

Onderzoeksartikel

Master Sport- en Beweginginnovatie

Is het zelfoordeel van basisschoolleerlingen een geschikte tool voor docenten bewegingsonderwijs?

Mattijs Mussert
www.gymladder.nl

Samenvatting

Het beoordelen van bewegingsonderwijs is al jaren een lastig probleem. Formatief beoordelen, met zelfbeoordeling als een van de belangrijkste componenten, wordt gezien als een mogelijke oplossing. Zelfbeoordeling bevordert het leren, geeft de leerlingen het gevoel van autonomie en zorgt voor een hogere kwaliteit. Ondanks de voordelen blijkt dat zelfbeoordeling nog niet veel wordt gebruikt. Een van de redenen hiervoor is dat de docenten twijfels hebben over de nauwkeurigheid en daarom de bruikbaarheid van de zelfbeoordeling. Een kwantitatief onderzoek is uitgevoerd onder 90 leerlingen van een basisschool. Er is gekeken of er een verschil is tussen het oordeel van de docent en het zelfoordeel van de leerling bij de activiteiten badminton en kastspringen. Uit de resultaten is gebleken dat leerlingen zichzelf hoger inschatten dan de docent zou doen. Er is een sterk positief verband gevonden tussen de beoordeling van de docent en de zelfbeoordeling van de leerling. Hieruit wordt geconcludeerd dat zelfbeoordeling een geschikte tool is voor docenten bewegingsonderwijs mits er rekening wordt gehouden met de overschatting van de eigen vaardigheid door de leerlingen. Het geslacht en de leeftijd van de leerlingen hebben geen significante invloed gehad op het gemeten verschil tussen docent-oordeel en het zelfoordeel van de leerling. Daarnaast is er door de leerlingen een zelfbeeld vragenlijst (CBSK) ingevuld. Leerlingen die zichzelf in de vragenlijst hoger ingeschat hebben blijken zichzelf ook hoger te beoordelen met de zelfbeoordeling.

Sleutelwoorden: Beoordelen, zelfbeoordeling, bewegingsonderwijs, zelfbeeld

Inleiding

Het beoordelen van leerlingen in het vak bewegingsonderwijs is een van de meest beladen en lastige problemen waar docenten bewegingsonderwijs de afgelopen veertig jaar mee te maken hebben gehad (López-Pastor et al., 2013). Dit terwijl evalueren van leerlingen, samen met pedagogiek en het leerplan, het fundament is van goed bewegingsonderwijs (Penney et al., 2009). De kwaliteit van beoordelen in het bewegingsonderwijs is niet gewaarborgd, Van Dokkum (2011) geeft aan: "eigenlijk doen we maar wat..." (Borghouts et al., 2017; Savelkoul, 2015; Van Dokkum, 2008, 2011). Savelkoul (2015) concludeert in haar onderzoek naar beoordeling van bewegingsonderwijs op de basisschool dat er nauwelijks sprake is van kwaliteitsnormen op het gebied van de beoordeling. Hoewel het beoordelen van bewegingsonderwijs op dit moment onvoldoende is, is wel duidelijk waar een goed beoordelingssysteem aan moet voldoen. Dit bestaat uit vijf normen; een helder doel, duidelijke normen, goed ontwerp, effectieve communicatie en leerling betrokkenheid (Melograno, 2007).

Bij het beoordelen in het bewegingsonderwijs op de basisschool wordt vooral op een summatieve manier beoordeeld. Door middel van een cijfer wordt aangegeven wat het behaalde niveau van de leerling is. De prestatie van de leerling wordt vergeleken met klasgenoten of met criteria en normen die voor iedere leerling gelijk zijn (Krijgsman & Maes, 2015). De afgelopen jaren is er kritiek ontstaan over deze manier van beoordelen. Er blijkt geen goede afstemming te zijn tussen de leerdoelen, de leeractiviteiten en de manier van evalueren (AIESEP, 2020; Borghouts et al., 2017; Weeldenburg & Slingerland, 2019). De internationale vereniging voor bewegingsonderwijs, AIESEP, stelt dat betekenisvol bewegingsonderwijs altijd formatieve beoordeling zou moeten bevatten (2020). Beoordelen is formatief als de informatie over de prestaties van de leerling wordt gebruikt om betere beslissingen te nemen over de vervolgstappen in het leerproces (Black & Wiliam, 1998). Uit onderzoek blijkt dat docenten bewegingsonderwijs formatief beoordelen wel belangrijk vinden, maar dat het nog niet uitgevoerd wordt (Borghouts et al., 2017; Panadero et al., 2014). Formatief beoordelen sluit beter aan bij de vijf kwaliteitsnormen van Melograno omdat de doelen en normen meer duidelijk worden en leerlingen meer betrokken worden (Melograno, 2007). Formatief

beoordelen is een manier om beoordelen te verbeteren (Borghouts & Slingerland, 2014; Krijgsman & Maes, 2015; Weeldenburg, Borghouts, & Slingerland, 2015). Zelfbeoordeling is een van de belangrijkste elementen van formatief beoordelen (AIESEP, 2020; Brown et al., 2015; Panadero et al., 2014)

Zelfbeoordeling

Het doel van formatieve zelfbeoordeling is het genereren van feedback die het leren en verbetering in resultaten stimuleert (Andrade, 2019). Voor leerlingen van alle leeftijden geldt dat zelfbeoordeling het leren verbetert, het de leerlingen het gevoel geeft dat ze controle hebben over hun beoordeling, het de autonomie ontwikkelt, de kwaliteit van het werk verbetert en zorgt voor een beter begrip (Papanthymou & Darra, 2019). Zelfbeoordeling betreft leerlingen in het denken over de kwaliteit van hun eigen werk in plaats van te vertrouwen op de docent als enige bron van evaluatie (Andrade & Valtcheva, 2015). Brown en Harris (2013) bevestigen dat zelfbeoordeling de schoolresultaten en het leren van de leerlingen verbetert. Ook bevordert zelfevaluatie de motivatie (McMillan & Hearn, 2008).

Motivatie

Volgens de zelfdeterminatietheorie (Ryan & Deci, 2000) bepalen autonomie, betrokkenheid en competentie een goede ervaring van de lessen bewegingsonderwijs (Haerens, n.d.). Als er tijdens de les voldaan wordt aan de behoefte aan autonomie, betrokkenheid en competentie zal de autonome motivatie toenemen (Niemi & Ryan, 2009). Indien niet aan deze behoefte voldaan kan worden, kan er een behoeftefrustratie ontstaan (Vansteenkiste & Ryan, 2013).

In de lessen bewegingsonderwijs moet vrijheid zijn om eigen keuzes te maken en moet men initiatiefnemer zijn van het eigen gedrag (behoefte aan autonomie). Daarnaast moet het bewegingsonderwijs de mogelijkheid bieden om te oefenen en om capaciteiten te laten zien (behoefte aan competentie). Tot slot is binnen de lessen behoefte aan verbondenheid. Dit verwijst naar zich geliefd en verbonden voelen door en met anderen. Als (vak)docent kan men op deze drie behoeften inspelen zodat voor een positieve ervaring met sport en bewegen wordt gezorgd.

(Collard, n.d.)

De inzet van zelfbeoordeling zorgt dat leerlingen meer autonomie en betrokkenheid ervaren (AIESEP, 2020). Papanthymou en Darra concluderen in hun review dat het gebruik van zelfbeoordeling leerlingen in het basisonderwijs motiveert (2019).

Leerresultaten

Ross, Hogaboam-Gray en Rolheiser (2001) beschrijven studieprestaties als een product van doelen en inzet. Bij zelfbeoordeling observeren, beoordelen en waarderen leerlingen hun bereikte niveau. Deze perceptie draagt bij aan hun verwachting dat ze soortgelijke taken in de toekomst ook succesvol kunnen uitvoeren. Deze verwachting heeft invloed op de doelen en de inzet van de leerlingen. Leerlingen die verwachten dat ze succesvol zullen zijn, stellen hogere doelen voor zichzelf en blijven doorzetten bij moeilijkheden (Bandura, 1997).

Als zelfbeoordeling wordt ingezet in het bewegingsonderwijs valt daarom te verwachten dat de leerlingen een hogere vaardigheid weten te halen (Ní Chrónín & Cosgrave, 2013; Panadero et al., 2014; Peyton, 2017).

Ondanks de genoemde voordelen blijkt dat zelfbeoordeling in de praktijk nog niet op grote schaal wordt ingezet (Panadero et al., 2014). De belangrijkste reden van docenten om geen zelfbeoordeling

te gebruiken is de twijfel over de nauwkeurigheid van zelfbeoordeling of de consistentie tussen het oordeel van de leerling en het oordeel van de docent (Brown & Harris, 2013; Panadero et al., 2014; Ross, 2006). Uit diverse onderzoeken blijkt dat leerlingen de neiging hebben tot overschatten van hun prestaties (Brown & Harris, 2013; Minelgaite et al., 2019).

Om zelfbeoordeling in het primair onderwijs in te zetten voor het vak bewegingsonderwijs is het van belang een beeld te hebben van de nauwkeurigheid. Verreweg de meeste onderzoeken naar zelfbeoordeling zijn gedaan in het hoger en middelbaar onderwijs (Wong, 2016). Andrade (2019) schrijft in haar review dat bij onderzoeken naar de nauwkeurigheid van zelfbeoordeling vaak naar voren komt dat oudere leerlingen zichzelf nauwkeuriger beoordelen. Daarnaast scoren de leerlingen met een hoger niveau beter dan de leerlingen met een lager niveau. Jongens blijken zichzelf vaker te overschatten dan meisjes. Er is gebrek aan onderzoek naar zelfbeoordeling in het primair onderwijs (Papanthymou & Darra, 2019).

Dat jonge kinderen zichzelf overschatten lijkt niet bewust te zijn, maar een gevolg van de cognitieve beperking van het ontwikkelingsstadium waar het kind in zit (Boseovski, 2010; Piaget, 1976). Naarmate kinderen zich ontwikkelen, worden hun cognitieve processen geavanceerder en minder vatbaar voor overschatting, wat resulteert in een meer nauwkeurige zelfbeoordeling (Keane & Griffin, 2015). Kinderen in het primair onderwijs zitten in de concreet operationele fase, deze fase loopt vanaf het zevende levensjaar tot elf- of twaalfjarige leeftijd. In deze fase worden de kinderen steeds minder beïnvloed door 'wishfull thinking', hetgeen een positieve invloed heeft op de nauwkeurigheid van de zelfbeoordeling (Butler, 1990).

Het zelfbeeld is de manier waarop iemand naar zichzelf kijkt. Het is het geheel van gedachten, ideeën en oordelen dat iemand over zichzelf heeft. Het algemeen en motorisch zelfbeeld wat iemand heeft kan van invloed zijn op de zelfbeoordeling van motorische vaardigheden. Ehrlinger (2008) geeft aan dat het zelfbeeld een mogelijke bron is voor fouten in zelfbeoordeling. Met name bij een lager zelfbeeld schat men zichzelf lager in (Ehrlinger, 2008)

Van alle onderzoeken naar zelfbeoordeling zijn er maar weinig gedaan in het vak bewegingsonderwijs. Deze onderzoeken in het bewegingsonderwijs zijn vooral gericht op de mogelijke voordelen en niet op de nauwkeurigheid van het zelfbeoordelen. Peyton (2017) beschrijft dat in het secundair onderwijs zelfbeoordeling in het vak bewegingsonderwijs helpt om toekomstige vaardigheden te verbeteren en zorgt voor een betere motivatie en zelfeffectiviteit. Uit de review van Papanthymou en Darra (2019) blijkt dat de voordelen van zelfbeoordeling in cognitieve vakken hetzelfde zijn voor alle leerlingen van primair en secundair onderwijs. In deze studies worden verhoogde motivatie en een verbetering in het leren en de prestaties genoemd. Het is daarom aannemelijk dat de voordelen voor het bewegingsonderwijs die Peyton (2017) in het secundair onderwijs heeft gevonden ook gelden voor het primair onderwijs.

Het hoofddoel van dit onderzoek is om inzicht te krijgen in de nauwkeurigheid van zelfbeoordeling bij het vak bewegingsonderwijs op de basisschool. Om de nauwkeurigheid te bepalen moet er gekeken worden naar de consistentie tussen docent-oordeel en leerling-oordeel. De verwachting is dat de leerlingen zichzelf hoger inschatten dan de docent.

Uit de review van Andrade blijkt dat bij leerlingen in het secundair onderwijs leeftijd een rol speelt bij de nauwkeurigheid van zelfbeoordeling (2019). Ook blijkt dat jongens zichzelf vergeleken met meisjes vaker hoger inschatten dan de docent. Bij leerlingen op de basisschool is de verwachting dat het zelfbeoordelen nauwkeuriger wordt naarmate de leerlingen ouder zijn. In dit onderzoek wordt de afwijking van leerlingen in groep zes met de afwijking van leerlingen in groep acht vergeleken.

Daarnaast wordt de afwijking van de meisjes ten opzichte van de docent met de afwijking van de jongens vergeleken.

In het onderzoek worden twee motorische vaardigheden ingezet om de nauwkeurigheid van zelfbeoordeling te bepalen. Deze vaardigheden verschillen in moeilijkheid van het zelfbeoordelen. De resultaten van beide onderdelen worden met elkaar vergeleken.

Naast de nauwkeurigheid van zelfbeoordeling is ook onderzocht of het zelfbeeld van de leerling invloed heeft op de zelfbeoordeling en de nauwkeurigheid van de zelfbeoordeling.

Methode

Om meer inzicht te krijgen in de consistentie tussen docent-oordeel en leerling-zelfoordeel is een kwantitatief experiment uitgevoerd. Voor het onderzoek zijn de activiteiten kasthurkspringen en badminton gebruikt. Met behulp van een rubric hebben de leerlingen hun eigen beweegniveau bepaald. In de rubric kunnen de leerlingen de opeenvolgende stappen zien van de leerlijn. De docent en een tweede beoordelaar hebben onafhankelijk van elkaar ook een oordeel over het beweegniveau van de leerlingen gegeven. Om inzicht te krijgen in de invloed van het zelfbeeld van de leerling is een zelfbeeld vragenlijst gebruikt.

Populatie

De onderzoeksgroep bestond uit 90 leerlingen (47 jongens en 43 meisjes) van groep 6 en groep 8 van een Daltonschool uit Zuid-Holland. De gemiddelde leeftijd was 10,45 jaar, variërend van 9 tot 13 jaar. In tabel 1 is een overzicht van de leerlingen per klas en het onderzochte onderdeel.

| Klas | Leerlingen | Jongens | Meisjes | Onderdeel |
|---------------|-------------------|----------------|----------------|------------------|
| 8A | 17 | 11 | 6 | Kasthurksprong |
| 8B | 25 | 12 | 13 | Badminton |
| 6A | 23 | 11 | 12 | Kasthurksprong |
| 6B | 25 | 13 | 12 | Badminton |
| totaal | 90 | 47 | 43 | |

Meetinstrument

De nauwkeurigheid van de zelfbeoordeling wordt gemeten door het verschil tussen het docent-oordeel en het leerling-oordeel te nemen. Zowel de leerling als de docent geeft een oordeel over de vaardigheid op basis van een rubric met 10 niveaus.

Vijf verschillende docenten bewegingsonderwijs hebben een bijdrage gehad in het tot stand komen van de rubrics. Deze docenten hebben ieder een 10-stappen plan voor de vaardigheden gemaakt. De onderzoeker heeft deze plannen samengevoegd en hier de rubric op gebaseerd. De rubric is vervolgens naar de docenten teruggestuurd om nogmaals feedback te geven. De niveaus zoals beschreven in de rubric corresponderen met de voortgang van de leerling in de leerlijn van die betreffende vaardigheid. Hierbij wordt uitgegaan dat niveau 1 haalbaar is voor leerlingen in groep 3 en niveau 10 uitdagend is voor de leerlingen van groep 8.

Het zelfbeeld van de leerlingen is gemeten met behulp van de CompetentieBelevingsschaal voor kinderen (CBSK)(Harter, 2012). De CBSK is een zelfrapportage vragenlijst die een beeld geeft van hoe een kind zichzelf ervaart en hoe het zijn eigen vaardigheden inschat. Deze vragenlijst heeft zes sub schalen: Schoolvaardigheden, Sociale Acceptatie, Sportieve Vaardigheden, Fysieke Verschijning, Gedragshouding en Gevoel van Eigenwaarde. Elke subschaal bevat zes items. Bij het beantwoorden van de items kiezen de kinderen tussen twee stellingen. Vervolgens geven ze aan of dit een beetje of helemaal van toepassing is.

Procedure

Tijdens een reguliere les bewegingsonderwijs, gegeven door een 3^e jaars student ALO, zijn de activiteiten badminton, kastspringen en lummelspel aan bod gekomen. Elke activiteit duurde ongeveer 8 minuten. Voorafgaand aan de lessen hebben de ALO studenten instructie gekregen over de manier van beoordelen en zijn de niveaus besproken. Aan het begin van de les werd de leerlingen verteld van welk onderdeel zij zelf hun vaardigheid moesten noteren. De niveaus zijn door de lesgever voorgedaan zodat de leerlingen een duidelijk beeld hadden van de verschillende niveaus. Tijdens het onderzoek hebben de leerlingen zicht gehad op de 10 'levels' van de rubric. Indien ze een level gehaald hadden, konden de leerlingen kijken op de rubric om hun volgende doel te bepalen. De leerlingen gaven na het oefenen middels de rubric aan welk level ze hadden gehaald. De docent en tweede beoordelaar hebben ook genoteerd welk level de leerling had gehaald. De leerling, docent en tweede beoordelaar wisten niet van elkaar welke score gegeven werd. Tijdens het onderzoek is er geen overleg over de beoordeling geweest tussen de leerlingen, de docent en de tweede beoordelaar om eventuele beïnvloeding uit te sluiten. De klas is verdeeld in drie groepen van acht tot tien leerlingen, die gelijktijdig bij de verschillende onderdelen aan het werk zijn gegaan. De rubric met de verschillende niveaus was op een duidelijk zichtbare plek opgehangen bij het onderdeel wat beoordeeld moest worden.

Aan het einde van de oefenronde moesten de leerlingen hun behaalde niveau noteren op een daarvoor gemaakte lijst. Op deze lijst stonden geen namen, enkel een nummer per leerling, dit om een eventuele bias uit te sluiten. De docent en tweede beoordelaar kregen ieder een leerlingenlijst om daarop de behaalde niveaus aan te kruisen.

Zelfbeeld

In dezelfde week van het praktijkgedeelte van het onderzoek is de CompetentieBelevingsschaal voor kinderen (CBSK) bij de leerlingen afgenomen(Harter, 2012). De onderzoeker heeft in het reguliere klaslokaal uitleg gegeven over deze vragenlijst en deze is vervolgens klassikaal afgenomen. Vooraf zijn enkele oefenvragen besproken en daarna zijn alle vragen van het onderzoek voorgelezen aan de leerlingen. Na het afronden van de vragenlijst zijn de ingevulde lijsten ingenomen door de onderzoeker. Deze procedure is in alle vier de klassen herhaald.

Data-analyse

De genoteerde niveaus van het praktijkgedeelte van het onderzoek zijn in eerste instantie in een Excel-document verzameld. De punten uit de zelfbeeldvragenlijst zijn hieraan toegevoegd. Na controle zijn vervolgens de gegevens uit het Excel document geïmporteerd in SPSS statistics 25. De verdere analyses zijn allen uitgevoerd in SPSS.

Als eerste is er een Wilcoxon test uitgevoerd om de beoordeling van de docent en de tweede docent te vergelijken, met als doel een eventuele bias van de docent uit te sluiten. Een tweede Wilcoxon test is uitgevoerd om eventuele significante verschillen tussen docent-beoordeling en leerling-beoordeling te beschrijven. Ook is de correlatie tussen docent-oordeel en leerling-oordeel getoetst

met behulp van een Spearman test.

Het verschil tussen docent en leerling is omgezet naar absolute waarden, de verschillcores. Met deze absolute waarden is een Mann-Whitney test uitgevoerd voor de variabelen geslacht, groep en onderdeel.

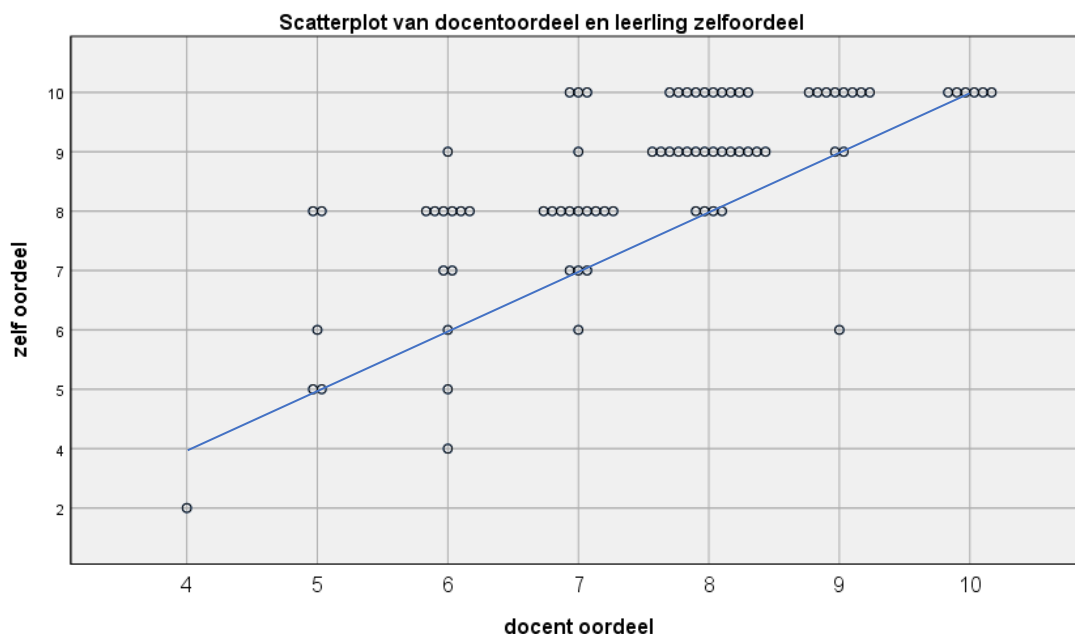
De antwoorden van de CBSK-vragenlijst zijn geordend naar sub-schalen en er is een totaalscore voor de CBSK gemaakt. Voor de sub schalen Sportieve Vaardigheid, Gevoel van Eigenwaarde en de totaalscore zijn groepen gemaakt op basis van de behaalde punten. De verschillcores van deze groepen zijn onderling vergeleken met een Mann-Whitney test om te kijken of het zelfbeeld van de leerlingen een voorspeller kan zijn voor de nauwkeurigheid van de zelfbeoordeling.

Resultaten

Op basis van een Wilcoxon test is geen verschil gevonden tussen het oordeel van de docent (Md=8 [4-10]) en de tweede beoordelaar (Md=7 [4-10]) ($p=.783$).

Een tweede Wilcoxon test is uitgevoerd om te bepalen of er een verschil was tussen het oordeel van de docent en het oordeel van de leerling. De leerling (Md=9 [2-10]) geeft zichzelf een significant hoger oordeel dan de docent (Md=8 [4-10]) ($p<.001$).

De Spearman-correlatie wees uit dat er een significant, sterk positief verband bestaat tussen docent oordeel en het zelfoordeel van de leerling. $R = .740$ $p < .001$. In figuur 1 is het verband tussen docentoordeel en het zelfoordeel van de leerling grafisch weergegeven middels een scatterplot.



Figuur 1: Scatterplot van docentoordeel en leerling zelfoordeel.

Het verschil tussen docent en leerling is gebruikt om het verschil tussen jongens (Md=1,0000[-3-3]) en meisjes (Md = 1,00 [0-3]) aan te tonen. Uit de Mann-Whitney test bleek dat er geen significant verschil ($p = .275$) is tussen jongens en meisjes. Ook is er geen verschil ($p = .750$) gevonden tussen leerlingen van groep 6 (Md= 1,00 [-3-3]) en leerlingen van groep 8 (Md= 1,00 [-2-3]). Tussen beide onderdelen, badminton (Md= 1,00 [-2-3]) en kastspringen (Md= 1,00 [-3-3]), is ook geen significant verschil ($p = .707$) gevonden.

Leerlingen die zichzelf hoger inschatten op gebied van sportieve vaardigheden (Md = 9 [5-10]) beoordelen zichzelf significant hoger dan leerlingen die zichzelf lager inschatten op gebied van sportieve vaardigheden (Md = 8 [2-10]) ($p = .020$). In het oordeel van de docent is er geen significant

verschil tussen leerlingen die zichzelf hoger inschatten op gebied van sportieve vaardigheden (Md = 8 [5-10]) en leerlingen die zichzelf lager inschatten op gebied van sportieve vaardigheden (Md = 8 [4-10]) ($p = .111$). Kijkend naar de afwijking ten opzichte van de docent is er geen significant verschil tussen leerlingen met een hogere score voor sportieve vaardigheden (Md = 1,00 [-1 max = 3]) en leerlingen met een lagere score voor sportieve vaardigheden (Md = 1,00 [-3-3]) ($p = .754$). Tabel 2 geeft een overzicht van de resultaten.

Een spearman-correlatie test wees uit dat er een significant, zwak positief verband bestaat tussen de score op CBSK-sportieve vaardigheid en het zelfoordeel van de leerling ($R = .312$ $p = 0.007$).

Tabel 2

Resultaten van de Mann-Whitney test: Zelfbeoordeling en docent beoordeling voor leerlingen met een hoge sportieve vaardigheid en een lage sportieve vaardigheid

| | Hoge sportieve vaardigheid | Lage sportieve vaardigheid | p-waarde |
|---------------------------|----------------------------|----------------------------|-------------|
| Zelfbeoordeling | | | |
| N | 29 | 44 | .020 |
| Mediaan (min-max) | 9 (5-10) | 8 (2-10) | |
| Docentbeoordeling | | | |
| N | 31 | 47 | .111 |
| Mediaan (min-max) | 8 (5-10) | 8 (4-10) | |
| Verskil met docent | | | |
| N | 29 | 44 | .754 |
| Mediaan (min-max) | 1 (-3 - 3) | 1 (-3 - 3) | |

Leerlingen die een hogere score hebben op het gevoel van eigenwaarde (Md = 9 [2-10]) beoordelen zichzelf niet significant hoger dan leerlingen met een lagere score op gevoel van eigenwaarde (Md = 9 [4-10]) ($p = .350$). Ook voor het docentoordeel is er geen verschil tussen leerlingen met een hogere score voor gevoel van eigenwaarde (Md = 8 [4-10]) en leerlingen met een lagere score voor gevoel van eigenwaarde (Md = 8 [5-10]) ($p = .588$). In tabel 3 staat een overzicht van de resultaten.

Er is geen significant verband gevonden tussen het gevoel van eigen waarde en de zelfbeoordeling van de leerlingen.

Tabel 3

Resultaten van de Mann-Whitney test: Zelfbeoordeling en docent beoordeling voor leerlingen met een hoger gevoel van eigenwaarde en een lager gevoel van eigenwaarde

| | Hoger gevoel van eigenwaarde | Lager gevoel van eigenwaarde | p-waarde |
|--------------------------|------------------------------|------------------------------|-------------|
| Zelfbeoordeling | | | |
| N | 34 | 40 | .350 |
| Mediaan (min-max) | 9 (2-10) | 8 (4-10) | |
| Docentbeoordeling | | | |
| N | 36 | 43 | .588 |
| Mediaan (min-max) | 8 (4-10) | 8 (5-10) | |

Discussie

Uit het onderzoek is gebleken dat verschillende docenten vergelijkbaar beoordelen, maar dat de leerlingen zichzelf significant hoger inschatten dan de docent. Er is echter wel een sterk verband gevonden tussen het oordeel van de docent en het zelfoordeel van de leerling.

Dat leerlingen zichzelf hoger inschatten dan de docent is te verklaren door de cognitieve fase waar leerlingen van deze leeftijd in zitten. Kinderen tussen de 7 en 12 jaar hebben de neiging om zichzelf te overschatten, al neemt in deze periode de invloed van "wishfull thinking" wel af (Butler 1990). Aangezien leerlingen in de concreet operationele fase hun vermogen tot zelfevaluatie ontwikkelen, werd verwacht dat er een verschil zou zijn tussen leerlingen van groep 6 en leerlingen van groep 8 (Boseovski 2010; Piaget 1976). In dit onderzoek is echter geen verschil gevonden in de nauwkeurigheid van zelfbeoordeling tussen deze groepen. Dit valt mogelijk te verklaren door het geringe leeftijdsverschil tussen leerlingen van groep 6 en groep 8. Een vergelijking met jongere leerlingen, bijvoorbeeld groep 4, zou mogelijk wel een verschil in nauwkeurigheid kunnen opleveren. Ook de ervaring die leerlingen van een Daltonschool hebben met zelfevaluatie kan invloed hebben op het ontbreken van verschil tussen de leerlingen van beide groepen.

Het geringe verschil tussen het oordeel van de docent en het zelfoordeel van de leerling kan veroorzaakt zijn door het gekozen meetinstrument. Voor dit onderzoek zijn rubrics ontwikkeld waarin de vaardigheden van leerlingen van groep 3 tot groep 8 beschreven worden. Deze rubrics zijn bewust eenvoudig gehouden, omdat een grote hoeveelheid criteria het zelfbeoordelen voor leerlingen complex kan maken (Weeldenburg, Borghouts & Slingerland 2015). Hoewel het bereik van het meetinstrument in theorie voor alle leerlingen van niveau 1 tot en met niveau 10 loopt, blijkt in de praktijk dat de leerlingen van groep 6 en groep 8 bijna allemaal tussen de 6 en de 9 scoren. Hierdoor is de kans dat een leerling een grote afwijking heeft ten opzichte van de docent kleiner.

Het was de verwachting dat de leerlingen het zelfbeoordelen van een spelsport (badminton) lastiger zouden vinden dan het zelfbeoordelen van een individuele vaardigheid (kastspringen). De verwachting was dat de nauwkeurigheid van het zelfoordeel bij badminton lager zou zijn dan bij het kastspringen. Er is echter geen verschil gevonden tussen beide onderdelen in dit onderzoek. Een mogelijke verklaring voor de overeenkomst kunnen de gebruikte rubrics zijn. Voor beide onderdelen is een rubric op dezelfde wijze tot stand gekomen met medewerking van dezelfde docenten bewegingsonderwijs. Aangenomen kan worden dat beide rubrics even duidelijk waren voor de leerlingen waardoor mogelijk de complexiteit van het onderdeel een minder grote rol heeft gespeeld.

Zelfbeeld

Opvallend is dat alle leerlingen relatief hoog scoorden op de CBSK. Binnen de CBSK wordt doorgaans gewerkt met een "onderste 15 percentiel" en een "bovenste 15 percentiel". Geen van de leerlingen scoorde in de onderste 15 procent. Bijna de helft van de leerlingen scoorde in de bovenste 15 procent. Er is daarom gekozen om de tweedeling te maken op basis van de mediaan van de CBSK scores. Zo is er een groep die zichzelf hoger inschat en een groep die zichzelf lager inschat. Leerlingen die zichzelf hoger inschatten op het gebied van sportieve vaardigheden in de CBSK geven zichzelf een hogere score bij de zelfbeoordeling. Aangezien deze leerlingen onderling geen significant verschil laten zien bij het verschil ten opzichte van de docent kan geconcludeerd worden dat het zelfbeeld van deze leerlingen juist is. Kijkend naar de scores voor gevoel van eigenwaarde zijn er geen significante verschillen tussen beide groepen.

Zowel gevoel van eigenwaarde als de score op sportieve vaardigheden zijn geen indicatieve variabelen voor de nauwkeurigheid van de zelfbeoordeling.

In een “normale” lessituatie heeft de docent niet de mogelijkheid om zich volledig te richten op de beoordeling van een groepje leerlingen. De docent is immers aan het lesgeven en moet daarom zijn/haar aandacht verdelen. Hierdoor zou een extra variabele ontstaan, de aandacht van de docent, welke van invloed kan zijn op de resultaten. In het onderzoek is daarom gekozen om de les door een vierdejaars student te laten geven, zodat de docent zich kon richten op de beoordeling van de leerlingen.

Externe validiteit en vervolgonderzoek

Dit onderzoek is uitgevoerd op een Daltonschool. Een van de kenmerken van Daltononderwijs is dat leerlingen al vanaf de kleuters kennismaken en oefenen met zelfreflectie. Er is geen kennis beschikbaar over het verschil in zelfevaluatie bij leerlingen van een Daltonschool ten opzichte van leerlingen van scholen met een andere onderwijsvorm. Mogelijk hebben leerlingen die niet op een Daltonschool leskrijgen minder voorkennis wat betreft zelfbeoordeling. Het is daarom aan te bevelen om dit onderzoek uit te voeren bij andere scholen waarin de nauwkeurigheid van zelfbeoordeling bij verschillende onderwijsvormen vergeleken wordt.

De gemiddelde score van de school op de centrale eindtoets (cito) van de afgelopen vijf jaar is 540,5. Dit is 5,1 punt hoger dan het landelijk gemiddelde van de afgelopen vijf jaar, deze was namelijk 535,3 (College voor toetsing en Examens 2021; Klaverweide 2021). Uit onderzoek blijkt dat leerlingen met hogere academische prestaties zichzelf realistischer zullen inschatten (Andrade 2019; Boud & Falchikov 1989). Er bestaat derhalve een kans dat leerlingen van een school met lagere cito-scores een grotere afwijking laten zien ten opzichte van de docent. In een eventueel vervolgonderzoek is het aan te raden scholen mee te nemen met verschillende cito-resultaten om het eventuele effect in kaart te brengen.

Daarnaast is het ook wenselijk om te kijken of de zelfbeoordeling in het bewegingsonderwijs nauwkeuriger wordt indien dit voor een langere periode wordt gebruikt. Als leerlingen oefenen met zelfbeoordeling verhoogt dit de nauwkeurigheid en de neiging tot het ophogen van de beoordeling neemt af (Wong 2016).

Aanbevelingen voor de praktijk.

Ondanks het verschil in beoordeling tussen leerling en docent is het aan te raden zelfbeoordeling te gebruiken in het bewegingsonderwijs. De voordelen van zelfbeoordeling wegen op tegen het mogelijk afwijkende oordeel van de leerling. Het gebruiken van zelfbeoordeling zorgt ervoor dat de kwaliteit van zelfbeoordelen verbetert (González-Betancor, Bolívar-Cruz, & Verano-Tacoronte 2017).

Door het sterke verband tussen het docent-oordeel en het zelfoordeel van de leerling zijn de uitkomsten van de zelfbeoordeling geschikt om te gebruiken door de docent. Om succesvol gebruik te maken van zelfbeoordeling is het essentieel dat de criteria waarop de leerlingen zichzelf beoordelen duidelijk zijn en moeten de leerlingen de ruimte krijgen om zichzelf te verbeteren (Andrade & Valtcheva 2009; Ross 2006). De verkregen data is, behalve als beoordeling, ook te gebruiken om het (begin)niveau van een klas te bepalen, het effect van de leraar en zijn/haar lesmethode te controleren, leerdoelen te bepalen voor leerlingen en leerlingen te motiveren binnen het bewegingsonderwijs. Dit sluit aan bij de stelling van het AIESEP dat leerlingen voor een optimale leerervaring actief betrokken moeten zijn in het beoordelingsproces (AIESEP 2020).

Het innovatieve systeem van Gymladder zou hier een rol in kunnen spelen door de administratie van de zelfbeoordeling te automatiseren en minder tijdsintensief te maken. Gymladder maakt gebruik van eenvoudige rubrics, wat zorgt voor een hogere motivatie bij leerlingen (Weeldenburg, Borghouts & Slingerland 2015).

Conclusie

Bij het afsluiten van dit artikel moeten de volgende beperkingen worden opgemerkt. Het onderzoek is op één school uitgevoerd, om de gevonden data te kunnen generaliseren zal er vervolgonderzoek gedaan moeten worden op meerdere scholen. Daarnaast heeft het gebruikte meetinstrument slechts een beperkt bereik waardoor het verschil tussen docent en leerling mogelijk is beïnvloed.

Leerlingen schatten zichzelf hoger in dan de docent. Het sterke verband tussen het oordeel van de docent en het zelfoordeel van de leerling zorgt ervoor dat het gebruik van zelfbeoordeling waardevolle informatie kan geven aan de docent bewegingsonderwijs. De docent moet dan wel rekening houden met de overschatting bij zelfbeoordeling. Daarnaast biedt het inzetten van zelfbeoordeling nog andere voordelen waardoor het gebruik van zelfbeoordeling zeker aan te bevelen is.

Literatuur

- AIESEP. (2020). *Standpunt Beoordeling bij Lichamelijke Opvoeding*. <https://aiesep.org/scientific-meetings/position-statements/>
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, 4(August), 1–13. <https://doi.org/10.3389/educ.2019.00087>
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12–19. <https://doi.org/10.1080/00405840802577544>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*. <https://doi.org/10.1002/hrm>
- Borghouts, L., Slingerland, M., & Haerens, L. (2017). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 473–489. <https://doi.org/10.1080/17408989.2016.1241226>
- Boseovski, J. J. (2010). Evidence of “rose-colored glasses”: An examination of the positivity bias in young children’s personality judgments. *Child Development Perspectives*, 4, 212–218. <https://doi.org/10.1111/j.1750-8606.2010.00149.x/full>
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education*, 18(5), 529–549. <https://doi.org/10.1007/BF00138746>
- Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy and Practice*, 22(4), 444–457. <https://doi.org/10.1080/0969594X.2014.996523>
- Brown, & Harris. (2013). *Student Self Assessment*.
- Butler, R. (1990). *The Effects of Mastery and Competitive Conditions on Self-Assessment at Different Ages*.
- Collard, D. (n.d.). *Effecten van bewegingsonderwijs op sport- en beweeggedrag op latere leeftijd*. Retrieved November 14, 2018, from <https://www.kennisbanksportenbewegen.nl/?file=8069&m=1504190751&action=file.download>
- Cvte.nl. (n.d.). *Leerlingresultaten Centrale Eindtoets 2021*. Retrieved June 3, 2021, from <https://www.centraleeindtoetspo.nl/actueel/nieuws/2021/05/25/leerlingresultaten-centraleeindtoets>
- Ehrlinger, J. (2008). Skill Level, Self-Views and Self-Theories as Sources of Error in Self-Assessment. *Social and Personality Psychology Compass*, 2(1), 382–398. <https://doi.org/10.1111/j.1751-9004.2007.00047.x>
- González-Betancor, S. M., Bolívar-Cruz, A., & Verano-Tacoronte, D. (2017). Self-assessment accuracy in higher education: The influence of gender and performance of university students. *Active Learning in Higher Education*, 20(2), 101–114. <https://doi.org/10.1177/1469787417735604>
- Haerens, L. (n.d.). *Lichamelijke opvoeding: “leuk” is het doel?!* Retrieved December 12, 2018, from <http://www.sportknowhowxl.nl/nieuws-en-achtergronden/open-podium/item/87345/>
- Harter, S. (2012). Social Support Scale for Children : Manual and Questionnaires. *University of Denver*, 44.

- Keane, L., & Griffin, C. (2015). Testing the limits of self-assessment: A critical examination of the developmental trajectories of self-assessment processes. *Irish Teachers' Journal*, 3(1), 37–52.
- Klaverweide, D. (2021). *Schoolgids*.
- Krijgsman, C., & Maes, J. (2015). Help, ik krijg vandaag een cijfer. *LO Magazine*.
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18(1), 57–76. <https://doi.org/10.1080/13573322.2012.713860>
- Melograno, V. J. (2007). Grading and Report Cards for Standards-Based Physical Education. *Journal of Physical Education, Recreation & Dance*, 78(6), 45–53. <https://doi.org/10.1080/07303084.2007.10598041>
- Minelgaite, I., Nedzinskaitė-Mačiūnienė, R., Kristinsson, K., & Gudjonsson, S. (2019). “The emperor is naked!”: Exposing (in) efficiency of self-assessment and group – assessment in higher education. *Pedagogika*, 134(2), 45–60. <https://doi.org/10.15823/p.2019.134.3>
- Ní Chrónín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18(2), 219–233. <https://doi.org/10.1080/17408989.2012.666787>
- Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education*, 7(2), 133–144. <https://doi.org/10.1177/1477878509104318>
- Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: a survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy and Practice*, 21(4), 365–383. <https://doi.org/10.1080/0969594X.2014.919247>
- Papanthymou, A., & Darra, M. (2019). *Student Self-Assessment in Primary and Secondary Education in Greece and Internationally*. 6(1). <https://doi.org/10.22158/wjer.v6n1p50>
- Penney, D., Brooker, R., Hay, P., & Gillespie, L. (2009). Curriculum, pedagogy and assessment: three message systems of schooling and dimensions of quality physical education. *Sport, Education and Society*, 14(4), 421–442. <https://doi.org/10.1080/13573320903217125>
- Peyton, C. (2017). *Student's Perception of The Self-Assessment Process In High School Physical Education*.
- Piaget, J. (1976). Piaget's Theory. *Piaget and His School*, 11–23. https://doi.org/10.1007/978-3-642-46323-5_2
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11(10), 1–13. <https://tspace.library.utoronto.ca/handle/1807/30005>
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2001). Self-evaluation in Grade 11 Mathematics. *American Educational Research Association*.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Savelkoul, V. (2015). *(HOE)WORDEN LEERLINGEN IN HET BASISONDERWIJS BEOORDEELD OP BEWEGINGSONDERWIJS?*
- Van Dokkum, G. (2008). Beoordelen binnen LO, vooral een agogisch probleem. *Lichamelijke Opvoeding*, 96, nr 11, 11–13.

- Van Dokkum, G. (2011). Het beoordelingspectrum: adaptief beoordelen in de les lo. *Kvlo*, *LO-6*, 23–25.
<https://www.kennisbanksportenbewegen.nl/?file=5144&m=1445331562&action=file.download>
- Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration*, *23*(3), 263–280. <https://doi.org/10.1037/a0032359>
- Weeldenburg, G., & Slingerland, M. (2019). Formatief evalueren. *Kvlo Magazine*.
<https://slo.nl/thema/meer/formatief-evalueren/>
- Weeldenburg, G., Borghouts, L., & Slingerland, M. (2015). *Motiverend beoordelen*. *September*.
- Wong, H. M. (2016). I can assess myself: Singaporean primary students' and teachers' perceptions of students' self-assessment ability. *Education 3-13*, *44*(4), 442–457.
<https://doi.org/10.1080/03004279.2014.982672>