

# MASTER'S THESIS

## Values als Invloedrijke Factoren voor Docenten bij de Toepassing van Assessment for Learning

Van Alphen, Thijmen

**Award date:**  
2019

[Link to publication](#)

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 04. Dec. 2021

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



Values als Invloedrijke Factoren voor Docenten bij de Toepassing van  
Assessment for Learning

Values as Influential Factors for Teachers in the Application of Assessment for  
Learning

Thijmen A. van Alphen

Master Onderwijswetenschappen

Open Universiteit

Datum: 23 augustus 2019

Begeleiding: dr. Ron Pat-El

Procesbegeleiding: dr. Kim Dirkx

### Samenvatting

Onderwijs waarbij de nadruk ligt op de eindbeoordeling kan negatieve gevolgen hebben voor de motivatie van leerlingen en de kwaliteit van hun leerproces. Assessment for Learning (AfL) is een onderwijsbenadering die een tegenbeweging vormt voor dergelijk onderwijs, maar in de praktijk weinig wordt toegepast. De keuze van docenten voor de toepassing van dergelijk innovatief onderwijs hangt mogelijk samen met de values (waarden) die docenten er op nahouden. Het doel van deze studie was om te onderzoeken of de vier values (Intrinsic, Attainment, Utility en Cost) uit de Expectancy-Value Theory van invloed zijn op de keuze van docenten om AfL toe te passen in hun lespraktijk.

Met behulp van convenience sampling en snowball sampling zijn 2250 docenten in het Nederlandse middelbaaronderwijs via e-mail bereikt. Hiervan hebben 15 (0.67%) docenten geparticipeerd aan drie digitale meetmomenten en de toegepaste interventie. De interventie bestond uit een 7,5 minuten durende Webinar, waarin docenten spraken over hun positieve ervaring met de toepassing van AfL. Deze werd ingezet na de afname van het tweede meetmoment met als doel de values omtrent de toepassing van AfL positief te beïnvloeden. De afgenomen vragenlijst is samengesteld uit schalen van meerder gevalideerde vragenlijsten, waaronder de Assessment for Learning Questionnaires for Teachers (TAFL-Q), Intrinsic Motivation Inventory (IMI) en de Instructional Efficacy Scale. De items van deze schalen zijn in overleg met docenten verwoord zodat zij passen bij de praktijk en de omgang met AfL.

Dit pretest-posttest pre-experiment met drie meetmomenten had als doel om natuurlijke groei in de values te detecteren en het effect van de interventie zichtbaar te maken. De data werd gemodelleerd in een Structural Equation Model raamwerk om zo de relaties tussen de variabelen te onderzoeken.

Bij het in verband brengen van de values en de keuze van docenten bij de toepassing van AfL werd de cost-value apart opgenomen in het model. Deze value bleek op een andere wijze effect uit te oefenen op de keuze. De attainment, intrinsic en utility-values leken wel een significante relatie te hebben met de keuze van docenten. Om een functionerend groeimodel te kunnen maken waarbij de verandering van de values over de drie meetmomenten werd opgenomen, moest de cost-value volledig buiten beschouwing worden gelaten. Het groeimodel, zonder de cost-value, suggereerde dat de interventie mogelijk effect heeft gehad op de wijze waarop docenten kijken naar de toepassing van AfL.

Op basis van deze resultaten kan men deze studie zien als een eerste stap richting het in kaart brengen van het effect van values op de toepassing van AfL in de lespraktijk. Dit verband lijkt namelijk aanwezig, met uitzondering van de cost-value, welke nader onderzocht dient te worden om te achterhalen wat de exacte relatie hiervan is in verhouding tot de andere constructen. De resultaten en de literatuur suggereren dat de cost-value niet direct van invloed is op de keuze van mensen, maar wellicht indirect via de andere values invloedrijk is. Daarnaast is een eerste stap gemaakt richting de

ontwikkeling en het testen van een interventie op de values rond AfL. Aanvullend onderzoek met een grotere steekproef en een controle groep is echter nodig om conclusies te trekken.

*Sleutelwoorden : Values, Assessment for Learning, Teachers, Intervention*

#### Abstract

Teaching focused on test results can negatively influence pupil motivation and the quality of their learning. Assessment for Learning (AfL) is a teaching approach that acts as a counter-movement to this form of summative assessment, but to date, is little used in practice. Whether teachers choose to apply such innovative teaching, may be related to the values teachers uphold concerning the use of AfL. The aim of this study was to investigate whether the four values (Intrinsic, Attainment, Utility and Cost) from the Expectancy-Value Theory influence teachers' choice in the application AfL.

With the help of convenience sampling and snowball sampling, 2250 teachers in Dutch secondary education were approached by e-mail. Of these, 15 (0.67%) teachers participated in all three online measurements and the applied intervention. The intervention consisted of a 7.5-minute Webinar, in which teachers talked about their positive experience with the application of AfL. This intervention was administered after the second round of measurements, with the aim of positively influencing the values of the participants concerning the application of AfL. The questionnaire used to measure the variables (the four individual values, expectancy and the outcome of teacher choice) consists of scales from several validated questionnaires, including the Assessment for Learning Questionnaires for Teachers (TAFL-Q), Intrinsic Motivation Inventory (IMI), and the Instructional Efficacy Scale. The items of these scales were reformulated with the help of teachers, in order to better fit with this AfL orientated research.

This pretest-post-test pre-experiment with three measurements was intended to detect natural growth in the values and to make the effect of the intervention visible. The data was modelled in a Structural Equation Model framework to investigate the relationships between the variables.

When linking the values and the choice of teachers to apply AfL, the cost value was included as a separate component in the model. This value appeared to have a different effect on the choices made. The attainment, intrinsic, and utility values seemed to have a significant relationship with the choice of teachers. In order to create a functioning growth model, in which the change of values over the three measurement periods was included, the cost value had to be disregarded in its entirety. This growth model, without the cost-value, showed that the intervention may have had an effect on the way in which instructors look at the application of AfL.

On the basis of these results, this study can best be seen as a preliminary step towards mapping the effect of values on the application of AfL in teaching practice. This connection appears to exist in accordance with the literature, with the exception of the cost-value, which needs to be further investigated. These results and other research suggest that the cost-value does not directly influence people's choice, but may have an indirect role in influencing other values. In addition, a first step has

been made towards the development and testing of an intervention on the values around AfL. Additional research with a larger sample size and the use of a control group is recommended.

*Keywords : Values, Assessment for Learning, Teachers, Intervention*

## Inhoudsopgave

Samenvatting.....	2
Abstract.....	3
1. Inleiding.....	6
1.1 De trend in onderwijsevaluatie en het effect op leerlingen.....	7
1.2 Assessment for Learning als reactie op de summatieve evaluatiecultuur.....	8
1.3 Values vormen de houding van docenten.....	9
1.4 Expectancy-Value Theory.....	10
1.5 Vraagstellingen.....	12
2. Methode.....	12
2.1 Ontwerp.....	12
2.2 Participanten.....	12
2.3 Materialen.....	13
2.3.1 Attainment-Value.....	13
2.3.2 Utility-Value.....	13
2.3.3 Intrinsic-Value.....	13
2.3.4 Cost-Value.....	14
2.3.5 Expectations.....	14
2.3.6 AfL.....	14
2.4 Interventie.....	15
2.5 Procedure.....	15
2.6 Data-analyse.....	16
3. Resultaten.....	18
4. Discussie.....	22
4.1 Values, kosten en Assessment for Learning.....	23
4.2 De verandering van values.....	24
4.3 Het meten van values.....	25
4.4 Implicaties.....	26
Referenties.....	27
Bijlage A.....	30

## Values als Invloedrijke Factoren voor Docenten bij de Toepassing van Assessment for Learning

### 1. Inleiding

De reviewstudie van Harlen en Crick (2003) naar de invloed van toetsen, beschrijft dat het gebruik van toetsen op strikt beoordelende wijze (e.g., *assessment of learning*) negatieve gevolgen kan hebben voor de motivatie van leerlingen. Aanvullend daarop concludeert Kohn (2011) dat uitsluitend cijfers ontvangen ter eindboordeling van het verkregen onderwijs niet alleen de motivatie doet afnemen, maar ook de kwaliteit van het leerproces omlaag haalt. Het gebruik van toetsen op enkel beoordelende wijze is hiermee een belangrijk aandachtspunt voor onderzoek. Birenbaum et al. (2015) beschrijven dat internationaal veel inspanning wordt verricht om de balans te vinden tussen toetsen die gedurende het leerproces ondersteuning bieden en beoordelende toetsen die pas aan het eind van het leerproces worden ingezet om het leerresultaat te meten. Het blijkt dat deze laatste vorm van onderwijsevaluatie regelmatig de overhand heeft.

Een onderwijsbenadering die een tegenbeweging vormt voor het alleen beoordelend evalueren is *Assessment for Learning* (AfL). AfL omvat naast toetsen ook andere activiteiten die docenten gedurende het leerproces ondernemen om informatie te verzamelen over het vermogen, het kennisniveau en de onderwijsdoelen van de leerling (Black & Wiliam, 2010). De docent geeft niet alleen terugkoppeling aan de leerling, maar past idealiter met deze verkregen informatie ook de eigen lespraktijk aan (Wiliam, 2011). James en Pedder (2006) geven aan dat er sterke wetenschappelijke ondersteuning is voor de positieve effecten op de leerkwaliteit door het systematische gebruik van AfL. De toepassing van AfL verloopt echter niet overal even gemakkelijk. Uit het onderzoek van Kitiashvili (2014) bleek dat docenten regelmatig het gevoel hebben te weinig expertise te bezitten voor het succesvolle gebruik van AfL. Deze docenten benoemden dat daardoor een terugval op een bekende beoordelende evaluatievorm verleidelijk is. Een eerdere survey van Neesom (2000) toonde aan dat docenten formatieve evaluatie zien als een complexe methode en als een extra activiteit, bovenop de reguliere lestaak. Docenten uitten daarbij onder meer hun zorg over de benodigde voorbereidingstijd en de opbrengsten van deze onderwijsbenadering.

Ondanks deze zorgen met betrekking tot de toepassing van AfL, verzamelde Gipps (1994) al eerder empirisch bewijs dat de achterliggende ideeën van AfL door docenten als authentieke uitdrukking van onderwijswaarden worden gezien; docenten zien de focus van AfL op het *leerproces* namelijk als kernwaarde van goed onderwijs. (Onderwijs)waarden kunnen daarmee belangrijk zijn in de manier waarop docenten AfL ervaren. Vanuit de Expectancy Value Theory (EVT) van Eccles (1983) werd gesteld dat values (waarden) de keuzes van mensen (i.e., *achievement-Related Choices*) dusdanig kunnen beïnvloeden dat zij meer volharden en daarmee een taak vaker tot een succes zullen volbrengen. Het idee dat de values bij docenten belangrijk zijn in het uitvoeren van (innovatieve) onderwijstaken komt ook naar voren in het werk van Brand-Gruwel, Bos en Van der Graaf (2019). Zij

stellen dat een groot deel van de onderwijsinnovaties niet succesvol zijn, als de docent geen heil ziet in de uitvoering daarvan.

Mogelijk bepalen de values die docenten over een onderwijsbenadering zoals AfL erop nahouden, op een vergelijkbare manier ook de mate waarin docenten gedreven zijn om AfL succesvol toe te passen. Het doel van dit onderzoek is dan ook om te achterhalen in welke mate de values uit de EVT voorspellend zijn voor de keuze van docenten om zich in te zetten voor de toepassing van AfL en of deze values van docenten positief te beïnvloeden zijn door middel van een interventie. Met deze kennis kunnen scholen, opleidingen en onderwijsinstanties docenten voorzien van gerichte ondersteuning wanneer de toepassing van formatieve evaluatie niet succesvol of moeizaam verloopt.

### **1.1 De trend in onderwijsevaluatie en het effect op leerlingen**

Het moment waarop een toets gebruikt wordt en de manier hoe met de verkregen informatie wordt omgegaan, is bepalend voor het karakter van de toets. Wiliam (2011) geeft aan dat een toets aan het begin of halverwege het leerproces kan dienen als meetmoment en zowel de docent als leerling voorziet van belangrijke informatie over de voortgang van het leerproces. Met deze informatie kan de route naar het einddoel beter worden bepaald. Wanneer toetsinformatie op deze manier wordt gebruikt, spreekt men van formatief toetsen (i.e., formatieve evaluatie). Toetsen kunnen ook gebruikt worden aan het einde van het proces om te achterhalen of de leerling genoeg heeft geleerd. Deze vorm van toetsen heeft een beoordelend karakter waarbij een specifiek leerproces wordt afgerond; de verkregen informatie kan niet worden gebruikt om het proces verder bij te sturen. Wanneer een toets op deze beoordelende wijze wordt gebruikt, spreekt men van summatief toetsen (i.e., summatieve evaluatie).

Kenmerkend is dat leerlingen, die alleen op summatieve wijze worden getoetst, minder begrijpen en onthouden van de leerstof, dan leerlingen die niet met een cijfer werken als eindresultaat (Kohn, 2011). Het enkel toetsen op summatieve wijze, zorgt ervoor dat leerlingen minder geïnteresseerd zijn in de inhoud van de leerstof (Harlen & Crick, 2003) en zich meer richten op de beoordeling. Hierdoor geven zij vaak eerder de voorkeur aan een makkelijke opdracht of taak om daarmee de kans op een betere beoordeling te vergroten (Pulfrey, Buchs, & Butera, 2011). Dit leidt mogelijk tot een afname in de kwaliteit van het denken, doordat de leerlingen zich enkel richten op de onderdelen van de leerstof die noodzakelijk zijn voor het behalen van een eindbeoordeling (Kohn, 2011). Zo tonen verschillende onderzoeken dat enkel summatief toetsen van leerlingen negatieve gevolgen met zich mee kan brengen voor de motivatie, kwaliteit van het denken en de kwaliteit van het gemaakte schoolwerk.

Wereldwijd onderzoek naar het spanningsveld tussen formatieve en summatieve evaluatie laat zien dat het formatieve deel in veel landen onder druk staat door grootschalige summatieve toetsen



(Birenbaum et al., 2015). Voorbeelden hiervan zijn de eindexamens aan Nederlandse middelbare scholen of de 'General Certificate of Secondary Education' examens in Engeland. Segers benoemde in haar oratie van 2004 al dat de evaluatie trend in Nederland en andere Europese landen leek te bewegen richting een toename in summatief toetsen; een volgens haar zorgwekkende ontwikkeling gezien de meerwaarde van formatieve evaluatie voor de het onderwijs. Ondanks de door Birenbaum et al. (2015) gedocumenteerde initiatieven van de betrokken landen om meer formatieve evaluatie toe te passen, is de door Segers benoemde trend nog altijd in volle gang en is er sprake van een overheersend summatieve evaluatiecultuur (Birenbaum et al., 2015).

### **1.2 Assessment for Learning als reactie op de summatieve evaluatiecultuur**

Een reactie op deze summatieve toetscultuur vormt Assessment for Learning (AfL) (Black & Wiliam, 1998). AfL richt zich op de uitwisseling van informatie over het onderwijsproces tussen docenten en leerlingen. Daarbij is er een algemene consensus dat drie aspecten een voorwaardelijke basis vormen voor een gedegen uitvoering van AfL (Brown, 2004; Gulikers & Baartman, 2017; James & Pedder, 2006; MacLellan, 2010; The National Council of Teachers of Mathematics, 2007; Wiliam, 2011). Wiliam benoemde in zijn oratie van 2010 als voorwaarden dat (1) de docent bezig is met het achterhalen van het onderwijsdoel van de leerling en het curriculum. Dit wordt mede gedaan door in gesprek met de leerlingen te gaan en de leerintenties te verduidelijken. Hierbij kan de docent in discussie gaan over de noodzakelijk doelen vanuit school, maar ook in gaan op de persoonlijke wensen van de leerlingen. (2) Daarbij is de docent ook bezig met het achterhalen van het huidige kennisniveau van de leerling. Dit doet de docent door discussies, taken en andere activiteiten te organiseren die bewijs ontlokken wat de leerling leert of heeft geleerd. Op dit punt is het niet de docent die feedback geeft, maar ontvangt. De manier waarop leerlingen antwoorden en reageren op de opdrachten van de docent is informatief. Hiermee kan de docent zijn of haar eigen lespraktijk anders organiseren, waarop deze beter aansluit op de gestelde leerdoelen. (3) Tot slot is AfL gecentreerd rond het geven van feedback aan de leerling. Ook hier is de feedback gericht op het bereiken van de eerder bepaalde doelen. Met dit onderdeel ondersteunt de docent de leerling bij het leerproces. Dit kan de docent doen door de lesvorm aan te passen op de behoefte van de leerling, gericht leerstrategieën aan te bieden of informatie aan te reiken die nuttig lijkt te zijn op dat punt van het leerproces. Op deze manier wordt getracht de uiteindelijke doelen van AfL te realiseren, namelijk de bevordering van dieper leren en leerlingautonomie (James & Pedder, 2006). Wanneer AfL gedegen wordt toegepast voorkomt het daarmee mogelijk de eerder benoemde nadelige effecten van summatieve evaluatie.

De drie voorwaarden laten zien dat het AfL-proces gericht is op de behoefte van en de interactie met individuele leerlingen. Het is wellicht daarom dat Wiliam de zorgen aanspreekt die auteurs sinds Bloom (1984) al opperden over de financiering en tijd van dergelijk individualistisch

onderwijs. In zijn oratie (Wiliam, 2010) over de stand van zaken omtrent AfL geeft Wiliam aan dat docenten niet harder moeten gaan werken, maar slimmer. Hoewel Wiliam adviseert over de invoering van AfL, benoemen docenten die met AfL werken nog altijd dat het formaat van de klas en de benodigde tijd voor het geven van (individuele) feedback belangrijke factoren zijn die de toepassing van AfL bemoeilijken. Naast de eveneens door Kitiashvili (2014) gevonden problemen met betrekking tot tijdsgebrek en onvoldoende expertise die ervaren wordt door sommige docenten, benoemt Neesom (2000) dat het hierdoor gevormde moreel (i.e., de wil om door te zetten) onder docenten een mogelijke rol heeft bij de succesvolle toepassing van formatieve evaluatie. De docenten uit dit onderzoek gaven aan dat dit moreel ook beïnvloed wordt door het gevoel dat formatieve evaluatie een taak is die veel extra werk oplevert en dat het niet altijd even zichtbaar is wat de opbrengsten zijn van deze onderwijsbenadering.

### **1.3 Values vormen de houding van docenten**

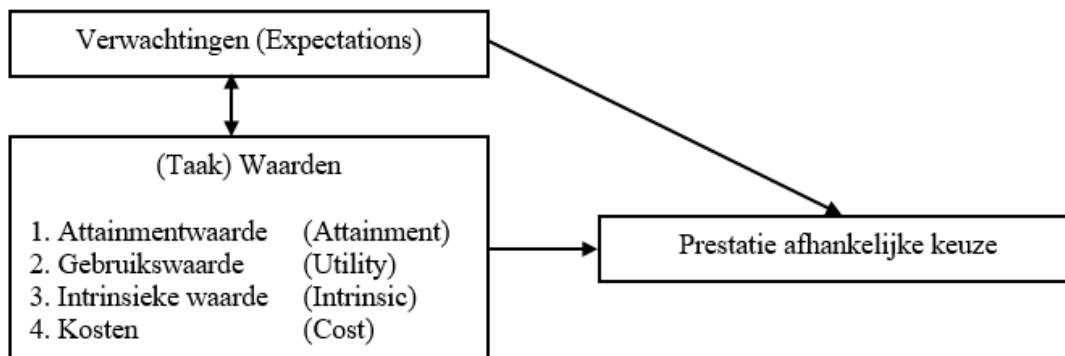
Dat dit moreel en de waarden (i.e., het belang van een taak) die docenten erop na houden belangrijk zijn in de uitvoering van hun werk, was een aandachtspunt in het werk van James en Pedder (2006). Zij benoemden dat de houding, mede gevormd door het belang dat de docent hecht aan de onderdelen van formatieve evaluatie, een onderbelicht aspect is bij de verklaring waarom docenten een bepaalde vorm van onderwijs adopteren. Zij keken naar de toepassing van summatieve en formatieve evaluatie en daarbij vooral naar het verschil tussen de waarde die een docent toebedeelde aan een bepaalde evaluatievorm. Ondanks de wens van docenten om meer formatief te evalueren, bleek in de praktijk toch meer summatief te worden gevalueerd. James en Pedder (2006) gaven aan dat er een sterke behoefte is van docenten om te werken aan het realiseren van formatieve onderwijswaarden.

Het onderzoek van James en Pedder (2006) laat zien dat docenten keuzes maken in hun manier van evalueren op basis van het belang dat zij hechten aan de vorm waarin dit gebeurt. Aansluitend op het eerder genoemde onderzoek van Gipps (1994) waarbij AfL door docenten werd gezien als een vertegenwoordiging van authentieke onderwijswaarden, gaf 83 procent van de docenten in het onderzoek van James en Pedder aan dat ze het belangrijk vinden om meer over het leerproces van hun leerlingen te leren. Hoewel dit onderzoek naar de waarden die docenten hechten aan de verschillende onderdelen van AfL inzicht geeft in de manier waarop docenten denken over AfL, laat het niet zien hoe het komt dat zij deze aspecten als waardevol beschouwen. De vraag waarom docenten de keuze maken op basis van hun waarden, ofwel gemotiveerd zijn om met AfL te werken blijft onderbelicht in het onderzoek van James en Pedder. Juist deze kennis over de manier hoe docenten denken over AfL kan bepalend zijn in de manier waarop ze omgaan met een dergelijke onderwijsbenadering. Om te onderzoeken hoe het komt dat docenten waarden hechten aan een

onderwijbenadering zoals AfL, wordt gebruik gemaakt van het Expectancy-Value model, afkomstig uit de Expectancy-Value Theory (EVT) van Eccles (1983).

#### 1.4 Expectancy-Value Theory

In het Expectancy-Value model wordt verondersteld dat keuzes worden beïnvloed door zowel negatieve als positieve eigenschappen van taken (Wigfield & Eccles, 2002). De inzet voor een taak, ofwel de prestatieafhankelijke keuze (achievement related choices) voor de uitvoering van een taak, is volgens de EVT mede afhankelijk van de values (waarden) die een persoon bij deze taak erop nahoudt (Figuur 1). De prestatieafhankelijke keuze is volgens deze theorie de mate waarin iemand gemotiveerd is om aan de taak te werken waarbij de term motivatie gezien wordt als een beweegreden, ofwel de drijvende kracht om actie te ondernemen (Wigfield & Eccles, 2002).



*Figuur 1.* De relaties tussen values (values), verwachtingen (Expectations) en prestatie afhankelijke keuzes uit het Expectancy-Value model van Wigfield en Eccles (2002).

Het Expectancy-Value model werd in eerste instantie door Eccles (1983) ontwikkeld om de keuzes die kinderen en jongvolwassenen maken tijdens hun schoolcarrière te kunnen verklaren. Maar recentere toepassingen van het Expectancy-Value model die de motivatie van leraren onderzoeken, hebben de oorspronkelijke EVT-constructen van Eccles (1983) opnieuw bekeken en geconceptualiseerd (Watt & Richardson, 2007). Het begrip value zoals toegepast op de motivatie van docenten, omvat (1) waarde van de persoonlijke bruikbaarheid (attainment-value), met nadruk op het persoonlijke belang of nut van de taak, (2) waarde van het maatschappelijk belang (utility-value), (3) intrinsic-value die het plezier en de tevredenheid met het resultaat beschrijft en (4) de kosten, een breed begrip dat inspanningen omvat zoals financiële en emotionele kosten. Deze beschrijving van docenten values sluit nauwlettend aan op de door Wigfield en Eccles (2002) oorspronkelijk gegeven definities van het begrip values, een indicatie dat het model inderdaad toepasbaar lijkt te zijn bij de verklaring van keuzes die docenten maken.

Eccles (1983) beschrijft de attainment-value als de behoefte van een persoon om een taak goed te doen. Zij veronderstelt hierbij dat deze behoefte niet enkel een interne beleving is, maar mede

gevormd wordt door de omgeving waarin de persoon zich bevindt. De definitie van deze value raakt hiermee aan de beschrijving die Ryan en Deci (2000) geven van de noodzaak om aan psychologische behoeften te voldoen. Deze behoefte zou een sterke beweegreden kunnen zijn bij het aangaan van uitdagingen of nieuw taken. Voor docenten kan dit betekenen dat erkenning van leerlingen, ouders, collega's en eventueel schoolbestuur voor het succesvol toepassen van AfL een drijfveer kan zijn om hiermee verder te gaan.

De utility-value (nutswaarde) wordt bepaald door de mate waarin een taak zich verhoudt tot huidige en toekomstige vooronderstellingen van het nut dat de uitvoering van de taak met zich mee brengt. Een taak kan daarmee een positieve waarde voor een persoon hebben, omdat het belangrijke toekomstige doelen mogelijk maakt. Wigfield en Eccles (2002) geven als voorbeeld dat studenten vaak lessen nemen die ze niet echt leuk vinden, maar die ze volgen om andere interesses na te streven. Waar de attainment en intrinsieke values meer gebaseerd zijn op interne prikkels, is de utility-value een meer 'extrinsieke' drijfveer om aan een taak deel te nemen. Bij docenten zou deze values beïnvloed kunnen worden door het huidige of toekomstige onderwijssucces van leerlingen wat bereikt zou kunnen worden door de toepassing van AfL.

De intrinsic-value (intrinsieke waarde) verwijst naar het plezier dat een individu ervaart door het uitvoeren van de taak of het belang dat het individu hecht aan de uitvoering daarvan. Ook bij deze value heeft het werk van Ryan en Deci (2000) een verdiepende rol met betrekking tot de definitie en de invloed op de beweegredenen van een persoon. Ryan en Deci spreken in hun werk over 'intrinsieke motivatie', ofwel een drijvende kracht die uit een persoon zelf komt. Deze kracht is volgens hen belangrijk bij het aangaan van nieuwe uitdagingen en het opzoeken van de grenzen. Ook hier is het mogelijk dat deze drijvende kracht voortkomt uit het voldoen aan psychologische behoeften. Dat zou een verklaring kunnen zijn voor de 'genotsbeleving' waarover wordt gesproken in Wigfield en Eccles (2000). In het kader van AfL zou intrinsieke motivatie voort kunnen komen uit een geslaagde toepassing van deze onderwijsbenadering.

Tot slot benoemen Eccles en Wigfield de kosten als een belangrijke value (Eccles, 1983; Wigfield & Eccles, 2000). De cost-value (kostenwaarde) wordt beschreven als de negatieve gevolgen of belastende aspecten die een taak tot gevolg kan hebben. De hoeveelheid inspanning die nodig is om te slagen en daarmee de energie en tijd die worden uitgegeven, zijn voorbeelden van kosten die verbonden kunnen zijn bij het uitvoeren van een taak (Wigfield & Eccles, 2002). Docenten hebben inderdaad tijd en expertise benoemd als factoren die invloedrijk zijn in het toepassingssucces. Een docent die geen tijd heeft om de principes van AfL te leren of toe te passen in de les, schat zijn of haar kansen voor succes mogelijk klein in (Kitiashvili, 2014).

### **1.5 Vraagstellingen**

Om de invloed van de beschreven values bij docenten te onderzoeken, is de centrale vraag in dit onderzoek: “Wat is het effect van een professionaliseringsmaatregel (Webinar) op de values (waarden) die docenten erop na houden over AfL en heeft een verandering in deze values invloed op de keuze van docenten om AfL toe te passen in hun lespraktijk?”. Deze centrale vraag wordt onderzocht met behulp van een interventie in de vorm van een Webinar. Deze interventie fungeert als manipulatie en is in het bijzonder gericht op het positief bijstellen van de values omtrent AfL. Om de verschillende onderdelen van deze vraagstelling gedegen te kunnen beantwoorden zijn de volgende deelvragen opgesteld:

Deelvraag 1: Is er sprake van een natuurlijke groei in de values die docenten erop na houden over het gebruik van AfL?

Deelvraag 2: In welke mate hebben de values invloed op de keuze van docenten om AfL toe te passen in hun lespraktijk?

Deelvraag 3: In welke mate zijn de values bij docenten te beïnvloeden door het gebruik van een professionaliseringsmaatregel in de vorm van een webinar?

## **2. Methode**

### **2.1 Ontwerp**

Om de onderzoeksvraag te beantwoorden is een pretest-posttest pre-experiment uitgevoerd, waarbij een interventie is ingezet en de observaties door middel van digitale vragenlijsten zijn uitgevoerd. De vragenlijsten dienden eventuele verandering bij docenten te meten ten aanzien van hun values en verwachtingen bij de keuze om AfL toe te passen. De vragenlijsten zijn op drie verschillende momenten ingezet, met een elk tussentijdse periode van twee weken.

De predictor variabelen zijn de vier values (attainment, utility, intrinsic en cost) en de outcome variabele is de keuze van docenten om met AfL te werken. De docenten die participeerden in dit onderzoek genereerden hun eigen identificatiecode. Hiermee zijn zij dusdanig geordend dat alleen de docenten die aan alle drie de metingen participeerden zijn meegenomen in de uiteindelijke analyse. Overige docenten die niet aan het volledige experiment participeerden, veroorzaakten hierdoor minder ruis in de analyse. Tot slot zijn de verwachtingen (expectancy) van docenten meegenomen als confounder, omdat er theoretische ondersteuning bestaat voor de samenhang tussen de verwachtingen, values en keuze van mensen om te werken aan een bepaalde taak (Wigfield & Eccles, 2002).

### **2.2 Participanten**

De steekproef is tot stand gekomen door binnen een grote scholengroep in de stad Amsterdam de docenten, via de schoolbesturen, per email te benaderen. Hierbij zijn docenten bereikt die lesgeven aan

meerdere middelbare schoolniveaus, zoals enkel VMBO (15%), HAVO/VWO (54%), VMBO/HAVO/VWO (23%) en enkel VWO (5%). Deze aanpak heeft naar schatting 1550 docenten bereikt. Daarnaast zijn door middel van snowball-sampling aanvullend 700 docenten van verschillende scholen buiten Amsterdam bereikt. Van deze totale groep ( $N = 2250$ ) hebben 39 docenten (1.73%) geparticipeerd in de eerste meting, 17 (0.75%) in de tweede meting en 15 (0.67%) in de laatste meting. De gemiddelde leeftijd van de docenten uit deze steekproef is 42.9 jaar ( $SD = 12.4$ ). De docenten uit deze steekproef hebben gemiddeld 14.3 jaar ( $SD = 10.9$ ) ervaring in het voortgezet onderwijs.

## 2.3 Materialen

De onafhankelijke variabelen in dit onderzoek zijn gebaseerd op de Expectancy-Value Theory (Eccles & Wigfield, 2002) en betreffen de verwachtingen (expectancy) en de vier values (attainment, utility, intrinsic en cost) die docenten erop na houden bij de toepassing van AfL. De schalen voor de onderzoeksvariabelen zijn in een vragenlijst gemeten met behulp van een 5-punts Likertschaal die loopt van 1= *zeer mee oneens*, tot 5= *zeer mee eens*. De samengestelde vragenlijst is voorafgaande aan de meting door zeven docenten uit het middelbare onderwijs voorzien van feedback. Op basis hiervan zijn onduidelijke items herschreven. De volledige vragenlijst is opgenomen in Bijlage A. De interne consistentie van de onderstaande schalen is bepaald met behulp van de eerste meting. Bij de overige metingen waren te weinig respondenten om deze betrouwbaarheidsanalyse uit te voeren.

**2.3.1 Attainment-Value.** De schaal om de attainment-value (attainment) te meten is samengesteld op basis van gesprekken met docenten in het middelbare onderwijs. In de literatuur bleek geen schaal voor de attainment-value toepasbaar op deze specifieke vorm van het construct. Als uitgangspunt voor deze gesprekken is gevraagd aan docenten op welke manier zij plezier beleefden aan het werken met AfL. De schaal bestaat uit vier items. Een voorbeelditem uit deze schaal is: *“ik ben een goede docent als ik mijn leerlingen betrek bij het stellen van persoonlijke leerdoelen”*. De betrouwbaarheidsanalyse van deze schaal toonde een maximale interne consistentie van  $\alpha = .65$ ,  $CI = [.44, .83]$ .

**2.3.2 Utility-Value.** Om de gebruikswaarde (utility) bij docenten te meten is met behulp van het werk van Neesom (2000) een schaal hiervoor samengesteld. Deze schaal meet het belang dat docenten hechten aan de opbrengsten die werken met AfL kan leveren en bestaat uit negen items. Een voorbeelditem uit deze schaal is: *“De toepassing van Assessment for Learning heeft voor mij als meerwaarde dat er sprake is van een toename in de kwaliteit van mijn lesgeven”*. De betrouwbaarheidsanalyse van deze schaal toonde een interne consistentie van  $\alpha = .89$ ,  $CI = [.88, .91]$ .

**2.3.3 Intrinsic-Value.** De intrinsieke waarde (intrinsic) gaat over het genot dat een docent kan beleven aan het werken met AfL. Dit is gemeten door gebruik te maken van de enjoyment-schaal uit de Intrinsic Motivation Inventory, gevalideerd door McAuley, Duncan, en Tammen (1989). De

oorspronkelijke schaal toonde een interne consistentie met  $\alpha = .78$  en bestaat uit vijf items. Een voorbeelditem uit deze schaal is: “Door de toepassing van Assessment for Learning ontwikkel ik interesse voor procesgericht onderwijs”. De betrouwbaarheidsanalyse van deze schaal toonde een interne consistentie van  $\alpha = .82$ , CI = [.73, .91].

**2.3.4 Cost-Value.** De kostenwaarde (cost) is gemeten met behulp van een schaal samengesteld op basis van de door Neesom (2000) geïnventariseerde lasten die docenten ervoeren bij de omgang met formatieve evaluatie. Deze schaal bestaat uit zeven items. Een voorbeelditem uit deze schaal is: “Door de toepassing van Assessment for Learning kom ik in de problemen doordat mij dit extra werkdruk oplevert”. De betrouwbaarheidsanalyse van deze schaal toonde een interne consistentie van  $\alpha = .88$ , CI = [.82, .93].

**2.3.5 Expectations.** Om de verwachtingen (expectations) van docenten te meten over het succes dat zij zullen hebben bij het toepassen van AfL, is gemeten met een aangepaste versie van de Instructional Efficacy-schaal van Tschannen-Moran en Hoy (2001). De oorspronkelijke schaal toonde een interne consistentie met  $\alpha = .91$ . De schaal bestaat uit acht items. Een voorbeelditem uit deze schaal is: “Ik verwacht het leerproces van de leerling gericht te kunnen aansturen”. De betrouwbaarheidsanalyse van deze schaal toonde een interne consistentie van  $\alpha = .85$ , CI = [.77, .92].

De confirmatory factor analysis (CFA) voor de schalen van verwachtingen en values is uitgevoerd in R 3.5.3 met het Lavaan pakket (v.0.6-3) om de verwachte oblique vijf-factorstructuur van de Expectancy-Value schalen te toetsen. De fit van het model was zeer zwak  $\chi^2(454) = 1451.989$ ,  $p < .001$ , CFI = .504, RMSEA = .166, SRMR = .127. Gezien de verwachtingen van docenten als versturende variabelen is meegenomen in deze schaal, is de schaal voor verwachtingen verwijderd. Dit nieuwe vier-factormodel was gaf een verbetering in de fit-indicatoren  $\chi^2(246) = 476.002$ ,  $p < .001$ , CFI = .619, RMSEA = .157, SRMR = .122, maar de fit bleef matig en deze verbetering bleek ook niet significant ( $p = .305$ ).

**2.3.6 AfL.** De outcome-variabele in dit onderzoek is de keuze om AfL toe te passen in de lespraktijk. Dit is gemeten door een aangepaste versie van de Teacher Assessment For Learning Questionnaire (TAFL-Q) (Pat-El, Tillema, Segers, & Vedder, 2013). De items uit deze vragenlijst zijn omgezet naar stellingen die de keuze meten om verschillende onderdelen van AfL toe te passen in de les. De oorspronkelijke schaal van 28 vragen toonde een interne consistentie met  $\alpha = .82$ . Een voorbeelditem uit deze schaal is: “Ik kies ervoor om met mijn leerlingen mee te denken over de manier waarop ze willen leren op school”. De betrouwbaarheidsanalyse van deze schaal toonde een interne consistentie van  $\alpha = .90$ , CI = [.85, .94]. Een CFA voor deze schaal is uitgevoerd om de verwachte oblique één-factorstructuur te toetsen. De fit van dit één-factormodel was  $\chi^2(378) = 719.258$ ,  $p < .001$ , CFI = .402, RMSEA = .164, SRMR = .141.

## 2.4 Interventie

De interventie had de vorm van een Webinar en was gebaseerd op de twaalf principes van de Multimedia Learning Theory, zoals beschreven in het werk van Mayer (2009). Mayer geeft aan dat wanneer men deze principes toepast, dit zorgt voor ‘deep learning’ doordat er meerdere cognitieve kanalen (auditief en visueel) tegelijk worden aangesproken. Aanvullend zijn ook andere principes uit het werk van Mayer (2009) gebruikt om dieper leren aan te sporen. Zo volgt de gehele interventie het voice-principe (het gebruik van een menselijk stem), het coherence-principe (geen gebruik onnodige begrippen), het signaling-principe (belangrijke begrippen benadrukken), het redundancy-principe (alleen de combinatie van twee informatiecomponenten om de respondent niet te overbelasten), het multimediaprincipe (niet alleen tekst, maar ook beeld gebruiken) en het personalisatie-principe (spraak is in conversatievorm en niet formeel).

Docenten die in de webinar figureerden deelden hun persoonlijke en positieve ervaringen met het gebruik van AfL. Het delen van deze ervaringen was gericht op het specifiek beïnvloeden van de values omtrent het gebruik van AfL. Acee en Weinstein (2010) gaven aan dat het specifiek richten op individuele values beter werkt dan de values al geheel proberen te beïnvloeden met een interventie. Daarbij is de interventie dusdanig ontworpen dat alle manipulaties op de aparte values direct van toepassing zijn op de onderwijspraktijk en context van AfL. Als voorbeeld hiervan waarbij de docenten specifiek de values benadrukken: “het geeft mij een goed gevoel geeft om samen met leerlingen leerdoelen te stellen en te werken aan het onderwijsproces”. Ook dit contextueel aanbieden versterkt de kracht van de interventie (Wigfield, Rosenzweig, & Eccles, 2017)

## 2.5 Procedure

De schoolbesturen binnen de onderwijsorganisatie zijn benaderd met behulp van een per e-mail verzonden brief. Deze brief beschreef de informatie over de lengte van het onderzoek, de omgang met verkregen data, de eventuele belasting en opbrengsten voor het personeel en de school. In de brief werd gevraagd om de e-mailgegevens van de docenten te verstrekken en of het bestuur dit onderzoek wilde aankondigen in het weekbericht.

Voorafgaand aan het invullen van de eerste meting is digitaal gevraagd om consent, waarbij de ethische omgang met gegevens werd toegelicht. In de brief werd benadrukt dat de te verzamelen gegevens niet ter herleiden zijn naar de respondent. Er werd voor dit onderzoek door respondenten een persoonlijke code aangemaakt om de data te kunnen verbinden tussen de herhaalde metingen. Deze code was gebaseerd op enkele basisvragen, zoals: “*de laatste letter van uw postcode en de eerste letter van uw achternaam*”. Op deze manier konden de respondenten de code onthouden en bij elke meting reproduceren.



Voorafgaand aan de interventie (Webinar) is opnieuw gevraagd aan de docenten om de vragenlijst in te vullen. Direct na het invullen van de tweede meting zijn de docenten gevraagd om online deel te nemen aan de 7,5 minuut durende interventie. Deze was beschikbaar via een snelkoppeling, aangeleverd via de online survey en direct na het afronden van de tweede meting. De respondenten werden erop gewezen dat deze link te kopiëren is en later ook beschikbaar zou zijn. Twee weken na de interventie is een laatste keer gevraagd om de survey in te vullen. Ook hier is na een week een herinnering gestuurd met daarin het verzoek alsnog deze laatste vragenlijst in te vullen. Voor het online verzamelen van de data is gebruik gemaakt van *Limesurvey*. De docenten hebben in hun eigen tijd de vragenlijsten ingevuld en de deelgenomen aan de Webinar, hiervoor ontvingen ze geen compensatie.

De Webinar werd gekenmerkt door een interview setting, waarbij de interviewer begon met een introductie van de belangrijkste begrippen. Begrippen zoals ‘formatief evalueren’ en ‘Assessment for Learning’ hadden hierin een hoofdrol. Door dit ‘pre-training’-principe toe te passen werden de respondenten bekend gemaakt met deze begrippen en was de verwachting dat zij de hoofdconcepten van de interventie beter konden volgen. Na de introductie volgden vier interviewsegmenten waarin gefilmde docenten in beeld kwamen en hun positieve ervaringen deelden met de toepassing van AfL.

## 2.6 Data-analyse

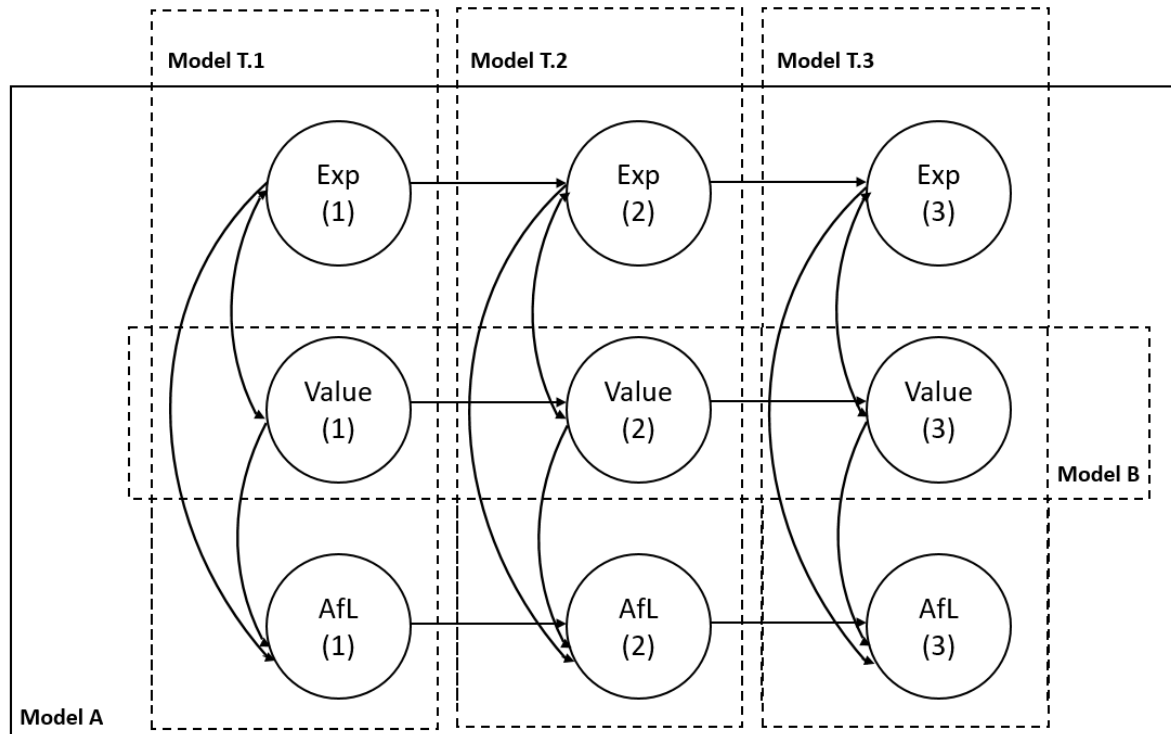
Om outliers te detecteren in de data is gebruik gemaakt van de Robust Squared Mahalanobis afstanden uit het MVN (5.7) pakket. Tevens is dit pakket gebruikt om de multivariate normaliteit te toetsen bij de drie datasets. In R 3.5.3 zijn de drie meetmomenten samengevoegd tot één dataset. Om een longitudinale meting te verkrijgen zonder missing values zijn de data van de respondenten met behulp van hun persoonlijke identificatiecode gesorteerd. Dit resulteerde in een dataset waarbij enkel de respondenten zijn opgenomen die aan alle metingen en de interventie hebben deelgenomen.

Om de onderzoeksvragen te kunnen beantwoorden zijn de correlaties en regressies tussen de constructen getoetst met Latent Modelling binnen het Structural Equation Model (SEM) raamwerk in R (Lavaan pakket). Figuur 2 toont de relaties tussen de latente constructen zoals deze in de Expectancy-Value theorie worden beschreven. Daarbij zijn de relaties tussen de verschillende tijdstmomenten afgebeeld bij een model dat alle variabelen met elkaar in verband brengt (A). Om de vraag te beantwoorden welke relatie er bestaat tussen de afzonderlijke constructen, wordt met de bijbehorende sub-modellen T.1, T.2 en T.3 onderzocht of de data passend is bij de individuele meetmomenten. Daarbij worden de kosten als apart latent construct meegenomen aangezien verwacht wordt dat deze, in tegenstelling tot de andere values, een negatieve invloed heeft op de keuze van docenten. Tevens toont figuur 2 de values als één construct, al zijn deze in de analyse echter als aparte latente variabelen meegenomen om de onderlinge verschillen hiertussen te kunnen onderzoeken.

Vervolgens is met een Latent Growth Model (B) gekeken naar de verandering van de intercepts en de slopes bij de values van docenten over de drie verschillende meetmomenten. Hiermee is getracht duidelijkheid te scheppen over de eventuele verandering in values ten aanzien van AfL bij docenten die zijn ontstaan gedurende het experiment.

De fit van de modellen op de data is bepaald met behulp van RMSEA (root mean squared error of approximation), SRMR (standardized root mean squared residual) en CFI (Comparative Fit Index) indicatoren. Wanneer RMSEA en SRMR een score tonen onder .05 en CFI een score boven .95 is dit indicatief voor een goede fit van het model (Browne & Cudeck, 1992). RMSEA en SRMR scores onder .08 en een CFI score boven .90 zijn indicatief voor een acceptabele fit (Hu & Bentler, 1999).

Het aantal benodigde respondenten voor een SEM is een onderwerp waarover nog hevig wordt gedebatteerd (Weston, 2006). Gezien de complexiteit van dit model is het lastig in te schatten hoeveel respondenten uiteindelijk nodig zijn om de correlaties en regressies voldoende te kunnen schatten. Weston geeft als indicatie 100 tot 200 respondenten voor een model met voldoende power. Het werk van Wolf, Harrington, Clark, en Miller (2013) noemt in eerste instantie ditzelfde bereik als vuistregel. In hun onderzoek simuleerden zij verschillende SEM's en kwamen zij tot de conclusie dat de veelzijdigheid van de modellen eerder een bereik van 30 tot 460 benodigde cases toont. Enkele van vele voorbeelden van invloedrijke factoren op het benodigde aantal respondenten zijn de factorladingen, het aantal indicatoren in het model, het aantal missing cases en het aantal latente variabelen. Ondanks deze lopende discussie, mag worden aangenomen dat het lage aantal respondenten van het huidige onderzoek gevolgen heeft voor de uiteindelijke conclusie.



*Figuur 2.* De veronderstelde latente modellen met de submodellen (T.1, T.2 en T.3) welke de verschillende meetmomenten representeren. Model A geeft de relatie tussen alle variabelen weer, waarbij model B de relatie beschrijft tussen de values over de drie verschillende meetmomenten. Om de weergave overzichtelijk te houden zijn de errortermen en de geobserveerde variabelen weggelaten uit deze weergave.

### 3. Resultaten

Vier cases (27, 28, 37 en 38) zijn geïdentificeerd via Mahalanobis afstand als multivariate uitschieters ( $p < .001$ ). Omdat er geen waarneembaar patroon kon worden geïdentificeerd en de multivariate uitbijterafwijkingen klein waren, zijn geen cases verwijderd uit de eerste meting. De tweede en derde meting bestonden uit te weinig cases, respectievelijk  $n = 17$  en  $n = 15$ , om een multivariate analyse op uit te voeren.

Tabel 1 geeft een overzicht van de univariate testen voor normaliteit op de eerste meting. Hieruit bleken alle variabelen, behalve de attainment-value binnen deze dataset normaal verdeeld te zijn. Aanvullende testen naar multivariate normaliteit toonden aan dat de multivariate scheefheid niet binnen de normale parameters valt ( $p = .012$ ). De multivariate platheid van de data bleek echter wel te voldoen aan de criteria voor normaliteit ( $p = .119$ ).

Tabel 2 toont de correlatiestructuur van de variabelen in de samengevoegde dataset, waarbij enkel de respondenten in voorkomen die aan alle meetmomenten hebben deelgenomen ( $n = 15$ ). Dit resultaat geeft een eerste indruk van de relaties tussen de constructen. In lijn met de theorie blijken de

kosten een significant negatieve relatie te hebben met de intrinsic-value die docenten er op na houden bij de keuze om te werken met AfL. Een andere significante samenhang lijkt te bestaan tussen de intrinsieke waarde en de attainment-value. Zo is bij alle metingen de relatie  $r = .70$  of hoger, waarbij de derde meting  $r = .80$  bereikt. Een significantie negatieve correlatie tussen de kosten en de intrinsieke values lijkt bij elk meetmoment aanwezig te zijn. Deze relatie is relatief stabiel en varieert over de verschillende metingen tussen  $r = -.52$  en  $r = -.66$ .

Tabel 1.

*Beschrijvende statistieken voor de univariate normaliteit bij de data ( $n = 36$ ) van het eerste meetmoment, 'AfL' geeft hier de keuze van docenten weer om met Assessment for Learning te willen werken.*

	M.	SD.	Shapiro-Wilk	P-waarde	Normaliteit
Expectancy	3,92	0,49	.970	.371	YES
Attainment-Value	3,99	0,50	.923	.011	NO
Utility-Value	3,98	0,61	.970	.382	YES
Intrinsic-Value	3,86	0,55	.970	.383	YES
Cost-Value	2,87	0,79	.945	.058	YES
AfL (Task Choice)	3,90	0,38	.978	.632	YES

Zo is bij alle metingen de relatie  $r = .70$  of hoger, waarbij de derde meting  $r = .80$  bereikt. Een significantie negatieve correlatie tussen de kosten en de intrinsieke values lijkt bij elk meetmoment aanwezig te zijn. Deze relatie is relatief stabiel en varieert over de verschillende metingen tussen  $r = -.52$  en  $r = -.66$ .

Het model voor de eerste meting (T.1) is weergegeven in Figuur 3, waarbij de relaties tussen de latente constructen zijn gemodelleerd. De fit van dit model was  $\chi^2(6) = 9.214$ ,  $p < .05$ , CFI = .902, RMSEA = .189, SRMR = .058. De fit voor dit en de andere drie sub-modellen zijn weergegeven in tabel 3.

Tussen de latente constructen zijn de gestandaardiseerde regressiecoëfficiënten weergegeven. Zo toont het model dat de keuze om met AfL te werken significant negatief beïnvloed wordt door de values van meting 1 met  $\beta = .70$ ,  $z = 1.989$ ,  $p < .05$ . De verwachtingen en kosten blijken in deze meting echter geen significante invloed uit te oefenen op de keuze van docenten. De theoretisch veronderstelde samenhang tussen verwachtingen en values is tevens niet significant ( $z = -.345$ ,  $p = .730$ ) bevonden in deze analyse.

Tabel 2.

*De relatie tussen Expectancy, Attainment, Utility, Intrinsic, Cost en Task choice (AFL). Waarbij voor elke variabele het respectievelijke meetmoment is opgenomen, aangegeven door het getal naast het construct.*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
1. Expectancy (1)	3.93(.40)																	
2. Attainment (1)	-0,10	4.10(.41)																
3. Utility (1)	0,21	<b>0,73**</b>	4.16(.47)															
4. Intrinsic (1)	0,14	<b>0,71**</b>	<b>0,57*</b>	3.96(.44)														
5. Cost (1)	-0,10	-0,50	<b>-0,67**</b>	-	2.83(.88)													
6. Task choice (AFL)(1)	0,3	<b>0,57*</b>	<b>0,64*</b>	0,40	-0,44	3.97(.34)												
7. Expectancy (2)	-0,16	-0,04	-0,3	0,16	0,27	-0,29	3.86(.33)											
8. Attainment (2)	-0,25	-0,03	-0,04	-0,08	-0,2	0,00	0,08	4.00(.40)										
9. Utility (2)	0,32	-0,12	-0,18	0,01	-0,02	-0,06	0,46	0,48	4.10(.35)									
10. Intrinsic (2)	-0,26	0,37	0,05	0,24	-0,08	0,18	0,46	<b>0,69**</b>	<b>0,57*</b>	3.80(.62)								
11. Cost (2)	-0,28	-0,33	-0,36	-0,12	0,41	-0,30	-0,07	-0,24	<b>-0,54*</b>	<b>-0,52*</b>	2.90(.85)							
12. Task choice (AFL) (2)	0,21	0,37	0,43	0,15	-0,08	0,32	0,30	0,18	0,42	0,36	-0,36	4.00(.28)						
13. Expectancy (3)	0,14	0,15	-0,05	0,34	0,06	-0,20	0,63*	0,06	0,56*	0,28	-0,12	0,45	4.00(.26)					
14. Attainment (3)	0,13	<b>0,64**</b>	0,34	0,50	<b>-0,52*</b>	0,50	0,03	0,14	0,27	0,31	-0,37	0,28	0,44	4.02(.47)				
15. Utility (3)	0,04	0,17	0,00	0,18	-0,09	0,01	0,18	0,25	0,22	0,07	0,14	0,2	<b>0,59*</b>	<b>0,65**</b>	4.03(.31)			
16. Intrinsic (3)	-0,08	0,21	-0,01	0,09	-0,03	-0,22	0,47	0,05	0,24	0,09	-0,08	0,34	<b>0,66**</b>	<b>0,54*</b>	<b>0,8***</b>	3.91(.53)		
17. Cost (3)	-0,17	0,03	0,13	0,08	0,09	0,11	-0,15	-0,11	-0,36	-0,08	0,33	-0,06	-0,30	-0,44	<b>-0,62*</b>	<b>-0,64**</b>	2.62(.84)	
18. Task choice (AFL) (3)	0,10	0,37	<b>0,46*</b>	<b>0,55*</b>	<b>-0,54*</b>	0,41	0,30	0,09	0,15	0,23	-0,28	0,08	0,22	0,46	0,37	0,34	-0,40	4.06(.29)

Note. De gemiddelde en standaardafwijkingen (tussen haakjes) zijn op de diagonaal weergegeven

\*  $p < .05$

\*\*  $p < .01$

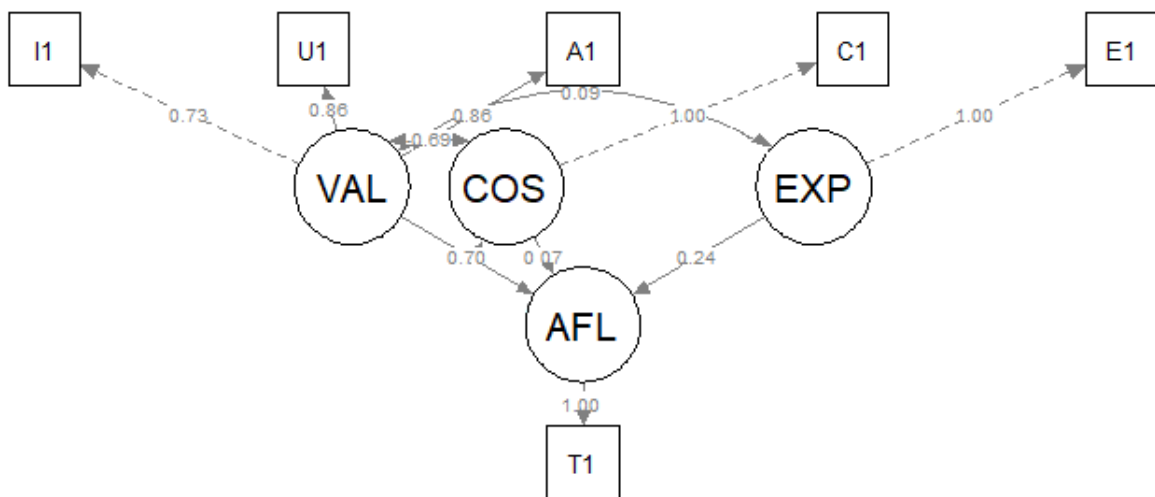
\*\*\*  $p < .001$

De modellen voor de tweede (T.2) en derde (T.3) meting laten geen significant goede fit zien met de data (Tabel 3). Daarnaast zijn de regressiecoëfficiënten tussen de latente constructen bij alle drie de modellen verschillend van elkaar en in veel gevallen niet significant. Model T.2 en T.3 geven beide lagere regressiecoëfficiënten voor de relatie tussen de values en kosten met de keuze van docenten om met AfL te werken. Waarbij model T.3 een voor geen enkele regressie tussen de latente constructen een significante relatie toont.

Tabel 3.

*De fitmaten voor de fit van modellen A en B op ieder meetmoment*

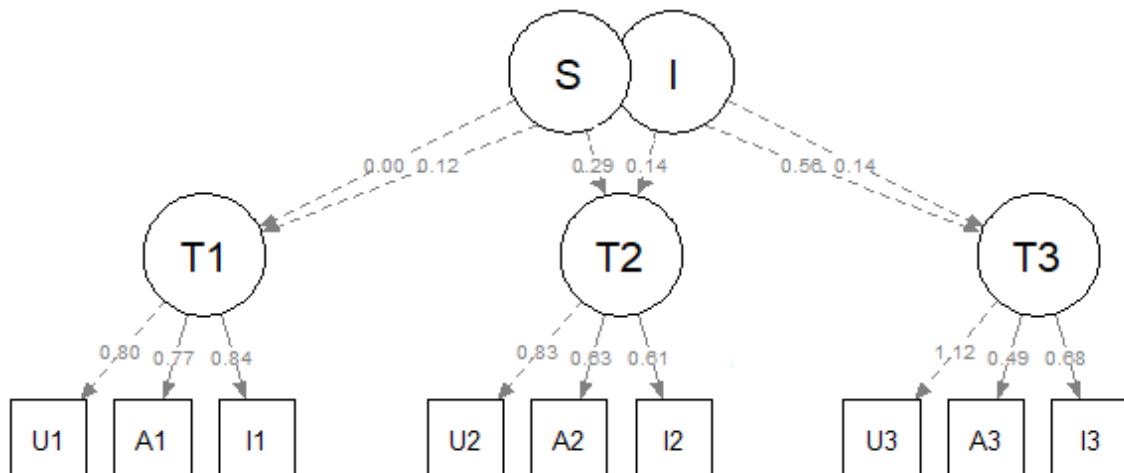
	$\chi^2$	df	<i>p</i>	CFI	RMSEA	SRMR
Model A: T.1	9.214	6	<.05	.902	.189	.058
Model A: T.2	14.636	6	<.05	.740	.274	.088
Model A: T.3	4.646	6	.843	1.000	.000	.041
Model B	52.280	30	<.05	.681	.223	.181



*Figuur 3.* De eerste (model T.1) van de drie metingen met gestandaardiseerde waarden in een Structural Equation Model, waarbij de geobserveerde variabelen zijn weergegeven als, C = Cost, I = Intrinsic, U= Utility, A= Attainment, E= Expectancy, T= Task choice (AfL). De latente constructen in het model worden weergegeven met COS = Cost , VAL = values, EXP = Expectancy en AFL = Task choice voor AfL. Om de weergave overzichtelijk te houden zijn de errortermen weggelaten uit deze weergave.

Om de verandering van de values over de drie verschillende metingen te onderzoeken is model B getest. Dit model bleek een slechte fit met de data te hebben,  $\chi^2(60) = 189.044$ ,  $p < .001$ , CFI = .170, RMSEA = .379, SRMR = .345. Gezien de slechte fit, is een alternatief model getoetst. In dit

model zijn de kosten niet opgenomen. De keuze om de kosten voor docenten in dit uiteindelijke model niet mee te nemen in de analyse, berust op de beschrijving van dit construct door Wigfield en Eccles (2002). Zij geven aan te erkennen dat de kosten niet op vergelijkbare wijze de keuze van docenten beïnvloeden zoals de andere values dat doen. Het uiteindelijke model (Figuur 4) gaf een significante ( $p < .001$ ) verbetering in fit met de data ten aanzien van het model met de cost-value.



*Figuur 4.* Het latent growth model van de drie metingen voor de values van docenten. Waarbij de geobserveerde variabelen zijn weergegeven als, A= attainment, I = Intrinsic, U= Utility. De latente variabelen zijn weergegeven als T 1, 2 en 3 = meting 1, 2 en 3, S= slopes en I= intercept. Om de weergave overzichtelijk te houden zijn de errortermen weggelaten uit deze weergave.

Ondanks significante verbetering in fit, toonde ook dit groeimodel geen goede fit,  $\chi^2(30) = 52.280$ ,  $p < .005$ , CFI = .681, RMSEA = .223, SRMR = .181. In R konden geen significanties en standaard afwijkingen worden berekend voor de relaties tussen verschillende metingen en constructen. Dit probleem wordt veroorzaakt door de eerder genoemde *non positive definite* matrix waarop het model is gebaseerd. Desondanks suggereert het latente groeimodel voor de gemiddelden, dat er geen toename in de intercept van het model plaatsvindt, de coëfficiënten blijven tussen de  $\beta = .12$  en  $\beta = .14$  hangen over de verschillende metingen. Daarentegen is wel een geleidelijke toename in de slopes zichtbaar, waarbij de coëfficiënten toenemen van de eerste meting ( $\beta = .00$ ) naar de tweede meting ( $\beta = .29$ ) en de derde meting ( $\beta = .56$ ).

#### 4. Discussie

De attitude van docenten speelt een belangrijke rol bij de succesvolle implementatie van onderwijsinnovaties (Brand-Gruwel et al., 2019). Onderzoek naar de houding van docenten in relatie tot de uitvoering van onderwijs is daarmee een belangrijke schakel tussen de onderwijswetenschappen en de praktijk. In het onderhavige onderzoek is gekeken naar de attitudes van docenten met betrekking tot het gebruik van de onderwijsbenadering Assessment for Learning vanuit een Expectancy-Value

(Wigfield & Eccles, 2002) benadering. Het doel van deze studie was om via de values (waarden) van docenten te kijken naar de motivatie van docenten om te werken met AfL. Daarnaast is ook onderzocht of deze values positief te beïnvloeden zijn met behulp van een interventie. Vanuit de Expectancy-Value theorie wordt namelijk verondersteld dat een positieve toename in de values bij mensen een positief effect heeft op de keuze die gemaakt wordt bij de uitvoering van de gerelateerde taak (Wigfield & Eccles, 2000), in dit geval AfL

De gestelde onderzoeksvraag in dit onderzoek was dan ook: “Wat is het effect van een professionaliseringsmaatregel (Webinar) op de values (waarden) die docenten erop na houden over AfL en heeft een verandering in deze values invloed op de keuze van docenten om AfL toe te passen in hun lespraktijk?”. In het algemeen ondersteunen de bevindingen uit dit onderzoek het idee dat de values van invloed zijn op de keuze van docenten om al dan niet met AfL te werken. Het is daarbij van belang om te stellen dat niet alle values op dezelfde manier invloedrijk zijn. Zo blijkt dat de cost-value mogelijk anders de keuze van docenten beïnvloedt dan de attainment-value, intrinsic-value en utility-value. Eveneens kan op basis van de resultaten gesteld worden dat mogelijk een deel van de values bij docenten positief te beïnvloeden zijn met behulp van de gebruikte interventie. Ook hier blijkt de cost-value buiten de algemene beschouwing te vallen. Het bleek voor deze value niet mogelijk om mee genomen te worden in een model, het lage aantal respondenten en de in de resultaten genoemde problemen met de correlatiematrix zijn hier mogelijk mede veroorzakers van. Deze bevindingen worden hieronder nader toegelicht, met de nadruk op de rol van de cost-value in dit onderzoek, de effecten van de interventie en de gevolgen van de kleine hoeveelheid respondenten.

#### **4.1 Values, kosten en Assessment for Learning**

De uitgevoerde analyse op de drie aparte meetmomenten toont dat de attainment-value, intrinsic-value, utility-value invloed hebben op de keuze van docenten om met AfL te werken. Dit blijkt enkel het geval te zijn wanneer de cost-value apart wordt opgenomen in het model. Wigfield en Eccles (2002) geven aan dat de keuzes die men maakt beïnvloed worden door de negatieve en positieve eigenschappen van een taak en dat elke keuze geassocieerde kosten met zich meebrengt. Dit impliceert dat de keuze van mensen afhankelijk is van de task-values (taakwaarden) en dat er een samenspel bestaat tussen de psychologische behoefte van de mensen en deze task-values. De vraag is echter in welke mate de kosten die docenten associëren met de uitvoering van een taak werkelijk passen in ditzelfde idee van psychologische behoefte. Het werk van Flake, Barron, Hulleman, McCoach, en Welsh (2015) suggereert dat de kosten eerder gezien kunnen worden als een factor die de andere values beïnvloedt, in plaats van zelf onderdeel van de values te zijn. Aanvullend laten de correlaties van het huidige onderzoek zien dat er inderdaad een negatieve relatie lijkt te bestaan tussen de kosten en de intrinsic- en attainment-value. Ook bij docenten is het dus mogelijk zo dat kosten van een taak niet zozeer hun motivatie direct beïnvloeden, maar dit doen via de andere values rond de uitvoering van AfL. In de praktijk blijkt dat docenten vooral tijdgebrek en werkdruk benoemen (James & Pedder,



2006; Kitiashvili, 2014) als invloedrijke kostenfactoren wanneer zij kiezen voor een bepaalde werkvorm. Dit huidige onderzoek suggereert dat een dergelijk gebrek aan tijd en een verhoogde werkdruk mogelijk een negatieve invloed heeft op de values van docenten, in plaats van een direct effect op de keuze om met AfL te werken. Onderzoek naar de cost-value rond AfL zou zich daarom verder kunnen verdiepen in de relatie tussen de kosten en de invloed op de attainment-value, intrinsic-value en utility-value, in plaats van de directe invloed op de keuze van docenten.

#### **4.2 De verandering van values**

De vraag of de values beïnvloed zijn door het gebruik van de interventie is niet eenduidig te beantwoorden. De drie individuele metingen tonen verschillende mate waarin de values invloed hebben op de keuze van docenten en geven daarmee een wisselend beeld. Er is sprake van verandering, maar niet consistent met de verwachting waarop dit zou gebeuren. Er werd verondersteld dat er geen verschil zou zijn in values tussen de eerste en tweede meting, toch bleek dit verschil wel aanwezig. Deze toename in values doet vermoeden dat er sprake is geweest van een natuurlijke groei of een toename door blootstelling aan het onderzoek. Dit onderzoek heeft vooraf niet geïnventariseerd in welke mate de deelnemende docenten bekend waren met de concepten rond AfL. Daarentegen zijn de concepten wel kort toegelicht in het begin van de vragenlijst en werd het belang van formatief onderwijs sterk weerspiegeld in de vragenlijst. Gezamenlijk heeft deze aanvullende toelichting en nadruk op deze onderwijsvorm mogelijk geleid tot een toename in interesse in AfL. Docenten hebben zich hierdoor bij aansluitende metingen, al dan niet bewust, kunnen verdiepen in de concepten en werking van AfL. Het is daarmee denkbaar dat de docent een verandering in de values heeft ondergaan, die niet veroorzaakt is door de interventie. Om dit te voorkomen zou vervolgonderzoek zich sterken door aanvullend te meten in welke mate docenten bekend zijn met, of zelfs actief gebruik maken van formatieve onderwijsbenaderingen zoals AfL. Hiermee kan een belangrijke nuance worden gemaakt in de verschillende vertrekpunten, waarmee het effect van de interventie zichtbaarder is.

Het idee dat het belangrijk is om inzage te hebben in de mate van bekendheid met AfL, wordt verder ondersteund door het groeimodel. Een interessante bevinding daarbij is dat dit model suggereert dat er niet direct sprake is van verandering in de initiële values, maar wel in de veranderlijkheid van de values. Omdat de werkwijze van AfL positief werd omschreven in de vragenlijstitems (e.g. “Door de toepassing van Assessment for Learning ervaar ik veel plezier in het geven van informeel, procesgericht onderwijs”), is mogelijk de zienswijze van de docent dusdanig beïnvloed dat hij of zij meer open staat voor het gebruik van deze onderwijsbenadering. Daarmee is niet gezegd dat een docent direct AfL zal toepassen in de eigen lespraktijk, maar mogelijk wel dat wanneer de invoering voldoende gefaciliteerd zou worden er meer bereidheid is ontstaan bij de docenten die aan de interventie hebben geparticipeerd. Aanvullende metingen over een langere periode zijn wenselijk om het gedrag van docenten met betrekking tot de bereidheid in het gebruik van AfL (of elementen daarvan) te meten.

De gebruikte interventie is ontworpen zodat alle values apart zijn aangesproken. Dit zorgde naar verwachting voor vergroting in effect (Acee & Weinstein, 2010). De values werden daarnaast ook specifiek in de context van AfL besproken. Ondanks deze aandachtspunten, zou de interventie mogelijk nog effectiever zijn door rekening te houden met de context waar de docenten zich in bevinden. Wigfield et al. (2017) geven namelijk aan dat de leerling-leraren relatie, leeftijdsgroep en schooltype als modererende factoren worden gezien in het onderzoek naar de effecten van values. Deze variabelen van het tweede niveau zouden beter in kaart gebracht kunnen worden door het gebruik van multilevel analyses, waarvoor grotere steekproeven zijn vereist.

#### **4.3 Het meten van values**

De dataverzameling van dit onderzoek vond plaats op het einde van het schooljaar, een moment waarop docenten hoge werkdruk ervaren. Het eerder genoemde werk van James en Pedder (2006) benoemt dat docenten onder (tijds)druk eerder geneigd zijn terug te grijpen naar bewezen en meer traditionele onderwijsbenaderingen. Gezien de beperkte hoeveelheid beschikbare tijd voor dit onderzoek, was het niet mogelijk een ander meetmoment te benutten. Naar verwachting is de interventie gedurende deze periode minder effectief gebleken. Het strekt dan ook tot aanbeveling om de populatie gedurende een langere periode te onderzoeken. Waarbij ook de interventie gedurende een rustigere periode in het schooljaar wordt ingezet, waardoor de docenten wellicht ontvankelijker zijn voor een dergelijke interventie.

Ook bij de verandering in de values is de cost-value anders werkzaam. Dit bleek doordat het groeimodel kon niet worden opgesteld wanneer de cost-value mee werd genomen in de analyse. De keuze om deze weg te laten uit het model werd daarom een gedwongen feit. Het onvoldoende aantal respondenten is naar verwachting een oorzaak van het niet functioneren van het model met de verzamelde data. Wolf et al. (2013) geven aan dat de robuustheid van een latent model in gevaar komt wanneer er te weinig data is. Een te kleine steekproef kan daarmee de verschillen van de vergelijkbare modellen verklaren. Wanneer meer respondenten worden geworven voor vergelijkbaar onderzoek, kan met meer zekerheid uitspraak gedaan worden over de relaties tussen de onderzochte variabelen. Om meer respondenten te verkrijgen zou het raadzaam zijn om de gebruikte vragenlijst te herzien en het aantal items te verminderen. Een deel van de respondenten heeft de vragenlijst niet volledig ingevuld. Daarnaast werd tijdens de vragenlijstevaluatie met de docentenfocushoofdgroep aangegeven dat zij de vragenlijst lang vonden. Om de vragenlijst te verkleinen en zoveel mogelijk van de psychometrische kwaliteiten te behouden, wordt aangeraden om naar lage factorladingen te kijken van de originele lijsten en waar mogelijk vragen te verwijderen. Hierbij moet rekening gehouden worden met de validiteit van de vragenlijsten en dient de achterliggende theorie meegenomen te worden in de uiteindelijke besluitvorming.

#### 4.4 Implicaties

De resultaten van dit onderzoek geven waardevolle inzichten in de relaties tussen de values, kosten en de keuze van docenten om met AfL te werken. Belangrijke inzichten zijn dan ook dat de kosten mogelijk indirect de values van docenten beïnvloeden, in plaats van direct de keuze van docenten te beïnvloeden. Verder onderzoek naar de values van docenten omtrent hun onderwijs is van belang, omdat deze mogelijk een belangrijke factor zijn in de wijze waarop docenten omgaan met onderwijsinnovaties en de toepassing hiervan. Zeker ook gezien er behoefte is aan innovatief onderwijs om te kunnen voldoen aan de toenemende diversiteit van leerlingen (Oolbekkink-Marchan, 2018). Values zijn daarbij weliswaar een belangrijk onderdeel, maar vormen geenszins het gehele beeld rond de motivatievorming van docenten. De aangrenzende ‘expectancies’ (verwachtingen) hebben daarbij immers ook een rol en dit construct dient dan ook meegenomen te worden bij verder onderzoek dat het geheel van de Expectancy-Value theorie en de relatie met AFL-keuze tracht te onderzoeken.

Daarnaast is een eerste stap gedaan richting de ontwikkeling en het testen van een interventie zodat de values rond AfL positief beïnvloed kunnen worden. Gezien de beschreven limitaties rond het aantal respondenten en de inzichten rond de werking van de cost-value moet dit onderzoek gezien worden als een voorlopige rapportage en wordt verder onderzoek met meer respondenten, langere metingen en een controle groep aanbevolen. Daarmee kan hopelijk een nog duidelijk beeld worden geschetst van de onderlinge relaties tussen de values en hun effect op de keuze van docenten, dit is noodzakelijk mede omdat het eerder genoemde werk van Brand-Gruwel et al. (2019) aangeeft aan dat de invoering van onderwijsinnovatie minder tot niet succesvol is zonder de docent als belangrijkste uitvoerder hiervan. Daarmee is dit en ander onderzoek naar de waarden die docenten hechten aan de manier waarop zij hun onderwijs vormgeven waardevol voor het huidige en toekomstige onderwijs.

## Referenties

- Acee, T. W., & Weinstein, C. E. (2010). Effects of a Value-Reappraisal Intervention on Statistics Students' Motivation and Performance. *The Journal of Experimental Education*, 78(4), 487-512. doi:10.1080/00220970903352753
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117-140. doi:10.1177/1478210314566733
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. doi:10.1080/0969595980050102
- Black, P., & Wiliam, D. (2010). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan*, 92(1), 81-90. doi:10.1177/003172171009200119
- Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4-16. doi:10.2307/1175554
- Brand-Gruwel, S., Bos, N. & van der Graaf, A., (2019). Het vergroten van studiesucces in het hoger onderwijs: belang van overtuigingen van docenten. *Pedagogische Studiën*, 96(1), 1-14. doi:10.1017/1478210319000001
- Brown, S. (2004). Assessment for Learning. *Learning and Teaching in Higher Education* (1), 81-89.
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258. doi:10.1177/0049124192021002005
- Eccles, J. (1983). *Expectancies, Values, and Academic Behaviors*. San Francisco, CA: Freeman.
- Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, 41, 232-244. doi:10.1016/j.cedpsych.2015.03.002
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Gulikers, J., & Baartman, L. (2017). *Doelgericht professionaliseren: formatieve toetspraktijken met effect! Wat DOET de docent in de klas? (405-15-722)*. Geraadpleegd van Nationaal Regieorgaan Onderzoek website: [https://www.nro.nl/wp-content/uploads/2015/09/Inhoudelijke-eindrapport\\_NRO-PPO-405-15-722\\_DEF.pdf](https://www.nro.nl/wp-content/uploads/2015/09/Inhoudelijke-eindrapport_NRO-PPO-405-15-722_DEF.pdf)
- Harlen, W., & Crick, R. (2003). Testing and Motivation for Learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169-207. doi:10.1080/0969594032000121270
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi:10.1080/10705519909540118
- James, M., & Pedder, D. (2006). Beyond method: assessment and learning practices and values. *Curriculum Journal*, 17(2), 109-138. doi:10.1080/09585170600792712

- Kitiashvili, A. (2014). Teachers' attitudes toward assessment of student learning and teacher assessment practices in general educational institutions: The case of Georgia. *Improving Schools*, 17(2), 163-175. doi:10.1177/1365480214534543
- Kohn, A. (2011, November). The Case Against Grades [Blog post]. Geraadpleegd van <https://www.alfiekohn.org/article/case-grades/>
- MacLellan, E. (2010). Assessment for Learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318. doi:10.1080/02602930120063466
- Mayer, R. E. (2009). *Multimedia Learning*. New York, NY: Cambridge University Press.
- McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Res Q Exerc Sport*, 60(1), 48-58. doi:10.1080/02701367.1989.10607413
- Neesom, A. (2000). *Report on Teachers' Perception of Formative Assessment*. London: Qualification and Curriculum Authority: QCA's Assessment for Learning Project.
- Oolbekkink-Marchan, H. (2018). *Leraren veranderen*. Geraadpleegd van de Hogeschool van Arnhem en Nijmegen website: [https://www.han.nl/onderzoek/kennismaken/kenniscentrum-kwaliteit-van-leren/omdat-leren-verandert/leraren-veranderen/\\_attachments/helma\\_oolbekkink\\_web.pdf](https://www.han.nl/onderzoek/kennismaken/kenniscentrum-kwaliteit-van-leren/omdat-leren-verandert/leraren-veranderen/_attachments/helma_oolbekkink_web.pdf)
- Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of Assessment for Learning Questionnaires for teachers and students. *British Journal of Educational Psychology*, 83(Pt 1), 98-113. doi:10.1111/j.2044-8279.2011.02057.x
- Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103(3), 683-700. doi:10.1037/a0023911
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78.
- Segers, M. S. R. (2004). *Assessment en leren als een twee-eenheid: Onderzoek naar de impact van assessment op leren*. Geraadpleegd van Universiteit Leiden website: <https://openaccess.leidenuniv.nl/bitstream/handle/1887/5367/OR170.pdf?sequence=1>
- The National Council of Teachers of Mathematics. (2007). *Five "Key Strategies" for Effective Formative Assessment* (Research Brief 04). Reston, VA: William, D.
- Tschannen-Moran, M., & Hoy, A. (2001). Teacher Efficacy: Capturing an Elusive Construct. *Teaching and Teacher Education*, 17(7), 783-805. doi:10.1016/S0742-051X(01)00036-1
- Watt, H. M. G., & Richardson, P. W. (2007). Motivational Factors Influencing Teaching as a Career Choice: Development and Validation of the FIT-Choice Scale. *The Journal of Experimental Education*, 75(3), 167-202. doi:10.3200/JEXE.75.3.167-202
- Weston, R. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist*, 34(5). doi: 10.1177/0011000006286345

- Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25(1), 68-81. doi: 10.1006/ceps.1999.1015
- Wigfield, A., & Eccles, J. S. (2002). Motivational Beliefs, Values and Goals. *Annual Review of Psychology*, 53. doi:10.1146/annurev.psych.53.100901.135153
- Wigfield, A., Rosenzweig, E., & Eccles, J. (2017). *Achievement Values*. New York, NY: Guilford Publications.
- Wiliam, D. (2010). *Assessment for Learning: why, what and how* [PowerPoint]. Geraadpleegd van [https://www.dylanwiliam.org/Dylan\\_Wiliams\\_website/Presentations\\_files/Oslo.pptx](https://www.dylanwiliam.org/Dylan_Wiliams_website/Presentations_files/Oslo.pptx)
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14. doi:10.1016/j.stueduc.2011.03.001
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and psychological measurement*, 76(6), 913-934. doi:10.1177/0013164413495237

## Bijlage A

## Teacher Expectancy-Value and Assessment for Learning Questionnaire

5-punts Likertschaal welke loopt van 1= zeer mee oneens, tot 5= zeer mee eens.

**Expectancy****Items**

“Ik verwacht ...

Exp_1	Te kunnen reageren op moeilijke vragen van mijn leerlingen
Exp_2	in te kunnen schatten of mijn leerlingen begrepen hebben wat mijn leerlingen hebben geleerd
Exp_3	vragen te kunnen bedenken voor mijn leerlingen waaruit blijkt wat ze hebben geleerd
Exp_4	mijn lessen aan te kunnen passen aan het individuele niveau van mijn leerlingen
Exp_5	Op verschillende manieren de voortgang van mijn leerlingen te kunnen evalueren
Exp_6	andere uitleg te kunnen geven wanneer mijn leerlingen in de war raken door de stof
Exp_7	geschikte uitdagingen te bieden aan zeer bekwame studenten
Exp_8	Het leerproces van de leerling gericht te kunnen aansturen

**Attainment****Items**

Att_1	door het gebruik van Assessment for Learning in de les word ik een goede docent
Att_2	ik ben een goede docent als ik mijn leerlingen betrek bij het stellen van persoonlijke leerdoelen
Att_3	ik ben een goede docent wanneer ik op de hoogte ben van de voortgang van mijn leerlingen
Att_4	ik geef goed les zodra ik mijn instructie aanpas aan de behoefte van de leerling

**Utility****Items**

“De toepassing van Assessment for learning heeft voor mij als meerwaarde dat...”

Uti_1	er sprake is van een toename in de kwaliteit van mijn lesgeven
Uti_2	onderwijs een gedeelde verantwoordelijkheid wordt tussen mijzelf en mijn leerlingen
Uti_3	de feedback over het onderwijs direct en passend is voor mijn leerlingen
Uti_4	ik nauw kan samenwerken met mijn leerlingen aan hun voortgang
Uti_5	ik nauw kan samenwerken met mijn collega's aan ons onderwijs
Uti_6	mijn leerlingen zelfvertrouwen krijgen in hun leerproces
Uti_7	mijn leerlingen gemotiveerd raken voor school
Uti_8	mijn leerlingen een positieve houding hebben ten aanzien van het onderwijs dat ik geef
Uti_9	Ik beter word in het begeleiden van leerlingen

**Intrinsic****Items**

“Door de toepassing van Assessment for Learning...”

Intr_1	ervaar ik veel plezier in het geven van informeel, procesgericht onderwijs
Intr_2	vind ik mijn lesgeven leuk
Intr_3	haal ik plezier uit het volgen van het onderwijsproces van mijn leerlingen
Intr_4_R	vind ik mijn lesgeven saai
Intr_5	ontwikkel ik interesse voor procesgericht onderwijs

<b>Cost</b>	<b>Items</b>
	“Door de toepassing van Assessment for Learning...”
Cost_1	ben ik te veel tijd kwijt aan de voorbereiding en planning van mijn onderwijs
Cost_2	neem ik het risico dat de kwaliteit van mijn onderwijs zal leiden
Cost_3	raak ik gestrest
Cost_4	kom ik in de problemen doordat mij dit extra werkdruk oplevert
Cost_5	ben ik bang dat mijn andere werkzaamheden op de achtergrond raken
Cost_6	moet ik meer moeite doen om een les te geven dan normaal
Cost_7	kom ik niet toe aan de behandeling van alle verplichte lesstof

<b>Achievement choice</b>	<b>Items</b>
	“Ik kies ervoor om...”
Achi_1	mijn leerlingen te stimuleren na te denken over hoe zij hun schoolwerk kunnen verbeteren
Achi_2	na het nakijken van een toets met iedere leerling zijn of haar antwoorden te bespreken tijdens het maken van het schoolwerk aan leerlingen te vragen hoe ze vinden dat het tot nu toe gaat op school
Achi_3	met mijn leerlingen mee te denken over de manier waarop ze willen leren op school.
Achi_4	leerlingen de gelegenheid te geven zelf hun leerpunten te bepalen
Achi_5	mijn leerlingen te vragen om aan te geven wat ze goed en minder goed hebben gedaan in hun schoolwerk.
Achi_6	leerlingen te stimuleren om terug te kijken op hun leerproces en te bedenken wat ze een volgende keer anders kunnen doen.
Achi_7	mijn leerlingen te laten weten wat hun sterke punten zijn op het gebied van leren.
Achi_8	mijn leerlingen te laten weten wat hun zwakke punten zijn op het gebied van leren.
Achi_9	mijn leerlingen aan te moedigen om hun leerproces te verbeteren.
Achi_10	mijn leerlingen aanwijzingen te geven die hen helpen bij het leren.
Achi_11	het gemaakte werk met mijn leerlingen te bespreken, zodat zij de lesstof beter begrijpen.
Achi_12	met mijn leerlingen hun vorderingen te bespreken.
Achi_13	na een toets-moment mijn leerlingen te vertellen hoe zij hun zwakke prestaties kunnen verbeteren.
Achi_14	met mijn leerlingen te bespreken hoe zij hun sterke kanten kunnen gebruiken om hun werk te verbeteren.
Achi_15	samen met mijn leerlingen een strategie vast te stellen om hun zwakke punten te verbeteren
Achi_16	
Achi_17	mijn instructie aan te passen als ik merk dat leerlingen een onderdeel niet begrijpen.
Achi_18	de leerlingen aanwijzingen te geven om hen te helpen de lesstof te begrijpen.
Achi_19	tijdens de les de leerlingen laten kunnen zien wat ze hebben geleerd
Achi_20	mijn vragen op een begrijpelijke manier te stellen
Achi_21	door vragen te stellen tijdens de les de leerlingen te helpen om de lesstof begrijpen.
Achi_22	open te staan voor de inbreng van mijn leerlingen in de klas.
Achi_23	met mijn leerlingen in discussie over de antwoorden te gaan.
Achi_24	ervoor te zorgen dat de leerling weet aan welke punten hij of zij moet werken om zijn of haar resultaten te verbeteren.
Achi_25	leerlingen de mogelijkheid te geven om vragen te stellen.
Achi_26	ervoor te zorgen dat de leerling weet aan welke eisen zijn of haar werk moet voldoen.
Achi_27	ervoor te zorgen dat de leerlingen weten wat ze kunnen leren van een opdracht.
Achi_28	te controleren of de leerling het leerdoel heeft behaald