



Universiteit Utrecht

Masterthesis

De invloed van relatie op de uitvoering van peer-assessments

Naam: Walther van Mechelen (4005848)

Cursus: Masterthesis, themagebied Samenwerkend Leren (200700224)

Begeleider: Dr. J. Janssen

Tweede begeleider: Dr. B. Slof

Datum: 11 januari 2015

Aantal woorden: 7461

Samenvatting

Samenwerkend leren en peer-assessments worden steeds belangrijker (Oostdam, Peetsma & Blok, 2007). Dit onderzoek bestudeert de totstandkoming van peer-assessments (Greguras, Robie & Born, 2001; Phielix, 2012) en richtte zich specifiek op de invloed van relaties tussen peers bij de uitvoering van peer-assessments. Er is gebruik gemaakt van het Social Relations Model (Kenny, 1994). Met behulp van een gestandaardiseerd instrument en een Round Robin Design naar voorbeeld van Phielix (2012) zijn beoordelingen van peers over een samenwerkingsopdracht geanalyseerd.

Met behulp van de applicatie SOREMO (Kenny, 1998) werd de totale variantie van de assessments berekend, waarna deze verdeeld werd in actor- en partnervariantie, evenals dyadische (relatie-) variantie. Ook berekende SOREMO correlaties tussen de beoordelingen van peers onderling. $N = 99$ respondenten uit de eerste en tweede klas in het voortgezet onderwijs namen deel aan het onderzoek.

De variantieanalyse liet een invloed van relatie zien, maar bestudering van de samenhang tussen beoordelingen van peers (*dyadic* en *generalized reciprocity*) liet geen significante invloed van relatie zien. Er werd wel een lichte samenhang waargenomen tussen de beoordelingen van peers onderling (*dyadic reciprocity*). Ook werd een samenhang gezien tussen de beoordelingen van een peer door groepsgenoten en de beoordelingen van die groepsgenoten door de beoordeelde peer (*generalized reciprocity*). Mogelijke oorzaken voor de zwakke samenhang worden besproken.

Hoewel geen directe conclusies over betrouwbaarheid en validiteit kunnen worden getrokken, is dit belangrijk fundamenteel onderzoek waarmee de basis wordt gelegd voor verdere onderzoeken naar de totstandkoming van peer-assessments en uitspraken kunnen worden gedaan over de invloed van relaties op peer-assessments. Scholen in het voortgezet onderwijs krijgen inzicht in de invloed van relatie tussen peers op de uitvoering van peer-assessment en kunnen eventueel hun werkwijze aanpassen. Dat de resultaten geen significante invloed van relatie laten zien is mogelijk te wijten aan de grootte van de respondentengroep, samenstelling van de respondenten, het verkeerd interpreteren van de criteria door de respondenten. Goed vervolgonderzoek zou zich kunnen richten op meerdere afnamen in plaats van één afname en kwalitatief onderzoek zou meer inzicht kunnen geven in beweegredenen van respondenten om bepaalde beoordelingen te geven.

Inleiding

Binnen het hedendaagse onderwijs is samenwerkend leren niet meer weg te denken, evenals verantwoordelijkheid bij de leerlingen leggen voor hun eigen leerproces (Oostdam, Peetsma & Blok, 2007). Bij samenwerkend leren werken leerlingen samen met hun medeleerlingen ('peers') met als doel het samen construeren van kennis (Johnson & Johnson, 2009). Bij opdrachten die tijdens samenwerkend leren moeten worden volbracht, kan ook een beoordeling door peers worden ingebouwd: peer-assessment. Peer-assessment heeft meer dan één kenmerk, maar de meest belangrijke is dat leerlingen elkaar en/of elkaars werk beoordelen (Topping, 1998; Sluijsmans, 2002; Van Gennip, Segers & Tillema, 2010; Van Zundert, Sluijsmans en Van Merriënboer, 2010; Gielen, Dochy & Onghena, 2011). De groepsgenoten beoordelen elkaar met behulp van verschillende criteria die zijn opgesteld. Zo heeft de docent een betere kijk op het samenwerken binnen groepjes, maar leerlingen kunnen bewuster leren en met samenwerken bezig zijn (Van Gennip, Segers & Tillema, 2010). Doordat nagedacht moet worden over wanneer iemand goed of minder goed kan samenwerken, leren leerlingen ook naar zichzelf te kijken. Daarnaast geven de beoordelingen van groepsgenoten inzicht in wat anderen vinden van het groepslid. Zeker wanneer meerdere beoordelingsmomenten worden ingebouwd kan een leerling leren van de uitkomst van de peer-assessments (Gielen, Dochy & Onghena, 2011).

Net als alle andere soorten assessments moeten peer-assessments betrouwbaar en valide zijn (Greguras, Robie & Born, 2001). Volgens Fastré, Van der Klink, Sluismans & Van Merriënboer (2013) is het voor een betrouwbaar en valide peer-assessment van vitaal belang om criteria vast te stellen op basis waarvan beoordeeld wordt. Zo kan ieder groepslid op dezelfde manier beoordelen en beoordeeld worden (betrouwbaarheid) en zullen er in het ideale geval correcte weergaven van het samenwerkingsproces weergegeven worden (validiteit). Elke peer gaat op zijn of haar eigen manier met de assessment-criteria om en dat is mogelijk van invloed op de uitkomst van het peer-assessment (Fastré et al., 2013).

Er blijken daarnaast meerdere sociale variabelen van invloed te zijn op de uitkomst van een peer-assessment (Gielen et al., 2011). De relatie tussen assessor en assessee kan invloed hebben: het vergt vaardigheid om een objectieve beoordeling te geven, ondanks een goede relatie tussen beide personen. Gielen et al. (2011) wijzen ook op vertrouwen tussen peers: bij een groot vertrouwen is een beoordeling gemakkelijker positief dan bij weinig vertrouwen. Wanneer peers elkaar goed kennen en positieve gevoelens koesteren kan dit invloed hebben op de beoordeling. Andersom is natuurlijk ook mogelijk: bij overheersend negatieve gevoelens kan de uitkomst van het peer-assessment wellicht lager uitvallen dan hoe die werkelijk had moeten zijn. Evenals bij de interpretatie

van criteria kan dit leiden tot een validiteits- en/of betrouwbaarheidsprobleem, wat gevolgen kan hebben voor het gebruik van de assessments (Greguras et al., 2001; Phielix, 2012).

Bij samenwerken werken leerlingen met elkaar aan een eindproduct en bij peer-assessments beoordeelt men elkaar over ieders bijdrage tijdens het samenwerken. Hier spelen veel sociale processen een rol: deze beïnvloeden het verloop en de uitkomst van het samenwerken en de peer-assessments (Tollema, 2003). Sociale processen kunnen bijvoorbeeld vriendschap of vijandschap zijn, maar ook afhankelijkheid van elkaar (hoewel peer-assessments juist in onafhankelijkheid moeten worden uitgevoerd om de objectiviteit te bewaken), de mate van vertrouwen in elkaar als assessor en assessee, veiligheid binnen groepen, percepties van leerlingen en populariteit van de leerlingen (Van Gennip, Segers & Tillema, 2010).

Het doel van dit onderzoek is inzicht krijgen in deze sociale processen en de mogelijke invloed op de uitkomst van het assessment. Er is nog weinig onderzoek gedaan naar en dus weinig bekend over deze processen en hun invloed op peer-assessments. Met behulp van het Social Relations Model van Kenny (1994) kunnen de verschillende processen worden geanalyseerd. Door groepsleden elkaar te laten beoordelen op samenwerking bij een langlopende opdracht én hierbij een onderdeel zelfbeoordeling in te voegen kan geanalyseerd worden in hoeverre percepties van de groepsleden overeen komen. De uitkomsten van deze analyses kunnen implicaties met zich meebrengen voor het gebruik en de uitvoering van peer-assessments in de toekomst. In het theoretisch kader wordt de bij dit onderzoek behorende theorie uiteengezet, worden de factoren benoemd die belangrijk zijn voor dit onderzoek met betrekking tot de sociale processen achter peer-assessment, en zal de centrale vraag worden geformuleerd.

Wetenschappelijke en maatschappelijke relevantie

De relevantie van dit onderzoek is groot. Peer-assessment is een vorm van toetsing die al met regelmaat gebruikt wordt in het onderwijs, zowel het basis- als het hoger onderwijs, maar ook in het voortgezet onderwijs (Lindsay & Clarke, 2001). Daarnaast zijn voor zowel basis- als voortgezet onderwijs plannen gemaakt om het leren meer sociaal-constructivistische (Lagerweij & Lagerweij-Voogt, 2004) kenmerken te geven, waarbinnen peer-assessment een belangrijke rol moet gaan vervullen (Oostdam et al., 2007). Omdat peer-assessments ook moeten voldoen aan bepaalde eisen om valide en betrouwbaar te zijn, is het van belang dit te testen, net als bij andere soorten van assessment wordt gedaan (Nitko & Brookhart, 2011). Uitkomsten van dit onderzoek kunnen weer gebruikt worden voor nieuw onderzoek of voor het ontwikkelen van wetenschappelijk onderbouwde instrumenten voor peer-assessments, rekening houdend met de hierboven beschreven sociale processen.

Voor de maatschappij is dit onderzoek niet onbelangrijk omdat het hele onderwijsstelsel gestoeld is op assessments: met behulp van assessments wordt de beheersing van de kennis gemeten. Met het toenemen van samenwerking is het belangrijk dit te monitoren met behulp van bijvoorbeeld peer-assessments. Om een accurate meting te doen moeten de toetsinstrumenten adequaat zijn en voldoen aan eisen om een zo goed mogelijk assessment uit te voeren. Scholen in het voortgezet onderwijs kunnen met behulp van dit onderzoek bepalen of er aanpassingen in hun lesprogramma's moeten worden gedaan en of er genoeg rekening wordt gehouden met de invloed van relatie tussen peers. Zo nodig kan met behulp van training jongeren geleerd worden om objectieve beoordelingen uit te spreken.

Theoretisch kader

Sluismans (2002) noemt peer-assessment 'het proces waar individuen hun peer(s) evalueren en beoordelen', waarbij peers de medestudenten van de student zelf zijn. Assessment wordt gebruikt om leerlingen (en docenten) inzicht te verschaffen in hun kennis, vaardigheden en verbeterpunten (Van Gennip, Segers & Tillema, 2010). Het is niet de bedoeling dat peers hierbij de volledige controle krijgen over de beoordeling, maar dat ze participeren in het beoordelingsproces, waarbij ze dus inspraak hebben op het uiteindelijke resultaat. De docent houdt de uiteindelijke verantwoordelijkheid voor de becijfering (Sluismans, 2002). Het beoordelen van elkaars werk is een sociale activiteit waarbij peers elkaar feedback geven op hun presteren, zowel individueel als groepsmatig (Van Gennip, Segers & Tillema, 2009). Hiermee kan geoefend worden op verschillende vaardigheden, zoals kritisch denken, kritisch analyseren en reflectie (Birenbaum, 1996). Ook het in staat zijn om te beoordelen zelf is een belangrijke vaardigheid, die met het oog op de eisen van de huidige maatschappij steeds belangrijker wordt: steeds meer werkgevers zien dat als een belangrijke competentie (Johnson & Johnson, 2009).

Voor het uitvoeren van een valide en betrouwbaar assessment is eenduidigheid belangrijk. Om een betrouwbaar en valide assessment uit te voeren moeten duidelijke criteria zijn geformuleerd, zo ook bij peer-assessment (Nitko & Brookhart, 2011). Wanneer deelnemers aan het assessment dezelfde persoon verschillend beoordelen op bijvoorbeeld gedrag, zoals samenwerking, zorgvuldigheid en vriendelijkheid, kan het zijn dat criteria niet duidelijk genoeg waren. Daarom moeten de criteria zorgvuldig worden opgesteld (Nitko & Brookhart). Een andere mogelijkheid zijn uiteenlopende percepties van peers: de één interpreteert criteria anders dan de andere, bijvoorbeeld omdat criteria niet goed begrepen worden (Van Gennip et al., 2010).

Weinig onderzoek is uitgevoerd naar het verband tussen relaties tussen peers en de uiteindelijke beoordeling tijdens het peer-assessment. Dit is opmerkelijk, omdat het aannemelijk is dat zulke relaties een rol spelen (Van Gennip et al., 2010). Een simpel voorbeeld hiervan is de rol van

vriendschap: in alle menselijkheid vindt men het soms moeilijk om iemand waarmee een goede band is opgebouwd een slechte beoordeling te geven (Van Gennip et al., 2010; Phielix, 2012). Anderzijds kan een slechte band het, bewust of onbewust, vergemakkelijken om een slechtere beoordeling te geven dan wanneer men neutraal was. Het moge duidelijk zijn dat zulke invloeden in principe buitengesloten zouden moeten worden om een objectieve uitkomst te verkrijgen.

Er zijn studies die ANOVA's gebruikten om de betrouwbaarheid van peer-assessments aan te tonen (Magin, 2001; Friedman, Cox & Mahler, 2008). Hierbij wordt de belangrijke assumptie van onafhankelijkheid vaak geschonden (Kenny, 1994; Greguras et al., 2001): er is afhankelijkheid tussen peers (logischerwijs heeft de beoordeling van peer A effect op de beoordelingen die peer A over andere peers uitvoert). Bij 'traditionele' toetsen zoals ANOVA's en t-toetsen vormt deze afhankelijkheid een probleem. Hierdoor kunnen peer-assessment niet worden getoetst op betrouwbaarheid en validiteit en kunnen nooit generaliseerbare uitspraken worden gedaan. Het Social Relations Model is ontworpen om juist die afhankelijkheid te onderzoeken (Kenny, 1994).

Social Relations Model

Het Social Relations Model (SRM) geeft informatie die ontbreekt bij analysemethoden als ANOVA's, maar zeer bruikbaar is om meer te weten te komen over de totstandkoming van de beoordelingen van peers (Greguras et al., 2001). Met behulp van het SRM kan ook de invloed van relaties worden onderzocht (Greguras, Robie, Born & Koenigs, 2007). In verschillende onderzoeken is al gebruik gemaakt van het SRM (Greguras et al., 2001; Phielix, 2012), waarbij rechtstreeks verwezen wordt naar Kenny (1994), die het model ontworpen heeft. Bij SRM worden variantieanalyses gedaan, maar de assumptie van afhankelijkheid is hierbij niet meer belangrijk. Er wordt juist met nadruk gekeken naar die afhankelijkheid omdat daar veel interessante informatie uit te halen is.

SRM gebruikt nog steeds de varianties van beoordelingen, maar benadert deze anders dan een ANOVA. De variatie van de scores die behaald zijn op de assessments wordt in drie varianties opgedeeld: actorvariantie, partnervariantie, en relatie- of dyadische variantie. Voor een betrouwbare en valide assessmentuitkomst is een hoge variantie in testcores nodig, want een hoge variantie betekent dat er sprake is van spreiding van antwoorden tussen personen. Variantie ontstaat bij verschil in antwoorden en aangezien niemand in principe exact hetzelfde presteert of beoordeelt, moet er wel spreiding ontstaan (Reader Testtheorie, 2013; Nitko & Brookhart, 2013).

Actorvariantie ontstaat door de neiging van leerlingen om hun peers allemaal gelijk te beoordelen. Dit wordt ook wel het actoreffect genoemd, waarbij dan sprake is van *assimilatie* (Phielix, 2012). Volgens de testtheorie is dit niet wenselijk en moet de totale variantie zo min mogelijk bepaald worden door de actorvariantie (Kenny, 1994). Partnervariantie is de variantie die zo hoog mogelijk moet zijn: een hoge variantie betekent een spreiding in scores tussen groepsleden,

waarbij de meeste variantie in de assessmentscore dus toe te schrijven is aan het gedrag van een groepslid (Kenny, 1994). Het zou bijvoorbeeld betekenen dat één peer van alle anderen een hoge score krijgt, de tweede peer van alle anderen een gemiddelde score en de laatste peer van alle anderen een lagere score. De beoordelingen van de peers over anderen komen dan overeen, zodat er sprake is van *consensus*. Dit wordt ook wel het partnereffect genoemd. De relatie- of dyadische variantie ontstaat door de unieke relatie tussen de beide peers (Kenny, 1994). Wanneer peer A een goede beoordeling geeft aan peer B, dit andersom ook zo is en de andere peers geven peer B een minder goede beoordeling, is de relatie tussen beide peers zeer waarschijnlijk de oorzaak. Dit geeft geen betrouwbaar beeld van de werkelijkheid. De relatievariantie wordt ook wel het relatie-effect genoemd.

Naast bovenstaande mogelijkheden maakt SRM ook een schatting van de correlatie tussen de beoordelingen van een peer enerzijds en de beoordeling van degenen die beoordeeld zijn anderzijds. Bij een correlatie van $r = 1.00$ zou er een volledige overeenstemming zijn. Zo kan een beeld worden weergegeven van de percepties en kan onderzocht worden in hoeverre de percepties van de peers invloed hebben op het peer-assessment. Kenny (1994) en Greguras et al. (2001) noemen dit *reciprocity* en onderscheiden de *generalized* en de *dyadic reciprocity*. SRM geeft voor deze twee 'wederkerigheden' correlaties weer. *Generalized reciprocity* is de correlatie tussen actor- en partnereffecten. Wanneer peer A anderen als coöperatief beoordeelt wordt peer A door de groepsleden ook als coöperatief beoordeeld. De *dyadic reciprocity*, voor dit onderzoek vooral van belang, is de correlatie tussen de beoordelingen van peers onderling, afgezet tegen de beoordelingen van de andere peers. Wanneer peer A positief beoordeeld wordt op samenwerking door peer B, terwijl de andere peers dat niet doen zou er sprake kunnen zijn van het relatie-effect.

Dimensies van sociaal en cognitief gedrag

Bij beoordeling van groepsgenoten door andere groepsgenoten zijn vijf dimensies bepalend voor het uiteindelijke resultaat. Deze vijf dimensies zijn verkregen met behulp van wetenschappelijk onderzoek naar de belangrijkste en bepalende aspecten van samenwerken en interactie (Wubbels, Créton & Hooymayers, 1985; Bales, 1988; Kenny, 1994; Den Brok, Brekelmans & Wubbels, 2006). Het gaat hierbij om sociaal gedrag ('invloed', 'vriendelijkheid', 'coöperatie', 'betrouwbaarheid') en cognitief gedrag ('productiviteit'). Deze vijf dimensies, gevalideerd door Phielix, Prins & Kirschner (2009), zullen in dit onderzoek gebruikt worden, en in de Methode beschreven en geoperationaliseerd. Een uitgebreide omschrijving van deze dimensies is terug te vinden in Phielix (2012).

Onderzoeksvraag en hypothesen

Op basis van bovenstaande bevindingen wordt de volgende onderzoeksvraag gesteld:

In hoeverre zijn relaties tussen peers van invloed op de uitvoering van peer-assessments (in het voortgezet onderwijs)?

Om deze onderzoeksvraag te kunnen beantwoorden is een aantal deelvragen geformuleerd, waarbij onderscheid wordt gemaakt tussen de variantiecomponent en de correlatiecomponent. Dit leidt tot de volgende verdeling.

Variantie

Door te verklaren waar de variantie vandaan komt (actor-, partner- of dyadische variantie) kunnen conclusies worden getrokken over de betrouwbaarheid van peer-assessments (een belangrijke voorwaarde voor de validiteit (Phielix, 2012)). Hoewel de dyadische variantie in dit onderzoek de meest belangrijke is, moet eerst getest worden aan welke bronnen, namelijk actor, partner of relatie, de variantie wordt toegeschreven. Wanneer er geen dyadische variantie wordt gevonden zou geconcludeerd kunnen worden dat de relatie tussen peers geen invloed heeft. Verwacht wordt echter dat die dyadische variantie er wel is, maar dit moet wel worden ondersteund door de statistische analyse. De volgende deelvragen worden gesteld in dit onderzoek.

1. Ziet een peer alle andere peers als hetzelfde?

Bij deze deelvraag gaat het om de actorvariantie, die voor een betrouwbaar peer-assessment zo laag mogelijk zou moeten zijn. Wanneer de actorvariantie laag is zijn criteria waarschijnlijk duidelijk geweest en wordt een peer door elke andere peer ongeveer gelijk beoordeeld: persoonlijke interpretatie heeft geen grote rol gespeeld bij het uitvoeren van het peer-assessment. Wanneer de actorvariantie hoog is zal partnervariantie lager uitvallen en vertonen de peer-beoordelaars de tendens dat ze alle peers gelijk beoordelen: assimilatie.

2. Wordt degene die beoordeeld wordt door elke andere peer hetzelfde beoordeeld?

Deze deelvraag betreft de partnervariantie. Partnervariantie zou hoog moeten zijn voor een betrouwbaar peer-assessment, omdat de beoordeling door peers dan bepaald wordt door het cognitief en sociaal gedrag van de beoordeelde peer: goed gedrag is goede beoordeling en minder goed gedrag is minder goede beoordeling. Er is dan sprake van consensus tussen de peer-beoordelaars en zal een groter deel van de variantie verklaard kunnen worden door de partnervariantie.

3. Wijkt de beoordeling van peer A door peer B af van de beoordeling van peer A door alle andere peers?

Deze deelvraag geeft zicht op de invloed van de relatie tussen peers. Wanneer de beoordeling van peer A door peer B stelselmatig anders is dan de beoordeling die de andere groepsgenoten aan

peer A geven, zal de relatie tussen de peers invloed hebben gehad. Wanneer dit op gebied van verschillende beoordelingsgebieden (de eerder beschreven dimensies) gebeurt kan hier geen sprake zijn van een actor- of partnereffect, omdat het stelselmatig gebeurt. Dit wordt de unieke variantie genoemd, omdat deze alleen bij de combinatie van deze twee peers tot stand komt. Een grote unieke, dyadische variantie duidt op een grote invloed van relatie.

Correlatie

Naast het berekenen van de varianties is kan bij het SRM ook een correlatie uitgerekend worden waarmee verdiept wordt in de relatie tussen peers. Hierin wordt onderscheid gemaakt tussen de *generalized reciprocity* en de *dyadic reciprocity*. Beide termen worden in een aparte deelvraag gevat.

1. Is er een samenhang tussen de beoordelingen die twee peers over elkaar uitspreken?

Deze deelvraag beantwoordt het vraagstuk over de *dyadic reciprocity*, ofwel het verband tussen de beoordelingen van twee peers onderling. Met behulp van deze deelvraag kan de onderzoeksvraag beantwoord worden: wanneer er een correlatie wordt gevonden betekent dit dat de relatie tussen de twee peers effect heeft gehad op de uitvoering van het peer-assessment. Een hoge correlatie wordt bijvoorbeeld gevonden wanneer peer A alle andere peers minder goed beoordeelt op bijvoorbeeld samenwerking en peer B juist goed, en ook andersom: peer B beoordeelt alle andere andere peers minder goed dan peer A op samenwerking.

2. Is er een samenhang tussen de beoordeling door peer A van groepsgenoten en hoe groepsgenoten peer A beoordelen?

De laatste deelvraag betreft de correlatie tussen hoe een peer beoordeeld is en hoe diezelfde peer de andere peers beoordeelt. Wanneer peer A door iedereen goed beoordeeld wordt op productiviteit, beoordeelt peer A de andere peers dan ook goed op productiviteit? Het productieve gedrag van peer A ontlokt dan hetzelfde productieve gedrag bij de andere peers.

In de Methode zullen de respondenten beschreven worden, evenals de gebruikte instrumenten en de uitgevoerde analyses.

Methode

Onderzoeksdesign

Voor dit onderzoek is gebruik gemaakt van leerlingen van verschillende scholen, die langlopende (minimaal één dag) samenwerkingsopdrachten moesten uitvoeren in groepen van vier. Aan het eind van hun samenwerkingsopdracht vulden de participanten een vragenlijst in over de samenwerking. Om het Social Relations Model zichtbaar te kunnen maken werd een *Round Robin Design* gebruikt. Bij zo'n model kunnen de verschillende beoordelingen van peers onderling per dimensie in beeld worden gebracht (zie Figuur 1). De antwoorden op de vragenlijsten zijn met behulp

van SPSS geanalyseerd en getest op betrouwbaarheid en validiteit en vervolgens is met behulp van SOREMO de variantie opgedeeld.

Round Robin Design						
Perceiver/ Actor	Target/Partner					
	A	B	C	D	E	F
A	-	x	x	x	x	x
B	x	-	x	x	x	x
C	x	x	-	x	x	x
D	x	x	x	-	x	x
E	x	x	x	x	-	x
F	x	x	x	x	x	-

Figuur 1. Round Robin Design (matrix)
(<http://www.persoc.net/Toolbox/Designs>)

Deelnemers

Voor dit onderzoek kon er gebruik worden gemaakt van twee scholen, waarvan zes klassen. Dit gaf $N = 99$ leerlingen. De deelnemers werkten allen in groepen van vier tot zes leerlingen aan projecten die langer dan één les duurden, waarbij de nadruk lag op het werken in groepen. Er waren zeventien groepen van vier leerlingen, vijf groepen van vijf leerlingen en één groep van zes leerlingen. De deelnemers waren leerlingen van de eerste en tweede klas van het voortgezet onderwijs. Van alle deelnemers gingen 76 leerlingen (vijf klassen) naar het VMBO ($n = 76$) en 23 leerlingen (één klas) naar HAVO ($n = 23$). Enkele ingevulde vragenlijsten waren onvolledig, waardoor een aantal groepen niet kon worden ingevoerd. Sommige klassen hadden een relatief klein totaal aantal leerlingen in verband met door de school gemaakte keuzes.

Instrumenten

Tabel 1. Vijf dimensies van samenwerken

Schaal	Dimensies	Beschrijving
Sociaal gedrag	Invloed (<i>influence</i>)	De mate van invloed die een groepslid heeft op andere groepsgenoten, werkwijze, beslissingen, etc.
	Vriendelijkheid (<i>friendliness</i>)	Gedrag ten opzichte van anderen: hulpvaardig, flexibel en bijdragend aan positief groepsklimaat.
	Coöperatie (<i>cooperation</i>)	In hoeverre een groepslid kan samenwerken, samen denken over oplossingen en werkwijzen, en initiatief nemen tot nieuwe voorstellen.
	Betrouwbaarheid (<i>reliability</i>)	Het nakomen van afspraken, doen wat hij/zij moet doen en het werk op tijd afhebben.
Cognitief gedrag	Productiviteit (<i>productivity</i>)	Mate van bijdrage aan groepsproduct, aandragen van oplossingen, inzet voor groepsopdracht.

Vijf dimensies van samenwerken, naar Phielix (2012).

Voor dit onderzoek is gebruik gemaakt van de vragenlijst zoals gebruikt door Phielix (2012), bestaande uit vijf items. Er werd een onderverdeling gemaakt in de vijf dimensies zoals eerder beschreven. De operationalisatie van de dimensies is weergegeven in Tabel 1.

De vragenlijst bestond uit vijf vragen en elke vraag toetste één van de vijf dimensies, waarbij de peers elkaar moesten becijferen op hun prestatie tijdens de samenwerking, op zowel sociaal als cognitief gedrag. De vragenlijst is opgebouwd aan de hand van een 10-punts Likert schaal. Er is voor deze schaal gekozen omdat deze verdeling bekend is voor de leerlingen: hun prestaties worden met een cijfer van één tot tien beoordeeld. Zie Bijlage I voor de vragenlijst.

Procedure

Verschillenden scholen waren betrokken bij het onderzoek. Leerlingen werkten in groepsverband aan een samenwerkingsopdracht die langer dan één les duurde. Alle leerlingen beoordeelden op de vragenlijsten van Phielix (2012) de samenwerking met hun groepsgenoten. Die vragen zijn gebaseerd op de sociale en cognitieve dimensies zoals beschreven door Phielix (2012). Na het invoeren van de respons in SPSS kon deze worden ingevoerd in SOREMO. Op basis van de uitkomsten uit SOREMO konden de resultaten worden opgemaakt en geïnterpreteerd.

Analyse

Om de varianties in te delen is gebruik gemaakt van SOREMO. Dit is een applicatie die de herkomst van de variantie kan schatten en deze indeelt onder de actor-, partner- en dyadische variantie. Daarnaast rekent SOREMO de correlatie uit tussen beoordelingen die de respondenten geven over zichzelf en beoordelingen die anderen over die respondent geven (de *generalized* en *dyadic reciprocity*). De ingevulde vragenlijsten werden in SPSS ingevoerd, om ze zo in SOREMO te kunnen inlezen. Dit gebeurde telkens per dimensie, dus eerst alle beoordelingen voor 'invloed', daarna voor de volgende dimensies. Dit databestand werd als textbestand ingelezen in SOREMO, die de varianties en correlaties uitrekende. Voor actor- en partnervariantie werd ook gekeken naar significantie. Voor de correlaties werden ook t-toetsen uitgevoerd om significantie te kunnen berekenen. Aan de hand van de opgedeelde varianties en de correlatiematrix kunnen de varianties worden geïnterpreteerd en kan de eventuele invloed van relatie worden bepaald.

Resultaten

Om een zo nauwkeurig mogelijk beeld te geven van de gevonden resultaten, zal in de volgorde van de deelvragen worden gewerkt, zoals genoemd in het theoretisch kader. Daarom zal eerst kort worden ingegaan op respectievelijk de actorvariantie, partnervariantie en relatievariantie. De laatste twee deelvragen over *dyadic* en *generalized reciprocity* en vooral de vierde deelvraag over de *dyadic reciprocity* zijn het meest van belang en het interessantst voor de beantwoording van de

onderzoeksvraag van belang zijn. Daarom zullen deze twee deelvragen steeds iets uitgebreider worden besproken.

In Tabel 2 en 3 zijn de absolute en relatieve variantieverdelingen weergegeven, waar telkens naar verwezen zal worden. Voor de interpretatie van de resultaten is telkens gebruikt gemaakt van het manuscript van Kenny (1998) en Kenny, Deborah en Cook (2006). Tabel 3 geeft de relatieve varianties weer, waarvan in Tabel 4 een schatting van de betrouwbaarheid van actor- en partnereffecten wordt weergegeven (bij de resultaten hieronder zullen voorbeelden worden gegeven van de effecten). De betrouwbaarheid kan als effectgrootte worden gezien, waarbij de gangbare maatstaven voor effectgrootte gebruikt kunnen worden (Greguras et al., 2001). Omdat de groepsgroottes voornamelijk overeen komen, kan Pearsons correlatiecoëfficiënt gehanteerd worden als maatstaf voor de effectgrootte (Field, 2011). In de Conclusie & Discussie zullen de gevonden resultaten vervolgens worden samengevat en de implicaties van de onderzoeksuitkomsten worden besproken. Daarnaast zullen beperkingen van dit onderzoek worden besproken en mogelijkheden voor vervolgonderzoek worden genoemd.

Actorvariantie

Voor alle dimensies is de absolute actorvariantie significant groter dan nul ($p < .05$, eenzijdig). De relatieve actorvarianties voor de vijf dimensies liggen rond de 15 procent, met uitzondering van de dimensie 'betrouwbaarheid': die is 22 procent. Dit betekent dat er een aanzienlijk actoreffect is aangetroffen: leerling A heeft een aanzienlijke neiging om leerling B hetzelfde te beoordelen op de dimensies. Ook bij het beoordelen onderling worden peers gelijk beoordeeld: er treedt assimilatie op, wat de betrouwbaarheid van het peer-assessment niet ten goede komt. Het is voor dit onderzoek niet belangrijk om dieper op de actoreffecten in te gaan. Het betrouwbaarheidscoëfficiënt ligt tussen $r = .46$ en $r = .61$, wat duidt op een grote betrouwbaarheid van de gevonden (relatieve) varianties (Field, 2011).

Tabel 2. *Absolute variantieverdeling voor de vijf variabelen*

Variable	ABSOLUTE VARIANCE PARTITIONING		
	Actor	Partner	Relationship
Influence	83.75	235.02	253.79
Friendliness	83.01	238.40	286.52
Cooperation	87.83	239.97	249.88
Reliability	132.70	216.76	251.64
Productivity	86.47	235.49	241.51

Partnervariantie

De relatieve partnervarianties voor de vijf dimensies lopen uiteen van 36 tot 42 procent. Ook de partnervarianties voor de vijf dimensies zijn significant bevonden ($p < .05$). Het betrouwbaarheidscoëfficiënt ligt voor alle dimensies boven $r = .70$, wat op een grote effectgrootte duidt (Field, 2011),

dus een groot partnereffect. Dat houdt in dat de beoordeling van een leerling voor een aanzienlijk deel tot stand is gekomen door het sociaal en cognitief gedrag van de leerling. Zoals gezegd is er dan sprake van *consensus* tussen de peers over de beoordeling van hun medeleerling, wat duidt op een betrouwbaar peer-assessment, kijkende naar de partnervariantie. De hoge betrouwbaarheidscoëfficiënten versterken dit beeld. Net als bij de actoreffecten is het voor dit onderzoek niet van belang om dieper in te gaan op de partnereffecten.

Tabel 3. *Relatieve variantieverdeling voor de vijf variabelen*

Variable	RELATIVE VARIANCE PARTITIONING		
	Actor	Partner	Relationship
Influence	0.15*	0.41*	0.44
Friendliness	0.14*	0.39*	0.47
Cooperation	0.15*	0.42*	0.43
Reliability	0.22*	0.36*	0.42
Productivity	0.15*	0.42*	0.43

* $p < .05$: componenten significant groter dan 0.

Tabel 4. *Betrouwbaarheid van actor- en partnereffecten*

Variable	REALIBILITY OF EFFECT ESTIMATES	
	Actor	Partner
Influence	0.492	0.731
Friendliness	0.463	0.713
Cooperation	0.506	0.737
Reliability	0.606	0.716
Productivity	0.522	0.749

(Gebaseerd op gemiddelde groepsgrootte 4.30)

Dyadische variantie

Aangezien de actor- en partnervarianties samen telkens ongeveer 50 tot 60 procent bedragen, blijft er nog een groot deel over: de dyadische (relatie-)variantie loopt voor de vijf dimensies uiteen van 42 tot 48 procent, waarmee de relatiecomponent op zichzelf een flink deel van de totale variantie bedraagt. Naar aanleiding van de variantie-analyse door SOREMO kan dus worden gezegd dat de unieke relatie tussen peers invloed heeft op de totstandkoming van het peer-assessment.

Er ontstaat dyadische variantie als de beoordeling van peer A door peer B voor de verschillende dimensies telkens anders is dan de beoordeling die de andere groepsgenoten aan peer A geven. Deze variantie is uniek: alleen voor die personen onderling van toepassing en niet toe te schrijven aan actor- of partnervariantie. Bijna de helft van de totale variantie is in de analyse toegeschreven aan de invloed van relatie tussen peers. Deze kan natuurlijk zowel vriendschappelijk als 'vijandig' zijn. Er wordt door SOREMO voor de aan relatie toegeschreven variantie niet getest op

significantie. Over de samenhang tussen de beoordeling van peers onderling moet gekeken worden naar de *dyadic reciprocity*.

Dyadic reciprocity

Wanneer er geen *dyadic reciprocity* zou zijn, had SOREMO deze ook niet uitgerekend (Kenny, 1998), maar de uitkomsten zijn niet significant. SOREMO geeft de t-waarden, behorende bij de berekende correlaties. Er kwamen geen significante waarden uit. Ter verduidelijking van de correlatiematrix (Tabel 5) zijn deze t-waarden opgenomen, zie hiervoor Tabel 6. Er is dus weinig samenhang gevonden tussen beoordelingen, maar er is wel een kleine trend te zien in de samenhang tussen de beoordelingen die peers over elkaar uitspreken. In Tabel 5 is te zien hoe de gegeven beoordelingen tussen twee peers met elkaar correleren. De diagonaal met onderstreepte waarden geeft voor elke dimensie weer wat de invloed van relatie is in de totstandkoming van de beoordelingen. Die samenhang is niet groter dan $r = .10$. De dimensie 'coöperatie' (cooperation) heeft $r = .10$ *dyadic reciprocity*, wat betekent dat er een zwakke samenhang is tussen de beoordeling van peer A aan peer B op coöperatie en andersom.

Tabel 5. *Correlatie-matrix (dyadic reciprocity)*

	Influence	Friendliness	Cooperation	Reliability	Productivity
Influence	<u>0.082</u>				
Friendliness	0.016	<u>0.022</u>			
Cooperation	0.078	-0.006	<u>0.101</u>		
Reliability	0.015	-0.129	0.037	<u>0.093</u>	
Productivity	-0.013	-0.095	-0.014	0.008	<u>-0.058</u>

N.B. diagonaal is *dyadic reciprocity* (onderstreepte waarden).

Tabel 6. *T-test resultaten (dyadic reciprocity)*

	Influence	Friendliness	Cooperation	Reliability	Productivity
Influence	<u>0.8764</u>				
Friendliness	0.2343	<u>0.2562</u>			
Cooperation	0.8816	-0.1020	<u>1.2029</u>		
Reliability	0.2064	-2.1690	0.5161	<u>0.8240</u>	
Productivity	-0.1603	-1.1007	-0.1364	0.0855	<u>-0.6742</u>

N.B. $p \leq .05$, $Df = 22$.

Hoewel het correlatiecoëfficiënt voor productiviteit (*productivity*) erg klein is, valt wel op dat het een negatief getal is. Dit duidt erop dat wanneer peer A als minder productief wordt beoordeeld, peer B als productiever wordt beoordeeld. De coëfficiënten om de diagonaal heen laten zien wat het verband is tussen gedrag van peers: vertoont peer A betrouwbaar gedrag (*reliability*) ten opzichte van peer B, dan vertoont peer B minder vriendelijk gedrag (*friendliness*) ten opzichte van peer B. Bij positieve waarden is het beeld andersom: wanneer peer A bijvoorbeeld veel invloed (*influence*) heeft zal peer B zich vriendelijker opstellen (*friendliness*). Maar omdat de correlatiecoëfficiënten dusdanig

laag zijn kan er geen generaliseerbare uitspraak worden gedaan. Uit de waarden in Tabel 5 blijkt wel dat zowel negatieve als positieve relaties tussen peers invloed hebben. Dit komt bijvoorbeeld in uiting bij vriendschap en vijandschap.

Generalized reciprocity

Naast de *dyadic reciprocity* rekent SOREMO ook de *generalized reciprocity* uit (zie Tabel 7). Dit laat zien in hoeverre er samenhang is tussen de beoordeling van een groep peers over peer A, en de beoordelingen van peer A over de groepsgenoten. Ook hier werden geen significante correlaties gevonden aan de hand van de berekende t-waarden. Correlaties kunnen als effectgrootte worden geïnterpreteerd (Field, 2012). De correlaties met gemiddelde effectgrootte (in Tabel 7) wijzen op een tendens dat de beoordelingen van peer A door groepsgenoten samenhangt met de beoordeling van peer A over de groepsgenoten.

Tabel 7. *Correlatiematrix (generalized reciprocity)*

Actor (rij) en partner (kolom)	Influence	Friendliness	Cooperation	Reliability	Productivity
Influence	<u>-0.375</u>	-0.337	-0.333	-0.349	-0.292
Friendliness	-0.075	<u>-0.022</u>	0.095	0.050	0.202
Cooperation	-0.173	-0.051	<u>-0.101</u>	0.030	0.032
Reliability	0.088	0.113	0.018	<u>-0.001</u>	0.108
Productivity	-0.032	0.143	0.059	-0.100	<u>0.038</u>

N.B. diagonaal is *generalized reciprocity* (onderstreepte waarden)

Voor de variabele 'invloed' (*influence*) kan gezegd worden dat er een negatieve samenhang is tussen peer A en de resterende groep peers. Wanneer peer A veel invloed vertoont, gaan andere groepsgenoten minder invloed vertonen. Deze samenhang is gemiddeld tot sterk volgens Pearsons correlatiecoëfficiënt (Field, 2011). Voor elke variabele geldt dat de samenhang af te lezen is uit de diagonaal en dat dus alleen de dimensies invloed en coöperatie enige (negatieve) samenhang vertonen. Waar peer A invloedrijk en coöpererend gedrag vertoont, beoordeelt peer A de overige groepsgenoten duidelijk minder goed op deze dimensies. De correlatiematrix behorende bij de *generalized reciprocity* is niet symmetrisch: de correlaties tussen peer A en B enerzijds en peer B en A anderzijds komen niet overeen, zoals bij de *dyadic reciprocity* wel het geval is. Coöperatie van peer A doet de invloed van peer B afnemen ($r = -.17$), terwijl coöperatie van peer B de invloed van peer A sterker doet afnemen ($r = -.33$).

Conclusie & Discussie

Een redelijk hoge actorvariantie (15 tot 22 procent, $p < .05$) werd gevonden bij de analyse door SOREMO. Dit duidt op de aanwezigheid van een actoreffect: groepsleden vertonen de neiging om andere groepsleden allemaal op dezelfde manier te beoordelen. Op de eerste deelvraag (*ziet een*

peer alle andere peers als hetzelfde?) kan gezegd worden dat de respondenten de neiging hadden om elkaar hetzelfde te beoordelen, waarbij dus sprake is van assimilatie (Phielix, 2012).

De partnervariantie was hoog (36 tot 42 procent, $p < .05$), wat laat zien dat beoordelingen voor een aanzienlijk deel tot stand zijn gekomen door het sociale gedrag van de beoordeelde peer. Dit versterkt de betrouwbaarheid van het peer-assessment. Bij de analyse werd een opvallend hoger effect voor de dimensie 'betrouwbaarheid' gevonden. Dit is mogelijk te verklaren door het abstractieniveau van deze dimensie. Sommige respondenten vonden het moeilijk om precies te vatten wat elke dimensie inhield en hoe je dat kan beoordelen. Betrouwbaarheid is zo'n typisch abstracte dimensie waarbinnen vaak zwart-wit wordt gekeken: je bent betrouwbaar of je bent het niet (Borman, 1997). De tweede deelvraag (*wordt degene die beoordeeld wordt door elke andere peer hetzelfde beoordeeld?)* kan wederom bevestigend worden beantwoord. De respondenten beoordeelden elkaar, ondanks hun neiging om een actoreffect te vertonen, dus wel redelijk gelijk. De peers vertoonden consensus (Phielix, 2012) en waren eensgezind over de beoordelingen van groepsgenoten.

Wanneer de relatieve actor- en partnervariantie (Tabel 3) vergeleken worden met de resultaten van Greguras et al. (2001) valt op dat in dat onderzoek een hogere actorvariantie werd gevonden, wat erop wees dat een aantal beoordelaars hun peers telkens anders beoordeelden dan de rest van de beoordelaars. Dit zou kunnen betekenen dat in het huidige onderzoek de criteria beter geformuleerd waren, zodat er eenduidigheid bestaat voor de peers (Fastré et al., 2013). Een objectieve beoordeling aan de hand van goede criteria zorgt voor objectiviteit en helpt de beoordelaars om een objectieve beoordeling te geven over de prestatie van hun peers (Sluijsmans, 2002; Phielix, 2012). Kijkend naar de theorie in Kenny (1994) en Kenny, Deborah en Cook (2006) kunnen deze criteria de assimilatie tegengaan en de actoreffecten verminderen. Er is geen theorie bekend waarin aangegeven wordt wat een 'goede' hoeveelheid actorvariantie zou zijn, maar Phielix (2012) benoemt wel het feit dat de actorvariantie zeker niet de grootste component mag zijn. De actorvariantie is in dit onderzoek de kleinste component gebleken. In het onderzoek van Greguras et al. (2007) is die verdeling al meer overeenkomstig de resultaten die in dit onderzoek naar voren zijn gekomen. Actorvariantie was relatief de kleinste component, waarna voor de meeste beoordelingsconstructen die zij gebruikten de partnervariantie volgde en de relatiecomponent de grootste component vormde (Greguras et al., 2007).

De resultaten in het huidige onderzoek laten een invloed van relatie op de uitvoering van peer-assessments zien. Bij bestudering van de variantieverdelingen bleek dat de relatiecomponent bijna de helft van de uitkomst verklaarde. De volgende anekdote ondersteunt deze bevindingen: tijdens het afnemen van het onderzoek kwam een groepje respondenten terug om op te biechten

dat de vragenlijsten niet helemaal naar waarheid waren ingevuld, wegens een stoeve samenwerking met een groepslid. Dit typeert de uitkomsten van deze analyses: peer-assessments zijn, zeker bij jongeren, zeer vatbaar voor relationele invloeden. De resultaten van de analyse voor deelvraag drie (*wijkt de beoordeling van peer A door peer B af van de beoordeling van peer A door alle andere peers?*) wijzen dus op een afwijking van de beoordeling van peer A door peer B ten opzichte van de beoordeling van peer A door de andere peers in de groep. Relatie kan hier een oorzaak in zijn. De resultaten in dit onderzoek komt grotendeels overeen met de bevindingen van Greguras et al. (2007). Zij vonden het relatiecomponent voor de meeste beoordelingsconstructen als de grootste variantiecomponent.

Bij nadere bestudering van de invloed van relatie aan de hand van Tabel 5 (*dyadic reciprocity*) bleek echter geen grote samenhang tussen de beoordelingen van peers onderling (de vierde deelvraag: *is er een samenhang tussen de beoordelingen die twee peers over elkaar uitspreken?*). Wanneer die samenhang tussen twee peers groot was geweest, zou dat betekend hebben dat twee peers elkaar aanzienlijk beter of slechter beoordelen ten opzichte van de beoordelingen van overige groepsgenoten. In dat geval zou de invloed van relatie heel duidelijk zijn geweest. Deze uitkomsten komen niet overeen met de bevindingen van Greguras et al., (2007), want zij vonden wel enige significante correlaties. Daar viel overigens wel op dat er voor de meeste constructen geen significante correlatie voor de *dyadic reciprocity* werd gevonden. Zij vonden de significante correlaties op inter-personele constructen (*teamwork* en *affiliation*, wat verwantschap betekent) wat aan kan geven dat juist op de constructen waar veel inter-persoonlijk contact komt kijken, de relatiecomponent van grote invloed is. Het is interessant om te zien dat in het huidige onderzoek de dimensie coöperatie de meeste correlatie vertoonde: ook een interpersoonlijke dimensie. In dat opzicht vertoont dit onderzoek overeenkomst met dat van Greguras et al. (2007).

Voor dit onderzoek kunnen geen goed onderbouwde uitspraken worden gedaan over de precieze invloed van relatie tussen peers op de uitvoering van peer-assessments. Er kunnen wel voorzichtige conclusies worden getrokken en aanbevelingen worden gedaan en daarnaast lijkt er ruim draagvlak te zijn voor een groter onderzoek om de samenhang beter en beter in kaart te brengen.

De laatste deelvraag (*is er een samenhang tussen de beoordeling door peer A van groepsgenoten en hoe groepsgenoten peer A beoordelen?*) is vanuit de analyses met sterkere argumenten te beantwoorden. Hoewel de correlaties van de *generalized reciprocity* (Tabel 7) niet significant zijn, is bij veel variabelen toch een behoorlijke effectgrootte te vinden. Dat maakt het verantwoord om de gevonden waarden te interpreteren. Er is een samenhang gevonden tussen de beoordelingen die een groep peers over peer A van die groep uitspreekt en de beoordeling die peer

A over diezelfde groep peers uitspreekt. Wanneer peer A anderen als coöperatief beoordeelt, wordt peer A door de groepsleden ook als coöperatief beoordeeld. Waar dit onderzoek sterkere correlaties vond voor de *generalized reciprocity*, vonden ook Greguras et al. (2007) sterkere, soms significante correlaties. Het huidige onderzoek vond voor de grootste correlaties vooral negatieve samenhang: dat betekent dat als peer A anderen als productief ziet, dezelfde peer door de groepsgenoten als minder productief wordt gezien. Greguras et al. (2007) vonden juist vooral positieve correlaties voor de constructen die zij gebruikten. Gielen et. al (2011) wijzen op het belang van vertrouwen wanneer peer-assessments worden uitgevoerd, omdat bij een groter vertrouwen beoordelen sneller positief uitvallen. Wellicht was er onder de respondenten maar weinig vertrouwen in elkaar. De telkens negatief uitvallende correlaties kunnen hierdoor wellicht worden verklaard.

Als antwoord op de onderzoeksvraag (*n hoeverre zijn relaties tussen peers van invloed op de uitvoering van peer-assessments (in het voortgezet onderwijs)?*) kan gezegd worden dat relatie tussen peers invloed heeft op de uitvoering van peer-assessments. Kijkende naar bijvoorbeeld het onderzoek van Greguras et al. (2007) blijkt wederom dat relatie van invloed is, hoewel in dit onderzoek geen significante resultaten zijn gevonden. Er lijken echter ook andere variabelen van invloed te zijn, gezien het percentage onverklaarde variantie enerzijds (Greguras et al., 2001; Greguras et al., 2007; Phielix, 2012) en de discrepantie tussen de hoeveelheid dyadische variantie in en de niet significantie correlaties bij bestudering van de *dyadic reciprocity* in dit onderzoek.

Implicaties

Aan de hand van de bevindingen uit dit onderzoek kan gezegd worden dat relatie tussen peers van invloed is op de uitvoering van peer-assessments, in het voortgezet onderwijs. De data komt echter alleen uit de eerste en tweede klas en niet uit de bovenbouw. Ook het onderwijstype VWO is niet vertegenwoordigd in dit onderzoek. Het biedt wel mogelijkheden voor de toekomst: om een groter onderzoek uit te voeren, onder meer scholen, onder meer leerlingen van alle onderwijsniveaus en alle klassen. Uit dit onderzoek is gebleken dat er wel degelijk invloed is van relatie tussen peers op de uitvoering van peer-assessments, zeker als men kijkt naar de variantieverdeling (minimaal 45 procent relatiecomponent). Het lijkt nodig dit nader te onderzoeken, omdat peer-assessments steeds belangrijker worden in het voortgezet onderwijs en dus moeten peer-assessments ook een betrouwbare assessmentmethode zijn (Greguras et al., 2001; Oostdam et al., 2007).

Dit onderzoek is te klein om goede generaliseerbare uitspraken te kunnen doen over de invloed van relatie tussen peers op de uitvoering van peer-assessment. De resultaten duiden echter wel een richting: relatie heeft invloed en er is dyadische samenhang tussen de beoordelingen. Daarom kunnen wel voorzichtige adviezen worden gegeven. Voor de scholen waar peer-assessments

worden gebruikt moet veel aandacht worden besteed aan de training van leerlingen in het objectief uitvoeren van peer-assessments (Van Zundert, Sluijsmans & Van Merriënboer, 2010), voordat ze werkelijk plaatsvinden. De anekdote over het groepje leerlingen dat terugkwam om te vertellen dat ze niet helemaal eerlijk waren geweest in het beoordelen ondersteunt deze aanbeveling. Ook het investeren in vertrouwen is belangrijk, om beoordelingen niet lager te laten uitvallen dan ze in werkelijkheid zouden moeten zijn (Gielen et al., 2011).

Daarnaast zijn de overeenkomsten met Greguras et al. (2007) een interessante aanleiding voor onderzoek gericht op de invloed van alleen inter-persoonlijke variabelen op de uitvoering van peer-assessment. De correlatiecoëfficiënt van de inter-persoonlijke dimensie coöperatie was de hoogst correlerende dimensie voor de relatiecomponent, terwijl Greguras et al. (2007) soortgelijke variabelen als significante constructen vond. Het zou tevens een stap zijn in de richting van het verklaren van onverklaarde varianties bij eerdere onderzoeken (Greguras et al., 2001; Greguras et al., 2007; Phielix, 2012).

Beperkingen en mogelijk vervolgonderzoek

Een mogelijke beperking van dit onderzoek is de samenstelling van de respondenten. Dit zijn leerlingen uit de onderbouw van het voortgezet onderwijs en niet uit de bovenbouw. Oudere leerlingen hebben al vaker samengewerkt en kunnen dat wellicht beter dan leerlingen uit de onderbouw (Oostdam et al., 2007). Dit kan een vertekend beeld geven en biedt mogelijkheden voor grotere en bredere onderzoeken. Van Zundert (2012) heeft aangetoond dat leeftijd en de beheersing van de stof een voorwaarde is voor een effectief peer-assessment. Mogelijkerwijs was de doelgroep nog te weinig onderlegd of te jong om een goed peer-assessment te kunnen uitvoeren. Bij vervolgonderzoek zou een grotere groep respondenten kunnen worden gebruikt, met een betere spreiding over de jaarlagen en niveaus. De groep respondenten was voor dit onderzoek wat dat betreft mogelijk te eenzijdig.

In aansluiting op bovenstaande, kan de interpretatie van de criteria door de leerlingen een probleem zijn geweest. Deze zouden door de leerlingen verschillend kunnen zijn geïnterpreteerd. Er werd door de onderzoeker wel uitleg gegeven voorafgaand aan het peer-assessment, maar ook na die uitleg bleek regelmatig dat respondenten het nog erg moeilijk vonden om een ingewikkeld begrip als productiviteit of invloed te beoordelen. Ook bleek bij nadere bestudering dat veel deelnemers aan het onderzoek zichzelf structureel lager beoordeelden dan dat ze anderen beoordeelden. Dit kan te maken hebben met het verlaagde zelfbeeld dat pubers veelal hebben (Woolfolk et al., 2008), of bijvoorbeeld een gebrek aan zelfreflectie (Meijers, Kuijpers & Winters, 2009).

Daarnaast zijn de vragenlijsten overal maar op één moment afgenomen, omdat het planmatig niet mogelijk was om meerdere momenten bij dezelfde groepen in te plannen. Met behulp

van het Social Relations Model kan hier wel belangrijke informatie mee worden ingewonnen (Greguras et al., 2007). Bij vervolgonderzoek kan hier een slag in worden gemaakt om een completer beeld te geven. In dit onderzoek wordt alleen gekeken naar verbanden met behulp van de vijf dimensies van samenwerken (Phielix, 2012).

Zoals in vele onderzoeken die met het Social Relations Model zijn uitgevoerd, stuitte men telkens weer op een deel onverklaarbare variantie (Greguras et al., 2001; Greguras et al., 2007; Phielix, 2012). In dit onderzoek laat de analyse van de varianties zien dat veel variantie veroorzaakt wordt door de dyadische component. De bestudering van de *dyadic reciprocity* bevestigde dit beeld niet. Wellicht zijn er andere variabelen van invloed op de totstandkoming van peer-assessments, zoals ook bovenstaande onderzoeken ondervonden. Vervolgonderzoek zou zich specifiek daarop kunnen richten.

Kwalitatief onderzoek zou meer bloot kunnen leggen dat alleen met kwantitatief onderzoek mogelijk is. Hiermee kunnen beoordelingen die worden gegeven ondersteund en uitgelegd worden. Met alleen de kwantitatieve data zijn keuzes die gemaakt zijn door leerlingen niet goed te interpreteren. Gemakzucht is er één van: 'ik weet het niet meer hoor', werd weleens gezegd door respondenten. Na uitleg gingen ze dan wel weer door, maar het is de vraag of dat nog wel serieus gebeurde. Kwalitatief onderzoek kan helpen om de beoordelingen te begrijpen, maar ook om leerlingen te helpen begrijpen hoe ze moeten beoordelen.

Onderhevig aan interessante veranderingen op het gebied van onderwijs staan zowel de wetenschap als de praktijk voor de uitdagende taak om het onderwijs in te richten naar de nieuwe eisen (zoals bijvoorbeeld het implementeren van 21st Century Skills), met behulp van nieuwe methoden ten opzichte van traditioneel onderwijs. In dit onderzoek is geprobeerd om bij te dragen aan deze herinrichting van het onderwijs, al is het maar op bescheiden wijze. Maar ook kleine stappen op weg naar onderwijsvernieuwing mogen niet onderschat worden.

Referenties

- Bales, R. F. (1988). A new overview of the SYMLOG system: Measuring and changing behavior in groups. In R. B. Polly, A. P. Hare, & P. J. Stone (Eds.), *The SYMLOG practitioner: Applications of small group research* (pp. 319-344). New York: Preager.
- Borman, W. C. (1997). 360 ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7(3), 299-315.
- Brok, P. den, Brekelmans, M. & Wubbels, Th. (2006). Multilevel issues in studies using students' perceptions of learning environments: The case of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 9, 199-213.
- Birenbaum, M. (1996). Assessment 2000: Towards a Pluralistic Approach to Assessment. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 3-29). Boston: Kluwer Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NY: Lawrence Earlbaum Associates.
- Emans, B., Koopman, P., Rutte, C. & Steensma, H. (1996). Teams in organisaties: Interne en externe determinanten van resultaatgerichtheid [Teams in organizations: Internal and external determinants of outcome orientation]. *Gedrag en Organisatie*, 9, 309–327.
- Fastré, G.M.J., Klink, M. R. van der, Sluismans, D. & Merriënboer, J.J.G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, 38 (5), 611-630.
- Field, A. (2011). *Discovering Statistics Using SPSS (Third Edition)*. London: SAGE Publications Ltd.
- Friedman, B. A., Cox, P. L., & Maher, L. E. (2008). An expectancy theory motivation approach to peer assessment. *Journal of Management Education*, 32(5), 580-612.

- Gennip, N. A.E. van, Segers, M. S., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41-54.
- Gennip, N.A.E. van, Segers, M.S.R. & Tillema, H.H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning & Instruction, 20*. 280-290.
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137-155. doi:10.1080/02602930903221444
- Greguras, G.J., Robie, C. & Born, M. Ph. (2001). Applying the social relations model to self and peer evaluations. *Journal of Management Development, 20*(6), 508-525.
- Greguras, G. J., Robie, C., Born, M. P. & Koenigs, J. (2007). A Social Relations Analysis of Team Performance Ratings. *International Journal of Selection and Assessment, 15*(4), 434-448.
- Johnson, D.W. & Johnson, R.T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher, 38*. 365-379. DOI: 10.3102/0013189X09339057
- Kenny, D.A. (1994). *Interpersonal Perception: A Social Relations Analysis*. New York: Guilford.
- Kenny, D. A. (1998). SOREMO Version V. 2. *Unpublished manuscript, University of Connecticut, Storrs, CT*.
- Kenny, D.A. (2007). Estimation of the SRM Using Specialized Software. In: Kenny, D. A., & Livi, S. (2009). A componential analysis of leadership using the Social Relations Model. In F. J. Yammarino & F. Dansereau (Eds.), *Multi-Level Issues in Organizational Behavior and Leadership* (Vol. 8 of *Research in Multi-level Issues*; pp. 147-191). Bingley, UK: Emerald (see pages 174-183).
- Kenny, D.A., Deborah, A. K., & Cook, W.L. (2006). *Dyadic Data Analysis*. New York: Guilford.

Lagerweij, N., & Lagerweij-Voogt, J. (2004). *Anders kijken: De dynamiek van een eeuw onderwijsverandering*. Antwerpen/Apeldoorn: Garant.

Lindsay, C., & Clarke, S. (2001). Enhancing primary science through self and paired assessment. *Primary Science Review*, 68, 15-18.

Magin, D. J. (2001). A novel technique for comparing the reliability of multiple peer assessments with that of single teacher assessments of group process work. *Assessment & Evaluation in Higher Education*, 26(2), 139-152.

Meijers, F., Kuijpers, M., & Winters, A. (2009). Loopbaanbegeleiding en loopbaandialoog in het onderwijs.

Nitko, A.J., & Brookhart, S.M. (2011). *Educational Assessment of Students*, sixth edition. Amsterdam: Pearson.

Oostdam, R., Peetsma, T., & Blok, H. (2007). Het nieuwe leren in basisonderwijs en voortgezet onderwijs nader beschouwd: een verkenningsnotitie voor het Ministerie van Onderwijs, Cultuur en Wetenschap. *Amsterdam: SCO-Kohnstamm Instituut*.

Phielix, C. (2012). *Enhancing Collaboration through Assessment & Reflection, Doctoral Dissertation*. Verkregen via: <http://dspace.library.uu.nl/handle/1874/255570> (29/11/2013).

Sluismans, D. (2002). Student involvement in assessment. The training of peer-assessment skills. *Unpublished doctoral dissertation*. Verkregen via <http://hdl.handle.net/1820/1034> (29/11/2013).

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.

Topping, K. J. (2003). Self and peer assessment in school and university: reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards (55-87)*. Dordrecht, The Netherlands: Kluwer.

Masterthesis Walther van Mechelen (4005848)

Woolfolk, A., Walkup, V., & Hughes, M. (2008). *Psychology in Education*. Longman.

Wubbels, T., Créton, H.A. & Hoymayers, H.P. (1985). *Discipline problems of beginning teachers, interactional teacher behaviors mapped out*. Paper gepresenteerd tijdens de jaarlijkse vergadering van de American Educational Research Association, Chicago.

Zundert, M. van. (2012). *Conditions of peer assessment for complex learning*. Maastricht: Universitaire Pers Maastricht

Zundert, M van., Sluijsmans, D. & Merriënboer, J. van. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*. 270-279.

Bijlagen

Bijlage I Vragenlijst bij samenwerking

Bijlage I: vragenlijst over samenwerken

Naam:

Je hebt vandaag in groepjes samengewerkt en bent samen bezig een opdracht te maken, of op een andere manier iets te maken of doen. Samenwerken is goed: het wordt steeds belangrijker en het is nog helemaal niet zo makkelijk om goed samen te werken.

Wat niet zo vaak gebeurt is bekijken hoe jullie samenwerking is verlopen: dat wordt reflecteren genoemd. Terwijl reflecteren wel heel erg belangrijk is. Je leert dan welke dingen goed gingen, welke dingen minder goed gingen en je leert wellicht van groepsleden, want iedereen heeft een eigen werkwijze.

Het is de bedoeling dat je je groepsleden beoordeelt op vijf verschillende punten:

Invloed gaat over de positie die iemand binnen jullie groepje innam. Was je groepsleden iemand die veel beslissingen nam, leiderschap en initiatief toonde? Dan geef je een hogere score. Iemand die niet zoveel van zich liet horen, geef je een minder hoge score.

Betrouwbaarheid gaat over het vertrouwen dat je hebt in elkaar. Kun je er van op aan dat je groepsleden hun werk doen en rekent iedereen daarop en *komt iedereen dat ook na*, dan geef je een hogere score. Als je telkens niet weet of iemand wel iets doet voor de opdracht en zijn/haar spullen op orde heeft, dan geef je een lagere score.

Productiviteit gaat niet echt over hoe je samenwerkt, maar vooral over de kwaliteit van het werk dat iemand doet. Heeft je groepsleden goed werk afgeleverd, dan geef je een hogere score. Is de kwaliteit minder goed (bijvoorbeeld onvolledig, niet goed of te weinig) dan geef je een lagere score.

Vriendelijkheid heeft te maken met hoe iemand zich tegenover de anderen gedraagt. Toont hij/zij respect en is hij/zij vriendelijk? Dan geef je een hogere score. Mensen die hoog scoren op vriendelijkheid zijn vaak flexibel en zijn best bereid om zich aan te passen. Als iemand zich niet vriendelijk, aardig en respectvol opstelde geef je een lagere score.

Coöperatie is een moeilijk woord voor samenwerking. Vond je iemand behulpzaam en fijn om mee samen te werken, kon je ook echt samen aan de opdracht werken? Dan geef je een hogere score. Als iemand vooral met zijn eigen deel bezig was en niet goed kon overleggen om samen tot een eindproduct te komen, geef je een lagere score. Ook groepsleden die niet veel deden geef je een lagere score.

Beoordeel nu je groepsleden op de bovenstaande vijf kenmerken. Geef een cijfer van 1 tot 10. Schrijf eerst de namen van je groepsleden op en vul dan telkens het cijfer in dat je je groepsleden geeft. Doe dit voor ALLE groepsleden!

VUL NU OP DE ACHTERKANT JE BEOORDELINGEN IN!!!

Alle beoordelingen zijn anoniem: niemand komt dus te weten wat je hebt ingevuld. Voor het onderzoek is het belangrijk namen in te kunnen vullen.

Z.O.Z.

Let op: beoordeel ook jezelf! Dat lijkt gek, maar dit kan belangrijk zijn voor het onderzoek. Doe dat op precies dezelfde manier zoals je dat bij anderen doet.

Geef een cijfer van 1 – 10. Vergeet niet de namen van je groepsgenoten in te vullen.

Naam:						JEZELF
Invloed
Vriendelijkheid
Coöperatie
Betrouwbaarheid
Productiviteit

Klaar?

Heel erg bedankt voor je deelname!

Tijdpad

September 2014	Oppakken onderzoeksplan	
26 september 2014	Concept onderzoeksplan	Gesprek met Jeroen Janssen op UU.
20 oktober	Inleveren definitief onderzoeksplan	
November	Verzamelen data op scholen	Vragenlijst benodigd
December	Verwerken data, analyses uitvoeren	
... december	Bespreken data + analyses	Gesprek met J. Janssen op UU.
December & Kerstvakantie	Schrijven resultaten, conclusie, etc.	
5 januari	Inleveren concept thesis Inleveren abstract voor conferentie	
5-8 januari	Feedback concept thesis	
11 januari	Inleveren definitieve thesis	
21 januari	Presenteren masterthesis	Masterthesisconferentie