Vandaar dat in deze paragraaf met kwalitatieve resultaten nog een tabel met kwantitatieve resultaten wordt gegeven.

Tabel 6.14 geeft aan dat het criterium oriëntatie inderdaad een van de twee criteria is waarbij de gemiddelde score van beoordelaar 2 lager is dan die van beoordelaar 1. Bovendien is het verschil daar ook groter dan bij het andere criterium, de bronverwijzingen. Op de andere criteria scoort beoordelaar 2 gemiddeld steeds iets hoger dan beoordelaar 1 of gelijkwaardig.

Tabel 6.14

Gemiddelde scores en standaarddeviaties voor de beoordelingen met de scoringsrubriek bij de opleiding Mediastudies van de UvA

Criterium	Beoordelaar 1	Beoordelaar 2
1 Oriëntatie (1-6)	4,1 (0,97)	3,8 (1,15)
2 Literatuurlijst (1-6)	4,5 (1,03)	5,2 (1,01)
3 Primaire bronnen (1-6)	4,3 (1,11)	4,4 (1,30)
4 Bronverwijzingen (1-6)	4,0 (1,00)	3,8 (1,27)
5 Informatieverwerking (1-6)	3,8 (0,98)	3,8 (1,29)
7 Secundaire bronnen (1-6)	4,1 (1,17)	4,4 (1,01)
Cijfer (1-10)	6,5 (1,48)	6,6 (1,67)

Consequenties voor studenten

De beoordelaars hadden in een eerder stadium studenten begeleid bij de uitvoering van de betreffende opdracht maar daarbij geen beoordelingen opgemaakt. Ze waren sowieso nooit eerder ingezet als beoordelaar van opdrachten in het onderwijs. Dat maakte de betrokkenheid bij het project voor hen aantrekkelijk. Voortbouwend op ervaringen die ze eerder als student-assistent hadden opgedaan konden ze nu een stap verder zetten in hun ontwikkeling als expert op het gebied van mediastudies. Gedurende het project kwam met name beoordelaar 1 er echter ook achter dat ze het wel plezierig vond dat haar beoordelingen geen consequenties zouden hebben voor de betrokken studenten. Het ontbreken van de verantwoordelijkheid om studenten eventueel af te wijzen, werd door haar dan ook niet als negatief ervaren maar eerder als een gunstige omstandigheid om ervaring op te doen met het beoordelen van studentenwerk.

Volgorde waarin de beoordelingscriteria worden gepresenteerd

Tijdens de laatste nabespreking deden de twee beoordelaars de suggestie om bij de scoringsrubriek de volgorde van de criteria te wijzigen. Daarbij ging het met name om de meer formele criteria: criterium 2 (literatuurlijst) en criterium 4 (bronverwijzingen). Dit betreft juist de criteria die bij beoordelaar 2 lager op de factoranalyse laadden.

6.6 Conclusies

Doel van het onderzoek dat in dit hoofdstuk is beschreven, was het vaststellen van de betrouwbaarheid en validiteit van de Scoringsrubriek voor Informatievaardigheden. In het onderzoek zijn daarvan weer verschillende vormen onderscheiden. Ze worden hierna afzonderlijk behandeld.

Onderzoeksvraag a): Interbeoordelaarsbetrouwbaarheid

Op basis van de resultaten uit paragraaf 6.4 kan worden geconcludeerd dat de onderlinge overeenstemming tussen de twee beoordelaars ('interbeoordelaarsbetrouwbaarheid') in dit geval goed was te noemen. Aanpalende overeenstemming voor de afzonderlijke criteria was met 80% of hoger zonder meer acceptabel; de aanpalende overeenstemming voor de eindcijfers was iets lager maar had dan ook betrekking op een 10 punts schaal in plaats van een 6 punts schaal. De Intraclass Correlation ICC(3) was alleen voor het eerste criterium (oriëntatie) onder het gewenste niveau van 0,60. Die voor de eindcijfers benaderde de excellente waarde van 0,75.

Voor het feit dat de ICC(3) voor het criterium 'oriëntatie' onder het gewenste niveau was, gaf een van de beoordelaars de verklaring dat hij juist bij dat criterium extra streng is geweest. Een andere verklaring is echter dat beoordelaars de neiging hebben om juist bij dit criterium hun eigen accenten te leggen. Het criterium geeft, meer dan veel van de andere criteria, daarvoor relatief veel vrijheid: het gaat om de kwestie of de student goede vragen heeft gesteld en of hij/zij een duidelijke eigen invalshoek heeft gekozen. Het enige andere criterium uit de scoringsrubriek met zo'n meer subjectief karakter is criterium 5, de 'verwerking van relevante informatie tot nieuwe kennis en/of inzichten'.

In tegenspraak met bovenstaande constatering dat de twee beoordelaars het over het algemeen goed met elkaar eens waren, is de constatering dat de overeenstemming over de beslissing voldoende-onvoldoende wel lager was dan gewenst. Voor vier van de zes afzonderlijke criteria was ICC(3) type A lager dan 0,60, voor ICC(3) type C gold dat zelfs voor vijf van de zes criteria. Gelukkig waren beide typen ICC(3) bij de beslissing voldoende-onvoldoende voor de eindcijfers wel hoger dan 0,60, zij het dat de waarden nog altijd 0,10 lager waren dan bij de score op de volledige schaal van 1-10. Een verklaring voor dit verschijnsel kan worden gevonden in het ontbreken van ervaring bij de beoordelaars met het nemen van deze voor studenten zo belangrijke beslissingen. In de nabesprekingen heeft één van de beoordelaars aangegeven dat zij zich op dat gebied nog erg onzeker voelde en dat ze blij was dat er voor studenten geen consequenties verbonden waren aan de beoordelingen.

Een andere verklaring was volgens de beoordelaars echter dat heel veel van de papers die werden beoordeeld, op het randje van de voldoende of onvoldoende balanceerden. “Dan kan het heel snel zijn dat ze een tik de ene kant op krijgen of net de andere kant op” (beoordelaar 2). Dat beeld van het grote aantal ‘twijfelgevallen’ komt trouwens weer niet overeen met de cijfers die in werkelijkheid door de echte docenten waren toegekend. Die hadden echter gebruik gemaakt van het volledige nakijkformulier uit bijlage 1 en bestonden zelf weer uit een groep van vijf beoordelaars terwijl de aantallen beoordeelde werkstukken zeer onevenwichtig over hen waren verdeeld. Vanwege de vele variabelen waarover de onderzoeker geen controle had, is besloten om van die gegevens geen gebruik te maken.

Onderzoeksvragen b) en c): Interne consistentie en homogeniteit (constructvaliditeit)

Waren de waarden voor de indicatoren van de interbeoordelaarsbetrouwbaarheid over het algemeen ‘goed’ te noemen, die voor de interne consistentie van de beoordelingen vielen nog wat beter uit. Vooral beoordelaar 1 wekte de indruk met de scoringsrubriek zeer consistent te kunnen werken. Cronbach’s Alpha benaderde in haar geval de 0,90, wat wordt beschouwd als excellent. Daar staat tegenover dat je in haar geval ook zou kunnen denken dat zij zich bij het scoren van afzonderlijke criteria ook heeft laten leiden door het ‘algemene beeld’ dat zij van het werkstuk had en dat beoordelaar 2 wat meer analytisch en gedifferentieerd was met zijn scores. Het feit dat de standaarddeviatie bij beoordelaar 2 over het algemeen wat hoger is (zie tabel 6.14), bevestigt dat. Voor complexe en samengestelde opdrachten zoals dit onderzoeksvoorstel dat de studenten bij de opleiding Mediastudies dienden te schrijven, geven onderwijskundigen overigens de voorkeur aan een meer holistische benadering. De analytische benadering kan namelijk leiden tot het louter afvinken van afzonderlijke criteria terwijl bij dit soort opdrachten het geheel vaak meer is dan de som van de delen (Hogeschool van Amsterdam, 2014). Het is wel plezierig om te merken dat het verschil in benadering tussen de twee beoordelaars in dit project niet heeft geleid tot extreme verschillen in de beoordelingen.

Met betrekking tot de constructvaliditeit kan in ieder geval gesteld worden dat de item-totaal statistieken voldoende aanwijzingen geven voor de conclusie dat de zes beoordelingscriteria gezamenlijk betrekking hebben op één en het zelfde construct. Dat werd nog eens onderschreven door de uitkomsten van de factoranalyses die werden uitgevoerd op de gegevens van de afzonderlijke beoordelaars. Ook die gaven aan dat de zes beoordelingscriteria verwijzen naar één en de zelfde dimensie. Daarmee is overigens nog steeds niet gezegd dat de beoordelingscriteria verwijzen naar het bedoelde construct informatievaardigheden en dat het model in afbeelding 6.1 correct is. De theoretische exercities in hoofdstuk 4 tijdens het ontwerpen van de scoringsrubriek maken dat echter wel waarschijnlijk. Cronbach en Mehl (1995, p. 294) zeggen in zo’n geval dat de validiteit “acceptabel” is.

Het is dan ook aannemelijk dat de scoringsrubriek een valide afbeelding is van het construct Informatievaardigheden. Wel is door de beoordelaars opgemerkt dat criterium 2 (Bronnenoverzicht/literatuurlijst) en criterium 4 (Bronverwijzingen in de tekst) inhoudelijk meer met elkaar te maken hebben dan in de huidige presentatie van de scoringsrubriek wordt gesuggereerd. Daaruit volgt de aanbeveling om deze twee criteria in een volgende versie van de scoringsrubriek in elkaars context

te presenteren. Het ligt voor de hand om criterium 4 (Bronverwijzingen in de tekst) dan te verplaatsen naar positie 2, zodat de volgorde tussen het huidige criterium 2 (Bronnenoverzicht / literatuurlijst) en criterium 3 (Kwaliteit van de gebruikte primaire informatiebronnen) blijft gehandhaafd.

Onderzoeksvraag d): Criteriumvaliditeit

Om te achterhalen in hoeverre de scoringsrubriek ook daadwerkelijk meet wat ze pretendeert te meten, namelijk de informatievaardigheden van een student, zijn de beoordelingen met de rubric vergeleken met beoordelingen die werden opgemaakt met het reeds bestaande beoordelingsinstrument dat door de opleiding zelf werd gebruikt. Vooraf mocht worden aangenomen dat het UvA-nakijkformulier goed zou aansluiten bij de manier waarop de beoordelaars gewend zijn naar de uitwerkingen te kijken. Anders dan de scoringsrubriek kwam het UvA-nakijkformulier immers uit de context waarin de beoordelaars zelf waren opgeleid en waarin ze gewend waren te werken.

Voor beide beoordelaars gold dat de twee verschillende soorten beoordelingen hoog met elkaar correleerden, voor een van de beoordelaars zelfs zeer hoog. Maar ook met betrekking tot deze correlaties geldt dat we dan nog steeds niet weten of de scoringsrubriek ook werkelijk het beoogde construct informatievaardigheden meet. Ook hier geldt echter dat de eerder uitgevoerde theoretische exercities dat wel aannemelijk maken.

6.7 Discussie

De algehele conclusie van dit onderzoek bij de opleiding Mediastudies van de UvA luidt dat er voldoende aanwijzingen zijn dat de scoringsrubriek in deze casus een betrouwbaar en valide instrument is geweest voor het beoordelen van informatievaardigheden. Voor de betrekkelijk lage overeenstemming over de beslissing voldoende-onvoldoende, kon geen andere verklaring worden gevonden dan het feit dat de betrokken beoordelaars allebei in het geheel geen ervaring hadden met het uitspreken van summatieve beoordelingen. Tekenend is dat de overeenstemming over deze certificerende beslissingen bij het gebruik van het UvA-formulier nog wat lager uitvielen.

De aanpak die is gebruikt bij het ontwerpen van de scoringsrubriek en die werd beschreven in hoofdstuk 4, heeft zeer waarschijnlijk bijgedragen aan de betrouwbaarheid en de validiteit ervan. De bruikbaarheid van de scoringsrubriek wordt echter niet alleen bepaald door de gekozen criteria, de omschrijvingen van het gedrag op het professionele en inadequate niveau of de gebruikte schaalverdeling. Ook een logische presentatie van de criteria heeft invloed op het gebruiksgemak (Zimmaro, 2004). Dat geldt met name als niet alleen beoordelaars gebruik maken van de scoringsrubriek maar ook studenten. Ter bevordering van het gebruik van de scoringsrubriek zou je de criteria ook in een willekeurige volgorde kunnen aanbieden aan studenten en hen vervolgens zelf de definitieve volgorde laten bepalen.

De twee beoordelaars die er bij de opleiding Mediastudies mee hebben gewerkt gaven in ieder geval aan dat de logische volgorde verbeterd zou kunnen worden. De onderzoeker zelf kwam tot de conclusie dat het dan de voorkeur heeft om beoordelingscriterium 4 naar de 2^e positie te verplaatsen. In die nieuwe volgorde van de criteria zou de samenhang tussen de bronverwijzingen in de tekst en

de bronnenlijst zelf dan beter tot haar recht komen. Het is zelfs niet uitgesloten dat de volgorde van de beoordelingscriteria invloed heeft op de scores die worden toegekend.

Het nadeel dat de scoringsrubriek wat minder gelegenheid geeft tot nuancering van het oordeel - een argument dat door de beoordelaars werd genoemd tijdens een van de nabesprekingen - kan worden opgeheven door het aantal punten per categorie op de Likertschaal ook enigszins te spreiden. Zo zouden studenten die op het eerste criterium dan “zeer goed” scoren niet per se 20 punten scoren maar een aantal punten van 17-20 en studenten met een “goed” een aantal punten van 13-16. Een tweede alternatief is om de 6 punts Likertschaal te vervangen door een 10 punts schaal.

Te verwachten is dat de scoringsrubriek voor informatievaardigheden ook nog heel andere bezwaren zal oproepen bij een grote groep vakdocenten. Een veel voorkomende bedenking is namelijk dat expliciete aandacht voor informatievaardigheden ten koste zal gaan van de aandacht voor vakinhoudelijke onderwerpen (Godwin, 2005, pp. 70-71). Deze discussie lijkt overigens op die rond de 21st century skills die rond 2010 is losgebarsten (zie bijvoorbeeld Van der Velden 2011, p. 383). Tijdens de nabesprekingen van de beoordelingsrondes met de scoringsrubriek voor informatievaardigheden lieten de beoordelaars weten dat ook zij wel eens het gevoel hadden dat vakinhoud ondergesneeuwd zou raken. De hoge correlaties tussen de beoordelingen met de scoringsrubriek en die met het formulier van Mediastudies, geven echter geen aanleiding om daar bezorgd over te zijn. De scoringsrubriek geeft bij verschillende criteria ook voldoende ruimte om die met kennis en vaardigheden uit het eigen vakgebied in te vullen. Ook de eerder genoemde Van der Velden geeft aan dat de verwerving van generieke vaardigheden niet ten koste hoeft te gaan van de verwerving van specifieke kennis of vaardigheden. De generieke vaardigheden worden namelijk het best verworven binnen de context van een specifiek beroep of vakgebied (Van der Velden, 2011, pp. 386-387). De opmerkingen van de beoordelaars bij de opleiding Mediastudies van de UvA geven echter wel aan dat hier bij de ingebruikname van de scoringsrubriek voldoende voorlichting over gegeven dient te worden.

In het voorgaande is steeds uitgegaan van de scoringsrubriek als één geheel. Daarbij dient wel opgemerkt te worden dat, zoals beschreven in paragraaf 6.1 bij ‘Methode’, een van de beoordelingscriteria in het onderzoek uiteindelijk niet is gebruikt (criterium 6, ‘Zoektermen’). Tijdens de eerste beoordelingsronde bleek namelijk dat dit element van het zoekproces onvoldoende onderdeel was geweest van de instructie aan studenten en dat verantwoording van gebruikte zoektermen in veel gevallen dan ook afwezig was. Te verwachten is dat dergelijke situaties zich vaker zullen voordoen bij de implementatie van de scoringsrubriek. Lang niet alle studieopdrachten behelzen alle facetten van de scoringsrubriek. Uiteraard staat het opleidingen dan vrij om criteria te laten vervallen. Wel dient ervoor gewaarschuwd te worden dat er voldoende beoordelingscriteria zouden moeten overblijven. Ook in dit onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam trad namelijk het verschijnsel op dat de overeenstemming tussen twee beoordelaars op de afzonderlijke criteria minder hoog is dan die op een groep van criteria. Dit verschijnsel dat meetfouten op afzonderlijke onderdelen worden gecompenseerd als gebruik wordt gemaakt van een samengestelde schaal, werd ook bij de casus van de Nederlandse Defensie Academie opgemerkt.

De aansluiting van de scoringsrubriek op bestaande studieopdrachten heeft echter ook een andere kant. De scoringsrubriek voor informatievaardigheden kan namelijk ook gebruikt worden voor het bedenken van goede studieopdrachten. In de literatuur wordt regelmatig aandacht besteed aan het feit dat een scoringsrubriek niet alleen een functie heeft als evaluatie- of beoordelingsinstrument maar ook gebruikt kan worden om de instructie en het leren zelf te verbeteren. Een voorbeeld daarvan is een veel geciteerd overzichtsartikel van Reddy en Andrade (2010). In hoofdstuk 8 van dit proefschrift zal onder andere op basis van dat overzichtsartikel worden ingegaan op de vraag op welke verschillende manieren docenten van De Haagse Hogeschool de scoringsrubriek gebruiken om de instructie te optimaliseren. Voorbeelden zijn gevonden van onderwijsteams die de Scoringsrubriek voor Informatievaardigheden hebben gebruikt bij het formuleren van leerdoelen en eindtermen. Van het gebruik van een scoringsrubriek als input voor het formuleren van goede studieopdrachten is in het onderzoek echter geen voorbeeld gevonden.

Een methodologische beperking van het onderzoek bij de opleiding Mediastudies was dat de beoordelaars in de derde beoordelingsronde het gevoel hadden dat ze niet geheel onbevangen naar de werkstukken hadden kunnen kijken omdat ze zich de beoordelingen van de zelfde werkstukken uit de tweede beoordelingsronde nog te goed meenden te herinneren. Die werkstukken waren in de vorige beoordelingsronde immers met het UvA-nakijkformulier beoordeeld (zie ook het schema in tabel 6.3). Consequentie van deze opmerking zou kunnen zijn dat de correlatie tussen de beoordelingen uit de 2^e beoordelingsronde en die uit de 3^e beoordelingsronde kunstmatig hoog uit viel omdat de beoordelingsmomenten te dicht op elkaar lagen.

Een proef op de som wees echter uit dat de correlatie tussen de eindcijfers uit periode 2 en 3 in het geheel niet hoger uitpakte dan die tussen de eindcijfers in periode 1 en 4. Voor beoordelaar 2 was die laatste correlatie zelfs iets hoger.

De algehele conclusie van dit onderzoek bij de opleiding Mediastudies van de UvA luidt dat er voldoende aanwijzingen zijn dat de scoringsrubriek in deze casus een betrouwbaar, valide en praktisch bruikbaar instrument is geweest voor het beoordelen van informatievaardigheden. De versie die gebruikt is in dit onderzoek en die bij dit hoofdstuk als bijlage 6.1 is afgedrukt, bleek op twee punten vatbaar voor verbetering:

Beoordelingscriterium 4 zou moeten worden verplaatst naar de 2^e positie. In de nieuwe presentatievolgorde komt de samenhang tussen de bronverwijzingen in de tekst en de bronnenlijst zelf, beter tot haar recht.

De scores die bij de categorieën op de Likertschaal worden vermeld, zouden moeten worden gewijzigd in een cijferreeks. Dat geeft de beoordelaars de gelegenheid om iets meer nuance in de score in de laatste kolom aan te brengen. In de instructie bij de scoringsrubriek moet dan wel duidelijk worden aangegeven hoe hiermee moet worden omgegaan. Een alternatieve oplossing zou zijn om de 6 punts Likertschaal te vervangen door een 10 punts Likertschaal.

Bijlage 6.1: UvA-nakijkformulier van de opleiding Mediastudies voor 'Toets II: Het onderzoeksvoorstel'

Paper nummer:

Naam beoordelaar:

Datum:

Aspect	Te behalen	Behaald	Totaal
Poortwachter (4.0)			
Incorrect aantal woorden (10%, 2250- 750)			
> 5 taalfouten			
1. Probleemstelling	10		
Kwaliteit centrale vraag (of stelling)			
Kwaliteit deelvragen (logisch, functioneel)			
Uitleg alternatieve formuleringen			
Voorlopige inhoudsopgave van het onderzoek			
2. Verantwoording	5		
Uitleg van de relevantie van de probleemstelling			
3. Theoretisch kader	20		
Situering probleemstelling in het relevante kennisgebied (focus en afbakening)			
Bespreking begrippen en theorieën			
4. Methode	30		
Kwaliteit clusters (logisch, werkbaar)			
Presentatie 25 bronnen verspreid over clusters			
Uitleg per cluster over inzet bronnen			
Evaluatie zoektocht naar bronnen (waar/hoe)			
Evaluatie gevonden bronnen (betrouwbaar)			
Nog benodigde informatie?			
Zinvolle en diverse selectie van bronnen			
Gezocht in verschillende catalogi en databases			
5. Evaluatie	10		
Terugblik op het verloop van het onderzoek			
Inzicht in mediahistorisch onderzoek doen			
Gebruik hoorcolleges en literatuur			
6. Planning	5		
Activiteitenplan en tijd- en faseringsschema			
Mogelijke obstakels en valkuilen			
7. Accuratesse	20		
Referentielijst (MLA)			
Technische verzorging (incl. correct citeren en parafraseren)			
Totaal	100		
Eindcijfer			

Opmerkingen docent:

Toelichting die de opleiding Mediastudies geeft bij het nakijkformulier

Algemeen

- Toets II telt voor 40% van het eindcijfer mee
- Toets II levert een deelcijfer op en kan dus alleen bestaan uit halve en hele cijfers tussen tenminste 1 en ten hoogste 10. De volgende uitslagen zijn niet toegestaan: 4,2; 7,85; 0,5; NAV
- We ronden naar beneden af. Een 88 wordt 8,5
- Als een student niet heeft deelgenomen aan de toets: 0. Dit geldt ook als de toets niet fysiek is ingeleverd of niet op Ephorus is ingeleverd.
- Toets II dient zowel fysiek als bij Ephorus ingeleverd te worden
- De deadline van toets II is 2 april om 15 uur. Onderhandel niet over de deadline en accepteer geen gemailde werkstukken (ook niet als de deadline gemist dreigt te worden)

Toelichting op de aspecten

Poortwachter (4.0)

- De omvang (2.250-2.750 woorden) wordt bepaald exclusief referentielijst en notenapparaat
- Ga niet onder de 4 als een werkstuk niet 'aan de voorwaarden voldoet'. Een student zou bij de herkansing wel aan de voorwaarden moeten kunnen voldoen

Technische verzorging (minpunten bij ontoereikende verzorging)

- Formele gegevens aangegeven (naam, werkgroepsnummer etc.)
- Nietje/ niet in plastic mapje, en dergelijken
- Lay-out:
 - kopjes zijn vet (niet vet én cursief)
 - alinea's zijn herkenbaar doordat ze afgesloten worden met een harde return en een tab aan het begin van de nieuwe regel van de volgende alinea
 - Grotere gehelen van alinea's, zoals paragrafen, worden onderscheiden door een regel wit
 - slechts één lettertype, en wel in 11- of 12-puntsgrootte
 - de regelafstand is anderhalf
 - de marge is ongeveer drie centimeter
 - teksten die meerdere pagina's omvatten, hebben een paginanummering

Bijlage 6.2: UvA-nakijkformulier zoals gebruikt in het onderzoek*

Paper nummer:

Naam beoordelaar:

Datum:

Aspect	Te behalen	Behaald	Totaal
Poortwachter (4.0)			
Incorrect aantal woorden (10%, 2250- 750)			
> 5 taalfouten			
1. Probleemstelling	10		
Kwaliteit centrale vraag (of stelling)	6		
Kwaliteit deelvragen (logisch, functioneel)	4		
Uitleg alternatieve formuleringen			
Voorlopige inhoudsopgave van het onderzoek			
2. Verantwoording	5		
Uitleg van de relevantie van de probleemstelling	5		
3. Theoretisch kader	20		
Situering probleemstelling in het relevante kennisgebied (focus en afbakening)	12		
Bespreking begrippen en theorieën	8		
4. Methode	30		
Kwaliteit clusters (logisch, werkbaar)			
Presentatie 25 bronnen verspreid over clusters			
Uitleg per cluster over inzet bronnen			
Evaluatie zoektocht naar bronnen (waar/hoe)	5		
Evaluatie gevonden bronnen (betrouwbaar)	10		
Nog benodigde informatie?			
Zinvolle en diverse selectie van bronnen	10		
Gezocht in verschillende catalogi en databases	5		
5. Evaluatie	10		
Terugblik op het verloop van het onderzoek			
Inzicht in mediahistorisch onderzoek doen			
Gebruik hoorcolleges en literatuur			
6. Planning	5		
Activiteitenplan en tijd- en faseringsschema			
Mogelijke obstakels en valkuilen			
7. Accuratesse	20		
Referentielijst (MLA)	10		
Technische verzorging (incl. correct citeren en parafraseren)	10		
Totaal	85		
Eindcijfer			

* De grijs afgedrukte onderdelen zijn in het onderzoek niet gebruikt. De vet gedrukte omschrijvingen en cijfers verwijzen naar de gebruikte hoofdcriteria. De omschrijvingen en cijfers in normale typografie hebben betrekking op de subcriteria. De cijfers verwijzen in beide gevallen naar de maximaal te behalen score.

Bijlage 6.3:

Scoringsrubriek zoals gebruikt in het onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam*

Onderdeel: studentenproduct

Criterium		Professioneel gedrag	Inadequaate gedrag	Score 1-20=
1	Oriëntatie op het onderwerp	☐ Uit het product blijkt dat de student zich goed op het vraagstuk of het onderwerp heeft georiënteerd en op basis daarvan een eigen invalshoek (focus) heeft geformuleerd. Dat blijkt ook uit het feit dat de student duidelijk heeft geformuleerd welke (onderzoeks)vragen hij/zij wil beantwoorden.	☐ Uit het product blijkt dat de student is uitgegaan van de oorspronkelijke vraag of de oorspronkelijke formulering van het onderwerp (zoals die zijn gegeven in het assignment) zonder dat hij/zij die verder heeft uitgediept. Dat blijkt bijvoorbeeld uit het feit dat de betekenis van kernbegrippen bekend wordt verondersteld en deze niet nader worden gedefinieerd.	
Beoordeling:		0 zeer goed (20) 0 goed (16) 0 voldoende (12)	0 zwak (10) 0 slecht (6) 0 zeer slecht (2)	

Criterium		Professioneel gedrag	Inadequaate gedrag	Score 1-10=
2	Bronnen-overzicht / literatuurlijst	☐ Het product bevat een bronnenoverzicht dat volledig is en waarbij een standaard citation style correct is toegepast. Aan de hand van het bronnenoverzicht zijn de gebruikte bronnen eenduidig te achterhalen. Toelichting: het laatste is belangrijker dan het format van de titelbeschrijvingen (de 'citation style'). Voor de score 'uitstekend' dient ook de citation style echter correct te zijn toegepast.	☐ In het product ontbreekt een bronnenoverzicht en / of ☐ het bronnenoverzicht is verre van volledig (van verschillende documenten waar in de tekst naar wordt verwezen ontbreken de gegevens) en / of ☐ cruciale bibliografische gegevens (titel, auteur, publicatiejaar) ontbreken. Voorbeeld dat vaak voorkomt in de praktijk: van internetbronnen zijn alleen de url's vastgelegd.	
Beoordeling:		0 zeer goed (10) 0 goed (8) 0 voldoende (6)	0 zwak (5) 0 slecht (3) 0 zeer slecht (1)	

Criterium		Professioneel gedrag	Inadequaate gedrag	Score 1-20=
3	Kwaliteit van de gebruikte primaire informatie-bronnen (boeken, artikelen, websites etc.)	☐ Uit het bronnenoverzicht dat bij het product behoort, blijkt dat de student gebruik heeft gemaakt van relevante, betrouwbare (bij voorkeur zo authentiek mogelijke) en actuele informatiebronnen die het onderwerp bovendien vanuit verschillende gezichtspunten behandelen.	☐ De gebruikte externe bronnen zijn te licht, verouderd of onvoldoende relevant. Een voorbeeld van 'te licht' is als de student alleen internet-sites als informatiebron heeft gebruikt. En / of ... ☐ De gebruikte literatuur is eenzijdig. Er is bijvoorbeeld alleen gebruik gemaakt van overheidsinformatie (.gov-sites) of van publicaties van één bepaalde auteur.	
Beoordeling:		0 zeer goed (20) 0 goed (16) 0 voldoende (12)	0 zwak (10) 0 slecht (6) 0 zeer slecht (2)	

Criterium		Professioneel gedrag	Inadequaate gedrag	Score 1-10=
4	Bron-verwijzingen in de tekst	☐ In de tekst van het product zelf heeft de student eenduidig verantwoord welke informatiebronnen hij of zij heeft gebruikt. Dat geldt ook voor afbeeldingen of voor audiovisuele illustraties in het geval van een digitaal product.	☐ De student heeft werk van anderen (stukken tekst, afbeeldingen, geluidsfragmenten) verwerkt in zijn/haar eigen product zonder dat hij/zij de gebruikte bron heeft vermeld. Hoewel vaak niet opzettelijk, is er strikt gesproken sprake van 'plagiat'.	
Beoordeling:		0 zeer goed (10) 0 goed (8) 0 voldoende (6)	0 zwak (5) 0 slecht (3) 0 zeer slecht (1)	

Criterium		Professioneel gedrag	Inadequaaf gedrag	Score 1-20=
5	Verwerking van relevante informatie tot nieuwe kennis en/of inzichten	☐ Uit het product blijkt dat de student informatie uit verschillende bronnen heeft geanalyseerd en op basis daarvan nieuwe toepassingen, inzichten of hypothesen heeft geformuleerd. Toelichting: in praktijk blijkt dat veel studenten wel toekomen aan de analyse en vergelijking van verschillende informatiebronnen, maar niet in staat zijn om de gevonden gegevens te synthetiseren tot een nieuw inzicht of nieuwe toepassing. Dit criterium wordt in dat geval beoordeeld als 'voldoende' of 'zwak'.	In het studentenproduct heeft de student ☐ de inhoud van de gebruikte informatiebronnen niet correct of niet helder weergegeven, en / óf ☐ in het geheel geen aandacht besteed aan analyse van de opgespoorde informatiebronnen, en / óf ☐ zich gebaseerd op slechts één informatiebron en hij/zij heeft de daar gevonden informatie niet ter discussie gesteld terwijl daar wel aanleiding toe is.	
Beoordeling:		0 zeer goed (20) 0 goed (16) 0 voldoende (12)	0 zwak (10) 0 slecht (6) 0 zeer slecht (2)	

Onderdeel: procesverslag / zoekstrategie

Criterium		Professioneel gedrag	Inadequaaf gedrag	Score 1-10=
6	Zoektermen	☐ De student heeft de zoektermen gebruikt die relevant zijn voor het te onderzoeken onderwerp / vraagstuk. Hij/zij heeft daarbij rekening gehouden met het vakjargon, de synoniemen en (zoek)termen in een andere taal (meestal Engels).	☐ De zoektermen die de student heeft gebruikt zijn te algemeen of komen uit het algemeen spraakgebruik. en / óf ☐ de student heeft geen rekening gehouden met synoniemen, verwante begrippen en (zoek)termen in een andere taal (meestal Engels).	
Beoordeling:		0 zeer goed (10) 0 goed (8) 0 voldoende (6)	0 zwak (5) 0 slecht (3) 0 zeer slecht (1)	

Criterium		Professioneel	Inadequaaf gedrag	Score 1-10=
7	Gebruik van diverse secundaire informatiebronnen	☐ De student heeft gebruik gemaakt van diverse typen secundaire bronnen (search engines, boeken, wetenschappelijke tijdschriften, databases, sociale netwerken). Indien van toepassing heeft hij/zij bovendien gebruik gemaakt van mogelijkheden buiten de eigen locatie (IBL, documentleverantie).	☐ De student heeft vooral informatiebronnen gebruikt die gemakkelijk toegankelijk zijn. Voorbeeld: de student heeft alleen gebruik gemaakt van • het internet via het 'eenvoudige zoek-scherm' van een grote algemene zoek-machine en / of • lesmateriaal dat door docenten is verstrekt.	
Beoordeling:		0 zeer goed (10) 0 goed (8) 0 voldoende (6)	0 zwak (5) 0 slecht (3) 0 zeer slecht (1)	

Totaalscore (maximaal 100) =



Op dit werk is een Creative Commons licentie van toepassing:

Naamsvermelding – Niet commercieel – Gelijk delen

Jos van Helvoort – De Haagse Hogeschool

* Criterium 6 is uiteindelijk niet gebruikt omdat rapportage van de gebruikte zoektermen geen onderdeel was van de instructie

Evalueren met studenten uit de deeltijdopleiding idm van De Haagse Hogeschool¹³

Abstract: Met als doel inzicht te krijgen in de wijze waarop studenten de scoringsrubriek voor informatievaardigheden (zouden kunnen) gebruiken, werd een focusgroepsgesprek gevoerd met vijf deeltijdstudenten van de opleiding idm van De Haagse Hogeschool. Voorafgaand aan het focusgroepsgesprek werd een literatuurstudie uitgevoerd die input gaf voor de onderwerpen die in het gesprek aan de orde werden gesteld. Werving van de deelnemers geschiedde met behulp van een korte online enquête die werd toegestuurd aan alle deelnemers van een regulier onderwijsblok van de opleiding idm.

Uit de resultaten valt af te leiden dat de studenten de scoringsrubriek vooral gebruikten als een self-assessment instrument voordat ze de uitwerking van een studietaak ter beoordeling opstuurden maar ze pasten het principe van de scoringsrubriek ook toe in andere situaties, bijvoorbeeld tijdens hun werk. Gevraagd naar de mate waarin zij de scoringsrubriek waarden, geven ze aan dat ze vooral de feedback die ze op papier krijgen bijzonder op prijs stellen.

¹³ De inhoud van hoofdstuk 7 is eerder gepubliceerd in: How adult students in information studies use a scoring rubric for the development of their information literacy skills. *The Journal of Academic Librarianship*, 38 (3), 165-171 (Van Helvoort, 2012).

Inleiding

Scoringsrubrieken kunnen op veel verschillende manieren tijdens het leerproces worden ingezet. De kwantitatieve onderzoeken in hoofdstuk 5 en 6 beperkten zich echter tot de functie van de Scoringsrubriek voor Informatievaardigheden als *beoordelingsinstrument*, met vooral docenten in de rol van beoordelaar. Het feit dat geen aandacht werd besteed aan de wijze waarop het gebruik van de scoringsrubriek werd ervaren door studenten (bijvoorbeeld als instrument voor feedback, instructie of self-assessment) kan worden beschouwd als een beperking van de daar beschreven onderzoeken.

In het najaar van 2010 werd daarom een vervolgonderzoek onder studenten uitgevoerd dat in dit hoofdstuk wordt beschreven. Studenten die werden benaderd voor deelname aan het onderzoek studeerden allen aan de opleiding idm van De Haagse Hogeschool. Doel van het onderzoek was te achterhalen hoe de scoringsrubriek daadwerkelijk door studenten wordt gebruikt. Daarbij ging de interesse niet alleen uit naar het gebruik tijdens studieactiviteiten maar ook naar het gebruik van de scoringsrubriek bij het reguliere werk (idm Den Haag verzorgt ook een deeltijdopleiding op 2 avonden per week waaraan studenten deelnemen die overdag werken).

In het onderzoek werd zowel gebruik gemaakt van een online enquête als van een focusgroepsgeprek. De studenten die benaderd werden om de enquête in te vullen, hadden vooraf deelgenomen aan een workshop waarin de scoringsrubriek was behandeld en waarin ze ook één maal hadden geoefend met het gebruik van de scoringsrubriek in het kader van peer-assessment. De enquête was vooral bedoeld om snel wat meer feitelijke informatie in te winnen bij studenten die hadden deelgenomen aan het onderwijsblok Digital Library waarvan de workshop deel had uitgemaakt. Het focusgroepsgeprek werd gebruikt om uitgebreidere informatie te verkrijgen van de studenten die in de enquête hadden aangegeven dat ze wel wilden meewerken aan een vervolgonderzoek in de vorm van een interview. Vooral het focusgroepsgeprek was gericht op het inwinnen van gedetailleerde informatie over de manier waarop de studenten de scoringsrubriek op een later moment en vanuit zichzelf gebruiken.

Voorafgaand aan het empirisch onderzoek met behulp van de online enquête en het focusgroepsgeprek werd een literatuurstudie uitgevoerd naar het gebruik van scoringsrubrieken in het hoger onderwijs in het algemeen. De gegevens uit deze literatuurstudie werden gebruikt bij de formulering van de enquête- en interviewvragen.

7.1 Literatuur

Reddy en Andrade publiceerden een metastudie naar het gebruik van scoringsrubrieken in het hoger onderwijs (Reddy & Andrade, 2010). Op basis van de twintig studies die zij bespreken, onderscheiden ze de wijze waarop studenten het gebruik van scoringsrubrieken beleven en de ervaringen die docenten er mee hebben. Vervolgens gaan ze in op de positieve effecten die het gebruik van scoringsrubrieken heeft op het leren van studenten en bespreken ze het gebruik van scoringsrubrieken voor curriculumverbetering.

Volgens Reddy en Andrade waarderen studenten het gebruik van scoringsrubrieken om drie redenen (2010, p. 438):

- ze maken duidelijk wat de student aan het eind moet kunnen of kennen
- ze stellen de studenten in staat om hun eigen leerproces goed in te delen en te reguleren en
- ze zorgen voor eerlijke en transparante beoordelingen.

Het laatste punt dat Reddy en Andrade noemen (“eerlijke en transparante beoordelingen”) zegt meer iets over de redenen waarom studenten het gebruik van een scoringsrubriek op prijs stellen dan over de manier waarop zij zelf een scoringsrubriek gebruiken. Transparantie en eerlijkheid van de beoordelingen worden dan ook vooral door docenten genoemd als argumenten voor het gebruik van een scoringsrubriek. Wel is het zo dat duidelijkheid omtrent de beoordelingscriteria de studenten kan helpen in hun streven om zo hoog mogelijk te scoren bij de uitvoering van de taak. Doordat studenten goed weten wat van hen wordt verwacht, zijn ze ook in staat om te streven naar excellente prestaties.

In hun artikelen over het gebruik van een scoringsrubriek voor het beoordelen van informatievaardigheden, verwijzen ook Scharf, Elliot, Huey, Briller en Joshi (2007) en Oakleaf (2008) naar de rol die zo’n schema speelt bij het verhelderen van de assessmentcriteria. In beide publicaties wordt benadrukt dat het van belang is om de beoordelingscriteria vooraf met studenten te bespreken en daarbij ook in te gaan op de vraag hoe ze geïnterpreteerd dienen te worden. Dit helpt de studenten om in te schatten wat de docenten van hen verwachten (Oakleaf, 2008, p. 245; Scharf et al., 2007, p. 471).

In de studies van Bolton (2006, p. 6) en Andrade en Du (2005, p. 6) komen studenten zelf aan het woord. De antwoorden die de studenten geven in een enquête (Bolton, 2006) en in focusgroepen (Andrade & Du, 2005) bevestigen dat het gebruik van een scoringsrubriek en de toelichting daarop voorafgaand aan de uitreiking van een studieopdracht, hen minder onzeker maakt doordat ze weten wat er van hen wordt verwacht. In het bijzonder voor volwassenen die hun studie combineren met een reguliere baan (zoals deeltijdstudenten in het Nederlandse hbo) geldt volgens Bolton dat een scoringrubriek hen ook helpt om de relevantie van bepaalde leerdoelen te onderkennen. Dit is volgens hem belangrijk omdat juist deze studenten in het volwassenenonderwijs sterk geleid worden door een intrinsieke motivatie. Een scoringsrubriek helpt hen om de relevantie en de validiteit te onderkennen van de onderwerpen waarin ze les krijgen (Bolton, 2006, p. 5).

Het is belangrijk dat studenten vertrouwd raken met de eisen die aan de uitwerking van een opdracht worden gesteld, niet alleen omdat dat hen minder onzeker maakt en hen meer betreft bij de taken die ze geacht worden uit te voeren maar ook omdat het hen in de gelegenheid stelt om het werk dat aan een studieopdracht is verbonden beter te plannen (Andrade & Du, 2005, p. 4). Van de veertien undergraduate studenten die door hen werden ondervraagd in focusgroepen, antwoordden de meesten dat zij scoringsrubrieken inderdaad gebruiken om hun werkzaamheden en hun aanpak van een opdracht te plannen. De onderzoekers merkten daar echter wel bij op dat ze de indruk hadden dat studenten voor die planning en aanpak vooral de beschrijvingen van het excellente werk gebruikten en niet die van de valkuilen en veel gemaakte fouten. Dat is jammer want juist van dat laatste zouden de studenten ook veel kunnen leren (Andrade & Du, 2005, pp. 6-7).

Bolton (2006, p. 5) beschrijft dat studenten een scoringsrubriek ook gebruiken om hun eigen werk te beoordelen voordat ze het ter beoordeling toesturen aan een docent. In deze gevallen gaat het dus om een soort van 'self-assessment' dat studenten uit zichzelf toepassen zonder dat daarvoor een instructie is gegeven. Verder rapporteert hij dat studenten deze strategie ook gebruiken voor de verbetering van hun resultaten voor studieopdrachten die elkaar opvolgen (2006, p. 6). Volgens Andrade en Du gebruikten de studenten die zij interviewden een vergelijkbare methode van self-assessment met behulp van scoringsrubrieken, door van hun peers, docenten en onderwijs-assistenten feedback te vragen op concept-uitwerkingen van hun papers en daar vervolgens op te reflecteren (Andrade & Du, 2005, p. 5). Het verkrijgen van feedback is voor studenten sowieso een van de belangrijkste functies van een scoringsrubriek (Reddy & Andrade, 2010, p. 437). Daarbij gaat het niet alleen om een reactie op fouten of zaken die verbeterd kunnen worden maar ook om bevestiging van zaken die goed zijn aangepakt zodat studenten weten dat ze het een volgende keer op de zelfde manier kunnen doen (Andrade & Du, 2005, p. 5).

De literatuur geeft aan dat scoringsrubrieken ook door studenten op verschillende manieren gebruikt kunnen worden. Hoewel meestal de nadruk ligt op de mate waarin scoringsrubrieken door studenten worden 'beleefd' of gewaardeerd, kunnen ook conclusies worden getrokken over de manier waarop zij ze daadwerkelijk gebruiken:

- a. een scoringsrubriek maakt duidelijk wat een docent van de studenten verwacht, wat er weer voor zorgt dat bij hen onzekerheid wordt weggenomen en dat zij in staat worden gesteld om zich goed voor te bereiden en vervolgens excellent te presteren;
- b. scoringsrubrieken worden door studenten gebruikt om de activiteiten die ze voor een studieopdracht moeten doen, te plannen;
- c. studenten gebruiken scoringsrubrieken om te reflecteren op het belang van de leerinhoud voor het functioneren in de beroepspraktijk, wat hen helpt om betrokken te raken bij de leerdoelen van de betreffende cursus en de gedoeerde onderwerpen.

Scoringsrubrieken maken studenten niet alleen duidelijk wat er van hen verwacht wordt (zie a.) ze worden door studenten ook (en vooral) gebruikt voor het ontvangen of generen van verschillende vormen van feedback. In het artikel van Reddy en Andrade worden de volgende soorten van feedback genoemd:

- d. studenten beoordelen zelf hun eigen werk met een scoringsrubriek en genereren op die manier dus hun eigen feedback voordat ze het werk ter beoordeling naar de docent sturen ('self-assessment');
- e. de feedback die studenten ontvangen van hun beoordelaars wordt door hen gebruikt om na te gaan welke (deel)taken ze inmiddels beheersen (positieve feedback die hen doet concluderen dat ze iets goed hebben aangepakt en dat ze het een volgende keer op dezelfde manier zullen doen) én op welke deeltaken ze zich nog dienen te verbeteren;
- f. scoringsrubrieken worden door studenten gebruikt om informele feedback te krijgen van hun medestudenten, onderwijsassistenten en anderen ('peer-assessment');
- g. en studenten gebruiken scoringsrubrieken om hun eigen progressie in de uitvoering van een bepaald type taak of activiteit te monitoren aan de hand van opeenvolgende studieopdrachten.

Deze samenvatting van de verschillende manieren waarop studenten een scoringsrubriek (kunnen) gebruiken, werd gehanteerd als uitgangspunt voor het focusgroeps gesprek dat wordt beschreven in paragraaf 7.3.

7.2 Enquête

Met behulp van een online vragenlijst die via e-mail werd toegestuurd (bijlage 7.1) werden data verzameld onder een kleine groep studenten van de opleiding idm van De Haagse Hogeschool. Deze opleiding wordt verzorgd in een voltijd- en een deeltijd-variant. De enquête werd toegestuurd aan 21 studenten die in het schooljaar 2009-2010 hadden deelgenomen aan een van de twee varianten (10 voltijdstudenten en 11 deeltijdstudenten). Het ging in deze enquête dus om kleine aantallen, wat werd veroorzaakt door het feit dat er de voorafgaande jaren überhaupt weinig inschrijvingen waren voor deze opleiding én door het feit dat het onderzoek werd afgenomen onder studenten die hadden deelgenomen aan het onderwijsblok Digital Library, het blok waarin de scoringsrubriek voor het eerst was gebruikt. Het betreffende onderwijsblok was onderdeel van de laatste hoofdfase van de opleiding, een fase waarin al een behoorlijk aantal studenten de studie heeft gestaakt.

In het onderwijsblok Digital Library is de scoringsrubriek door studenten gebruikt in een workshop waarin ze voor het eerst werd uitgedeeld en besproken. Na een introductie en een toelichting deden de studenten een eerste oefening met de scoringsrubriek die eruit bestond dat ze een voorbeeldessay van een van hun groepsgenoten beoordeelden. Tijdens de workshop beoordeelden alle deelnemers hetzelfde voorbeeldessay. De beoordeling werd vervolgens klassikaal besproken en de workshop werd afgesloten met het uitdelen van de beoordelingen van hun eigen essays (die voorafgaand aan de workshop aan de docent waren toegestuurd) en de gelegenheid om vragen te stellen naar aanleiding van die beoordelingen.

Alle studenten aan wie in het najaar van 2010 de enquête was toegestuurd hadden in de periode februari-april 2010 (deeltijd) of mei-juli 2010 (voltijd) deelgenomen aan het onderwijsblok waarvan deze workshop deel uitmaakte én hadden deelgenomen aan de afsluitende opdrachten en toetsen van het onderwijsblok. Vier van de 10 voltijdstudenten en 6 van de 11 deeltijdstudenten hebben de enquête ingevuld en teruggestuurd. De leeftijd van de voltijdstudenten die hebben gereageerd varieerde van 21 tot 23 jaar, de deeltijdstudenten waren tussen de 31 en 48 jaar oud. Van de 10 respondenten op de enquête had er 1 niet deelgenomen aan de workshop. Deze student van de voltijdgroep gaf aan dat ze de scoringsrubriek wel in een later stadium van het onderwijsblok had gebruikt.

Tijdens de workshops waarin de scoringsrubriek werd geïntroduceerd zijn de studenten gestimuleerd om het scoringsschema te blijven gebruiken voor zelfbeoordelingen en beoordelingen van elkaars werk ('self- en peer-assessment'). Met de online enquête werd geprobeerd te achterhalen welke studenten de scoringsrubriek na afloop van de workshop inderdaad nog minstens één keer hadden gebruikt. Twee van de vier respondenten uit de voltijd groep antwoordden dat ze het schema na afloop van de workshop nog eens hadden opgespoord. Eén van hen was de student die had aangegeven dat ze de workshop zelf niet had bijgewoond. Van de deeltijdgroep gaven vijf van de zes

respondenten aan dat ze de scoringsrubriek in een later stadium nog eens hadden gebruikt. Tabel 7.1 geeft een samenvatting van de antwoorden op de enquêtevraag nummer 2, zowel voor de voltijd als voor de deeltijd studenten.

Tabel 7.1

Antwoorden van voltijd- en deeltijdstudenten idm op enquêtevraag 2

	Na die ene workshop waarin we de scoringsrubriek hebben gebruikt ...		Totaal
Welke variant van de opleiding idm volg je?	heb ik haar zelf nog minstens één keer opgezocht om haar opnieuw te gebruiken	heb ik er niets meer mee gedaan	
Voltijd	2	2	4
Deeltijd	5	1	6

Geen van de respondenten uit de voltijdgroep was bereid om deel te nemen aan een vervolgin-
terview. De vijf deeltijdstudenten die hadden geantwoord dat ze de scoringsrubriek in een later
stadium nog eens hadden gebruikt, waren allen wel bereid om aan een interview deel te nemen.
De enquête zelf en alle gegeven antwoorden worden weergegeven in bijlage 7.1 bij dit hoofdstuk.

Hoewel het aantal respondenten klein was, kan uit het feit dat vijf deeltijdstudenten en niet één van
de voltijdstudenten bereid was om deel te nemen aan een vervolgonderzoek, worden afgeleid dat
deeltijdstudenten beter gemotiveerd zijn om de scoringsrubriek te gebruiken dan voltijdstuden-
ten. Hun bereidheid om in hun eigen tijd deel te nemen aan een vervolgin-terview kan echter ook
worden uitgelegd als een bevestiging van het feit dat deeltijdstudenten überhaupt meer intrinsiek
gemotiveerd zijn voor hun studie dan de jongere voltijdstudenten, wat ook gesuggereerd wordt
door Bolton (2006).

7.3 Focusgroepsgesprek

In de oorspronkelijke opzet van het onderzoek zou de enquête worden gevolgd door een aantal indi-
viduele interviews, waarin zou worden nagegaan in hoeverre het gebruik van ‘rubrics’ zoals beschre-
ven in de literatuur ook van toepassing is voor de Scoringsrubriek voor Informatievaardigheden.
Toen uit de enquête bleek dat alleen deeltijdstudenten bereid waren om deel te nemen aan een
interview, werd er gekozen voor het gebruik van één focusgroepsgesprek met vijf studenten
tegelijkertijd.

Radcliff, Jensen, Salem Jr., Burhanna en Gedeon definiëren een ‘focus group’ als een “face-to-face
interview with a group of six to twelve people that focuses on a single subject or topic” (2007, p. 73).
Focusgroepsgesprekken zijn als methode vooral geschikt als het onderzoek zich richt op het verkrij-
gen van uitgebreide informatie over een bepaald onderwerp. Doordat het gesprek wordt gevoerd
met een betrekkelijk kleine groep deelnemers die een gemeenschappelijke belangstelling hebben,

kan het onderwerp in een focusgroeps gesprek diepgaand worden bediscussieerd en kunnen de gesprekspartners elkaar aanvullen (zie onder andere Shoaf, 2003, p. 124 en Radcliff et al., 2007, p. 73). De interactie in een focusgroeps gesprek kan echter ook ongewenste effecten opleveren. Te denken valt daarbij aan dominante personen die meer ingetogen deelnemers belemmeren in hun deelname aan het gesprek (Ketelaar, Hentenaar en Kooter, 2011, p. 27). Een goede moderator zou dit effect echter moeten kunnen beperken.

De studenten die hadden aangegeven dat ze wilden meedoen aan een vervolginterview konden worden beschouwd als zo'n 'focusgroep' omdat het in alle gevallen ging om deelnemers aan de deeltijd avondstudie. Dat betekent dat zij ouder zijn dan de voltijdstudenten en dat ze hun studie zelf financieren en combineren met een reguliere baan. Hoewel de groep van vijf studenten misschien wat klein was - Radcliff et al. geven zoals eerder gezegd aan dat een focusgroep zou moeten bestaan uit zes tot twaalf personen en Shoaf (2003, p. 127) noemt vijf deelnemers als het minimum - werd het risico toch genomen en werd het aanvankelijke plan om individuele vervolginterviews af te nemen gewijzigd in de organisatie van één gezamenlijk focusgroeps gesprek. Dat was niet alleen aantrekkelijk om redenen van efficiency maar had ook het bijkomende voordeel dat het studenten de gelegenheid zou geven om hun meningen daadwerkelijk uit te wisselen en op die manier tot beter onderbouwde opinies te komen.

De mogelijke groepseffecten in een focusgroeps gesprek werden in dit geval bovendien niet louter negatief ingeschat. Natuurlijk lag het gevaar op de loer van 'onderonsjes' omdat de deelnemers aan het gesprek elkaar al langer kenden. Ook zou een wat onzekere student geneigd kunnen zijn om zich te schikken naar de meerderheid van de groep. Aan de andere kant had de vorm van een groeps gesprek het voordeel dat de deelnemers zich in zo'n groep wellicht toch wat vrijer zouden durven uitlaten dan in een individueel gesprek met de onderzoeker die ook nog eens hun eigen docent was. In het focusgroeps gesprek kon de nadruk dan ook meer op de algehele opinie van de groep worden gelegd dan op die van het individu (zie ook Ketelaar, Hentenaar en Kooter, 2011, pp. 31-32). Bovendien bood de organisatie van één groeps gesprek de kans tot het aantrekken van een ervaren neutrale gespreksleider, een mogelijkheid die bij afzonderlijke gesprekken veel moeilijker te realiseren was geweest. De persoon die in dit geval als moderator werd gevraagd, was een van de lectoren van De Haagse Hogeschool. Deze lector was weliswaar bekend met het fenomeen informatievaardigheid en haar verschillende aspecten maar was niet betrokken geweest bij de constructie van de scoringsrubriek. Vanwege de beschikbaarheid van deze moderator hoefde de onderzoeker zelf niet deel te nemen aan het gesprek. Uiteindelijk was hij dan ook slechts als observator aanwezig, maakte aantekeningen en bediende een memorecorder voor de opname van het gesprek. Het gevaar van ongewenste beïnvloeding van het proces werd op die manier tot het minimum beperkt.

De eerste stap die gezet werd bij de organisatie van de focusgroep was de uitnodiging van de studenten zelf. Ze ontvingen een e-mail met een aantal voorstellen voor plaats en tijd. Uiteindelijk werd gekozen voor een tijdstip om 6 uur in de avond in een gewoon klaslokaal. Dit tijdstip werd aan de studenten bevestigd en ook pas op dat moment werd de moderator benaderd met het verzoek om zijn bijdrage. Zowel moderator als studenten ontvingen 10 dagen voor de afspraak een herin-

nering. In de herinneringsmail werd ook aangegeven dat er een maaltijd beschikbaar zou zijn en werd zowel het doel van het focusgroeps gesprek als dat van het volledige onderzoek nog eens kort toegelicht. Bovendien werd aan de studenten gevraagd om vooraf na te denken over het antwoord op de vraag uit de eerste ronde van het focusgroeps gesprek, de vraag hoe en met welk doel hij of zij de scoringsrubriek voor informatievaardigheden concreet (heeft) gebruikt. Als opfrisser werd met de herinneringsmail een artikel over de scoringsrubriek uit het blad *InformatieProfessional* meegestuurd (Van Helvoort, 2009). Aan de moderator werd een korte uitleg van het onderzoek toegestuurd met een script voor het focusgroeps gesprek.

Het gesprek zelf werd georganiseerd in twee ‘rondes’ waarin verschillende thema’s aan de orde kwamen, zoals peer-assessment, self-assessment, transfer van vaardigheden, feedback van de docent en de scoringsrubriek als planningstool. De vragen waren gebaseerd op het eerder toegestuurde script maar de moderator heeft er wel een eigen kleur aan gegeven. De eerste ronde van het interview richtte zich op het daadwerkelijke gebruik van de scoringsrubriek.

De openingsvraag van de moderator luidde als volgt:

- Kun je zo concreet mogelijk beschrijven in welke situatie je de scoringsrubriek gebruikt hebt?

Vervolg vragen waren:

- Was dat in het kader van je studie of in het kader van je reguliere werk?
- Waren er ook andere mensen bij betrokken?
- Met welk doel gebruikte je de scoringsrubriek (bijvoorbeeld om haar uit te leggen aan anderen, om je eigen werk te beoordelen of voor het beoordelen van het werk van een van je groepsgenoten)?
- Hoe zinvol was het gebruik van de scoringsrubriek in die situatie volgens jou?

De tweede ronde van het gesprek ging over mogelijke alternatieve manieren om de scoringsrubriek te gebruiken, manieren die in de eerste ronde van het gesprek nog niet aan de orde waren geweest. De moderator deelde een lijst uit met mogelijke opties die gebaseerd was op het literatuuronderzoek dat in paragraaf 7.1 is besproken. Aan de studenten werd vervolgens gevraagd welke van de daar genoemde mogelijkheden volgens hen realistisch waren en van toepassing op de Scoringsrubriek voor Informatievaardigheden.

Resultaten van het focusgroeps gesprek

Deze samenvatting van het focusgroeps gesprek is gemaakt op basis van een inhoudelijke analyse van de letterlijke tekst van het focusgroeps gesprek met de commentaarfunctie van MS Word. Aan de hand van de analyse zijn uiteindelijk negen thema’s onderscheiden, vier thema’s in de eerste gespreksronde en vijf thema’s in de tweede. Bij de inhoudelijke weergave van de thema’s wordt de mening van de groep als geheel geïllustreerd met kenmerkende uitspraken van individuele studenten. Daarnaast worden eventuele afwijkende standpunten rond het betreffende thema weergegeven. Om redenen van vertrouwelijkheid zijn de namen van de betreffende studenten gewijzigd. Aan het gesprek namen drie mannelijke studenten deel in de leeftijd van 31, 36 en 39 jaar en twee vrouwen van 32 respectievelijk 34 jaar oud.

7.3.1 Daadwerkelijk gebruik van de scoringsrubriek

Thema 1: Self-assessment van het schoolwerk

Alle vijf de studenten gaven aan dat ze hun eindopdracht voor het onderdeel Digital Library (een visiedocument met betrekking tot een digitale bibliotheek voor een zelf te kiezen organisatie) met de scoringsrubriek hadden geëvalueerd voordat ze het ter beoordeling hadden ingezonden. Erik gebruikte de scoringsrubriek zelfs als vervanging voor feedback van de docent zelf: “Doordat ik geen feedback had gekregen van de docent omdat ik te laat was met inleveren van mijn concepten, heb ik hem gebruikt om er mezelf feedback mee te geven, om te zien of mijn document wel aan de eisen zou voldoen. En dusdanig steeds aangepast tot ik dacht, ja, ik heb nu wel een behoorlijke score en toen heb ik ingeleverd. Daarna heb ik hem eigenlijk niet meer gebruikt want sindsdien heb ik geen onderwijs meer gevolgd.”

Andere studenten (Susanne, Astrid, William) gebruikten hem niet alleen in het blok Digital Library maar ook in onderwijsblokken die ze daarna nog volgden. Susanne: “In het blok Digital Library gebruikte ik hem als een checklist. Ik heb hem er naast gehouden en de punten zeg maar afgelopen. En dat heeft blijkbaar geholpen want ik had relatief een hoog cijfer. Ik denk dat dat ook is omdat ik mezelf hier steeds op getoetst heb. In ieder geval heb ik goed gelet op de literatuurlijst en dat soort dingen... Voor de rest heb ik hem eigenlijk in het algemeen gebruikt. Ik heb hem er niet echt bijgehouden maar er wel aan teruggedacht, eigenlijk iedere keer als ik iets van een verslag moest maken. Voor school of voor mijn werk, in wat voor zin dan ook.”

William beaamde dat de scoringsrubriek hem had geholpen om zijn schrijfproducten te verbeteren maar hij vond ook dat de criteria zelf wel van een heel verschillend niveau waren: “Bij het ene is het gewoon een checklistje, en bij de andere is het veel meer kwalitatief, weet je wel. Ook heel inhoudelijk. Ik vind eigenlijk dat bijvoorbeeld criterium 2 [bronnenoverzicht / literatuurlijst] en 4 [bronverwijzingen in de tekst], die staan niet in verhouding tot 3 [kwaliteit van de gebruikte primaire informatiebronnen] en al helemaal niet tot 1 [oriëntatie op het onderwerp] en 5 [verwerking tot nieuwe kennis en / of inzichten]. Dat blijkt ook wel uit de rechter kolom met de scores waar er verschil is tussen 10 of 20 punten per criterium, maar in praktijk zie je dat toch wel snel over het hoofd.” Een andere opmerking die William maakte was dat vooral voor criterium 1 en 5 (de meer cognitieve criteria) geldt dat je toch wel enigszins ingevoerd moet zijn in de materie om überhaupt een oordeel over je eigen werk te kunnen uitspreken: “Het gaat niet alleen over informatievaardigheden sec, het gaat ook over je inhoudelijke beheersing van het vakgebied. Als je die niet hebt, weet je immers niet welke vragen je vergeet te stellen”. William benoemde hier een interessant punt want het impliceert dat, in ieder geval volgens hem, de competentie informatievaardigheid niet geheel los kan worden gezien van expertise op het betreffende domein.

Thema 2: Transfer van het gebruik van de scoringsrubriek naar andere (werk)situaties

Voor de deeltijdstudenten uit de focusgroep gold dat ze de scoringsrubriek niet alleen gebruikten voor school maar ook voor hun dagelijkse werk, met name als ze werkzaam waren in een kennis- of informatierijke organisatie. Met uitzondering van één student die als treinconducteur bij de Nederlandse Spoorwegen werkt en voor zijn reguliere werk niet zo heel veel te maken heeft met informatieverwerking, gaven de gespreksdeelnemers aan dat ze de criteria van de scoringsrubriek

voor informatievaardigheden ook in hun achterhoofd hadden als ze in hun gewone baan aan een rapport of een verslag werkten. Sander lichtte dat als volgt toe: “Het is inderdaad wel een heel goede stok achter de deur om aan een minimale kritische waarde te komen voor je stuk of je opdracht. Je wordt gedwongen om er naar te kijken en dat is wel de meerwaarde”. Hij gaf echter ook aan dat men bij hem op het werk al gewend was om te werken met dit soort lijstjes, zelfs met checklists die nog veel uitgebreider zijn. Jammer genoeg kon hij hier niet verder op ingaan omdat zijn werkgever dat beschouwt als vertrouwelijke informatie die niet met anderen mag worden gedeeld.

Astrid gaf aan dat ze sinds de kennismaking met de scoringsrubriek veel meer gespitst is op bronvermelding dan voor die tijd. “Ik beheer zelf een website waarbij ik gewend was om plaatjes van andere sites gewoon te gebruiken op onze eigen site. Nu heb ik alle plaatjes die ik heb gebruikt voorzien van een bronvermelding. En, ja dat is wel grappig, nu zijn we er opeens op aangesproken want we hebben twee plaatjes moeten verwijderen.”

Het is het vermelden waard dat alle studenten die deelnamen aan het gesprek het erover eens waren dat de scoringsrubriek hen geholpen heeft om kritischer te worden op hun eigen (schrijf)werk, zowel dat voor school als dat voor hun reguliere werk of privé-activiteiten. Het voorbeeld van Astrid met betrekking tot het beheer van de website maakt echter ook duidelijk dat je het voor jezelf niet altijd gemakkelijker maakt met zo’n kritische houding.

Thema 3: Kritische beoordeling van het werk van anderen, bijvoorbeeld dat van collega’s op het werk
De studenten vertelden ook dat ze, sinds hun eigen werk was beoordeeld met de scoringsrubriek, kritischer waren geworden op het werk van anderen. Volgens hen was het vooral de nadruk die in de scoringsrubriek wordt gelegd op de onderbouwing van beweringen met authentieke, relevante en betrouwbare informatiebronnen (de criteria 2, 3 en 4) die hen heeft geholpen bij het ontwikkelen van een academische en kritische houding naar andermans werk. Zo vertelde Astrid dat ze sindsdien een veel kritischer lezer is: “Ja ik let er nu veel meer op of iemand wel heeft onderbouwd wat hij beweert. Steeds vaker denk ik: is dat wel zo, waar staat de bron?” Ook William merkt wel dat hij kritischer is geworden maar hij weet niet of dat te maken heeft met zijn ervaringen met de scoringsrubriek of gewoon door het feit dat hij studeert aan een hbo-opleiding. “Die kritische houding” zegt hij “zit wel in mijn ‘systeem’ maar ja je doet dan ook al drie jaar deze studie idm”.

Thema 4: De ontwikkeling van een kritische houding naar opdrachtgevers

Dit was misschien wel de meest onverwachte uitkomst van het focusgroeps gesprek: de problemen die studenten hebben met het formuleren van goede onderzoeksvragen.

Studenten van de opleiding idm (maar niet alleen die van deze opleiding) werken tijdens hun studie vaak aan projecten waarbij een van de docenten fungeert als opdrachtgever. Later in de studie werken ze ook vaak voor echte opdrachtgevers. De deeltijdstudenten die deelnamen aan het focusgroeps gesprek hebben daarnaast vaak te maken met projecten in het kader van hun reguliere baan. Projecten die idm-studenten uitvoeren voor hun studie liggen vaak op het gebied van het inrichten van een archief- of documentatiesysteem maar ook opdrachten met betrekking tot literatuuronderzoek en –rapportage komen voor.

De scoringsrubriek werd geïntroduceerd tijdens de feedback op een studieopdracht om een kort essay te schrijven. In de betreffende workshop ervoeren de studenten aan den lijve hoe belangrijk het is om op zoek te gaan naar de vraag achter de oorspronkelijke vraag en er niet klakkeloos van uit te gaan dat een opdrachtgever of vraagsteller precies weet wat hij wil hebben. Voor al de studenten die deelnamen aan het focusgroepsgebesprek gold dat ze nu herkenden dat die herformulering van de opdracht onderdeel dient te zijn van de oriëntatiefase (criterium 1 van de scoringsrubriek). Voor Susanne was dat echter iets dat ze pas in die betreffende workshop had geleerd: “Wat voor mij nieuw was, was die oriëntatie op het onderwerp. Dat je dus de vraagstelling vrij moet interpreteren. [...] Tot het moment dat we die oefening gingen doen met die scoringsrubriek was ik er op gebrand de vraag zo goed mogelijk te lezen en zo goed mogelijk antwoord te geven op de vraag en juist helemaal niet af te wijken van de vraagstelling. Dat is niet echt iets wat ik nu gebruik maar iets wat me een heel nieuwe kijk heeft gegeven op dit soort werk.”

Anderen benadrukten dat grondige kennis van een onderwerp niet voldoende is om goede vragen te kunnen stellen maar dat het óók om een bepaalde houding gaat, de juiste communicatieve vaardigheden en het lef om kritisch te zijn. Toen Sander vroeg wie er echt kritische vragen aan zijn collega's durft te stellen, drukte William dat als volgt uit: “Ik ga niet tegen mijn directeur zeggen dat hij er geen snars van begrijpt. Maar tegen mijn collega's durf ik dat wel.”

Al met al waren alle gespreksdeelnemers het er over eens dat het gebruik van de scoringsrubriek tijdens de workshop en de toelichting erbij hen had overtuigd van de noodzaak om de formulering van een (studie)opdracht kritisch te beschouwen en zo nodig aan te passen.

7.3.2 Potentieel gebruik van de scoringsrubriek

Tijdens de tweede ronde van het focusgroepsgebesprek ging de aandacht niet zozeer uit naar de manier waarop de studenten de scoringsrubriek daadwerkelijk hadden gebruikt maar meer naar hun oordeel over de potenties ervan. Daarbij werd ook ingegaan op de beperkingen die de scoringsrubriek volgens de studenten had. De resultaten van de tweede gespreksronde worden hieronder opnieuw rond thema's gepresenteerd (thema 5 tot en met 9).

Thema 5: Feedback van de docent

Alle vijf de studenten beaamden dat feedback een van de belangrijkste functies is van de scoringsrubriek voor informatievaardigheden. Erik merkte echter op dat een ingevuld scoringsformulier niet als vervanging kan dienen voor de face-to-face feedback van de docent. Hij zou naast het ingevulde scoringsformulier ook nog graag een mondelinge toelichting krijgen bij de toegekende scores en de aanvullende opmerkingen van de beoordelaar of de docent zelf. “Ik denk dat die feedback een heel goede toepassing ervan is. Ik neem wel aan dat je er dan daarna ook nog samen over communiceert.” Deze toevoeging van Erik was opvallend want hij was ook degene die in het begin van het gesprek had aangegeven dat hij de scoringsrubriek had gebruikt om zijn eigen werk te beoordelen omdat hij te laat was geweest met het insturen van zijn concept versies. De ervaring dat hij in staat was om de opdracht zonder de feedback van de docent toch naar behoren uit te voeren en af te ronden, had hem er blijkbaar niet van overtuigd dat de mondelinge feedback van de docent overbodig was.

Alle gesprekspartners gaven overigens aan dat de schriftelijke feedback met de scoringsrubriek vooral zinvol is omdat die zo goed is toegespitst op de zeven afzonderlijke criteria én omdat je hem op papier krijgt [...] “zodat je hem later nog eens rustig kunt nalezen” (Astrid). Ze waren het echter ook zonder uitzondering eens met Erik over het feit dat de feedback op het formulier aanvullend zou moeten zijn op de face-to-face feedback die ze gewend zijn te krijgen en dat hij deze laatste niet kan vervangen.

Thema 6: Beoordeling van elkaars werk (peer-assessment)

Anders dan de onderzoeker vooraf had verwacht gaven de studenten aan dat ze de scoringsrubriek niet hadden gebruikt om elkaars werk te beoordelen. De reden daarvoor was volgens hen dat de gelegenheid zich daarvoor niet had aangediend in het blok Digital Library (dit was een blok met vooral individuele opdrachten) of een van de andere onderwijsblokken die ze daarna volgden. Natuurlijk hadden ze zelf wel afspraken kunnen maken om elkaars werk te reviewen maar de tijdsdruk stond dat niet toe. Aan de andere kant gaven de respondenten aan dat ze commentaar op elkaars werk over het algemeen zeer waardeerden en dat deze voor hen even zinvol is als de feedback die ze krijgen van een docent. William: “Er wordt nu onderscheid gemaakt tussen jezelf reviewen, peer-review en de docent. Maar volgens mij is dat onderscheid in ieder geval bij deze scoringsrubriek niet zo relevant. Ze is gewoon bruikbaar voor feedback en je kan dat zelf doen, door een collega die er een beetje in thuis is en door een docent.” Erik was dat wel met hem eens maar maakte er de kanttekening bij dat degene die de review doet in ieder geval zelf ook wel voldoende informatievaardig moet zijn.

Voor de onderzoeker was ook deze conclusie van de respondenten nogal verrassend. Hij had verwacht dat er meer waarde zou worden gehecht aan de feedback van de docent dan aan die van medestudenten aangezien de docent aan het einde van het project ook de beoordelaar van het werk is. De antwoorden van de studenten bij dit thema ‘peer-assessment’ kunnen geïnterpreteerd worden als een tweede aanwijzing voor het feit dat volwassen deeltijdstudenten relatief hoog gemotiveerd zijn voor hun studie zoals eerder gesuggereerd in het literatuuroverzicht in paragraaf 7.1 (Bolton, 2006).

Thema 7: De scoringsrubriek is geen planningstool

Het literatuuroverzicht in paragraaf 7.1 suggereerde dat scoringsrubrieken voor studenten ook hun nut bewijzen bij het plannen van de activiteiten die voor een bepaalde studieopdracht nodig zijn. De studenten uit de focusgroep meenden echter unaniem dat de scoringsrubriek voor informatievaardigheden daarvoor niet geschikt is. In deze rubriek ligt de nadruk immers op de kenmerken van het eindproduct en niet op de manier waarop het product tot stand komt. Om die zelfde reden twijfelden ze over het nut van de scoringsrubriek bij het geven van instructie. Deze mening werd door William als volgt verwoord: “Als scoringsrubriek is ze denk ik wel goed maar als je het op andere plekken ook wil gebruiken dan wil je ook meer informatie, zoals: hoe moet ik het dan wel doen?”

Thema 8: De scoringsrubriek is niet het instrument bij uitstek om mensen te overtuigen van het belang van informatievaardigheden

Ook het gebruik van een scoringsrubriek voor het inzichtelijk maken van de validiteit en de relevantie van de leerstof, een toepassing ervan die volgens Bolton vooral bij het leren van volwassenen effect heeft en hen helpt om gemotiveerd te blijven voor de studie (Bolton, 2006, p. 6), werd door de gespreksdeelnemers niet herkend. De studenten gaven aan dat de scoringsrubriek voor informatievaardigheden hen wel had geholpen om het belang van correct informatiegebruik te leren inzien (zoals correct citaatgedrag en het doen van onderzoek naar bestaande informatie voordat je zelf een oplossing bedenkt voor een probleem dat moet worden opgelost) maar ze zagen dat niet als een doel waarvoor de scoringsrubriek was gemaakt. Sander vatte hun opinie als volgt samen: “Er wordt gesuggereerd dat de scoringsrubriek een brug slaat tussen de leerstof en de beroepspraktijk. Dat is niet onwaar. Maar we zijn er niet van overtuigd dat juist de scoringrubriek daarvoor het beste middel is.”

Thema 9: Monitoren van progressie in informatievaardigheden

Tijdens de discussie over het eerste thema (‘self-assessment van het schoolwerk’) werd door de studenten verteld dat ze de scoringsrubriek voor informatievaardigheden niet alleen hadden gebruikt in het onderwijsblok Digital Library waar zij was geïntroduceerd maar ook in onderwijsblokken die ze daarna nog hadden gevolgd. Opvallend was echter dat ze aangaven de scoringsrubriek niet te gebruiken om te controleren of ze ook beter werden in hun vaardigheden op informatiegebied maar eigenlijk alleen als een checklist om ervoor te zorgen dat ze een goede beoordeling zouden behalen voor de opdracht waar ze op dat moment aan werkten. Aan het einde van het focusgroepsgesprek gaven de studenten aan dat de scoringsrubriek ook wel gebruikt zou kunnen worden voor het meten van de eigen progressie maar geen van hen vond dat echt zinvol. Sander was daarin zeer uitgesproken: “Je weet van jezelf dat je ergens zwak in bent. Wat is de meerwaarde om dat expliciet te maken?”

7.4 Conclusies en discussie

De studenten die aan het focusgroepsgesprek hebben deelgenomen, gaven aan dat ze de scoringsrubriek hadden gebruikt voor self-assessment gedurende het onderwijsblok waarin de rubriek tijdens een workshop was geïntroduceerd. De scoringsrubriek stelde hen in de gelegenheid om de uitwerking van hun eigen studieopdracht goed te evalueren voordat ze hem instuurden voor een beoordeling. Vier van de deelnemers aan het gesprek vertelden dat ze de scoringsrubriek ook gebruikt hadden in latere onderwijsblokken maar dat ze haar er dan niet daadwerkelijk naast gehouden hadden. Wel hadden ze de criteria in hun achterhoofd als ze nog een laatste check deden op hun schoolwerk voordat ze het instuurden. Dit laatste is een indicatie dat zij dat evaluatieproces hebben geïntegreerd in hun werk dat ze voor school uitvoeren.

Dezelfde vier studenten gaven aan dat ze de criteria van de scoringsrubriek, nadat ze er in het onderwijs dat ze volgden mee bekend waren geraakt, ook gebruikten als ze werk gerelateerde rapporten of notities schreven. Bovendien vertelden ze dat ze kritischer geworden waren op het werk van anderen, bijvoorbeeld hun collega's. De scoringsrubriek bleek dus overdraagbaar te zijn van

schoolse situaties naar de beroepspraktijk. Deze overdraagbaarheid van de scoringsrubriek was voor de onderzoeker die de rubriek had ontwikkeld een van de meest informatieve uitkomsten van het focusgroepsgesprek.

Daar staat tegenover dat uit de enquête - die aan het focusgroepsgesprek vooraf ging - blijkt dat gebruik van de scoringsrubriek bij uitvoering van anderen taken dan die waarbij ze werd geïntroduceerd, niet voor alle studenten vanzelfsprekend is. Drie van de tien respondenten gaven aan dat ze haar niet meer hadden gebruikt. Dit gegeven is van belang bij de implementatie van de Scoringsrubriek voor Informatievaardigheden in een curriculum. Het geeft immers aan dat er op verschillende momenten aandacht aan de scoringsrubriek besteed zou moeten worden als je wilt dat studenten het gebruik ervan onderdeel maken van hun studiehouding.

Hoewel de rubriek slechts één keer in een workshop was gebruikt, waren alle deelnemende studenten het erover eens dat een ingevulde scoringsrubriek een uitstekend instrument zou zijn voor het geven van feedback door de docent, vooral omdat je de feedback dan op schrift krijgt en hem later nog eens kan nalezen. Ze vonden echter ook dat de feedback via zo'n scoringsformulier de mondelinge feedback die ze gewoon zijn om van hun docent te krijgen, niet kan vervangen.

Bovenstaande uitkomsten van het focusgroepsgesprek bevestigen het nut van de scoringsrubriek voor de ontwikkeling van informatievaardigheden. De opmerkingen van de studenten zijn echter ook aanleiding voor kritische kanttekeningen bij de rubriek zelf en de manier waarop zij door de docent / onderzoeker in het onderwijs werd toegepast.

Eén van de studenten merkte tijdens de bespreking van thema 1 (self-assessment van het schoolwerk) op dat voor sommige van de criteria geldt dat voor de hogere niveaus vaardigheden op informatiegebied niet voldoende zijn maar dat daarvoor ook beheersing van het vakgebied zelf nodig is. Deze verwevenheid van informatievaardigheid met het kennisdomein zelf werd ook opgemerkt in hoofdstuk 5 bij de casus van de Open Universiteit. Daar bleek dat de beoordelaar die zelf niet werkzaam was in het vakgebied van de betreffende studentenproducten relatief gezien meer moeite had om de beoordelingsinstrumenten consistent toe te passen dan de beoordelaar die wel goed thuis was in het betreffende vakgebied.

Volgens alle deelnemers aan het focusgroepsgesprek (dus inclusief de 'neutrale' gespreksleider) had de scoringsrubriek één belangrijke beperking: zij geeft een beschrijving van de kenmerken van het gewenste niveau van informatievaardigheid (respectievelijk de veel voorkomende fouten bij het ongewenste niveau) maar het blijft onduidelijk hoe je het gewenste niveau van informatievaardigheid kunt bereiken. Deze tekortkoming van de scoringsrubriek wordt onderkend door de onderzoeker maar is een logisch gevolg van de beslissing om de scoringsrubriek te richten op de eigenschappen van het product dat de studenten opleveren en niet op de manier waarop ze eraan gewerkt hebben, een beslissing die in hoofdstuk 4 is gemotiveerd met het argument dat er bij het oplossen van informatieproblemen vaak verschillende strategieën van toepassing kunnen zijn. Dit laat onverlet dat ook uit dit deelonderzoek kan worden geconcludeerd dat het gebruik van de scoringsrubriek alleen de studenten nog niet informatievaardig maakt. Daarvoor is ook instructie en

oefening nodig. Bij het ontwerpen van die instructie en tijdens de oefeningen kan de scoringsrubriek weer wel een goede dienst bewijzen. Dit is onder andere onderwerp van de studie die in het volgende hoofdstuk wordt besproken.

Tenslotte werd tijdens het focusgroepsgesprek een interessante opmerking gemaakt met betrekking tot het eigen onderwijs van de opleiding idm. De workshop waarin de scoringsrubriek aan de studenten werd geïntroduceerd valt betrekkelijk laat in het curriculum (voor de deelnemende deeltijdstudenten ging het in praktijk om het einde van hun tweede studiejaar; voor voltijdstudenten wordt de workshop gegeven aan het einde van het derde studiejaar). De deeltijdstudenten die deelnamen aan het focusgroepsgesprek vonden unaniem dat er veel vroeger in het curriculum aandacht aan besteed zou moeten worden. Hoewel in het gesprek niet duidelijk werd op welk moment de rubriek dan wel geïntroduceerd zou moeten worden of welke vorm het college zou moeten hebben waarin dat gebeurt, is dit wel een belangrijke opmerking. Door hun aanbeveling op te volgen zou immers de mogelijkheid gecreëerd worden dat studenten de scoringsrubriek gedurende een langere periode in hun opleiding gebruiken om hun eigen ontwikkeling op het gebied van informatievaardigheid te monitoren, wat nu niet het geval bleek te zijn (thema 9).

Het onderzoek waarvan in dit hoofdstuk verslag werd gedaan is op een kleine schaal uitgevoerd. Aan de online enquête namen slechts tien studenten deel, aan het focusgroepsgesprek slechts vijf. De uitkomsten ervan hebben echter wel implicaties voor het onderwijs. De resultaten van zowel de enquête als het focusgroepsgesprek geven aan dat het de voorkeur heeft om de Scoringsrubriek voor Informatievaardigheden op verschillende plekken in een curriculum aan de orde te stellen. Een eenmalige workshop lijkt vooral voor voltijd studenten niet voldoende.

Bijlage 7.1 *Vragen en resultaten van de enquête onder idm-studenten van De Haagse Hogeschool*

1)*

Ik was aanwezig bij de workshop waarin we de scoringsrubriek voor informatievaardigheden hebben gebruikt.

- ☐ Ja (9x)
- ☐ Nee (1x)
- ☐ Weet ik niet meer (0x).

2)*

Na die ene workshop waarin we de scoringsrubriek hebben gebruikt...

- ☐ heb ik haar zelf nog minstens een keer opgezocht om haar opnieuw te gebruiken (7x)
- ☐ heb ik er niets meer mee gedaan (3x).

3)*

Welke variant van de opleiding idm volg je?

- ☐ Voltijd (4x)
- ☐ Deeltijd (6x).

4)*

Wat is je leeftijd?

21 (2x), 23 (2x), 31, 32, 34, 36, 39, 48.

5)*

Wat is je geslacht?

- ☐ Man (6x)
- ☐ Vrouw (4x).

6)*

Ben je bereid deel te nemen aan een vervolginterview over het gebruik van de scoringsrubriek?

- ☐ Ja (5x)
- ☐ Nee (5x).

7)

Als je bij vraag 6 "Ja" hebt ingevuld, geef dan s.v.p. hier je naam op.

Alle studenten die vraag 6 positief beantwoordden vulden bij vraag 7 ook hun naam in.

* Geeft aan dat een antwoord verplicht was.

Evalueren met docenten van De Haagse Hogeschool¹⁴

Abstract: Om inzicht te krijgen in de wijze waarop docenten de scoringsrubriek voor informatievaardigheden (zouden kunnen) gebruiken, zijn vier semi-gestructureerde interviews gehouden en is één vragenlijst via e-mail beantwoord. De vijf respondenten waren docenten van vijf verschillende academies van De Haagse Hogeschool. De topiclijst c.q. vragenlijst die werd gebruikt, was gebaseerd op literatuuronderzoek over het gebruik van scoringsrubrieken in het algemeen en scoringsrubrieken voor informatievaardigheden in het bijzonder. De selectie van de participerende docenten was gebaseerd op eerdere contacten van de onderzoeker. Van alle gespreksdeelnemers was bekend dat ze in het verleden hadden geïnformeerd naar de scoringsrubriek.

De docenten gaven unaniem aan dat ze (delen) van de scoringsrubriek hadden aangepast voordat ze hem gebruikten voor het beoordelen van studieproducten. Ook hadden ze de ervaring dat het gebruik van de scoringsrubriek meer effect sorteert als het instrument ook tijdens een van de lessen met de studenten wordt besproken. Alle respondenten gaven aan dat implementatie van de scoringsrubriek in het curriculum het best kan gebeuren in het kader van aandacht voor onderzoeksvaardigheden, een onderwerp dat binnen het Nederlandse hbo zeer actueel is.

¹⁴ De inhoud van hoofdstuk 8 is eerder gepubliceerd in: How faculty in The Hague University of Applied Sciences uses the scoring rubric for information literacy. *Communications in Computer and Information Science*, 397, 436-442 (Van Helvoort, 2013).

Inleiding

De Scoringsrubriek voor Informatievaardigheden wordt volgens de theoretische uitgangspunten niet alleen geacht bij te dragen aan een betrouwbare beoordeling van studieproducten maar ook aan de bevordering van professioneel informatiegebruik door studenten.

Om het theoretisch gefundeerde antwoord op de centrale onderzoeksvraag van het onderzoek bevestigd te krijgen, was er echter ook behoefte aan ‘evidence based’ argumenten voor het gebruik van de scoringsrubriek voor informatievaardigheden in de onderwijspraktijk. Met het doel die argumenten te verzamelen, zijn in de zomer en in het najaar van 2012 interviews gehouden met docenten van vier verschillende opleidingen c.q. academies van De Haagse Hogeschool. Een vijfde docent beantwoordde de hoofdvraag van dit deelonderzoek via e-mail. De gehanteerde onderzoeksvraag was vergelijkbaar met de vraag uit hoofdstuk 7, zij het dat in dit geval niet aan *studenten* werd gevraagd hoe ze de scoringsrubriek gebruiken maar aan *docenten*. Het onderscheid met de onderzoeken onder docenten en beoordelaars uit hoofdstuk 5 en 6 is dat in hoofdstuk 8 met name de vraag aan de orde is of de docenten de scoringsrubriek überhaupt gebruiken, de motieven die ze daarvoor hebben en de manier waarop ze dat dan doen. De topiclijst die voor de interviews werd gebruikt (bijlage 8.1) was gebaseerd op de literatuur die in de eerste paragraaf van dit hoofdstuk wordt beschreven.

8.1 Literatuur

De uitgangspublicatie voor het literatuuronderzoek was het overzicht van Reddy en Andrade dat ook als sleutelpublicatie is gebruikt voor het literatuuronderzoek dat werd uitgevoerd in het kader van hoofdstuk 7 (Reddy & Andrade, 2010). In dit geval werd de lijst van artikelen uit hun overzicht aangevuld met conferentiepapers en artikelen over scoringsrubrieken op het specifieke terrein van informatievaardigheden. De verwachting was namelijk dat toepassing van een scoringsrubriek voor *informatievaardigheden* extra moeilijkheden zou opleveren als dat gebeurt binnen andere vakgebieden. De gebruikte zoektermen om de aanvullende publicaties op het spoor te komen waren “information literacy”, rubric* en “higher education” (alle drie gecombineerd met behulp van een Booleaanse AND operator). De zoekacties werden uitgevoerd in LISA, LISTA en het Web of Science waarbij de resultatenlijst werd beperkt tot publicaties vanaf 2002. De zoekactie werd uitgevoerd en afgesloten op 11 mei 2012.

Reddy en Andrade beginnen hun bespreking van de manier waarop docenten scoringsrubrieken ervaren met de opmerking dat de meeste docenten in het hoger onderwijs een beperkte opvatting hebben over het nut ervan en ze vooral gebruiken in het kader van beoordelingen en summatieve assessments (Reddy & Andrade, 2010, p. 437 en p. 439). Deze toepassing van een scoringsrubriek wordt ook wel aangeduid als ‘assessment of learning outcomes’ (zie bijvoorbeeld Kesselman & Sherman, 2009, p. 307) wat - zoals in dit onderzoek eerder werd opgemerkt - vaak gesteld wordt tegenover ‘assessment for learning’. Reddy en Andrade rapporteren verschillende onderzoeken waaruit zou blijken dat docenten bij het beoordelen van studieproducten (een vorm van beoordelen die in haar aard aanleiding geeft tot subjectiviteit) de ervaring hebben dat het beoordelingsproces

betrouwbaarder, efficiënter en consistentier wordt als zij daarbij gebruik maken van een scoringsrubriek (Reddy & Andrade, 2010, p. 439). Op de zelfde pagina van hun literatuuroverzicht constateren Reddy en Andrade echter ook dat veel docenten juist niet bereid zijn om zich in het gebruik van scoringsrubrieken te verdiepen. Hun suggestie is dat die docenten veel positiever tegenover het gebruik van scoringsrubrieken zouden staan als ze zouden begrijpen dat deze niet alleen kunnen worden gebruikt voor het *beoordelen* van studentenwerk maar ook uitstekend ingezet kunnen worden in het onderwijs zelf om een leerproces op gang te brengen (Reddy & Andrade, 2010, p. 439).

Ook in de literatuur over het gebruik van scoringsrubrieken voor informatievaardigheden komen de onderwerpen betrouwbaarheid, consistentie en efficiëntie van het beoordelingsproces met enige regelmaat aan de orde. Voorbeelden zijn de publicaties van Fagerheim en Shrode (2009), Oakleaf (2009b), Perruso Brown en Kingsley (2010) en Scharf, Elliot, Huey, Briller en Joshi (2007). Maar ook aan de andere, meer didactische argumenten voor het gebruik van scoringsrubrieken wordt in die publicaties aandacht besteed. Ze worden hieronder besproken.

Het gebruik van scoringsrubrieken in het kader van 'assessments for learning'

In hun literatuuroverzicht beschrijven Reddy en Andrade verschillende onderzoeken die suggereren dat het gebruik van scoringsrubrieken in de klas door de studenten zelf (ofwel in het kader van self-assessment ofwel in het kader van peer-assessment) tot betere studieprestaties leidt op het gebied van academische vaardigheden en kritisch denken (Reddy & Andrade, 2010, pp. 439-440). Om die reden kunnen scoringsrubrieken behalve als een beoordelingsinstrument ook worden beschouwd als een didactisch hulpmiddel dat het leerproces van studenten bevordert. Reddy en Andrade benadrukken wel dat dat vooral geldt als het gebruik van de scoringsrubriek gepaard gaat met een instructie. Voor het gebruik van een scoringsrubriek als hulpmiddel bij het leren is het dus niet voldoende om de rubriek simpelweg beschikbaar te stellen via bijvoorbeeld de elektronische leeromgeving maar moet zij tevens worden toegelicht (zie ook Andrade, 2001). Ook Oakleaf benadrukt hoe belangrijk het is om een scoringsrubriek vooraf met studenten te bediscussieren (2008, p. 245).

In een eerdere publicatie die de basis vormde voor het vorige hoofdstuk (Van Helvoort, 2012) beschreef de onderzoeker zelf hoe een eenmalige workshop in het gebruik van de Scoringsrubriek voor Informatievaardigheden deeltijdstudenten aanzette om de scoringsrubriek uit zichzelf te gaan gebruiken voor self-assessment vóórdat ze hun definitieve uitwerking van een studieopdracht ter beoordeling instuurden of uploaden. De meerderheid van de studenten in de focusgroep uit dat onderzoek meldde bovendien dat ze de scoringsrubriek ook in latere cursussen of onderwijsblokken hadden gebruikt maar dan meer met de criteria van de scoringsrubriek in het achterhoofd, wat werd geïnterpreteerd als een indicatie voor het feit dat het gebruik van de scoringsrubriek door hen was geïnternaliseerd. Dat laatste maakt duidelijk dat de scoringsrubriek hen bovendien heeft geholpen bij het zelfstandig leren.

Feedback

Scoringsrubrieken zijn bij uitstek geschikt voor het geven van feedback. Volgens Megan Oakleaf is de feedbackfunctie dan ook een van de belangrijkste argumenten voor het gebruik van scorings-

rubrieken in het onderwijs (Oakleaf, 2008) wat onderschreven werd door de studenten uit het focusgroepsgeprek dat in het vorige hoofdstuk werd beschreven (zie ook Van Helvoort, 2012). Feedback is zeker in het kader van onderwijs in informatievaardigheden van belang aangezien het oplossen van informatieproblemen een 'complexe vaardigheid' is (Brand-Gruwel, Wopereis & Vermetten, 2005; Scharf et al., 2007) die het best kan worden geleerd door regelmatig te oefenen, te leren van opmerkingen en te verbeteren (Wilson & Blankenship, 2010).

Verbetering van de instructie

De hierboven besproken onderzoeken benadrukken de positieve effecten van het gebruik van scoringsrubrieken op het *leren* door studenten. Er is echter ook onderzoeksliteratuur die juist de positieve effecten op het *instructieproces* benadrukt. Knight beschrijft hoe een scoringsrubriek werd gebruikt voor het beoordelen en analyseren van door studenten samengestelde bibliografieën maar de gegevens die dat opleverde werden in haar geval gebruikt voor het ontwikkelen van voorstellen ter verbetering van een bibliotheekinstructie en een tutorial (2006, p. 47 en p. 52).

Ook Floyd, Colvin en Bodur beschrijven het gebruik van een rubric in een bibliografie-project. In hun geval ging het echter om een rubric met criteria die studenten zelf konden gebruiken voor de selectie van geschikte literatuur (Floyd et al., 2008). Ze concluderen dat de toevoeging van de rubric aan het cursusmateriaal, samen met de uitdrukkelijke instructie van de vakdocent om wetenschappelijke bronnen te gebruiken én een presentatie van een informatiespecialist met bibliotheekinstructies, een positieve invloed had op de kwaliteit van de bibliografieën die door de studenten werden samengesteld. Onduidelijk is echter in welke mate precies het gebruik van de rubric daaraan heeft bijgedragen.

Tenslotte kan in het kader van instructieverbetering hier Oakleafs 'Information Literacy Instruction Assessment Cycle' (ILIAC) genoemd worden (Oakleaf, 2009a). Het model heeft volgens Oakleaf zelf weliswaar zijn wortels in de hierboven genoemde benadering van 'assessments for learning' (en is dus een didactisch instrument, in dit geval gericht op het verbeteren van de vaardigheden van studenten in het evalueren van websites) maar is ook ontworpen om een tweede doel te bereiken, de verbetering van de instructievaardigheden van informatieprofessionals.

Longitudinaal onderzoek naar progressie en vergelijkende studies

De besproken voorbeelden van het gebruik van een scoringsrubriek (zowel dat voor verbetering van de instructie als dat voor summatieve en formatieve beoordelingen door docenten, studiegenoten of door de studenten zelf) waren steeds gebaseerd op afzonderlijke cursussen of onderwijseenheden. Maar scoringsrubrieken kunnen natuurlijk ook gebruikt worden in het kader van een longitudinaal onderzoek om de ontwikkeling van individuele studenten over een bepaalde periode te meten of juist om een cursusprogramma van een bepaald instituut te evalueren, bijvoorbeeld in het kader van een accreditatie (Diller & Phelps, 2008). Het al eerder genoemde onderzoek van Scharf et al. (2007) beschrijft pogingen voor dergelijk longitudinaal onderzoek naar studentenportfolio's. De auteurs waarschuwen zelf echter voor de valkuilen bij de interpretatie van hun onderzoeksresultaten: hun scoringsrubriek is onder andere ontworpen voor het meten van de effecten van instructie in informatievaardigheden op de lange termijn (p. 463) maar er spelen over zo'n langere periode ook andere factoren een rol die het lastig maken om het zuivere effect van de instructie vast

te stellen, zoals het simpele feit dat studenten volwassener worden en het feit dat ook andere delen van het curriculum in de loop der tijd gewijzigd kunnen zijn (p. 471).

Opmerkingen kunnen ook worden geplaatst bij het onderzoek van Wilson en Blankenship (2010). Zij gebruikten een scoringsrubriek om de 'final papers' voor een schrijfpdracht te beoordelen. Anders dan voorgaande jaren maakten een workshop in onderzoeks- en schrijfvaardigheden en een aantal assignments deel uit van de betreffende cursus. Na afloop werden de beoordelingen van de eindopdrachten vergeleken met de beoordelingen van voorgaande jaren en kwamen de onderzoekers tot de conclusie dat de kwaliteit van de uitwerkingen inderdaad vooruit was gegaan (Wilson & Blankenship, 2010, p. 26). In hun geval hebben echter niet meer dan 9 van de 14 studenten die aanvankelijk aan de cursus waren begonnen de eindopdracht ook daadwerkelijk afgerond, een erg klein aantal om een statistisch betrouwbare vergelijking met voorgaande jaren te maken. Bovendien wordt uit hun verslag niet duidelijk op welke manier de 'final papers' uit voorgaande jaren zijn beoordeeld en ontbrak een controlegroep met studenten die het oude programma had gevolgd in plaats van het vernieuwde programma.

Green en Bowser beschreven een onderzoek waarbij studentenproducten van twee verschillende instituten met elkaar werden vergeleken (2006). Ze construeerden een scoringsrubriek voor literatuuroverzichten die, naar hun mening, zowel gebruikt zou kunnen worden voor formatieve en summatieve assessments als voor instructies ("to communicate clear guidelines for expectations of well-constructed literature reviews", p. 192) en evaluaties daárvan (p. 186). Zij kwamen gedurende het onderzoek echter ook tot de conclusie dat hun scoringsrubriek niet zonder meer kon worden toegepast op een ander instituut en dat zij aangepast diende te worden aan de specifieke omstandigheden en de kenmerken van de betreffende studentenpopulatie (p. 195).

8.2 Methode

In de zomer en het najaar van 2012 zijn vier docenten van verschillende opleidingen van De Haagse Hogeschool geïnterviewd over de manier waarop zij de Scoringsrubriek voor Informatievaardigheden in hun eigen onderwijs gebruikten. Een vijfde collega heeft deze vraag uiteindelijk beantwoord via e-mail, nadat eerdere pogingen om tot een afspraak te komen verschillende malen waren gestrand op agendaproblemen. De deelnemende respondenten waren allemaal werkzaam als docent of hogeschooldocent en werkten niet alleen op vijf verschillende opleidingen (Voeding en Diëtetiek, Sociaal Pedagogische Hulpverlening, Communicatie, Business IT & Management en Werktuigbouwkunde) maar ook op verschillende academies van De Haagse Hogeschool die op dat moment in totaal dertien academies telde.

De respondenten zijn in het voorjaar van 2012 benaderd omdat zij in het verleden hadden laten blijken dat ze in de Scoringsrubriek voor Informatievaardigheden geïnteresseerd waren.

De onderwerpen die in de interviews aan de orde zijn gesteld, werden afgeleid van het literatuuroverzicht in paragraaf 8.1 en in de vorm van een 'topiclijst' vooraf toegestuurd aan de respondenten zodat zij zich op het interview konden voorbereiden. Onderwerpen waren het gebruik van de

scoringsrubriek voor summatieve en diagnostische toetsing, het geven van feedback, het verbeteren van de instructie en het curriculum, gebruik voor peer- en self-assessment, instructie met de scoringsrubriek, monitoring van studenten en de implementatie van de scoringsrubriek. De volledige topiclijst is opgenomen als bijlage 8.1. De gesprekken zijn door de onderzoeker zelf gevoerd, werden opgenomen met een memorecorder en vervolgens woord voor woord getranscribeerd ten behoeve van verdere analyse. De analyse zelf is uitgevoerd met behulp van de commentaarfunctie van MS Word.

8.3 Resultaten

Summatieve beoordeling

Wellicht de meest opvallende uitkomst van de interviews was dat slechts één van de respondenten aangaf de Scoringsrubriek voor Informatievaardigheden ook daadwerkelijk toe te passen voor een summatieve beoordeling. In de opleiding business it & management (bim) werd de scoringsrubriek gebruikt in het allereerste onderwijsblok bij de beoordeling van een essay. “Het doel van die essay-opdracht is dat studenten aan de gang gaan met informatievaardigheden en het tweede doel is om ze zelf kennis te laten maken met het vakgebied” (docent bim). Zij vervolgt: “Studenten moeten dus zelf uitzoeken wat business intelligence of iets anders is. Maar we willen er ook hun taalvaardigheid mee toetsen, diagnostisch dan, met dezelfde opdracht. Daarvoor hebben we de oorspronkelijke scoringsrubriek uitgebreid met criteria voor zinsbouw, tekststructuur en publieksgericht schrijven.” De docent legde verder uit dat de scoringsrubriek niet alleen werd gebruikt voor het *beoordelen* van de essays maar ook voor het geven van feedback, aanvullend bij de opmerkingen die ze ook al krijgen op de uitwerkingen zelf.

Aanpassen van bestaande beoordelingsmodellen

De andere geïnterviewden gaven aan dat ze de scoringsrubriek niet gebruikten voor het beoordelen van studieproducten. Bij verdere discussie over dit thema bleek echter dat het schema wel gebruikt is om bestaande beoordelingscriteria en –instrumenten aan te passen, bijvoorbeeld die voor het beoordelen van schrijfoopdrachten en de bachelor thesis (docent voeding en diëtetiek). De docent sociaal pedagogische hulpverlening (sph) gaf aan dat men binnen haar opleiding nog altijd worstelt met de plek van Information Problem Solving in het curriculum en dat het gebruik van alle criteria uit de scoringsrubriek daar voorlopig nog niet aan de orde is.

Curriculumontwerp

Met de vorige opmerking verwees de docent sph naar een onderwerp dat in de interviews het vaakst genoemd werd, het gebruik van de Scoringsrubriek voor Informatievaardigheden door curriculumontwerpers voor het selecteren van relevante leerinhoud. Zo schreef de docent van de opleiding werktuigbouwkunde die per e-mail reageerde, dat de scoringsrubriek bij zijn opleiding wordt gebruikt als “een van de achterliggende documenten voor het Plan van Aanpak om onderzoeksvaardigheden een meer prominente plaats in het curriculum te geven”. De docent van de opleiding communicatie en de docent sph voegden daar aan toe dat dergelijke vaardigheden voor docenten vaak impliciete kennis zijn en dat de scoringsrubriek voor hen duidelijker heeft gemaakt

wat zij eigenlijk willen dat de studenten zouden moeten kunnen en zouden moeten doen. Dat hielp hen bovendien om een duidelijker beeld te krijgen van de onderwerpen waarover les gegeven zou moeten worden.

Instructie in gebruik van de scoringsrubriek

De opleidingen die de scoringsrubriek (of onderdelen daaruit) op een of andere manier gebruikten bij het beoordelen van studieproducten (bim, communicatie, sph en voeding en diëtetiek) stellen de scoringsrubriek of de betreffende beoordelingscriteria ook beschikbaar in de studiehandleiding of op de elektronische leeromgeving Blackboard. Zowel de docent sph als de docent voeding en diëtetiek heeft echter de ervaring dat dat niet voldoende is en dat de scoringsrubriek of de lijst met beoordelingscriteria ook met de studenten besproken moet worden. “Je zult studenten nog steeds moeten leren wáárom een correcte citaatvermelding en het gebruik van meer wetenschappelijke informatiebronnen zo belangrijk zijn” (docent sph). Van de vijf opleidingen die betrokken waren bij dit onderzoek was bim opnieuw de enige waar de studenten expliciet werden geïnstrueerd in het gebruik van de scoringsrubriek.

Peer-assessment

Het schema werd bij bim ook gebruikt voor peer-assessment. De docent legde uit hoe dat in zijn werk gaat: “De studenten hebben twee aan twee hun concepten uitgewisseld en elkaars werk met de scoringsrubriek beoordeeld. De concepttekst dienden ze vervolgens, samen met de commentaren van hun medestudent, op te nemen in hun eigen dossier dat uiteindelijk werd beoordeeld door de docent”. Bij de andere opleidingen werd een dergelijke vorm van peer-assessment niet toegepast. De geïnterviewde docenten gingen er wel vanuit dat studenten de beoordelingscriteria gebruikten om hun eigen werk te beoordelen voordat ze hun definitieve uitwerking inleveren, maar geen van hen had er zicht op of dat ook daadwerkelijk gebeurde.

Implementatie

Voor de meesten van de geïnterviewden gold dat juist ook de implementatie in de opleiding als problematisch werd ervaren. De docent sph gaf aan dat haar collega's enthousiast waren toen de scoringsrubriek voor het eerst met hen werd besproken maar dat ze twijfelde of zij de scoringsrubriek nu nog in het achterhoofd hebben als ze studentenwerk beoordelen. Bij voeding en diëtetiek was de afspraak tussen docenten dat de scoringsrubriek zou worden opgenomen in de handleiding bij de eerstejaars studieopdrachten maar recent had zij de ervaring dat meer dan de helft van de docenten en tutoren uit dat eerste jaar niet eens weet had van het bestaan van de scoringsrubriek. “We zijn dus zeker niet geslaagd bij de ingebruikneming van de scoringsrubriek” voegde ze daar aan toe en ze deed de suggestie om ook de docenten te trainen in het gebruik van zo'n beoordelingsinstrument. Volgens de docent van de opleiding communicatie is echter juist het gebrek aan tijd daarvoor het grootste probleem. Zij had echter ook de recente ervaring dat een aankomende accreditatie van de opleiding haar wél de mogelijkheid had gegeven om het formuleren van goede beoordelingscriteria onder de aandacht te brengen bij haar collega's.

De enige van de vijf opleidingen die bij het onderzoek waren betrokken en waar de implementatie tot dan toe succesvol verlopen leek, was opnieuw business it & management. Daar moet echter bij worden opgemerkt dat deze docent bij bim tevens de manager van het betreffende team was. Zij

zelf drukte het als volgt uit: “Als hoofd van de opleiding heb ik zelf interesse in de scoringsrubriek en dat helpt natuurlijk wel”.

8.4 Conclusie en discussie

Uit de resultaten kan worden geconcludeerd dat de Scoringsrubriek voor Informatievaardigheden potentie heeft om gebruikt te worden in het onderwijs van De Haagse Hogeschool maar dat zij ten tijde van het onderzoek nergens werd gebruikt in de vorm zoals zij is gepubliceerd. Docenten van verschillende academies en opleidingen gebruikten onderdelen van de scoringsrubriek om beoordelingscriteria bij de eigen studieopdrachten te (her)formuleren of te verfijnen. Daaruit blijkt dat zij de afzonderlijke beoordelingscriteria beschouwden als zogenaamde “primary trait rubrics” (Center for Advanced Research on Language Acquisition : University of Minnesota, 2012) maar dat zij de oorspronkelijke formulering van de criteria en de gedragsbeschrijvingen wel aanpasten aan hun eigen praktijk. De enige opleiding die wel gebruik maakte van de oorspronkelijke scoringsrubriek breidde deze vervolgens uit met beoordelingscriteria op het gebied van schrijfvaardigheid. Deze constatering die werd gedaan op basis van de antwoorden tijdens de interviews bevestigt de conclusie uit het literatuuroverzicht van Green en Bowser (2006) dat een scoringsrubriek die wordt gebruikt binnen een bepaald instituut niet zonder meer kan worden toegepast op een ander instituut. Zelfs binnen één hogeschool bleek dat opleidingen verschillende doelen hebben met de studieopdrachten waarin informatievaardigheden een rol spelen.

Het gebruik van de scoringsrubriek of een van de individuele criteria voor een longitudinaal assessment (ten behoeve van het meten van meerjarige ontwikkeling op het gebied van informatievaardigheid) werd door geen van de geïnterviewde docenten genoemd. Dat kan echter ook verklaard worden uit het feit dat binnen alle betreffende opleidingen pas recent aandacht voor informatievaardigheden was ontstaan.

Implicaties voor het onderwijs

Uit de interviews werd duidelijk dat docenten die binnen hun opleiding (onderdelen van) de scoringsrubriek willen gaan gebruiken bij het beoordelen van studieopdrachten er voor zullen moeten zorgen dat er ook in het onderwijs extra aandacht aan wordt besteed. Ze moeten er bovendien rekening mee houden dat ook hun collega's die geacht worden er mee te gaan werken gemotiveerd en getraind dienen te worden in het gebruik van de scoringsrubriek. De opmerkingen van de docenten sph en voeding en diëtetiek dat het niet voldoende is om de scoringsrubriek op te nemen in de studiehandleiding of het studiemateriaal, werd bevestigd door de literatuur in paragraaf 8.1 (Reddy & Andrade, 2010, p. 440). Ook de teleurstelling van de geïnterviewde docenten over de geringe betrokkenheid van hun collega's is herkenbaar voor wie bekend is met de literatuur over de implementatie van informatievaardigheden in domeingerelateerde curricula. Keer op keer blijkt hoe lastig het is om docenten uit een specifieke academische discipline enthousiast te krijgen voor de instructie van dergelijke generieke vaardigheden (zie bijvoorbeeld Godwin, 2005, pp. 70-72). Het is dan ook niet verwonderlijk dat het in dit onderzoek juist de opleiding business it & management was waar het gebruik van de scoringsrubriek wel werd geaccepteerd, aangezien de manager van de opleiding in dat geval zelf een pleitbezorger van het gebruik ervan was.

De respondenten die waren betrokken bij dit deelonderzoek gaven allen aan dat de scoringsrubriek ook gebruikt was om onderzoeksvaardigheden beter te integreren in het curriculum. Deze opmerking dient te worden begrepen in de context van de veranderde rol die hogescholen zichzelf hebben toegedicht: van het opleiden voor een specifiek beroep met de nadruk op beroepsvaardigheden naar het opleiden van professionals die ook in staat zijn om, binnen de context van hun beroep, zelfstandig praktijkgericht onderzoek uit te voeren (UASnet, n.d., p. 8). Deze verandering in de missie van hogescholen is mede veroorzaakt door de ontwikkeling van de West-Europese economieën tot een kenniseconomie en maakt onderdeel uit van de Europese en Nederlandse ambitie om “uit te groeien tot de meest dynamische kennissamenleving van de wereld”(AWT, 2005, p. 8). Op het moment dat de interviews met de docenten werden gevoerd was de ontwikkeling van een leerlijn onderzoeksvaardigheden een belangrijk onderwerp binnen de verschillende opleidingen. Niet duidelijk is of de geïnterviewden op een ander tijdstip (bijvoorbeeld vijf jaar eerder of juist later) aan de scoringsrubriek ook de functie toegekend zouden hebben van katalysator bij het ontwikkelen van een onderzoeksleerlijn. Opvallend is echter wel dat ze allen uitdrukkelijk beweerden dat de scoringsrubriek hen daar op dat moment goed bij ondersteunde. In die zin kan worden gesteld dat de Scoringsrubriek voor Informatievaardigheden inderdaad aansluit bij recente ontwikkelingen in het hoger onderwijs.

Kanttekening

Een beperking van het in dit hoofdstuk beschreven deelonderzoek is het feit dat de scoringsrubriek binnen De Haagse Hogeschool nooit consequent en systematisch is gepromoot. In het najaar van 2009 verscheen een interview met de onderzoeker over de scoringsrubriek in het personeelsblad van de hogeschool. Ook verzorgde hij in die periode zowel een workshop als een presentatie over de scoringsrubriek voor zijn collega's van andere opleidingen. Naar aanleiding van die workshop en de presentatie konden collega's die geïnteresseerd waren de onderzoeker benaderen voor extra informatie of tips maar deze had geen tijd beschikbaar om zijn collega's uitgebreid te ondersteunen bij de implementatie van de scoringsrubriek in hun curricula. Deze beperking van het hier beschreven onderzoek had echter ook een voordeel: de geïnterviewde collega's hadden er blijk van gegeven oprecht geïnteresseerd te zijn in het onderwerp informatievaardigheid en hoewel ze allen een positieve houding hadden tegenover de scoringsrubriek is er geen enkele indicatie geweest dat zij zich bij het beantwoorden van de vragen hebben laten leiden door sociale wenselijkheid.

Bijlage 8.1 *Topic lijst voor de interviews met docenten van De Haagse Hogeschool*

- Je hebt de scoringsrubriek gebruikt voor een summatieve beoordeling ('voor een cijfer').
Vervolg vragen:
 - Soort product dat wordt beoordeeld (rapport, advies, bibliografie etcetera)
 - Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
 - Is de scoringsrubriek vooraf besproken; hebben studenten er eerder mee gewerkt?
 - Hoe nuttig is het gebruik van de scoringsrubriek voor dit doel? Welke voordelen heeft het?
- Je hebt de scoringsrubriek gebruikt voor het geven van feedback.
Vervolg vragen:
 - Alleen bij summatieve beoordelingen of ook bij formatieve/diagnostische?
 - Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
 - Soort product dat wordt beoordeeld (rapport, advies, bibliografie etcetera)
 - Hoe nuttig is het gebruik van de scoringsrubriek voor dit doel? Welke voordelen heeft het?
- Je hebt de scoringsrubriek gebruikt voor een diagnostische toets om te achterhalen op welke gebieden de instructie verbeterd zou moeten worden.
Vervolg vragen:
 - Soort product dat wordt beoordeeld (rapport, advies, bibliografie etcetera)
 - Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
 - Hoe nuttig is het gebruik van de scoringsrubriek voor dit doel? Welke voordelen heeft het?
- Je hebt de scoringsrubriek gebruikt om tekortkomingen in het curriculum te achterhalen.
Vervolg vraag:
 - Op welke punten zijn verbeteringen voorgesteld?
- Je hebt de scoringsrubriek door studenten laten gebruiken om hun eigen werk te beoordelen voordat ze het inleveren ('self-assessment').
Vervolg vragen:
 - Is de scoringsrubriek vooraf besproken; hebben studenten er eerder mee gewerkt?
 - Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
 - Hoe nuttig is het gebruik van de scoringsrubriek volgens de studenten? Welke voordelen heeft het?

- Je hebt de scoringsrubriek door studenten laten gebruiken om elkaars werk op dit facet te beoordelen ('peer-assessment').

Vervolg vragen:

- ☐ Is de scoringsrubriek vooraf besproken; hebben studenten er eerder mee gewerkt?
- ☐ Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
- ☐ Hoe nuttig is het gebruik van de scoringsrubriek volgens de studenten? Welke voordelen heeft het?

- De scoringsrubriek is behandeld in een college en is onderdeel van de instructie.

Vervolg vragen:

- ☐ Soort product dat in het college is besproken / waarover de instructie is gegeven (rapport, advies, bibliografie etcetera)
- ☐ Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
- ☐ Welk effect meen je waar te nemen naar aanleiding van de instructie?

- Je gebruikt de scoringsrubriek om voortgang van de student op langere termijn te monitoren.

Vervolg vragen:

- ☐ Soort product dat wordt beoordeeld (rapport, advies, bibliografie etcetera)
- ☐ Heb je de scoringsrubriek aangepast vanwege je eigen vakgebied, studentenpopulatie etcetera?
- ☐ Worden ingevulde scoringsformulieren opgenomen in het portfolio van de student?
- ☐ Hoe nuttig is het gebruik van de scoringsrubriek voor dit doel? Welke voordelen heeft het?

- Je hebt collega's geprobeerd te overtuigen om ook met de scoringsrubriek te gaan werken.

Vervolg vraag:

- ☐ Op welke weerstand ben je daarbij gestuit?

9

Algemene conclusies en discussie

Inleiding

Het streven in dit onderzoek was het ontwikkelen en testen van een generiek toepasbaar instrument waarmee kan worden beoordeeld of studenten in het hoger onderwijs voldoende informatievaardig zijn en waarmee hun informatievaardigheid tevens wordt bevorderd. In het onderzoek is daarvoor een mogelijke oplossing aangereikt, de Scoringsrubriek voor Informatievaardigheden. Voor het ontwikkelen en testen daarvan is een aantal deelonderzoeken uitgevoerd. Paragraaf 9.1 geeft een overzicht van de resultaten uit die onderzoeken en de antwoorden die daarmee kunnen worden gegeven op de onderzoeksvragen. Deze samenvatting van de onderzoeksresultaten en de conclusies die daaruit kunnen worden getrokken, worden gevolgd door een reflectie in paragraaf 9.2 op de aanpak die bij het onderzoek werd gehanteerd. In paragraaf 9.3 worden in verschillende subparagrafen de reikwijdte en de beperkingen van het onderzoek besproken. Het hoofdstuk wordt afgesloten met paragrafen over de inzichten die het onderzoek heeft opgeleverd zonder dat daar aanvankelijk onderzoek naar zou worden gedaan, een aantal aanbevelingen voor de onderwijspraktijk en suggesties voor nader onderzoek.

9.1 Resultaten en conclusies van het onderzoek

De centrale onderzoeksvraag waarop met de verschillende onderzoeken een antwoord werd gezocht, luidde: Wat is een betrouwbaar, valide en praktisch bruikbaar instrument dat generiek toepasbaar is voor het beoordelen én het bevorderen van de informatievaardigheden van individuele studenten in het hoger onderwijs?

Om een antwoord te vinden op bovenstaande centrale onderzoeksvraag is onderzoek gedaan aan de hand van de volgende onderliggende onderzoeksvragen:

1. Wat is de meest geschikte assessmentmethode voor het beoordelen en bevorderen van informatievaardigheden?
2. Hoe ziet een scoringsrubriek eruit die geschikt is voor het beoordelen en bevorderen van informatievaardigheden bij studenten in het hoger onderwijs?
3. Hoe betrouwbaar en valide is de scoringsrubriek?
4. Hoe wordt de scoringsrubriek door studenten en docenten in het hoger onderwijs gebruikt en hoe ervaren zij het gebruik ervan?

De antwoorden op deze onderzoeksvragen waren als volgt.

Onderzoeksvraag 1: Wat is de meest geschikte assessmentmethode voor het beoordelen en bevorderen van informatievaardigheden?

Een van de deelstudies was een literatuurstudie naar verschillende beoordelingsmethodes. Uit de gevonden literatuur kwam naar voren dat authentieke studieopdrachten die nauw gerelateerd zijn aan het beroep of het vakgebied waarin een student wordt opgeleid, bij uitstek geschikt zijn om te beoordelen of de betreffende student in staat is om complexe informatieproblemen op te lossen.

Bijkomend voordeel van dergelijke opdrachten waarbij studenten hun eigen informatievaardigheden dienen in te zetten, is dat zij ook een leerproces op gang brengen (Knight, 2006; Oakleaf, 2008). In het hoger onderwijs worden dan ook veel onderwijsvormen toegepast waarbij gebruik wordt gemaakt van dergelijke informatievraagstukken, zoals projectonderwijs, probleemgestuurd onderwijs en competentiegericht onderwijs (Brand-Gruwel, Wopereis & Vermetten, 2005). Eveneens op basis van literatuuronderzoek is gedurende het onderzoek voor dit proefschrift een model voor het Information Problem Solving-proces geconstrueerd dat goed aansluit bij de didactische uitgangspunten van zulk onderwijs. Het model fungeert in de rest van het onderzoek als theoretisch kader bij het ontwerpen van een beoordelingsinstrument. Het idee dat het oplossen van informatieproblemen bijdraagt aan de constructie van persoonlijke kennis is geoperationaliseerd aan de hand van verschillende soorten kennis op een specifiek vakgebied en metacognitieve kennis over leren en cognitieve taken in het algemeen.

De motivatie voor het gebruik van studieopdrachten voor het beoordelen van informatievaardigheden van studenten, is tevens gestoeld op het gegeven dat die assessmentvorm zich meer dan andere beoordelingsinstrumenten richt op 'hele taken' uit een realistische context. In het Information Problem Solving-model komt dat idee van 'hele taken' terug. In het model wordt uitgedrukt dat bij het oplossen van informatieproblemen de verschillende deelactiviteiten die worden uitgevoerd, onderling sterk samenhangen. Het beoordelen van dergelijke deelactiviteiten als 'discrete vaardigheden' zoals gebeurt in meer traditionele tests, is dan minder op zijn plaats.

Deze argumentatie voor het gebruik van studieopdrachten is vooral gebaseerd op de inhoudsvaliditeit van de beoordelingsmethode. Naast de validiteit speelt bij de keuze van een beoordelingsmethode ook de betrouwbaarheid een belangrijke rol. Die kan onder andere worden bevorderd door het gebruik van een 'scoring rubric' als beoordelingskader (Oakleaf, 2008, p. 246). Zo'n scoringsrubriek geeft een overzicht van de te hanteren beoordelingscriteria aangevuld met omschrijvingen van het verwachte gedrag van studenten op verschillende niveaus. Andere voordelen van het gebruik van scoringsrubrieken, naast de betrouwbaarheid van de beoordelingen, zijn volgens Oakleaf (2008) de uitgebreide feedback die ze bieden aan studenten en de duidelijkheid die de omschrijvingen van het verwachte gedrag vooraf verschaffen. Die laatste twee kenmerken zorgen ervoor dat een scoringsrubriek tevens fungeert als een instrument dat de *verwerving* van vaardigheden bevordert.

Onderzoeksvraag 2: Hoe ziet een scoringsrubriek eruit die geschikt is voor het beoordelen en bevorderen van informatievaardigheden bij studenten in het hoger onderwijs?

De kern van dit proefschrift is de Scoringsrubriek voor Informatievaardigheden waarvan het prototype is opgenomen als bijlage 4.1. De scoringsrubriek geeft gedragsomschrijvingen voor diverse beoordelingscriteria op twee niveaus, het professionele en het inadequate gedrag. Het prototype blijkt in het algemeen goed te functioneren als beoordelingsinstrument. Op basis van de resultaten uit het onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam zijn wel twee suggesties gedaan om het prototype te verbeteren.

- De volgorde van de beoordelingscriteria werd door de twee beoordelaars die het prototype in dat project hadden gebruikt niet als logisch ervaren. Verplaatsing van criterium 4

(“Bronverwijzingen in de tekst”) naar de 2^e positie zou een verbetering zijn. Het criterium gaat dan vooraf aan het beoordelingscriterium met betrekking tot het bronnenoverzicht zelf. De onderlinge samenhang van de twee beoordelingscriteria met betrekking tot de bronverwijzingen komt in die nieuwe volgorde beter tot uitdrukking.

- De twee beoordelaars van het project bij Mediastudies hadden ook een duidelijke voorkeur voor een grotere spreiding van de beoordelingen dan die met de 6 punts Likertschaal. Ze gaven aan veel liever te werken met een 10 punts schaal, onder andere omdat die beter aansluit bij het in Nederland gebruikelijke systeem van rapportcijfers. Bij de aanpassing van het prototype is het idee van 10 punts Likertschaal overgenomen maar wordt nog wel steeds gebruik gemaakt van omschrijvingen in de vorm van *woorden* in plaats van cijfers. Het gebruik van een 10 punts Likertschaal in de vorm van *woorden* doet namelijk meer dan scores in de vorm van cijfers recht aan de feedbackfunctie die de scoringsrubriek ook heeft.

Bijlage 9.1 geeft een versie 2.0 van de scoringsrubriek waarin de twee bovenstaande wijzigingen zijn doorgevoerd. Voor de verwoording van de 10 punten op de Likertschaal is de in Nederland gebruikelijke aanduiding van de rapportcijfers op een schaal van 1 tot 10 - zoals vermeld op Wikipedia.nl - aangehouden (Schoolrapport, n.d.).

De scoringsrubriek wordt geacht generiek toepasbaar te zijn voor verschillende vakgebieden en bij verschillende opdrachten. De meerwaarde van deze generieke toepasbaarheid is dat daarmee een kader wordt geboden dat door verschillende opleidingen en instituten kan worden gebruikt om de informatievaardigheden van hun studenten met elkaar te vergelijken. Daarnaast biedt een meer algemeen voorbeeld van een beoordelingsmodel aan collega-docenten in het hoger onderwijs die nog niet over zo'n instrument beschikken, de mogelijkheid om gebruik te maken van een 'beproefde techniek' die indien gewenst kan worden aangepast aan de specifieke eigen context. In welke mate de Scoringsrubriek voor Informatievaardigheden zichzelf heeft bewezen en dus ook werkelijk een 'beproefde techniek' is, was onderdeel van het onderzoek naar de validiteit en de betrouwbaarheid waarvan de resultaten bij onderzoeksvraag 3 aan de orde komen.

Onderzoeksvraag 3: Hoe valide en betrouwbaar is de scoringsrubriek?

Het gebruik van een scoringsrubriek wordt over het algemeen geacht bij te dragen aan een consistente beoordeling van studieopdrachten (Ito, 2015). Hoewel dit in zijn algemeenheid wel kan worden gezegd, is het voor een specifieke scoringsrubriek die wordt ontwikkeld nog wel nodig te onderzoeken hoe valide en betrouwbaar die éne scoringsrubrieken dan is. Deze eisen met betrekking tot betrouwbaarheid en validiteit van een beoordelingsinstrument worden ook wel beschouwd als “klassieke” psychometrische kenmerken (Van de Watering & Dierick, 2002, p. 63). Daarbij spelen verschillende vormen van validiteit en betrouwbaarheid een rol. De Scoringsrubriek voor Informatievaardigheden is op verschillende van deze kenmerken onderzocht.

De inhoudsvaliditeit is met name gecontroleerd gedurende het ontwikkelproces van de scoringsrubriek dat werd beschreven in hoofdstuk 4. Een van de fases gedurende het ontwikkelproces was de collegiale toetsing. Tijdens drie rondes werd aan collega-docenten gevraagd om concepten van

de scoringsrubriek te voorzien van commentaar. Op die manier is er tijdens de constructie van de scoringsrubriek voor gezorgd dat de te gebruiken beoordelingscriteria en gedragsomschrijvingen werden herkend door ervaren docenten en beoordelaars uit het hoger onderwijs.

Twee andere vormen van validiteit die door Cronbach en Meehl (1955) worden onderscheiden zijn criteriumvaliditeit en constructvaliditeit. Constructvaliditeit heeft betrekking op de mate waarin verschillende items van een beoordelingsinstrument verwijzen naar één en het zelfde construct. In het onderzoek naar de scoringsrubriek is een betrouwbaarheidsanalyse in SPSS uitgevoerd om deze vast te stellen. Hoewel de naam van de procedure suggereert dat daarmee alleen de betrouwbaarheid van een schaal wordt vastgesteld, levert deze ook gegevens op waarmee uitspraken kunnen worden gedaan over de constructvaliditeit. Cronbach's Alpha en de Item-Totaal Statistieken voor de tests bij de Nederlandse Defensie Academie en de Open Universiteit gaven aan dat de afzonderlijke beoordelingscriteria verwijzen naar één gemeenschappelijk construct 'informatievaardigheden'. Daar waar dat voor afzonderlijke beoordelaars een enkele keer niet gold - het criterium 'Secundaire bronnen' voor beoordelaar 2 bij de Nederlandse Defensie Academie en het criterium 'Zoektermen' voor beoordelaar 1 bij de Open Universiteit - kon een legitieme verklaring gegeven worden in de opzet van de studieopdracht of de achtergrond van de beoordelaar. In geen van deze twee contexten kwam het voor dat een beoordelingscriterium bij beide beoordelaars leidde tot een extreem lage of negatieve correlatie tussen het betreffende item en de totale schaal.

De procedure betrouwbaarheidsanalyse van SPSS is ook gebruikt om de interne consistentie vast te stellen van de beoordelingen die zijn opgemaakt bij de opleiding Mediastudies van de Universiteit van Amsterdam. Ook daar gaven Cronbach's Alpha en de Item-Totaal Statistieken aan dat er sprake was van één gemeenschappelijk construct. In die analyse werden de beoordelingscriteria van de scoringsrubriek ook nog eens beschouwd als de gemeten variabelen van het achterliggende construct Informatievaardigheid. Door middel van factoranalyses met SPSS werd onderzocht of er toch geen tweede factor of 'dimensie' kon worden vastgesteld. De resultaten voor allebei de beoordelaars gaven aan dat met de scoringsrubriek één ongedeeld construct werd gemeten en er dus sprake is van één samenhangend beoordelingsmodel. De resultaten bij beoordelaar 2 leken in eerste instantie te suggereren dat er een tweede factor kon worden onderscheiden maar nadere analyse gaf aan dat beperking tot één factor meer voor de hand ligt.

Bij de derde vorm van validiteit, criteriumvaliditeit, wordt onderzocht in welke mate de score die wordt bereikt met het éne instrument correleert met de score die wordt bereikt met een ander (concurrerend) beoordelingsinstrument dat min of meer gelijktijdig wordt toegepast (Cronbach & Meehl, 1955). In het onderzoek bij de opleiding Mediastudies van de UvA werden de beoordelingen met de scoringsrubriek vergeleken met de scores die werden toegekend met behulp van het beoordelingsformulier voor informatievaardigheden van de opleiding zelf. De gevonden waarden voor Pearson correlatie waren zo hoog dat inderdaad kon worden geconcludeerd dat de scoringsrubriek in die context een geschikt model is voor het beoordelen van de betreffende werkstukken. Dat betekent uiteraard nog niet dat we ook zeker weten dat met die twee beoordelingsmodellen het construct 'informatievaardigheden' wordt gemeten. Er kan echter wel worden geconcludeerd dat dat laatste zeer aannemelijk is omdat

de hoge correlaties uit het onderzoek naar de criteriumvaliditeit niet op zich zelf staan maar worden aangevuld met

- a) de inhoudsvaliditeit van de scoringsrubriek die werd geborgd door het raadplegen van literatuur en experts tijdens het ontwikkeltraject en
- b) de resultaten uit het onderzoek naar de constructvaliditeit.

Voor het vaststellen van de *betrouwbaarheid* van de scoringsrubriek voor informatievaardigheden werd gebruik gemaakt van het idee van interbeoordelaarsbetrouwbaarheid, de mate waarin twee beoordelaars overeenstemmen als ieder van hen de zelfde set uitwerkingen in de zelfde periode beoordeelt. Met betrekking tot deze interbeoordelaarsbetrouwbaarheid, uitgedrukt in het percentage aanpalende overeenstemming tussen twee beoordelaars, werd in het onderzoek aan de Nederlandse Defensie Academie geconcludeerd dat die zonder twijfel voldoende was. In het onderzoek aan de Faculteit Psychologie en Onderwijswetenschappen van de Open Universiteit werd de conclusie getrokken dat de mate van interbeoordelaarsbetrouwbaarheid te lijden had gehad onder het gegeven dat een van de beoordelaars onvoldoende thuis was in het vakgebied waarop de opdrachten betrekking hadden. Het was onder andere deze onduidelijkheid over de interbeoordelaarsbetrouwbaarheid van de beoordelingen bij de Open Universiteit die heeft geleid tot het besluit om een grootschaliger onderzoek uit te voeren bij de opleiding Mediastudies van de UvA. Bij dit laatste onderzoek werden bewust beoordelaars ingezet met een sterke inhoudelijke betrokkenheid. Bovendien werd voorafgaande aan het eigenlijke beoordelingsproces veel aandacht besteed aan het trainen van de beoordelaars. Deze inspanningen hebben zich uitbetaald in die zin dat de mate van interbeoordelaarsbetrouwbaarheid ook bij deze UvA-studie voldoende tot goed was. Dat werd niet alleen uitgedrukt in het percentage aanpalende overeenstemming tussen de twee beoordelaars maar ook door de waarden voor de Intraclass Correlation Coefficient. Gezien het aantal beoordeelde werkstukken kon bij de opleiding Mediastudies van de UvA tevens gebruik gemaakt worden van deze statistische maat die, net als Cronbach's Alpha en de Item-Totaal Statistieken, kan worden vastgesteld met de procedure betrouwbaarheidsanalyse in SPSS. Voor de afzonderlijke criteria werd over het algemeen een ICC-waarde bereikt die wordt beschouwd als 'goed', voor de eindscore was die zelfs excellent.

Wel werd opgemerkt dat gebruik van één of slechts een zeer beperkt aantal criteria van de scoringsrubriek ten koste zal gaan van de interbeoordelaarsbetrouwbaarheid. Hoe meer van de beoordelingscriteria worden gebruikt, hoe hoger de onderlinge overeenstemming tussen twee beoordelaars uitpakt. Het gebruik van losse beoordelingscriteria als "primary traits" wordt voor summatieve beoordelingen dan ook afgeraden. Voor het gebruik van losse beoordelingscriteria in het kader van instructie, peer- of self-assessment geldt dit bezwaar uiteraard veel minder.

Onderzoeksvraag 4: Hoe wordt de scoringsrubriek door studenten en docenten in het hoger onderwijs gebruikt en hoe ervaren zij het gebruik ervan?

Naast de klassieke psychometrische eisen die aan een beoordelingsinstrument mogen worden gesteld, is het tevens van belang dat de beoordelaars die er mee moeten werken er goed mee uit de voeten kunnen. Aan acht docenten van de opleiding Commerciële Economie van Saxion werd onder andere de vraag voorgelegd hoeveel tijd zij nodig hadden gehad om in het kader van de

beoordeling van individuele procesverslagen de scoringsrubriek in te vullen. Zes van hen gaven aan daarvoor minder dan vijf minuten nodig te hebben gehad, twee docenten antwoordden dat ze er vijf tot tien minuten per beoordeling aan hadden besteed. Informatiever was een opmerking van de twee beoordelaars van de opleiding Mediastudies van de UvA. Deze twee student-assistenten waren wel eerder betrokken geweest bij het begeleiden van de opdrachten maar nooit ingezet voor het beoordelen van de uitwerkingen. In het kader van het onderzoek dienden zij dan ook zowel te leren werken met de scoringsrubriek als met het beoordelingsformulier van de opleiding zelf. Voor beiden gold dat ze de scoringsrubriek gebruikersvriendelijker vonden dan het UvA nakijkformulier. Vooral de Likertschalen van de scoringsrubriek zorgden er volgens hen voor dat je geen onderdelen van de beoordeling vergat in te vullen en dat ze betrekkelijk makkelijk en snel tot een score kwamen. Hoewel deze efficiency in de onderwijskundige vakliteratuur niet vaak wordt genoemd als argument om over te stappen op het gebruik van een scoringsrubriek, worden soortgelijke ervaringen als deze van de student-assistenten wel regelmatig genoemd als een bijkomend voordeel (Stevens & Levi, 2013, pp. 14-15).

Bovenstaande opmerkingen over de praktische bruikbaarheid van de scoringsrubriek kunnen worden beschouwd als 'bijvangst' uit de kwantitatieve onderzoeken in hoofdstuk 5 en hoofdstuk 6. In de hoofdstukken 7 en 8 werden onderzoeken besproken die zich explicieter richtten op de *motieven en ervaringen van gebruikers* van de scoringsrubriek en daardoor veel meer kwalitatief van aard waren.

Uit het onderzoek dat werd beschreven in hoofdstuk 7 bleek dat studenten van de deeltijdopleiding idm van De Haagse Hogeschool de scoringsrubriek vooral waarden vanwege de goede feedback die ze oplevert. Deze studenten bevestigden daarmee de conclusies uit het onderzoek van Oakleaf (2008, p. 245). Verder gebruikten ze het formulier ook in het kader van self-assessment voordat een opdracht ter beoordeling definitief werd ingestuurd en bleek dat ze de principes van de scoringsrubriek ook toepasten in andere situaties, zoals hun werk. Dit onderzoek onder deeltijd studenten gaf dus aan dat de scoringsrubriek niet alleen een functie heeft als beoordelingsinstrument maar ook door studenten zelf wordt gebruikt om hun informatievaardigheden te verbeteren.

De geïnterviewde docenten uit hoofdstuk 8 gaven allen aan dat ze de scoringsrubriek wél hadden aangepast voordat ze haar gebruikten in hun eigen lessen. Ook kwam bij hen duidelijk naar voren dat het instrument het meeste succes sorteert als het vooraf met studenten wordt besproken. De docent die het meest uitgebreid gebruik maakte van de scoringsrubriek gaf het formulier ook een duidelijke plek in het onderwijs door het tevens te gebruiken in het kader van peer-assessment. Daarnaast vertelden de docenten dat ze de scoringsrubriek gebruikten als input bij het formuleren van een leerlijn onderzoeksvaardigheden. Ook bij hen bleek dus dat de scoringsrubriek op verschillende manieren en voor verschillende doelen gebruikt kan worden.

De kwalitatieve onderzoeken uit hoofdstuk 7 en 8 geven aan dat de functie van de scoringsrubriek verder reikt dan alleen het beoordelen van informatievaardigheden. Hoewel de groepen respondenten in deze studies klein waren, gaven de resultaten voldoende aanwijzingen voor het feit dat de scoringsrubriek, mits goed geïmplementeerd, fungeert als een instrument dat leren bevordert en dat tevens kan worden ingezet bij *onderwijsontwikkeling*. De verhalen van de studenten en docen-

ten onderschrijven dat dit niet alleen een theoretische mogelijkheid is zoals die ook in de literatuur wordt beschreven - bijvoorbeeld door Stiggins (2002 en 2005) en Oakleaf (2008) - maar dat de scoringsrubriek ook daadwerkelijk in de praktijk zo werkt.

9.2 Reflectie op de onderzoeksaanpak

Tijdens het onderzoek onder docenten van De Haagse Hogeschool (hoofdstuk 8) bleek dat de scoringsrubriek ook anders werd toegepast dan alleen voor het doel waarvoor zij door de onderzoeker was ontwikkeld (zie ook paragraaf 9.4.3 over het gebruik van de scoringsrubriek bij onderwijsontwikkeling). Hier was dus sprake van voortschrijdend inzicht gedurende het onderzoek. Iets vergelijkbaars vond plaats in de beginfase van het onderzoek. De bestudering van onderwijskundige literatuur en dan met name die over toetsen en beoordelen in het hoger onderwijs, leidde ertoe dat de onderzoeker zijn opvattingen hierover steeds meer nuanceerde en overtuigd raakte van de verwevenheid van toetsen en leren. Dat had uiteindelijk gevolgen voor de nadere eisen die aan het te ontwikkelen instrument werden gesteld (paragraaf 1.3). De aanvankelijke nadruk op summatieve toetsing ontwikkelde zich tijdens het vooronderzoek naar aandacht voor toetsing gedurende het gehele onderwijsproces en dus ook voor formatieve toetsing die geïntegreerd is in het onderwijs zelf (Nitko & Brookhart, 2007, p. 8). Dergelijk voortschrijdend inzicht is overigens een normaal verschijnsel bij kwalitatief onderzoek dat vaak een cyclisch karakter heeft (Wester en Peters, 2004, p. 84). Dat de keuze voor een scoringsrubriek aanvankelijk werd ingegeven door het idee een betrouwbare, valide en praktisch bruikbare beoordelingsmethode te ontwikkelen en dat er pas later in het onderzoek zicht ontstond op de andere voordelen van scoringsrubrieken, komt overeen met de constatering van Reddy en Andrade dat veel docenten zich onvoldoende realiseren dat ze ook gebruikt kunnen worden als een didactisch instrument “to enhance teaching and learning” (Reddy & Andrade, 2010, p. 439).

Het onderzoek dat in dit proefschrift is besproken had naast een cyclisch karakter tevens de kenmerken van een ‘ontwerpgericht’ onderzoek. Voor de opeenvolgende stappen van het ontwerpproces zijn verschillende onderzoeksmethodes gebruikt die zowel kwalitatief als kwantitatief van aard waren.

De nadruk lag in de hoofdstukken 1 tot en met 3 op een theoretische verkenning van het begrip informatievaardigheden en de achtergronden van toetsen en beoordelen in het kader van het hoger onderwijs. Dit kwalitatieve onderzoek mondde in hoofdstuk 4 uit in de constructie van een scoringsrubriek volgens de Step-by-Step Procedure van Mertler (2001) waarbij gebruik werd gemaakt van analyse van literatuur, interviews met experts en eigen observaties tijdens het verzorgen van onderwijs. Het ontwikkelde prototype werd vervolgens in een kwantitatief onderzoek getest. Tijdens de onderzoeken die zijn beschreven in hoofdstuk 5 en 6 werden cijfermatige gegevens verzameld die vervolgens met behulp van statistische procedures werden geanalyseerd. Deze kwantitatieve onderzoeksmethodes leverden resultaten op met betrekking tot betrouwbaarheid en validiteit van de scoringsrubriek, eigenschappen die kenmerkend zijn voor meer klassiek georiënteerd psychometrisch onderzoek. Om meer inzicht te krijgen in andere eigenschappen van de scoringsrubriek - zoals de motieven van studenten en docenten

om er gebruik van te maken en de ervaringen die zij er mee hebben opgedaan - werden de kwantitatieve onderzoeken op hun beurt weer aangevuld met meer kwalitatieve evaluaties op basis van interviews en een focusgroepsgebesprek. Tijdens het gehele onderzoek is met andere woorden gebruik gemaakt van een mixed methods-aanpak waarbij ieder type onderzoek haar eigen soort gegevens opleverde. Deze verschillende soorten gegevens bewezen ieder hun waarde in het onderzoek. De constructie van de scoringsrubriek kon niet succesvol worden afgerond zonder de input uit de interviews van de collega's en de waarnemingen tijdens de practica, voor het testen van het prototype was een kwantitatieve analyse van een omvangrijke set beoordelingen onmisbaar.

9.3 Reikwijdte van het onderzoek

9.3.1 Verhouding tussen het IPS-model en de Scoringsrubriek voor Informatievaardigheden

Met de Scoringsrubriek voor Informatievaardigheden wordt beoordeeld met welk resultaat een student het Information Problem Solving-proces heeft uitgevoerd. Tabel 9.1 geeft aan hoe de afzonderlijke criteria van de scoringsrubriek corresponderen met de verschillende fasen uit het conceptueel kader, het Information Problem Solving-model uit figuur 2.6.

Tabel 9.1

Stappen uit het IPS-model met corresponderende criteria uit de Scoringsrubriek voor Informatievaardigheden

IPS-model	Scoringsrubriek voor informatievaardigheden
1 Onderkennen en verwoorden van de informatiebehoefte	1 Oriëntatie op het onderwerp
2 Zoeken naar informatie	6 Zoektermen
	7 Gebruik van diverse secundaire informatiebronnen
3 Selecteren en beoordelen van informatie	3 Kwaliteit van de gebruikte primaire informatiebronnen
4 Verwerken van informatie	5 Verwerking van relevante informatie tot nieuwe kennis en/of inzichten
5 Vervaardigen en verspreiden van nieuwe informatie	2 Bronnenoverzicht/literatuurlijst
	4 Bronverwijzingen in de tekst

Uit tabel 9.1 blijkt dat de afzonderlijke criteria uit de scoringsrubriek inhoudelijk corresponderen met de verschillende fasen van het Information Problem Solving-model uit figuur 2.6. Wel blijft de beoordeling van stap 5 uit het IPS-model ("Vervaardigen en verspreiden van nieuwe informatie") in de scoringsrubriek beperkt tot het beoordelen van de literatuurlijst (criterium 2 van de scoringsrubriek) en de bronverwijzingen in de tekst (criterium 4). Het beoordelen van schrijfvaardigheid is in de Scoringsrubriek voor Informatievaardigheden niet meegenomen. In hoofdstuk 8 bleek dat dat door minstens één van de geïnterviewde docenten als een tekortkoming was ervaren en dat

zij de scoringsrubriek daarvoor had aangevuld met extra beoordelingscriteria voor schriftelijke taalvaardigheid. Het IPS-model uit figuur 2.6 kent nog een zesde stap in de regulatiebalk. Ook deze zesde stap wordt met de scoringsrubriek niet expliciet beoordeeld.

Dat de criteria uit de scoringsrubriek de stappen van het IPS-model niet één op één volgen komt onder andere door het feit dat de meeste van de criteria die in hoofdstuk 4 voor de scoringsrubriek werden vastgesteld betrekking hebben op de *uitkomsten* van het Information Problem Solving-proces, de kwaliteiten van het product dat studenten maken. Alleen de laatste twee criteria gaan over de uitvoering van afzonderlijke activiteiten en dan specifiek die bij stap 2 van het IPS-model, zoeken naar informatie ofwel het proces van ‘information seeking’ (Wilson, 2000). In de scoringsrubriek gaat het om de kwaliteit van de gebruikte zoektermen (criterium 6) en de secundaire informatiebronnen die zijn gebruikt zoals databases, sites of naslagwerken (criterium 7). Dat proces van informatie zoeken is echter alleen te beoordelen als de studenten daarover ook rapporteren. Daarbij is dan sprake van een vorm van indirecte waarneming door de beoordelaar aan de hand van proces- of reflectieverslagen. In lang niet alle gevallen is het bijhouden van zo’n proces- of reflectieverslag echter onderdeel van de studieopdracht. In zulke situaties zal het vaak zo zijn dat de beoordelingscriteria met betrekking tot het zoekproces niet gebruikt kunnen worden.

9.3.2 Beoordeling van het proces van ‘information seeking’

In paragraaf 3.2.3 zijn diverse andere methodes besproken om het proces van ‘information seeking’ vast te leggen. Die waren allemaal gebaseerd op directe observatie en liepen uiteen van screen-captures en eye-tracking tot ‘hardop denken-protocols’. Voor al deze onderzoeken gold echter dat ze als doel hadden om vast te leggen wat er daadwerkelijk gebeurde zonder dat er voor de deelnemende studenten consequenties aan waren verbonden in de vorm van een beoordeling. Er zijn geen voorbeelden gevonden van dergelijke directe observatiemethodes die succesvol werden toegepast in een authentieke onderwijssituatie.

In de casussen die gebruikt werden in hoofdstuk 5 bleek echter dat ook als de studenten wel gevraagd werd een verantwoording van het zoekproces op te stellen, juist deze criteria 6 (Zoektermen) en 7 (Gebruikt van diverse secundaire bronnen) aanleiding gaven tot misverstanden en verschillende interpretaties. Bij de casus van de Nederlandse Defensie Academie was het gebruik van bepaalde (met naam genoemde) wetenschappelijke databases onderdeel van de instructie bij de opdracht. Dat leidde er toe dat een van de beoordelaars steeds betrekkelijk hoge scores toekende voor het betreffende criterium terwijl de andere beoordelaar zijn aandacht bij dit criterium juist had gericht op de vraag of de student ook andere wetenschappelijk bronnen had gebruikt dan die welke in de opdracht als verplicht te gebruiken vermeld hadden gestaan. Bij de vergelijking van de beoordelingen die in het kader van de casus van de Open Universiteit waren opgemaakt, bleek een vergelijkbaar probleem te spelen: daar was het gebruik van één specifieke bibliografische database verplicht gesteld en bleek criterium 7 niet goed bruikbaar. Op grond van deze ervaringen is tijdens de trainingssessies in het kader van het laatste kwantitatieve onderzoek, dat bij de opleiding Mediastudies van de UvA, juist aan de interpretatie van dit criterium extra aandacht besteed.

Geconcludeerd kan worden dat een beoordeling van het zoekproces pas mogelijk is als de criteria die daarvoor gebruikt worden, goed zijn afgestemd op de instructie die bij de opdracht hoort. De criteria die met betrekking tot het zoekproces in de scoringsrubriek zijn geformuleerd, zijn algemeen van aard. Juist voor die criteria zal in veel gevallen gelden dat ze moeten worden aangepast aan de inhoud van de opdracht die met betrekking tot het zoekproces wordt gehanteerd. Hoe ze dan precies geformuleerd zouden moeten worden, is in zijn algemeenheid niet te zeggen. In het ene geval zal gelden dat de student in ieder geval de sleutelpublicaties moet noemen van waaruit citatie- of sneeuwbalonderzoek is gestart, in een ander geval zal worden gevraagd om een inhoudelijk gegroepeerde lijst van zoektermen of een lijst van zoekmachines / bibliografische databases die zijn gebruikt. Te verwachten is dat de exacte formulering van deze criteria ook per vakgebied zal verschillen. In paragraaf 9.4.2 wordt daarop teruggekomen.

9.3.3 De Scoringsrubriek voor Informatievaardigheden en het Persoonlijk Kennissysteem

Het oplossen van informatieproblemen is in het hoger onderwijs meestal niet een doel op zich zelf. Het proces heeft ook tot doel dat studenten hun kennis vermeerderen, zowel die op het eigen vakgebied als metacognitieve kennis met betrekking tot leren in het algemeen. Met de scoringsrubriek wordt echter niet beoordeeld of er ook daadwerkelijk sprake is van kennisvermeerdering. Om ook dat laatste te beoordelen zou een assessmentmethode nodig zijn die over een langere termijn inzetbaar is en meer de nadruk legt op persoonlijke ontwikkeling van een student. Daarbij valt met name te denken aan een ontwikkelingsgericht portfolio met aandacht voor reflectieactiviteiten door de student. Voor de beoordeling zou dan ook een ander instrument nodig zijn dan de scoringsrubriek voor informatievaardigheden die in dit onderzoek is ontwikkeld. Dat instrument zou zich meer dan de scoringsrubriek dienen te richten op het leervermogen van de student en zou thuis horen in een onderwijskundige context.

De scoringsrubriek zelf zou wel over een langere periode ingezet kunnen worden voor het monitoren van progressie in het Information Problem Solving-proces. In het focusgroeps gesprek dat onderwerp was van hoofdstuk 7, gaven de studenten aan dat zij zelf meenden dat dat niet zo zinvol was ("Je weet van jezelf dat je ergens zwak in bent. Wat is de meerwaarde om dat expliciet te maken?") maar naar aanleiding van die discussie kwam de onderzoeker tot de conclusie dat de Scoringsrubriek voor Informatievaardigheden, in ieder geval in de opleiding idm waarin hij zelf werkzaam is, veel vroeger in het curriculum aan de orde gesteld zou moeten worden. Aandacht ervoor tijdens verschillende fases van de studie zou de meerwaarde bieden dat studenten dan gewend raken om hun eigen progressie op het gebied van informatiegebruik te monitoren en hun leerproces indien noodzakelijk bij te sturen.

9.4 Additionele inzichten

9.4.1 Transfer van informatievaardigheden

Zoals bij de bespreking van de verschillende onderzoeksbenaderingen in hoofdstuk 2 al werd opgemerkt, speelt het idee van 'transfer' binnen de constructivistische onderwijsopvattingen een

belangrijke rol. Volgens Kirschner en Van Merriënboer is juist het gebruik van zogenaamde ‘hele taken’, zoals ook in het onderhavige onderzoek naar het beoordelen van informatievaardigheden is bepleit, bevorderlijk voor het vermogen om de betreffende vaardigheden ook in andere situaties toe te passen (Kirschner & Van Merriënboer, 2013, pp. 7-8).

In het focusgroepsgesprek (hoofdstuk 7) gaven de deeltijd studenten idm aan dat ze de scoringsrubriek niet alleen hadden gebruikt in het blok Digital Library maar dat ze de principes ervan ook toepasten in onderwijsblokken die ze daarna nog volgden en in situaties op hun reguliere werk. In die andere contexten was het gebruik ervan wel meer impliciet. Ze gaven aan dat ze het gebruik van de scoringsrubriek hadden geabstraheerd van de context waarin ze ermee hadden leren werken en de inhoud vervolgens hadden toegepast in vergelijkbare omgevingen. Voor het onderzoek naar het beoordelen en bevorderen van informatievaardigheden is dit een belangrijke conclusie omdat de bevordering van zelfgestuurd en onafhankelijk leren juist een van de argumenten is om in het hoger onderwijs aandacht aan informatievaardigheden te besteden (Bruce, 2004).

Hoe die transfer dan plaatsvindt is in dit onderzoek buiten beschouwing gebleven. Onderzoek van Walraven, Brand-Gruwel en Boshuizen (2008) geeft aan dat een van de theorieën aangaande de transfer van vaardigheden betrekking heeft op de opbouw van metacognitieve kennis zoals die ook wordt uitgedrukt in het Persoonlijk Kennissysteem dat in hoofdstuk 2 werd besproken. In hun onderzoek dat werd uitgevoerd onder scholieren uit de derde klas van het voortgezet onderwijs werd de opbouw van metacognitieve kennis bij een van de groepen gestimuleerd door het invullen van zogenaamde ‘proceswerkbladen’. Daarin staan de leerlingen expliciet stil bij de stappen die ze doorlopen tijdens het uitvoeren van een taak. In dat zelfde onderzoek geven leerlingen echter ook aan dat ze het invullen van zo’n proceswerkblad niet altijd plezierig vinden.

De andere transfertheorie die Walraven et al. behandelen wordt door hen ‘kennisnetwerk theorie’ genoemd. Die gaat er vanuit dat transfer van een vaardigheid wordt gestimuleerd als leerlingen of studenten de onderliggende concepten van zo’n vaardigheid goed kennen. Het lijkt zo te zijn dat de opmerkingen van de idm-studenten over de transfer van informatievaardigheden het best past in die kennisnetwerk theorie: doordat de scoringsrubriek hen expliciet heeft geconfronteerd met de operationalisering van begrippen als ‘betrouwbare informatiebronnen’ en ‘goede onderzoeksvragen’ en de manier waarop die met elkaar samenhangen, waren zij in staat om die kennis ook in een andere context toe te passen. In hun geval was in ieder geval ook van belang dat zij in hun werksituatie gelegenheid hadden om het geleerde toe te passen. Onderzoek van Testers, Gegenfurtner en Brand-Gruwel (2015) toont aan dat die gelegenheid een belangrijke factor is om studenten überhaupt tot transfer te motiveren.

9.4.2 Afstemming van de scoringsrubriek op specifieke disciplines

Een van de argumenten aan het einde van hoofdstuk 3 om te besluiten tot het ontwikkelen van een scoringsrubriek voor informatievaardigheden, was dat er nog geen voorbeeld bekend was van zo’n scoringsrubriek die meer generiek toepasbaar was. Uit de onderzochte literatuur waren al wel voorbeelden van vergelijkbare scoringsrubrieken bekend (Knight, 2006; Oakleaf, 2009) maar deze waren allebei op maat gemaakt en bedoeld om te worden toegepast op een specifieke opdracht of

in een specifieke cursus. Het streven in dit onderzoek om een meer generiek toepasbaar beoordelingsinstrument te ontwikkelen, zou de volgende voordelen opleveren:

- opleidingen of instituten die nog moeten beginnen met het ontwikkelen van onderwijs in informatievaardigheden kunnen zo'n instrument - na eventuele kleine aanpassingen - snel en gemakkelijk implementeren
- over de resultaten van beoordelingen kan worden gecommuniceerd met collega's van andere cursussen / instituten en
- de gevonden gegevens kunnen tussen opleidingen onderling worden uitgewisseld en vergeleken (benchmarking).

Gedurende het onderzoek is een aantal malen naar voren gekomen dat toepassing van de scoringsrubriek in verschillende vakken en onderwijssituaties echter niet altijd vanzelfsprekend is. Tijdens de interviews met de vakdocenten van De Haagse Hogeschool (hoofdstuk 8) bleek dat geen van de vijf respondenten de scoringsrubriek gebruikte zoals zij in dit onderzoek gepresenteerd is. De meesten gebruikten onderdelen van het prototype uit hoofdstuk 4 voor uitbreiding of verfijning van een eigen, bestaand beoordelingsmodel. Dat betekent uiteraard niet dat de scoringsrubriek niet bruikbaar is geweest voor hen. Het gebruik ervan week wel af van het gebruik dat de ontwikkelaar ervan voor ogen stond. Uit het empirisch onderzoek naar het gebruik van de scoringsrubriek bij de Nederlandse Defensie Academie maar ook uit het onderzoek bij de opleiding Mediastudies van de UvA, blijkt dat daar ook nog wel een kanttekening bij kan worden geplaatst. Daar kwam naar voren dat overeenstemming over de beslissing voldoende-onvoldoende in sommige gevallen problematisch was als alleen aan de hand van één afzonderlijk criterium werd beoordeeld. Gebruik van de complete scoringsrubriek leverde wel een betrouwbaar en consistent beeld op. De verklaring die daarvoor werd gegeven was dat mogelijke beoordelingsfouten op afzonderlijke criteria minder zwaar doorwegen indien alle zeven criteria gezamenlijk worden gebruikt. Daaruit kan worden geconcludeerd dat er risico's kleven aan de praktijk om afzonderlijke criteria over te nemen in bestaande beoordelingsinstrumenten indien ze niet kunnen worden gecompenseerd met verwante andere criteria. Dit is feitelijk de reden dat men een construct in het algemeen wil meten met verschillende items; toevallige fouten spelen dan minder een rol en de betrouwbaarheid van de beoordelingen neemt toe.

Aanpassing van de scoringsrubriek kan echter niet alleen plaatsvinden door het laten vervallen van een of meer beoordelingscriteria, ook kunnen meer generieke beschrijvingen specifieker worden gemaakt voor toepassing in het betreffende vakgebied. Tijdens het onderzoek naar het gebruik van de scoringsrubriek bij de faculteit Psychologie en Onderwijswetenschappen van de Open Universiteit kwam naar voren dat de scoringsrubriek voor een beoordelaar moeilijk te hanteren is als deze onbekend is met het vakgebied. In de literatuur gaan dan ook regelmatig stemmen op om het onderwijs in informatievaardigheden stevig in te bedden in de verschillende disciplines (Dirkx et al., 2013; Pinto & Sales, 2015, p. 205). Ook uit onderwijspsychologisch onderzoek is bekend dat de manier waarop mensen informatie verwerken (in de scoringsrubriek criterium nummer 5) per vakgebied verschilt. Zo toont Wineburg (1991) aan dat het oplossen van informatieproblemen in de historische wetenschappen of het geschiedenisonderwijs principieel verschilt van het oplossen van bijvoorbeeld wiskunde problemen of problemen op het gebied van fysica. Daaruit kan wor-

den geconcludeerd dat de scoringsrubriek voor informatievaardigheden kan worden toegepast op verschillende vakgebieden maar dat de exacte hantering ervan vaak nog wel afstemming vergt op die specifieke discipline. In het onderzoek bij de opleiding Mediastudies van de UvA is dat uitgangspunt onder andere aanleiding geweest om de beoordelingen vooraf te laten gaan door twee trainingssessies van ieder ongeveer een halve dag.

De conclusie dient dan ook te luiden dat de scoringsrubriek kan worden toegepast in verschillende contexten maar dat de meer algemene omschrijvingen vaak nog gefinetuned dienen te worden voor gebruik binnen het eigen vakgebied. Dat sluit ook aan bij de opmerking in hoofdstuk 6.7 dat de verwerving van generieke vaardigheden het best plaatsvindt binnen de context van een specifiek beroep of vakgebied. Zoals al bleek bij de bespreking van de manier waarop de scoringsrubriek door vakdocenten gebruikt wordt op De Haagse Hogeschool kan deze dan toch een functie hebben als basis en uitgangspunt voor de formulering van een eigen beoordelingsmodel. Dat geldt bovendien voor het ‘concept’ waar de scoringsrubriek voor informatievaardigheden op is gebaseerd, de beschrijvingen op twee niveaus en vervolgens de vertaalslag naar een score op de Likertschaal. De versie van de scoringsrubriek die beschikbaar is op de HBO Kennisbank is dan ook gepubliceerd onder de Creative Commons licentie ‘Naamsvermelding – Niet Commercieel – Gelijk Delen’ waarmee impliciet wordt aangegeven dat ze voorafgaande aan het hergebruik mag worden bewerkt. Het literatuuronderzoek dat is verricht voor hoofdstuk 7 en 8 suggereert zelfs dat aanpassing van een scoringsrubriek in samenspraak met de studenten zelf de meeste positieve effecten zal hebben op het leerproces van studenten (Andrade, Du & Wang, 2008, p. 4; Reddy & Andrade, 2010, p. 445). De voordelen van een generiek beoordelingsmodel dreigen dan wel verloren te gaan.

9.4.3 Gebruik van de scoringsrubriek voor onderwijsontwikkeling

Het onderzoek heeft zich voornamelijk beperkt tot de ontwikkeling, test en evaluatie van een *beoordelingsinstrument* voor informatievaardigheden. Zoals beschreven in hoofdstuk 8 kan de scoringsrubriek ook fungeren als een startpunt voor het ontwerp en herontwerp van trainings- of instructieprogramma's op het gebied van informatie- en onderzoeksvaardigheden, het formuleren van geschikte studieopdrachten of de integratie van informatievaardigheden in curricula. De beschikbaarheid van de scoringsrubriek alleen is echter niet voldoende om goed onderwijs in informatievaardigheden tot stand te brengen. Dat was ook een boodschap van de deeltijdstudenten tijdens het focusgroeps gesprek dat is beschreven in hoofdstuk 7: als beoordelings- en feedbackinstrument is de scoringsrubriek zeker zinvol, als informatiebron veel minder. Daarvoor geeft de scoringsrubriek te weinig antwoord op de vraag hoe je het oplossen van informatieproblemen wel zou kunnen aanpakken.

Als opleidingen aan de hand van de scoringsrubriek wijzigingen aanbrengen in hun curriculum of onderwijsprogramma, dan moeten zij dus nog altijd nadenken over de vraag hoe het onderwijs met betrekking tot die nieuwe onderwerpen wordt ingericht. Zoals al eerder aangegeven in hoofdstuk 1 dienen leeruitkomsten, te gebruiken leeractiviteiten en toetsvormen immers nauwkeurig op elkaar afgestemd te worden, het idee van “constructive alignment” (Biggs, 2003). Het onderzoek dat is uitgevoerd beperkt zich tot een keuze voor inbedding van de beoordeling van informatievaardig-

heden in de beoordeling van (bestaande) studieopdrachten. Het gaat niet in op de vraag welke onderwerpen dan onderwezen zouden moeten worden en ook niet op de vraag hoe die dan aan de orde gesteld moeten worden. In het Nederlandse taalgebied bestaan inmiddels overigens verschillende onderwijsmethodes die voor het onderwijs in informatievaardigheden kunnen worden gebruikt en die goed aansluiten bij de scoringsrubriek (Van Veen & Westerkamp, 2010; Brand-Gruwel & Wopereis, 2014). In de tweede editie van het boek van Van Veen en Westerkamp is de tekst van het prototype zelfs letterlijk opgenomen (pp. 216-222).

Een van de conclusies van het onderzoek was dan ook dat de ontwikkelde scoringsrubriek niet alleen een antwoord is op de onderzoeksvraag (“Wat is een betrouwbaar, valide en praktisch bruikbaar instrument dat generiek toepasbaar is voor het beoordelen én het bevorderen van de informatievaardigheden van individuele studenten in het hoger onderwijs?”) maar dat ze ook hulp kan bieden bij het ontwerpen van onderwijs en het formuleren van geschikte opdrachten.

9.5 Aanbevelingen voor het onderwijs

Tijdens de interviews met de docenten van De Haagse Hogeschool kwam naar voren hoe lastig het is om systematische aandacht voor het beoordelen van informatievaardigheden te implementeren in het onderwijs. Het is een illusie om te denken dat studenten en docenten een scoringsrubriek (of een andere vorm van een checklist) zullen gebruiken zodra deze is opgenomen in het studiemateriaal of de -handleiding. De praktijk is dat zo'n checklist na verloop van tijd aan de aandacht ontsnapt, vooral als die slechts is opgenomen als ‘bijlage’. De kwantitatieve onderzoeken uit hoofdstuk 5 en hoofdstuk 6 maakten bovendien duidelijk dat het de voorkeur heeft om de nodige tijd aan training en normering te besteden voordat de scoringsrubriek als beoordelingsinstrument wordt ingezet.

Voor vakdocenten geldt vaak echter nog een heel andere bewuste overweging: ze besteden hun tijd (en die van hun studenten) liever aan hun eigen vak dan aan het aanleren van meer algemene vaardigheden (Godwin, 2005, p. 70). Wat dat betreft is er voor veel vakdocenten niet veel verschil tussen het aanleren van informatievaardigheden en de verwerving van andere algemene vaardigheden op het gebied van bijvoorbeeld communicatie, projectmanagement en samenwerking. Allemaal nuttig maar het gaat wel ten koste van de tijd die ook besteed kan worden aan onderwijs in het eigen vak.

Het opsporen en inhoudelijk verwerken van relevante informatie is echter, zoals ook Grafstein (2002) beweert, onderdeel van het domein waarvoor studenten worden opgeleid. Bovendien zorgt de analyse van hoog gekwalificeerde vakliteratuur dat studenten betere leerresultaten behalen op hun eigen vakgebied. De scoringsrubriek is daarom weliswaar ontwikkeld met het doel te functioneren als een generiek model, op basis van de in hoofdstuk 3 besproken literatuur is er wel bewust voor gekozen om te streven naar toepassing binnen verschillende specifieke vakgebieden. De eerdere discussie over de generieke toepasbaarheid van de scoringsrubriek ging juist over het gegeven dat de scoringsrubriek niet een kant en klare oplossing biedt voor het beoordelen van

informatievaardigheden in een heel specifieke situatie maar wel de bouwstenen om die oplossing te creëren. Volgens Van Aken & Andriessen (2011, p. 16) is dat overigens een kenmerk van vrijwel alle resultaten van ontwerpgericht wetenschappelijk onderzoek.

Een van de uitdagingen voor opleidingen die de scoringsrubriek voor informatievaardigheden willen implementeren, is dat van vakdocenten verwacht zal worden dat ze hun studieopdrachten zodanig kiezen en formuleren dat de informatievaardigheden met betrekking tot hun specifieke onderwerpsgebied goed tot hun recht komen. Daarbij kan vooral gedacht worden aan domeinspecifieke informatiebronnen en analysetools die binnen een vakgebied gangbaar zijn. Hoewel vakdocenten in de meeste gevallen wel bekend zullen zijn met de betreffende bronnen en instrumenten, blijken ze voor hun eigen informatievoorziening toch vaak op het internet te vertrouwen. Het gevaar ligt echter op de loer dat ze menen dat ook hun studenten daar voldoende aan hebben. Wat die docenten over het hoofd zien, is dat zij voor het bepalen van hun zoekingen op het internet sterk vertrouwen op de kennis die zij in hun netwerken van collega's en vakgenoten hebben opgedaan en dat het voor hun studenten veel moeilijker is om de betrouwbare vakinformatie te vinden (Godwin, 2005, p. 70).

Een voorbeeld uit de onderzoeksliteratuur waarin wordt beschreven hoe goed onderwijs in informatievaardigheden domeinspecifiek is te maken, is een artikel van Walraven, Brand-Gruwel en Boshuizen (2013). De auteurs beschrijven hoe een instructie op het gebied van geschiedenisonderwijs waarbij gebruik werd gemaakt van 'information problem solving'-taken, zowel leidde tot verbetering van de resultaten op het gebied van informatievaardigheden als van die op het gebied van geschiedenis zelf. In hun geval had de instructie en de evaluatie echter betrekking op een groep docenten en leerlingen uit het derde jaar van het voortgezet onderwijs.

Onderzoeken van Cooney en Hiris (2003) met betrekking tot een business-course en Kesselman en Sherman (2009) op het gebied van voeding ("Food and Nutrition") zijn voorbeelden van studies waarbij informatievaardigheden en vakkennis goed in het *hoger* onderwijs zijn geïntegreerd. In beide case studies is er tevens sprake van een project waarbij docenten en informatieprofessionals intensief samenwerken, in het geval van Cooney en Hiris ook bij de beoordeling van het studentenwerk. Wat dat betreft is het tekenend dat de succesvolle implementatie van de scoringsrubriek op De Haagse Hogeschool juist ook plaatsvond bij een opleiding (business it & management) die intussen eveneens sterk was gaan samenwerken met een informatieprofessional van de hogeschoolbibliotheek. Betrokkenheid van informatieprofessionals bij het onderwijs en bij toetsen en beoordelen van informatievaardigheden kan blijkbaar een positieve bijdrage leveren aan implementatie van dat onderwijs in de opleiding. Een meer cruciale factor lijkt echter nog altijd de vraag te zijn of de verantwoordelijke vakdocenten ook betrokkenheid uitstralen bij de manier waarop informatie wordt opgespoord en verwerkt. Zij zijn binnen de schoolwereld immers meer een rolmodel waar de studenten zich op zullen richten dan de professionals uit de hogeschoolbibliotheek. Dat deze betrokkenheid van de docenten bij de opleiding business it & management wel tot stand kwam, was zeer waarschijnlijk ook te danken aan het feit dat het gebruik van de Scoringsrubriek voor Informatievaardigheden in dat geval sterk werd ondersteund door de manager van de opleiding zelf.

Samenvattend wordt gesteld dat implementatie van de Scoringsrubriek voor Informatievaardigheden in een opleiding dient plaats te vinden op verschillende niveaus:

- Op het niveau van de onderwijseenheid dient de scoringsrubriek te worden geïntegreerd met studieopdrachten op het vakgebied waarvoor de opleiding opleidt. In verschillende onderzoeken wordt aangetoond dat juist zo'n geïntegreerde toepassing leidt tot transfer van de vaardigheden en kennis naar een andere context (Brand-Gruwel & Wopereis, 2006; Oakleaf, 2008; Kirschner & Van Merriënboer, 2013). Bovendien motiveert de integratie in het vakgebied de studenten om aan hun informatievaardigheden te werken. Wel is van belang dat de scoringsrubriek zelf ook onderdeel van de instructie is.
- Op het niveau van het curriculum is het van belang dat aandacht voor informatievaardigheden en de scoringsrubriek doorlopend terugkeert. Op die manier worden informatievaardigheden duurzaam aangebracht en is te verwachten dat ze beklijven, ook nadat de student de opleiding heeft afgerond en de school of universiteit heeft verlaten.
- De docent dient het goede voorbeeld te geven en er voor te zorgen dat het eigen lesmateriaal en de communicatie - zoals e-mails - voldoen aan de normen die in de scoringsrubriek worden gehanteerd. Daarbij kan hij of zij zich laten ondersteunen door informatieprofessionals van de hogeschool- of universiteitsbibliotheek maar zijn / haar eigen rol als voorbeeld voor studenten mag niet worden onderschat.
- Zonder steun van het management is implementatie van de scoringsrubriek in een opleiding niet te realiseren. Docenten worden voortdurend geconfronteerd met vernieuwingen in opleidingsprogramma's en met nieuwe didactische theorieën en hulpmiddelen. Voor al die vernieuwingen geldt dat ze niet kunnen worden doorgevoerd zonder dat de mensen die er mee moeten werken er vooraf in zijn getraind. Ook bij gebruik van de scoringsrubriek bleek dat training en normering vooraf een voorwaarde voor succes is. Als het management dit niet onderkent en hiervoor geen tijd beschikbaar stelt, is het niet erg waarschijnlijk dat implementatie van de scoringsrubriek in een opleiding duurzaam tot stand komt.

9.6 Eindwaardering en suggesties voor nader onderzoek

Alle resultaten en overwegingen overziend, wordt geconcludeerd dat gebruik van de scoringsrubriek - eventueel aangepast voor het betreffende vakgebied - voor het beoordelen van informatievaardigheden kan worden aanbevolen. In de praktijk van het onderwijs kan het gebruik ervan worden verankerd door de aanbevelingen uit de vorige paragraaf ter harte te nemen. Hierna volgt nog een aantal suggesties voor vervolgonderzoek dat het beoordelen en bevorderen van informatievaardigheden van een nog steviger fundament kan voorzien.

9.6.1 Onderzoek naar het beoordelen van het 'Information Problem Solving'-proces

Eerder is opgemerkt dat het bij gebruik van de scoringsrubriek nog niet duidelijk is hoe het 'information problem solving'-proces van de student is verlopen. Beoordelingscriteria 6 ('Zoektermen') en 7 ('Gebruik van diverse secundaire bronnen') hebben wel betrekking op het 'information seeking'-proces maar zijn in praktijk niet altijd te beoordelen omdat daar in de studieopdrachten

niet naar wordt gevraagd. Bovendien is het zoekproces maar een beperkt onderdeel van het volledige ‘information problem solving’-proces dat verder onder andere nog bestaat uit het evalueren van resultatenlijsten, evalueren van afzonderlijke bronnen en het inhoudelijk verwerken ervan. Bij het gebruik van de scoringsrubriek kijkt een beoordelaar voor deze laatste aspecten alleen naar het eindresultaat, het kennisproduct dat de student oplevert.

In het literatuuronderzoek van hoofdstuk 3 maar ook in de discussie eerder in dit hoofdstuk, is een aantal dataverzamelmethodes genoemd waarmee dergelijke processen wel waarneembaar kunnen worden gemaakt. Voorbeelden daarvan zijn: screen-capturing, eye-tracking, het downloaden van logfiles en het gebruik van hardop-denken protocols. De vraag is echter hoe deze methodes ook in een normale onderwijssituatie kunnen worden ingezet waarbij de uitkomsten dan niet alleen een beschrijvend maar ook een beoordelend en mogelijk zelfs een certificerend karakter hebben.

Onderzoek dat daarvoor suggesties doet zal een gat opvullen waarvan gedurende het huidige onderzoek is gebleken dat het bestaat. Vanuit de literatuur over assessments lijkt een “proeve van bekwaamheid” en dan meer specifiek die in de vorm van een simulatie (Dochy & Nickmans, 2005, p. 83) een voor de hand liggende methode om het proces van Information Problem Solving te beoordelen. De leertaken die deel uitmaken van zo’n simulatie kunnen worden gebaseerd op het Information Problem Solving-model dat voor het onderhavige onderzoek is ontwikkeld. Onderzoeksvragen die bij zo’n onderzoek horen zouden kunnen lijken op de hier gebruikte onderzoeksvragen:

- Hoe ziet een simulatie eruit die geschikt is voor het beoordelen en bevorderen van informatievaardigheden bij studenten in het hoger onderwijs?
- Hoe betrouwbaar en valide is de simulatie?
- Hoe ervaren studenten en assessoren het gebruik van de simulatie?

Een kanttekening hierbij is dat Emmett en Emde al eerder gebruik maakten van dergelijke simulaties en dat zij daar na twee jaar van afstapten vanwege de grote tijdsbesteding die deze vorm van beoordelen vergt (Emmett & Emde, 2007).

9.6.2 Informatievaardigheden van docenten

Een tweede suggestie is een onderzoek naar de informatievaardigheden van de docenten zelf. Bij de bespreking van de aanbevelingen voor het onderwijs is opgemerkt dat docenten voor veel studenten een rolmodel zijn. Ze zijn in de ogen van studenten representatief voor de professie waarvoor zij een opleiding volgen. Gezien die rol is het relevant om docenten met behulp van de scoringsrubriek bewust te maken van de manier waarop zij zelf verantwoording afleggen over het gebruik van bestaande kennis bij de informatieproducten die zij creëren zoals syllabi en presentaties, artikelen die zij publiceren of hun tweets en blogposts.

De onderzoeksvraag voor dit onderzoek zou kunnen luiden: hoe beoordelen de docenten van opleiding of faculteit x zelf hun informatievaardigheden als ze hun eigen kennisproducten beoordelen met de Scoringsrubriek voor Informatievaardigheden?

Door bij dit onderzoek te kiezen voor self- of peer-assessment als onderzoeksmethode zal van het gebruik van de scoringsrubriek een stimulerende werking kunnen uitgaan. Uitgangspunt dient niet zo zeer het vaststellen van deficiënties te zijn maar het bewust maken van het eigen gedrag. Ook kunnen docenten op deze wijze ideeën opdoen voor de formulering van nieuwe studieopdrachten en het gebruik van de scoringrubriek tijdens de instructie, bijvoorbeeld als formatief toetsinstrument.

9.6.3 Digitale versie van de scoringsrubriek

Tenslotte is het aan te bevelen om een vervolgonderzoek uit te voeren naar de vorm van de scoringsrubriek. Het prototype is op dit moment beschikbaar als MS Word-document op de HBO-Kennisbank en in een alternatieve versie als onderdeel van een publicatie in InformatieProfessional (Van Helvoort, 2009). Recent is de scoringsrubriek ook opgenomen als een van de modules in het online trainingssysteem HIT (Hogeronderwijs Informatievaardigheden Training) dat door een aantal samenwerkende hogeschoolbibliotheken aan hun studenten en medewerkers wordt aangeboden. Dit betreft echter een read only-versie waarin de beoordelingscriteria en gedragsomschrijving alleen maar worden toegelicht, zonder dat de module kan worden gebruikt om eigen werk of dat van anderen online te beoordelen.

Ontwikkeling van een interactieve digitale versie zou het gebruik ervan wellicht veel aantrekkelijker maken. De constructie van zo'n digitaal beoordelingsinstrument zou bovendien goed aansluiten bij een recent afgesloten project van de opleiding idm van De Haagse Hogeschool dat betrekking had op de ontwikkeling van een zogenaamde 'blended minor' op het gebied van informatievaardigheden. In die minor worden instructies en opdrachten op een speciaal ontwikkeld digitaal platform afgewisseld met wekelijkse fysieke bijeenkomsten. Bij de opdrachten wordt gebruik gemaakt van de scoringsrubriek maar die heeft zelf nog geen digitaal interactief karakter. Een volgende stap zou zijn om uit te zoeken hoe een digitale en interactieve versie van de scoringsrubriek die goed kan worden geïntegreerd met het digitale platform, er uit zou moeten zien. De ervaring leert dat het gebruik van digitale hulpmiddelen om in een eigen tempo en op zelf te kiezen tijdstippen aan studietaken te kunnen werken, zowel ten goede komt aan de aantrekkelijkheid van het onderwijs voor studenten als aan de studieresultaten (Becker & Hiskes, 2014). De digitale scoringsrubriek zou bij voorkeur echter ook beschikbaar worden gesteld als zelfstandig instrument - los van het speciaal ontwikkelde platform - voor de beoordeling van andere studieopdrachten. Voordelen van een interactieve digitale versie zijn, naast de aantrekkelijkheid ervan in het gebruik, de mogelijkheid om direct feedback te geven en de mogelijkheid om beoordelingen voor langere tijd te bewaren en op die manier een portfolio op te bouwen met betrekking tot de voortgang van een student op het gebied van informatievaardigheden.

Bijlage 9.1: Versie 2.0 van de Scoringsrubriek voor Informatievaardigheden**Onderdeel: studentenproduct****Datum:**

Criterion	Professioneel gedrag	Inadequaate gedrag	
1	☐ Uit het product blijkt dat de student zich goed op het vraagstuk of het onderwerp heeft georiënteerd en op basis daarvan een eigen invalshoek (focus) heeft geformuleerd. Dat blijkt ook uit het feit dat de student duidelijk heeft geformuleerd welke (onderzoeks)vragen hij/zij wil beantwoorden.	☐ Uit het product blijkt dat de student is uitgegaan van de oorspronkelijke vraag of de oorspronkelijke formulering van het onderwerp (zoals die zijn gegeven in het assignment) zonder dat hij/zij die verder heeft uitgediept. Dat blijkt bijvoorbeeld uit het feit dat de betekenis van kernbegrippen bekend wordt verondersteld en deze niet nader worden gedefinieerd.	Score 1-20=
Beoordeling: uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht			

Criterion	Professioneel gedrag	Inadequaate gedrag	
2	☐ In de tekst van het product zelf heeft de student eenduidig verantwoord welke informatiebronnen hij of zij heeft gebruikt. Dat geldt ook voor afbeeldingen of voor audiovisuele illustraties in het geval van een digitaal product.	☐ De student heeft werk van anderen (stukken tekst, afbeeldingen, geluidsfragmenten) verwerkt in zijn/haar eigen product zonder dat hij/zij de gebruikte bron heeft vermeld. Hoewel vaak niet opzettelijk, is er strikt gesproken sprake van 'plagiat'.	Score 1-10=
Beoordeling: uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht			

Criterion	Professioneel gedrag	Inadequaate gedrag	
3	☐ Het product bevat een bronnenoverzicht dat volledig is en waarbij een standaard citation style correct is toegepast. Aan de hand van het bronnenoverzicht zijn de gebruikte bronnen eenduidig te achterhalen. Toelichting: het laatste is belangrijker dan het format van de titelbeschrijvingen (de 'citation style'). Voor de score 'uitstekend' dient ook de citation style echter correct te zijn toegepast.	☐ In het product ontbreekt een bronnenoverzicht en / óf ☐ het bronnenoverzicht is verre van volledig (van verschillende documenten waar in de tekst naar wordt verwezen ontbreken de gegevens) en / óf ☐ cruciale bibliografische gegevens (titel, auteur, publicatiejaar) ontbreken. Voorbeeld dat vaak voorkomt in de praktijk: van internetbronnen zijn alleen de url's vastgelegd.	Score 1-20=
Beoordeling: uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht			

Criterion	Professioneel gedrag	Inadequaate gedrag	
4	☐ Uit het bronnenoverzicht dat bij het product behoort, blijkt dat de student gebruik heeft gemaakt van relevante, betrouwbare (bij voorkeur zo authentiek mogelijke) en actuele informatiebronnen die het onderwerp bovendien vanuit verschillende gezichtspunten behandelen.	☐ De gebruikte externe bronnen zijn te licht, verouderd of onvoldoende relevant. Een voorbeeld van 'te licht' is als de student alleen internetsites als informatiebron heeft gebruikt. En / óf ... ☐ De gebruikte literatuur is eenzijdig. Er is bijvoorbeeld alleen gebruik gemaakt van overheidsinformatie (.gov-sites) of van publicaties van één bepaalde auteur.	Score 1-10=
Beoordeling: uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht			

Criterium		Professioneel gedrag	Inadequaat gedrag	
5	Verwerking van relevante informatie tot nieuwe kennis en/of inzichten	☐ Uit het product blijkt dat de student informatie uit verschillende bronnen heeft geanalyseerd en op basis daarvan nieuwe toepassingen, inzichten of hypothesen heeft geformuleerd. Toelichting: in praktijk blijkt dat veel studenten wel toekomen aan de analyse en vergelijking van verschillende informatiebronnen, maar niet in staat zijn om de gevonden gegevens te synthetiseren tot een nieuw inzicht of nieuwe toepassing. Dit criterium wordt in dat geval beoordeeld als 'voldoende' of 'zwak'.	In het studentenproduct heeft de student ☐ de inhoud van de gebruikte informatiebronnen niet correct of niet helder weergegeven, en / óf ☐ in het geheel geen aandacht besteed aan analyse van de opgespoorde informatiebronnen, en / óf ☐ zich gebaseerd op slechts één informatiebron en hij/zij heeft de daar gevonden informatie niet ter discussie gesteld terwijl daar wel aanleiding toe is.	Score 1-20=
Beoordeling:		uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht		

Scoringsrubriek voor Informatievaardigheden

Onderdeel: procesverslag / zoekstrategie

Criterium		Professioneel gedrag	Inadequaat gedrag	
6	Zoektermen	☐ De student heeft de zoektermen gebruikt die relevant zijn voor het te onderzoeken onderwerp / vraagstuk. Hij/zij heeft daarbij rekening gehouden met het vakjargon, de synoniemen en (zoek)termen in een andere taal (meestal Engels).	☐ De zoektermen die de student heeft gebruikt zijn te algemeen of komen uit het algemeen spraakgebruik.. en / óf ☐ de student heeft geen rekening gehouden met synoniemen, verwante begrippen en (zoek)termen in een andere taal (meestal Engels).	Score 1-10=
Beoordeling:		uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht		

Criterium		Professioneel	Inadequaat gedrag	
7	Gebruik van diverse secundaire informatiebronnen	☐ De student heeft gebruik gemaakt van diverse typen secundaire bronnen (search engines, boeken, wetenschappelijke tijdschriften, databases, sociale netwerken). Indien van toepassing heeft hij/zij bovendien gebruik gemaakt van mogelijkheden buiten de eigen locatie (IBL, documentleverantie).	☐ De student heeft vooral informatiebronnen gebruikt die gemakkelijk toegankelijk zijn. Voorbeeld: de student heeft alleen gebruik gemaakt van • het internet via het 'eenvoudige zoek-scherm' van een grote algemene zoekmachine en / of • lesmateriaal dat door docenten is verstrekt.	Score 1-10=
Beoordeling:		uitstekend – zeer goed – goed – ruim voldoende – voldoende zwak – onvoldoende – ruim onvoldoende – slecht – zeer slecht		

Totaalscore (maximaal 100) =



Op dit werk is een Creative Commons licentie van toepassing:
 Naamsvermelding – Niet commercieel – Gelijk delen
 Jos van Helvoort – De Haagse Hogeschool

Referenties

- ACRL (2000). *Information literacy competency standards for higher education*. Chicago: American Library Association. Geraadpleegd op 1-2-2009 op <http://www.ala.org/ala/mgrps/divs/acrl/standards/standards.pdf>.
- ACRL (2015). *Framework for information literacy for higher education*. Geraadpleegd op 16 februari 2015 op <http://acrl.ala.org/ilstandards/wp-content/uploads/2015/01/Framework-MW15-Board-Docs.pdf>.
- Adviesraad voor het Wetenschaps- en Technologiebeleid (2005). *Ontwerp en ontwikkeling: De functie en plaats van onderzoeksactiviteiten in hogescholen*. (Adviezen van de AWT nr. 65). Geraadpleegd op 13 juli 2012 op <http://www.awt.nl/upload/documents/publicaties/tekst/a65.pdf>.
- Aken, J. v. & Andriessen, D. (2011). *Handboek ontwerpgericht wetenschappelijk onderzoek*. Den Haag: Boom Lemma uitgevers.
- ALA (1989). *Presidential committee on information literacy: Final report*. Chicago: American Library Association. Geraadpleegd op 31 januari 2009 op <http://www.ala.org/ala/mgrps/divs/acrl/publications/whitepapers/presidential.cfm>.
- Alexander, B. (2006). Web 2.0: A new wave of innovation for teaching and learning. *Educause review*, 2006 (March / April), 33-44.
- Alexander, P., Schallert, D. & Hare, V. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61 (3), 315-343.
- Allen, J. & Velden, R. v. d. (2011). The flexible professional in the knowledge society: Required competences and the role of higher education. In: J. Allen & R. van der Velden (Red.), *The flexible professional in the knowledge society: New challenges for higher education* (pp. 15-53). Dordrecht etc.: Springer.
- Anderson, C. (2006). *The long tail : Why the future of business is selling less of more*. New York: Hyperion.
- Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, 4 (4). Geraadpleegd op 25 juni 2010 op <http://cie.asu.edu/volume4/number4/>.
- Andrade, H. & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10 (3), 1-11. Geraadpleegd op 25 juni 2010 op <http://www.pareonline.net/pdf/v10n3.pdf>.

- Andrade, H., Du, Y. & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27 (2), 3-13.
- Ballator, N., Farnum, M. & Kaplan, B. (1999). *NAEP 1996 trends in writing: Fluency and writing conventions*. Washington, DC: National Center for Education Statistics.
- Becker, P. & Hiskes, A. (2014). Blended learning: A new approach for teaching information literacy. *5th International Symposium on Information Management in a Changing World*. Antalya, Turkije. 26 november 2014.
- Bell, D. (1976a). The coming of the post-industrial society. *The Educational Forum*, 40 (4), 575-579.
- Bell, D. (1976b). Welcome to the post-industrial society. *Physics Today*, 29 (2), 46-49.
- Berkowitz, R. & Eisenberg, M. (1990). *Information problem-solving : The Big Six Skills approach to library & information skills instruction*. Norwood, N.J.: Ablex Pub. Corp.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32 (3), 347-364.
- Biggs, J. (2003). Aligning teaching for constructing learning. *Higher Education Academy*. Geraadpleegd op 1 november 2013 op http://www.bangor.ac.uk/adu/the_scheme/documents/Biggs.pdf
- Bloom, B. (Red.) (1956). *Taxonomy of educational objectives : The classification of educational goals. Handbook I, cognitive domain*. London: Longman.
- Boekhorst, A. (2000). *Informatievaardig worden in het onderwijs, een informatiewetenschappelijk perspectief: Een vergelijkende gevallenstudie in Nederland en Zuid-Afrika*. Pretoria: University of Pretoria.
- Boekhorst, A., Koers, D. & Kwast, I. (1999). *Informatievaardigheden*. Utrecht: Lemma.
- Boekhorst, A., Kwast, I. & Wevers, D. (2004). *Informatievaardigheden*. (3^e druk). Utrecht: Lemma.
- Boer, F. d. (2006). Mixed methods: Een nieuwe methodologische benadering. *Kwalon*, 11 (2).
- Bolton, F. (2006). Rubrics and adult learners: Andragogy and assessment. *Assessment Update*, 18 (3), 5-6.
- Borgdorff, H., Staa, A. v. & Vos, J. v. d. (2007). Kennis in context: Onderzoek aan hogescholen. *Th&ma, Tijdschrift voor Hoger Onderwijs & Management*, 14 (5), 10-17.
- Brand-Gruwel, S. & Wopereis, I. (2006). Integration of the information problem-solving skill in an educational programme: The effects of learning with authentic tasks. *Technology, Instruction, Cognition, and Learning*, 4, 243-263.
- Brand-Gruwel, S. & Wopereis, I. (2014). *Word informatievaardig! Digitale informatie selecteren, beoordelen en verwerken*. (2^e druk). Groningen/Houten: Noordhoff Uitgevers.

- Brand-Gruwel, S., Wopereis, I. & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21 (3), 487-508.
- Brown, A. & Gratch-Lindauer, B. (2004). Developing a tool to assess community college students. In: I. Rockman & Associates (Red.), *Integrating information literacy into the higher education curriculum: Practical models for transformation* (pp. 165-206). San Francisco: Jossey-Bass.
- Bruce, C. (1997). *The seven faces of information literacy*. Adelaide: Auslib Press.
- Bruce, C. (1998). The phenomenon of information literacy. *Higher Education Research & Development*, 17 (1), 25-43.
- Bruce, C. (1999). Workplace experiences of information literacy. *International journal of information management*, 19 (1), 33-47.
- Bruce, C. (2000). Information literacy research : Dimensions of the emerging collective consciousness. *Australian Academic and Research Libraries (AARL)*, 31 (2), 91-109. Geraadpleegd op 9 juni 2012 op <http://eprints.qut.edu.au/46274/>.
- Bruce, C. (2004). Information literacy as a catalyst for educational change: A background paper. *Proceedings "Lifelong Learning: Whose Responsibility and what is Your Contribution?"*, the 3rd International Lifelong Learning Conference. Yeppoon, Queensland.
- Buchanan, L., Luck, D. & Jones, T. (2002). Integrating information literacy into the virtual university: A course model. *Library Trends*, 51 (2), 144-166.
- Bundy, A. (Ed.) (2004). *Australian and New Zealand information literacy framework: Principles, standards and practice* (2nd ed.). Adelaide: Australian and New Zealand Institute for Information Literacy.
- Burkhardt, J., Kinnie, J. & Cournoyer, C. (2008). Information literacy successes compared: Online vs. face to face. *Journal of Library Administration*, 48 (3), 379-389.
- Cameron, L., Wise, S. & Lottridge, S. (2007). The development and validation of the information literacy test. *College & Research Libraries*, 68 (3), 229-236.
- Castells, M. (1996). *The rise of the network society*. Malden, Mass.: Blackwell Publishers.
- Castells, M. (2010). *The rise of the network society*. (2e druk). Chichester: Wiley-Blackwell.
- Center for Advanced Research on Language Acquisition : University of Minnesota (2012). *Types of rubrics: Primary trait and multiple trait*. Geraadpleegd op 27 december 2012 op <http://www.carla.umn.edu/assessment/VAC/Evaluation/rubrics/types/traitRubrics.html>.
- Centraal Bureau voor de Statistiek (2012). *Arbeidsvolume*. Geraadpleegd op 20 januari 2014 op <http://www.cbs.nl/nl-NL/menu/themas/arbeid-sociale-zekerheid/publicaties/arbeidsmarkt-vogelvlucht/structuur-arbeidsmarkt/2006-arbeidsmarkt-vv-arbeidsvolume-art.htm>.

- Cherry, B., Fiegen, A. & Watson, K. (2002). Reflections on collaboration: Learning outcomes and information literacy assessment in the business curriculum. *Reference Services Review*, 30 (4), 307-318.
- Cheuk, B. (1998). An information seeking and using process model in the workplace: A constructivist approach. *Asian Libraries*, 7 (12), 375-390.
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6 (4), 284.
- Cochrane, C. (2006). Embedding information literacy in an undergraduate management degree: Lecturers' and students' perspectives. *Education for Information*, 24 (2), 97-123.
- Cooney, M. & Hiris, L. (2003). Integrating information literacy and its assessment into a graduate business course: A collaborative framework. *Research Strategies*, 19 (3-4), 213-232.
- Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Decoo, W. (2003). "Heb jij dat zelf geschreven?" : Plagiaat bij de internetgeneratie. *Info-Frans*, 30 (4), 19-32. Geraadpleegd op 1 november 2013 op http://webh01.ua.ac.be/DIDASCALIA/downloads%5COnlineartikels%5CDecoo%5CDecoo_Plagiaat.pdf.
- Deursen, A. v. & Dijk, J. v. (2008). *Digitale vaardigheden van Nederlandse burgers: Een prestatie-meting van operationele, formele, informatie en strategische vaardigheden bij het gebruik van overheidswebsites*. Enschede: Universiteit van Twente.
- Dierick, S., Watering, G. v. d. & Muijtens, A. (2002). De actuele kwaliteit van assessment: Ontwikkelingen in de edumetrie. In: F. Dochy, L. Heylen & H. van de Mosselaer (Red.), *Assessment in onderwijs: Nieuwe toetsvormen en examinering in studentgericht onderwijs en competentiegericht onderwijs* (pp. 91-122). Utrecht: Lemma.
- Dijk, J. v. (2001). *De netwerkmaatschappij : Sociale aspecten van nieuwe media*. Alphen aan den Rijn: Samsom.
- Dijk, J. v. (2011). Social media in de netwerkmaatschappij. In: D. van Osch & R. van Zijl (Red.), *Basisboek social media* (pp. 15-43). Den Haag: Boom Lemma uitgevers.
- Diller, K. & Phelps, S. (2008). Learning outcomes, portfolios, and rubrics, oh my! Authentic assessment of an information literacy program. *Portal: Libraries and the Academy*, 8 (1), 75-89.
- Dirkx, A., Drent, M., Jonkers, J., Knippenberg, H., Vermijs, H., Walraven, A., et al. (2013). *Leren omgaan met wetenschappelijke informatie: Informatievaardigheid in het Nederlandse wetenschappelijk onderwijs*. Ongepubliceerd manuscript. Geraadpleegd op 6 september 2013, op http://annekedirkx.files.wordpress.com/2013/09/informatievaardigheid-hoger-onderwijs-versie-maart-2013_2.pdf.
- Dochy, F. & McDowell, L. (1997). Assessment as a tool for learning. *Studies In Educational Evaluation*, 23 (4), 279-298.

- Dochy, F. & Nickmans, G. (2005). *Competentiegericht opleiden en toetsen: Theorie en praktijk van flexibel leren*. Utrecht: Lemma.
- Dochy, F., Segers, M. & Rijdt, C. d. (2002). Nieuwe ontwikkelingen: De assessmentcultuur. In: F. Dochy, L. Heylen & H. van de Mosselaer (Red.), *Assessment in onderwijs: Nieuwe toetsvormen en examinering in studentgericht en competentiegericht onderwijs* (pp. 11-31). Utrecht: Lemma.
- Doyle, C. (1992). *Outcome measures for information literacy within the National Education Goals of 1990. Final Report to National Forum on Information Literacy. Summary of Findings*. Geraadpleegd op 6 november 2015 op <http://files.eric.ed.gov/fulltext/ED351033.pdf>.
- Driessen, E. & Vleuten, C. v. d. (2000). Matching student assessment to problem-based learning: Lessons from experience in a law faculty. *Studies in Continuing Education*, 22 (2), 235-248.
- Drigas, A. & Tsolaki, M. (2015). Lifelong learning and ICTs. *International Journal of Recent Contributions from Engineering, Science & IT*, 3 (2), 15-20.
- Drucker, P. (1992). Be data literate: know what to know. *Wall Street Journal*, 1992 (3 december). Geraadpleegd op 15 juni 2012 op <http://azcost.net/article3.htm>.
- Drucker, P. (2001a). The new workforce. *The Economist*. Geraadpleegd op 15 juni 2013 op <http://www.economist.com/node/770847>.
- Drucker, P. (2001b). The next society. *The Economist*. Geraadpleegd op 15 juni 2013 op <http://www.economist.com/node/770819>.
- Dunn, K. (2002). Assessing information literacy skills in the California State University: A progress report. *The Journal of Academic Librarianship*, 28 (1-2), 26-35.
- Eisenberg, M. (1996). *Comparison of information skills process models*. Geraadpleegd op 20 april 2009 op <http://www.shambles.net/ITinset/Appendix/ISmod.html>.
- Elshout-Mohr, M. & Oostdam, R. (2001). *Assessment van competenties in een dynamisch curriculum*. Amsterdam: Universiteit van Amsterdam, SCO-Kohnstamm Instituut. (SCO-rapport nr. 605).
- Elshout-Mohr, M., Oostdam, R. & Overmaat, M. (2002). Student assessment within the context of constructivist educational settings. *Studies In Educational Evaluation*, 28 (4), 369-390.
- Emmett, A. & Emde, J. (2007). Assessing information literacy skills using the ACRL standards as a guide. *Reference Services Review*, 35 (2), 210-229.
- European Commission (2014). *Lifelong Learning Programme 2007-2013*. Geraadpleegd op 30 november 2015 op http://eacea.ec.europa.eu/llp/index_en.php.
- European Commission (2015). *Online trust*. Geraadpleegd op 11 november 2015 op <http://ec.europa.eu/digital-agenda/en/online-trust>.

- Everitt, B. (2002). *Cambridge dictionary of statistics*. (2e druk). Cambridge [etc.]: Cambridge University Press.
- Fagerheim, B. & Shrode, F. (2009). Information literacy rubrics within the disciplines. *Communications in Information Literacy*, 3 (2), 158-170.
- Floyd, D. M., Colvin, G. & Bodur, Y. (2008). A faculty–librarian collaboration for developing information literacy skills among preservice teachers. *Teaching and Teacher Education*, 24 (2), 368-376.
- Fourie, I. & Niekerk, D. v. (1999). Using portfolio assessment in a module in research information skills. *Education for Information*, 17 (4), 333-352.
- Fourie, I. & Niekerk, D. v. (2001). Follow-up on the use of portfolio assessment for a module in research information skills: An analysis of its value. *Education for Information*, 19 (2), 107-126.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American psychologist*, 39 (3), 193-202.
- Godwin, P. (2005). Making life easier for academics: How librarians can help staff weather the technological storm. *Journal of eLiteracy*, 2 (2), 68-79.
- Gormally, C., Brickman, P. & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education*, 11, 364-377.
- Grafstein, A. (2002). A discipline-based approach to information literacy. *Journal of Academic Librarianship*, 28 (4), 197-204.
- Gratch-Lindauer, B. (2003). Selecting and developing assessment tools. In: E. Avery (Red.), *Assessing student learning outcomes for information literacy instruction in academic institutions* (pp. 22-39). Chicago: American Library Association.
- Green, R. & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, 45 (1-2), 185-202.
- Gross, M. & Latham, D. (2007). Attaining information literacy: An investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library and Information Science Research*, 29 (3), 332-353.
- Haan, J. d. & Huysmans, F. (2006). Informatievaardigheden in een kennissamenleving. *Investeren in vermogen* (pp. 91-115). Den Haag: Sociaal en Cultureel Planbureau. Geraadpleegd op 23-11-2012 op http://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2006/Investeren_in_vermogen.
- Hallgren, K. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8 (1), 23.

- Hamelink, C. (1989). Informatisering van de samenleving. In: Cees Hamelink & Jaap van der Stel (Red.), *Informatisering van de samenleving* (pp. 13-26). Groningen: Boekwerk.
- Helvoort, J. v. (2009). Informatiegebruik bij studenten gemeten. *InformatieProfessional*, 2009 (10), 30-35.
- Helvoort, J. v. (2010). A scoring rubric for performance assessment of information literacy in Dutch Higher Education. *Journal of Information Literacy*, 4 (1), 22-39.
- Helvoort, J. v. (2012a). Competenties van reference librarians nu en in de toekomst. *InformatieProfessional*, 2012 (3), 26-29.
- Helvoort, J. v. (2012b). How adult students in information studies use a scoring rubric for the development of their information literacy skills. *The Journal of Academic Librarianship*, 38 (3), 165-171.
- Helvoort, J. v. (2013). How faculty in The Hague University of Applied Sciences uses the scoring rubric for information literacy. *Communications in Computer and Information Science*, 397, 436-442.
- Helvoort, J. v. (2014). The personal knowledge base conception of information literacy. *Communications in Computer and Information Science*, 492, 31-36.
- Hignite, M., Margavio, T. & Margavio, G. (2009). Information literacy assessment: moving beyond computer literacy. *College Student Journal*, 43 (3), 812-821.
- Hogeschool van Amsterdam (2014). *Toetscyclus : Beoordelen*. Geraadpleegd op 6 mei 2015 op <https://score.hva.nl/docent/toetscyclus/Paginas/Beoordelen.aspx>.
- Hook, S., Bracke, M., Greenfield, L. & Mills, V. (2003). In-house training for instruction librarians. *Research Strategies*, 19 (2), 99-127.
- Irving, A. (1985). *Study and information skills across the curriculum*. London: Heinemann educational books.
- Ito, H. (2015). Is a rubric worth the time and effort? Conditions for its success. *International Journal of Learning, Teaching and Educational Research*, 10 (2), 32-45.
- Ivanitskaya, L., O'Boyle, I. & Casey, A. (2006). Health information literacy and competencies of information age students: Results from the interactive online Research Readiness Self-Assessment (RRSA). *Journal of Medical Internet Research*, 8 (2). Geraadpleegd op 17 januari 2003 op <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550696/>.
- Janssen-Noordman, A. & Merriënboer, J. v. (2002). *Innovatief onderwijs ontwerpen: Via leertaken naar complexe vaardigheden*. Groningen;Houten: Wolters-Noordhoff.
- Jenkins, A. & Healy, M. (2005). *Institutional strategies to link teaching and research*. Heslington (York, UK): The Higher Education Academy. Geraadpleegd op 6 april 2014 op http://textweb.livjm.ac.uk/partnership/collab_partner_docs/pf_jan_07_martyn_stewart_jenkins_and_healey.pdf.

- Johnson, H. (2003). The SCONUL task force on information skills. In: Allan Martin & Hannelore B. Rader (Red.), *Information and IT literacy: Enabling learning in the 21st century* (pp. 45-52). London: Facet.
- Johnston, B. & Webber, S. (2003). Information literacy in higher education: A review and case study. *Studies in Higher Education*, 28 (3), 335-352.
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144.
- Kelly, K. (1998). *New rules for the new economy : 10 radical strategies for a connected world*. New York: Viking Penguin.
- Keshav, S. (2007). How to read a paper. *ACM SIGCOMM Computer Communication Review*, 37 (3), 83-84.
- Kesselman, M. & Sherman, A. (2009). Linking information to real-life problems: An interdisciplinary collaboration of librarians, departments, and food businesses. *Journal of Agricultural & Food Information*, 10 (4), 300-318.
- Ketelaar, P., Hentenaar, F. & Kooter, M. (2011). *Groepen in focus : In vier stappen naar toegepast focusgroeponderzoek*. Den Haag: Boom Lemma uitgevers.
- Kirschner, P. & Berg, E. v. d. (2012). Pas op: Hypes! In: Leo Plugge (Red.), *De bakens verzetten*. Wetenschappelijk Technische Raad SURF. Geraadpleegd op 28 oktober 2012 op <http://www.wtrtrendrapport2012.nl/hoofdstukken.php?h=5>.
- Kirschner, P. & Merriënboer, J. v. (2013). *Ten steps to complex learning: A systematic approach to four-component instructional design*. (2e druk). New York: Routledge.
- Knight, L. (2006). Using rubrics to assess information literacy. *Reference Services Review*, 34 (1), 43-55.
- Kolowich, S. (2011). What students don't know. *Inside Higher ED*, 2011 (22 augustus). Geraadpleegd op 23 augustus 2011 op http://www.insidehighered.com/news/2011/08/22/erial_study_of_student_research_habits_at_illinois_university_libraries_reveals_alarmingly_poor_information_literacy_and_skills.
- Kuhlthau, C. (1993). *Seeking meaning : A process approach to library and information services*. Norwood, N.J.: Ablex.
- Kuhlthau, C. & Tama, S. (2001). Information search process of lawyers: A call for 'just for me' information services. *Journal of Documentation*, 57 (1), 25-43.
- Kurbanoglu, S., Akkoyunlu, B. & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation*, 62 (6), 730-743.
- Kvavik, R. (2005). Convenience, communications, and control: How students use technology. In: D. Oblinger & J. Oblinger (Red.), *Educating the net generation* (pp. 7.1-20). [S.L.]: Educause. Geraadpleegd op 12 maart 2008 op www.educause.edu/educatingthenetgen.

- Landers, R. (2011). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *NeoAcademic*, 2011 (16 november). Geraadpleegd op 16 november 2013 op <http://neoacademic.com/2011/11/16/computing-intraclass-correlations-icc-as-estimates-of-interrater-reliability-in-spss/>.
- Lazonder, A. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology*, 36 (3), 465-475.
- Lee, A. & So, C. (2014). Media literacy and information literacy: Similarities and differences. *Communicar*, 42 (XXI), 137-145.
- Liefsoens, L. (2013). Personal information management: Oplossing voor het 'rommelig-bureau-syndroom'. *InformatieProfessional*, 2013 (1), 24-28.
- Limberg, L. (1999). Experiencing information seeking and learning: A study of the interaction between two phenomena. *Information Research*, 5 (1). Geraadpleegd op 16 mei 2008 op <http://informationr.net/ir/5-1/paper68.html>.
- Mediawijzer.net Over Mediawijzer.net. Geraadpleegd op 30 april 2014 op <http://www.mediawijzer.net/over-mediawijzer/>.
- Merriënboer, J. v., Clark, R. & Croock, M. d. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50 (2), 39-61.
- Merriënboer, J. v., Klink, M. v. d. & Hendriks, M. (2002). *Competenties: Van complicaties tot com-promis: Over schuifjes en begrenzers*. Den Haag: Onderwijsraad.
- Mertler, C. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7 (25), 72-81.
- Mittermeyer, D. & Quirion, D. (2003). *Information literacy: Study of incoming first-year undergraduates in Québec*. Montréal: CREPUQ. Geraadpleegd op 20 januari 2013 op <http://library.concordia.ca/services/users/faculty/infolit/infolit-crepuq.pdf>.
- Monoï, S., O'Hanlon, N. & Diaz, K. (2005). Online searching skills: Development of an inventory to assess self-efficacy. *The Journal of Academic Librarianship*, 31 (2), 98-105.
- Moskal, B. & Leydens, J. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7 (10). Geraadpleegd op 30 maart 2008 op <http://pare-online.net/getvn.asp?v=7&n=10>.
- Murphy, E. (1997). *Constructivist epistemology*. Geraadpleegd op 2 december 2012 op <http://www.ucs.mun.ca/~emurphy/stemnet/cle2.html>.
- Neely, T. (2006). *Information literacy assessment: Standards-based tools and assignments*. Chicago: American Library Association.
- Nitko, A. & Brookhart, S. (2007). *Educational assessment of students*. (5^e druk). Upper Saddle River, N.J.: Pearson Education.

- Novotny, E. & Cahoy, E. (2006). If we teach, do they learn? The impact of instruction on online catalog search strategies. *Portal: Libraries and the Academy*, 6 (2), 155-167.
- Nuffic (2012). *Competenties en leeruitkomsten*. Geraadpleegd op 1 december 2012 op <http://www.nuffic.nl/expertise/bolognaproces/achtergrondinformatie/competenties-en-leeruitkomsten>.
- Nutefall, J. (2005). Paper Trail: One method of information literacy assessment. *Research Strategies*, 20 (1-2), 89-98.
- Oakleaf, M. (2008). Dangers and opportunities: A conceptual map of information literacy assessment approaches. *Portal: Libraries and the Academy*, 8 (3), 233-253.
- Oakleaf, M. (2009a). The information literacy instruction assessment cycle: A guide for increasing student learning and improving librarian instructional skills. *Journal of Documentation*, 65 (4), 539-560.
- Oakleaf, M. (2009b). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, 60 (5), 969-983.
- Oakleaf, M. (2014). A roadmap for assessing student learning using the new Framework for Information Literacy for Higher Education. *The Journal of Academic Librarianship* 40 (5), 510-514.
- Partridge, H. (2007). *Establishing the human perspective of the information society*. Brisbane, Australia: Queensland University of Technology.
- Perruso Brown, C. & Kingsley-Wilson, B. (2010). Assessing organically: Turning an assignment into an assessment. *Reference Services Review*, 38 (4), 536-556.
- Petty, R. & Cacioppo, J. (1981). The theory of reasoned action. In: Richard Petty & John Cacioppo (Red.), *Attitudes and persuasion: Classic and contemporary approaches* (pp. 193-212). Dubuque, Iowa: W.C. Brown Company Publishers.
- Pinto, M. & Sales, D. (2015). Uncovering information literacy's disciplinary differences through students' attitudes: An empirical study. *Journal of Librarianship and Information Science*, 47 (3), 204-215.
- Pintrich, P. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 40 (4), 219-225.
- Radcliff, C., Jensen, M., Salem Jr., J., Burhanna, K. & Gedeon, J. (2007). *A practical guide to information literacy assessment for academic librarians*. Westport, Conn.: Libraries Unlimited.
- Reddy, Y. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35 (4), 435-448.
- Rockman, I. (2002). Strengthening connections between information literacy, general education, and assessment efforts. *Library Trends*, 51 (2), 185-198.

- Roth, L. (1999). Educating the cut-and-paste generation. *Library Journal*, 124 (18), 42-44.
- SAILS (2012). *Sample questions from our information literacy assessment*. Geraadpleegd op 3 januari 2013 op <https://www.projectsails.org/SampleQuestions>.
- Schaap, H., de Bruijn, E., Van der Schaaf, M. & Kirschner, P. (2009). Students' personal professional theories in competence-based vocational education: The construction of personal knowledge through internalisation and socialisation. *Journal of Vocational Education and Training*, 61 (4), 481-494.
- Scharf, D., Elliot, N., Huey, H. A., Briller, V. & Joshi, K. (2007). Direct assessment of information literacy using writing portfolios. *Journal of Academic Librarianship*, 33 (4), 462-477.
- Schoolrapport. (n.d.). *Wikipedia*. Geraadpleegd op 13 juni 2015 op <http://nl.wikipedia.org/wiki/Schoolrapport>.
- Shirky, C. (2010). *Cognitive surplus : How technology makes consumers into collaborators*. New York, London: Penguin Press.
- Shoaf, E. (2003). Using a professional moderator in library focus group research. *College and Research Libraries*, 64 (2), 124-132.
- Sieverts, E. (2013). Meer informatie en beter zoeken: Ook centraal in 2023. In: G. J. van Bussel (Red.), *De informatiemaatschappij van 2023: Perspectieven op de nabije toekomst*. Amsterdam: Hogeschool van Amsterdam. Geraadpleegd op 7 oktober 2013 op <http://sieverts.pbworks.com/f/Zoeken2023.pdf>.
- Smith, K. & Tillema, H. (2003). Clarifying different types of portfolio use. *Assessment & Evaluation in Higher Education*, 28 (6), 625-648.
- Sonley, V., Turner, D., Myer, S. & Cotton, Y. (2007). Information literacy assessment by portfolio: A case study. *Reference Services Review*, 35 (1), 41-70.
- Stemler, S. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9 (4). Geraadpleegd op 17 september 2008 op <http://www.pareonline.net/getvn.asp?v=9&n=4>.
- Stevens, D. & Levi, A. (2013). *Introduction to rubrics : An assessment tool to save grading time, convey effective feedback, and promote student learning*. (2^e druk). Sterling, Va.: Stylus Pub.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83 (10), 758-765.
- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87 (4), 324-328.
- Stoffberg, E. & Blignaut, A. (2008). A case for multimodal training of electronic databases at a higher education institution. *South African Journal of Libraries & Information Science*, 74 (1), 1-8.

- Straetmans, G. (2006). *Bekwaam beoordelen en beslissen: Beoordelen in competentiegerichte beroepsopleidingen*. Enschede: Saxion Hogescholen.
- Streatfield, D. & Markless, S. (2008). Evaluating the impact of information literacy in higher education: Progress and prospects. *Libri*, 58 (2), 102-109. Geraadpleegd op 17 januari 2013 op <http://www.librijournal.org/pdf/2008-2pp102-109.pdf>.
- Sundin, O. (2008). Negotiations on information-seeking expertise: A study of web-based tutorials for information literacy. *Journal of Documentation*, 64 (1), 24-44.
- Testers, L., Gegenfurtner, A. & Brand-Gruwel, S. (2015). Motivation to transfer learning to multiple contexts. *The School Library Rocks: Living it, Learning it, Loving it: 44th International Association of School Librarianship International Conference, Incorporating the 19th International Forum on Research in School Librarianship*. Maastricht. 28 juni - 2 juli 2015.
- Thijssen, J. (2001). Loopbaanontwikkeling in verandering: Post-industriële achtergronden, persoonlijke ontwikkelingsplannen en (leer) competenties. *Opleiding en Ontwikkeling*, 14 (11), 19-27.
- Thirion, P. & Pochet, B. (2009). Information literacy in students entering higher education in the French speaking community of Belgium: Lessons learned from an evaluation. *IFLA Journal*, 35 (2), 152-170.
- Thompson, C. (2011). Why kids can't search. *Wired*, 2011 (1 november). Geraadpleegd op 6 september 2013 op http://www.wired.com/magazine/2011/11/st_thompson_searchresults/.
- Tigelaar, D., Dolmans, D., Wolfhagen, I. & Vleuten, C. v. d. (2005). Quality issues in judging portfolios: Implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education*, 30 (5), 595-610.
- Timmers, C. & Glas, C. (2010). Developing scales for information-seeking behaviour. *Journal of Documentation*, 66 (1), 46-69.
- Tweede Fase Adviespunt (2005). *Zeven jaar tweede fase, een balans: Evaluatie tweede fase*. Den Haag: Tweede Fase Adviespunt. Geraadpleegd op 1 november 2013 op http://www.nvon.nl/sites/nvon.nl/files/oud/tf_2007/2005_10_zeven%20jaar%20tweede%20fase%20een%20balans.pdf.
- UASnet (zonder datum). *Country report in the framework of the Balama study: The Netherlands*. UASnet. Geraadpleegd op 15 juni 2012 op <https://docs.google.com/file/d/0B841P4yDYEDsNDkyZjQ2ZTgtZmFmZS00MzY4LTg5NTQtZWMyY2I2NzBjNmFj/edit>.
- Unesco (2003). *Prague declaration: Towards an information literate society*. Geraadpleegd op 16 juli 2015 op <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/PragueDeclaration.pdf>.
- Veen, M. v. & Westerkamp, K. (2010). *Deskresearch*. (Tweede editie). Amsterdam: Pearson Education.
- Velden, R. v. d. (2011). Generiek of specifiek opleiden? *Tijdschrift voor Arbeidsvraagstukken*, 27 (4), 380-397.

- Vermunt, J. & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, 9 (3), 257-280.
- Verschuren, P. (2005). *De probleemstelling voor een onderzoek*. (9^e druk). Utrecht: Uitgeverij het Spectrum.
- Voogt, J. & Pareja Roblin, N. (2010). *21st century skills: Discussienota*. Enschede: Universiteit Twente.
- Walhout, J., Brand-Gruwel, S. & Martens, R. (2011). Op weg naar een model voor het leren organiseren van digitale informatie. *39th Research Onderwijs Dagen (ORD)*. Maastricht. Juni 2011.
- Walraven, A. (2008). *Becoming a critical websearcher: Effects of instruction to foster transfer*. [Heerlen]: Open Universiteit.
- Walraven, A., Brand-Gruwel, S. & Boshuizen, E. (2008). Internetinformatie leren beoordelen: Twee onderwijsmethodes vergeleken. *OnderwijsInnovatie*, 2008 (1), 17-25.
- Walraven, A., Brand-Gruwel, S. & Boshuizen, H. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, 52 (1), 234-246.
- Walraven, A., Brand-Gruwel, S. & Boshuizen, H. (2013). Fostering students' evaluation behaviour while searching the internet. *Instructional Science*, 41 (1), 125-146.
- Walsh, A. (2009). Information literacy assessment: Where do we start? *Journal of Librarianship and Information Science*, 41 (1), 19-28.
- Watering, G. v. d. & Dierick, S. (2002). Kwaliteit van assessment: De bruikbaarheid van klassieke toetsen binnen studentgericht- en competentiegericht onderwijs. In: F. Dochy, L. Heylen & H. Van de Mosselaer (Red.), *Assessment in onderwijs: Nieuwe toetsvormen en examinering in studentgericht onderwijs en competentiegericht onderwijs* (pp. 61-89). Utrecht: Lemma.
- Webber, S. & Johnston, B. (2000). Conceptions of information literacy: New perspectives and implications. *Journal of Information Science*, 26 (6), 381-397.
- Webster, F. (2014). *Theories of the information society*. (4^e druk). London;New York: Routledge.
- Weel, A. v. d. (2012). Van waardeketen naar waardeweb. In: Arianne Baggerman, Paul Hoftijzer, Gerard Unger & Adriaan van der Weel (Red.), *Schakels in de keten van het boek: Boeketje boekwetenschap II* (pp. 29-35). Den Haag; Amsterdam: Dr. P.A. Tiele-Stichting; AUP. Geraadpleegd op 27 april 2014 op <http://hdl.handle.net/1887/20451>.
- Wester, F. & Peters, V. (2004). *Kwalitatieve analyse : Uitgangspunten en procedures*. Bussum: Coutinho.
- Wetenschappelijke Raad voor het Regeringsbeleid (2013). *Naar een lerende economie: Investeren in het verdienvermogen van Nederland*. Amsterdam: Amsterdam University Press. Geraadpleegd op 10 maart 2014 op http://www.wrr.nl/fileadmin/nl/publicaties/PDF-Rapporten/2013-11-01__WRR_Naar_een_lerende_economie.pdf.

- Wilson, E. & Blankenship, J. (2010). Integrating information literacy instruction in an upper-division writing-intensive class. *The Southeastern Librarian*, 58 (3), 19-27.
- Wilson, T. (2000). Human information behavior. *Informing Science*, 3 (2), 49-56.
- Windham, C. (2006). *Getting past Google : Perspectives on information literacy from the Millennial mind*. [S.L.]: Educause. Geraadpleegd op 12 april 2008 op <http://connect.educause.edu/Library/ELI/GettingPastGooglePerspect/39341?time=1208003180>.
- Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83 (1), 73-87.
- Zimmaro, D. (2004). *Developing grading rubrics*. Geraadpleegd op 6 mei 2015 op http://slo.sbccc.edu/wp-content/uploads/developing_grading_rubrics.pdf.

Beoordelen van informatievaardigheden in het hoger onderwijs

1 Probleemstelling en context van het onderzoek

Informatievaardigheden (in het Engels “Information Literacy”) hebben betrekking op iemands bekwaamheid in het gebruik van informatie waarbij de volgende deelvaardigheden worden onderscheiden:

- onderkennen van een informatiebehoefte
- formuleren van een zoekvraag
- selecteren van geschikte informatiebronnen (zoals het internet, een digitaal bestand of een bibliotheekcollectie)
- gebruik van ICT bij het raadplegen van de informatiebronnen
- selecteren en beoordelen van de informatie die is gevonden
- verwerken (analyseren en synthetiseren), hergebruiken en verspreiden van de gevonden informatie (Boekhorst, 2000, p. 103; Webber & Johnston, 2000, p. 382).

Sinds de jaren ‘80 is het belang van informatievaardigheden voor deelname aan de samenleving sterk toegenomen. Ten eerste omdat de productie, verwerking en de distributie van informatie een steeds belangrijker rol spelen in de economie maar ook omdat de toegang tot informatie sterk is gedemocratiseerd. Informatie-eindgebruikers dienen steeds vaker zelf over de vaardigheden te beschikken om de weg te vinden in de informatieovervloed, terwijl zij voorheen konden vertrouwen op poortwachters zoals uitgevers en bibliotheken. Bovendien zijn ontwikkelingen in de informatie- en communicatietechnologie van dien aard dat de beroepskennis van professionals steeds sneller veroudert wat er eveneens voor zorgt dat zij dienen te beschikken over het vermogen om zelfstandig te leren op basis van elders beschikbare informatie.

In het hoger onderwijs is een vergelijkbare ontwikkeling gaande: de gangbare didactiek hecht steeds meer waarde aan zelfstandig leren waarbij gebruik wordt gemaakt van informatieverwerking door de lerende zelf. Onderwijsvormen waarbij de constructie van kennis door de lerenden centraal staat, staan bekend onder namen als probleemgestuurd onderwijs, projectonderwijs en competentiegericht onderwijs. Deze ontwikkeling in onderwijsvormen, maar ook het gegeven dat ook aan het hbo inmiddels een behoorlijke rol wordt toegekend bij het uitvoeren van vraaggericht onderzoek, heeft er voor gezorgd dat de verwerving van informatievaardigheden in het hoger onderwijs een centrale plaats verdient. Als die rol van informatievaardigheden serieus genomen wordt, zal er niet alleen aandacht aan besteed moeten worden in het onderwijs maar dient ook de toetsing daarmee in overeenstemming te worden gebracht. Die overtuiging is aanleiding geweest tot het onderzoek naar het beoordelen van informatievaardigheden dat in dit

proefschrift wordt besproken. Het onderzoek zelf richtte zich op de constructie en het testen van een algemeen toepasbaar assessmentinstrument maar ook de vraag naar de manier waarop zo'n instrument in de praktijk gebruikt wordt, is in het onderzoek betrokken. Gezien de hedendaagse onderwijskundige inzichten waarbij toetsing niet beperkt blijft tot de selectie en kwalificatie van studenten maar ook een rol speelt tijdens het leerproces zelf, is in het onderzoek bovendien de vraag meegenomen hoe zo'n assessmentinstrument de verwerving van informatievaardigheden kan bevorderen. De volledige onderzoeksvraag kwam daarmee als volgt te luiden: Wat is een betrouwbaar, valide en praktisch bruikbaar instrument dat generiek toepasbaar is voor het beoordelen én het bevorderen van de informatievaardigheden van individuele studenten in het hoger onderwijs?

2 Wetenschappelijk onderzoek naar informatievaardigheden

In het wetenschappelijk onderzoek naar informatievaardigheden worden drie benaderingen onderscheiden waarvan er twee relevant zijn voor dit onderzoek. De eerste relevante is een benadering die de nadruk legt op de te bereiken leeruitkomsten. Omdat die leeruitkomsten in de betreffende benadering worden gedefinieerd in termen van waarneembaar gedrag, wordt dit ook wel de “behavioural approach” (in het Nederlands “gedragsbenadering”) genoemd. De uitkomstgerichte gedragsbenadering van informatievaardigheden is in het kader van het eigen onderzoek relevant omdat de beoogde leeruitkomsten een belangrijke rol spelen bij het ontwerpen van een beoordelingsinstrument.

De tweede relevante benadering is de procesgerichte constructivistische. In deze benadering ligt de nadruk meer op het proces van het oplossen van informatieproblemen dan op de te bereiken resultaten. Kenmerkend voor deze benadering is dat men ervan uitgaat dat een goede beheersing van het informatieverwerkingsproces bijdraagt aan het leerproces van studenten in het algemeen: door het oplossen van informatieproblemen wordt tevens nieuwe kennis geconstrueerd. Deze opvatting van informatievaardigheden sluit meer dan de uitkomstgerichte benadering aan bij hedendaagse opvattingen over leren die in het hoger onderwijs worden gehanteerd (bijvoorbeeld in het kader van competentiegericht leren, projectonderwijs en probleemgestuurd leren).

In het onderhavige onderzoek speelt de uitkomstgerichte benadering vooral een rol bij de operationalisering van het construct informatievaardigheden en de nadere invulling van ‘informatievaardig gedrag’. Voor de keuze van een geschikt beoordelingsinstrument wordt veel meer aangesloten bij de procesgerichte benadering van informatievaardigheden, die zelf immers veel beter aansluit bij de gangbare ideeën over onderwijs en leren. Een kenmerk daarvan is onder andere dat beoordelen niet alleen een rol speelt aan het einde van het leerproces (bij het onderscheiden van goed en minder goed presterende studenten) maar ook gedurende het leerproces zodat studenten tijdig feedback ontvangen over hun voortgang en de onderdelen waar ze (extra) aandacht aan dienen te besteden

Behalve naar wetenschapstheoretische benaderingen van het fenomeen informatievaardigheden wordt in hoofdstuk 2 tevens gekeken naar verschillende modellen die in de context van onderzoek naar informatievaardigheden worden gehanteerd. Daarbij wordt onderscheid gemaakt tussen twee verschillende facetten:

- het vermogen om een informatieprobleem op te lossen aan de hand van een specifieke voorliggende (studie)taak (vaak ook aangeduid als “Information Problem Solving”) en
- de constructie van metacognitieve kennis (“Information Literacy as a way of learning”; Bruce, 2000).

Voor informatievaardigheden opgevat als “Information Problem Solving” is uiteindelijk een model opgesteld dat teruggrijpt op een model van Brand-Gruwel, Wopereis en Vermetten (2005) die zich zelf weer lieten inspireren door het Big Six Skills-model van Berkowitz en Eisenberg. Het vermogen om informatieproblemen op te lossen, wordt daar voorgesteld als een complexe vaardigheid die is samengesteld uit 6 stappen en verschillende deelactiviteiten die gecoördineerd en geïntegreerd uitgevoerd dienen te worden. Dat laatste wordt uitgedrukt door de horizontale regulatiebalk, het punt waarop het model van Brand-Gruwel, Wopereis en Vermetten zich onderscheidt van dat van Berkowitz en Eisenberg en dat een belangrijk argument was om het model van Brand-Gruwel et al. te gebruiken als uitgangspunt voor het eigen model.

Voor de constructie van metacognitieve kennis is gebruik gemaakt van het idee van een Personal Knowledge Base (Bruce, 1998), in figuur 2.7 aangeduid als een Persoonlijk Kennissysteem. Dit construct van een persoonlijk kennissysteem is aan de hand van wetenschappelijke literatuur over leertheorieën verder geoperationaliseerd, waarbij weer onderscheid wordt gemaakt tussen de kennis die behoort bij een vakgebied en vakoverstijgende metacognitieve kennis.

Het model van het Information Problem Solving-proces uit figuur 2.6 heeft tijdens het volledige onderzoek gefungeerd als conceptueel kader voor het onderzoek naar de wijze waarop de informatievaardigheden van studenten kunnen worden gemeten.

3 Meten van informatievaardigheden

Voor het meten van informatievaardigheden worden in de literatuur verschillende soorten beoordelingsinstrumenten onderscheiden: tests, enquêtes, studieopdrachten, rechtstreekse observatie, logboeken en portfolio's. Tests en enquêtes zijn minder geschikt voor het toetsen van het werkelijke gebruik van informatie door studenten. Ze zijn daarvoor teveel gericht op wat studenten weten en begrijpen (tests) of op hetgeen studenten zeggen dat ze doen (enquêtes), zonder dat zichtbaar wordt hoe ze zich daadwerkelijk gedragen. Bij enquêtes is juist ook het gevaar van sociaal wenselijke antwoorden aanwezig.

Een alternatieve vorm van beoordelen is die met behulp van authentieke opdrachten die hun wortels hebben in de beroepspraktijk. Dergelijke toetsen zijn dan niet gericht op het meten van deelvaardigheden maar beoordelen de uitvoering van een complexe taak waarbij de student zijn aandacht moet verdelen over verschillende facetten en deeltaken tegelijk. Dergelijke vormen van toetsen worden veel gebruikt bij de alternatieve onderwijsvormen die zijn onderscheiden in

hoofdstuk 1 (probleemgestuurd onderwijs, projectonderwijs en competentiegericht onderwijs). Bijkomend voordeel zou zijn dat de vaardigheden die worden opgedaan in dergelijke realistische arbeidssituaties ook makkelijker getransfereerd kunnen worden naar andere studie- of werksituaties.

In de literatuur worden voor dergelijke vormen van “performance assessment” verschillende mogelijkheden genoemd. Om te beginnen zijn dat studieopdrachten waarbij het informatiegebruik wordt beoordeeld aan de hand van de eindproducten (bijvoorbeeld de opgenomen literatuur) of aan de hand van een reflectie over het onderzoeksproces zelf. Het werk aan zo’n studieopdracht kan ook worden gedocumenteerd met behulp van een portfolio. Behalve het eindproduct, een logboek en een reflectie kunnen daarin dan ook nog beoordelingen door derden worden opgenomen. Nog beter is natuurlijk als studenten een portfolio over langere tijd bijhouden zodat ook de ontwikkeling van de student inzichtelijk wordt. Nadeel van reflecties en logboeken is echter dat bij deze vormen van beoordeling, net als bij enquêtes, niet het daadwerkelijke gedrag wordt beoordeeld maar slechts datgene wat de student zegt dat hij / zij doet. Directe observatie van het informatiegebruik is wel mogelijk, onder andere door gebruik te maken van beeldschermopnamen en hardop denken-protocollen, maar is zo arbeidsintensief dat deze in de literatuur nergens als beoordelingsmethode wordt beschreven. Wel wordt directe observatie gebruikt als onderzoeksmethode bij met name experimenteel onderzoek.

In de hierboven beschreven vormen van performance assessment van informatievaardigheden wordt steeds benadrukt dat een checklist voor het vaststellen van de beoordeling zelf een onmisbaar hulpmiddel is. Een meer geavanceerde vorm van zo’n checklist is een zogenaamde ‘analytische scoringsrubriek’. Dergelijke “rubrics” bevatten zowel de criteria zelf voor de beoordeling van een bepaalde taak, als de normen die worden gehanteerd bij het gebruik van die criteria. Ze worden steeds vaker toegepast bij het beoordelen van de uitvoering van authentieke beroepstaken en horen dus echt thuis in de sfeer van performance assessments. In de literatuur over het beoordelen van informatievaardigheden worden aan analytische scoringsrubrieken onder andere de volgende voordelen toegekend (Oakleaf, 2008):

- ze verminderen de subjectiviteit van de beoordelingen doordat beoordelaars systematisch kijken aan de hand van een vastgestelde lijst met criteria en duidelijk omschreven normen
- ze zijn geschikt voor het geven van feedback
- als ze vooraf worden verspreid en worden besproken bieden ze de studenten inzicht in wat er van hen verwacht wordt
- studenten kunnen een scoringsrubriek ook gebruiken om elkaars werk en het eigen werk te beoordelen (peer- en self-assessment).

Om een antwoord te vinden op de onderzoeksvraag die in dit onderzoek centraal staat, is de keuze gemaakt voor het ontwerp van een zo authentiek mogelijke vorm van toetsing. Op basis van de literatuurstudie is gekozen voor de constructie van een analytische scoringsrubriek die gebruikt kan worden voor het beoordelen van het Information Problem Solving-proces bij de uitvoering van beroep gerelateerde kennistaken. Dat betekent onder andere dat de beoordeling van informatie-

vaardigheden wordt ingebed in het domeinspecifieke onderwijs. Er is gekozen voor een *analytische* scoringsrubriek omdat een dergelijk beoordelingsinstrument bij uitstek ook geschikt is voor het *bevorderen* van vaardigheden in het gebruik van informatie.

4 Constructie van de scoringsrubriek

De scoringsrubriek zelf is geconstrueerd in zeven stappen die zelf weer waren verdeeld over 3 ‘fasen’:

Verkennde fase

1. Vaststellen van de gewenste criteria
2. Beschrijving van het ‘professionele gedrag’ voor ieder criterium
3. Beschrijving van het ‘inadequaat gedrag’ voor ieder criterium

Uitwerking

1. Keuze voor het aantal niveaus dat gebruikt gaat worden in de scoringsrubriek
2. Beschrijving van het gedrag op de tussenliggende niveaus voor ieder criterium

Testfase

1. Testen van de eerste concept versie met uitwerkingen van studenten op verschillende niveaus
2. Aanpassing van de scoringsrubriek als dat gewenst is.

Tijdens de verkennende fase is in een iteratief proces gebruik gemaakt van literatuuronderzoek (met name de ACRL standaarden zijn geanalyseerd), monitoring van het zoekgedrag van studenten gedurende practica en consultatie van collega's over eerste ontwerpen van de scoringsrubriek. Voor de keuze van de gewenste criteria is uiteindelijk de nadruk komen liggen op kenmerken van het eindproduct en minder op de beschrijving van het gevolgde proces, hoewel dat laatste nog wel tot uitdrukking kwam in de laatste van de twee gekozen criteria. Argument voor de nadruk op het eindproduct was dat het oplossen van informatieproblemen een betrekkelijk generieke vaardigheid is waarbij het reëel is om studenten, binnen bepaalde grenzen, zelf te laten kiezen hoe ze een opdracht aanpakken.

De normen voor het ‘professionele gedrag’ zijn met name ontleend aan de literatuur, de beschrijving van het inadequate gedrag is vooral gebaseerd op waarnemingen tijdens de practica. Uiteindelijk zijn de formuleringen van de diverse criteria en gedragsbeschrijvingen ook besproken in review sessies met collega's van de opleiding idm en docenten van drie andere opleidingen, waarna ze opnieuw zijn geformuleerd om te komen tot een conceptversie die kon worden getest.

Het eerste concept van de scoringsrubriek kende gedragsbeschrijvingen op drie niveaus (tabel 4.1). De resultaten van een test waarbij 15 studentproducten met dit eerste concept werden beoordeeld door twee verschillende beoordelaars, waren aanleiding om het te reviseren. Het definitieve prototype kent voor 7 criteria beschrijvingen op twee niveaus, aangevuld met een 6 punts Likertschaal (bijlage 4.1).

5 Evalueren van de bruikbaarheid bij Saxion; Testen van betrouwbaarheid en validiteit bij de Nederlandse Defensie Academie en Open Universiteit

In hoofdstuk 5 wordt beschreven hoe het prototype is getest in drie verschillende contexten:

- de opleiding Commerciële Economie van Saxion te Enschede (School of Marketing & International Management)
- de opleiding Militaire Bedrijfs Wetenschappen van de Nederlandse Defensie Academie en
- de cursus Informatievaardigheden voor sociale wetenschappers van de Open Universiteit (faculteit Psychologie en onderwijswetenschappen).

Bij Saxion is de scoringsrubriek gebruikt door acht tutoren die er individuele procesverslagen mee beoordeelden op het aspect 'informatiegebruik'. Zeven van de acht tutoren meenden dat de criteria uit de scoringsrubriek gezamenlijk een goede en een volledige afspiegeling vormen van het construct informatievaardigheid. Eveneens zeven van hen ervoeren de scoringsrubriek als een goed hulpmiddel om goed gefocust te blijven tijdens het beoordelen van de verslagen. Op de vraag of ze de scoringsrubriek in de toekomst nog eens zouden gebruiken voor een summatie (selecterende) beoordeling antwoordde de meerderheid echter negatief of met een twijfel. Over het gebruik in het kader van instructie of formatieve toetsing waren 6 van 8 respondenten wel positief.

Bij de Nederlandse Defensie Academie beoordeelden twee docenten informatiekunde ieder de zelfde 27 werkstukken van studenten, in dit geval literatuuroverzichten. In het kader van de test van de scoringsrubriek werd in kaart gebracht in welke mate de beoordelingen van de twee afzonderlijke beoordelaars met elkaar overeenstemden en in welke mate er consistentie was tussen de beoordelingen van iedere beoordelaar afzonderlijk. Het bleek dat de twee docenten het betrekkelijk zelden precies met elkaar eens waren over de beoordeling met de afzonderlijke criteria op de Likertschaal maar dat ze ook niet veel van elkaar verschilden. De zogenaamde aanpalende overeenkomst (niet meer dan één punt verschil op de 6 punts Likertschaal) was aanzienlijk (voor alle criteria op één na 85% of hoger). Ook de eindbeoordelingen op een schaal van 1-10 bevestigden dit beeld van een voldoende aanpalende overeenstemming. De consensus over de beslissing wel of niet geslaagd (pass/fail-decision) was voor de eindbeoordelingen zelfs 93%. Voor drie afzonderlijke criteria was de consensus met betrekking tot de beslissing wel of niet geslaagd teleurstellend (minder dan 75%). Dit was aanleiding om te concluderen dat het gebruik van afzonderlijke criteria voor summatieve beoordelingen niet aan te raden is. Tevens werd geconcludeerd dat de betrekkelijk hoge betrouwbaarheid van de eindbeoordelingen wordt bereikt doordat mogelijke beoordelingsfouten op de afzonderlijke criteria met elkaar worden gecompenseerd als alle zeven criteria gezamenlijk worden gebruikt.

De Item-Totaal Statistieken die met SPSS werden vastgesteld, gaven aan dat vooral beoordelaar 1 in staat was om met behulp van de scoringsrubriek beoordelingen op te maken die een heel hoge mate van interne consistentie vertonen.

De twee beoordelaars van de Nederlandse Defensie Academie waren het eens met de meerderheid van de respondenten van Saxion dat de zeven criteria gezamenlijk een goede afspiegeling vormen

van het construct informatievaardigheid en dat de opzet van de scoringsrubriek hen hielp om bij het beoordelen goed te blijven letten op de relevante aspecten.

De opzet van de test met de casus van de Open Universiteit volgde in grote lijnen die welke gebruikt was bij de Nederlandse Defensie Academie. Van de cursus Informatievaardigheden voor sociale wetenschappers werden 19 dossiers met betrekking tot een literatuuronderzoek en bijbehorend essay beoordeeld door twee beoordelaars, één beoordelaar van de Open Universiteit en de onderzoeker zelf. Ook bij deze casus gold dat de aanpalende overeenstemming voor de meeste criteria voldoende was, met uitzondering van die voor criterium 2, de kwaliteit van het bronnenoverzicht. Verder bleek dat vooral de onderzoeker zelf moeite had om de verschillende criteria onderling consistent toe te passen. Zo bleken de beoordelingen op criterium 6 (de kwaliteit van de zoektermen) behoorlijk af te wijken van die op de andere criteria. Een mogelijke verklaring daarvoor is zijn onbekendheid met het vakgebied waar het literatuuronderzoek betrekking op had.

Een verschil met de casus van de Nederlandse Defensie Academie was dat bij de Open Universiteit de zelfde dossiers ook beoordeeld werden met een model dat bij de Open Universiteit reeds eerder in gebruik was. Dat gaf de mogelijkheid om ook een onderzoek te doen naar de correlatie tussen de beoordelingen met het OU-model en die met de scoringsrubriek. Deze samenhang bleek voor beide beoordelaars aanwezig: $r=0,68$ voor beoordelaar 1 ($p<0,01$) en $r=0,55$ voor beoordelaar 2 ($p<0,05$).

6 Testen van betrouwbaarheid en validiteit bij de opleiding Mediastudies van de Universiteit van Amsterdam

Bij de opleiding Mediastudies van de Universiteit van Amsterdam werden 80 uitwerkingen van een studieopdracht door twee beoordelaars ieder twee maal beoordeeld: één maal met de scoringsrubriek en één maal met een beoordelingsmodel van de opleiding zelf. De twee beoordelaars waren zelf recent afgestudeerd aan dezelfde opleiding. Ze waren in het verleden zelf bij de opleiding betrokken geweest in de rol van student assistent maar hadden geen ervaring met het beoordelen van werkstukken.

De 80 uitwerkingen zijn random geselecteerd uit een totaal van 120. Het aantal van 80 was groot genoeg om de interbeoordelaarsbetrouwbaarheid niet alleen vast te stellen aan de hand van het percentage aanpalende overeenstemming maar ook met het statistische kengetal ICC(3). Deze Intraclass Correlation Coefficient was voor bijna alle beoordelingscriteria van de scoringsrubriek hoger dan 0,60 wat betekent dat de interbeoordelaarsbetrouwbaarheid 'goed' is. Voor de eindcijfers kwam de ICC(3) zelfs maar juist onder de excellente waarde van 0,75. Alleen voor beoordelingscriterium 1 ('Oriëntatie') kwam ICC(3) uit onder 0,60. Een verklaring voor de matige overeenstemming bij dit criterium is dat het inderdaad wat meer vrijheid geeft aan beoordelaars om er een eigen interpretatie aan te geven. Beoordelingscriterium 6 van de scoringsrubriek (de gebruikte 'zoektermen') kon bij de test niet worden betrokken omdat in de opdracht aan studenten niet voldoende duidelijk was gevraagd om daarover te rapporteren.

Behalve naar de interbeoordelaarsbetrouwbaarheid werd bij de casus van de opleiding Mediastudies ook onderzoek gedaan naar de constructvaliditeit en de criteriumvaliditeit van de Scoringsrubriek voor Informatievaardigheden. Cronbach's Alpha was voor allebei de beoordelaars hoger dan 0,8 en ook de Item-Totaal Statistieken gaven aan dat de interne consistentie van hun beoordelingen als goed kan worden beschouwd. Samen met de uitkomsten van de exploratieve factoranalyse die werd uitgevoerd, was dat een bevestiging van het feit dat de zes gebruikte beoordelingscriteria alle verwijzen naar één en het zelfde construct. De criteriumvaliditeit werd net als bij het onderzoek naar de beoordelingen bij de Open Universiteit vastgesteld door voor de afzonderlijke beoordelaars de correlatie te berekenen tussen de eindbeoordelingen met de scoringsrubriek en die met het beoordelingsmodel van de opleiding zelf. De vastgestelde correlaties van $r=0,76$ ($p<0,01$) en $r=0,93$ ($p<0,01$) worden beschouwd als hoog respectievelijk zeer hoog.

Naar aanleiding van het gebruik van de Scoringsrubriek voor Informatievaardigheden gaven de twee beoordelaars van de opleiding Mediastudies de volgende feedback. Ze vonden de scoringsrubriek gemakkelijker en efficiënter in gebruik dan het model van de opleiding zelf. Wel zouden ze zelf de voorkeur geven aan een 10 punts Likertschaal in plaats van de toegepaste 6 punts schaal en deden ze de suggestie om de presentatievolgorde van de beoordelingscriteria in de scoringsrubriek te wijzigen. Daarbij ging het er dan vooral om dat criterium 4 ('bronverwijzingen in de tekst') en criterium 2 ('bronnenoverzicht / literatuurlijst') elkaar in de presentatie direct opeen zouden moeten volgen.

7 Evalueren met studenten uit de deeltijdopleiding idm van De Haagse Hogeschool

Ominzichttekrijgen in de wijze waarop studenten de Scoringsrubriek voor Informatievaardigheden (zouden kunnen) gebruiken, werd een onderzoek uitgevoerd onder studenten van de opleiding idm die in een eerder stadium een workshop hadden bijgewoond waarin de scoringsrubriek was toegelicht en waarin ze er kort mee gewerkt hadden. Vijf studenten van de deeltijd opleiding idm werden ongeveer een half jaar na het bijwonen van die workshop geïnterviewd in een focusgroepsgesprek.

De studenten gaven aan dat ze de scoringsrubriek vooral gebruikt hadden als een checklist om hun eigen werk te beoordelen voordat ze het instuurden ter beoordeling. Deze vorm van self-assessment is inderdaad ook een van de meest in de literatuur genoemde toepassingen van scoringsrubrieken in het algemeen. Het gebruik van de scoringsrubriek hoeft echter niet beperkt te blijven tot zulke schoolsituaties. De deeltijdstudenten uit het focusgroepsgesprek vertelden dat ze de kritische reflectie over eigen en andermans informatieproducten nu ook toepasten bij andere gelegenheden zoals hun reguliere werk of hobby. Een andere toepassing van de scoringsrubriek die bijzonder gewaardeerd werd, is het geven van feedback. Juist het feit dat die kan worden toegespitst op de zeven afzonderlijke criteria en dat hij is vastgelegd zodat hij later nog eens kan worden teruggezocht, wordt door de studenten bijzonder gewaardeerd. Sommigen waren van mening dat het daarbij niet veel uitmaakt of de feedback wordt gegeven door de docent of door een van de medestudenten of collega's ('peers'), anderen meenden dat je er in ieder geval van

overtuigd moet zijn dat de reviewer zelf wel voldoende informatievaardig is. Voor het plannen van een studietaak, een vaak genoemde toepassing van scoringsrubrieken in het algemeen, is de Scoringsrubriek voor Informatievaardigheden volgens de geïnterviewde studenten niet geschikt. De verklaring die ze daarvoor gaven was dat de scoringsrubriek voor informatievaardigheden te veel de nadruk legt op de eigenschappen van het eindproduct en niet op de manier hoe dat tot stand komt.

8 Evalueren met docenten van De Haagse Hogeschool

Net zoals voor studenten werd onderzocht hoe de Scoringsrubriek voor Informatievaardigheden in de praktijk werd gebruikt, werd bij docenten nagegaan hoe zij de scoringsrubriek gebruiken in hun eigen onderwijs. Daarvoor werden vijf docenten van even zoveel verschillende academies van De Haagse Hogeschool ondervraagd. De docenten werden benaderd omdat ze in het verleden op een of andere manier hadden laten horen dat ze in de scoringsrubriek geïnteresseerd waren.

De meesten van de geïnterviewde docenten gaven aan dat ze onderdelen van de scoringsrubriek hadden gebruikt voor het aanscherpen van bestaande beoordelingscriteria en –instrumenten, bijvoorbeeld die voor het beoordelen van afstudeerverslagen. Slechts één respondent vertelde dat bij haar opleiding de scoringsrubriek in haar geheel werd gebruikt maar daar hadden ze de rubriek vervolgens weer uitgebreid met criteria op het gebied van schrijfvaardigheid. In de betreffende opleiding (Business IT & Management) kregen studenten expliciet les in het hoe en waarom van de scoringsrubriek en de daar gebruikte criteria. Ook werd de scoringsrubriek daar in het onderwijs ingezet als instrument voor peer-assessment waarvan de beoordelingen door studenten moesten worden opgenomen in hun trainingsdossiers.

De scoringsrubriek werd door docenten echter niet alleen gebruikt voor het toets- en beoordelingsproces, respondenten legden uit dat zij haar ook gebruikten voor het bepalen van relevante leerinhouden in het kader van curriculumontwikkeling. Deze toepassing bij curriculumontwerp is vooral actueel omdat veel opleidingen druk zijn met de ontwikkeling van onderwijs in onderzoeksvaardigheden, een onderwerp dat recent veel aandacht heeft gekregen in het hoger beroepsonderwijs.

9 Algemene conclusies en discussie

De centrale vraag waar in dit proefschrift onderzoek naar is gedaan luidde: Wat is een betrouwbaar, valide en praktisch bruikbaar instrument dat generiek toepasbaar is voor het beoordelen én het bevorderen van de informatievaardigheden van individuele studenten in het hoger onderwijs? Tijdens het onderzoek kwam naar voren dat voor het beoordelen van het daadwerkelijke informatiegebruik door studenten (“performance assessment”) het best gebruik gemaakt kan worden van voor het vakgebied authentieke studieopdrachten die op het aspect informatiegebruik worden beoordeeld met een speciaal daarvoor ontwikkelde scoringsrubriek. In de praktijk blijkt dat de scoringsrubriek die daarvoor in het onderzoek is ontwikkeld, meestal

nog wordt aangepast en toegespitst op de opdracht waarbij zij wordt ingezet. In haar opzet is het instrument echter bruikbaar gebleken voor een betrouwbare en valide beoordeling van de informatievaardigheden van individuele studenten en bewijst zij bovendien haar diensten bij de ontwikkeling van opdrachten, lessen en curriculumonderdelen op het gebied van onderzoeks- en informatievaardigheden.

Wat men zich wel dient te realiseren, is dat de scoringsrubriek zich beperkt tot een beoordeling van de vaardigheid in het oplossen van informatieproblemen en dat de constructie van nieuwe kennis - uiteindelijk ook een doel van het Information Problem Solving-proces in het hoger onderwijs - er niet mee beoordeeld wordt. Ook is van belang dat de meeste criteria betrekking hebben op het informatieproduct dat de studenten maken en dat aandacht voor het proces is beperkt tot *informatiezoekproces*.

Een conclusie is ook dat de verwerving van informatievaardigheden binnen één specifieke context geen belemmering is voor de transfer ervan naar andere professionele situaties. Dit is in overeenstemming met de literatuur over het leren van complexe vaardigheden (bijvoorbeeld Kirschner & Van Merriënboer, 2013). Gedurende het onderzoek is bovendien een aantal malen naar voren gekomen dat de beoordeling van informatievaardigheden bij voorkeur geïntegreerd plaatsvindt met de beoordeling van vakgerelateerde kennis en vaardigheden. In veel gevallen zal de Scoringsrubriek voor Informatievaardigheden dan nog wel specifiek gemaakt moeten worden voor dat betreffende vakgebied, bijvoorbeeld door het noemen van specifieke databases en analysetools. De bestaande scoringsrubriek kan de vakdocenten helpen bij het formuleren van de meer specifieke gedragsbeschrijvingen. Het domeinspecifiek maken van de scoringsrubriek zal ook helpen om de interesse voor het onderwerp informatievaardigheden onder deze docenten te vergroten. Andere aanbevelingen voor het onderwijs zijn de suggestie om de aandacht voor informatievaardigheden onder te brengen in een doorlopende leerlijn gedurende het volledige curriculum en de suggestie om docenten vooraf te trainen in het gebruik van de scoringsrubriek. Het management dient de docenten daarin dan natuurlijk wel te faciliteren.

Vervolgonderzoek ter verbetering van het beoordelen van informatievaardigheden in het hoger onderwijs zou kunnen gaan over instrumenten voor het beoordelen van de verschillende *stappen* in het Information Problem Solving-proces, een aspect dat in het onderhavige onderzoek onderbelicht is gebleven. In eerste instantie kan daarbij worden gedacht aan de vraag wat voor soort simulaties daarvoor kunnen worden ingezet. Andere suggesties voor vervolgonderzoek zijn een onderzoek naar de informatievaardigheden van docenten zelf en de ontwikkeling van een digitale interactieve versie van de Scoringsrubriek voor Informatievaardigheden.

Assessment of information literacy skills in higher education

1 Research question and context of the research

Information literacy is the ability to use information. Sub-skills of this competence according to Boekhorst (2000, p. 103) and Webber and Johnston (2000, p. 382) include the ability to:

- recognize an information need
- formulate a question
- select adequate information sources (for instance: internet, a digital database or a library collection)
- use IT to access the information sources
- select and evaluate the information found
- analyze and synthesize the information, reuse and disseminate it.

Information literacy skills have become increasingly important to society since the 1980s. This is not only because the production, processing and distribution of information have become more of an important factor in the economy, but also because access to information has been democratized and information became readily available to all people in Western societies. Contemporary information users often have to find their own way through the abundance of available information, where in the past they could trust gatekeepers like publishers and libraries to do this for them. Developments in ICT are also the cause of knowledge becoming outdated quickly; all these together make for the fact that people need essential independent learning skills to be able to process the information for use in their working environment.

Similar processes are taking place in higher education: today's didactics emphasize the advantages of autonomous learning, where students process information themselves. Educational models, in which the construction of knowledge by the learner plays a central role, are known under various names such as problem-based learning, project-based learning and competence-based learning. These developments in higher education, but also the stronger focus on applied research at the 'hogescholen' in The Netherlands, show that the acquisition of information literacy skills deserves much more attention than it is the case right now. However, if this attention is to be taken seriously, it cannot be restricted to instruction only - measuring students' knowledge and performance should form an inherent part. This very belief was the motive for researching the evaluation of students' information literacy skills in this thesis. The research emphasized the construction and the test of a generic applicable assessment instrument, although how such an instrument is used in the practice of instructing was also analyzed. Current education concepts emphasize that assessment is not only relevant for the selection and qualification of students but should also play a role

for the learning process. The question of how such an instrument could stimulate the acquisition of information literacy skills was therefore part of the research. In the process, the main research question became: What is a reliable, valid and usable instrument that could be used to generically measure and stimulate individual students' information literacy skills in higher education?

2 Research in information literacy

In research papers on information literacy, there are three main research approaches. Two of them are relevant for this thesis. The first relevant approach focuses on the learning outcomes. Because these are mostly expressed in terms of observable behavior, the outcomes-based approach is also called the 'behavioral approach'. It is relevant for this thesis because the formulated learning outcomes play an important role in the construction of an assessment instrument.

The second approach relevant for this research is process oriented and constructivist. The focus thus shifts to the process of solving information problems rather than the results. What is characteristic for this approach is that, according to the researchers, information processing skills help students to learn: through solving information problems they also construct new knowledge. This view is more in agreement with modern concepts of higher education such as competence-based, project-based and problem-based learning, than the outcomes-based conception of information literacy.

In this research, the outcomes-based approach plays a role when the construct of information literacy is operationalized and 'information literate behavior' becomes a more concrete concept. For the selection of an appropriate assessment instrument, the choice of the process-oriented constructivist approach - which is better connected with modern ideas about education and learning - was more appealing. One of the characteristics of current higher education system is that assessment plays a role not only at the end of the learning process - where well-performing students are distinguished from bad performers - but also throughout the process so that students receive timely feedback on their progress and get an understanding of the areas which require more of their attention.

Chapter 2 is not exclusively dedicated to the research approaches to information literacy but also to the different models that are used in research. There is a distinction between two facets of information literacy:

- the ability to solve a specific, information problem-solving task;
- the construction of metacognitive knowledge ("Information literacy as a way of learning"; Bruce, 2000).

Information literacy as 'Information Problem Solving' is in this thesis based on a model by Brand-Gruwel, Wopereis and Vermetten (2005), who themselves were inspired by the Big Six Skills model by Berkowitz and Eisenberg. Information Problem Solving is presented as a complex skill which consists of six steps and several sub-activities which should be performed

in an integrated and coordinated way. The latter is expressed by the horizontal regulation bar, the point in which the model by Brand-Gruwel et al. differs from the one by Berkowitz and Eisenberg. In the following research, the idea of ‘regulation’ was also an important argument in choosing the model by Brand-Gruwel et al. as a starting point for the design of the information literacy as ‘Information Problem Solving’ model.

For the construction of metacognitive knowledge, the concept of Personal Knowledge Base applies (Bruce, 1998). This construct is operationalized with the help of learning theories while domain-based knowledge was distinguished from metacognitive knowledge which goes beyond specific domains or subjects.

The Information Problem Solving model has served throughout the project the role of a conceptual framework for the research on how students’ information literacy skills can be measured.

3 Measurement of information literacy skills

In the research literature, different types of instruments that measure information literacy skills have been identified: tests, surveys, assignments, direct observation, logs and portfolios. Tests and surveys fail to be fully adequate to assess the students’ actual use of information. They are too much oriented on what students know and understand (tests) or what they say how they behave (surveys), without providing reliable evidence on how they actually perform. With surveys, there is also the risk of students producing socially desirable answers.

A possible alternative is the use of authentic assignments which have their roots in the professional work environment. These kinds of assessments are not oriented on the measurement of sub-tasks or sub-skills but on the assessment of ‘whole tasks’. While completing these complex tasks, students have to split their attention between different facets and sub-tasks simultaneously. These types of assessment are often used in alternative educational settings as distinguished in Chapter 1 (problem-based learning, project-based learning and competence-based learning). An additional benefit is that the skills learned in such a practical work setting can then easily be transferred to other study or work environments.

Different types of measurement are mentioned in the literature on ‘performance assessments’. Firstly, there are assignments where student products are assessed on the use of information (e.g. by evaluating the reference list) or reflective reports on the research process itself. Work on the assignment can also be documented by students in a portfolio. With the exception of the product itself, a log report and a reflection report can also be supplemented with evaluations by others including peers and critical friends. Yet it is even more informative when students can work on their portfolio for a longer time so that their development becomes clearly visible. The disadvantages of reflection and log reports are that these types of assessment do not measure the actual performance of the students but only what they say has been done. It is possible to observe students’ performances directly, for instance with screen capture software and think-

aloud protocols, but these methods require so much effort that they are hardly spoken of in the literature as an assessment method. It has to be said, however, that direct observation has been described as a research method in experimental research.

In the abovementioned types of information literacy performance assessment, a need for some kind of checklist for the grading scheme itself is prevalent. A more advanced type of a checklist is an 'analytical scoring rubric'. This rubric contains criteria for the evaluation of a task performance and these should also apply to the grading itself. Rubrics are used more and more often for the performance assessment of authentic work tasks. In the literature regarding performance assessment of information literacy, the following advantages are mentioned in favor of analytical scoring rubrics (Oakleaf, 2008):

- they facilitate consistent scoring and prevent inaccuracy and bias because of the detailed descriptions of criteria and norms on which graders can rely
- they provide excellent feedback to the students
- when provided in advance and discussed in the classroom, students understand what the teacher expects from them
- students can use the scoring to assess their own or their peers' work (peer and self-assessment).

In order to answer the main question of this research, it was decided to construct a tool for 'authentic assessment'. Based on the literature, the development of an analytical scoring rubric was chosen, which could be used for the evaluation of the Information Problem Solving process during the performance of work-related knowledge tasks. This means that the evaluation of information literacy skills will be embedded in specific domain-related courses. The choice of an analytical scoring rubric is based on the conviction that this instrument is also the most appropriate one for stimulation of information literacy skills.

4 Construction of the scoring rubric

The scoring rubric was designed in seven different steps which were divided into three stages:

Orientation

1. Determination of the criteria
2. Description of the 'professional behavior' for each criterion
3. Description of the 'inadequate behavior' for each criterion

Elaboration

4. Choice for the number of levels that will be used
5. Description of the behavior for the 'in between' levels for each criterion

Test

6. Test of the first draft version with student products at different levels
7. Revision of the scoring rubric, if needed.

Orientation was executed during an iterative process with the aid of literature (in particular the ACRL standards were analyzed), monitoring of student behavior during labs and consultations with colleagues concerning the first drafts of the rubric. In the end, the focus became more about the characteristics of the product than the description of the process, although the latter is still expressed in criteria 6 and 7 of the scoring rubric. The argument for the focus on the student product was that Information Problem Solving is a rather generic skill and therefore it is realistic to allow students, within certain boundaries, to make their own choices on how to approach the task.

The norms for 'professional behavior' were mainly derived from the literature, and the descriptions of 'inadequate behavior' from lab observations. At the end, the formulations of criteria and the descriptions of behavior were discussed in review sessions with colleagues from the department of information studies and three other departments, before the first draft version to be tested was determined.

This first draft version of the scoring rubric contained descriptions of behavior on three levels (Table 4.1). The results of the test with 15 student products were graded by two different raters with the motivation to revise the first draft version. The definitive prototype has seven criteria and descriptions of behavior on two levels, completed in accordance to the six-point Likert scale (bijlage 4.1).

5 Evaluation of usefulness at Saxion; Test of reliability and validity at the Netherlands Defense Academy and Open University

This chapter describes how the prototype was tested in three different contexts:

- Department Commercial Economy of Saxion in Enschede (School of Marketing & International Management)
- Department Military Business Studies of the Netherlands Defense Academy
- The Information Literacy course for social scientists at the Open University (Psychology and Educational Sciences faculty).

At Saxion, the scoring rubric was used by eight tutors who graded individual process reports on the facet of 'information use'. Seven of the eight tutors were of the opinion that all the criteria from the rubric together were a good and complete picture of the construct information literacy. Seven of them observed that the scoring rubric had helped them to stay focused during the grading process. When they were asked, however, whether they would use the rubric once again for summative (i.e. selective) evaluations, their answer was 'no' or they were in doubt. Regarding the use of the rubric for formative assessments or for instructions, six tutors had positive feedback.

At the Netherlands Defense Academy, two professors in Information Science graded the same 27 student papers, which in this case were literature reviews. To test the quality of the scoring rubric, it was examined how much the two professors agreed in their gradings and in which

sense the gradings of the individual professors were internally consistent. It seemed that the two professors seldom reached absolute agreement on the six-point Likert scale for the individual criteria, but that they also did not differ too much. The so-called adjacent agreement (not more than one point difference on the six-point Likert scale) was quite high (for all criteria but one 85% or higher). The final gradings on a scale from 1-10 confirmed this rather high adjacent agreement. Agreement about the pass/fail decisions was for the final gradings at 93%. For three individual criteria, however, the agreement on the pass/fail decision was disappointing (less than 75%). It was concluded that it is not advised to use the individual criteria for summative assessments. It was also concluded that a good reliability of the final gradings can be reached because potential mistakes in the grading of the individual criteria can be compensated when all seven criteria are applied together.

Item-total statistics were calculated with SPSS and indicated that one grader in particular was able to grade the student products with the help of the scoring rubric with a very high level of internal consistency.

The two graders at the Netherlands Defense Academy agreed with the majority of the Saxion respondents that the seven criteria together gave a good picture of the construct of information literacy, and that the scoring rubric helped them to pay attention to the relevant topics.

In the case of Open University, the research method used for the Netherlands Defense Academy remained largely unchanged. There were 19 literature reviews and associated essays submitted by students of an Information Literacy course for social scientists that were evaluated by two graders. One grader was an Open University professor, the other was the researcher himself. In this case, adjacent agreement was sufficient for most of the criteria. The exception was criterion 2, the quality of the reference list. It seemed that the researcher himself had found the individual criteria hard to use in an internally consistent way. It appeared that his gradings with criterion 6 (the quality of search terms) differed from those on the other criteria. A possible explanation is his unfamiliarity with the subject field.

Unlike the Netherlands Defense Academy, at the Open University the same dossiers were also graded using an existing Open University grading model. That made it possible to investigate the existing correlations between their model and that of the Scoring Rubric for Information Literacy. It appeared that this correlation looked as follows: $r=0.68$ for grader 1 ($p<0.01$) and $r=0.55$ for grader 2 ($p<0.05$).

6 Test of reliability and validity at the department of Media Studies of the University of Amsterdam

At the University of Amsterdam's Media Studies department, two student assistants each graded 80 pieces of an assignment twice: once with the scoring rubric and once with the departmental grading model. The student assistants had recently finished their Master's degree in Media Studies.

In the past they had been involved with helping and preparing students for their assignments but they did not have any experience with actual grading of student products.

80 assignments were randomly selected from a group of 120; the number was large enough to determine the interrater reliability, not only from the rate of adjacent agreement but also from the statistical coefficient ICC(3). This Intraclass Correlation Coefficient was higher than 0.60 for almost all criteria which means that interrater reliability was overall 'good'. For the final grades, the ICC(3) almost reached the highly desirable value of 0.75. For criterion 1 ('Orientation') only, the ICC(3) was beneath 0.60. A possible explanation for the moderate agreement on this criterion is that it leaves graders more room for individual interpretation. Criterion 6 from the scoring rubric could not be used as the original assignment had not asked explicitly enough to report on the search terms used.

At the department of Media Studies at the University of Amsterdam, it was not only the interrater reliability that was investigated but also the construct validity and criterion validity of the scoring rubric for information literacy. For both graders, Cronbach's Alpha was higher than 0.80. Item-total statistics also indicated that the graders had handled the scoring rubric in an internally consistent way. This was, together with the results from an explorative factor analysis, a confirmation of the fact that the six used criteria all refer to the same construct. Criterion validity was, as in the case of Open University in Chapter 5, determined by calculating correlations between the gradings with the scoring rubric, and those with the grading model of the department for both individual graders. The correlations were $r=0.76$ ($p<0.01$) and $r=0.93$ ($p<0.01$), and can be regarded as high and very high respectively.

The two graders were asked to provide some comments during the feedback sessions. They remarked that the scoring rubric was easier to use than the departments' grading model and provided sufficient support. However, they preferred a ten-point Likert scale instead of the six-point one which is used in the Scoring Rubric for Information Literacy. They further recommended to change the presentation order of the criteria. It was in particular criterion 4 ('in-text citations') and criterion 2 ('reference list') which should be presented one after another.

7 Student evaluation: part-time Information Studies students at The Hague University of Applied Sciences

To gather information about the way how students (could) use the Scoring Rubric for Information Literacy, research was carried out among a group of Information Studies students. These students had previously participated in a workshop in which the rubric was introduced and they learned how to work with it. Almost half a year after the workshop, five students of the part-time Information Studies course were interviewed in a focus group.

The students indicated that they had used the rubric above all for self-assessment before uploading their piece of work for grading. The appliance of this type of self-assessment is also mentioned

the most often in the general literature on scoring rubrics. The use of scoring rubrics should not, however, be limited to school-related situations only. The students from the focus group advised that they executed this critical attitude in their own or other people's work, and in situations such as their regular job or even free-time activities. Feedback is the second application of the scoring rubric, which appeared to gain high praise. It was particularly valued that it focuses on seven criteria and that it is written down in a consistent form so that it can be re-read afterwards. Some of the students were of the opinion that it does not matter very much who is giving the feedback - the teacher, a fellow student or perhaps a colleague. But other students insisted that the reviewer himself (herself) should be information literate. All of the participating students shared the opinion that the Scoring Rubric for Information Literacy was not a handy tool for planning work on the assignment, an application that is often mentioned in the literature on rubrics in general, the reason for that being the Scoring rubric for Information Literacy emphasizes the characteristics of the product much more than how it is created.

8 Teacher evaluation at The Hague University of Applied Sciences

The research in Chapter 8 is comparable with that in Chapter 7, except that Chapter 8 investigated how *teachers* use the Scoring Rubric for Information Literacy in their own instructions and student evaluations. With this aim in mind, five teachers from five different faculties were questioned. The teachers were requested to participate because they had shown interest in the scoring rubric in the past. Most of them explained that they had used parts of the scoring rubric to modify existing assessment instruments, such as those for the evaluation of bachelor theses. Only one respondent said that they used a complete rubric at her department but that they had supplemented it with criteria for writing skills. At her department (Business IT & Management), students were explicitly informed about the importance of the rubric, the individual criteria, and how and why they should be using it. They also used the rubric for peer assessment and the students inserted the peer evaluations in their training dossiers.

But the scoring rubric was not only used for assessments and evaluation. Respondents explained how they also used it for determining relevant learning content during the process of curriculum development. The appliance of the rubric for curriculum development was especially relevant because most of the departments were at that time busy with developing instructions in research skills, a topic which had and still has a lot of attention at universities of applied sciences in The Netherlands.

9 General conclusions and further discussion

The main question in this research was: What is a reliable, valid and usable instrument that could be used to generically measure and stimulate individual students' information literacy skills in higher education?

The results of the research indicate that for the evaluation of the actual information use by students, work-related, authentic assignments assessed with a scoring rubric constructed specifically for

the evaluation of information literacy skills are preferable. It appeared that the rubric which was developed in this research still needs small modifications when used for specific assignments. However, the design of the Scoring Rubric for Information Literacy appeared to be useful for a reliable and valid evaluation of individual students' information problem-solving skills. The rubric also helps with the development of assignments, lessons and learning content for a curriculum on research and information literacy skills.

Teachers who wish to use the rubric should, however, realize that it is restricted to the evaluation of information problem-solving skills and that the acquisition of new knowledge - the ultimate aim of the information problem solving process in higher education - cannot be assessed with it. It is also important to realize that most of the criteria are related with the products which the students create and that assessment of the process is restricted to the information seeking process.

It has also been concluded that the specific context where information literacy skills are acquired are not an obstacle for the transfer of these skills to other professional situations. This is confirmed by the literature on learning of complex skills (for instance Kirschner & Van Merriënboer, 2013). During the research, it also appeared that it is preferred to embed the evaluation of information problem-solving skills in the evaluation of more domain-related knowledge and skills. It is expected that the Scoring Rubric of Information Literacy should be adapted at a detailed level for those specific domains, for instance by referring to specific databases or analyzing tools. The example of the prototype in this research can help subject teachers with the formulation of specific student behavior. Adaption of the scoring rubric to specific domains will also help to stimulate engagement of subject teachers with information literacy in general. Other suggestions for the education sector are to create continuing learning lines for information literacy skills during a complete curriculum, and the suggestion to train teachers beforehand in the use of the scoring rubric. This should, of course, be facilitated by the management of the departments.

It is highly recommended to conduct further research on instruments to improve the evaluation of information literacy skills in higher education and to evaluate the different steps of the Information Problem Solving process, which is under-exposed in the current research. One of the first questions to be answered is which kind of simulations can be used for this aim. Other suggestions for future research are the information literacy skills of teachers themselves and the construction of a digital interactive version of the Scoring Rubric for Information Literacy.

Lijst van bijlagen

- 4.1 Prototype van de Scoringsrubriek voor Informatievaardigheden
- 5.1 Vragenlijst die werd gebruikt bij de opleiding Commerciële Economie van Saxion
- 5.2 OU-model voor het beoordelen van het literatuuronderzoek en het essay met betrekking tot Self Regulated Learning
- 5.3 Vergelijking van de beoordelingscriteria uit de Scoringsrubriek voor Informatievaardigheden met de beoordelingscriteria die worden gebruikt door de Open Universiteit in de cursus 'Informatievaardigheden voor sociale wetenschappers'
- 6.1 UvA-nakijkformulier van de opleiding Mediastudies voor 'Toets II: Het onderzoeksvoorstel'
- 6.2 UvA-nakijkformulier zoals gebruikt in het onderzoek
- 6.3 Scoringsrubriek zoals gebruikt in het onderzoek bij de opleiding Mediastudies van de Universiteit van Amsterdam
- 7.1 Vragen en resultaten van de enquête onder idm-studenten van De Haagse Hogeschool
- 8.1 Topic lijst voor de interviews met docenten van De Haagse Hogeschool
- 9.1 Versie 2.0 van de Scoringsrubriek voor Informatievaardigheden

Lijst van figuren

- 2.1 Model Informatievaardigheden van Boekhorst
- 2.2 Information Problem Solving-model van Brand-Gruwel et al. (2005)
- 2.3 Model van het Information Search Process volgens Kuhlthau (1993)
- 2.4 Information skills-model van Irving (1985)
- 2.5 Vergelijking van 4 Information Problem Solving-modellen
- 2.6 Het Information Problem Solving-proces zoals voorgesteld in het kader van het onderhavige onderzoek
- 2.7 Elementen waaruit een Persoonlijk Kennissysteem bestaat
- 3.1 Voorbeeld uit de ‘sample questions’ van SAILS (2012)
- 3.2 Testitem nummer 16 uit het “Research Readiness Self-Assessment” (Ivanitskaya et al., 2006)
- 3.3 Voorbeeld van een testitem uit de lokaal ontwikkelde test van Cherry et al. (2002)
- 3.4 Voorbeeld van een item uit de enquête van Timmers en Glas (20010)
- 3.5 Fragment uit de scoringsrubriek van Oakleaf (2009)
- 4.1 “Step-by-Step Procedure” voor het ontwerpen van scoringsrubrieken (Mertler, 2001)
- 4.2 Ontwerpproces van de Scoringsrubriek voor Informatievaardigheden
- 4.3 Vaststellen van de beoordelingscriteria en de beschrijving van professioneel en inadequaat gedrag tijdens de verkennende fase van het ontwerp van de scoringsrubriek
- 4.4 Onderdeel uit het prototype van de Scoringsrubriek voor Informatievaardigheden
- 6.1 Vergelijking van de deelvaardigheden behorende bij het construct informatievaardigheden met de indicatoren van het UvA-nakijkformulier en die van de Scoringsrubriek voor Informatievaardigheden

Lijst van tabellen

- 1.1 Hoofdstukindeling en opbouw van het proefschrift
- 3.1 Literatuur over het beoordelen van informatievaardigheden in het hoger onderwijs, periode 2000-2009
- 4.1 Eerste concept versie van de Scoringsrubriek voor Informatievaardigheden
- 5.1 Antwoorden op de stellingen en vragen die zijn voorgelegd aan de docenten van de opleiding Commerciële Economie van Saxion
- 5.2 Verschillende vormen van consensus tussen beoordelaar 1 en beoordelaar 2 bij de Nederlandse Defensie Academie
- 5.3 Item-Totaal Statistieken voor NDA beoordelaar 1
- 5.4 Item-Totaal Statistieken voor NDA beoordelaar 2
- 5.5 Absolute en aanpalende overeenstemming tussen beoordelaar 1 en beoordelaar 2 bij de Open Universiteit Nederland
- 5.6 Item-Totaal Statistieken voor OU beoordelaar 1
- 5.7 Item-Totaal Statistieken voor OU beoordelaar 2
- 6.1 Weging van de afzonderlijke indicatoren van zowel het UvA-nakijkformulier als van de Scoringsrubriek voor Informatievaardigheden
- 6.2 Wegingspercentages van de in het onderzoek bij Mediastudies daadwerkelijk gebruikte indicatoren
- 6.3 Tijdschema voor het beoordelen van de werkstukken bij de opleiding Mediastudies van de UvA
- 6.4 Verschillende vormen van consensus tussen beoordelaar 1 en beoordelaar 2 bij het gebruik van de scoringsrubriek bij de opleiding Mediastudies van de UvA
- 6.5 Intraclass Correlation Coefficient type Two-Way Mixed (ICC(3)) voor de beoordelingen met de scoringsrubriek bij de opleiding Mediastudies van de UvA
- 6.6 Intraclass Correlation Coefficient type Two-Way Mixed (ICC(3)) voor de beslissing voldoende-onvoldoende met de scoringsrubriek bij de opleiding Mediastudies van de UvA
- 6.7 Factor Matrices voor de scoringsrubriek bij extractie van één factor
- 6.8 Item-Totaal Statistieken voor de scoringsrubriek (UvA-beoordelaar 1)
- 6.9 Item-Totaal Statistieken voor de scoringsrubriek (UvA-beoordelaar 2)
- 6.10 Item-Totaal Statistieken voor het UvA-nakijkformulier (UvA-beoordelaar 1)
- 6.11 Item-Totaal Statistieken voor het UvA-nakijkformulier (UvA-beoordelaar 2)
- 6.12 Pearson correlatiematrix tussen de eindbeoordelingen met de scoringsrubriek en die met het UvA-nakijkformulier voor beoordelaar 1 en beoordelaar 2

-
- 6.13 Pearson correlatiematrix tussen de totaalscores met de scoringsrubriek (maximaal 90 punten) en die met het UvA-nakijkformulier (maximaal 85 punten) voor beoordelaar 1 en beoordelaar 2
 - 6.14 Gemiddelde scores en standaarddeviaties voor de beoordelingen met de scoringsrubriek bij de opleiding Mediastudies van de UvA
 - 7.1 Antwoorden van voltijd- en deeltijdstudenten op de enquêtevraag 2
 - 9.1 Stappen uit het IPS-model met corresponderende criteria uit de Scoringsrubriek voor Informatievaardigheden

Informatievaardigheden hebben betrekking op het oplossen van problemen of vraagstukken met behulp van informatie van het internet of uit een al of niet digitale bibliotheek. In het hoger onderwijs wordt er veel waarde aan gehecht dat studenten over deze vaardigheden beschikken. Maar hoe beoordeel je of studenten voldoende informatievaardig zijn? En wat voor soort beoordelingsinstrument is tevens geschikt om die vaardigheden te bevorderen?

In dit proefschrift wordt onderzoek gerapporteerd dat aannemelijk maakt dat een scoringsrubriek de voorkeur heeft als je een instrument zoekt waarmee je informatievaardigheden zowel kunt beoordelen als stimuleren. Op basis van literatuurstudie en de uitkomsten van onderzoek onder studenten en docenten is een scoringsrubriek ontwikkeld. Vervolgens is het beoordelingsinstrument door middel van empirische studies getest en geëvalueerd.

Jos van Helvoort (1957) studeerde Nederlandse taal- en letterkunde aan de Katholieke Universiteit Nijmegen en volgde de Opleiding tot Wetenschappelijk Bibliothecaris aan de Universiteit van Amsterdam. Dit onderzoek komt voort uit zijn werkzaamheden als hogeschooldocent bij de opleidingen IDM en HBO ICT van De Haagse Hogeschool en als onderzoeker bij het lectoraat Duurzame Talentontwikkeling. Jos publiceerde eerder over het onderwerp 'Beoordelen van informatievaardigheden' in het *Journal of Information Literacy*, het *Journal of Academic Librarianship* en *Communications in Computer and Information Science*. Hij is lid van het standing committee en programme committee van de European Conference on Information Literacy (ECIL) en reviewer voor het *Journal of Information Literacy*.