

Attitude, kwaliteit van peerfeedback en leereffect na peerassessmenttraining over stijl bij basisschoolleerlingen

Inge W.A. Hooijer, 3656837

Mara P. Rooijmans, 3446913

Daan W.M. Verhees, 3450198

Begeleider: Frans J. Prins

Utrecht, juni 2011

Universiteit Utrecht

Abstract: Peerfeedback is feedback die gegeven wordt door gelijkgestemden. In het basisonderwijs is er nog weinig onderzoek verricht naar peerfeedback. In dit artikel wordt getracht antwoord te krijgen op de vraag of een training over feedbackstijlen invloed heeft op de peerfeedback die basisschoolkinderen elkaar geven.

De effecten van de training zijn gemeten op (1) de kwaliteit van de peerfeedback, (2) de attitude jegens peerfeedback en (3) op het uiteindelijke leereffect van de schrijfpdracht. Het is een schrijfpdracht voor groep 7 en 8 van BS Meuleveld uit Horst geweest, waar 115 leerlingen, tussen de 10 en 12 jaar, aan deelgenomen hebben.

Uit het onderzoek is naar voren gekomen dat peerfeedback zorgt voor een beter leerresultaat bij leerlingen van de basisschool en dat bij groep 7 de training gezorgd heeft voor een beter gebruik van de exploratieve feedbackstijl, dit verschil is ook significant, $t(46) = -2.45$, $p = .02$. Hieruit kan geconcludeerd worden dat peerfeedback een positief effect heeft op basisschoolleerlingen om van de peerfeedback te leren en dat een training over feedbackstijlen invloed heeft op de kwaliteit van de peerfeedback.

Sleutelbegrippen: peerfeedback, attitude feedback, feedbackstijl, leereffect van peerfeedback, peerassessmenttraining, kwaliteit peerfeedback.

Introductie

Het geven van feedback en de laatste jaren ook peerfeedback zijn belangrijke factoren in het onderwijs. Feedback is informatie geven over iemands houding of product door een ander persoon en deze is bedoeld om studenten te helpen hun eigen sterke en zwakke punten te identificeren (Hattie & Timperly, 2007; Topping, 1998). Er zijn twee hoofdvormen van feedback te onderscheiden namelijk formatief en summatief. Formatieve feedback is de kennis die tussentijds aan studenten wordt gegeven om het gat tussen de huidige en de gewenste situatie te verminderen (Black en William, 1998). De meest gebruikte manier in het onderwijs is feedback door docenten, aangezien zij meer kennis hebben van de criteria en standaarden (Sadler, 1989), de prestaties van studenten beoordelen en hun leerproces evalueren (Silverman, Tyson & Kampitz, 1992).

Een andere manier is het gebruiken van peerfeedback, waarbij leerlingen elkaars prestatie beoordelen zodat beiden ervan leren en het product beter wordt (Van den Berg, Pilot & Admiraal, 2005). Peerfeedback draagt bij aan het nemen van verantwoordelijkheid voor het eigen leren, daarnaast is er meer betrokkenheid bij het leerproces en vergroot deze vorm van feedback de aandacht voor de onderwijshoud (Lindsay & Clarke, 2001). Het geven en ontvangen van peerfeedback beïnvloedt het gedrag en de attitude van leerlingen in het primair onderwijs positief (Briggs, Woodfield, Martin & Swatton, 2008) en doordat leerlingen van elkaar feedback ontvangen, bereiken de leerlingen een hoger niveau van zelfkennis en inzicht over de impact van een uitgevoerde prestatie (Berihun, Tesera & Desta, 2006).

Peerfeedback is vooral nuttig als deze van goede kwaliteit is (Leenknecht, 2011). Stijl vormt een onderdeel van de kwaliteit van de feedback. Er is een verdeling in vier feedbackstijlen, de autoritaire, interpreterende, exploratieve en collaboratieve stijl (Van den Berg, 2003). Deze feedbackstijlen kunnen gelinkt worden aan de manieren van commentaar geven, de eerste twee aan de gecontroleerde en leidende manier, de laatste twee stijlen aan de faciliterende manier (Straub, 1996).

Wanneer bij de feedback het oordeel benadrukt wordt en er bijna geen verklaring wordt gegeven, is de stijl autoritair. Bij zowel de interpreterende als de exploratieve stijl worden verklaringen en suggesties voor herziening gegeven van het product, maar iemand met een interpreterende stijl blijft bij zijn eigen ervaringen en ideeën. Iemand met een exploratieve stijl neemt het perspectief van de student over en geeft verklaringen en suggesties die corresponderen met de intentie van de student. Wanneer een collaboratieve stijl is gebruikt, overleggen de gever en de ontvanger over de prestatie en creëren ze een collaboratief product (Van de Berg, 2003).

Het volgen van een training omtrent peerassessment heeft bij basisschoolleerlingen een positief effect op de kwaliteit van de gegeven feedback (Meusen-Beekman & Joosten-Ten Brinke, 2010). Leerlingen die de training niet hebben gevolgd, blijven bij hun beoordelingen steken bij opmerkingen zoals 'de spreekbeurt was leuk' of 'super, 10+'. Bij de leerlingen die de training wel hebben gevolgd, worden de criteria die uitgewerkt zijn in de training expliciet benoemd, onderbouwd en geven de leerlingen objectief een uitspraak over het niveau van de presentatie. Het onderzoek van Meusen-Beekman en Joosten-Ten Brinke (2010) heeft zich niet gericht op de perceptie van feedback. De perceptie jegens peerfeedback in het hoger onderwijs zijn vooral de redenen van een veranderde houding ten opzichte van peerfeedback na het volgen van een training (Cheng & Warren, 1997). Beide onderzoeken vormen een interessant uitgangspunt voor nieuw onderzoek naar peerfeedback en de perceptie hiervan met verdere verdieping.

De criteria voor de spreekbeurt zijn in het onderzoek van Meusen-Beekman en Joosten-Ten Brinke (2010) expliciet benoemd en onderbouwd en de leerlingen geven objectief een uitspraak over het niveau van de prestatie. Tevens worden er verbeterpunten gegeven. Het geven en ontvangen van feedback kan gerelateerd worden aan de facetten criteria/kwaliteit, stijl en de voorkeuren van de gever en ontvanger (Prins, Sluijsmans & Kirschner, 2006).

Wanneer bovenstaande informatie wordt samengevat zou de volgende vraag geformuleerd kunnen worden: “Is de kwaliteit van de feedback van de groep leerlingen die een training gericht op feedbackstijl heeft gevolgd, beter dan die groep van leerlingen die geen training heeft gevolgd?”

De training zal gericht worden op de exploratieve stijl omdat deze samen met de collaboratieve stijl als meest effectief wordt gezien (Straub, 1996). Aangezien de peerfeedback anoniem wordt gegeven, kan er geen gebruik worden gemaakt van de collaboratieve stijl. De verwachting is dat een training over de verschillende soorten stijlen van feedback geven een effect heeft op de kwaliteit en de perceptie van peerfeedback.

De perceptie van peerfeedback wordt in dit onderzoek aangeduid als de attitude ten opzichte van peerfeedback. Bij studenten uit voornamelijk het hoger onderwijs is de attitude ten opzichte van peerfeedback al vaker onderzocht. Daar zijn verschillende resultaten uit gekomen, zowel positief als negatief. Studenten genieten bijvoorbeeld van het uitvoeren van de peerassessment en zien dit als voordelig voor hun leren (Orsmond & Merry, 1996). Daarnaast is peerassessment een waardering voor intern prijzen en bevordert het de inter-persoonlijke relaties in de klas (Keaten & Richardson, 1992). Studies laten ook een gebrek aan zelfvertrouwen zien wanneer studenten elkaar moeten beoordelen (Sullivan, Hitchcock & Dunnington, 1999) en zij hebben behoefte aan een reeds bestaande richtlijn of regel voor de assessmentactiviteit (Orsmond & Merry, 1996). Cheng en Warren (1997) hebben gezien dat studenten over het algemeen voorstander zijn van peerassessment, terwijl minder dan de helft van hen denkt dat eerstejaars studenten in staat zijn om peerassessment op een eerlijke en verantwoordelijke manier uit te voeren. In eerste instantie zijn de studenten niet helemaal overtuigd geweest van hun eigen kunnen om medestudenten te beoordelen. Na de oefening is er een positieve verschuiving in zowel de houding als het vertrouwen. Om te kijken of dit ook voor het primair onderwijs geldt, zal de attitude gemeten worden op de aspecten affectie, zelfvertrouwen en inter-persoonlijke relaties jegens peerfeedback.

Zal een kwalitatief betere feedback en positievere attitude jegens peerfeedback een leereffect opleveren? Leerlingen hebben door peerfeedback een grotere invloed op het leerproces en de resultaten daarvan (Lindsay en Clarke, 2001), daarnaast zorgt peerfeedback voor ondersteuning van het leerproces door een tussentijdse controle van de prestatie in vergelijking met criteria, vergezeld door feedback over sterktes, zwaktes en/of tips voor verbetering (Falchikov, 1996). Er kunnen ook leervoordelen zijn voor de feedbackgever doordat hij andere voorbeelden of benaderingen ziet en door het eigen maken van de criteria en standaarden (Topping, 1998).

Ook studenten die in het hoger onderwijs peerfeedback toepassen, verbeteren daarbij eigen inzichten en prestaties (Hargreaves, 1997). Peerassessments kunnen dus de kwaliteit van een prestatie bevorderen. Uit onderzoek van Meusen-Beekman en Joosten-Ten Brinke (2010) blijkt dat deze conclusie voor het geven van een spreekbeurt in het basisonderwijs nog niet getrokken kan worden. Klassikale spreekbeurten kunnen een vertekend beeld geven van het precieze leereffect, aangezien leerlingen die later hun spreekbeurt hebben gehouden, reeds hebben kunnen leren van eerder gegeven spreekbeurten.

Om de kwaliteit van feedback te verbeteren, zijn er vaardigheden nodig bij de feedbackgever en feedbackontvanger (Sluijsmans, 2002; Leenknecht, 2011). De verwachting is doordat de kwaliteit van de feedback beter is, de leerlingen peerfeedback positiever beoordelen en het als nuttiger ervaren. Daar er nog weinig onderzoek is gedaan naar peerfeedback in het primair onderwijs, zal deze studie zich richten op de perceptie van peerfeedback in het primair onderwijs, waarbij het effect van een training gericht op de stijl van feedback het startpunt zal zijn.

Sluijsmans (2002) onderscheidt de volgende drie vaardigheden die nodig zijn bij peerfeedback geven:

1. Definiëren van de feedback criteria;
2. Beoordelen van de prestatie van een peer;
3. Het geven van feedback.

Er blijkt een intensieve training nodig te zijn voor het eigen maken van deze vaardigheden (Sluijsmans, 2002; Prins, Sluijsmans & Kirschner, 2006). Aangezien leerlingen in het primair onderwijs nauwelijks gewend zijn om prestaties van medeleerlingen van feedback te voorzien (Meusen-Beekman & Joosten-Ten Brinke, 2010), is het belangrijk dat zij de rol van feedbackgever en -nemer kunnen oefenen.

Sluijsmans (2002) geeft in haar proefschrift aandachtspunten bij het vormgeven van een peerassessmenttraining, waarbij het uitgangspunt ligt op het leereffect hiervan. Deze uitgangspunten zijn onder te verdelen in de volgende procedure van 4 stappen (Stiggings, 1987) en zijn ook terug te vinden in de training zoals bij de procedure is omschreven:

1. Vaststellen van het doel van peerassessment: leerlingen leren elkaars prestatie te beoordelen;
2. Kiezen van een format voor een peerassessment: de vaardigheden die beoordeeld worden, moeten geïdentificeerd worden;
3. Formuleren van criteria gebaseerd op de te beoordelen vaardigheden: met behulp van de beoordelingscriteria wordt vastgesteld welke eisen gesteld worden aan de prestatie;
4. Creëren van een peerassessmentbeoordeling: de leerlingen beoordelen elkaars prestatie op basis van vooraf vastgestelde criteria.

Verder geven Meusen-Beekman en Joosten-Ten Brinke (2010) aan dat het belangrijk is om een herhalingsles te geven aangezien de aangeleerde peerassessmentvaardigheden afnemen na verloop van tijd. Als de vaardigheden niet herhaald worden, dan laat het resultaat zien dat de leerlingen weer terugvallen in hun oude gewoonten van oppervlakkige beoordelingen geven. Dit sluit aan bij de conclusie van Sluijsmans (2002) dat leerlingen na een enkele training in peerassessment niet als expert beschouwd mogen worden, maar dat herhaling van training noodzakelijk is.

Vraagstellingen en hypothesen

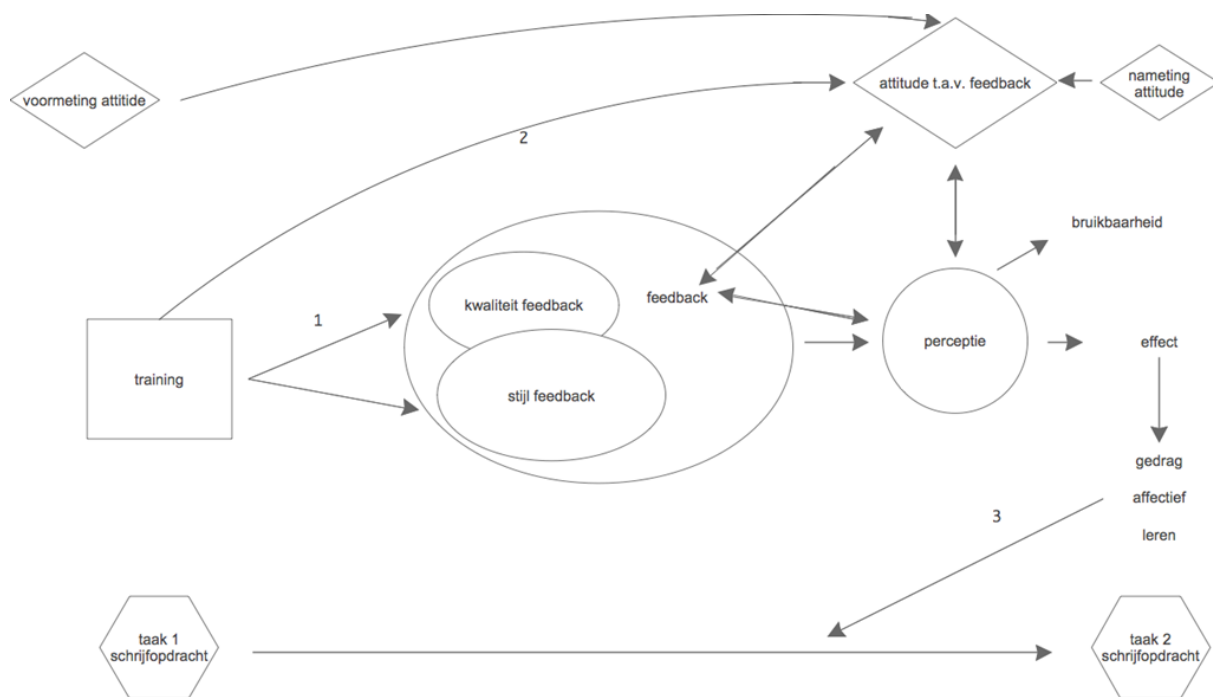
De stijl en kwaliteit van feedback, perceptie in de vorm van attitude en het leereffect staan centraal in dit onderzoek, de onderzoeksvragen die uit de introductie voortvloeien luiden als volgt:

1. Heeft het volgen van de peerassessmenttraining, gericht op de exploratieve feedbackstijl, een positief effect op de kwaliteit van de gegeven peerfeedback bij bovenbouwleerlingen van het primair onderwijs?;
2. Heeft het volgen van de peerassessmenttraining, gericht op de feedbackstijl, een positief effect op de attitude ten opzichte van peerfeedback bij bovenbouwleerlingen van het primair onderwijs?;
3. Heeft het volgen van de peerassessmenttraining, gericht op de feedbackstijl, een positief leereffect op het schrijven van een stelopdracht door bovenbouwleerlingen van het primair onderwijs?

Na analyse van de literatuur zijn de volgende verwachtingen ontstaan waarbij de hypothesen geformuleerd kunnen worden;

- De kwaliteit van de feedback van de groep leerlingen die de training hebben gevolgd, zal beter zijn dan die van de leerlingen die geen training hebben gevolgd;
- Leerlingen die de training hebben gevolgd omtrent peerassessment krijgen een positievere attitude ten opzichte van peerfeedback dan leerlingen die geen training hebben gevolgd;
- Hoe beter de gegeven feedback is, hoe positiever en nuttiger de leerlingen peerfeedback zien en hoe meer zij er uiteindelijk mee kunnen. Het gedrag zal uiteindelijk resulteren in een verbeterde stelopdracht.

In figuur 1 worden de onderzoeksvragen, zoals onderbouwd, weergegeven. De genummerde pijlen zijn gerelateerd aan de onderzoeksvragen. Pijl 1 geeft weer wat het effect van de training is op de feedbackkwaliteit, pijl 2 geeft weer dat de attitude verandert door de training. Pijl 3 verloopt via pijl 1 en 2, door een betere feedbackkwaliteit en een positievere attitude zal er een positievere perceptie volgen en dus een groter leereffect zijn.



Figuur 1. Schematische weergaven Onderzoeksvragen

Methode

Deelnemers

Participanten van dit onderzoek zijn 115 leerlingen van groep 7 en 8 van basisschool Meuleveld te Horst. Er hebben 57 meisjes en 58 jongens meegedaan in de leeftijd van 10 tot 12 jaar. Er is voor deze doelgroep gekozen omdat zij zichzelf op een adequate manier uit kunnen drukken. Alle kinderen hebben nog geen ervaring met het geven en ontvangen van peerfeedback.

Voor het onderzoek is er gebruik gemaakt van de bestaande groepen, waarbij de experimentele groepen random zijn bepaald. Groep 7a en 8b zijn de experimentele groepen die de training hebben gevolgd, groep 7b en 8a zijn de controlegroepen.

Design en procedure

Met dit onderzoek zijn op de 3 gebieden attitude, kwaliteit gegeven feedback en leereffect gegevens verzameld. Allereerst is er, door middel van een voormeting, aan zowel de experimentele als de controlegroepen gevraagd wat hun attitude is ten opzichte van peerfeedback. Door middel van deze vragenlijst is het referentiekader van de leerlingen bepaald. Ten tweede is er na deze voormeting data verzameld omtrent de kwaliteit van de gegeven feedback. Er heeft een interventie in de vorm van een training met 4 lessen plaatsgevonden overeenkomstig aan de 3 vaardigheden welke Slujsmans (2002) onderscheidt en de 4 stappen die Stiggings (1987) gebruikt.

Bij alle lessen is er rekening gehouden met het feit dat de leerlingen nog geen ervaring hebben met peerfeedback. Om het effect van de interventie te meten, hebben de kinderen een stelopdracht gemaakt. Groep 7 en 8 hebben verschillende onderwerpen qua opdracht gehad in verband met de leeftijd van de doelgroep. De opdracht van groep 7 is gericht op de manier van werken van groep 7 en groep 8, waarbij de leerlingen de leerkrachten van groep 8 moeten overtuigen om bepaalde dingen anders te doen als zij dat wensen. De opdracht van groep 8 is gericht op het opereren als

directeur van Meuleveld voor één dag, waarbij de kinderen moeten aangeven welke ideeën zij hebben en hoe zij dingen anders zullen aanpakken.

De eerste les hebben alle groepen gevolgd. De leerlingen hebben een mindmap gemaakt over het onderwerp van de schrijfpdracht, waarbij er door de leerlingen is gediscussieerd aan welke criteria de opdracht moet voldoen, wil deze als goed beoordeeld worden. Op deze manier zijn de feedbackcriteria gedefinieerd, zodat het voor de leerlingen duidelijk was waar zij op moesten letten bij het geven van de feedback. Deze les had een tijdsbestek van 30 minuten. Na deze les hebben alle leerlingen de opdracht geschreven.

Na de eerste les hebben alleen de experimentele groepen een tweede en een derde les gevolgd, waarbij de nadruk is gelegd op de exploratieve stijl en het bezitten van de vereiste kennis en vaardigheden om elkaars prestatie te kunnen beoordelen en adequaat feedback te kunnen geven.

Bij de tweede les kregen de leerlingen enkele films (Rooijmans, 2011) te zien van de verschillende feedbackstijlen (Van den Berg, 2003) waarop zij hun reacties gaven. Door middel van een leergesprek zijn de voordelen van exploratieve feedback besproken en zijn er, samen met de kinderen, aspecten opgesteld waar goede peerfeedback aan moet voldoen. Hieruit is gekomen dat de feedback zich moet richten op wat de schrijver wil zeggen en dat niet de interesses van de feedbackgever belangrijk zijn. Tevens hebben de leerlingen aangegeven dat er een toelichting gegeven moet worden in de feedback, die de verbeteringen aangeeft en zich richt op de inhoud.

In de derde les hebben de leerlingen van de experimentele groep geoefend met de rol van feedbackgever en -ontvanger door middel van dramaspel. Voorbeeld van een spelsituatie is: “ Je vriendin gaat naar de kapper toe, want jij hebt tegen haar gezegd dat een kort kapsel leuker is. Jouw vriendin komt terug van de kapper en je vindt het helemaal niet mooi! Hoe ga je reageren?” Voor het meer inzichtelijk maken van de opgestelde factoren waaraan goede feedback moet voldoen, zijn deze na het spelen van elke situatie met de kinderen doorgesproken. De tweede en de derde les hebben ieder een tijdsbestek van 40 minuten gehad.

De afsluitende vierde les hebben wederom alle groepen gevolgd, welke gefungeerd heeft als herhalingsles. Na afloop van deze les hebben alle leerlingen anoniem op een leeg blad feedback gegeven op de opdracht van een klasgenoot. Na het ontvangen van de peerfeedback hebben de leerlingen de mogelijkheid gehad om deze te verwerken en hun opdracht te verbeteren.

Ten derde is er data verzameld met betrekking tot het leereffect. De stelopdracht is, voor het geven van de peerfeedback, zonder medeweten van de leerlingen beoordeeld met een cijfer door de leerkrachten. Na het verwerken en eventueel verbeteren van de opdracht aan de hand van de ontvangen peerfeedback is de opdracht wederom beoordeeld door de leerkrachten.

Als laatste heeft er een nameting omtrent attitude plaatsgevonden bij alle leerlingen, na het verwerken van de ontvangen peerfeedback. Hiervoor is gebruik gemaakt van dezelfde vragenlijst als bij de voormeting.

Instrumenten

Beoordelingsformulier kwaliteit peerfeedback

Het beoordelingsformulier dat gebruikt is voor het beoordelen van de kwaliteit van de gegeven feedback op het gebruik van criteria en explorerende stijl, is vormgegeven aan de hand van de aspecten die volgens de leerlingen belangrijk zijn. Het beoordelingsformulier kent een indeling in twee gebieden, namelijk criteriagebruik en explorerende stijl.

Criteriagebruik analyseert of de criteria waaraan een goede schrijfpdracht moet voldoen, gebruikt zijn. Deze zijn in samenspraak met de leerlingen in de eerste les opgesteld. Een aantal criteriapunten die volgens de leerlingen belangrijk zijn bij het maken van een goede opdracht zijn: “Het onderwerp wordt in de inleiding genoemd”, “De tekst heeft een inleiding, kern en slot”, “De tekst heeft een logische volgorde”, “De lay-out is netjes” en “Er zitten geen spelfouten in de tekst en de tekst heeft minimaal 150 woorden”. Tevens wordt bij criteriagebruik beoordeeld of de feedback zich gericht

heeft op het leerproces en of er verbeterpunten worden gegeven. Explorerende stijl analyseert of de opmerkingen in de peerfeedback behulpzaam zijn, of er mogelijke oplossingen worden gegeven, of ze een toelichting geven en of de peerfeedback stimulerend is.

Om de kwaliteit van de peerfeedback te analyseren is er een scoringsrubriek verwerkt in het beoordelingsformulier, waarin de 7 gebieden beoordeeld kunnen worden op onvoldoende, voldoende en goed. In tabel 1 is een overzicht geschetst van het beoordelingsformulier met bijbehorende scoremogelijkheden. Twee onderzoekers hebben onafhankelijk van elkaar de gegeven feedback beoordeeld door middel van het beoordelingsformulier.

Vragenlijst attitude

Een instrument voor het meten van attitude bij huisartsen in opleiding (Prins, Sluijsmans en Kirschner, 2006) heeft als leidraad gefungeerd voor het samenstellen van een vragenlijst omtrent het meten van de attitude van de leerlingen. Het uitvoeren van een betrouwbaarheidsanalyse heeft uiteindelijk geresulteerd in een vragenlijst met 29 items. Wegens het hebben van een lage lading ($r = <. 3$) zijn 4 items uit de oorspronkelijke lijst van 33 items verwijderd.

De lijst representeert 3 factoren omtrent peerfeedback, te weten affectie ($N = 22$), inter-persoonlijk ($N = 6$) en zelfvertrouwen ($N = 3$) en laat een voldoende tot goede interne consistentie zien respectievelijk van $\alpha = .85$, $\alpha = .74$ en $\alpha = .65$ (Field, 2009).

Affectie heeft betrekking op voorkeuren en gevoelens bij het gebruiken van peerfeedback. Op het gebied van affectie zijn vragen gesteld als: "Ik vind het fijn om feedback te krijgen, zodat ik mijn werk kan verbeteren" en "Ik vind het prettig als in de feedback uitgelegd wordt waarom iets goed of fout is". De factor inter-persoonlijk representeert het gebruik van feedback en de omgang met klasgenoten. Er zijn vragen gesteld als: "Ik geef het liefst feedback aan een klasgenoot die ik goed ken" en "Als ik feedback geef houd ik er rekening mee of ik iemand aardig vind". De factor zelfvertrouwen geeft aan hoe het zelfbeeld is van de kinderen bij en na het gebruiken van peerfeedback. Hierbij zijn vragen gesteld als: "Ik ben zenuwachtig als ik iets moet zeggen over iemands werk" en "Ik ben meestal zenuwachtig als ik feedback krijg". Leerlingen hebben door middel van een 5-puntsverdeling kunnen aangeven in hoeverre zij het helemaal eens of helemaal oneens waren met een stelling.

Beoordelingsformulier leereffect

Zoals genoemd in de introductie van dit onderzoek kan het gebruik van peerfeedback ervoor zorgen dat het gat tussen de huidige en de gewenste situatie wordt verkleind (Black & William, 1998). Wanneer het gat verkleind wordt, is de peerfeedback effectief gebruikt (Leenknecht, 2010). Om te analyseren of er een leereffect heeft plaatsgevonden, hebben de leerkrachten de stelopdrachten van alle leerlingen twee maal beoordeeld. De leerkrachten hebben dezelfde criteria beoordeeld als degene die door de leerlingen tijdens de interventie (les 2) zijn opgesteld op het gebied van inhoud, opbouw en presentatie.

De leerkrachten hebben bepaald in welke mate de stelopdrachten aan de gestelde criteria hebben voldaan. De opdrachten zijn beoordeeld op 5 niveaus; onvoldoende, matig, voldoende, goed en zeer goed. De niveaus corresponderen met punten, zodat voor elke leerling een score bepaald kon worden. Onvoldoende is 1 punt, matig 2 punten, voldoende 3 punten, goed is 4 punten en zeer goed 5 punten (tabel 2). Aangezien er 14 criteria waren konden er 70 punten worden behaald. Bij de score worden 10 punten opgeteld waarna dit gedeeld is door 8, wat het uiteindelijke cijfer heeft gegeven.

Gebruik van film en drama

In de tweede les van de training is er voor het illustreren van de verschillende feedbackstijlen, waarbij de nadruk is gelegd op de explorerende stijl, gebruik gemaakt van videobeelden (Rooijmans, 2011). Om de leerlingen van de experimentele groepen bekend te maken met het geven van peerfeedback en om ze te laten oefenen met de rol van feedbackgever en -ontvanger, zijn er in de derde les verschillende dramawerkvormen gebruikt.

Tabel 1. Beoordelingsformulier voor de kwaliteit van de gegeven feedback gericht op criteriagebruik en gebruik explorerende stijl

Gebied	Aspect	Goed	Voldoende	Onvoldoende
Criteriagebruik	Gebruik alle criteria	Er worden 4 of meer criteria genoemd in de feedback.	Er worden 2-3 criteria genoemd in de feedback.	Er wordt maximaal 1 criteria genoemd in de feedback.
	Feedback gericht op leerproces	De feedback geeft minimaal 2 tips om het gat tussen de huidige en gewenste situatie te verbeteren.	De feedback geeft 1 tip om het gat tussen de huidige en gewenste situatie te verbeteren.	De feedback geeft geen tips om het gat tussen de huidige en gewenste situatie te verbeteren.
	Verbeterpunten	De feedback geeft 4 of meer verbeterpunten voor het gebruiken van de gestelde criteria.	De feedback geeft 2-3 verbeterpunten voor het gebruiken van de gestelde criteria.	De feedback geeft minimaal 1 verbeterpunt voor het gebruiken van de gestelde criteria.
Gebruik explorerende stijl	Behulpzame opmerkingen	Er worden meer dan 4 behulpzame opmerkingen geven waar de klasgenoot mee verder kan.	Er wordt een 1 tot 3 behulpzame opmerkingen gegeven waar de klasgenoot mee verder kan.	Er worden geen behulpzame opmerkingen geven waar de klasgenoot mee verder kan.
	Aandragen mogelijke oplossingen	De feedback geeft duidelijk en helder een minimaal 4 oplossingen voor het verbeteren van de opdracht.	De feedback geeft duidelijk en helder 2- 3 oplossingen voor het verbeteren van de opdracht.	De feedback geeft minimaal 1 oplossing voor het verbeteren van de opdracht.
	Geven van toelichting	De feedback geeft duidelijk meer dan 4 toelichtingen op de verbeterpunten door middel van een voorbeeld.	De feedback geeft een toelichting op de verbeterpunten door middel van een voorbeeld.	De feedback geeft geen toelichting op de verbeterpunten.
	Stimulerende peerfeedback	De feedback is erg stimulerend en er worden minimaal 2 complimenten gegeven.	De feedback is een beetje stimulerend en er wordt één compliment gegeven.	De feedback is niet stimulerend en geeft geen complimenten.

Tabel 2 . Voorbeeld van beoordelingsniveaus met bijbehorende punten op het gebied van leereffect

Gebied	Criteria	Onvoldoende	Matig	Voldoende	Goed	Zeer goed
		(1)	(2)	(3)	(4)	(5)
Opbouw	Het slot rond de tekst af?	Er is helemaal geen slot.		Er is een slot waarneembaar.		Het slot is duidelijk herkenbaar.
Inhoud	De tekst heeft een logische volgorde?	Tekst loopt erg door elkaar.		Tekst heeft een redelijke volgorde.		Tekst heeft een duidelijke volgorde.
Presentatie	Er zitten geen spelfouten in de tekst?	Meer dan 8 fouten.		Vijf fouten in de tekst.		De tekst heeft 1 of geen spelfouten.

Analyse

Het criteriagebruik en de stijl bij de peerfeedback zal geanalyseerd worden door middel van een t- toets om eventuele verschillen tussen de experimentele en controlegroep aan te tonen. Daarna zal door middel van een Anova geanalyseerd worden wat het effect van conditie is op de kwaliteit van de feedback bij groep 7 en wat het effect is van de conditie op stijl. Ook zal het effect van stijl op criteriagebruik berekend worden, waarbij als covariaat de conditie toegevoegd zal worden. Tevens zal door middel van een Anova het effect van geslacht worden geanalyseerd.

De vragenlijst voor het meten van de affectie jegens peerfeedback wordt allereerst geanalyseerd door middel van een betrouwbaarheids- en factoranalyse. Door middel van een onafhankelijke t-toets zullen de scores van zowel de experimentele- als de controlegroep vergeleken worden. Het veranderingseffect tussen de voor- en nameting zal geanalyseerd worden door middel van een afhankelijke t-toets.

Het leereffect wordt voor alle groepen apart bepaald door het verschil tussen eerste versie en de tweede versie te analyseren. Van beide leereffecten wordt met een gepaarde t-toets bepaald of er een leereffect is. Met behulp van een Ancova zal het effect van stelopdracht 1 op stelopdracht 2 geanalyseerd worden. Tevens zal conditie als covariaat getoetst worden.

Resultaten

Kwaliteit van de feedback

De kwaliteit van de feedback gegeven door de leerlingen is op te delen in de mate van criteriagebruik en de mate van gebruik van de explorerende stijl. Beide kwaliteitsbeoordelingen zijn gedaan door twee beoordelaars, op beide onderdelen is een hoge betrouwbaarheid (Cohen, 1960) waar te nemen, namelijk Cohen's Kappa crit = .76 en Cohen's Kappa stijl = .74.

Om te kunnen bepalen of de training effect heeft op de kwaliteit van de feedback wordt er eerst gekeken wat het gemiddelde per groep is voor het gebruik van de criteria en de explorerende stijl (tabel 3). De experimentele groep 7 heeft andere gemiddelden bij beide kwaliteitscriteria, de feedbackstijl van de controle groep 8 is juist veel hoger dan van de experimentele groep.

Tabel 3. Overzicht gemiddelde, standaard deviatie per groep voor criteriagebruik en gebruik voor explorerende stijl

Variabelen	Controlegroep						Experimentele groep					
	7 (n=23)		8 (n=26)		Totaal (n=49)		7 (n=25)		8 (n=23)		Totaal (n=48)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Kw_criteria	5.13	1.60	5.77	1.80	5.47	1.72	6.38	1.63	5.35	1.58	5.89	1.67
Kw_stijl	4.91	1.24	7.12	3.04	6.08	2.60	7.03	2.29	5.57	2.15	6.32	2.32

Verskil in kwaliteit van de feedback

De kwaliteit is op te delen in Kw_criteria en Kw_stijl. Op het gebied van Kw_criteria is het hoger gebruik van de experimentele groepen niet significant, $t(95) = -.33$, $p = .74$. Ook op het gebied van Kw_stijl is er geen significante toename voor de experimentele groep, $t(95) = .25$, $p = .81$.

Bij de twee groepen 7 scoort de experimentele groep hoger dan de controlegroep, daarom zal hier nader naar gekeken worden. Het verschil tussen de twee groepen 7 op het gebruik van de explorerende stijl is significant, $t(46) = -2.45$, $p = .02$. Bij de beoordeling van de kwaliteit van de feedback is geanalyseerd hoe groot het effect is van de conditie (controlegroep of experimentele groep). De conditie is significant gerelateerd aan het gebruik van de explorerende stijl, $F(1,47) = 15.28$, $p < .001$, $\eta^2 = .25$. Daarnaast is er ook een significant effect van criteriagebruik als covariaat op feedbackstijl, $F(1,45) = 79.33$, $p < .001$, $\eta^2 = .64$. De conditie blijkt nog steeds van invloed te zijn, $F(1,45) = 7.35$, $p = .01$, $\eta^2 = .14$. De totale effectgrootte van criteriagebruik en de conditie op feedbackstijl blijkt $\eta^2 = .73$ te zijn.

Criteriagebruik blijkt veel effect te hebben op het gebruik van de explorerende stijl. Er kan daarom ook worden afgevraagd of de explorerende stijl veel effect heeft op het criteriagebruik. Daarvoor wordt eerst beoordeeld of het volgen van een peerassessmenttraining effect heeft op criteriagebruik, dat blijkt zo te zijn, $F(1,47) = 7.16$, $p = .01$, $\eta^2 = .14$. Het daarna toevoegen van feedbackstijl als covariaat blijkt een significant effect te geven, $F(1,45) = 79.33$, $p < .001$, $\eta^2 = .64$. Alleen blijkt dan de peerassessmenttraining geen significant effect meer te hebben, $F(1,47) = 0.41$, $p = .41$, $\eta^2 = .01$.

In de tabel staan de gemiddelde scores voor de explorerende stijl van jongens en meisjes van zowel de experimentele als de controlegroep. Het blijkt dat de meisjes bij beide groepen beter scoren. Dit verschil is significant, $F(1,96) = 11.59$, $p < .01$, $\eta^2 = .11$.

Tabel 4. Verskil gebruik explorerende stijl tussen jongens en meisjes

Geslacht	Conditie	Gemiddelde	SD	N
Jongen	Controle	5.13	1.85	24
	Experimenteel	5.73	2.18	28
	Totaal	5.45	2.03	52
Meisje	Controle	7.00	2.90	25
	Experimenteel	7.15	2.32	20
	Totaal	7.07	2.63	45
Totaal	Controle	6.08	2.60	49
	Experimenteel	6.32	2.32	48
	Totaal	6.20	2.45	97

Attitude

Een principale componentenanalyse met oubliieke rotatie is uitgevoerd op 29 items, met als doel te bepalen of de samenhang tussen geobserveerde variabelen en de hypothetische variabelen correct is. Op deze manier is getracht een meer gedetailleerd inzicht te krijgen in de interne structuur van de vragenlijst. De factoranalyse laat zien dat er 9 componenten met een Eigenwaarde groter dan 1, en 3 componenten met een Eigenwaarde groter dan 2 te onderscheiden zijn. De 3 componenten met een Eigenwaarde groter dan 2 zijn te interpreteren als de 3 factoren; affectie, inter-persoonlijk en zelfvertrouwen. Zij hebben respectievelijk een Eigenwaarde van 6.25, 4.07 en 2.73. De factor affectie verklaart 20.17%, de factor inter-persoonlijk 11.20% en de factor zelfvertrouwen 7.93% van de variantie. Tezamen verklaren de 3 factoren 42.29%, wat opgevat kan worden als voldoende (Field, 2009). In tabel 5 wordt de componentenmatrix weergegeven.

Tabel 5. Componentenmatrix principale factoranalyse voor meting van attitude omtrent peerfeedback (ladingen > .3)

	Attitude	Inter-persoonlijk	zelfvertrouwen
ik ben zenuwachtig als ik iets moet zeggen over iemands werk			.79
ik ben meestal zenuwachtig als ik feedback krijg			.76
als ik feedback geef houd ik er rekening mee of ik iemand aardig vind		.59	
als ik feedback krijg dan waardeer ik dit	.68		
bij iemand die ik aardig vind geef ik positievere feedback	.40		
doordat ik feedback geef leer ik zelf ook	.58		
feedback geven is leuk	.61		
ik geef het liefst feedback aan een klasgenoot die ik goed ken		.49	
ik krijg het liefst feedback van een klasgenoot die ik goed ken		.59	
ik vind het fijn om feedback te krijgen omdat ik dan meer kan leren	.78		
ik krijg liever van mijn juf of meester feedback dan van mijn klasgenootjes		.38	
ik leer meer als ik complimenten krijg	.33		
ik vind dat je meer leert van strenge feedback	.36		
ik krijg liever van een klasgenoot die ik aardig vind feedback dan van een klasgenoot die ik niet zo aardig vind		.67	
ik vind het fijn om feedback te krijgen omdat ik dan mijn werk kan verbeteren	.72		
ik vind het fijn om met mijn klasgenoten te overleggen waarop we moeten letten als we feedback geven	.46		
ik vind het fijn om snel feedback te krijgen	.33		
ik vind het fijn om te praten over de feedback die ik heb gegeven	.52		
ik vind het prettig als in de feedback uitgelegd wordt waarom iets goed of fout is	.43		
ik vind het fijn om vaak feedback te geven	.66		
van gekregen feedback van een klasgenoot kan ik iets leren	.67		
ik vind het moeilijk als iemand iets negatiefs over mijn werk zegt		.52	
ik vind het prettig als feedback aangeeft of mijn werk goed of fout is	.57		
ik vind het fijn om te praten over de feedback die ik heb gekregen	.52		
ik zie het nut van elkaar feedback geven	.70		
van de feedback die ik kreeg heb ik meer zelfvertrouwen gekregen	.48		
ik vind het fijn om vaak feedback te krijgen	.61		
ik kan gerust iets zeggen over iemands werk			.48

Noot. N = 29

Bij het bepalen of er een effect is van de interventie op de attitude van de leerlingen zijn allereerst de gemiddelden en de standaarddeviaties van de groepen vergeleken, zie tabel 6.

Tabel 6. Gemiddelde en standaard deviatie per groep op de voor- en nameting van attitude

Variabelen	Controlegroep						Experimentele groep					
	7 (n=31)		8 (n=25)		Totaal (n=56)		7 (n=31)		8 (n=21)		Totaal (n=52)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Affectie 1	71.11	9.07	62.04	12.89	66.58	10.98	65.58	9.07	60.86	14.27	63.22	11.67
Affectie 2	66.16	7.83	60.07	11.27	63.12	9.55	65.00	7.83	60.00	18.72	62.5	13.28
Interpers. 1	16.26	3.79	16.81	4.06	16.54	3.93	21.22	3.79	17.54	4.86	19.38	4.33
Interpers. 2	14.29	3.99	15.26	5.01	14.78	4.50	18.79	3.99	14.63	5.84	16.71	4.92
Zelfv. 1	10.96	2.79	12.85	2.14	11.91	2.47	10.97	2.79	12.58	2.66	11.78	2.73
Zelfv. 2	11.17	2.74	12.89	2.26	12.03	2.50	11.16	2.74	13.00	2.67	12.08	2.71

De nameting van zowel affectie als inter-persoonlijk is bij elke groep duidelijk lager dan de voormeting. Het verschil in gemiddelde op het gebied van affectie van de experimentele groep is significant, $t(51) = 2.50$, $p = .01$. Echter voor de controlegroepen niet, $t(51) = .51$, $p = .61$. Deze afname voor inter-persoonlijk is voor zowel de experimentele ($t(57) = 3.39$, $p < .001$) als de controle groep ($t(55) = 4.57$, $p < .001$) significant. Het verschil in toename van het gemiddelde op het gebied van zelfvertrouwen blijkt voor zowel de experimentele ($t(54) = -.465$, $p = .32$) als de controlegroep ($t(55) = -.83$, $p = .41$) niet significant.

Vershil in attitude

Het gemiddelde van de controle groepen ($M = 63.12$, $SD = 9.77$) op de nameting op het gebied van affectie is groter dan van de experimentele groepen ($M = 62.50$, $SD = 13.60$). Er is geen significant effect van conditie op de afname in affectie waar te nemen, $t(113) = .20$, $p = .42$. Het gemiddelde van de controle groepen ($M = 14.74$, $SD = 4.84$) op het gebied van inter-persoonlijk is kleiner dan van de experimentele groepen ($M = 17.04$, $SD = 5.24$). Er is ook geen significant effect van conditie op de afname van inter-persoonlijk, $t(113) = -2.44$, $p = .08$.

De toename bij de controlegroepen ($M = 11.98$, $SD = 2.50$) op het gebied van zelfvertrouwen is groter dan het gemiddelde van de experimentele groepen, $M = 11.95$, $SD = 2.84$. Deze toename is niet significant, $t(111) = .07$, $p = .47$.

Leereffect

De stelopdrachten zijn beoordeeld door de leerkracht van de eigen groep. Versie 1 en 2 zijn apart van elkaar beoordeeld en in tabel 7 worden de statistieken gerapporteerd.

Tabel 7. Overzicht gemiddelde en standaard deviatie per groep voor versie 1 en 2 van de stelopdracht

	Controlegroep						Experimentele groep					
	7 (n=31)		8 (n=26)		Totaal (n=57)		7 (n=32)		8 (n=24)		Totaal (n=56)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Versie 1	7.10	1.30	7.76	1.67	7.41	1.50	6.01	1.21	5.78	1.26	5.92	1.23
Versie 2	7.57	1.10	8.34	1.37	7.92	1.28	6.45	1.31	6.44	1.16	6.45	1.24

Bij iedere groep is het punt van versie 2 van de schrijfopdracht significant toegenomen ten opzichte van het punt van versie 1. Voor de controlegroep 7 is versie 2 gemiddeld .47 hoger, deze toename is significant, $t(30) = 6.44$, $p < .001$. Voor de experimentele groep 7 is versie 2 gemiddeld .44 hoger, deze toename is significant, $t(31) = 3.77$, $p = .001$. Bij de controlegroep 8 gaat versie 2 gemiddeld .58 omhoog, deze toename is significant, $t(25) = 2.92$, $p = .01$. Bij de experimentele groep 8 gaat versie 2 gemiddeld .66 omhoog, ook deze toename is significant, $t(23) = 6.26$, $p < .001$.

Vershil in leereffect

De experimentele en controlegroepen zijn bij versie 1 verschillend beoordeeld, $t(109.27) = 5.83$, $p < .001$ bij tweezijdige toetsing. Ook versie 2 is verschillend beoordeeld bij de groepen, $t(111) = 6.20$, $p < .001$ bij tweezijdige toetsing. Omdat de opdrachten met verschillende cijfers zijn beoordeeld, wordt het verwachte hogere leereffect bij de experimentele groep getoetst met een ANCOVA, waarbij versie 1 als covariaat meegenomen wordt. De toename in leereffect bij de experimentele groepen blijkt bij eenzijdige toetsing significant te zijn, $F(1, 110) = 3.16$, $p = .04$, $\eta^2 = .03$. De effectgrootte van versie 1 blijkt significant te zijn en veel groter, $F(1, 110) = 424.20$, $p < .001$, $\eta^2 = .79$. De totale effectgrootte wordt dan $\eta^2 = .85$.

Omdat de scores van de afzonderlijke groepen op versie 1 en 2 uiteen blijken te liggen, is er getoetst of deze scores ook significant verschillen. De onafhankelijke t-toets geeft aan dat de experimentele en controlegroepen bij versie 1 verschillend zijn beoordeeld ($t(109,27) = 5.83$, $p < .001$ bij tweezijdige toetsing). Ook versie 2 is verschillend beoordeeld ($t(111) = 6.20$, $p < .001$ bij tweezijdige toetsing). Omdat de opdrachten met verschillende cijfers zijn beoordeeld, wordt er een

ANCOVA uitgevoerd tussen versie 2 van de controle en experimentele groep met als covariaat versie 1. De verschillen laten zien dat de experimentele groepen een groter leereffect hebben dan de controlegroepen. Met de ANCOVA lijkt de conditie een heel klein effect te hebben ($\eta^2 = .03$ met eenzijdige toetsing, $p = .04$), waarbij cijfer 1 een groot effect heeft ($\eta^2 = .79$, $p < .001$).

Voor de ANCOVA die alleen uitgevoerd wordt met groep 7, is er geen significant resultaat waar te nemen voor conditie, $F(1, 63) = 1.35$, $p = .13$ (eenzijdige toetsing). Bij de ANCOVA die uitgevoerd is bij groep 8, komt er ook geen significant resultaat uit voor de conditie, $F(1,50) = 2.26$, $p = .07$ eenzijdige toetsing.

Discussie

Het doel van dit onderzoek is om aan te tonen dat basisschoolleerlingen die een peerassessmenttraining hebben gevolgd een betere attitude krijgen jegens peerfeedback, betere peerfeedback geven en een groter leereffect hebben. In totaal hebben er vier groepen meegedaan, twee groepen 7 en twee groepen 8. Een groep 7 en een groep 8 behoren tot de experimentele groep en de andere twee groepen behoren tot de controle groep. De groepen komen allemaal van dezelfde school, dit kan invloed hebben op de betrouwbaarheid en de validiteit ook omdat er gebruik is gemaakt van bestaande klassen. Daarnaast is de betrouwbaarheid van de instrumenten wel gewaarborgd.

1. *Heeft het volgen van de peerassessmenttraining, gericht op de exploratieve feedbackstijl, een positief effect op de kwaliteit van de gegeven peerfeedback bij bovenbouwleerlingen van het primair onderwijs?*

Inzicht in de peerfeedbackvaardigheden van zowel de experimentele als de controlegroep zijn verkregen door middel van het analyseren van de gegeven peerfeedback door twee onderzoekers. De interbeoordelaarsbetrouwbaarheid is hierbij voldoende gebleken (Cohen's Kappa $>.70$; Cohen, 1960) waarbij gesteld kan worden dat de data op voldoende wijze hetzelfde beoordeeld is.

Resultaten van de kwaliteit van de gegeven peerfeedback laten zien dat de experimentele groep 7 van alle groepen naar verwachting het hoogste scoort op criteriagebruik en de controle groep 8 het hoogste scoort op het gebruik van de explorerende stijl wat niet volgens verwachting is. Toch blijken zowel conditie als groep geen significant effect te hebben op zowel het criteriagebruik als het gebruik van de explorerende stijl. Deze resultaten zijn natuurlijk niet volgens verwachting aangezien er bij het onderzoek vanuit wordt gegaan dat de conditie wel degelijk een significant effect geeft. Als je echter alleen kijkt naar de groepen 7 dan is naar verwachting te zien dat het verschil wel significant is in het voordeel van de experimentele groep. Bij de groepen 7 is de effectgrootte van conditie bij de explorerende stijl 25%. Als je bij de effectgrootte van conditie ook criteriagebruik meeneemt dan is de effectgrootte gegroeid naar 73%.

Het lage gemiddelde van de experimentele groep 8 geeft echter wel stof tot nadenken. Deze groep gebruikt minder behulpzame opmerkingen, oplossingen en toelichtingen in de peerfeedback dan de controlegroep 8. Tevens is de peerfeedback van deze groep tegen de verwachting in minder positief dan de feedback gegeven door de controlegroep 8. De experimentele groep 8 toont geen duidelijke exploratieve stijl in de gegeven feedback omdat zij of niet over de juiste vaardigheden bezitten om deze stijl effectief te kunnen toepassen of nog niet spontaan de juiste strategie hiervoor kunnen toepassen (Flavell,1976; Prins, Sluijsmans & Kirschner, 2006). Met de juiste vaardigheden wordt bedoeld het geven van verklaringen en suggesties voor herziening van het product en daarbij wordt het perspectief van de student overgenomen en geeft de feedbackgever verklaringen en suggesties die corresponderen met de intentie van de student.

De controlegroep 8 lijkt tevens in het voordeel te zijn om een tweetal redenen. Uit gesprekken met de leerkracht van deze groep blijkt er een positief groepsklimaat te heersen, wat zeer belangrijk is voor de onderlinge relaties (Wragg, 2001). Dit zal niet alleen invloed hebben op de attitude jegens peerfeedback, het zal ook invloed uitoefenen op de explorerende stijl. Een leerling welke zich sociaal veilig kan bewegen en de ondersteuning en structuur krijgt om zich

competent te voelen, zal immers makkelijker positieve explorerende feedback kunnen geven. Daarnaast blijkt uit dezelfde gesprekken dat de controlegroep 8 voor dit onderzoek wekenlang les heeft gehad in het geven van complimenten. Deze lessen zijn onderdeel van een onderzoek van de des betreffende docent dat een zestal weken in beslag heeft genomen. Ook dat kan verklaren waarom zij de explorerende stijl meer hanteren dan de experimentele groep 8.

Daarnaast geeft Sadler (1989) aan dat het noodzakelijk is dat er een behoorlijk aantal voorbeelden worden gegeven, willen deze voorbeelden effectief zijn. Tijdens de training hebben de leerlingen 3 voorbeelden gezien. Bij elke feedbackstijl is er een filmpje getoond. Hieruit kan geconcludeerd worden dat dit niet voldoende is en dus niet effectief genoeg om de explorerende stijl aan te leren.

Dat leerlingen in het primair onderwijs elkaar feedback kunnen geven, blijkt uit de hogere gemiddelden op zowel gebruik van criteria alsmede het gebruik van de explorerende stijl van de experimentele groep 7, dit is ook volgens verwachting. De experimentele groep 7 geeft bij de attitudemeting aan dat zij het nut van elkaar feedback geven inzien. Tevens vinden zij het fijn om peerfeedback te krijgen zodat zij hun werk kunnen verbeteren en meer kunnen leren. Hiermee kan een koppeling gemaakt worden naar Lindsay en Clarke (2001), die stellen dat leerlingen door het gebruik van peerfeedback verantwoordelijkheid nemen voor hun eigen leren. Dit leidt tot een kritischere zelfhouding, actieve deelname aan het leerproces en het stellen van hogere doelen. Om hier inhoudelijk vorm aan te kunnen geven, vereist dit ontwikkeling van feedbackvaardigheden van de leerlingen gedurende een groter tijdsbestek, waarbij die geïntegreerd zouden kunnen worden in het curriculum van de school. Deze ontwikkeling wordt gestimuleerd door middel van een grote hoeveelheid specifieke oefeningen (Sluijsmans, 2002).

Daarnaast is er ook nog gekeken naar het verschil van explorerende stijl voor zowel jongens als meisjes. Hieruit komt duidelijk naar voren dat meisjes hoger scoren dan jongens, dit verschil is significant. Het maakt hierbij bijna geen verschil of de jongen in de controle ($m = 5.13$) of experimentele groep ($m = 5.73$) zit, hetzelfde geldt bij de meisjes (controle $m = 7.00$ en experimenteel $m = 7.15$). Uit onderzoek van Leenknecht (2008) blijkt dat meisjes meer verificaties maken dan jongens. Ook is er bij dat onderzoek geconcludeerd dat meisjes meer uitgebreide feedback geven. Daarnaast hebben ze ook meer statements gegeven dan hun mannelijke klasgenoten. Dit kan misschien verklaren waarom bij dit onderzoek de meisjes hoger scoren op de explorerende stijl.

2. Heeft het volgen van de peerassessmenttraining, gericht op de feedbackstijl, een positief effect op de attitude ten opzichte van peerfeedback bij bovenbouwleerlingen van het primair onderwijs?

De resultaten laten zien dat bij de attitude jegens peerfeedback voor de factoren affectie en inter-persoonlijk niet geldt dat dit bij de leerlingen die een peerassessmenttraining hebben gevolgd, beter wordt. Bij alle groepen zowel de experimentele als de controlegroepen scoren ze lager dan de eerste meting op deze twee factoren, dit is niet in de lijn met wat van te voren verwacht is. Op de factor zelfvertrouwen scoren alle groepen, naar verwachting, wel hoger dan de vorige meting maar dit is voor alle groepen niet significant. Dit kan verschillende verklaringen hebben.

Zo kan er bij de literatuur al gelezen worden dat studies laten zien dat ook een gebrek aan zelfvertrouwen wordt gezien wanneer studenten elkaar moeten beoordelen (Sullivan, Hitchcock & Dunnington, 1999). Uit onderzoek blijkt dus dat leerlingen vaak niet tevreden zijn met de feedback die ze krijgen omdat het te moeilijk te interpreteren is maar ook omdat het een negatieve impact heeft op het zelfvertrouwen van de leerlingen (Carless, 2006). De leerlingen van dit onderzoek scoren echter hoger op zelfvertrouwen. Dit komt waarschijnlijk doordat ze elkaar feedback geven, hierdoor kunnen ze de feedback goed interpreteren. De feedback is ook erg positief uitgevallen met benamingen zoals 'alles is perfect', 'wauw' en 'goed gedaan'. De nadruk van de training heeft echter niet op zelfvertrouwen gelegen dat kan een reden zijn waarom het verschil niet significant is. Daarnaast hebben maar 3 items van de vragenlijst betrekking op de factor zelfvertrouwen, dat kan er voor zorgen dat de factor zelfvertrouwen geen significante resultaten geeft.

Tevens geeft de literatuur aan dat studenten behoefte hebben aan een reeds bestaande richtlijn of regel voor de assessmentactiviteit (Orsmond & Merry, 1996). Dit kan verklaren waarom de leerlingen niet hoger gaan scoren. Er is bij de training wel geoefend met peerfeedback geven maar dat is maar één keer gebeurd. Doordat de leerlingen maar één keer hebben kunnen oefenen, hebben ze niet voldoende kennis opgedaan om makkelijk en snel goede feedback te kunnen geven. Daardoor kost het geven van peerfeedback veel tijd en kunnen de leerlingen het als lastig ervaren hebben, dat kan effect hebben op de affectie jegens peerfeedback.

Ook voor het lager scoren op de factor inter-persoonlijk zijn verklaringen te geven. Zo geeft de literatuur aan dat het groepsklimaat wordt gezien als een krachtige druk uit de omgeving die het gedrag van individuele leden kan vergroten of belemmeren (Phipps & Zastowny, 1988). De leerlingen uit de experimentele groepen hebben op de voormeting hoger gescoord dan de leerlingen uit de controlegroep. Het verschil tussen de voor- en nameting bij de experimentele groepen is echter veel groter dan bij de controlegroep. Een groot deel hiervan is te verklaren door het groepsklimaat, zoals al eerder is gemeld, is uit gesprekken gebleken dat de controlegroep een positieve groepsklimaat heeft. Dit zou kunnen betekenen dat dit ook effect heeft op de factor inter-persoonlijk. Ook is er tijdens de training geen aandacht besteed aan het voordeel van peerfeedback voor de factor inter-persoonlijk of de samenhang hier tussen. Als hier meer aandacht aan zou zijn besteed, dan zouden de resultaten voor de experimentele groepen anders uit kunnen pakken.

3. *Heeft het volgen van de peerassessmenttraining, gericht op de feedbackstijl, een positief leereffect op het schrijven van een stelopdracht door bovenbouwleerlingen van het primair onderwijs?*

Bij het leereffect is duidelijk een significant verschil tussen de twee metingen waar te nemen voor alle groepen, waarbij de gemiddelde toename voor de experimentele groepen groter is dan voor de controle groepen. Conditie blijkt hierbij, naar verwachting, een significant verschil te geven. Het geven van peerfeedback heeft dus een positief effect.

De effectgrootte van conditie bij cijfer 2 is tegen de verwachting in echter niet zo groot, 3%. Dat conditie een klein effect geeft, kan afhankelijk zijn van het hanteren van de exploratieve feedbackstijl. Analyse van eerdere gegevens laat namelijk zien dat niet alle leerlingen de exploratieve stijl consequent gebruikten bij het verstrekken van de peerfeedback. In de feedback moeten behulpzame opmerkingen, oplossingen en toelichtingen gegeven worden, wil deze effectief gebruikt kunnen worden voor het leerproces (Black and William, 1998). Wanneer er in de gegeven feedback niet expliciet is aangegeven wat er aan de opdracht verbeterd moet worden, dan zal de leerling die de feedback terug krijgt niet tot nauwelijks een leereffect ervaren. Die leerling zal dan zelf de keuze moeten maken tussen verbeteren of niets aan de opdracht veranderen.

Als je bij deze effectgrootte echter cijfer 1 ook meeneemt dan wordt deze vergroot naar 79%. Dit geeft dus aan dat versie 1 erg veel effect heeft op het leereffect en dus op versie 2. Daarnaast is er gekeken of het verschil tussen de twee groepen 7 en de twee groepen 8 significant is, dit is echter niet het geval. Ook is naar voren gekomen dat zowel versie 1 als versie 2 bij de experimentele en controlegroepen verschillend zijn beoordeeld. Dit wetende, kan er worden afgevraagd of de resultaten van het leereffect wel betrouwbaar zijn.

Conclusie

Peerfeedback wordt niet vaak gebruikt in het primair onderwijs. Dit onderzoek laat zien dat peerfeedback in het primair onderwijs zeker effect kan hebben. Zo heeft het onderzoek laten zien dat een training omtrent peerfeedback effect heeft op de kwaliteit van de feedback en op het leereffect. Dit verschil is echter alleen van toepassing op de groepen 7. Zoals in de discussie al valt te lezen, zijn er verschillende oorzaken waarom deze effecten niet voor de twee groepen 8 gelden. Dus hieruit valt wel te concluderen dat er een effect optreedt als de twee groepen homogener aan elkaar zijn.

Volgens Sadler (1998) is kennis over standaarden en criteria een essentieel deel van de competentie om effectieve feedback te verzorgen. In dit onderzoek hebben alle leerlingen de kennis over de criteria gekregen en sommige leerlingen hebben ook de kennis over de standaarden gekregen. Hoewel bij de groepen 7 is te zien dat de kennis over de standaarden een verschil kan maken, laten de twee groepen 8 zien dat er ook andere invloeden zijn, zoals het groepsklimaat en de lessenserie complimenten geven.

Een vervolgonderzoek kan echter nuttig zijn om bepaalde bevindingen te verifiëren. Zo kan er bij vervolgonderzoek naar peerfeedback in het primair onderwijs gekeken worden naar de voordelen van het gebruik van peerfeedback gericht op de attitude en de effectiviteit van de feedback. Een evenredige verdeling van zelfvertrouwen en attitude bij de vragenlijst zal misschien betere resultaten opleveren. Hierbij lijkt een effectonderzoek naar het tijdsbestek waarin een training omtrent peerfeedbackvaardigheden kan plaatsvinden nuttig, wil deze feedback effectief zijn. Daarnaast is het ook belangrijk dat de leerlingen een bestaande richtlijn of regel krijgen aangereikt voor het geven van peerfeedback in de explorerende stijl omdat uit onderzoek (Orsmond & Merry, 1996) is gebleken dat leerlingen hier behoefte aan hebben. De leerlingen moeten weten hoe ze de criteria op een explorerende stijl kunnen overbrengen bij de peerfeedback en bepaalde begrippen moeten duidelijk zijn. Zo moet van te voren bijvoorbeeld duidelijk worden aangegeven dat de leerlingen stimulerende opmerkingen moeten gebruiken, zodat de feedbackontvanger ook echt gestimuleerd wordt om zijn tekst te verbeteren. Ook is het belangrijk dat ze weten wat stimulerende opmerkingen zijn. Ook is het interessant om te kijken naar samenwerking tussen jongens en meisjes aangezien meisjes hoger scoren op het gebruik van de explorerende stijl. Als laatste is het ook belangrijk om bij vervolgonderzoek een uitgebreidere training te maken. Uit onderzoek (Prins, Sluismans & Krischner, 2006) is gebleken dat de kwaliteit van de gegeven peerfeedback gericht op de explorerende stijl vergroot kan worden maar dan is er een uitgebreide training wel essentieel. Aangezien de training nu uit 4 lessen bestaat, is het voor vervolgonderzoek belangrijk dat er een uitgebreidere training komt zodat de effecten misschien groter zijn.

Naar aanleiding van dit onderzoek kan je concluderen dat het geven van peerfeedback nut heeft in het primair onderwijs. Het is echter aan te raden om de leerlingen een langere training te geven zodat ze genoeg tijd hebben om te oefenen met de explorerende stijl en ze op die manier ook meer aandacht kunnen besteden aan de attitude jegens peerfeedback.

Referenties

- Berg, v. d. (2003). *Peer assessment in universitair onderwijs*. Utrecht: IVLOS.
- Berg, V. d., Pilot, A., & Admiraal, W. (2005). *Peer assessment als leermiddel*. Utrecht: Universiteit Utrecht (IVLOS).
- Berihun, A., Tesera, A., & Desta, D. (2006). Enchancing active learning through teachers' peer and self reflections in selected primary schools. *Review of Educational Research, 29*, 172-187.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education: principles, policy & practice, 5*, hoofdstuk 1.
- Briggs, M., Woodfield, A., Martin, C., & Swatton, P. (2008). *Assessment for learning and Teaching in Primary Schools*. Exeter: Learning Matters.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31(2)*, 219-233.
- Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higer Education, 22*, 233-239.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20(1)*, 37-46.
- Falchikov, N. (1996). Involving students in feedback and assessment. *Paper presented at the Assessment Strategies in Scottish Higher Education (ASSHE) conference*. Stirling, UK.
- Field, A. (2009). *Discovering statistics using SPSS, third edition*. London: SAGE Publications Ltd.
- Flavell, J. (1976). Metacognitive aspects of problem solving. *The Nature of Intelligence, 231-236*.
- Hargreaves, A. (1997). Rethinking educational change: Going deeper and wider in the quest for success. In A. Hargreaves, *Rethinking educational change with heart and mind*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*, 81-112.
- Keaten, J., & Richardson, M. (1992). A field investigation of peer assessment as part of the student group grading process. *Paper presented at the Western Speech Communication Assocation Convention*. Albuquerque: NM.
- Leenknecht, M. (2011). Formative peer feedback in primary school: the effect of knowledge about standards and criteria on the feedback style. *Nog niet gepubliceerd*, -.
- Lindsay, C., & Clarke, S. (2001). Enchancing primary science through self and paired assessment. *Primary Science Review, 68*, 15-18.
- Meusen-Beekman, K., & Joosten-Ten Brinke, D. (2010). Peer assessment in het basisonderwijs. *Examens, 3*, 13-16.
- Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higer Education, 21*, 239-250.
- Phipps, L., & Zastowny, T. R. (1988). Leadership behavior, group climate and outcome in group psychotherapy: a study of outpatient psychotherapy groups. *GROUP, 12(3)*, 157-171.
- Prins, F., Sluijsmans, D., & Kirschner, P. (2006). Feedback for General Practioners in Training: Quality, Styles, and Preferences.
- Rooijmans, M. (2011, 04 25). *Homepage*. Opgeroepen op 04 26, 2011, van Homepage: <http://feedbackfilms.webklik.nl>
- Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

- Silverman, S., Tyson, L., & Krampitz, J. (1992). Teacher feedback and achievement in physical education: interaction with student practice. *Teaching & Teacher Education, 8*, 333-344.
- Sluijsmans, D. (2002). *Student involvement in assessment: The training of peer assessment skills*. Heerlen: Open Universiteit Nederland.
- Stiggings, R. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practice, 6*, 33-42.
- Straub, S. (1996). The concept of control in teacher response: defining the varieties of "directive" and "facilitative" commentary. *College Compositions and Communication, 47*, 223-251.
- Sullivan, M., Hitchcock, M., & Dunnington, G. (1999). Peer and self assessment during problem-based tutorials. *The American Journal of Surgery, 177*, 266-269.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249-276.
- Wragg, E. (2001). *Assessment and learning in the Primary School*. New York: RoutledgeFalmer.