

Een Onderzoek naar de Invloed van Drie Verschillende Beoordelingsmethoden in het mbo.

Universiteit Utrecht

Imke van Kats (3622908)

Inleverdatum: 12 juni 2015

Aantal woorden: 7983

1^o beoordelaar: dr. Renske de Kleijn

2^o beoordelaar: dr. Frans Prins

Externe begeleiders: Jorick Scheerens en Anne-Marie Gootjes

Themagebied: Feedback in Interactie

Opleiding: Onderwijskundig Ontwerp en Advisering

Faculteit: Sociale Wetenschappen

Abstract

In deze studie is met behulp van een experiment onderzocht welke invloed de manier van beoordelen heeft op de taakbeleving, ervaren autonomie, ervaren competentie, de bruikbaarheid van feedback en de docent- en studentervaringen over de bruikbaarheid van de beoordeling. De drie onderzochte beoordelingsmethoden zijn rubrics, de studentbetrokkenheid bij het opstellen van beoordelingscriteria en een controleconditie. In totaal zijn bij 72 studenten zowel kwantitatieve data (vragenlijsten) als kwalitatieve data (aanvullende vragen studenten en interview met de docent) verzameld. Uit de kwantitatieve analyses blijkt dat de bruikbaarheid van feedback, de mate van ervaren autonomie en competentie en taakbeleving niet significant verschillen tussen de drie condities. Deze resultaten komen niet overeen met de verwachtingen van het onderzoek. Ondanks het ontbreken van verschillen tussen de condities in de kwantitatieve data is wel duidelijk dat de docent de rubric een bruikbare beoordelingsmethode vindt en dat studenten beide nieuwe beoordelingsmanieren accepteren en waarderen. Verder laat het onderzoek zien dat experimenteren met beoordelingsmethoden kan leiden tot bewustwording bij docenten over het belang van een duidelijke en transparante beoordelingsmethode. In vervolgonderzoek kan onderzocht worden wat de invloed is van het gekregen cijfer op de waardering van de beoordelingsmethode.

Sleutelwoorden; beoordelen, mbo, feedback, autonomie, competentie, taakbeleving

Inleiding

Door prestaties van studenten te beoordelen kunnen docenten onder andere vaststellen of een student de leerstof heeft begrepen. Bij het beoordelen van prestaties is het van belang dat duidelijk is aan welke standaarden de prestatie moet voldoen. Deze standaarden, ook wel beoordelingscriteria genoemd, zijn de kenmerken waarmee de kwaliteit van een prestatie kan worden beoordeeld (Sadler, 1989). Beoordelen is onlosmakelijk verbonden met het werk van een docent en al zo oud als het onderwijs zelf. De voortdurende ontwikkelingen in het onderwijs zorgen ervoor dat de manier van beoordelen ook constant aan veranderingen onderhevig is. In het mbo is op dit moment het competentiegericht onderwijs de leidraad. Het beheersen van competenties kan op verschillende manieren worden aangetoond waardoor de beoordelingscriteria ruimer zijn opgesteld (Elshout-Mohr, Oostdam, & Overmaat, 2002). Door de ruimere beoordelingscriteria is het van belang dat de docent vooraf communiceert met de student zodat duidelijk is waarop wordt beoordeeld (Sol & Stokking, 2009).

In de praktijk lijkt het echter nog weleens te gebeuren dat de docent eerst de leertaak ontwerpt en achteraf pas nadenkt over de beoordeling. Zo ontstaat er een gebrek aan afstemming tussen de leertaak, de instructie en de beoordeling. Met als gevolg dat studenten negatief verrast kunnen worden door het gebrek aan congruentie tussen de beoordelingsmethode en de inhoud van de stof (Broekkamp & van Hout-Wolters, 2007; Sluijsmans, Straetmans, & van Merrienboer, 2008). Het gebrek aan transparantie en congruentie in de beoordelingscriteria kan onder andere leiden tot een verminderde taakinzet of verkeerde interpretatie van de feedback (Black & William, 1998; Boekaerts, 1992).

In dit onderzoek is onderzocht welke invloed de manier van beoordelen heeft op de taakbeleving, ervaren autonomie, ervaren competentie, de bruikbaarheid van feedback en de docent- en studentervaringen over de bruikbaarheid van de beoordeling. Het is voor mbo scholen relevant te weten welke invloed beoordelen op hun studenten heeft en welke mogelijkheden zij hebben om beoordelingscriteria duidelijker en transparanter te maken. Wellicht kan dit onderzoek bij docenten die nog geen transparante beoordelingsmethode hanteren, bijdragen aan een verandering van denken over het belang van het opstellen van een duidelijke beoordelingsmethode.

Theoretisch Kader

In de praktijk blijken docenten en studenten niet altijd dezelfde perceptie te hebben over de beoordelingscriteria (Carless, 2006; Merry, Orsmond, & Reiling, 2000). De beoordelingscriteria worden meestal individueel door een docent bepaald, vaak zelfs op het moment dat het gegeven onderwijs al voorbij is en alleen de beoordeling nog moet plaatsvinden (Sluijsmans, 2008). Worden de beoordelingscriteria wel van tevoren kenbaar gemaakt dan vaak in de vorm van een lijst met beoordelingscriteria (Lea & Street, 1998; Rust, Price, & O'Donovan, 2003). Het uitdelen van een lijst met beoordelingscriteria is echter niet genoeg om de 'tacit' kennis van de docent over te brengen op de student. De beoordelingscriteria zitten vooral in het hoofd van de docent waardoor de studenten de 'taal van de criteria' niet begrijpen (Rust et al., 2003; Sadler, 1989). Studenten weten niet waar ze aan moeten voldoen en kunnen de criteria niet gebruiken om hun leerprestaties te verhogen (Lea & Street, 1998). Onderzoek van Broekkamp, Van Hout-Wolters, Van den Bergh & Rijlaarsdam (2004) liet zien dat de verwachtingen van leerlingen in 5 VWO over een geschiedenisproefwerk nauwelijks overeen kwamen met de verwachtingen van de docent. Het is belangrijk dat de beoordelingscriteria vroegtijdig helder zijn voor alle betrokkenen om misconcepties in de verwachtingen te voorkomen (Broekkamp & van Hout-Wolters, 2007; Cluitmans & Klarus, 2005).

Belang van Transparante Beoordelingscriteria

Goed omschreven opdrachtdoelen en beoordelingscriteria kunnen het leerproces positief beïnvloeden (Black & William, 1998). Studenten zijn meer geneigd om hun taakinzet te vergroten als het beoogde doel van de opdracht helder is (Kluger & DeNisi, 1996). Onduidelijkheid kan ertoe leiden dat studenten zich minder inspannen om de discrepantie te verminderen tussen het huidige werk en het beoogde doel (Sadler, 1989). Als de beoordelingscriteria en standaarden van een opdracht worden geëxpliciteerd dan worden de studenten bewuster van het feit waarom zij iets moeten leren (Nicol & Macfarlane-Dick, 2006). Door te communiceren over het beoordelingsproces en de criteria zijn studenten beter voorbereid en kunnen ze de opdracht zelfstandiger uitvoeren (Baartman, 2008; Sluijsmans, 2008). Duidelijke beoordelingscriteria hebben ook tijdens en na afloop van de opdracht positieve effecten. De feedback over de vorderingen van de opdracht kan bijvoorbeeld gekoppeld worden aan de beoordelingscriteria. De feedback is daardoor beter te interpreteren dan wanneer de

feedback wordt gegeven op basis van normen ten opzichte van andere studenten (Marzano, Pickering en Pollock, 2001). Een manier om te communiceren is door studenten te betrekken bij het construeren van beoordelingscriteria (Nicol, 2009).

Studenten Betrekken bij Beoordelingscriteria

Door studenten te betrekken bij het opstellen, is de kans groter dat de studenten zich meer betrokken voelen bij de leerinhoud (Hinnel & Weeden, 2000). De student is niet per definitie verantwoordelijk voor de beoordeling, maar krijgt een rol in de bepaling van wat beoordeeld wordt (Sommervell, 1993). Het betrekken van studenten kan leiden tot betere interpretatie van de criteria dan wanneer ze zijn samengesteld door de docent (Nicol, 2010; Price & O'Donovan, 2006).

Studentbetrokkenheid bij het creëren van beoordelingscriteria kan door het aanbieden van voorbeeldopdrachten van verschillende prestatieniveaus, studenten kunnen op basis daarvan beoordelingscriteria ontwikkelen (Orsmond, Merry, & Reiling, 2002). De voorbeelden maken expliciet wat de vereisten van de opdracht zijn. Ze definiëren een standaard waarmee het eigen werk kan worden vergeleken. Voorbeelden laten zien dat kwaliteit niet een vaststaand iets is maar op verschillende manieren uitgewerkt kan worden (Sadler, 1989). Orsmond et al., (2002) vonden bij deze vorm van studentbetrokkenheid een significante verbetering van de leerprestaties ten opzichte van de controleconditie die niet betrokken werden bij de beoordelingscriteria. Het meedenken over de criteria leidt tot actieve betrokkenheid waardoor betere connecties gemaakt kunnen worden tussen de ontvangen feedback en het eigen werk (Sadler, 1998). Het aanbieden van uitgewerkte voorbeelden en vervolgens het gestructureerd zelf formuleren van beoordelingscriteria vormt de eerste experimentele conditie in dit onderzoek.

Rubrics als Beoordelingsinstrument

Een alternatieve manier van beoordelen die ook zorgt voor de benodigde communicatie over de beoordelingscriteria zijn *rubrics*. Rubrics geven docenten een raamwerk om de beoordelingscriteria vooraf te bespreken met studenten (Arter & Chappuis, 2007). Rubrics zijn beoordelingsschalen waarin de vaardigheden worden beoordeeld door middel van een aantal criteria (Rochford & Borchert, 2011). Per criterium zijn drie of vier niveaus te onderscheiden variërend van onvoldoende tot uitstekend. Per niveau is beschreven waar de student aan moet voldoen om het niveau te behalen (Andrade, 2000).

Rubrics zijn effectief voor het begrijpen van de eisen van een opdracht (Reddy & Andrade, 2010). Als rubrics worden gebruikt om een cijfer voor het geleverde werk te geven, hebben ze de functie van een summatief beoordelingsinstrument (Reddy & Andrade, 2010). Rubrics kunnen ook dienen als basis voor het geven van formatieve feedback door de docent (Wiggins, 2012). Rubrics bieden namelijk de mogelijkheid om tijdens het leerproces te vertellen welk niveau tot nu toe is gehaald en naar welk niveau gestreefd moet worden (Arter & Chappuis, 2007). Het wordt voor de docent gemakkelijker om goede feedback te geven en voor de student gemakkelijker om de feedback te interpreteren (Wiggins, 2012). In dit onderzoek vormen rubrics als beoordelingsinstrument de tweede experimentele conditie.

Naast de twee experimentele condities bevat dit onderzoek ook een controleconditie. In de controleconditie heeft de docent in de opdrachtbeschrijving een paar criteria opgesteld. Van tevoren heeft echter geen mondelinge communicatie plaatsgevonden over de beoordelingscriteria. Ook zijn de criteria niet concreet opgesteld. Een voorbeeld van een gehanteerd criterium is: *'Het verslag moet een duidelijke en overzichtelijke inhoud hebben'*.

Effecten van de Beoordelingsmethoden

De verwachting is dat de beoordelingsmethoden in de experimentele condities effect hebben op de studentperceptie over de bruikbaarheid van feedback, ervaren autonomie en competentie, taakbeleving en de docent- en studentervaringen over de bruikbaarheid van de beoordeling.

Studentperceptie over de bruikbaarheid van feedback. Feedback is een van de belangrijkste voorspellers van goede leerresultaten mits het op de juiste manier wordt ingezet (Hattie & Timperley, 2007; Voerman, 2014). Feedback is effectief als het zo specifiek mogelijk is en informatie en aanwijzingen geeft over de taakuitvoering en het te bereiken doel (Hattie & Timperley, 2007). Uit recent onderzoek van Voerman (2014) onder 78 voortgezet onderwijs docenten blijkt dat nog geen 20 % van de geobserveerde interventies van docenten in interactie met leerlingen te typeren valt als feedback. Was er wel sprake van feedback dan meestal in de vorm van niet specifieke uitingen. Terwijl juist specifieke feedback door studenten gezien wordt als bruikbaar (Weaver, 2006). Feedback wordt niet als bruikbaar ervaren wanneer het ontbreekt aan sturing, te negatief of algemeen is, of wanneer het niet gerelateerd is aan de beoordelingscriteria (Weaver, 2006). In beide experimentele condities worden de beoordelingscriteria van tevoren geëxpliciteerd waardoor ook het doel van de

opdracht duidelijker wordt. Daardoor is het beter mogelijk om specifieke feedback te geven op de aanpak en resultaten dan wanneer de beoordelingscriteria niet helder zijn zoals in de controleconditie. De verwachting is dat in beide experimentele condities de feedback specifiek kan worden gegeven en daardoor ook beter geïnterpreteerd kan worden door studenten (Black & William, 1998; Wiggins, 2012). Daarom wordt in dit onderzoek de studentperceptie over de bruikbaarheid van feedback als uitkomstvariabele meegenomen.

Ervaren autonomie en competentie. Volgens de zelfdeterminatietheorie van Ryan & Deci (2000) vormen autonomie en competentie samen met verbondenheid de drie psychologische basisbehoeften die een positieve invloed hebben op de intrinsieke motivatie. Studenten die intrinsiek gemotiveerd zijn hebben een betere kwaliteit van leren dan studenten die extrinsiek gemotiveerd zijn (Ryan & Deci, 2000). Het is dus van belang om de intrinsieke motivatie van studenten zoveel mogelijk te stimuleren door in het onderwijs te voldoen aan de drie psychologische basisbehoeften. In dit onderzoek worden de mate van ervaren autonomie en competentie onderzocht.

Autonomie wordt ervaren wanneer de mogelijkheid bestaat om zelf een keuze of beslissing te maken (Deci & Ryan, 2008). Autonomie kan worden aangeboden door actiekeuzes en optiekeuzes (Ryan & Deci, 2000). Een actiekeuze is een keuze om wel of niet deel te nemen aan een opdracht. De persoon bepaalt zelf of, wanneer en hoelang de activiteit wordt gedaan (Ryan & Deci, 2000). Een optiekeuze is een keuze uit een aantal opties die iemand anders aandraagt (Ryan & Deci, 2000). In de conditie waarbij de studenten zelf de criteria mogen opstellen krijgen studenten optiekeuzes aangeboden. Ze bepalen zelf waar ze op worden beoordeeld waardoor voldaan kan worden aan de beleving van autonomie. Door de actieve rol die studenten krijgen in het beoordelingsproces is de verwachting dat ze zichzelf gaan zien als eigenaar van het leerproces en dus meer autonomie ervaren (Castelijn & Andersen, 2013).

Competentie kan worden ervaren door het aanbieden van structuur (Ryan & Deci, 2000). Het aanbieden van structuur kan in de vorm van duidelijke instructie, het expliciteren van leerdoelen en tussentijdse feedback. Het gebruik van rubrics zorgt voor structuur door de heldere criteria en de communicatie over de verwachtingen. Studenten krijgen inzicht in hun persoonlijke vooruitgang en sterke en zwakke punten waardoor het kan bijdragen aan hun gevoel van competentie (Jonsson &

Svingby, 2007). De mate van ervaren competentie in combinatie met de taakeisen van een opdracht spelen een rol in hoe een taak wordt beleefd (Boekaerts, 1992).

Taakbeleving. Als studenten een leertaak krijgen aangeboden zullen ze een inschatting maken over de taakeisen en de benodigde competenties om aan de eisen te voldoen. Deze inschattingen vormen de taakbeleving en zijn afhankelijk van de taakkenmerken, de manier waarop de taak wordt aangeboden, de generieke motivatie en de stabiele eigenschappen van een persoon (Boekaerts, 1992).

De taakbeleving is de subjectieve beleving die bestaat uit gevoelens over het nut van de opdracht, de haalbaarheid, moeilijkheid en aantrekkelijkheid van een opdracht (Boekaerts, 2002). Als studenten de opdracht niet te moeilijk vinden en zich bekwaam voelen om aan de eisen te voldoen ervaren ze controle over de situatie (Boekaerts, 1992). Als de taakeisen echter hoger worden geschat dan de benodigde vaardigheden, wordt de opdracht als moeilijk en onhaalbaar ervaren waardoor taakangst kan ontstaan. De verwachting is dat de explicitering in de beoordelingscriteria in beide condities leidt tot een positievere taakbeleving. Door het verduidelijken van beoordelingscriteria en studentbetrokkenheid bij de criteria weten studenten beter wat er van ze verwacht wordt waardoor de taakangst minder wordt en de taakinzet en succesverwachting hoger (Hattie, 2009). Ook wordt door de explicitering en betrokkenheid bij de taakeisen, de moeilijkheidsgraad van de opdracht minder hoog ingeschaald waardoor de student zich competentier voelt om de opdracht uit te voeren (Boekaerts, 1992). Het is interessant om de invloed van beoordelingsmethoden op taakbeleving te onderzoeken want de taakbelevingen bepalen de bereidheid om tijd en energie te steken in het bereiken van het leerdoel (Seegers & Boekaerts, 1993). Een hoge taakbeleving is van belang want het draagt bij aan de wil om te leren (Boekaerts, & Simons, 1995).

Docent- en studentervaringen over de bruikbaarheid van de beoordeling. De bruikbaarheid van een beoordelingsmethode wordt bepaald door de validiteit, betrouwbaarheid, onderwijsconsequenties, acceptatie en kosten (van der Vleuten, 1996). De meeste elementen kunnen objectief worden vastgesteld behalve acceptatie. Acceptatie betekent in hoeverre betrokkenen zich kunnen vinden in de opzet en uitvoering van het gekozen assessment (Baartman, 2008). Acceptatie door studenten en docenten wordt bepaald door zowel praktische overwegingen als persoonlijke

ervaringen, overtuigingen en (mis)concepties (van der Vleuten, 1996). Acceptatie door de betrokkenen is een cruciaal element voor de aanvaarding en bruikbaarheid van een beoordelingsmethode (van der Vleuten, 1996). Zonder acceptatie kan een nieuwe beoordelingsmethode niet succesvol geïmplementeerd worden in het onderwijs. In dit onderzoek worden de docent- en studentervaringen over de beoordeling meegenomen om te onderzoeken in hoeverre de beoordelingsmethoden worden geaccepteerd en als bruikbaar worden ervaren.

Onderzoeksvraag en Hypothesen

De onderzoeksvraag die in dit onderzoek centraal staat is:

In hoeverre heeft de manier van beoordelen invloed op de studentperceptie over de bruikbaarheid van feedback, de mate van ervaren autonomie en competentie, de taakbeleving en invloed op de docent- en studentervaringen over de bruikbaarheid van de beoordeling?

Bijbehorende deelvragen zijn:

1. In hoeverre zorgen de verschillende beoordelingsvormen voor een verschil in de studentperceptie over de bruikbaarheid van feedback?

Hypothese I: De verwachting is dat in beide experimentele condities hoger gescoord wordt op de bruikbaarheid van feedback dan in de controleconditie.

2. In hoeverre zorgen de verschillende beoordelingsvormen voor een verschil in de mate van ervaren autonomie en ervaren competentie van studenten?

Hypothese II: Verwacht wordt dat in de betrokkenheidconditie hoger wordt gescoord op de ervaren autonomie dan in de andere twee condities. Ook wordt verwacht dat in de Rubric-conditie hoger wordt gescoord op de ervaren competentie dan de andere twee condities.

3. In hoeverre hebben de verschillende beoordelingsvormen een positieve invloed op de subschalen van taakbeleving?

Hypothese III: Verwacht wordt dat de subschalen van taakbeleving een grotere groei in score (nameting ten opzichte van voormeting) laten zien in de twee experimentele condities dan in de controleconditie.

4. In hoeverre hebben de verschillende beoordelingsvormen invloed op de docent- en studentervaringen over de bruikbaarheid van de beoordeling?

Methode

Onderzoeksdesign

Het onderzoek is zowel kwantitatief als kwalitatief van aard, ook wel ‘mixed-method design’ genoemd. Het kwantitatieve deel heeft de vorm van een quasi-experimenteel design met metingen op twee tijdstippen en drie verschillende condities. De studenten zijn op basis van de bestaande klassen ingedeeld in drie condities.

Conditie 1, zelf beoordelingscriteria opstellen. De studenten mochten in deze conditie zelf de beoordelingscriteria opstellen aan de hand van voorbeeldopdrachten en voorbeeldbeoordelingscriteria. Ook is een formulier met instructies en restricties uitgedeeld waarop de student de zelfbedachte criteria moest schrijven, zie bijlage A. Deze conditie wordt ook wel de betrokkenheidconditie genoemd.

Conditie 2, Rubrics. In deze conditie werd beoordeeld met een rubric, zie bijlage B. De rubric is opgesteld op basis van de leerdoelen en de eisen van de opdracht. De rubric is geconstrueerd door de onderzoeker in samenspraak met de docent en een onderwijskundig beleidsmedewerker van de school.

Conditie 3, controleconditie. In deze conditie is de beoordeling gegaan zoals de desbetreffende docent altijd beoordeeld. De criteria zijn niet specifiek geformuleerd en de schriftelijk opgestelde criteria zijn niet mondeling toegelicht aan de studenten, zie bijlage C.

De eerste meting werd verricht na globale uitleg van de opdracht. De tweede meting vond vijf weken later plaats, na het inleveren van de opdracht. De kwalitatieve gegevens bestonden uit aanvullende vragen aan de studenten bij de nameting en een interview met de docent.

Deelnemers

In totaal hebben 72 van de 80 eerstejaars studenten van drie bestaande klassen mediamanager en mediaredactie van een mbo in Amsterdam deelgenomen. Klas MR1A vormt de betrokkenheidconditie ($N = 25$), klas MM1B de rubric-conditie ($N = 26$) en klas MM1A de controleconditie ($N = 21$). Eén student uit MR1A, twee uit MM1B en vijf studenten uit MM1A hebben niet deelgenomen aan het onderzoek vanwege afwezigheid in de klas. De gemiddelde leeftijd van de studenten was 17.3 jaar ($SD = 1.08$). In totaal hebben 32 jongens en 40 meisjes geparticipeerd. Bij de

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

variabele taakbeleving hebben in totaal 62 studenten, 26 jongens en 38 meisjes zowel de voor- als nameting ingevuld. De gemiddelde leeftijd was 17.3 jaar ($SD = 1.10$). Er zijn geen significante verschillen tussen de groepen op basis van leeftijd en geslacht. Op basis van deze gegevens en de informatie van de docent over de leerprestaties lijken de klassen vergelijkbaar.

Instrumenten

Alle variabelen zijn gemeten met een 5-punts Likertschaal (1: *helemaal mee oneens tot 5: helemaal mee eens*). Zie Tabel 1 voor de Cronbach's Alfa's van alle variabelen. De gehele vragenlijst is weergegeven in bijlage D.

Tabel 1

Cronbach's Alfa's en Voorbeelditems van de Bruikbaarheid van Feedback, Autonomie, Competentie en de Subschalen van Taakbeleving

Variabele	Aantal items	Voormeting α	Nameting α	Voorbeelditem
Bruikbaarheid van feedback	8		.94	'De feedback was bruikbaar.'
Autonomie	7		.73	'Ik geloof dat ik in bepaalde zaken keuzevrijheid had tijdens deze opdracht.'
Competentie	6		.85	'Ik denk dat ik best goed was in deze opdracht.'
Taakangst	5	.82	.74	'Hoe voel je je nu de taak is afgerond?'
Succesverwachting	2	.79	.64	'Wat voor cijfer verwacht je te kunnen halen voor deze opdracht?'
Taak-aantrekkelijkheid	4	.79	.66	'Ik vond de taak leuk om aan te werken.'
Taakinzet	3	.60	.69	'Hoe zorgvuldig heb je de opdracht gedaan?'
Taakimportantie	2	.56	.44	'Hoe belangrijk heb je deze taak gevonden?'

Vragenlijst bruikbaarheid van feedback. De studentperceptie over de bruikbaarheid van feedback is gemeten met een doorontwikkelde enquête gebaseerd op de vragenlijst 'percepties over de acceptatie en bruikbaarheid van feedback' van Sol en Stokking (2009). De geconstrueerde vragenlijst bevat items over de bruikbaarheid, duidelijkheid, acceptatie en leerzaamheid van de feedback.

Vragenlijst ervaren autonomie. De mate van ervaren autonomie is een subschaal van de Intrinsic Motivation Inventory (IMI; Ryan, 1982). De zeven items maten vooral of er sprake was van

actiekeuzes terwijl in dit onderzoek alleen optiekeuzes werden aangeboden. Besloten is om drie items die actiekeuzes meten te vervangen door drie items die optiekeuzes meten. Met een pilot onder 24 respondenten van één klas eerstejaars Mediamanagers die niet meededen aan het experiment is onderzocht of de opnieuw geconstrueerde vragenlijst betrouwbaar is. In de pilot zijn vijf optiekeuze items geconstrueerd om vervolgens de drie met de beste Cronbach's Alfa's te kiezen. Twee extra items zijn opgenomen om de best geconstrueerde items te kunnen selecteren. Na verwijdering van de twee items met de laagste itemrestcorrelaties ($< .30$) had de vragenlijst een Cronbach's Alfa van $.72$.

Vragenlijst ervaren competentie. De mate van ervaren competentie is gemeten met een vertaalde versie van de subschaal van de IMI (Ryan, 1982). De subschaal heeft betrekking op de ervaringen van de student tijdens het maken van de opdracht dus kon alleen tijdens de nameting worden ingevuld.

Vragenlijst taakbeleving. Om het effect van taakbeleving te meten is een pre-taak en een post-taak vragenlijst ontworpen op basis van de gevalideerde Online Motivation Questionnaire van Boekaerts (2002) beschreven in Van Dijk (2013). Taakbeleving bestaat zoals eerder genoemd uit gevoelens over het nut, de haalbaarheid, moeilijkheid en aantrekkelijkheid van de opdracht (Boekaerts, 2002). Deze gevoelens zijn vertaald in de subschalen taakimportantie, succesverwachting, taakangst, taakaanrekkelijk en taakinzet. Alle schalen behalve taakimportantie hebben minstens een Cronbach's Alfa van $.60$, zie Tabel 1. Dit komt overeen met de Cronbach's Alfa's gevonden door Van Dijk (2013). Vanwege de lage Cronbach's Alfa is besloten de subschaal taakimportantie niet mee te nemen in verdere analyses.

Pilot vragenlijsten. In een pilot met twaalf studenten is onderzocht of de geconstrueerde vragenlijsten begrijpelijk en niet te lang waren. Naar aanleiding van de pilot is besloten om in de vragenlijst over autonomie te kiezen voor het toevoegen van optiekeuzes in plaats van actiekeuzes. De studenten vonden dat het hebben van actiekeuzes niet of nauwelijks voorkomt op school. Besloten is om de vragen van de verschillende constructen niet door elkaar te stellen omdat de studenten dat verwarrend.

Vragen studenten. Aan alle studenten zijn bij de vragenlijst aanvullende vragen gesteld om te achterhalen in hoeverre de beoordeling als bruikbaar werd ervaren. De antwoorden zijn kwalitatief

verwerkt. De vragen waren: (1) *‘Wat vond je van deze manier van beoordelen?’* (2) *‘Zou de manier van beoordelen zoals die nu was altijd zo moeten zijn of heb je liever een andere manier? Zo ja, welke manier?’* En (3) *‘Was je motivatie voor de opdracht hoger of lager omdat je wist waarop je werd beoordeeld? Waarom?’*.

Interview docent. Het interview met de docent heeft semigestructureerd plaatsgevonden. In het interview zijn onder andere de ervaringen met de beoordelingsmethoden besproken. Doel was om te achterhalen of de docent de onderzochte beoordelingsmethoden als bruikbaar heeft ervaren. De opgestelde interviewvragen zijn te vinden in bijlage E.

Procedure

Het onderzoek heeft plaatsgevonden tijdens het project ‘De Buurt’ bij het vak Burgerschap. De studenten moesten in groepjes van vijf een digitale flyer maken over hun fantasiestad. Tijdens de voormeting hebben de studenten een vijf minuten durende digitale vragenlijst ingevuld over taakbeleving. De vragenlijst werd afgenomen nadat de docent globaal had verteld wat de opdracht inhield. Aan alle studenten is duidelijk verteld dat de vragenlijst individueel moest worden invullen. De volgende les een week later, hebben de drie klassen uitleg gehad van de onderzoeker over de manier van beoordelen bij de opdracht, zie bijlage F. Nadat de studenten twee keer een uur in de les en thuis konden werken aan de opdracht hebben ze tijdens de les, mondelinge feedback gehad van de onderzoeker op hun vorderingen bij de groepsopdracht. De feedback was expliciet gerelateerd aan de beoordelingscriteria geldend in die specifieke klas. Deze manier van feedback geven was in alle klassen hetzelfde en gebaseerd op de taak- en procesgerichte feedback van het model van Hattie en Timperley (2007). De feedbackprocedure is beschreven in bijlage G. Een week later leverden de studenten de digitale flyer in. Ter afsluiting vulden alle studenten de digitale nameting vragenlijst in van ongeveer 10 minuten. De studenten hadden op dat moment nog geen beoordeling in de vorm van een cijfer gehad. Ten slotte is een week na afsluiting van het project het interview met de docent afgenomen.

Analyses

Kwantitatieve analyses. Om de vragen over in hoeverre de verschillende beoordelingsvormen zorgen voor een verschil in de bruikbaarheid van feedback en de ervaren autonomie en competentie te kunnen beantwoorden zijn drie ANOVA's uitgevoerd. Om te onderzoeken in hoeverre de verschillende beoordelingsvormen een positieve invloed hebben op de subschalen van taakbeleving is een mixed design ANOVA uitgevoerd. De samenhang tussen de uitkomstvariabelen is onderzocht met correlatie analyses.

Kwalitatieve analyses. In hoeverre de verschillende beoordelingsvormen invloed hebben op de docent- en studentervaringen over de bruikbaarheid van de beoordeling is geanalyseerd met de gestelde vragen aan de studenten en een interview met de docent. De antwoorden van de studenten zijn geanalyseerd met de grounded theory benadering (Strauss & Glaser, 1967). De open codering is gedaan door alle antwoorden te lezen en aan alle relevante fragmenten codes te verbinden in de vorm van steekwoorden die representatief waren voor het tekstfragment (Corbin & Strauss, 1990). Vervolgens is er axiaal gecodeerd: de grote hoeveelheid codes geformuleerd bij het open coderen werden gereduceerd en met elkaar geïntegreerd tot een lijst categorieën. Ten slotte is er selectief gecodeerd: de verschillende categorieën zijn met elkaar verbonden wat uiteindelijk heeft geleid tot een codeboom. In de codeboom is er één kerncategorie die theoretische verbanden heeft met de andere categorieën (Corbin & Strauss, 1990). Het proces van coderen is door twee personen gedaan om onafhankelijkheid en intersubjectiviteit te garanderen. Als er geen eenduidigheid was over een code tussen de onderzoekers is net zolang gediscussieerd tot overeenstemming werd bereikt over de desbetreffende code. Het interview met de docent is opgenomen en vervolgens getranscribeerd. De inhoud van het interview is per interviewvraag in samenvattende vorm weergegeven en ondersteund met citaten.

Resultaten

Beschrijvende Statistieken

Allereerst zijn de gemiddelde scores van alle variabelen uitgerekend en weergegeven in Tabel 2 en 3. De correlatiematrix voor de vier uitkomstvariabelen is te vinden in tabel 4.

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

Tabel 2

Beschrijvende Statistieken van de Variabelen Bruikbaarheid van Feedback, Autonomie en Competentie per Conditie

Afhankelijke variabele	Conditie (n)	M	SD	Missing
Bruikbaarheid van feedback	Betrokkenheid (21)	3.42	0.93	4
	Rubrics (26)	3.24	0.75	0
	Controle (20)	2.99	0.81	1
Autonomie	Betrokkenheid (22)	3.76	0.57	3
	Rubrics (25)	3.65	0.39	1
	Controle (21)	3.50	0.63	0
Competentie	Betrokkenheid (22)	3.83	0.74	3
	Rubrics (24)	3.62	0.40	2
	Controle (21)	3.98	0.62	0

Tabel 3

Beschrijvende Statistieken van de Subscales van Taakbeleving op de Voor- en Nameting per Conditie

Variabele		Voor		Na	
		M	SD	M	SD
Taakaantr.	Betrokkenheid	3.52	0.87	3.49	0.72
	Rubrics	3.02	0.56	3.31	0.66
	Controle	3.59	0.42	3.29	0.59
Taakinzet	Betrokkenheid	3.78	0.75	3.53	0.93
	Rubrics	3.37	0.82	3.59	0.62
	Controle	3.59	0.68	3.75	0.70
Taakangst	Betrokkenheid	3.96	0.88	4.15	0.59
	Rubrics	4.42	0.65	4.29	0.54
	Controle	4.32	0.61	4.33	0.64
Succes vw.	Betrokkenheid	3.58	0.61	3.88	0.76
	Rubrics	3.76	0.54	3.98	0.43
	Controle	4.06	0.61	4.06	0.92

Noot. Betrokkenheid; $n = 20$. Rubrics; $n = 25$. Controle; $n = 17$. De score voor taakangst is omgepoold (5=1).

Tabel 4

Correlatiematrix voor de Vier Uitkomstvariabelen

Variabelen	1	2	3
1. Bruikbaarheid feedback	---		
2. Autonomie	.38**	---	
3. Competentie	.13	.40**	---
4. Taakbeleving	.29*	.53**	.65**

Noot. * = $p < .05$. ** $p < .01$.

Uit de correlatiematrix blijkt dat alle variabelen met elkaar samenhangen behalve competentie en de bruikbaarheid van feedback. De verbanden onder de .3 zijn zwak. De verbanden tussen de .3 en .5 zijn te typeren als redelijk. De verbanden boven de .5 als sterk (Cohen, 1988).

Verschillen tussen de Conditie op de Bruikbaarheid van Feedback

Om te onderzoeken in hoeverre de verschillende beoordelingsvormen zorgen voor een verschil in de studentperceptie over de bruikbaarheid van feedback is een ANOVA uitgevoerd. Volgens de normaliteitstoets van Shapiro Wilk ($p < .05$) werd niet voldaan aan de assumptie voor een normaalverdeling. Vier outliers die de normaliteit hebben beïnvloed zijn gecorrigeerd naar het gemiddelde door deze te vermeerderen of verminderen met twee standaarddeviaties ($M \pm 2SD$). Na het corrigeren van de vier outliers werd wel voldaan aan een normaalverdeling. Uit de ANOVA kwam geen significant verschil tussen de groepen. $F(2, 64) = 1.49$, $p = .23$. De manier van beoordelen zorgt dus niet voor een verschil in studentperceptie over de bruikbaarheid van feedback.

Verschillen tussen de Conditie op de Ervaren Autonomie en Competentie

Om te onderzoeken in hoeverre de manier van beoordelen zorgt voor een verschil in de mate van ervaren autonomie en ervaren competentie zijn twee losse ANOVA's uitgevoerd. Bij de ANOVA voor ervaren competentie werd niet voldaan aan de assumptie voor homogeniteit van varianties ($F = 4.52$, $p = .02$). De ANOVA leidde niet tot significante resultaten $F(2, 64) = 2.03$, $p = .14$, $\eta^2 = .06$. Om te corrigeren voor het niet voldoen aan homogeniteit van varianties is ook getoetst met Welch's F en Brown-Forsythe F maar dat leidde niet tot andere conclusies. Ook de ANOVA voor ervaren autonomie bleek niet significant $F(2, 65) = 1.24$, $p = .30$, $\eta^2 = .04$. De manier van beoordelen zorgt dus niet voor een verschil in de mate van ervaren autonomie en competentie.

Verschillen tussen de Conditie op Taakbeleving

Op de voormeting bleek een significant verschil te zijn tussen de condities op de schalen taakaan-trekkelijkheid en succesverwachting. De ANOVA van taakaan-trekkelijkheid gaf de volgende resultaten, $F(2,59) = 5.09$, $p = .01$, $\eta^2 = .15$. Post hoc analyse met Gabriel (significantieniveau van $p = .05$) liet zien dat de rubric-conditie significant lager scoorde op taakaan-trekkelijkheid dan de betrokkenheid- en controleconditie. Effectgroottes waren respectievelijk $d = 0.37$ en $d = 0.33$. Gekozen is voor Gabriel's post hoc omdat de drie condities geen gelijke groepsgrootte hadden (Field, 2009).

Het resultaat van de ANOVA voor succesverwachting was $F(2,59) = 3.26$, $p = .045$, $\eta^2 = .10$. Een post hoc analyse liet zien dat de controleconditie significant hoger scoorde op succesverwachting dan de betrokkenheidconditie. De effectgrootte was $d = 0.33$. De rubric-conditie verschilde niet significant van de andere condities. Bij het interpreteren van de resultaten moet rekening worden gehouden met deze significante verschillen.

Een mixed design ANOVA is uitgevoerd om te onderzoeken in hoeverre de verschillende beoordelingsvormen een positieve invloed hebben op de subschalen van taakbeleving. Alle subschalen voldoen aan de assumpties passend bij een mixed design ANOVA. De resultaten zijn weergegeven in Tabel 5.

Tabel 5

Hoofdeffecten en Interactie-effecten van de Vier Subschalen van Taakbeleving

Variabele	Hoofdeffect tijd (pre-en posttest)			Hoofdeffect conditie			Interactie-effect		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>P</i>	η^2	<i>F</i>	<i>p</i>	η^2
Taakaantr.	0.02	.90	.00	2.60	.08	.08	3.04	.06	.09
Taakinzet	0.15	.70	.00	0.63	.54	.02	1.71	.19	.06
Taakangst	0.06	.82	.00	1.89	.16	.06	1.13	.33	.04
Succesvw.	2.90	.09	.05	1.94	.15	.06	0.73	.49	.02

Noot. F = toetsingsgrootte, p = significantieniveau, η^2 = effectgrootte.

Voor geen één subschaal was sprake van een hoofdeffect van tijd, hoofdeffect van conditie of een interactie-effect. Geen hoofdeffect van tijd wil zeggen dat de score van de subschalen op de nameting niet significant verschilde van de score op de voormeting. Geen hoofdeffect van conditie duidt op geen significant verschil in score tussen de drie verschillende beoordelingsmethoden. Het ontbreken van een interactie-effect wil zeggen dat geen enkele manier van beoordelen heeft gezorgd voor een grotere groei tussen de voormeting en nameting ten opzichte van een andere manier van beoordelen.

Resultaten Interview Docent Burgerschap

Bruikbaarheid van de beoordelingsmethoden. De betrokkenheidconditie werd door de docent niet als bruikbaar ervaren. De docent gaf aan dat het zelf bedenken van beoordelingscriteria te moeilijk en abstract was voor mbo studenten. *‘Hun eigen beoordeling maken, daar hebben studenten wat moeite mee, ik zie het niet als meerwaarde voor de opdracht’.*

De rubric ziet de docent wel als bruikbare beoordelingsmethode. *‘Deze formulieren (wijst op rubric) zijn wel heel goed om toe te passen’.* Volgens de docent bieden rubrics de studenten structuur en uitdaging. De docent gaf aan dat zijn eigen beoordelingscriteria te abstract waren, hij vroeg bijvoorbeeld naar een presentatie met een duidelijk verhaal, maar wat is de definitie van duidelijk? *‘Het formulier wat ik handhaaf is te vrijblijvend’.*

Vershil in kwaliteit en motivatie als gevolg van de interventie. Op de vraag of de manier van beoordelen invloed had op het niveau van de klassen antwoordde de docent ontkennend. *‘Er zat geen verschil in’.* Later iets nuancerend: *‘Ik vind het moeilijk te beoordelen, want de b klas (rubric) was wel heel goed maar is eigenlijk ook wel een voorbeeldklas, de a klas (controle) is een beetje een onrustige klas en de r1a klas (betrokkenheid) is een beetje een zeurklas’.* Een duidelijk verschil tussen de klassen in kwaliteit en motivatie zag de docent niet.

Leerervaringen van het experiment. De docent gaf aan dat het experiment hem bewust heeft gemaakt van het feit dat bij de ontwikkeling van een opdracht ook nagedacht moet worden over de beoordelingsmethode. *‘Het is voor mij leren, want ik deed het altijd op een eenvoudige manier. Ik heb me er nooit zo in verdiept. Heb ik in het begin wel gedaan maar het kostte me veel tijd en dus vergeet*

ik die hele cesuur. ' Hij vond het fijn dat hij bij experiment geholpen werd met de beoordeling. *'Ik had altijd gezegd, als ik nou iemand naast me had die me hielp met beoordelen, en deze rubrics bewijzen dat, ik zie de formuliertjes van je en denk dit wil ik ook.*

Resultaten Vragen Studenten

De antwoorden op de vragen zijn gecodeerd en gereduceerd tot drie hoofdcategorieën: de mening, motivatie en perceptie van studenten over de beoordeling. De antwoorden bij de categorie perceptie zijn niet meegenomen in de resultaten omdat deze categorie diende als interventiecheck. Zie bijlage V voor de bijbehorende codeboom. In totaal hebben 68 studenten de vragen ingevuld. 25 in elke experimentele conditie en achttien in de controleconditie.

De mening van studenten over de acceptatie en bruikbaarheid van de beoordeling. Deze categorie is te verdelen in positieve en negatieve meningen over de manier van beoordelen. Bij de positieve meningen lieten studenten antwoorden noteren zoals prettig, eerlijk, duidelijk, goed, en origineel. Tussen de klassen zat verschil. In beide experimentele klassen gaven minstens de helft van de studenten een positieve mening over de manier van beoordelen. In de controleconditie was dat maar 1/3^{de} van de respondenten. Voorbeelden van positieve uitingen van de experimentele condities waren: *'Ik vond dat het fijn was dat je zelf mocht kiezen wat belangrijker was'* (betrokkenheid). *'Ik vond de beoordeling erg duidelijk want alles wordt beoordeeld zo kun je goed bekijken wat voor cijfer je zou moeten krijgen.'* (rubric)

Naast positieve meningen over de bruikbaarheid van de beoordelingsmethode waren er ook positieve uitingen over de acceptatie van de beoordelingsmethode. In de Rubric-conditie gaven dertien van de 25 studenten aan dat deze manier van beoordelen altijd zo zou moeten zijn. Één respondent zei bijvoorbeeld *'Ja, hij zou altijd zo moeten zijn. Zo weet je of je de opdracht goed doet en of er nog iets aan veranderd moet worden.'* In de betrokkenheidconditie gaven acht van de 25 studenten aan deze methode vaker te willen. Daarnaast gaven nog zeven studenten uit dezelfde conditie aan dat het een goede methode is maar dat het afhankelijk is van de opdracht of de methode geschikt is. In de controleconditie gaven vijf van de achttien studenten aan deze manier van beoordelen te willen houden. Naast positieve meningen over de acceptatie van de beoordelingsmethode waren er ook negatieve uitingen. Vier studenten uit de rubrics conditie, zes uit betrokkenheidconditie en zeven uit

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

de controleconditie hebben liever een andere manier van beoordelen. De studenten die aangeven een andere beoordelingsmethode te willen, geven aan dat de feedback bij de beoordeling nog inhoudelijker en meer mag zijn. In alle condities werd minimaal twee keer een negatieve mening uitgesproken over de verkregen feedback. In totaal waren er vijf negatieve meningen in de rubric-conditie en vier negatieve meningen in de andere condities over de bruikbaarheid van de beoordelingsmethode. Enkele voorbeelden van negatieve meningen waren:

'Ik geloof niet dat ik van de criteria invullen iets heb geleerd. Ik zou liever gewoon hebben dat de leraar de criteria maakte.' (betrokkenheid)

'Ik mis iets meer diepgang en verbeterpunten, ofwel opbouwende kritiek.' (rubric)

'Er mag wel wat meer feedback gegeven worden bij de beoordeling.' (controle).

De motivatie van de studenten bij de beoordeling. De antwoorden met betrekking tot motivatie konden ingedeeld worden in hoger, lager of geen verschil in motivatie veroorzaakt door de beoordelingsmethode of een andere oorzaak. Zo gaven in elke conditie minimaal twee studenten aan dat hun motivatie hoger was door de leuke opdracht. De antwoorden over motivatie die gegeven zijn in relatie tot de beoordelingsmethode zijn gekwantificeerd in Tabel 6.

Tabel 6

Motivatie van Respondenten Gerelateerd aan de Beoordelingsmethode

Conditie	Lagere motivatie n (%)	Motivatie gelijk gebleven	
		n (%)	
Betrokkenheid klas	2 (8%)	10 (40%)	7 (28%)
Rubric klas	0 (0%)	12 (46%)	8 (31%)
Controle klas	1 (5%)	5 (24%)	2 (10%)

Noot. Rubrics; n=25. betrokkenheid; n=25. Controle; n=18

Voorbeelden van antwoorden waaruit een hogere motivatie blijkt zijn:

'Hoger, omdat ik beter mijn kwaliteiten kon gebruiken en wist waar ik aan moest werken'.

(betrokkenheid)

'Hoger, omdat ik precies wist wat er van me verwacht werd.' (rubric)

Een lagere motivatie werd door een student in de betrokkenheidconditie gerelateerd aan het gegeven dat hij achteraf niet zo blij was met de opgestelde beoordelingscriteria. De studenten die geen verschil in motivatie hadden letten niet op de beoordelingscriteria. Ze maakten de opdracht voor zichzelf en

niet voor een beoordeling. De hogere motivatie werd volgens de studenten in beide experimentele condities vooral veroorzaakt doordat ze wisten waarop ze werden beoordeeld. De beoordelingscriteria gaven richting aan de opdracht.

Discussie

In dit onderzoek is getracht antwoord te krijgen op de vraag: *In hoeverre heeft de manier van beoordelen invloed op studentperceptie over de bruikbaarheid van feedback, de mate van ervaren autonomie en competentie, de taakbeleving en invloed op de docent- en studentervaringen over de bruikbaarheid van de beoordeling?*

Op basis van de kwantitatieve resultaten kan gesteld worden dat de manier van beoordelen niet leidt tot een hogere taakbeleving, ervaren autonomie en ervaren competentie. Ook de bruikbaarheid van feedback werd niet beter gewaardeerd als gevolg van een specifieke manier van beoordelen. Op basis van de kwalitatieve data kan wel gesteld worden dat de verschillende beoordelingsvormen invloed hebben gehad op de docent- en studentervaringen over de bruikbaarheid van de beoordeling.

Studentperceptie over de Bruikbaarheid van Feedback

In tegenstelling tot de verwachting hebben de verschillende beoordelingsvormen niet gezorgd voor een verschil in de studentperceptie over de bruikbaarheid van feedback. De gemiddelden suggereren dat de feedback als bruikbaar werd ervaren. Dat er geen verschil was tussen de condities kan verklaard worden door de manier waarop feedback is gegeven. Weaver (2006) liet in onderzoek zien dat feedback door studenten als onbruikbaar wordt ervaren wanneer de feedback te algemeen is, aan sturing ontbreekt, een te negatieve focus heeft of niet gerelateerd is aan de beoordelingscriteria. Het enige verschil tussen de groepen was dat de feedback in de controleconditie minder gerelateerd was aan de beoordelingscriteria dan in de andere groepen. Wellicht is door het wegnemen van de andere onduidelijkheden door het toepassen van taakgerichte en procesgerichte feedback geen verschil ontstaan.

Ervaren Autonomie

De verschillende beoordelingsvormen zorgen in tegenstelling tot de verwachting niet voor een verschil in de mate van ervaren autonomie. In dit onderzoek is autonomie geboden in de vorm van optiekeuzes. Uit onderzoek blijkt echter dat vooral het aanbieden van actiekeuzes invloed heeft op het

gevoel van ervaren autonomie (Reeve et al., 2003). Actiekeuzes hebben de grootste invloed doordat de persoon een keuze heeft over of en hoe de activiteit wordt uitgevoerd. Bij optiekeuzes is de keuzevrijheid gelimiteerd doordat de keuze bestaat uit een aantal opgelegde opties waardoor het gevoel van autonomie minder groot is dan bij actiekeuzes (Reeve et al., 2003).

Daarnaast waren de geboden keuzes misschien te fragmentarisch om daadwerkelijk invloed te hebben op de mate van ervaren autonomie. Het is namelijk van belang dat het maken van keuzes plaatsvindt in een autonomie-ondersteunende omgeving (Reeve et al., 2003). In een autonomieondersteunende omgeving wordt ook niet-directief taalgebruik toegepast. De relevantie van de opdracht wordt aangetoond en eigen initiatief wordt aangemoedigd (Deci, Ryan, & Williams, 1996). Ten slotte was in de opdracht al sprake van relatief veel autonomie waardoor weinig ruimte was voor verbetering in autonomie door de beoordelingsmethode. De groepskeuze stond vrij, het te kiezen onderwerp had weinig restricties en ook in de invulling van de verplichte onderdelen hadden de studenten relatief veel vrijheid.

Ervaren Competentie

De verschillende beoordelingsvormen zorgen in tegenstelling tot de verwachting niet voor een verschil in ervaren competentie. In dit onderzoek is getracht te voldoen aan de mate van ervaren competentie door structuur te bieden en constructieve feedback te geven. Een verklaring voor het ontbreken van een hogere score in de rubric-conditie kan gezocht worden in het gegeven dat competentie ook gerelateerd is aan self-efficacy. Competentie is de bekwaamheid die men voelt om een gewenst resultaat te kunnen halen (Deci & Ryan, 2000). Self-efficacy is het vertrouwen in de eigen bekwaamheid om een activiteit met succes te volbrengen (Bandura, 1977). De persoonlijke overtuigingen over de benodigde competenties voor het uitvoeren van een taak worden beïnvloed door eerdere (succes)ervaringen, voorbeeldervaringen van andere personen, de verbale uitlatingen van anderen en de emotionele en fysieke spanning (Bandura, 1977). In dit onderzoek was de manier van beoordelen in beide experimentele condities nieuw wat misschien heeft gezorgd voor spanning en het ontbreken van eerdere (succes)ervaringen.

Taakbeleving

De studenten rapporteerden geen hogere score op taakbeleving door de verschillende beoordelingsvormen. Door de explicitering van de beoordelingscriteria in beide experimentele condities werd de taak anders gepresenteerd dan normaal. De verwachting was dat de studenten beter wisten wat er van ze werd verwacht, waardoor de taakangst minder zou worden en de taakinzet en succesverwachting hoger (Boekaerts, 1992; Hattie, 2009). Echter is de invloed van de manier van beoordelen niet groot genoeg geweest om een verschil in taakbeleving tussen de verschillende groepen te creëren. De taakbeleving wordt niet alleen bepaald door de manier waarop de taak wordt aangeboden maar ook door de taakkenmerken, de generieke motivatie en de stabiele eigenschappen van een persoon (Boekaerts, 1992). Deze variabelen hebben mogelijk een grotere invloed op de taakbeleving van studenten dan de manier van beoordelen.

Een kanttekening is dat er op de voormeting een verschil te zien was tussen de drie condities op de subschalen succesverwachting en taakaanrekkelijkheid. Op de nameting was dit verschil niet meer aanwezig. Een mogelijke verklaring voor de inhaalslag in de rubric-conditie bij taakaanrekkelijkheid is dat de verduidelijking van de beoordelingscriteria in de rubric ervoor heeft gezorgd dat het algemene oordeel over de opdracht positiever is uitgevallen dan bij de controleconditie. Bij succesverwachting heeft de betrokkenheidsconditie een inhaalslag gemaakt ten opzichte van de andere twee condities maar de succesverwachting blijft in deze conditie het laagst.

Ervaring Docent en Studenten over de Manier van Beoordelen

In termen van bruikbaarheid en acceptatie zijn de verschillende beoordelingsvormen anders ervaren door de studenten en de docent. De docent ziet door de geboden structuur de rubric als een bruikbare beoordelingsmethode om in de toekomst verder in te investeren. Het zelf opstellen van beoordelingscriteria ziet de docent niet als bruikbare methode. Zijn eigen gehanteerde beoordelingsmethode vindt hij te vaag en abstract om bruikbaar te zijn. Het experiment heeft bewustwording gecreëerd bij de docent over het belang van het opstellen van een duidelijke beoordelingsmethode.

Minstens de helft van de studenten in de experimentele condities heeft een positieve mening over de bruikbaarheid van de beoordeling en zou daarnaast de ervaren beoordelingsmethode vaker toegepast willen zien. Geconcludeerd kan worden dat ten opzichte van de controleconditie de

beoordelingsmethoden als meer bruikbaar en acceptabel worden ervaren. In de betrokkenheidconditie werd wel de kanttekening gemaakt dat de beoordelingsmethode goed is voor de afwisseling maar wel afhankelijk van de opdracht.

Beperkingen Onderzoek

Ondanks dat vooraf de vergelijkbaarheid van de klassen is geïnventariseerd, bleken ze toch te verschillen in initiële taakbeleving. Wellicht waren de klassen op voorhand toch niet zo vergelijkbaar als gedacht waardoor verschillend is gescoord op de voormeting en waardoor het effect van de interventie is vertroebeld. De mogelijke verschillen tussen de klassen hebben de vergelijking tussen de klassen bemoeilijkt doordat niet onderzochte variabelen wellicht ook met taakbeleving samenhangen. De vergelijkbaarheid vooraf is onder andere gebaseerd op het subjectieve oordeel van de docent over de leerprestaties in de klas. De vergelijkbaarheid bleek toch moeilijker te voorspellen dan gedacht. In het vervolg kunnen meer objectieve gegevens worden meegenomen zoals cijfers voor het vak en intelligentie om mogelijke aanvankelijke verschillen te corrigeren, waardoor het effect van de interventie zuiverder wordt.

Niet alle vragenlijsten van de nameting zijn op hetzelfde moment afgenomen. Door een studiereis hebben de studenten in de controleconditie de vragenlijst pas 1,5 week later ingevuld dan de andere twee condities. De volgende opdracht voor het vak was alweer begonnen waardoor de opdrachterevaringen wellicht meer naar de achtergrond zijn gezakt. Met als gevolg dat de gemeten taakbeleving in de controleconditie mogelijk niet alleen ging over de onderzochte opdracht en daardoor moeilijker te vergelijken is met de andere condities.

De uitval van 10 participanten werd veroorzaakt door ziekte of andere omstandigheden waardoor studenten niet op school aanwezig waren en maakte dat de totale steekproef slechts 62 studenten betrof. De kleine steekproef heeft negatieve gevolgen voor de statistische power. De kans op type II fouten wordt groter en de generaliseerbaarheid van de resultaten wordt beperkt (Rogelberg & Stanton, 2007).

Mogelijkheden voor Vervolgonderzoek

In de kwantitatieve data is geen verschil gevonden tussen de verschillende beoordelingsmethoden. In de kwalitatieve data blijken in termen van acceptatie en bruikbaarheid wel

verschillen te bestaan. De divergentie tussen de kwalitatieve en kwantitatieve data kan veroorzaakt zijn door de vraagstelling. In de kwantitatieve data was de link tussen de beoordelingsmethode en de items in de te meten constructen mogelijk te impliciet. Terwijl in de kwalitatieve data wel expliciet werd gevraagd naar de effecten van de beoordelingsmethode. In vervolgonderzoek kan meer congruentie in de vraagstelling worden toegepast waardoor de kans op een eenduidiger beeld tussen de twee typen data groter wordt.

In de kwantitatieve data werd vooral gekeken naar de effecten van de beoordelingsmethoden op de ervaringen en gevoelens bij de opdracht. Het is interessant om in vervolgonderzoek de prestaties van studenten mee te nemen. Onderzocht kan worden of bepaalde beoordelingsmethoden tot hogere cijfers leiden dan andere beoordelingsmethoden. Daarnaast is de nameting afgenomen voordat de studenten hun cijfer kregen voor de opdracht. In vervolgonderzoek is het interessant om te kijken hoe de manier van beoordelen wordt ervaren als het cijfer bijvoorbeeld lager is uitgevallen dan verwacht. De vraag is of het cijfer van invloed is op de waardering van de beoordelingsmethode.

Praktische Implicaties

Ondanks dat er geen verschillen bleken te zijn tussen de condities in taakbeleving, autonomie, competentie en bruikbaarheid van de feedback is wel duidelijk dat de docent de rubric een bruikbare beoordelingsmethode vindt en dat studenten beide nieuwe beoordelingsmanieren accepteren en waarderen. In elke conditie gaven minimaal twee studenten aan dat de feedback nog persoonlijker, directer of meer mag zijn. De perceptie over goede feedback kan verschillen per student waardoor docenten in het vervolg de dialoog aan kunnen gaan over de gegeven feedback (Sol & Stokking, 2009). Door de dialoog aan te gaan over de bruikbaarheid van de feedback, kan beter worden ingespeeld op de individuele feedbackbehoeften van de studenten. Met als gevolg dat docenten nog meer inhoudelijke en directe feedback kunnen geven die aansluit bij de feedbackwensen van de studenten.

Het belang van een transparante beoordelingsmethode is uit de literatuur duidelijk te herleiden maar in de praktijk niet altijd aan de orde (Broekkamp & van Hout-Wolters, 2007; Slujsmans et al., 2008). Dit onderzoek laat zien dat experimenteren met beoordelingsmethoden kan leiden tot bewustwording bij docenten over het belang van een duidelijke en transparante beoordelingsmethode.

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

De docent is in gaan zien dat duidelijke en transparante criteria moeten worden opgesteld om het gewenste leerresultaat uit studenten te halen. Dit sluit ook aan bij de antwoorden van de studenten. De studenten die aangeven een verhoogde motivatie te hebben door de beoordelingcriteria geven aan dat de hogere motivatie vooral werd veroorzaakt doordat ze wisten waarop ze werden beoordeeld. De beoordeling in de experimentele condities gaf meer duidelijkheid ten opzichte van de beoordeling in de controleconditie. De docent gaat de komende periode tijd investeren in het opzetten van een betere beoordelingsmethode in samenspraak met de studenten. Transparantie en duidelijkheid gaan daar in ieder geval een grote rol in krijgen.

Referenties

- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57, 13-18.
- Baartman, L. K. J. (2008). *Assessing the assessment: Development and use of quality criteria for Competence Assessment Programmes* (proefschrift). Utrecht: Universiteit Utrecht.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84, 191.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology: An International Review*, 41(4), 377-397.
- Boekaerts, M., & Simons, P. R. J. (1995). *Leren en Instructie: Psychologie van de leerling en het leerproces*. Assen: Van Gorcum.
- Boekaerts, M. (2002). The on-line motivation questionnaire: A self-report instrument to assess students' context sensitivity. In: P. R. Pintrich & M. L. Maehr (Eds.) *Advances in Motivation and Achievement, Volume 12: New Directions in Measures and Methods*. (pp. 77-120). New York: JAI / Elsevier Science.
- Broekkamp, H., Van Hout-Wolters, B. H. A. M., Van den Bergh, H., & Rijlaarsdam, G. (2004). Students' expectations about the processing demands of teacher-made tests. *Studies in Educational Evaluation*, 30, 281-304.
- Broekkamp, H., & Van Hout-Wolters, B. H. A. M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, 19, 401- 428.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31, 219-233.
- Castelijns, J., & Andersen, I. (2013). *Beoordelen om te leren: Leerlingen als mede-beoordelaars van hun eigen leerproces*. 's-Hertogenbosch: KPC Groep.
- Cluitmans, J. & Klarus, R. (2005). Competentiebeoordeling: Een pleidooi voor congruentie. *Tijdschrift voor Hoger Onderwijs*, 23, 221-238.

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13, 3-21.
- Deci, E. L., & Ryan R. M. (2008). Facilitating optimal motivation and psychological well-being across life's domains. *Canadian Psychology*, 49, 14-23.
- Deci, E. L., Ryan, R. M., & Williams, C. (1996). Need satisfaction and the self-regulation of Learning. *Learning and Individual Differences*, 8, 165-183.
- Elshout-Mohr, M., Oostdam, R. & Overmaat, M. (2002). Student assessment within the context of constructivist educational settings. *Studies in Educational Evaluation*, 28, 369-390.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage publications.
- Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Hinett, K., & Weeden, P. (2000). How am I doing?: Developing critical self-evaluation in trainee teachers. *Quality in Higher Education*, 6, 245 - 257.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review*, 2, 130-144.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 19, 254-284.
- Lea, M., & Street, B. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23, 157-172.
- Merry, S., Orsmond, P., & Reiling, K. (2000). Biological essays: how do students use feedback. In: C. Rust (Ed.) *Improving student learning through the disciplines*. Oxford, Engeland: OCSLD.

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.
- Nicol, D. (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education*, 34, 335-352.
- Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35, 501-517.
- Orsmond, P, Merry, S and Reiling, K (2002) The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 27, 309–323.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315-341.
- Pokorny, H., & Pickford, P. (2010). Complexity, cues and relationships: Student perceptions of feedback. *Active Learning in Higher Education*, 11, 21-30.
- Price, M., & O'Donovan, B. (2006). Improving performance through enhancing student understanding of criteria and feedback. In: C. Bryan & K. Clegg (Eds.) *Innovative assessment in higher education* (pp. 100-109). Londen, Engeland: Routledge.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35, 435-448.
- Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology*, 95, 375-392.
- Reeve, J., Deci, E. L., & Ryan, R. M. (2004). Self-determination theory: A dialectical framework for understanding socio-cultural influences on student motivation. In: D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (pp. 31-60). Greenwich, CT: Information Age Publishing.

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

- Rochford, L., & Borchert, P. S. (2011). Assessing higher level learning: Developing rubrics for case analysis. *Journal of Education for Business*, 86, 258-265.
- Rogelberg, S. G., & Stanton, J. M. (2007). Introduction understanding and dealing with organizational survey nonresponse. *Organizational Research Methods*, 10(2), 195-209.
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28, 147-64.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Seegers, G., & Boekaerts, M. (1993). Taskmotivation and mathematics achievement in actual task situations. *Learning and Instruction*, 3, 133-150.
- Slujsmans, D. M. A. (2008). *Betrokken bij beoordelen. Duurzaam beoordelen in vraaggestuurd leren*. Lectorale rede. Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Slujsmans, D. M. A., Straetmans, G., & Van Merriënboer, J. (2008). Integrating authentic assessment with competency based learning: The protocol portfolio scoring. *Journal of Vocational Education and Training*, 60, 159-172.
- Sol, Y. B., & Stokking, K. M. (2009). *Mondelinge feedback bij zelfstandig werken: interactie tussen docenten en studenten in het VO*. Utrecht: IVLOS.
- Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education*, 18, 221-233.
- Van Dijk, J. (2013). *Passend onderwijs voor de (hoog)begaafde leerling: Onderzoek naar excellentiebevordering en het effect van Flip de Klas op leerresultaten en taakbeleving wiskunde op het vwo* (Masterthesis). MEBIT, Universiteit van Maastricht, Maastricht.

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

Van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education, 1*, 41-67.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31*, 379-394.

Wiggins, G. (2012). Seven keys to effective feedback. *Educational Leadership, 70*, 10-16.

Bijlage A Invulformulier Betrokkenheidconditie

<p>Namen van de groepsleden: _____</p> <p>Klas: _____</p>	
<p>Layout digitale flyer (minimaal 2 criteria)</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> <p>4. _____</p>	<p>Inhoud digitale flyer (minimaal 2 criteria)</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> <p>4. _____</p>
<p>Beoordeling (cesuur)</p> <p>Aantal punten te verdienen (totaal 100)</p> <p>Layout digitale flyer</p> <p>Inhoud digitale flyer</p> <p>Verslag industriestadwoord volgens de richtlijnen</p> <p>Presentatie</p> <p>Samenwerking</p> <p>Volgende bij hoeveel punten</p> <p>(bepaal de docent)</p> <p>< 70 punten = onvoldoende</p> <p>70 punten = voldoende</p> <p>90 punten of meer = goed</p>	
<p>Presentatie (minimaal 1 criterium)</p> <p>1. _____</p> <p>2. _____</p> <p>Samenwerking (minimaal 1 criterium)</p> <p>1. _____</p> <p>2. _____</p>	<p>Verslag (minimaal 2 criteria)</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> <p>4. _____</p>

Bijlage B Rubric

		Beginner (1 punt)	Gevorderd (3 punten)	Expert (5 punten)	Score (1, 3 of 5)
1. Lay-out digitale flyer	1.1 <i>Voorpagina</i>	De voorpagina van de digitale flyer bestaat uit een foto en tekst die nauwelijks aansluiten bij het gekozen onderwerp van de fantasieplek.	De voorpagina van de digitale flyer bestaat uit een foto en tekst die aansluiten bij het gekozen onderwerp van de fantasieplek.	De tekst en afbeelding(en) op de voorpagina zijn zo geplaatst dat de voorpagina opvalt, aantrekkelijk is en een goede indruk geeft van de fantasieplek.	
	1.2 <i>Ontwerp van de gehele flyer</i>	De lay-out van de gehele digitale flyer is onverzorgd. De tekst is niet consequent opgemaakt en er is ook weinig gebruik gemaakt van kleur en foto's die aansluiten bij het gekozen onderwerp.	De lay-out is verzorgd. De tekst, foto's en het kleurgebruik passen bij elkaar en het gekozen onderwerp maar bevat weinig originaliteit of creativiteit.	De digitale flyer is zo vormgegeven dat het aansluit bij het gekozen onderwerp. Zowel de tekst, kleurgebruik als de foto's zijn één geheel. De lay-out is daarnaast ook origineel en/of creatief vormgegeven.	
2. Inhoud digitale flyer	2.1 <i>Onderdelen aanwezig</i>	Er zijn minder onderdelen aanwezig dan de verplichte 6. Verplichte onderdelen: <ul style="list-style-type: none"> - Voorpagina - Plattegrond - Winkels en/of bezienswaardigheden - Historische gebouwen - Cultuur - Foto's en illustraties 	Alle 6 de verplichte onderdelen zijn aanwezig maar niet alle onderdelen zijn even volledig uitgewerkt met zowel beeldmateriaal als tekst.	Alle 6 de onderdelen zijn beschreven en allemaal even volledig uitgewerkt met zowel tekst als foto's en/of ander materiaal zoals video of audio.	
	2.2 <i>Originaliteit onderdelen</i>	De onderdelen zijn op een standaard manier uitgewerkt, bijvoorbeeld alleen maar gebruik gemaakt van bestaande afbeeldingen. Er is weinig origineels of creatiefs aan toegevoegd.	Bij enkele onderdelen is iets origineels of creatiefs toegevoegd zoals een gefotoshopte afbeelding, zelf gemaakte foto, filmpje of een audiobestand.	Alle onderdelen zijn origineel en creatief uitgewerkt. Er is gebruik gemaakt van de fantasie en de verschillende mogelijkheden van Smore.	
	2.3 <i>Informatief</i>	De informatie op de flyer is niet volledig en interessant genoeg. De lezer wordt nauwelijks geïnformeerd over wat de fantasieplek te bieden heeft.	De flyer biedt de lezer informatie over wat de fantasieplek te bieden heeft. Alleen deze informatie is niet bij alle onderdelen even interessant en/of volledig	De flyer biedt de lezer bij alle onderdelen interessante en volledige informatie over wat de fantasieplek te bieden heeft.	
3. Verslag (1 A4)	3.1 <i>Verantwoording gekozen fantasieplek</i>	Het verslag bevat nauwelijks argumenten waarom gekozen is voor deze plek en deze manier van wonen/leven. Ook de beschrijving van de	Het verslag bevat een verantwoording waarom gekozen is voor de fantasieplek maar mist een goed onderbouwd verhaal over de	Het verslag bevat een goed gemotiveerde verantwoording waarom gekozen is voor deze plek. Ook het stukje over de participatie in de wijk is	

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

		participatie binnen de fantasieplek ontbreekt.	participatie binnen de fantasieplek.	duidelijk beschreven.	
		Beginner (1 punt)	Gevorderd (3 punten)	Expert (5 punten)	Score (1,3 of 5)
	3.2 <i>Slotwoord</i>	In het slotwoord wordt de gekozen fantasieplek nauwelijks aangeprezen. De lezer en dus ook de uitvoerder worden niet overtuigd om voor deze fantasieplek te kiezen.	In het slotwoord wordt geprobeerd de lezer & de uitvoerder te overtuigen om voor dit plan te kiezen. Maar de echte overtuigingskracht en inhoudelijk sterke argumenten missen.	In het slotwoord wordt de fantasieplek goed aangeprezen. De lezer & de uitvoerder worden met inhoudelijke sterke en creatieve argumenten overtuigd om te kiezen voor deze fantasieplek.	
	3.3 <i>Stijl / spelling</i>	Het verslag leest minder prettig door de vele taal- en spelfouten en de stijl van schrijven past ook niet bij het doel van beargumenteren en overtuigen.	Het verslag heeft een schrijfstijl die aansluit bij het doel van de tekst. Het bevat wel wat taal- en /of spelfoutjes, maar geen grote, storende fouten.	Het verslag heeft een goede schrijfstijl en bevat nauwelijks tot geen taal- en spelfouten	
4.	4.1 Presentatie <i>Overtuigingskracht presentatie</i>	De presentatie gaf geen goed beeld van wat de gekozen fantasieplek allemaal te bieden had. Ook werd de plek niet aangeprezen.	De presentatie gaf een goed beeld van wat de fantasieplek te bieden had maar miste het benodigde verkooppraatje om de plek echt aan te prijzen.	De presentatie gaf een goed beeld van de mogelijkheden van de fantasieplek en daarnaast werd de plek ook echt aangeprezen.	
5.	5.1 Samenwerking <i>Teamwork van de groep.</i>	Nauwelijks samenwerking tussen de leden van de groep. Geen taakverdeling, al het werk werd individueel gedaan en op het eind samengevoegd.	Merkbare samenwerking tussen de leden van de groep; de taken werden verdeeld.	Goede samenwerking tussen de leden van de groep, de taken zijn gelijk verdeeld. Er was merkbaar plezier in het samenwerken.	

Totaal aantal punten:

Cesuur:

0-29 punten is een onvoldoende

30-39 punten is voldoende

40-50 punten is goed

Bijlage C Beoordelingsformulier Controleconditie

Inleveren

- Een **Concept** voor een Feedback in de week van 23 maart.
- **Presentatie** in de week van 30 maart:
 - ✓ Een **Digitale Flyer** gemaakt in het programma **Smore.com** met de verplichte **ZES** onderdelen.
 - ✓ Een **Verslag** waarin duidelijk wordt beschreven waarom jullie voor deze fantasie-plek hebben gekozen.
 - ✓ Verslag niet langer dan 1x A4
 - ✓ Inhoud > verantwoording over de manier van wonen en werken in jouw fantasiestad.
 - ✓ Een korte beschrijving van de participatie in jouw fantasiestad (zorg voor en afspraken met bewoners).
 - ✓ Een slotwoord om jouw woonomgeving aan te prijzen.

Beoordeling

- Complete digitale Flyer in **Smore.com**.
- Alle **ZES** onderdelen aanwezig.
- Duidelijk en een overzichtelijke inhoud.
- Een verslag op A4 met een duidelijk verhaal en slotwoord.

Cesuur

Digitale Flyer in Smore.com	per onderwerp 1 punt	Max 6 punten	Totaal 12 punten = Goed	Totaal 10 – 11 punten = Voldoende	Totaal < 10 punten = Onvoldoende
	duidelijk in teksten beeld	Max 2 punten			
Verslag	het verhaal	Max 3 punten			
	slotwoord	Max 1 punten			

Bijlage D Vragenlijst Voor- en Nameting

Voormeting

Taakaantrekkelijkheid

-
1. Heb je zin om aan de opdracht te beginnen?
 2. Hoe enthousiast ben je over deze opdracht?
 3. Hoe leuk vind je deze opdracht?
 4. Hoe voel je je op dit moment over de opdracht die je te wachten staat?

Taakimportantie

-
1. Hoe belangrijk vind je deze opdracht?
 2. Hoe belangrijk vind je het om deze opdracht goed te doen?

Succesverwachting

-
1. Hoe goed verwacht je deze opdracht te gaan doen?
 2. Welk cijfer verwacht je te halen voor de opdracht?

Taakangst

Hoe voel je je op dit moment over de opdracht die je te wachten staat?

1. *Niet op mijn gemak 1 2 3 4 5 op mijn gemak*
2. *Niet nerveus 1 2 3 4 5 nerveus*
3. *Niet goed 1 2 3 4 5 goed*
4. *Niet bezorgd 1 2 3 4 5 bezorgd*
5. *Zonder zelfvertrouwen 1 2 3 4 5 met zelfvertrouwen*

Taakinzet

-
1. Met hoeveel aandacht denk je deze opdracht te gaan doen?
 2. Hoeveel energie ben je van plan in deze opdracht te gaan stoppen?
 3. Als je net zoveel tijd aan deze opdracht zou kunnen besteden als je zou willen, hoelang zou je er dan aan werken?

Ik begin gewoon en zie later wel hoe lang ik er aan ga werken.

Ik werk eraan totdat ik net een voldoende kan halen.

Ik werk eraan totdat ik geen zin meer heb.

Ik werk eraan totdat ik zeker weet dat ik een voldoende kan halen.

Ik werk er net zo lang aan tot ik echt tevreden over mezelf kan zijn.

Naamting

Taakaantrekkelijkheid

-
1. Hoe voel je je nu de opdracht is afgerond? *Ben het nog niet zat 1 2 3 4 5 ben het zat*
 2. Hoe voel je je nu de opdracht is afgerond? *Niet verveeld 1 2 3 4 5 verveeld*
 3. Ik vond de opdracht leuk om aan te werken
 4. Ik was enthousiast bij het werken aan de opdracht

Taakimportantie

-
1. Hoe belangrijk heb je deze opdracht gevonden?
 2. Hoe belangrijk vond je het om deze opdracht goed te doen?

Succesverwachting

-
1. Hoe voel je je nu de opdracht is afgerond? *Niet tevreden 1 2 3 4 5 tevreden*
 2. Wat voor cijfer verwacht je te kunnen halen voor deze opdracht?

Taakangst

Hoe voel je je nu de opdracht is afgerond?

1. *Niet opgelucht 1 2 3 4 5 opgelucht*
2. *Niet op mijn gemak 1 2 3 4 5 op mijn gemak*
3. *Niet nerveus 1 2 3 4 5 nerveus*
4. *Niet bezorgd 1 2 3 4 5 bezorgd*
5. *Niet zelfverzekerd 1 2 3 4 5 zelfverzekerd*

Taakinzet

-
1. Hoe zorgvuldig heb je de opdracht gedaan? Ik heb er:
 2. Hoe lang heb je aan de opdracht gewerkt?

Ik ben begonnen, maar halverwege ben ik afgehaakt

Ik heb er net zoveel aan gewerkt tot ik dacht een voldoende te kunnen halen

Ik heb er net zolang aan gewerkt totdat ik geen zin meer had

Ik heb er net zoveel aan gewerkt tot ik dacht een redelijk goed cijfer te kunnen halen

Ik heb net zolang gewerkt tot ik helemaal tevreden was over mezelf

3. Hoe goed heb je deze opdracht gedaan?

Open vragen studenten

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

-
1. Heb je gemerkt dat de manier van beoordelen anders was bij deze opdracht dan normaal? Zo ja, wat vond je van deze manier van beoordelen?
 2. Zou de manier van beoordelen zoals die nu was altijd zo moeten zijn of heb je liever een andere manier? Zo ja, welke manier?
 3. In welke mate heb je bij het maken van de opdracht gelet op de beoordelingscriteria van de opdracht?
 4. Was je motivatie voor de opdracht hoger of lager omdat je wist waarop je werd beoordeeld?
Waarom?
-

Ervaren competentie

1. Toen ik een tijdje met de opdracht bezig was, dacht ik: dit gaat me lukken.
 2. Dit was een opdracht waar ik NIET zo goed in was.*
 3. Ik had voldoende vaardigheden om de opdracht goed uit te voeren.
 4. Ik denk dat ik best goed was in deze opdracht.
 5. Ik denk dat ik vergeleken met andere studenten deze opdracht best goed heb gedaan.
 6. Ik ben tevreden met mijn prestaties bij deze opdracht.
-

*= omgekeerde items

Ervaren autonomie

1. Ik geloof dat ik in bepaalde zaken keuzevrijheid had tijdens deze opdracht.
 2. De opdrachtbeschrijving liet ruimte over om zelf te bepalen hoe ik de opdracht ging uitvoeren.
 3. Ik had het gevoel dat ik helemaal GEEN keuzevrijheid had in deze opdracht.*
 4. Ik deed mee aan deze opdracht, ook omdat ik het zelf wilde.
 5. In deze opdracht had ik het gevoel dat NIET alles vast lag, ik mocht zelf keuzes maken.
 6. De manier van beoordelen bij deze opdracht had een positieve invloed op mijn gevoel van keuzevrijheid.
 7. Ik deed mee aan deze opdracht, alleen omdat het moest.*
-

*= omgekeerde items

De bruikbaarheid van feedback

1. De gegeven feedback was duidelijk voor mij.
 2. Door de feedback wist ik wat ik moest doen om de opdracht te verbeteren.
 3. De feedback maakte duidelijk waar de opdracht aan moest voldoen.
 4. De feedback ging over wat goed ging of goed was gedaan.
 5. De feedback was bruikbaar.
 6. Ik heb van de feedback geleerd.
 7. De feedback ging over wat nog NIET goed was of nog NIET goed ging.
 8. De feedback heeft mij geholpen met het vervolg van de opdracht.
-

Bijlage E Interview Docent Burgerschap

Manier van beoordelen

- Klas voor klas bespreken; wat vond je van de manieren van beoordelen in de klassen? Mm1b kreeg structuur, mr1a mocht zelf meebepalen, mm1a zoals het altijd was. Welke manier vond je zelf het fijnst?
- Zie je mogelijkheden voor een andere manier van beoordelen dan de manier die je op dit moment doet?

Kwaliteit van opdrachten en motivatie studenten

- Kwaliteit van de opdrachten, merkte je verschil tussen de klassen? Zo ja is dat verschil te wijten aan de interventie of alleen aan het niveau van de klassen?
- Heb je een idee over in welke klas de studenten de manier van beoordelen het fijnste vonden?
- Was er verschil in motivatie voor de opdracht tussen de klassen?

Evaluatie project ansich

- Wat ging er minder goed in het project en kan de volgende keer anders?
- Wat ging er wel goed bij het project?
- Tijdsinspanning

Leerervaring project

- Heb je iets van het gehele project geleerd? Ga je de volgende keer bijvoorbeeld dingen anders aanpakken door deze interventie?

Bijlage F Beschrijvingen van de drie verschillende condities

Experimentele conditie 1: zelf criteria opstellen op basis van uitgewerkte voorbeelden.

Deel 1: algemene uitleg opdracht (10 minuten). In de algemene uitleg wordt de opdracht uitgelegd. Inclusief het doel en de verwachtingen. De studenten krijgen klassikaal enkele voorbeelden te zien van de uitwerking van de opdracht.

Deel 2: Wat zijn beoordelingscriteria? (5 minuten). Ondanks dat de studenten wel bekend zijn met het begrip beoordelingscriteria wordt eerst kort toegelicht wat beoordelingscriteria zijn en waarom die zo belangrijk zijn. Om zo laagdrempelig mogelijk te beginnen mogen de studenten beoordelingscriteria bepalen voor een onderwerp waar ze een mening over hebben of zich mee kunnen identificeren. In dit geval is gekozen voor de vraag: ‘Wat maakt iemand tot een goede docent?’. De studenten mogen vijf beoordelingscriteria bedenken waar een goede docent aan moet voldoen. Dit wordt klassikaal besproken. Als vervolgens voor iedereen duidelijk is wat beoordelingscriteria zijn wordt de volgende werkvorm toegelicht.

Deel 3: Uitleg over het zelf maken van beoordelingscriteria voor de opdracht (5 minuten). Aan de studenten wordt duidelijk gemaakt dat ze zelf beoordelingscriteria voor het project mogen verzinnen. Eerst wordt uitgelegd waar ze allemaal aan kunnen denken als het gaat om beoordelingscriteria. Om de studenten niet helemaal vrij te laten worden wel enige restricties aangegeven. Er moeten minimaal acht criteria worden bedacht, meer mag, minder niet. Er moeten criteria bedacht worden voor de lay-out (minimaal 2), de inhoud van de stadgids (minimaal 2), het verslag van 1 a4 (minimaal 2), de presentatie (minimaal 1) en één criterium voor de samenwerking in het groepje. Ten slotte wordt aan de studenten uitgelegd dat de beoordelingscriteria die ze zelf hebben bedacht de basis vormen voor de beoordeling van de opdracht.

Deel 4: gezamenlijk criteria opstellen (20 minuten). De studenten gaan in de groepjes zitten waarmee ze ook het project gaan uitvoeren. Vervolgens kunnen ze op hun computer de 2 voorbeelden van de digitale flyer openen. Ze ontvangen ook twee compleet ingevulde beoordelingsformulieren. Zo krijgen de studenten inzicht in de verschillende mogelijkheden wat betreft het opstellen van beoordelingscriteria. De werkvorm gaat als volgt:

EFFECTEN VAN DRIE VERSCHILLENDE BEOORDELINGSMETHODEN

Ieder subgroepje krijgt een vel papier. In de 4 hoeken van het vel papier staan de clusters lay-out, inhoud, verslag, presentatie en samenwerking. In het midden staat een tabel getekend waar de weging van de verschillende onderdelen van de opdracht kan worden aangegeven. De studenten bekijken de voorbeelden van de digitale flyers en schrijven ieder individueel op wat de beoordelingscriteria moeten zijn voor de opdracht. Vervolgens proberen de studenten tot een gemeenschappelijk antwoord te komen, ze beargumenteren hun antwoord en luisteren naar anderen. Uiteindelijk schrijven de studenten de gemeenschappelijke antwoorden op het vel papier. De uitkomst is dus minimaal acht beoordelingscriteria waar hun digitale flyer aan moet voldoen. Meer mag minder niet. De cesuur voor een voldoende staat al geschreven op het formulier, de weging mogen de studenten wel zelf bepalen.

Deel 5: uitleg vervolg opdracht (5 minuten). Ten slotte wordt aan de studenten nog een keer expliciet uitgelegd dat deze beoordelingscriteria die ze zelf hebben bedacht de basis vormen voor de beoordeling. Na het opstellen van de beoordelingscriteria mogen de studenten beginnen aan het uitwerken van de opdracht. Ook ontvangen ze eenmaal tussentijdse feedback die expliciet gerelateerd is aan de zelfopgestelde beoordelingscriteria.

Experimentele conditie 2: beoordelen door middel van hele uitgewerkte beoordelingscriteria

Deel 1: algemene uitleg opdracht (10 minuten). In de algemene uitleg wordt de opdracht uitgelegd. Inclusief het doel en verwachtingen. De studenten krijgen klassikaal enkele voorbeelden te zien van de uitwerking van de opdracht.

Deel 2: uitleg aan de studenten over de rubric (15 minuten). Aan de studenten wordt uitgelegd dat dit keer anders wordt beoordeeld dan normaal. Normaal gesproken verteld de docent wel aan welke criteria de opdracht moet voldoen maar niet zo uitgebreid. Dit keer wordt er wel gebruik gemaakt van hele uitgebreide beoordelingscriteria. Namelijk in de vorm van een rubric. Uitgelegd wordt wat een Rubric is. Vervolgens worden de criteria en standaarden van de rubric één voor één uitgelegd. Ook de cesuurbepaling wordt toegelicht.

Deel 3: Uitleg vervolg van de opdracht (5 minuten). Na de uitleg van de rubric mogen de studenten verder werken aan het project. Er wordt nog wel duidelijk gemaakt dat de beoordeling van de opdracht plaatsvindt op basis van de rubric. Ook ontvangen ze eenmaal tussentijdse mondelinge feedback op basis van de beoordelingscriteria in de rubric.

Controleconditie

Deel 1: algemene uitleg opdracht (10 minuten). In de algemene uitleg wordt de opdracht uitgelegd. Inclusief het doel en de verwachtingen. De studenten krijgen klassikaal enkele voorbeelden te zien van de uitwerking van de opdracht.

De beoordelingscriteria. In de controleconditie blijven de beoordelingscriteria, de tussentijdse feedback en het beoordelen op de manier zoals ze altijd zijn uitgevoerd door de docent van het vak burgerschap. De docent heeft een aantal criteria geformuleerd, een cesuur bedacht en de studenten mogen daaraan niets veranderen of toevoegen. Vervolgens gaan de studenten aan de slag met het project en ontvangen ze net zoals in de experimentele condities één keer tussentijdse mondelinge feedback. De feedback is gericht op wat ze tot dan toe hebben gemaakt voor het project maar gaat niet specifiek in op de beoordelingscriteria. Daarna werken de studenten weer verder aan het project en leveren ze de opdracht in.

Bijlage G Feedbackprocedure

De feedback tijdens dit experiment wordt gegeven door de onderzoeker. In dit onderzoek is gekozen om de gegeven feedback grotendeels te baseren op het model van Hattie en Timperley (2007). Volgens Hattie en Timperley (2007) is feedback pas effectief wanneer het antwoord geeft op de drie vragen: “Waar ga ik heen/naartoe?” (feed-up), “Wat heb ik tot nu toe gedaan?” (feedback) en “Wat ga ik verder doen om het gestelde doel te bereiken?” (feedforward). Aan de feed-up vraag wordt getracht te voldoen door de feedback aan te laten sluiten op de beoordelingscriteria. In alle drie de groepen in het onderzoek is de feedback gebaseerd op de beoordelingscriteria. In de ene experimentele groep betekent dat feedback op basis van de inhoud van de rubrics. In de andere experimentele groep betekent dat feedback op basis van de zelfopgestelde criteria van de studenten. In de controleconditie is de feedback gericht op de minimale eisen die de docent aan de opdracht heeft gesteld. Om de “Wat heb ik tot nu toe gedaan?” vraag te beantwoorden wordt feedback gegeven op taakniveau. De taakgerichte feedback gaat over hoe goed is de taak begrepen. De feedback geeft studenten informatie over hun presteren bij het werken aan een taak. De feedback biedt informatie over gemaakte fouten of verkeerde aanpakken (Hattie en Timperley, 2007).

Ten slotte wordt geprobeerd om aan te sluiten bij de feedforward vraag: “Wat ga ik verder doen om het gestelde doel te bereiken”. Daarom ontvangen de studenten procesgerichte feedback. Bij procesgerichte feedback gaat de feedback meer over de uitvoering van de taak, hoe het proces verliep dat nodig is om de taak goed uit te kunnen voeren. De studenten krijgen aanwijzingen over hun strategiegebruik of gebruikte aanpak. Volgens Hattie en Timperley (2007) is dit de meest effectieve vorm van feedback. Een voorbeeld van toegepaste resultaatgerichte feedback is aangeven wat tot nu toe goed is gedaan of nog niet goed is gedaan. Een voorbeeld van procesgerichte feedback is het stellen van open vragen aan de studenten over de gemaakte werkzaamheden en het geven van voorbeelden en hints.

Bijlage H Codeboom Vragen Studenten

