

Voorblad Onderzoeksvoorstel Subsidieaanvraag

Naam student: Jan Willem Boer
Studentnummer: 852414418
E-mail: ou@janwillemboer.nl
Cursuscode en cursusnaam: OM0323222344 Atelier
Datum (van inleveren): 10 juli 2023

In het Examenreglement van de Open Universiteit zijn definities alsmede een nadere toelichting opgenomen betreffende de begrippen fraude en plagiaat. Het Examenreglement vindt u op www.ou.nl/documenten

Verklaring eigen werk

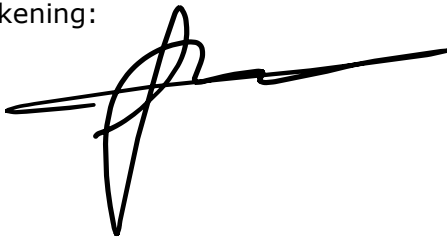
Ik heb de definities van fraude en plagiaat zorgvuldig gelezen en begrepen en ik verklaar hierbij dat het bovengenoemde werkstuk geheel mijn eigen werk is, en ik mij niet schuldig heb gemaakt aan fraude en/of plagiaat.

Toestemming gebruik

Daarnaast verleen ik de faculteit Onderwijswetenschappen toestemming om dit document in geanonimiseerde vorm te gebruiken voor:

- a. onderwijsdoeleinden ja
- b. onderzoeksdoeleinden ja

Plaats: Apeldoorn
Datum: juli 2023
Handtekening:



Onderzoeksvoorstel Subsidieaanvraag

1. Aanvrager(s)

Hoofdaanvrager / Contact persoon	
Naam	: Jan Willem Boer
Studentnummer	: 852414418
E-mailadres	: ou@janwillemboer.nl
Keuzeblok Atelier	: Effectief leren
Aantal woorden	: 2500

2. Titel en Samenvatting

Titel van het project en korte samenvatting van de probleemstelling	
Title (English):	Making introductory programming assessments for software engineering students more valid by adding multimedia elements to the assessment questions.
Title (Dutch):	Validiteit van toetsen voor beginnende software engineering studenten verhogen door het toevoegen van multimedia elementen aan de toets.
Samenvatting (max. 100 woorden):	Toetsen voor eerstejaars informaticastudenten in het hoger onderwijs bevatten elementen die de toets mogelijk minder valide en minder toegankelijk maken. Bij toetsitems wordt soms voorkennis verondersteld die niet met het getoetste te maken heeft, en niet bij alle studenten gelijk aanwezig is. Dit introduceert oneerlijke verschillen in <i>extraneous cognitive load</i> (ECL). In dit onderzoek worden ondersteunende afbeeldingen aan de toetsvragen toegevoegd. Naar verwachting zal dit de verwerkingstijd, ECL en toetsresultaten gunstig beïnvloeden. Het onderzoek is een experiment met een controlegroep en vier experimentele groepen, met verschillende typen afbeeldingen. De resultaten kunnen tot richtlijnen voor het opstellen van programmeertoetsen worden uitgewerkt.

3. Onderzoeksvoorstel

3.1 Inleiding tot het onderzoek en de onderzoeksvragen

Studenten in het Nederlandse hoger beroepsonderwijs worden opgeleid tot “beroepsbeoefenaars”. De beroepsbekwaamheid wordt getoetst door middel van representatieve beroepsopdrachten (Andriessen et al., 2017). Zo wordt aan eerstejaars studenten aan een bacheloropleiding informatica (hbo-ict) gevraagd om voor hun eerste toets een softwarematige oplossing voor een concrete casus uit de beroepspraktijk te programmeren. De student moet dus in staat zijn om zich in de voorgestelde situatie te kunnen inleven. Het interpreteren van de casus is echter geen leerdoel van de toets en valt daarmee onder *extraneous cognitive load* (ECL): cognitieve belasting die niet met het getoetste te maken heeft.

Onderzoek in onderwijs in het algemeen laat zien dat cognitieve belasting die door beroepstaken wordt geïntroduceerd, hoog is (Peng et al., 2017). Of dit ook voor beroepstaken in programmeertoetsen geldt, is niet duidelijk. In onderzoek naar de complexiteit van programmeertoetsen werd niet altijd een verschil in effect gevonden tussen het aanbieden van de toetsitems als casus of dezelfde items als abstracte opdracht (Bouvier et al., 2016; Craig et al., 2017). Een follow-up studie door Leinonen et al. (2021) vond bij sommige toetsvragen een verschil in effect. Als verklaring suggereren de auteurs dat bekendheid van de student met de casus of het onderwerp van de casus mee kan spelen. Als de student moeite heeft met het herkennen van de casus, dan vergt die meer cognitieve load dan een abstracte opdracht.

De toets moet daarom zo zijn opgesteld dat de casus aansluit bij de voorkennis en beleavingswereld van de student. Maar die verschillen per student, zeker op hogescholen waarbij een deel van de studentenpopulatie uit het buitenland afkomstig is. Deze verschillen kunnen tijdens de toets tot verschil in ECL leiden en daarmee *construct-irrelevant variance* introduceren: variatie in de uitkomst van de toets die niet met het getoetste zelf te maken heeft (Braun, 2019).

Als verschillen in achtergrond en voorkennis van studenten maken dat er verschillen zijn in ECL die geen relevantie in de toets zelf hebben, dan is dat een bedreiging voor de validiteit van de toets. Deze subsidieaanvraag stelt een onderzoek voor dat als doel heeft om de validiteit van de toets te vergroten. Bij validiteit van een toets is het belangrijk dat de toets meet wat deze beoogt te meten, en dat andere factoren zo min mogelijk invloed op de uitkomst hebben. In het onderzoek wordt een experimentele ingreep geïntroduceerd om de ECL die met de casus gepaard gaat voor alle studenten zo veel mogelijk te nivelleren en daarmee de construct-irrelevant variance te reduceren.

Om de ECL te verlagen, zullen voor dit onderzoek ondersteunende afbeeldingen aan de toetsvragen toegevoegd worden. Het multimedia principe in toetsing stelt dat een toetsvraag die tekst en afbeeldingen combineert, makkelijker te verwerken is dan een toetsvraag met alleen tekst (Lindner, 2021). De afbeeldingen richten zich specifiek op het ondersteunen van de casus, niet op het ondersteunen van de programmeertaak. Sheard et al. (2013) beargumenteren dat cognitieve belasting bij programmeertoetsen voor beginnende programmeurs bepaald wordt door de complexiteit van zowel de vraag als de moeilijkheid van de taak op zich. Zij stellen een framework voor waarmee de complexiteit in zes categorieën gemeten kan worden, zie Tabel 1, waarbij de eerste drie categorieën over de vraag gaan, en de laatste drie categorieën over de uit te voeren taak.

Tabel 1

De zes categorieën om complexiteit van items bij programmeertoetsen voor beginnende programmeurs te meten volgens Sheard et al. (2013).

Measure	Focus	Description
External domain references	Q	Reference to a domain beyond what one would reasonably expect introductory programming students to know
Linguistic complexity	Q	Length and sophistication of the natural language used to specify the question
Explicitness	Q	Extent of prescriptiveness as to how to answer the question
Conceptual complexity	Q & A	Classification of the individual programming concepts required to answer the question
Code length	Q & A	Whether code is up to half a dozen lines long, up to two dozen lines long, or longer
Intellectual complexity (Bloom level)	Q & A	Bloom’s taxonomy as applied to programming questions

Voor dit voorstel zijn de categorieën *external domain references* en *linguistic complexity* van belang, omdat deze categorieën de veroorzakers van ECL in de casus bevatten. Voor de overige categorieën geldt dat de complexiteit onderdeel is van de toets zelf. Ondersteunde afbeeldingen in de tekst van de casus zijn dus bedoeld om de complexiteit in de eerste twee categorieën te verlagen.

Afbeeldingen kunnen in toetsen verschillende functies hebben. Kerckhoffs et al. (2023) stellen hiervoor een categoriseringsstelsel voor. In dit stelsel worden voor de *pictorial representation* van onderdelen van een toets zeven verschillende functionele categorieën van afbeeldingen onderscheiden.

Voor dit onderzoeksvoorstel zijn hiervan drie categorieën van belang die bedoeld zijn om de verwerking van de tekst van de opgave te vereenvoudigen: *representational*, *organizational* en *interpretational*.

Representational afbeeldingen zijn afbeeldingen die de tekst verduidelijken en minder abstract maken door de tekst te visualiseren. Dit type afbeeldingen kan worden gebruikt worden voor het verlagen van linguistic complexity, als hiermee onbekendere termen geïllustreerd worden. Uit het onderzoek van Sheard et al. (2013) blijkt dat linguistic complexity de hoogste correlatie met de moeilijkheid van het toetsitem heeft. De verwachting is dus dat het toevoegen van representational afbeeldingen het meeste effect zal hebben.

Organizational afbeeldingen helpen de student de structuur van de casus te begrijpen. Dit type afbeeldingen is geschikt voor het verlagen van complexiteit door externe domeinreferenties. Een voorbeeld van een casus is een winkeleigenaar die haar boekhouding in elektronische vorm wil gaan doen. De tekst kan beschrijven wat dit inhoudt, maar deze tekst bevat veel nieuwe informatie voor studenten die niet bekend zijn met de boekhouding van een winkel. Het toevoegen van een afbeelding van een kassabon kan hierbij een organizational functie hebben, omdat een zorgvuldig samengestelde kassabon de structuur van een transactie overzichtelijk laat zien. Denk aan dat er btw wordt gerekend over producten, met mogelijk verschillende tarieven, dat prijzen keer het aantal artikelen gedaan worden en dat er een fooi kan worden gegeven.

Interpretational afbeeldingen helpen een tekst te interpreteren door processen of wijzigingen over tijd duidelijk te maken. Ook dit type afbeeldingen kan complexiteit door externe domeinreferenties reduceren. Bij het voorbeeld van de winkel kan een flow-diagram van de bedrijfsvoering het proces van het bedrijf in beeld brengen. De student hoeft daardoor minder moeite te doen om een cognitieve representatie van dit proces te construeren. Uit het flowdiagram kan bijvoorbeeld duidelijk worden dat de som van alle verkooptransacties van een dag vergeleken worden met de inkomsten, en dat kasverschillen daarna opgelost moeten worden.

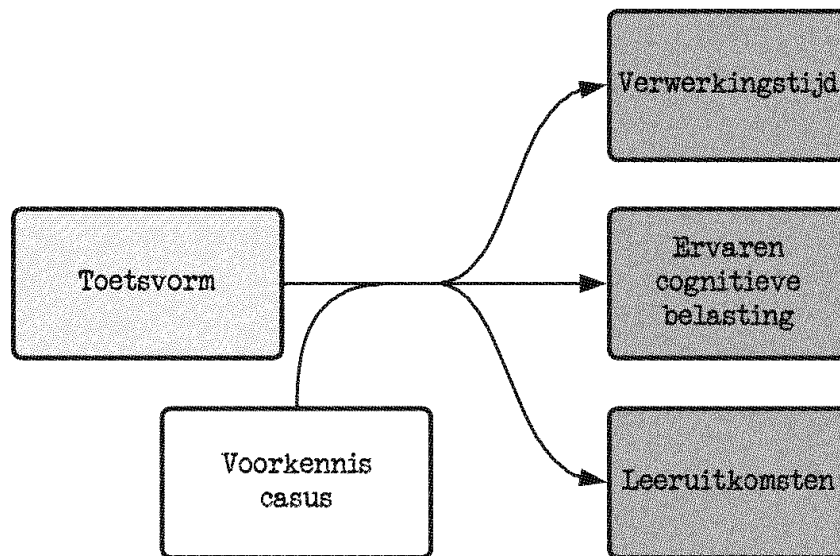
Om het effect van het toevoegen van ondersteunende afbeeldingen in de toetsvragen te onderzoeken, wordt gekeken naar de variabelen ECL, uitkomst van de toets en verwerkingstijd. Hierbij wordt voorkennis die de student rapporteert als covariaat meegenomen. In Figuur 1 zijn de variabelen als conceptueel model weergegeven. Voor de studie zijn de volgende onderzoeksvragen opgesteld. Elke vraag heeft drie bijbehorende subvragen, om ook het effect van elk afbeeldingstype afzonderlijk in kaart te brengen.

- RQ1: welk effect heeft het toevoegen van ondersteunende afbeeldingen aan de toetsvraag op de ECL die studenten ervaren bij het verwerken van de toetsvraag?
 - RQ1a: welk effect heeft het toevoegen van representational afbeeldingen hierbij?
 - RQ1b: welk effect heeft het toevoegen van organizational afbeeldingen hierbij?
 - RQ1c: welk effect heeft het toevoegen van interpretational afbeeldingen hierbij?
- RQ2: welk effect heeft het toevoegen van ondersteunende afbeeldingen aan de toetsvraag op de uitkomst van de toets? RQ2a, RQ2b en RQ2c brengen het effect van de afbeeldingstypes afzonderlijk in kaart op dezelfde wijze als bij RQ1.
- RQ3: welk effect heeft de aanpassing op de tijd die studenten bezig zijn met het verwerken van de vraag? RQ3a, RQ3b en RQ3c brengen het effect van de afbeeldingstypes afzonderlijk in kaart op dezelfde wijze als bij RQ1.

De verwachting bij alle drie de onderzoeksvragen is dat de experimentele condities minder spreiding in de uitkomstmaten laten zien dan de controleconditie. Dat zou de hypothese ondersteunen dat de verschillen tussen studenten minder groot zijn als de toetsvraag ondersteunende afbeeldingen heeft dan bij toetsvragen zonder afbeeldingen. Hierbij is de verwachting dat dit vooral voor representational afbeeldingen zal gelden, en in mindere mate voor organizational en interpretational afbeeldingen.

Figuur 1

Conceptueel model van de variabelen in het onderzoeksvoorstel



3.2 Methode (ca. 800 woorden)

De onderzoeksvragen zullen in een *between groups* experimenteel design beantwoord worden. De deelnemers worden geworven uit eerstejaarsstudenten Introductie Programmeren van cohort 2023-24. Het gaat om jongeren van 17-20 jaar, waarvan ongeveer 10% vrouwen. Ongeveer 20% van de deelnemers zijn internationale studenten, waarbij het aantal vrouwen iets hoger ligt, namelijk 20%. Bij een gemiddelde effectgrootte, een power van 0.8 en een significantieniveau (α) van 0.05 wordt gestreefd naar 265 deelnemers (cf. Faul et al., 2007). De deelnemers worden willekeurig ingedeeld in een van de vijf groepen. De controlegroep, groep A, krijgt de toets met de itemstam die alleen tekstueel is. Er zijn vier experimentele groepen die een herontworpen versie van de toets krijgen. Groep B krijgt de toets met representational afbeeldingen. Groep C met organizational afbeeldingen. Groep D met interpretational afbeeldingen. Groep E krijgt de toets met alle typen afbeeldingen. Een overzicht hiervan staat in Tabel 2.

Tabel 2

De condities in het experiment

Groep	Toetsvorm
A	Tekstueel
B	Tekstueel en afbeeldingen, representational
C	Tekstueel en afbeeldingen, organizational
D	Tekstueel en afbeeldingen, interpretational
E	Tekstueel en afbeeldingen, alle functies

Voorafgaand aan de toets wordt de voorkennis van de studenten in kaart gebracht. De studenten wordt gevraagd zichzelf te scoren op een schaal van 1-7 in hoeverre ze ervaring hebben met het domein dat in de toetsvraag als onderwerp wordt gebruikt.

Om ethische redenen zal het experiment niet tijdens een summatieve toets afgenomen worden. Het experiment zal plaatsvinden tijdens een oefentoets, waarbij de setting van de summatieve toets zo veel mogelijk wordt nagebootst: overleg is niet toegestaan, vooraf is bepaald welk materiaal als naslagwerk beschikbaar is, en de tijd is beperkt. De toets wordt in een online omgeving aangeboden.

De verwerkingstijd wordt tijdens de toets gemeten. Studenten worden vooraf geïnstrueerd om drie verschillende tijdstipmomenten aan te geven. Het eerste moment is dat de student aangeeft te willen beginnen. De toetsvraag, het volledige item, wordt dan zichtbaar. De student mag de vraag lezen, maar nog niet aan de opgave beginnen. Het tweede moment is als de student aangeeft klaar te zijn met lezen, en te willen starten met het maken van de opgave. Het verschil tussen deze twee momenten is de initiële verwerkingstijd. Tijdens het maken van de opgave blijft de toetsvraag zichtbaar. Als de student klaar is met de opgaven, levert die het werk in. Het is dan niet meer mogelijk om de opgave nog te wijzigen. Het moment van inleveren geldt als het derde moment. Het verschil tussen het eerste en derde moment geldt als de totale verwerkingstijd.

Na het inleveren wordt de student gevraagd een vragenlijst in te vullen over diens cognitieve verwerking tijdens de toets. Hiervoor worden de items over ECL gebruikt uit de vragenlijst van Krieglstein et al. (2023). Om de leerruitkomsten te bepalen wordt elke toets willekeurig aan twee verschillende docenten toegewezen, die deze anoniem en via rubrics nakijken. Om de consistentie tussen de beoordelaars inzichtelijk te maken wordt de *inter-rater agreement* berekend met behulp van Cohens

kappa, waarbij 0.8 of meer als acceptabel wordt beschouwd (cf. Graham et al., 2012). Als de kappa minder is dan 0.8, dan geldt dit als belangrijke kanttekening bij de analyse en conclusies rondom leeruitkomsten.

De data zal worden geanalyseerd met een MANCOVA met als onafhankelijke variabele de toetsvorm, als covariaat de voorkennis op het domein, en als uitkomstvariabelen de initiële verwerkingstijd, totale verwerkingstijd, ervaren cognitieve belasting en leeruitkomst.

Deelname aan het experiment is vrijwillig, en deelnemers kunnen te allen tijde afzien van deelname. Ze kunnen de oefentoets dan op de normale wijze doen, zonder aanvullende vragen en zonder dat hun gegevens in het experiment worden meegenomen. Vooraf worden de deelnemers geïnformeerd en wordt hun om toestemming gevraagd. Deelnemers jonger dan 18 krijgen een informatiebrief voor de ouders mee. Het experiment zal eerst aan een ethische commissie worden voorgelegd.

3.3 Wetenschappelijke relevantie

De uitkomsten van onderzoek naar het effect van de vormgeving van items van programmeertoetsen voor beginnende programmeurs op de leeruitkomsten zijn niet eenduidig. Er wordt niet altijd een effect gevonden bij het aanbieden van toetsitems als casus in plaats van als abstracte opdracht (Bouvier et al., 2016; Craig et al., 2017). Leinonen et al. (2021) denken dat contextuele voorkennis hierbij belangrijk is. Ook hun onderzoek geeft een gemengd beeld. De gevallen waarbij de casus een positief effect had, waren gevallen waarbij de opgaven meer wiskundig dan algoritmisch van aard waren, wat consistent is met onderzoek naar wiskundetoetsen. In de gevallen waarbij de casus tegenwerkte, speelde de factor tijd om de casus te verwerken een rol. Zij suggereren dat er meer onderzoek nodig is naar het effect van contextuele beschrijvingen in toetsen.

Dit onderzoeksvoorstel werpt mogelijk meer licht op wat het effect van de vormgeving van toetsitems voor beginnende programmeurs op de cognitieve belasting van studenten tijdens de toets is. Verder combineert het voorgestelde onderzoek twee bestaande modellen met elkaar, namelijk het categoriseringssysteem voor multimedia in toetsing van Kerckhoffs et al. (2023) en het complexiteitsframework voor toetsen voor beginnende programmeurs van Sheard et al. (2013). Hiermee geeft het onderzoek waardevol inzicht in de praktische toepasbaarheid van deze modellen, en geeft het vervolgonderzoeken handvatten in het gebruik van deze modellen voor soortgelijke onderwerpen.

3.4 Verwachte wetenschappelijke output

Van het onderzoek en de resultaten zal verslag gedaan worden in een open access artikel dat aan het canon van wetenschappelijk onderzoek naar programmeertoetsen en multimedia in toetsing kan worden toegevoegd. De resultaten van het onderzoek zullen geanonimiseerd als open data beschikbaar worden gesteld om de reproduceerbaarheid en transparantie van het onderzoek te vergroten.

Om de resultaten ook voor ontwikkelaars van toetsen voor de hbo-ict's toegankelijk te maken, zal het onderzoek samengevat worden in een blog post en een poster die op de website van hbo-ict zullen worden geplaatst, het samenwerkingsverband voor hbo-ict's in Nederland. Met een QR-code bij de blogpost en de poster is het mogelijk het complete onderzoek te downloaden.

Voor de online omgeving waarmee de data wordt verzameld, wordt een eigen systeem ontwikkeld. In deze omgeving wordt de consent geregeld, de tijdregistraties gedaan, de toets beschikbaar gesteld, en de vragenlijsten afgenomen. De broncode van deze software wordt voor gebruik voor andere onderzoeken als open source beschikbaar gesteld.

3.5 Maatschappelijke/Praktische relevantie op micro, meso, en/of macro niveau

De resultaten van het voorgestelde onderzoek kunnen licht werpen op welke factoren meewegen in de complexiteit van casussen bij programmeertoetsen voor beginnende programmeurs. Er zitten nu mogelijk oneerlijke verschillen in de toetsresultaten van studenten doordat de complexiteit door factoren beïnvloed wordt die van student tot student kunnen verschillen. Hierdoor is de validiteit van deze toetsen minder hoog: er worden onderdelen gemeten die niet met het getoetste te maken hebben, zoals bijvoorbeeld voorkennis en cultureel-maatschappelijke kennis, die niet voor iedereen gelijk zijn. Ook wordt de toegankelijkheid van toetsen minder als voor sommige studenten door irrelevante factoren drempels worden opgeworpen om hun kennis van programmeren te laten toetsen, bijvoorbeeld als snelheid van lezen meetelt in de uitkomst (Kettler et al., 2009).

Met de resultaten van het voorgestelde onderzoek kunnen richtlijnen opgesteld worden voor ontwikkelaars van toetsen om de toetsen eerlijker en toegankelijker te maken door het toevoegen van verschillende soorten afbeeldingen bij de tekst. De resultaten zijn nuttig voor alle hbo-ict's die beroepsauthentieke toetsen gebruiken om hun beginnende programmeurs te toetsen.

3.6 Referenties (Max. 2 pagina's)

Andriessen, D., Sluijsmans, D., Snel, M., & Jacobs, A. (2017). *Protocol Verbeteren en Verantwoorden van Afstuderen in het hbo 2.0*.

Hoe, waar vandaan	Gevonden via de beleidsdocumenten van onze hogeschool over toetsing. De wetenschappelijke papertrail van deze beleidsdocumenten is erg mager, dus het spoor is niet altijd rechtstreeks. In dit geval leidt er
-------------------	--

	uiteindelijk een spoor naar Merriënboer en de 4C/ID gedachten over toetsing.
Design & participanten	Nvt
Relevantie	Richtlijnen voor toetsen in het hbo.

Bouvier, D., Lovellette, E., Matta, J., Alshaigy, B., Becker, B. A., Craig, M., Jackova, J., McCartney, R., Sanders, K., & Zarb, M. (2016). Novice Programmers and the Problem Description Effect. *Proceedings of the 2016 ITiCSE Working Group Reports*, 103-118. <https://doi.org/10.1145/3024906.3024912>

Hoe, waar vandaan	Via forward search in Google Scholar op het artikel van Sheard et al.
Design & participanten	232 studenten, verspreid over 6 instellingen in 4 verschillende landen. Het onderzoek had een twee-groepen design, maar het is niet duidelijk hoe de groepen verdeeld zijn. Er is ook geen voormeting gedaan. Het lijkt dus een pre-experimenteel design te zijn.
Relevantie	De uitkomsten van het onderzoek laten geen verschil zien in opgaven met context en zuiver abstracte opgaven.

Braun, H. (2019). Performance assessment and standardization in higher education: A problematic conjunction? *British Journal of Educational Psychology*, 89(3), 429-440. <https://doi.org/10.1111/bjep.12274>

Hoe, waar vandaan	Via zoeken in EBSCO host, validity/assessment/higher education, vanaf 2013, peer-reviewed, full text online.
Design & participanten	nvt
Relevantie	Achtergrondinformatie over validiteitsbedreigers van toetsen in het hoger onderwijs. Heeft ook nog een interessante link naar een toetsen-repository die in Duitsland zou worden samengesteld met allemaal gevalideerde toetsen voor allerlei onderwerpen.

Craig, M., Smith, J., & Petersen, A. (2017). Familiar contexts and the difficulty of programming problems. *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*, 123-127. <https://doi.org/10.1145/3141880.3141898>

Hoe, waar vandaan	Via forward search in Google Scholar op het artikel van Sheard et al. Dit is een vervolgonderzoek op Bouvier et al. (Michelle Craig is een van de al.)
Design & participanten	879 studenten van een Finse universiteit kregen 10 opgaven waarvan 5 willekeurig met context en 5 abstract.
Relevantie	Ook de uitkomsten van dit onderzoek laten geen verschil zien in opgaven met context en zuiver abstracte opgaven.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. <https://doi.org/10.3758/BF03193146>

Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. In *Online Submission*. <https://eric.ed.gov/?id=ED532068>

Hoe, waar vandaan	Zoeken in Google Scholar naar "interrater reliability education assessment".
Design & participanten	Nvt
Relevantie	Achtergrond over interrater reliability in onderwijs. Auteurs beargumenteren dat je beter inter-rater agreement kunt gebruiken omdat het zinvoller is naar de absolute cijfers te kijken dan relatieve.

Kerckhoffs, A. M. H., Janssen, J. P. W., Pat-El, R. J., Schleepen, T. M. J., & Jarodzka, H. (2023). *Developing a comprehensive classification system of multimedia use in testing* [Unpublished manuscript]. Faculty of Educational Sciences, Open Universiteit.

Kettler, R. J., Elliott, S. N., & Beddow, P. A. (2009). Modifying Achievement Test Items: A Theory-Guided and Data-Based Approach for Better Measurement of What Students With Disabilities Know. *Peabody Journal of Education*, 84(4), 529-551. <https://doi.org/10.1080/01619560903240996>

Hoe, waar vandaan	Handboek van Beddow over "fairness bij toetsing" is niet toegankelijk via instellingsabonnementen. Dit artikel me Beddow als co-auteur gevonden.
Design & participanten	Nvt
Relevantie	Achtergrondinformatie over accessibility van toetsen. Mooie strakke definitie: dat een student tijdens de toets in staat is de getoetste kennis te etaleren.

Kriegelstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and Validation of a Theory-Based Questionnaire to Measure Different Types of Cognitive Load. *Educational Psychology Review*, 35(1), 9. <https://doi.org/10.1007/s10648-023-09738-0>

Hoe, waar vandaan	<i>Zoeken in Google Scholar naar "measuring cognitive load education".</i>
Design & participanten	5 studies voor het valideren van een nieuw meetinstrument voor CL met in totaal aantal participanten: 69 (PCA) + 158 (CFA) + 54 (3 validatieonderzoeken van de drie factoren).
Relevantie	Ontwikkeling van een meetinstrument voor het meten van ICL, ECL en GCL. De ECL items gebruik ik in dit onderzoek.

Leinonen, J., Denny, P., & Whalley, J. (2021). Exploring the Effects of Contextualized Problem Descriptions on Problem Solving. *Proceedings of the 23rd Australasian Computing Education Conference*, 30-39. <https://doi.org/10.1145/3441636.3442302>

Hoe, waar vandaan	Forward search in Google Scholar naar de artikelen van Bouvier en Craig, waar dit artikel een vervolg op is.
Design & participanten	917 studenten aan een universiteit in Nieuw-Zeeland. Twee experimentele groepen en een controlegroep kregen twee verschillende opgaven. De experimentele groepen met context, de controlegroep alleen de abstracte opgave.
Relevantie	Nuance in de uitkomst van eerder onderzoek: contextuele informatie maakt uit, maar is gekoppeld met voorkennis.

Lindner, M. A. (2021). Principles for Educational Assessment with Multimedia. In L. Fiorella & R. E. Mayer (Red.), *The Cambridge Handbook of Multimedia Learning* (3de dr., pp. 552-565). Cambridge University Press. <https://doi.org/10.1017/9781108894333.055>

Peng, J., Wang, M., & Sampson, D. (2017). Visualizing the Complex Process for Deep Learning with an Authentic Programming Project. *Journal of Educational Technology & Society*, 20(4), 275-287.

Hoe, waar vandaan	<i>Google Scholar, authentic summative programming assessment</i>
Design & participanten	29 derdejaars studenten van een universiteit in Hong Kong, in een pretest-posttest within subject design.
Relevantie	Achtergrondinformatie over cognitieve belasting van studenten bij project-based learning in programmeeronderwijs.

Saxion. (2022). Toetsbeleid ACT 2022-2023.

Hoe, waar vandaan	Beleidsdocument rondom toetsing van de academie van de auteur
Design & participanten	nvt
Relevantie	Beleidsinformatie over toetsen op de opleiding

Sheard, J., Simon, Carbone, A., Chinn, D., Clear, T., Corney, M., D'Souza, D., Fenwick, J., Harland, J., Laakso, M.-J., & Teague, D. (2013). How difficult are exams? A framework for assessing the complexity of introductory programming exams. *Proceedings of the Fifteenth Australasian Computing Education Conference - Volume 136*, 145-154.

Hoe, waar vandaan	<i>Google scholar, programming assessment complexity</i>
Design & participanten	20 toetsen van 10 instellingen, verdeeld over 5 landen.
Relevantie	Framework om complexiteit van eerstejaars programmeertoetsen te scoren.