

Praktijkonderzoek

Formatief & summatief peer assessment

Hoe betrouwbaar is peer-assessment en heeft het effect op de motivatie en prestaties tijdens de lessen speerwerpen aan 5 VWO?

Mathijs van de Vorst

Studentnummer: 2146619

Klas: 4C

Leerarrangement: Praktijkonderzoek

Opleiding: Fontys Sporthogeschool

School van onderzoek: Mencia de Mendoza Lyceum

Begeleider praktijkonderzoek: Hans van Kruijsdijk

Datum: 19 juni 2013

Samenvatting

Beoordelen kan op vele verschillende manieren, maar wat is nu een effectieve maar betrouwbare manier in de lessen lichamelijke opvoeding (LO), waarbij de leerlingen actief bij de lessen betrokken zijn en tevens gezorgd wordt voor het behoud of zelfs vergroten van de motivatie? Er is onderzoek gedaan bij twee 5 VWO klassen binnen een lessenreeks speerwerpen van drie weken of het formatief en summatief beoordelen van leerlingen door elkaar (peer-assessment) betrouwbaar was en of dit leidde tot meer motivatie en hogere cijfers (betere prestaties). Uit dit onderzoek blijkt dat peer-assessment een sterke en betrouwbare didactische manier is om te beoordelen zowel bij de jongens als de meisjes, omdat cijfers van de leerlingen vrijwel overeenkomen met de cijfers van de docent. Er zijn verschillen ontdekt in wat jongens en meisjes aantrekkelijk vinden met betrekking tot peer-assessment en opvallend is dat de jongens van de interventiegroep die elkaar in de lessen formatief hebben beoordeeld gemiddeld beter presteerden dan de jongens van de controlegroep.

Door: Mathijs van de Vorst

Inhoudsopgave

1 Inleiding	4
2 Opzet van onderzoek.....	6
3 Resultaten.....	7
4 Discussie en conclusie	10
5 Aanbevelingen.....	11
Bronnenlijst.....	12
Dankwoord	13

1 Inleiding

De aanleiding voor dit onderzoek was de ontevredenheid die regelmatig te constateren was onder de leerlingen van het Mencia de Mendoza Lyceum te Breda met betrekking tot het beoordelen van de cijfers bij de lessen LO. Daarom is een onderzoek gestart om te meten of leerlingen meer gemotiveerd worden wanneer ze elkaar feedback moeten verzorgen aan de hand van beschreven beoordelingscriteria (formatief beoordelen) tijdens de lessenreeks speerwerpen en elkaar in de laatste les beoordelen met een cijfer op basis van die beschreven beoordelingscriteria (summatief beoordelen).

Om meer te weten te komen over het onderwerp is er een literatuuronderzoek gedaan naar verschillende aspecten van motivatie en beoordelen.

Motivatie en beoordelen?

Motivatie is datgene wat ieder mens tot een bepaald gedrag drijft (Geerlings & Van der Veen, 2001). Ieder individu heeft zijn eigen motivaties, waardoor mensen verschillen in hoe ze reageren in verschillende situaties. Dit hangt af van hetgeen hij belangrijk vindt. Interessant is het om te kijken wat in het onderwijs leidt tot het vergroten van de motivatie met betrekking tot het beoordelen. Volgens Geerlings & Van der Veen (2001) zorgt peer-assessment ervoor dat leerlingen zich meer zelfverantwoordelijk gaan voelen, waardoor ze gemotiveerder worden en dus beter hun best gaan doen om zichzelf te ontwikkelen. Schellekens (1989) beschrijft dat het summatief beoordelen zowel door de docent als door middel van peer-assessment zorgt voor een extra prikkel, waardoor leerlingen beter hun best gaan doen om te komen tot een goed eindresultaat en dus de motivatie stijgt. Echter wanneer een opdracht te moeilijk lijkt te zijn en niet op maat wordt aangeboden dan zal dit leiden tot demotivatie.

Wat is beoordelen?

Volgens Schellekens (1989) zijn er meerdere synoniemen van beoordelen. Onder beoordelen vallen ook de begrippen evalueren en toetsen. Beoordelen is het toekennen van een cijfer aan een bepaalde vaardigheid of doel wat is gesteld om de mate van bekwaamheid vast te stellen. Leerlingen kunnen formatief of summatief beoordeeld worden. Formatief beoordelen wil zeggen dat er feedback wordt gegeven op basis van de vooraf beschreven beoordelingscriteria. Leerlingen krijgen feedback op hetgeen wat ze doen en kunnen zichzelf op basis daarvan verbeteren. Er zit geen cijfer aan vast. Bij summatief beoordelen wordt wel een cijfer aan de prestatie gekoppeld. Dit gebeurt aan het eind van het leerproces.

Waarom beoordelen?

Brouwer (1996) beschrijft dat beoordelen ervoor zorgt dat leerlingen uitgedaagd worden om het maximale te bereiken, zorgt voor bewaking van een bepaald level van een vak (kwaliteit) en het beoordelen geeft de docent informatie over de kwaliteit van de afgelopen lessenreeks die de leerling heeft ondergaan. Het beoordelen binnen het vak LO heeft niet als doel om leerlingen te kwalificeren en selecteren voor eventuele vervolgopleidingen. Wel is het vak LO van belang bij het geven van overgangsadvisen binnen het Voortgezet Onderwijs (Vermolen, 2008).

Wat is goed beoordelen?

Volgens Valcke (2010) is beoordelen een toets die voldoet aan de volgende criteria:

- Een toets moet betrouwbaar zijn. Betrouwbaarheid wil zeggen dat het beoordelen niet beïnvloed wordt door toevalligheden. Een beoordeling is betrouwbaar, wanneer deze niet of minimaal verschilt in de uitkomst ongeacht wie de toets afneemt of op welk tijdstip de toets wordt afgenomen.
- Een toets moet valide zijn. Valide betekent dat men meet/beoordeelt wat men ook echt wil meten/beoordelen.
- Een toets moet transparant zijn. Transparant wil zeggen dat het voor de leerlingen duidelijk moet zijn waarop hij/zij wordt beoordeeld en het moet voor de leerlingen duidelijk zijn welke vaardigheden (of kennis) bij welk cijfer hoort (Valcke, 2010).

Wie beoordeelt?

De vraag wie beoordeelt kan verschillende antwoorden krijgen. Meestal is de betreffende docent de eerste en enige beoordelaar, maar uit onderzoek blijkt dat de methoden om leerlingen elkaar te laten beoordelen ook positief werkt. Peer-assessment heet de methode waarbij leerlingen elkaar beoordelen. Er is onderzocht dat deze methode positief werkt (Valcke, 2010). Het positieve effect van peer-assessment is dat de leerlingen samenwerken tijdens de lessen. Ebbens (1996) beschrijft dat samenwerkend leren positieve voordelen heeft.

Probleemstelling:

De probleemstelling die door middel van dit onderzoek wordt onderzocht is: Zijn de leerlingen uit 5 VWO van het Mencia de Mendoza Lyceum in staat om elkaar betrouwbaar summatief te beoordelen aan de hand van vaste (vooraf beschreven) beoordelingscriteria binnen een lessenreeks van drie weken speerwerpen? En daarnaast de vraag of de bekendheid met en het werken met de criteria in het geven van feedback (formatief beoordelen) motiverend werkt en of ook leidt tot betere prestaties.

2 Opzet van onderzoek

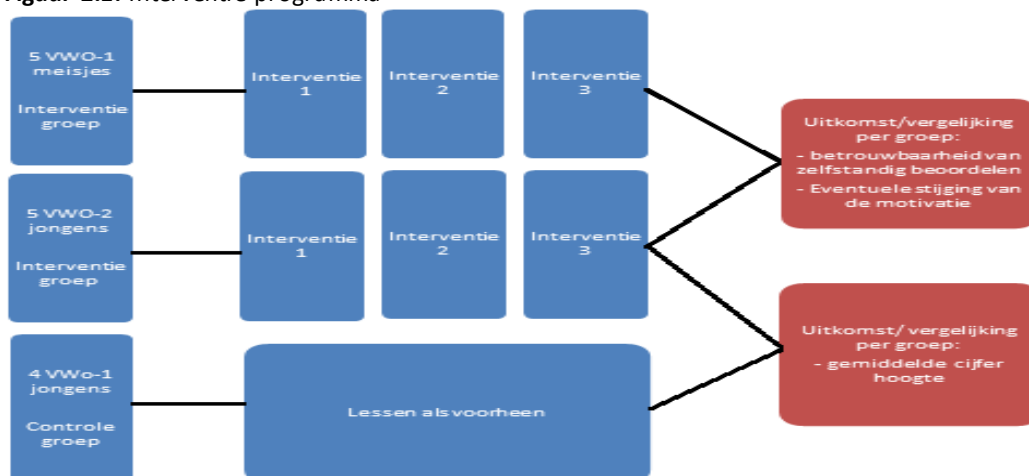
Het onderzoek heeft plaatsgevonden binnen een lessenreeks speerwerpen van drie weken aan twee 5 VWO klassen (een jongens klas en een meisjes klas); dit waren de interventiegroepen. Daarnaast was er als controlegroep een 4 VWO klas (jongens klas). Die controlegroep heeft een lessenreeks speerwerpen aangeboden gekregen zoals voorheen de lessen werden verzorgd, waarbij alleen de docent verantwoordelijk was voor het geven van feedback en de eindbeoordeling. De interventiegroepen zijn binnen het peer-assessment bezig geweest met het formatief en summatief beoordelen van elkaar. Om feedback te verzamelen over de lessenreeksen hebben de interventiegroepen na de laatste interventie les een vragenlijst van tien stellingen beantwoord en de controlegroep een vragenlijst van drie stellingen.

Beide interventiegroepen zijn in de eerste twee weken bezig geweest om 30 minuten per week elkaar formatief te beoordelen in tweetallen op basis van een uitgereikt beoordelingsformulier met vaste beoordelingscriteria voor zowel de techniek als afstand (zie bijlagen V). Door te werken met vaste beoordelingscriteria konden leerlingen elkaar gemakkelijker helpen en was het duidelijk wat leerlingen moesten doen om een beter cijfer te behalen. In week drie (de laatste interventie les) zijn de leerlingen 15 minuten bezig geweest met formatief beoordelen in tweetallen. Daarna zijn de leerlingen ingedeeld in groepjes van drie en hebben de leerlingen de tijd gehad om elkaar binnen 25 minuten summatief te beoordelen op basis van hetzelfde beoordelingsformulier voor zowel techniek als afstand (zie lesvoorbereidingen in bijlagen VI). Maar iedere leerling zou aan het einde van de lessenreeks volgens de vaksectie LO een voldoende moeten kunnen behalen als hij/zij zich tijdens de lessen actief inzet, want leerlingen zijn op het Mencia de Mendoza Lyceum te Breda niet toegelaten op basis van hun motorische vaardigheden, maar op hun cognitieve vermogens. Daarom werd ook de inzet beoordeeld. Die inzet werd beoordeeld door de docent en niet door de leerlingen, omdat er van uit wordt gegaan dat de docent de meest betrouwbare beoordeling hierover kan geven op basis van wat hij waarneemt in de lessen. Door het inzet cijfer te koppelen aan het cijfer voor techniek en voor de geworpen afstand zijn de mindere bewegers ook in staat geweest om een voldoende te behalen.

Onderzoeksopzet/ meetprocedure (zie ook Figuur 2.1):

- Betrouwbaarheid: Door het gemiddelde cijfer van beide interventiegroepen te vergelijken met het gemiddelde cijfer van de docent wordt er gekeken of leerlingen elkaar betrouwbaar kunnen beoordelen.
- Prestaties: Door de docent beoordeelde cijfers van de interventie 5 VWO-2 jongens klas te vergelijken met de docent beoordeelde cijfers van de controle 4 VWO-1 jongens klas wordt er gekeken of leerlingen beter presteren wanneer zij elkaar formatief hebben beoordeeld tijdens de lessen LO.
- Motivatie: Door het gemiddelde van stelling acht, negen en tien uit de vragen lijst (zie bijlagen) te vergelijken van beide interventiegroepen is er gekeken of de jongens en meisjes gemotiveerd werden door het formatief en summatief beoordelen en is het verschil tussen de jongens en meisjes waargenomen.

Figuur 2.1: Interventie programma

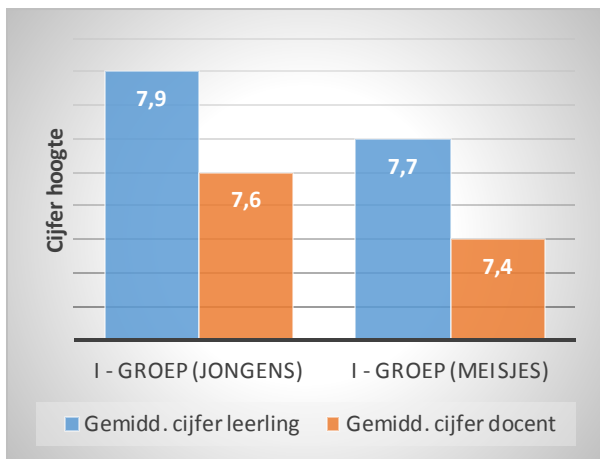


3 Resultaten

Een deel van de onderzoeksvraag was: Zijn de leerlingen uit 5 VWO van het Mencia de Mendoza Lyceum in staat om elkaar betrouwbaar summatief te beoordelen aan de hand van vaste (vooraf beschreven) beoordelingscriteria binnen een lessenreeks van drie weken speerwerpen?

Betrouwbaarheid:

Om de mate van betrouwbaarheid met betrekking tot het summatief beoordelen te meten werd gekeken naar het gemiddelde cijfer dat de leerlingen van de interventiegroepen elkaar hebben gegeven en het gemiddelde cijfer van de docent. Uit het onderzoek komt naar voren dat beide interventiegroepen (jongensgroep zowel als de meisjesgroep) elkaar gemiddeld 0,3 punt hoger geven dan de docent (Figuur 3.1).

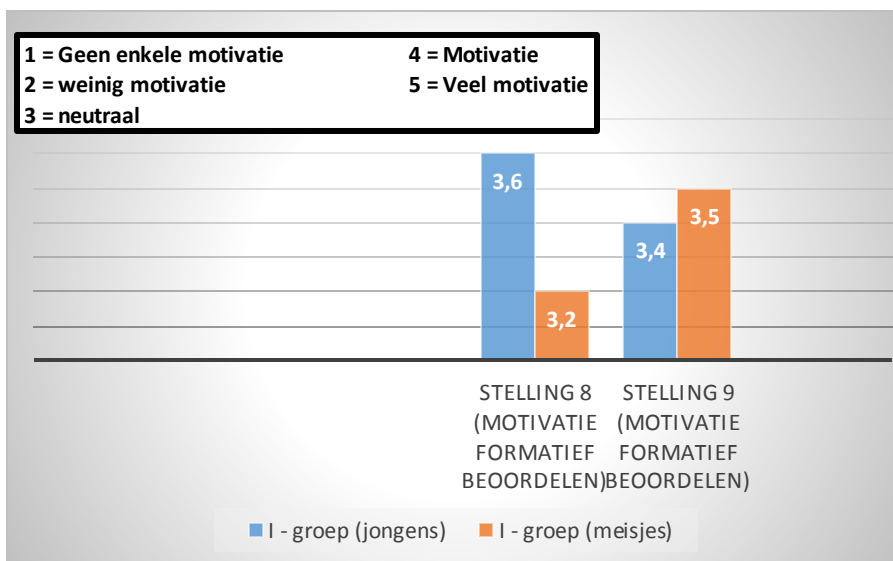


Figuur 3.1: Verschil jongens en meisjes in beoordelen met docent

Een ander deel van de onderzoeksvraag was: Werkte de bekendheid met en het werken met de criteria in het geven van feedback (formatief beoordelen) motiverend en of leidde het tot betere prestaties?

Motivatie met betrekking tot formatief beoordelen:

Beide interventiegroepen (jongens en meisjes) geven aan gemotiveerd te zijn geraakt om zich zelf tijdens de lessen meer als normaal te willen verbeteren. Dit blijkt uit stelling achten negen (Figuur 3.2), waarin wordt gevraagd of de leerlingen beter hun best zijn gaan doen om zich zelf te verbeteren, omdat zij de beoordelingscriteria kende en aan de hand hiervan feedback kregen.



Figuur 3.2: Gemiddelde motivatie bij formatief beoordelen

Prestaties:

Door de cijfers die zijn gegeven door de docent aan de interventie jongensklas en de controle jongensklas met elkaar te vergelijken komt naar voren dat de docent de interventiegroep met gemiddeld een 0,5 punt hoger beoordeelt dan de controlegroep.

Verder onderzochte resultaten:

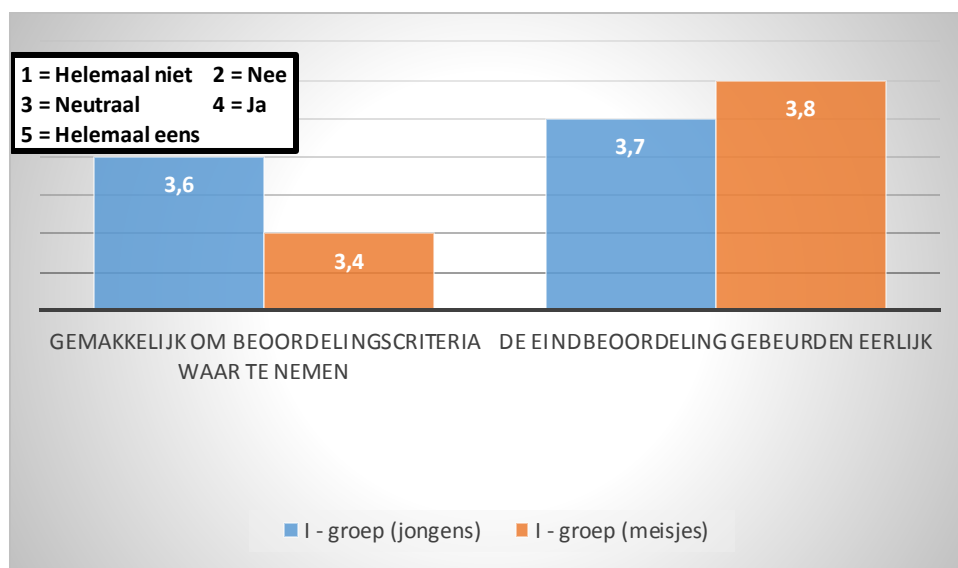
De leerlingen van beide interventiegroepen hebben de lessen door het formatief en summatief beoordelen als prettiger ervaren dan de controlegroep door gemiddeld het cijfer 3,7 te geven op een vijfpunt schaal tegen 3,4 van de controlegroep. Ook gaven zij een positiever oordeel met betrekking tot hun kennis van de beoordelingscriteria en hun tevredenheid met betrekking tot de gekregen hoeveelheid feedback (Figuur 3.3).



Figuur 3.3: Gemiddelde scores van beide interventiegroepen als controlegroep op de algemene vragen

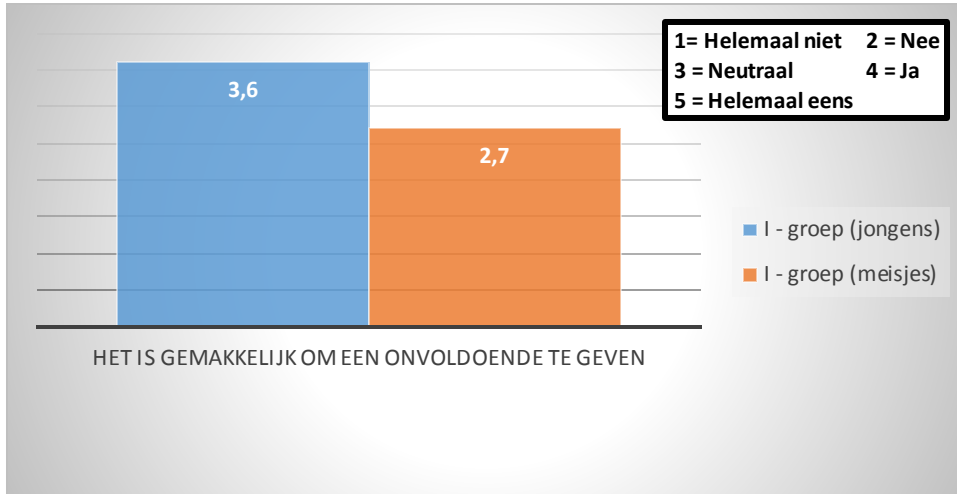
Betrouwbaarheid met betrekking tot summatief beoordelen:

Beide interventiegroepen gaven aan het gemakkelijk te vinden om te zien of iemand aan de beoordelingscriteria voldeed door gemiddeld 3,6 (jongens) en 3,4 (meisjes) te geven. Ook denken beide groepen elkaar betrouwbaar te hebben getoetst met betrekking tot het summatief beoordelen door gemiddeld 3,7 (jongens) en 3,8 (meisjes) te geven (Figuur 3.4).



Figuur 3.4: Gemiddelde scores van beide interventiegroepen met betrekking tot eerlijkheid bij het summatief beoordelen

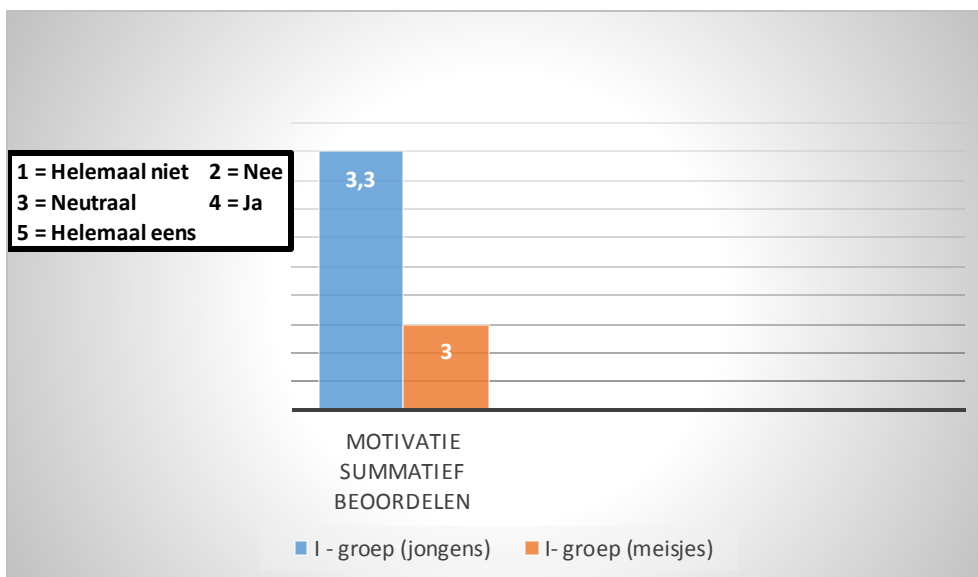
De jongens interventiegroep geeft aan het gemakkelijker te vinden om elkaar een onvoldoende te geven dan de meisjes interventiegroep. Dit blijkt doordat de jongens gemiddeld antwoord geven met het cijfer 3,6 en de meisjes met 2,7 (Figuur 3.5).



Figuur 3.5: Gemiddelde scores met betrekking tot het kunnen geven van een voldoende aan een klasgenoot

Motivatie met betrekking tot summatief beoordelen:

De jongens interventiegroep is meer gemotiveerd dan de meisjes interventiegroep om in de toekomst gebruik te maken van het elkaar summatief beoordelen volgens vaste beoordelingscriteria. Dit blijkt doordat de jongens gemiddeld het cijfer 3,3 geven en de meisjes het cijfer 3 (Figuur 3.6).



Figuur 3.6: Gemiddelde scores van beide interventiegroepen met betrekking tot de motivatie bij het summatief beoordelen

Verzorgen van feedback:

Door in de lessen te werken met vaste beoordelingscriteria zijn de leerlingen van de jongens interventiegroep meer bezig geweest met het verzorgen van feedback (tops en tips) naar elkaar als voorheen in de lessen. Dit blijkt doordat de jongens gemiddeld het cijfer 3,7 geven en de meisjes 3,6.

4 Discussie en conclusie

Betrouwbaarheid:

De jongens als meisjes van de interventiegroepen gaven ongeveer een drie tiende hoger cijfer dan de docent bij het summatief beoordelen met betrekking tot de techniek. Dit is een relatief klein verschil. Dit werd ook al eerder geconcludeerd in het onderzoek van Valcke (2010). Niet alle leerlingen gaven aan eerlijk te kunnen beoordelen als zijn of haar klasgenoot een onvoldoende zou verdienen. 50% van de meisjes kon dat niet en 18% van de jongens. Uit de beoordeling van de docent blijkt dat er geen enkele onvoldoende is behaald, waardoor de meisjes ook niet in de verleiding terecht zijn gekomen om van een werkelijke onvoldoende een voldoende te maken. De meisjes geven tegelijkertijd aan net als de jongens aan dat de beoordeling eerlijk/betrouwbaar is verlopen, wat ook blijkt uit de gegeven cijfers. Het zou kunnen zijn dat de meisjes vriendschap eerder boven eerlijkheid zouden verkiezen. Dat blijkt ook uit een onderzoek van Poullisse (2010).

Effect van peer-assessment op prestaties/beoordelingen:

De interventiegroep (jongens) presteerde gemiddeld 0,5 punt hoger dan de controlegroep (jongens). Zou dit komen doordat de jongens uit de interventiegroep meer feedback hebben gehad tijdens de lessen en de criteria goed kenden? Uit de lessen is gebleken dat de interventiegroep in staat was om elkaar in tweetallen van feedback te verzorgen op basis van de beschreven beoordelingscriteria. Dit blijkt uit de stelling zeven van de vragenlijst, waar werd gevraagd of leerlingen meer feedback zijn gaan geven naar elkaar door bekend te zijn met de vaste beoordelingscriteria. Daardoor hebben de leerlingen meerdere feedback momenten gekregen. De leerlingen van de controlegroep waren in de lessen erg ongemotiveerd door de minder aangename weersomstandigheden, waardoor er vaak moest worden opgetreden om de leerlingen actief bezig te houden, waardoor er minder feedback momenten zijn geweest om leerlingen individueel te verbeteren. Dit kan mede reden zijn geweest voor het gemiddeld lagere cijfer, omdat deze groep niet gebruik heeft kunnen maken van individuele feedback en de interventiegroep wel. Figuur 3.3 geeft ook weer dat de controlegroep gemiddeld lager scoort op de vragen of de leerlingen de criteria kenden en of zij genoeg feedback hebben gekregen.

Motivatie:

Uit de resultaten blijkt dat zowel de jongens als de meisjes meer gemotiveerd zijn geraakt tijdens de lessen door elkaar formatief te beoordelen. Ze gaven aan beter hun best te zijn gaan doen om zich zelf te verbeteren in de lessen als voorheen (Figuur 3.2). Daarnaast gaven de meisjes wel aan niet of weinig gemotiveerd te zijn voor het geven van cijfers aan elkaar volgens vaste beoordelingscriteria (summatief beoordelen). In de lessen gaven de meisjes de onderzoeker wel de indruk serieus bezig te zijn geweest met de beoordelingsformulieren en het verzorgen van feedback door elkaar bij iedere beurt hardop feedback te geven. De indruk bestond wel dat er onzekerheid was om uiteindelijk het juiste cijfer te geven, want zij vroegen tijdens het beoordelen vaak of ze wel het juiste cijfer gaven. Bij de jongens verliep dit beter. De meisjes gaven de indruk het ook niet leuk te vinden, omdat zij regelmatig overdreven aan het zuchten waren tijdens het beoordelen. De jongens namen sneller de verantwoordelijkheid en de meisjes moesten meer begeleid worden door de docent. Dit kwam in het experiment naar voren, doordat de meisjes veel vragen stelden en om veel bevestiging vroegen over of de techniek nu wel echt goed was. Terwijl dit bij de jongens beduidend minder het geval was. Volgens Harter (1998) zou dit kunnen komen, omdat jongens zelfverzekerder en onafhankelijker zijn en meisjes zijn meer gehoorzaam naar de docent.

Conclusies:

- Het summatief peer-assessment aan de hand van vaste beoordelingscriteria voor de techniek van het speerwerpen is betrouwbaar gebleken.
- De interventiegroepen gaven elkaar meer feedback en presteerden ook beter. Wanneer leerlingen meer feedback ontvangen leidt dit waarschijnlijk tot betere prestaties in een lessenreeks van drie weken.
- De methode van het peer-assessment door leerlingen elkaar formatief te laten beoordelen zorgt bij de jongens als meisjes voor extra motivatie.

5 Aanbevelingen

- Het is aan te raden om leerlingen te laten samenwerken in kleine groepjes (peer-assessment), waarbij leerlingen elkaar formatief beoordelen, waardoor leerlingen meerdere feedbackmomenten krijgen binnen een les en beter op de hoogte komen van de beoordelingscriteria wat waarschijnlijk bijdraagt aan een hogere motivatie en betere prestaties (Figuur 3.2 en 3.3).
- Wanneer een klas kundig wordt geschat door de docent om elkaar summatief te beoordelen dan is dit een gemakkelijke tool om lestijd te kunnen besparen, want er is geen garantie dat deze vorm van beoordelen ook als prettig wordt ervaren (dit blijkt ook het geval bij de meisjes interventiegroep geweest te zijn).
- Met betrekking tot de berouwbaarheid van het peer-assessment oordeel als docent altijd mee om eventueel een leerling te kunnen overrulen in het cijfer. Maak hierbij bijvoorbeeld de afspraak dat het cijfer niet meer mag schelen als één heel punt met het cijfer van de docent (dan wordt de leerling overruled).
- Wanneer een klas niet kundig wordt ingeschat met betrekking tot het summatief beoordelen van elkaar aan de hand van vaste beoordelingscriteria kan de docent er toch voor kiezen om in de lessen leerlingen formatief te laten beoordelen, waardoor er gebruik wordt gemaakt van activerende didactiek (bijvoorbeeld door mediakaarten).
- De docent dient eigenlijk altijd de criteria kenbaar te maken, waardoor leerlingen meer kennis op doen over de beoordelingscriteria en over de verwachtingen van de docent (transparantie), zodat leerlingen weten waar zij aan toe zijn.
- De binnen periode heeft meer voordelen als de buiten periode, want in de zaal kunnen de leerlingen het formulier met vaste beoordelingscriteria een vaste plaats geven, het formulier kan niet vies of nat worden en het formulier kan niet weg waaien (wat nu bij speerwerpen regelmatig gebeurden en wat storend was voor de les).

Bronnenlijst

Brouwer, C. (2009). *Opleiden, trainen en presenteren*. Baarn: Nelissen B.V.

Geerlings, T. & Veen, T. van der (2001). *Lesgeven en zelfstandig leren*. Assen: Gorcum B.V.

Harter, S. (1998). *The development of Self-Representations*. New York: Wiley

Poulisse, H. (2010). *Peer assessment in de lichamelijke opvoeding: Kunnen en willen leerlingen uit de vierde klas elkaar beoordelen bij het onderdeel hoogspringen?* Geraadpleegd op 10 mei 2013 via http://hbokennisbank.nl/nl/page/hborecord.view/?uploadId=fontys_didlmods%3Aoai%3Arepository.fontys.nl%3A25454

Schellekens, H.M.C. (1989). *Resultaatgericht doceren*. Groningen: Wolters-Noordhoff B.V.

Sluijsmans, D. (2008). *Betrokken bij beoordelen*. Tilburg: Mesoconsult

Valcke, M. (2010). *Onderwijskunde als ontwerpwetenschap: Een inleiding voor ontwikkelaars van instructie en voor toekomstige leerkrachten*. Gent: Academia Press

Vermolen, B. (2008). *Onderwijs met perspectief: deskresearch en aanbevelingen vanuit het onderwijsveld* (080619).

Dankwoord

Voor het mogelijk maken van de uitvoering voor het onderzoek wil ik graag het Mencia de Mendoza Lyceum te Breda bedanken waar het onderzoek heeft kunnen plaats vinden. Voor het verzorgen van tussentijdse feedback op dit onderzoek wil ik graag Hans van Kruijsdijk (begeleidend docent van de Fontys Sporthogeschool) bedanken.