

Attitudes van Psychologie Studenten over AI in Beoordeling: De Invloed van Educatie

Daniël Scheepstra

S4018613

Afdeling Psychologie, Rijksuniversiteit Groningen

PSB3A-BT15: Bachelor These

Supervisor: prof. dr. Henk Kiers

Tweede beoordelaar: dr. Lisette de Jonge-Hoekstra

In samenwerking met: Anouk Jansen, Margit Koldijk, Ellen Moll, Stan Velthuis en Vivian

Zandberg

13 April 2025

Een scriptie is een proeve van bekwaamheid voor studenten. De goedkeuring van de scriptie is het bewijs dat de student over voldoende onderzoeks- en rapportagevaardigheden beschikt om af te studeren, maar biedt geen garantie voor de kwaliteit van het onderzoek en de resultaten van het onderzoek als zodanig, en de scriptie is daarom niet per se geschikt als academische bron om naar te verwijzen. Als u meer wilt weten over het in deze scriptie besproken onderzoek en de daarop gebaseerde publicaties waarnaar u zou kunnen verwijzen, neem dan contact op met de genoemde begeleider.

Attitudes of Psychology Students Toward AI in Assessment: The Impact of Education

Abstract

Artificial Intelligence (AI) offers promising applications in higher education, including the automated assessment of essays and exams. This study examined the attitudes of Dutch psychology students toward AI-based assessment and the impact of an educational intervention. In an experimental design ($n = 221$), half of the students were exposed to a brief educational article about AI assessment, while the control group read a neutral article. Attitudes were measured using a 17-item questionnaire ($\alpha = .91$).

Results showed that, on average, students held more negative than positive attitudes toward AI as an assessor. The experimental group reported significantly more positive attitudes ($M = 4.60$, $SD = 1.40$) than the control group ($M = 3.70$, $SD = 1.49$), $t(219) = -2.84$, $p = .005$, 95% CI $[-0.94, -0.17]$. Thus, the intervention led to a small but significant improvement in attitude. A significant positive correlation was found between self-assessed knowledge of how AI grading works and attitude ($r = .291$, $p < .001$, 95% CI $[.165, .407]$), whereas knowledge of ethical issues related to AI showed no significant relationship with attitude ($r = .002$, $p = .971$, 95% CI $[-.130, .134]$).

These findings suggest that educating students on the technical functioning of AI may foster more positive attitudes. For successful implementation of AI assessment systems in higher education, transparent and accessible communication tailored to students' informational needs is essential.

Keywords: artificial intelligence, student attitudes, automated grading, higher education, educational intervention, psychology students, AI in assessment

Samenvatting

Kunstmatige intelligentie (AI) biedt veelbelovende toepassingen binnen het hoger onderwijs, zoals het automatisch beoordelen van essays en tentamens. Deze studie onderzocht de attitudes van Nederlandse psychologie studenten ten opzichte van AI-beoordeling en de invloed van een educatieve interventie hierop. In een experimenteel design ($n = 221$) werd de helft van de studenten blootgesteld aan een kort educatief artikel over AI-beoordeling, terwijl de controlegroep een neutraal artikel las. De attitude werd gemeten met een 17-item vragenlijst ($\alpha = .91$).

Resultaten tonen dat studenten gemiddeld eerder negatief dan positief stonden tegenover AI als beoordelaar. De experimentele groep scoorde significant hoger op attitude ($M = 4.60$, $SD = 1.40$) dan de controlegroep ($M = 3.70$, $SD = 1.49$), $t(219) = -2.84$, $p = .005$, 95% BHI [-0.94, -0.17]. De interventie leidde dus tot een kleine, maar significante toename in positieve houding. Daarnaast werd een significante positieve correlatie gevonden tussen zelfingeschatte kennis over de werking van AI en attitude ($r = .291$, $p < .001$, 95% BHI [.165, .407]), terwijl kennis over ethische aspecten van AI geen significante relatie vertoonde met attitude ($r = .002$, $p = .971$, 95% BHI [-.130, .134]).

De resultaten suggereren dat educatie over de functionele werking van AI kan bijdragen aan positievere houdingen onder studenten. Voor succesvolle implementatie van AI-beoordelingssystemen in het hoger onderwijs is het belangrijk om transparante en toegankelijke voorlichting te bieden, afgestemd op de informatiebehoefte van studenten.

Trefwoorden: kunstmatige intelligentie, attitude, beoordeling, hoger onderwijs, educatieve interventie, psychologie studenten, AI in assessment

Attitudes van Psychologie Studenten over AI in Beoordeling: De Invloed van Educatie

Kunstmatige intelligentie (AI) is een van de meest besproken en invloedrijke technologieën van de afgelopen decennia. Volgens experts is AI in staat om in de toekomst veel banen te vervangen (Grace et al., 2018). Zo zijn er binnen het medische veld al efficiënte AI modellen ontstaan die vroegtijdig trombose kunnen detecteren door subtiele patronen in medische data op te sporen (Stamate et al., 2024). Ook in het hoger onderwijs heeft AI de potentie om het hoger onderwijs te revolutioneren (Adiguzel et al., 2023).

AI is een vorm van technologie waar een machine zich gedraagt alsof het een mens is (He et al., 2020). Hiermee is het in staat taken uit te voeren die normaal gesproken menselijke intelligentie vereisen, zoals leren, redeneren, problemen oplossen en het verwerken van taal. Deze technologie is gebaseerd op geavanceerde algoritmen en maakt gebruik van grote hoeveelheden data om patronen te herkennen en autonome beslissingen te nemen (Zhang, 2023). De toepassingen van AI zijn breed en variëren van spraakherkenning en zelfrijdende auto's tot geautomatiseerde medische diagnostiek en onderwijs (Dwivedi et al., 2019).

In het onderwijs heeft AI een bijzondere positie verworven als hulpmiddel om processen te optimaliseren en leren te ondersteunen (Adiguzel et al., 2023). Specifieke toepassingen omvatten gepersonaliseerd leren, waarbij systemen zich aanpassen aan de behoeften van individuele studenten, en geautomatiseerde feedbacktools die snel en efficiënt begeleiding kunnen bieden (Chen et al., 2020). Deze ontwikkelingen hebben geleid tot de integratie van AI-modellen zoals ChatGPT en andere Natural Language Processing-systemen (d.w.z. systemen die in staat zijn menselijke talen te lezen), die studenten helpen met het genereren van tekst, samenvattingen, en het verbeteren van schrijfvaardigheden (Adiguzel et al., 2023).

Een van de meest veelbelovende toepassingen van AI in het hoger onderwijs is het gebruik ervan in beoordelings- en feedbackprocessen. Administratieve taken, zoals nakijken,

cijfers invoeren en feedback geven kosten docenten veel tijd. AI-systemen kunnen grote hoeveelheden studentenwerk analyseren en beoordelen op een snelheid en schaal die voor menselijke beoordelaars onhaalbaar zijn (Chassignol et al., 2018). Gebruik van AI zou kwaliteit van de lessen kunnen vergroten door dat docenten meer tijd kunnen gebruiken voor de inhoud van de lesstof, als administratieve zaken geautomatiseerd kunnen worden. Dit is met name waardevol in vakgebieden waar studenten regelmatig essays, onderzoeksrapporten of andere geschreven opdrachten indienen. Geautomatiseerde beoordelingssystemen maken gebruik van Natural Language Processing (NLP) om teksten te analyseren, grammaticale fouten te identificeren, en inhoudelijke structuur te beoordelen.

Een belangrijk voordeel van AI-gebaseerde beoordeling is de objectiviteit. Waar menselijke beoordelaars soms onbewust worden beïnvloed door persoonlijke bias, vermoeidheid of afleidingen, kunnen AI-systemen consistent dezelfde standaarden toepassen (Owan et al., 2023). Dit draagt bij aan eerlijke beoordelingen, ongeacht door wie of wanneer een opdracht wordt ingediend. Randall en Engelhard (2010) hebben in een observationele studie aangetoond dat er een bias aanwezig is op grond van de kennis die de docent heeft over de beoordeelde leerling, en dus niet alleen op het ingeleverde werk wordt beoordeeld. AI zou hier objectiever in kunnen zijn (Owan et al., 2023). Echter, dit is voor AI alleen mogelijk als de data waar ze op getraind worden ook zonder bias is (Chassignol et al., 2018). Anders zijn de standaarden die de AI-systemen consistent toepassen ook biased.

Recent onderzoek suggereert dat AI milder lijkt te zijn bij lagere prestatieniveaus van studenten en strenger bij hogere niveaus, maar dat de algehele nauwkeurigheid in beoordeling vergelijkbaar lijkt te zijn met die van docenten (Wetzler et al., 2024). Ander onderzoek heeft aangetoond dat docenten 92% van het door AI uitgevoerde beoordelingswerk als van hoge kwaliteit beschouwen (Bernius et al., 2022). Academics zelf geven ook aan dat het bruikbaar

is voor beoordeling van studenten, al waarschuwen ze wel voor ethische kwesties (Malik et al., 2024).

De attitude van studenten richting AI is over het algemeen positief. Zo hebben internationale farmacie studenten over het algemeen een positieve attitude richting AI in het algemeen en pleiten ze voor meer onderwijs over gebruik van AI in hun werkveld (Busch et al., 2023). Echter lijken farmacie studenten die meer *tech-savy* zijn wel zorgen te uiten over ethische en legale aspecten van deze technologie. Tech-savy refereert naar in welke mate je veel technologie gebruikt, wat in dit onderzoek beoordeeld is met een vragenlijst.

Chinese psychologie studenten voelen zich gemiddeld eerlijker beoordeeld wanneer zij door een AI worden geëvalueerd in plaats van door een docent (Chai et al., 2024). In het onderzoek van Chai et al. (2024) hebben studenten realistische scenario's gerelateerd aan beoordelingen binnen het onderwijs gekregen. Daarna moesten de studenten rapporteren wat hun gevoel van eerlijkheid was. Hierin was de onafhankelijke variabele of ze door AI of door een leraar beoordeeld werden. Uit dit onderzoek bleek dat de modererende variabele was of er uitleg bij de beoordeling was. Als er geen goede uitleg was, dan bleek de beoordeling van AI als eerlijker te voelen. Echter, is het is niet aannemelijk om de attitudes van Chinese studenten over te nemen wegens het cultuur verschil. In Nederland mist er nog onderzoek naar de attitude van studenten ten opzicht van beoordeling door AI.

Attitudes zouden mogelijk veranderd kunnen worden door wetenschappelijke kennis te leveren. Ranney en Clark (2016) hebben al laten zien dat dit mogelijk was met klimaatverandering. In een onderzoek met zeven experimenten is er een oorzakelijk verband gevonden tussen kennis die geleverd wordt en de mate van acceptatie voor klimaatverandering. In het eerste experiment werd met open vragen aangetoond dat participanten niet correct konden aangeven hoe klimaatverandering werkt. In experimenten 2-5 werden verschillende groepen participanten over verschillende perioden van tijd

geïnformeerd over klimaatverandering. Hierin werden sommige groepen met maar een heel kort artikel van 400 woorden geïnformeerd, terwijl andere groepen lezingen hadden van 45 minuten. Hieruit bleek dat de informatie hielp met het begrijpen van de gevaren van klimaatverandering. In experimenten 6 en 7 werd de attitude die participanten hadden ten opzichte van klimaatverandering veranderd met statistieken over klimaatverandering. Deze kennis is gebruikt om een website aan te passen om mensen te informeren op een korte, bondige manier, zelfs tot een samenvatting van het mechanisme van klimaatverandering met 35 woorden (Ranney et al., 2013). Deze vorm van informatie kort ter beschikking stellen zou ook bij studenten over AI in beoordeling gebruikt kunnen worden. Als studenten meer zouden weten over AI in beoordeling, dan zou dit positief beïnvloed kunnen worden, al lijkt dit in strijd te zijn met de uitkomsten van het eerder genoemde onderzoek van Busch et al. (2023).

Onderzoeksvraag

Voor dit onderzoek staan er drie onderzoeksvragen centraal: Ten eerste: Wat zijn de attitudes van psychologie studenten in Nederland ten opzichte van beoordeling door AI? Ten tweede: Hoe verandert deze attitude na het lezen van een educatief stuk over de effectiviteit van AI in beoordeling? En ten derde: Wat is de relatie tussen de attitude ten opzichte van beoordeling door AI en de bestaande kennis over beoordeling door AI bij psychologie studenten in Nederland?

Het is van belang om deze vragen te beantwoorden, omdat de antwoorden van deze vragen bijdragen aan het ontwikkelen van beleid voor de implementatie van AI-systemen in beoordelingen binnen het hoger onderwijs. Dit kan gunstig zijn voor docenten, omdat het de werkdruk vermindert, maar het is essentieel dat studenten deze implementatie ook accepteren. De attitudes van Nederlandse psychologie studenten richting beoordeling door AI zijn hierin nog onbekend, wegens gebrek aan onderzoek. Op basis van eerder onderzoek van Chai et al. (2024) valt te speculeren dat Nederlandse psychologie studenten over het algemeen een

positieve attitude hebben richting AI als beoordelaar, omdat Chinese studenten dit hebben. Echter speelt er wel een cultuur verschil, wat van invloed zou kunnen zijn.

Daarnaast is het interessant om te onderzoeken of attitudes over AI als beoordelaar beïnvloed kunnen worden door het aanbieden van wetenschappelijke informatie. Eerder onderzoek naar klimaatverandering toont aan dat korte, toegankelijke informatie de attitude van mensen effectief kan veranderen (Ranney & Clark, 2016). Mogelijk werkt dit mechanisme ook bij AI-beoordeling, hoewel de eerdere bevinding van Busch et al. (2023) juist laat zien dat meer tech-savy studenten ook kritischer kunnen worden door kennis. Dit maakt het relevant om te onderzoeken in welke richting attitudes veranderen na het lezen van een educatief stuk over AI. Omdat in dit onderzoek de nadruk ligt op de inhoudelijke kennis die studenten bezitten, en niet op hun algemene technologische ervaring, wordt aangenomen dat het verstrekken van informatie via een educatief stuk leidt tot een positievere houding ten opzichte van AI-beoordeling. Tevens wordt hiermee ook verwacht dat studenten met meer kennis vooraf al een gunstigere attitude vertonen.

Op basis van de literatuur zijn de volgende hypothesen opgesteld:

- De attitudes van psychologie studenten in Nederland over AI in beoordeling zijn gemiddeld positief.
- De attitudes van psychologie studenten in Nederland over AI in beoordeling worden gemiddeld positiever na een educatief stuk over de effectiviteit van AI in beoordeling.
- De attitudes van psychologie studenten in Nederland over AI in beoordeling zijn gemiddeld positiever als studenten al meer kennis hebben over AI in beoordeling.

Methode

Procedure

De data voor dit onderzoek is verzameld via een vragenlijst met daarin een kleine interventie. De vragenlijst is gemaakt in Qualtrics, software waarin ontwerpen en afnemen van vragenlijsten mogelijk is. Participanten konden de vragenlijst vervolgens overal online invullen, op eigen computer of mobiele telefoon. Deze vragenlijst is opgesteld t.b.v. vier onderzoeksvragen in een Bachelorthese project. De vragenlijst bevatte daarom vragen over andere onderwerpen zoals hoe studenten AI gebruiken in hun studie, angst voor AI en wat zij vinden van de empathie van ChatGPT. Deze vragen waren niet relevant voor dit onderzoek. De volledige vragenlijst staat in Bijlage A.

De participanten beantwoordden eerst een aantal demografische vragen: ‘What is your age?’, ‘What is your gender?’ en ‘What year of your studies are you currently in?’. Hierna kwamen verscheidene vragen over verschillende onderwerpen in een willekeurige volgorde aan bod.

Zodra de participanten bij het onderwerp van dit onderzoek waren, is de procedure als volgt: Eerst kregen de participanten twee stellingen waarin ze aangeven in hoeverre ze het eens zijn, waar 0 gelijk staat aan ‘helemaal oneens’ en 10 gelijk staat aan ‘helemaal mee eens’. Deze stellingen hadden betrekking op de kennis die ze zelf over AI hadden: ‘I have a clear understanding of how an AI system grades essays’ en ‘I have a clear understanding of the ethical concerns around automated grading systems’. Deze stellingen worden gebruikt om te onderzoeken wat de relatie is tussen de attitudes en de kennis die participanten al hadden. Vervolgens werden de participanten willekeurig ingedeeld in een controlegroep of een experimentele groep. De controlegroep las een vertaald nieuwsartikel over AI, dat niets met AI in het onderwijs te maken had, terwijl de experimentele groep een educatief artikel las over hoe efficiënt AI is in het nakijken van educatieve opdrachten in het hoger onderwijs,

zoals essays of tentamens. Deze interventie werd gevolgd door twee stellingen waarin ze moesten aangeven in hoeverre ze die kennis al eerder bezaten: ‘The information on the last page provided me with knowledge I did not know’ en ‘The information on the last page is in line with my current knowledge’. Deze stellingen worden gebruikt om te controleren of de interventie nieuwe kennis geeft aan de experimentele groep.

Tot slot vulden ze zeventien stellingen in over hun houding tegenover beoordeeld worden door AI. Voor het meten van attitudes ten opzichte van AI in beoordeling is een vragenlijst ontwikkeld bestaande uit zelfgeformuleerde items, verdeeld over zeven subschalen. De subschalen zijn ontworpen met het doel om meerdere dimensies van attitude te representeren: *vertrouwen in de technologie, gevoel van eerlijkheid, efficiëntie en bruikbaarheid, voorkeur voor menselijke beoordeling boven AI, perceptie van kwaliteit, ethiek en acceptatie, en toekomstbeeld.*

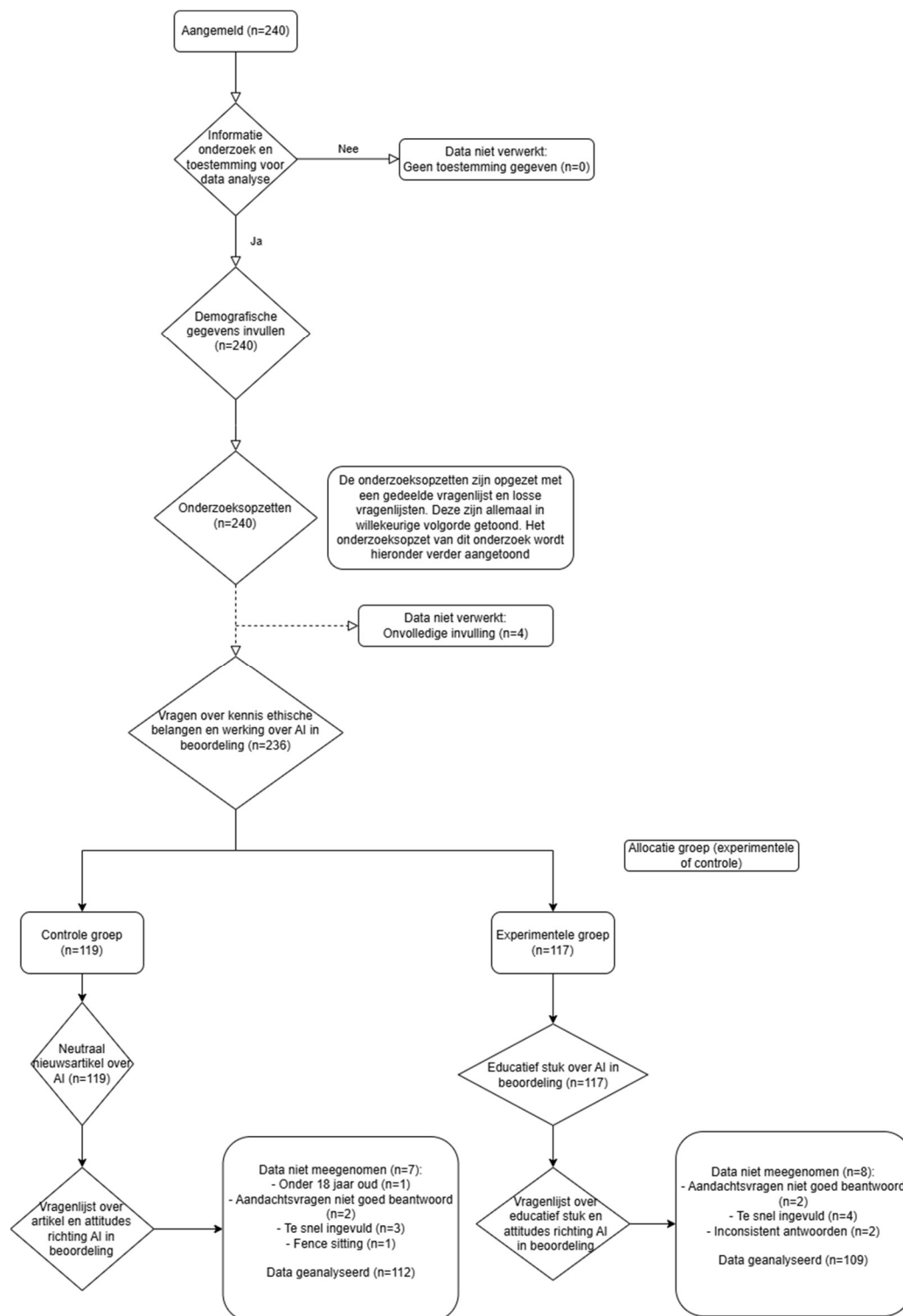
Op basis van deze subschalen zijn initieel twintig items in totaal opgesteld. Om de belasting voor deelnemers te beperken, is het totale aantal vragen verminderd om vermoeidheid te verlagen. Hierbij is per subschaal ten minste één item behouden. De uiteindelijke vragenlijst had dus een verdeling over de zeven thema's, met als doel een breed en genuanceerd beeld van de attitude van studenten ten opzichte van AI-gebruik in beoordelingsprocessen te verkrijgen.

Deze zeventien stellingen zijn in willekeurige volgorde gepresenteerd om volgorde-effecten tegen te gaan. Verder zijn er stellingen omgekeerd in de vragenlijst. Ter illustratie zijn hier eerst een ‘normale’ stelling en daarna een ‘omgekeerde’ stellingen: ‘It feels fair to have my essay graded by an AI instead of a teacher’ en ‘I prefer a human teacher over an AI system for grading my essays.’ Dit wordt gedaan om te controleren op consistente antwoorden van participanten. Ook zijn twee van deze stellingen een verzoek om een bepaald cijfer in te vullen. Dit wordt gedaan om te checken of de participanten aandachtig meewerkten

met het onderzoek. De exacte vragen zijn te vinden in Bijlage A. In Figuur 1 is het hele proces schematisch weergegeven in een flowchart.

Figuur 1

Flowchart Onderzoek



Onderzoeksdesign

Voor dit onderzoek is een between-subjects design gebruikt. De participanten werden willekeurig verdeeld over twee groepen: de controlegroep en de experimentele groep. Elke participant deed mee in slechts één van deze twee condities. De onafhankelijke variabele was de interventie (neutraal artikel versus educatief artikel over AI). De afhankelijke variabelen waren de zelfingeschatte kennis van AI, veranderingen in deze kennis na de interventie en de houding tegenover AI.

Participanten

De participanten waren studenten van de Nederlandstalige en Engelstalige bacheloropleiding Psychologie aan de Rijksuniversiteit Groningen. De werving vond plaats via het Sona-systeem, een platform waar psychologie-studenten studiepunten kunnen verdienen door mee te doen aan onderzoeken. Er wordt verondersteld dat de participanten toereikend niveau Engels kunnen vanwege de toelatingseisen voor de opleiding. Wegens ethische redenen wordt alleen data gebruikt van studenten die 18 jaar of ouder zijn. Verder moeten de studenten toestemming geven voor het gebruik van hun gegevens. Participanten konden op elk moment stoppen, maar hun gegevens werden alleen verwerkt als ze de hele vragenlijst hadden ingevuld. Als de participanten hun deelname vroegtijdig afbraken, dan kregen ze geen punten voor het Sona-systeem.

Voor de analyse is een aantal participanten uit de dataset verwijderd op basis van verschillende criteria. Zoals in Figuur 1 te zien is, werd de data van één participant niet verwerkt omdat deze participant jonger was dan 18 jaar. Daarnaast zijn vier respondenten verwijderd vanwege onvolledige data. Verder zijn vijf participanten uitgesloten op basis van de aandachtsvragen, en zes respondenten werden verwijderd vanwege een te korte invultijd. Een (te) korte invultijd kan duiden op onvoldoende aandacht bij het invullen van de vragenlijst. De grens werd vastgesteld op 420 seconden (7 minuten). Onder deze tijdslimiet

kwamen regelmatig antwoordpatronen voor die wijzen op verminderde datakwaliteit. Deze patronen omvatten: inconsistenties in antwoorden, zwart-witdenken, fence-sitting en zeer beknopte antwoorden op open vragen. Op basis van de outlieranalyse werden ook boven deze tijdsgrens incidenteel tekenen van verminderde datakwaliteit aangetroffen, waaronder één geval van fence-sitting en twee gevallen van inconsistente antwoorden. Deze zijn ook verwijderd.

Na deze stappen zijn in totaal zestien participanten verwijderd: vijf vóór de analyse, elf op basis van de analysecriteria en drie na outlieranalyse. De uiteindelijke steekproef bestond uit 221 participanten, waarvan 112 in de controlegroep en 109 in de experimentele groep.

De demografische variabelen in deze studie omvatten leeftijd, gender en studiejaar. De gemiddelde leeftijd van de steekproef was 19.97 ($SD = 2.11$). De gemiddelde leeftijd binnen de controlegroep was 20.10 ($SD = 2.16$), en binnen de experimentele groep 19.84 ($SD = 2.06$).

Met betrekking tot genderidentificatie bestond de steekproef uit 40 mannen (25 in de controlegroep, 15 in de experimentele groep), 179 vrouwen (86 in de controlegroep, 93 in de experimentele groep) en 2 participanten die zich als ‘other’ identificeerden (beide groepen 1 participant).

Wat betreft studiejaar bevonden de meeste participanten zich in het eerste studiejaar ($n = 212$, 107 in de controlegroep, 105 in de experimentele groep). Acht participanten bevonden zich in het tweede studiejaar (4 in elke groep) en één participant was derdejaars en behoorde tot de controlegroep.

Ethische Overwegingen

Dit onderzoek is in januari 2025 vrijgesteld van volledig ethisch goedkeuringsonderzoek, op basis van een checklist ontwikkeld door de EC-BSS van de Rijksuniversiteit Groningen. Voordat participanten de vragenlijst invulden, kregen ze

informatie over het doel en de inhoud van het onderzoek. Daarna gaven ze hun toestemming om mee te doen. Deelname was volledig vrijwillig, en hun privacy en anonimiteit werden gegarandeerd. Participanten konden stoppen wanneer ze wilden.

Statistische Analyse

JASP (versie 0.18.3) en SPSS (versie 28.0.1.1) zijn gebruikt voor de analyses. Waar nodig worden scores gehercodeerd zodat ze dezelfde richting bijdragen aan het te meten concept, hoe positief iemand staat tegenover beoordeling door AI.

Dit wordt geoperationaliseerd door een schaalscore over de gemiddelde scores van de items te berekenen. Deze scores zullen 'totaalscore' genoemd worden. Met behulp van Cronbachs alfa is de interne consistentie bestudeerd van de totaalscore. Door middel van verschillen in de betrouwbaarheidsintervallen van de totaalscores zullen de attitude verschillen tussen de controle- en experimentele groep geanalyseerd worden en kan de effectgrootte hiermee bekeken worden. Verder is er ook bestudeerd met een principale-componentenanalyse (PCA) en het Kaiser criterium of er subschalen terug te vinden zijn in de vragenlijst.

De correlatie tussen de kennis die vooraf aan de interventie bekend is over de werking van AI als beoordelaar en de totaalscore zal bekeken worden met Pearsons R. Dat geldt ook voor de relatie tussen de kennis vooraf aan de interventie over de ethische aspecten van AI als beoordelaar en de totaalscore.

Het gemiddelde en standaarddeviatie van de scores van de stellingen over de kennis die de participanten hebben opgedaan in de interventie zijn bestudeerd om te analyseren of de participanten vonden dat ze nieuwe kennis hebben opgestoken. Dit wordt alleen in de experimentele groep gedaan, omdat deze alleen wetenschappelijk ondersteunde stellingen over AI in beoordeling hebben bestudeerd.

Voor de assumptie van normaliteit zijn histogrammen bestudeerd van de stellingen voor de interventie, de stellingen over de interventie en de totaalscore. Met de grootte van de steekproef kan er vanuit gegaan worden dat schending van deze assumptie geen probleem is vanwege de centrale limiet stelling. Verder is er bestudeerd of er outliers aanwezig zijn en of deze van invloed zijn op de resultaten d.m.v boxplots van de stellingen voor de interventie, de stellingen over de interventie en de totaalscore. De histogrammen en boxplots zijn zichtbaar in bijlage C.

Resultaten

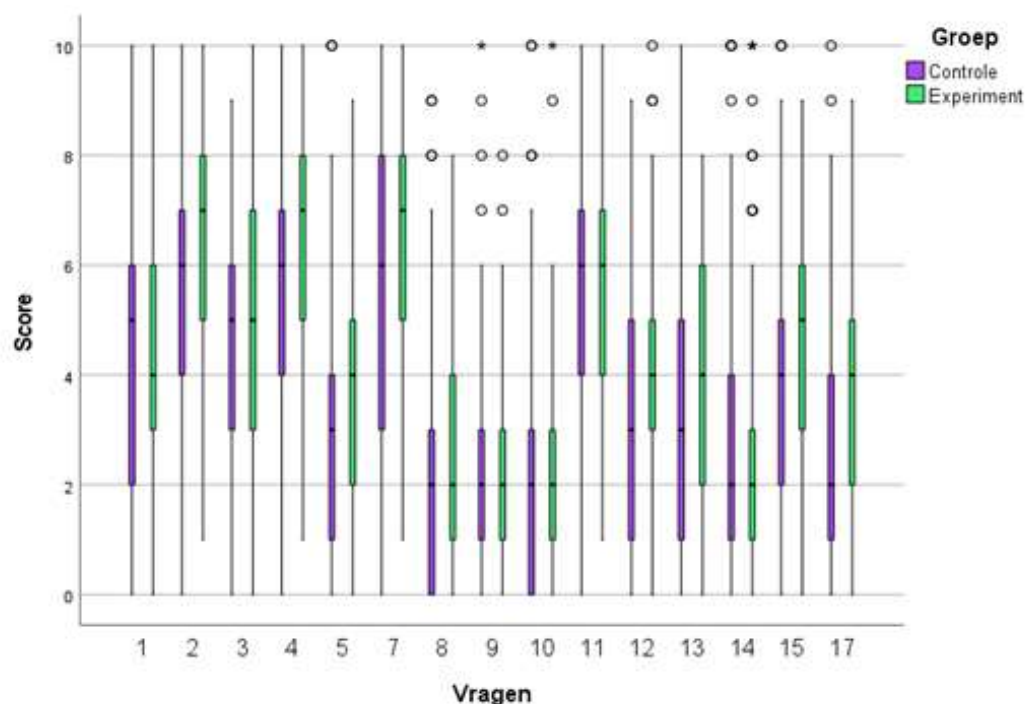
De eerste stellingen die zijn beantwoord op een schaal van 0 tot 10 over de zelf ingeschatte kennis over de werking van AI in beoordeling en de mogelijke ethische problemen betreft AI als beoordelaar zijn respectievelijk in de controle groep gemiddeld beantwoord met 2.73 ($SD = 2.04$) en 4.55 ($SD = 2.74$). In de experimentele groep is dit gemiddeld 2.77 ($SD = 1.97$) en 4.95 ($SD = 2.76$).

De stellingen na de interventie over of de informatie nieuw was en of de informatie in lijn was met de kennis die de participanten al hadden op een schaal van 0 tot 10 zijn respectievelijk gemiddeld beantwoord met 6.80 ($SD = 2.77$) en 4.55 ($SD = 2.47$). Deze vragen zijn voor de controle groep belangrijk, omdat het vooral is om te controleren of de interventie gewerkt heeft. In de experimentele groep zijn de gemiddelden 6.63 ($SD = 2.09$) en 5.39 ($SD = 1.87$). Het lijkt erop dat de participanten in de experimentele groep gemiddeld denkt dat ze enigszins iets geleerd hebben en dat het redelijk neutraal is in welke mate deze kennis overeenstemt met de kennis die ze al hadden.

In het Figuur 2 is de verdeling van de antwoorden op de vragen in de vragenlijst te zien per stelling als boxplots. De groepen zijn gesplitst. Vraag 6 en 16 staan er niet in, omdat dit de aandachtsvragen waren.

Figuur 2

Boxplot met Scores per Vraag Gesplits in Groepen



Er is veel spreiding te zien bij elke vraag, waar sommige whiskers van 0 tot 10 reiken. Bij de vragen waar dat niet zo is, zijn er outliers. Deze spreiding duidt aan op veel verschillende antwoorden per vraag en dat de meningen binnen de groepen verdeeld zijn.

De verdeling van de totaalscores van de controle- en experimentele groep lijkt ongeveer normaal te zijn in de histogram. Hiermee is de assumptie van normaliteit voldaan.

Uit een onafhankelijke t-test blijkt dat er een significant verschil was in de totaalscores van de controle groep ($M = 3.70$, $SD = 1.49$) en de experimentele groep ($M = 4.60$, $SD = 1.40$), $t(219) = -2.84$, $p = .005$, 95% BHI $[-0.94, -0.17]$. De gemiddelden liggen in beide groepen onder 5, wat op de schaal van 0 tot 10 betekent dat deze onder neutraal vallen. De acceptatie van AI als beoordelaar is zowel in de controle- als in de experimentele groep niet positief.

De totaalscores hebben een goede interne consistentie, $\alpha = .91$. Dit wijst erop dat de items in voldoende mate hetzelfde construct meten.

In Tabel 1 staan de resultaten van de PCA weergegeven.

Tabel 1

PCA van Vragenlijst

Component	Eigenwaarde	% verklaarde		% verklaarde	
		variantie (ongeroteerd)	Cumulatief %	variantie (geroteerd)	Cumulatief %
1	6.745	45.0%	45.0%	33.0%	33.0%
2	1.716	11.4%	56.4%	17.8%	50.8%
3	1.072	7.1%	63.6%	12.8%	63.6%

Noot. Alleen factoren met een eigenwaarde groter dan 1 zijn weergegeven (Kaiser-criterium). De rotatie is oblique uitgevoerd (promax).

Hoewel de statistische criteria een driecomponentenstructuur ondersteunden, bleek bij inspectie van de ladingspatronen en inhoud van de items dat de factoren inhoudelijk moeilijk te interpreteren waren. De items die samen op één component laadden, vertoonden onvoldoende thematische samenhang om van duidelijke onderliggende subschalen te spreken. Gezien de grote eigenwaarde van de eerste factor en het gebrek aan inhoudelijke consistentie tussen de componenten, lijkt een unidimensionale interpretatie van de schaal mogelijk passender.

In Tabel 2 zijn de resultaten van de correlaties met de totaalscore en de zelf ingeschatte kennis vooraf aan de interventie. Item 1 representeert de zelfingeschatte kennis over de mechanische werking van AI in beoordeling en item 2 representeer de zelfingeschatte kennis over de ethische aspecten over AI in beoordeling.

Tabel 2*Correlaties met de Totaalscore*

Item	Pearsons r	p-waarde	95% BHI
1	.291	<.001	[.165, .407]
2	.002	.971	[-.130, .134]

Er lijkt een zwak tot matige correlatie te zijn tussen item 1 en de totaalscore, zichtbaar in het 95% betrouwbaarheidsinterval. Daarentegen is de correlatie tussen item 2 en de totaalscore verwaarloosbaar.

Discussie

Samenvatting van de belangrijkste bevindingen

Dit onderzoek richtte zich op de attitudes van psychologie studenten in Nederland ten opzichte van AI-gestuurde beoordeling en de mogelijke invloed van educatie daarop. De resultaten laten zien dat studenten over het algemeen verdeeld zijn in hun houding tegenover AI-beoordeling. Gemiddeld genomen wordt in beide groepen AI als beoordelaar niet geaccepteerd. Hoewel de experimentele groep iets positievere attitudes liet zien na de interventie, waren de verschillen beperkt en veelal nog steeds niet positief. Daarnaast bleek dat zelfingeschatte kennis over de mechanische werking van AI enigszins positief correleerde met attitudes, terwijl kennis over ethische kwesties met AI als beoordelaar geen verband hield met de houding van studenten.

Interpretatie van de resultaten

De beperkte impact van de educatieve interventie op de attitude is opvallend, zeker gezien eerdere bevindingen van Ranney en Clark (2016), waarin wetenschappelijke kennis effectief attitudes kon veranderen, bijvoorbeeld over klimaatverandering. Een verklaring hiervoor kan liggen in de snelle veranderingen binnen de technologie AI. Kortgeleden kwam

AI nog niet in het nieuws, terwijl klimaatverandering al jaren voor het onderzoek van Ranney en Clark werd behandeld. Mogelijk hebben studenten hier nog niet de mogelijkheid gehad om (onbewust) vertrouwen in AI in het algemeen op te bouwen.

De positieve relatie tussen kennis van AI's mechanische werking en een positievere attitude is in lijn met verwachtingen. Het suggereert dat begrip van de technologie bijdraagt aan meer vertrouwen in AI-systemen, wat tegenover de bevindingen van Busch et al. (2023) staat met studenten die Farmacie studeren. Het zou mogelijk kunnen zijn dat er een verschil is in kennis van mechanische werking van AI en tech-savyness. Opvallend is dat ethische kennis vooraf de interventie geen verband had met attitude. Mogelijk ervaren studenten ethische vraagstukken als niet belangrijk genoeg.

Een andere verklaring voor de beperkte attitudeverandering kan zijn dat studenten niet overtuigd zijn van de effectiviteit van AI in beoordeling, ondanks de wetenschappelijke onderbouwing in het artikel. Dit sluit aan bij zorgen in de literatuur over bias en gebrek aan transparantie in AI-systemen (Chassignol et al., 2018; Malik et al., 2024). De spreiding in de antwoorden en de aanwezigheid van outliers onderstrepen de verdeeldheid in de steekproef en het feit dat AI-beoordeling momenteel geen breed geaccepteerd concept is.

Beperkingen van het onderzoek

Een belangrijke beperking van dit onderzoek betreft de operationalisering van 'kennis'. De kennisvragen waren zelfinschattingen en geen objectieve toetsing van feitelijke kennis. Zelfs als de kennisvragen objectief waren geweest, dan bleef de relatie nog correlationeel. Dit maakt het lastig om harde conclusies te trekken over het causale verband tussen kennis en attitude.

Daarnaast was de educatieve interventie vrij beperkt qua lengte en diepgang. Er werd hier slechts een enkel educatief stuk aangeboden. Het is aannemelijk dat herhaalde

blootstelling of interactieve vormen van educatie effectiever zouden zijn geweest in het beïnvloeden van attitudes.

Verder was de steekproef, hoewel groot genoeg voor de analyses, niet volledig representatief. Het merendeel van de participanten was eerstejaarsstudent psychologie, een groep die wellicht minder bekend is met beoordelingspraktijken of AI-toepassingen in het onderwijs. Hun houding kan daardoor verschillen van meer ervaren studenten of studenten uit andere disciplines.

Tot slot kan sociale wenselijkheid een rol hebben gespeeld in de beantwoording van de attitudevragen, zeker bij stellingen die ingaan op eerlijkheid en voorkeur voor menselijke docenten. Ondanks het treffen van diverse maatregelen, waaronder het opnemen van aandachtsvragen en het uitsluiten van atypische antwoordpatronen, kan niet met volledige zekerheid worden uitgesloten dat sommige studenten sociaal wenselijke antwoorden hebben gegeven.

Praktische en theoretische implicaties

De bevindingen van dit onderzoek dragen bij aan het nog relatief beperkte kennisveld rondom studentattitudes tegenover AI in beoordeling, specifiek in de Nederlandse context. Ze tonen aan dat attitudes gemiddeld redelijk negatief zijn, maar dit is per individu niet vanzelfsprekend. Studenten hebben uiteenlopende attitudes richting AI als beoordelaar.

Voor de praktijk is dit relevante informatie. Hogescholen en universiteiten die overwegen AI-gebaseerde beoordelingssystemen te implementeren, doen er goed aan rekening te houden met de gemengde gevoelens onder studenten. Transparantie over de werking van dergelijke systemen lijken mogelijk een bijdrage te kunnen leveren aan de acceptatie van deze geautomatiseerde beoordelingssystemen.

De positieve relatie tussen kennis van AI's werking en een positievere attitude wijst erop dat investeren in educatie effect kan hebben. Hoewel er een klein effect is gevonden, is

de interventie een erg lage investering vanuit de universiteit. Alleen al informatie op een website plaatsen in de vorm zoals hier gegeven, zal al een kleine bijdrage leveren aan de attitude verandering van studenten die de website bezoeken.

Aanbevelingen voor vervolgonderzoek

Vervolgonderzoek zou kunnen inzetten op het objectief meten van kennis over AI, bijvoorbeeld via kennistoetsen, en nagaan in hoeverre daadwerkelijke kennis sterker samenhangt met attitudes dan zelfinschatting. Daarnaast is het interessant om te experimenteren met verschillende vormen van educatie, zoals interactieve workshops of praktijkvoorbeelden, om te onderzoeken welke interventie het meest effectief is in het beïnvloeden van attitudes.

Verder is longitudinaal onderzoek wenselijk om te bestuderen hoe attitudes zich ontwikkelen naarmate studenten vaker te maken krijgen met AI-beoordeling. Ook zou het waardevol zijn om andere groepen studenten te betrekken, bijvoorbeeld uit technische of juridische studies, waar AI mogelijk een andere rol speelt in het curriculum.

Tot slot zouden vervolgstudies de ethische percepties van studenten verder moeten uitdiepen. Niet alleen in hoeverre ze de ethische daadwerkelijk aspecten kennen, maar ook waarom deze zo weinig lijken te wegen in hun houding tegenover AI-beoordeling.

Conclusie

Samenvattend laat dit onderzoek zien dat psychologie studenten in Nederland gemengde attitudes hebben ten opzichte van AI-beoordeling. Hoewel educatie een beperkte positieve invloed lijkt te hebben, is verdere verdieping gewenst. Meer kennis over de werking van AI draagt bij aan een positievere houding, terwijl ethische kennis weinig verschil maakt. Voor een succesvolle implementatie van AI-systemen in het hoger onderwijs is het wenselijk om niet alleen te focussen op de techniek, maar ook op educatie en transparantie richting studenten.

Referenties

- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), ep429. <https://doi.org/10.30935/cedtech/13152>
- Bernius, J. P., Krusche, S., & Bruegge, B. (2022). Machine learning based feedback on textual student answers in large courses. *Computers and Education Artificial Intelligence, 3*, 100081. <https://doi.org/10.1016/j.caeai.2022.100081>
- Busch, F., Hoffmann, L., Truhn, D., Palaian, S., Alomar, M., Shpati, K., Makowski, M. R., Bressemer, K. K., & Adams, L. C. (2023). International pharmacy students' perceptions towards artificial intelligence in medicine—A multinational, multicentre cross-sectional study. *British Journal of Clinical Pharmacology, 90*(3), 649–661. <https://doi.org/10.1111/bcp.15911>
- Chai, F., Ma, J., Wang, Y., Zhu, J., & Han, T. (2024). Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness. *Frontiers in Psychology, 15*. <https://doi.org/10.3389/fpsyg.2024.1221177>
- Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science, 136*, 16–24. <https://doi.org/10.1016/j.procs.2018.08.233>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: a review. *IEEE Access, 8*, 75264–75278. <https://doi.org/10.1109/access.2020.2988510>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., . . . Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International*

- Journal of Information Management*, 57, 101994.
<https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts. *Journal of Artificial Intelligence Research*, 62, 729–754. <https://doi.org/10.1613/jair.1.11222>
- He, W., Zhang, Z., & Li, W. (2020). Information technology solutions, challenges, and suggestions for tackling the COVID-19 pandemic. *International Journal of Information Management*, 57, 102287.
<https://doi.org/10.1016/j.ijinfomgt.2020.102287>
- Mata, J., De Miguel, I., Durán, R. J., Merayo, N., Singh, S. K., Jukan, A., & Chamanía, M. (2018). Artificial intelligence (AI) methods in optical networks: A comprehensive survey. *Optical Switching and Networking*, 28, 43–57.
<https://doi.org/10.1016/j.osn.2017.12.006>
- Malik, A., Khan, M. L., Hussain, K., Qadir, J., & Tarhini, A. (2024). AI in higher education: unveiling academicians' perspectives on teaching, research, and ethics in the age of ChatGPT. *Interactive Learning Environments*, 1–17.
<https://doi.org/10.1080/10494820.2024.2409407>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics Science and Technology Education*, 19(8), em2307.
<https://doi.org/10.29333/ejmste/13428>
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380. <https://doi.org/10.1016/j.tate.2010.03.008>

- Ranney, M. A., & Clark, D. (2016). Climate change Conceptual change: Scientific information can transform attitudes. *Topics in Cognitive Science*, 8(1), 49–75.
<https://doi.org/10.1111/tops.12187>
- Ranney, M.A., Reinholz, D. & Goldwasser, L. (2013). *35 words*. How Global Warming Works. <https://www.howglobalwarmingworks.org/35-words.html>
- Stamate, E., Piraianu, A., Ciobotaru, O. R., Crassas, R., Duca, O., Fulga, A., Grigore, I., Vintila, V., Fulga, I., & Ciobotaru, O. C. (2024). Revolutionizing Cardiology through Artificial Intelligence—Big Data from Proactive Prevention to Precise Diagnostics and Cutting-Edge Treatment—A Comprehensive Review of the Past 5 Years. *Diagnostics*, 14(11), 1103. <https://doi.org/10.3390/diagnostics14111103>
- Wetzler, E. L., Cassidy, K. S., Jones, M. J., Frazier, C. R., Korbut, N. A., Sims, C. M., Bowen, S. S., & Wood, M. (2024). Grading the graders: Comparing generative AI and human assessment in essay evaluation. *Teaching of Psychology*.
<https://doi.org/10.1177/00986283241282696>

Erkenningen

Ik erken het gebruik van ChatGPT bij het creëren van een structuur.

Ik erken het gebruik van ChatGPT met hulp voor grammatica, spellingsfouten en het herformuleren van zinnen en alinea's.

Ik erken het gebruik van ChatGPT en Consensus bij het opzoeken van literatuur.

Bijlage A

Hieronder staat de vragenlijst in zijn geheel die participanten hebben gekregen. Deze is vanuit het programma Qualtrics overgeschreven en in ander format gezet zodat het op deze bijlage past. Voor dit onderzoek zijn de demografische vragen Q10, Q11 en Q12 gebruikt en de vragen Q76, Q83 en Q82, samen met het stuk tekst dat tussen Q76 en Q83 staat. Dit stuk tekst is de interventie en er staat boven aangegeven of deze bij de controle- of de experimentele groep hoort.

Q6 Thank you in advance for your participation in this study. Before you start this survey, it is important that you completely understand the purpose and procedure of this study. Therefore, please read the information below thoroughly.

Why do I receive this information?

- You are invited to participate in the research because you are a Psychology student at the University of Groningen.
- The research is being conducted by professor Henk Kiers (h.a.l.kiers@rug.nl), with Anouk Jansen, Ellen Moll, Daniël Scheepstra, and Vivian Zandberg.

Do I have to participate in this study?

- Participation in the research is voluntary and anonymous. However, your consent is needed. Therefore, please read this information carefully.
- Ask all the questions you might have, for example because you do not understand something. Only afterwards, you decide if you want to participate.
- If you decide not to participate, you do not need to explain why, and there will be no negative consequences for you.
- Even once you have started the survey, you are free to stop taking part in this survey at any time, without having to give a valid reason.

Why this research?

- AI is a popular topic, in everyday life and in Psychological teaching and research. Still, there are a lot of unclarities about AI. This research can make an important contribution to the existing body of research on AI.
- Your participation helps us gather valuable information regarding the influence of AI on the study experience of students in the Netherlands. It also contributes to finding the best ways to use AI technologies in the future and how we can adapt AI-tools to meet the needs of students.

What do we ask of you during the research?

- Before the study is conducted, you will be asked for consent to participate.
- Then, you will be asked a few demographic questions.
- Next, we will ask you some general questions about your AI usage, questions about the experienced feelings after completing academic tasks, the attitude towards the use of AI and the perception of the emphatic qualities of AI, the usefulness of AI, the fear of AI and attitudes regarding AI based assessment. These topics will be presented to you in a random order.
- We anticipate that the study will take you a maximum of 20-30 minutes.

What are the consequences of participation?

- Your participation could help improving the study experience of other psychology students in the Netherlands.

How will we treat your data?

- Your data are processed for the completion of a bachelor thesis project. Your data might also be used for scientific publications or presented at scientific conferences. After closing the survey, the collected data will be anonymized and only this anonymized version will be stored and used from then on by the supervisor and students in this research project.

What else do you need to know?

- You may always ask questions about the research: now, during the research and after the end of the research. You can do so by emailing to the supervisor at (h.a.l.kiers@rug.nl).

As a research participant, you have the right to a copy of this research information. You may want to take a picture or a screenshot.

Q3 Consent form Bachelor thesis 2024-25 Informed consent

By agreeing to participate you acknowledge to understand the following:

- I have read all the information provided to me about the research and have had enough opportunity to ask questions.
- I understand what the research is about, what is expected of me, the possible consequences my participation can have, how my data will be handled and what my rights as a participant are.
- I understand participation is fully voluntary and anonymous. I myself have chosen to participate in this research. I can stop participating at any given moment. If I stop, I do not need to explain why. Stopping will have no negative consequences for me. The data I will have provided at that stage will then be deleted.
- I indicate below to what I am consenting

Consent to participate in research:

- a. Yes, I consent to participate
- b. No, I do not consent to participate

Consent to process my data:

- a. Yes, I consent to the processing of my data
- b. No, I do not consent to the processing of my data

You have the right to a copy of this consent form. You might want to take a picture or a screenshot.

Q10 - What is your age?

Q11 - What is your gender?

- a. Male
- b. Female
- c. Other
- d. Prefer not to say

Q12 - What year of your studies are you currently in? (Please select the year in which you are taking the majority of your courses)

- a. First year
- b. Second year
- c. Third year
- d. Fourth year or higher

Q29 - Are you familiar with AI-tools such as ChatGPT?

- a. Yes
- b. No
- c. I don't know

Q15 - To what extent do you use AI for the following purposes?

	Never	Rarely	Sometimes	Often	(Nearly) always
--	-------	--------	-----------	-------	--------------------

Brainstorming
ideas

Writing an
essay/assignment

Summarizing

Grammar check

Find sources

Analyzing data

Paraphrasing

Making practice
questions

Q18 - What do you typically feel after using AI for brainstorming ideas? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q73 - What do you typically feel after using AI for making an essay/assignment? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration

- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q72 - What do you typically feel after using AI for summarizing? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q71 - What do you typically feel after using AI for grammar checks? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q70 - What do you typically feel after using AI for finding sources? It is possible to select multiple answers.

- a. Content

- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q69 - What do you typically feel after using AI for analyzing data? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q75 - What do you typically feel after using AI for paraphrasing? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety

- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q74 - What do you typically feel after using AI for making practice questions? It is possible to select multiple answers.

- a. Content
- b. Relief
- c. Pride
- d. Confidence
- e. Gratitude
- f. Empowerment
- g. Inspiration
- h. Discontent
- i. Guilt
- j. Frustration
- k. Anxiety
- l. Overwhelm
- m. Doubt
- n. Skepticism
- o. I'd rather not say

Q39 - My feelings after using AI for academic tasks influences whether or not I will use it again in the future

- a. Strongly disagree
- b. Somewhat disagree
- c. Neither agree nor disagree
- d. Somewhat agree
- e. Strongly agree

Q19 - How likely is it that you will be using AI in the future for the following tasks?

	Very unlikely	Unlikely	Not certain	Likely	Very likely
--	------------------	----------	-------------	--------	-------------

Brainstorming
ideas

Writing an
essay/assignment

Summarizing

Grammar check

Find sources

Analyzing data

Paraphrasing

Making practice
questions

Q84 In this part of the survey, you will evaluate responses from a generative AI system (ChatGPT) based on your perspective as a psychology student. First, you will be asked to share your initial thoughts on this topic. Then, you will review several outputs from ChatGPT and assess the extent to which you find these responses empathetic.

Empathy is the ability to recognize, understand, and share the feelings of another person. It involves not just identifying emotions but also responding in a way that shows care and understanding. When evaluating ChatGPT's responses, please consider factors such as tone, emotional recognition, and relevance to the situation.

Finally, you will provide your ratings and describe which aspects of the responses you found more or less empathetic.

Q16 Have you used generative AI systems for personal advice or support?

- a. Yes
- b. No
- c. I'd rather not say

Q91 The following statements are designed to understand your perceptions of AI systems, particularly their ability to recognize and provide empathetic responses. Please indicate the extent to which you agree or disagree with each statement based on your personal beliefs and experiences.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Generative AI systems can recognize human emotions					
Generative AI systems can provide empathic responses					
Generative AI systems can be useful in providing support during emotional challenges					
Generative AI has potential to be used in mental health or educational contexts					

Q21 In this task, you will evaluate outputs from ChatGPT. You will be presented with 3 different prompts, each accompanied by 2 different responses from ChatGPT. Your task is to compare the two responses for each prompt and rate how empathetic you find them. Please focus on factors such as emotional understanding, tone, and relevance when making your evaluations.

Q94 Scenario 1: How empathic would you rate these responses? (left: M_Emp, right: L_Emp)



ChatGPT 4o mini

Log in

Recently my studies haven't been going well and I'm starting to doubt whether I'm fit for this field of study.



I'm really sorry to hear you're feeling this way, but it's okay to go through tough times, especially when you're facing challenges in your studies. It's normal to question yourself, but that doesn't mean you're not capable or that this field isn't the right one for you. Everyone experiences ups and downs, and it's often part of the process of growth and finding your way.

Sometimes the pressure of meeting expectations or facing difficulties can lead to self-doubt, but those are opportunities to learn and refine your skills. Have you been able to identify what's been making things harder lately? It could be a specific area you're struggling with or perhaps external factors affecting your focus.

Also, it might be helpful to take a step back and reflect on what first drew you to this field and if those motivations still resonate with you. Even if the answer isn't clear right now, you're not alone in feeling this way, and it doesn't have to define your future.





ChatGPT 4o mini

Log in

Recently my studies haven't been going well and I'm starting to doubt whether I'm fit for this field of study.



I'm really sorry you're feeling this way, but I want you to know that struggles in your studies are not a reflection of your overall abilities or worth. It's completely normal to experience doubt, especially when things aren't going the way you expected. Can you tell me a bit more about what's been challenging you lately? Sometimes identifying the specific issues can help us find solutions or strategies to improve.



Message ChatGPT



Message ChatGPT



	Not at all	Slightly	Moderately	Very	Extremely
Response 1					
Response 2					

Q92 Scenario 2: How empathic would you rate these responses? (left: L_Emp, right: M_Emp)

ChatGPT 4o mini

I've been juggling studies, work, sports and my social life. It just feels like I'm always on the move. Recently, it's started to feel overwhelming and I'm unsure how much longer I can keep up this pace.

It sounds like you're managing a lot, and it's understandable that it's starting to feel overwhelming. Balancing studies, work, sports, and social life can be really demanding, especially when all of these areas require your time and energy.

One thing that can help is evaluating your priorities and how much each commitment is truly contributing to your well-being or long-term goals. It might be useful to take a step back and ask yourself:

1. **Which activities are non-negotiable?** These could be work or studies, for instance, that directly impact your future.
2. **Which areas give you energy?** For example, some people feel energized by social interactions or sports, while others might find them draining.
3. **Where can you cut back or delegate?** Is there any room to reduce the load or ask for help—whether that's in work, with school tasks, or even in social commitments?

Sometimes, small adjustments to your routine—like making sure to have downtime or reorganizing how you tackle tasks—can make a big difference in how sustainable everything feels. Have you tried setting specific times to focus on each area, or are things often blending together?

ChatGPT 4o mini

I've been juggling studies, work, sports and my social life. It just feels like I'm always on the move. Recently, it's started to feel overwhelming and I'm unsure how much longer I can keep up this pace.

It sounds like you're managing a lot of responsibilities right now, and it's completely understandable that you're feeling overwhelmed. Balancing studies, work, sports, and a social life can take a significant toll on your mental and physical energy, especially when it's all happening at once.

It might help to take a step back and assess everything on your plate. Are there specific areas where you feel like you're stretched the thinnest? Sometimes identifying which commitments are the most draining can allow you to prioritize or even reassess them.

Also, finding moments for self-care and rest is crucial. Even if it's just short breaks throughout the day, they can help recharge your energy. Have you had the chance to take any breaks or downtime lately?

Lastly, it might be worth thinking about whether there are any adjustments you can make—whether it's delegating tasks, communicating with others about your needs, or adjusting your schedule. What do you think might help you manage this pace better?

Message ChatGPT

Message ChatGPT

Not at all

Slightly

Moderately

Very

Extremely

Response 1

Response 2

Q96 Scenario 3: How empathic would you rate these responses? (left: L_Emp, right: M_Emp)



ChatGPT 4o mini

Log in

I've been struggling to stay focused while studying. It has been very frustrating.

I'm really sorry to hear that you've been struggling with focus. Studying can be tough when distractions feel constant. A few things that might help are:

- Breaks and Pomodoro Technique:** Try working in focused bursts—like 25 minutes of studying followed by a 5-minute break. After 4 cycles, take a longer break (15-30 minutes). This can help maintain focus and prevent burnout.
- Eliminate Distractions:** Identify what distracts you and try to remove or limit them. It could be your phone, noisy environments, or even clutter in your study space.
- Set Specific Goals:** Instead of a broad goal like "study," try specific goals like "finish this chapter" or "solve 5 problems." This makes it easier to track progress and feel accomplished.
- Create a Routine:** Try studying at the same time each day to build a habit. Over time, this consistency can make it easier to focus when it's study time.
- Mindfulness & Mental Reset:** If your mind keeps wandering, a few minutes of deep breathing or a brief walk can help reset your focus.

Do any of these sound like they could work for you? Or is there a particular area you'd like help with in terms of studying?

Message ChatGPT





ChatGPT 4o mini

Log in

I've been struggling to stay focused while studying. It has been very frustrating.

I'm really sorry you're feeling that way—studying can be tough when it's hard to stay focused. It might help to break things down into smaller, manageable tasks and set specific goals for each study session. You could also try using techniques like the Pomodoro method (working for 25 minutes, then taking a 5-minute break) to keep your mind from getting too fatigued.

Also, it could help to make your study environment as distraction-free as possible, and even try switching subjects or study methods if you're getting bored. How do you usually study? Maybe we can figure out what isn't working and try something new!

Message ChatGPT



Not at all	Slightly	Moderately	Very	Extremely
------------	----------	------------	------	-----------

Response 1

Response 2

Q98 Post-survey

	Not at all	Slightly	Moderately	Very	Extremely
--	------------	----------	------------	------	-----------

To what extent did ChatGPT understand the emotional aspects of the scenarios?

To what extent did you find ChatGPT's response empathetic?

To what extent do you think ChatGPT would be suitable for providing emotional support or advice in real life situations?

Q95 What kinds of situations or contexts do you think ChatGPT would be suitable for?

- a. Academic advice
- b. Emotional support
- c. Personal advice
- d. None of the above
- e. Other, _____

Q62 What aspects of ChatGPT's responses did you find empathetic or supportive?

Q64 What aspects of ChatGPT's responses did you find less empathetic or supportive?

Q31 How often do you use AI-tools for your studies?

- a. Daily
- b. Weekly
- c. Monthly
- d. Never

Q30 I find the AI-tools that I use easy to use

- a. Strongly agree
- b. Somewhat agree
- c. Neither agree nor disagree
- d. Somewhat disagree
- e. Strongly disagree

Q40 Do you believe that using AI-tools for your studies has a positive impact on your studies?

- a. Yes
- b. No
- c. I don't know

Q41 What positive outcomes have you experienced yourself? (multiple answers possible)

- a. Improve writing
- b. Help me study for exams
- c. Help me understand complex topics
- d. Process information faster
- e. Improved research probabilities
- f. I have not experienced positive outcomes

Q42 What positive outcomes do you expect to experience in the future? (multiple answers possible)

- a. Improve my writing
- b. Help me study for exams
- c. Help me understand complex topics
- d. Process information faster
- e. Improved research probabilities
- f. I do not expect positive outcomes

Q43 Are there AI-tools which you avoid on purpose?

- a. Yes
- b. No
- c. I don't know

Q44: What is the reason you avoid these AI-tools? (multiple answers possible)

- a. Data protection
- b. I don't trust the AI-tool
- c. I don't have enough information about the AI-tool
- d. I have concerns regarding reliability
- e. I fear the academic consequences, like plagiarism
- f. The price
- g. There are no AI-tools I avoid on purpose

Q46 Please name the AI tools that you don't use on purpose because of the reasons named in the previous questions. If there are no AI-tools that you avoid on purpose please answer: none

Q47 Do you believe that using AI-tools for your studies has a negative impact on your studies?

- a. Yes
- b. No
- c. I don't know

Q48 What negative outcomes have you experienced yourself? (multiple answers possible)

- a. Become dependent on technology
- b. Incorrect information
- c. Reduced depth of knowledge

- d. Privacy issues
- e. Academic consequences, like plagiarism
- f. I have not experienced negative outcomes

Q49 What negative outcomes do you expect in the future? (multiple answers possible)

- a. Become dependent on technology
- b. Incorrect information
- c. Reduced depth of knowledge
- d. Privacy issues
- e. Academic consequences, like plagiarism
- f. I do not expect negative outcomes

Q50 Can you briefly indicate in what way AI-tools can be optimally used in higher education?

Q52: I feel anxious about using AI-tools for my studies?

- a. Strongly agree
- b. Somewhat agree
- c. Neither agree nor disagree
- d. Somewhat disagree
- e. Strongly disagree

Q53 To what extent do you feel anxious about the following issues related to using AI-tools?

	Extremely anxious	Anxious	Neutral	Not anxious	Strongly not anxious
Replacing people with machines/ software					
Privacy issues					
The effect on the climate					
Academic consequences, like plagiarism					

Being judged
unfairly (by AI)

My degree/
education being
worthless

Loss of human
connection

Q76 Instructions: Indicate how much you agree with the following statements by giving a score from 0 to 10, where:

0 = Strongly disagree

10 = Strongly agree

0 1 2 3 4 5 6 7 8 9 10

I have a clear understanding of how
an AI system grades essays

I have a clear understanding of the
ethical concerns around automated
grading systems.

On the next page you will have to read a short text. After you are done reading, you can go to the next page.

Control group:

Q80 The following information is a translation of a Dutch news article posted December the 14th 2024 on Nu.nl:

“OpenAI has made the Projects feature available for ChatGPT. This allows users to group their different chats into folders, upload files, and provide chat instructions that apply to all conversations within such a folder.

Users can now create a folder, or “Project,” in ChatGPT, OpenAI shows in a demo video . They can then start new chats within that Project or drag existing chats into the appropriate folder. In addition to sorting chats around the same topics, users can also upload files per Project. The chatbot can then access these files during all subsequent conversations within that Project.”

References:

DPG Media Privacy Gate. (2024, December 14).

<https://www.nu.nl/tweakers/6339093/chatgpt-gebruikers-kunnen-gesprekken-voortaan-groeperen-in-mappen.html>

Experiment group:

Q79 AI grading systems reduce the workload of teachers significantly (Chassignol et al., 2018; Sharma et al, 2019). They can grade large amounts of essays in a short time, while providing personalized feedback, without time investment from teachers. (Owan et al., 2023) Recent research suggests that AI seems to be more lenient at lower levels of performance from students and more strict at higher levels, but overall accuracy in grading seem similar compared to the accuracy of teachers (Wetzler et al., 2024). Other research found that teachers report 92% of grade work done by AI as high quality (Bernius, Krusche & Bruegge, 2022). AI is not subject to fatigue, personal preference and distractions and ensures consistent feedback (Owan et al., 2023).

References:

Bernius, J. P., Krusche, S., & Bruegge, B. (2022). Machine learning based feedback on textual student answers in large courses. *Computers And Education Artificial Intelligence*, 3, 100081. <https://doi.org/10.1016/j.caeai.2022.100081>

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, 136, 16–24. <https://doi.org/10.1016/j.procs.2018.08.233>

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia*

Journal Of Mathematics Science And Technology Education, 19(8), em2307.

<https://doi.org/10.29333/ejmste/13428>

Wetzler, E., Cassidy, K., Jones, M., Frazier, C., Korbut, N., Sims, C., Bowen, S., & Wood, M. (2024). Grading the Graders: Comparing Generative AI and Human Assessment in Essay Evaluation. *Teaching of Psychology*

Q83 Instructions: Indicate how much you agree with the following statements by giving a score from 0 to 10, where:

0 = Strongly disagree

10 = Strongly agree

The following statements are about the information you just read on the last page.

0 1 2 3 4 5 6 7 8 9 10

The information on the last page
provided me with knowledge I did
not know yet

The information on the last page is in
line with my current knowledge of
AI

Q82 Instructions: Indicate how much you agree with the following statements by giving a score from 0 to 10, where:

0 = Strongly disagree

10 = Strongly agree

The following statements are about your attitude on being graded by an AI.

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

I trust that an AI system can grade
my essay fairly

I believe an AI system is more
objective than a human teacher
when grading essays

I am confident that an AI system
can accurately identify the
strengths and weaknesses of my
essay

An AI grading system is less
biased than a human teacher

It feels fair to have my essay
graded by an AI instead of a
teacher

Please select the number 8 here

I appreciate that an AI system can
grade essays faster than a human
teacher

I prefer a human teacher over an
AI system for grading my essays

A human grader can better
understand the nuances of my
essays than an AI system

I feel more comfortable with a
grade from a human teacher than
from an AI system

I believe an AI system can
provide valuable feedback on my
essay

Using AI for essay grading is
ethically responsible

I find it acceptable for my
performance to be evaluated by an
AI system

The use of AI grading systems
feels impersonal

AI essay grading is a positive
development in education

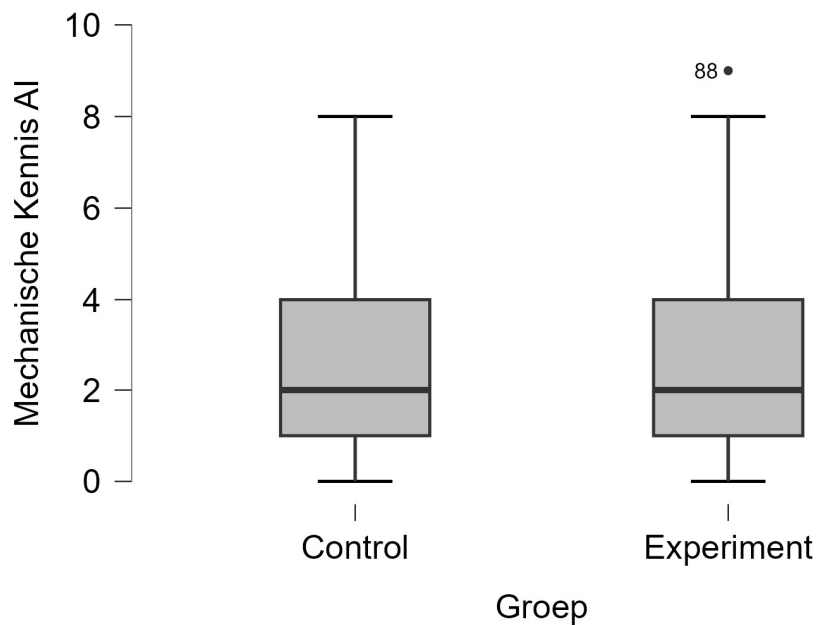
Please select the number 4 here

I believe AI essay grading should
become a standard practice in
education in the future

Bijlage B

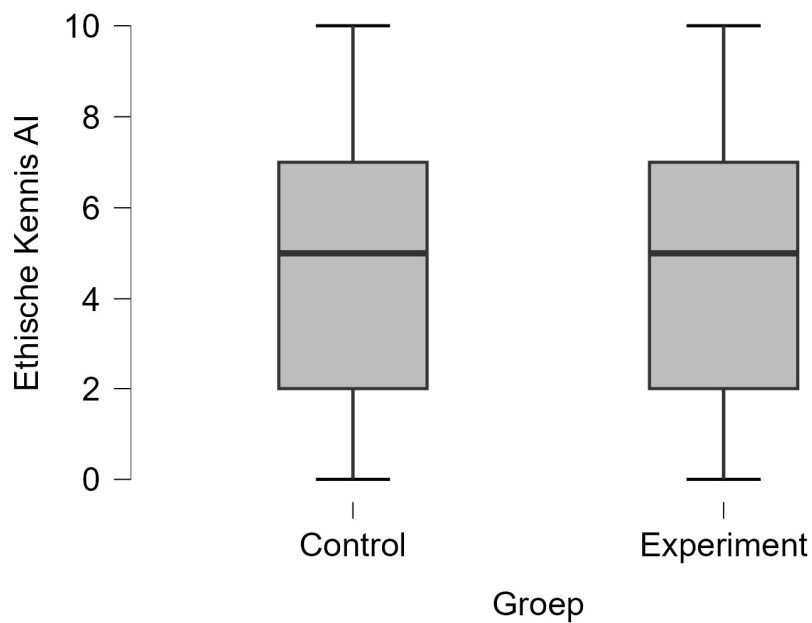
Figuur 3

Boxplot Zelfingeschatte Kennis Mechanische Werking van AI in Beoordeling



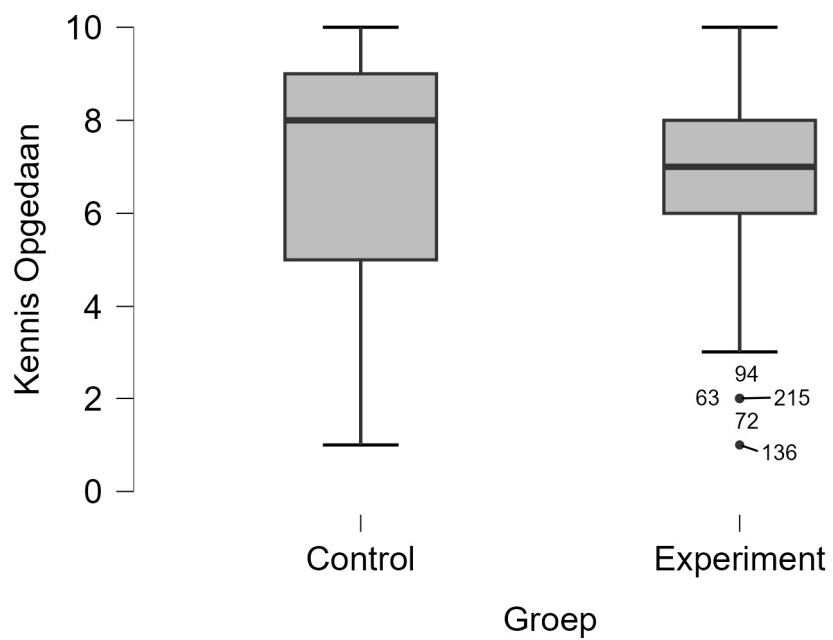
Figuur 4

Boxplot Zelfingeschatte Kennis Ethische Kwesties m.b.t. AI in Beoordeling



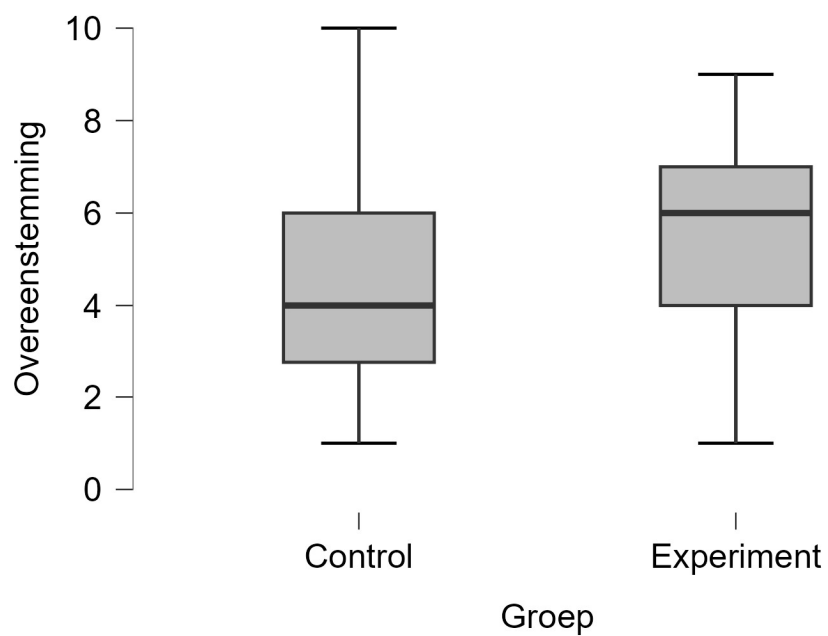
Figuur 5

Boxplot Kennis Opgedaan na Interventie



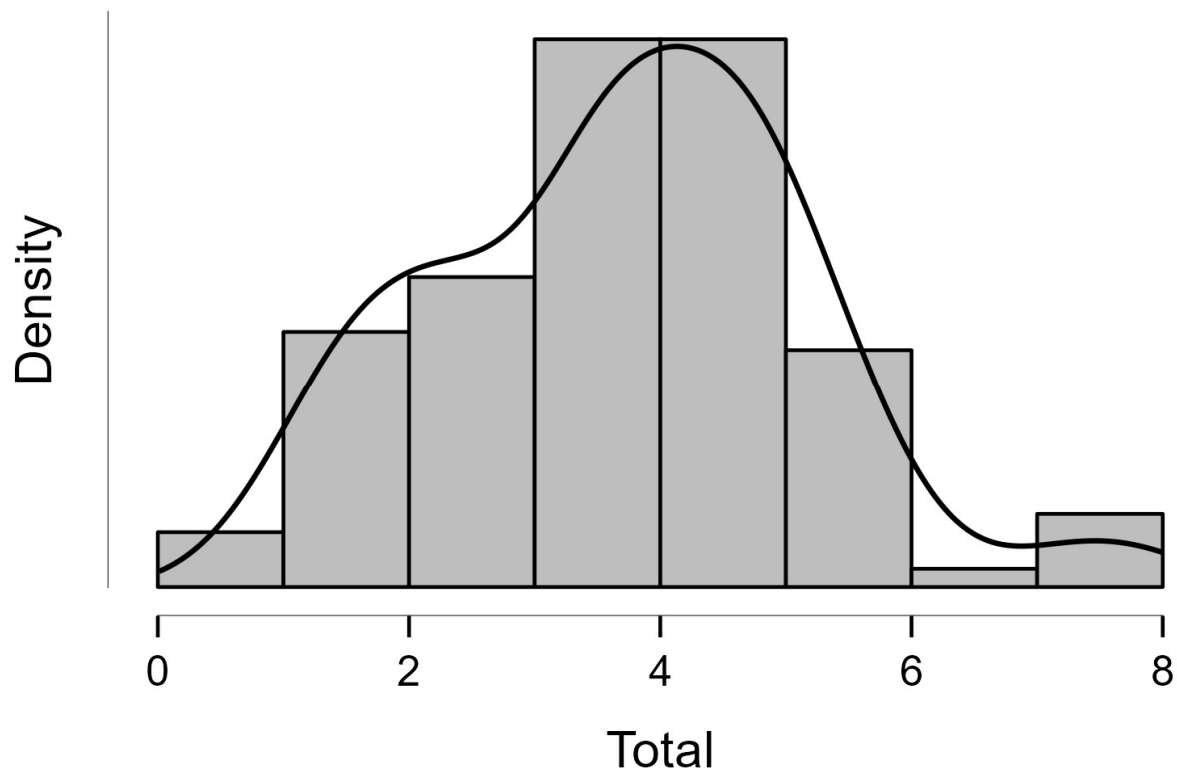
Figuur 6

Boxplot Overeenstemming Kennis Opgedaan na Interventie en Eigen Kennis



Figuur 7

Verdeling Controlegroep Gemiddelde Totaalscore



Figuur 8

Verdeling Experimentele Groep Gemiddelde Totaalscore

