

8. Past drama in een hokje? Een praktijkonderzoek naar de validiteit en toepasbaarheid van een beoordelingsinstrument

Debbie Klarenbeek en Borius van der Meulen

Lok (2013) heeft een beoordelingsinstrument ontworpen voor dramalessen in het voortgezet onderwijs (zie hfst. 7 in deze bundel). Het beoogt de belangrijkste dramavaardigheden te beoordelen met een rubric. Zoals bij alle beoordelingsinstrumenten kan men vragen stellen over de kwaliteit ervan, bijvoorbeeld over de betrouwbaarheid (hoe objectief of subjectief zijn de gegeven oordelen?), de transparantie (voor leerlingen en collega's), de eerlijkheid en de praktische toepasbaarheid. Maar de meest basale kwaliteitsvraag is die naar de validiteit van de beoordeling: meet het instrument wat het bedoelt te meten?

In deze bijdrage doen we verslag van ons onderzoek naar de validiteit van het door Lok ontworpen instrument (Klarenbeek en Van der Meulen, 2014). We zijn beiden dramadocenten en we vinden een dergelijk veelomvattend instrument van belang voor het dramaonderwijs. Met dit onderzoek willen we een verdere bijdrage aan de ontwikkeling van het instrument geven. We richten ons op de validiteit van de onderscheiden criteria en de niveaubeschrijvingen van deze criteria. Daarnaast komt de praktische toepasbaarheid van het instrument aan de orde.

8.1 De onderzoeksvraag

Er zijn vele definities voor validiteit. Drenth & Sijtsma (1990, p.173) omschrijven de validiteit van een test bijvoorbeeld als 'de mate waarin de test aan zijn doel beantwoordt'. Bovendien zijn er verschillende vormen van validiteit, zoals predictieve validiteit, constructvaliditeit (Drenth & Sijtsma, 1990), begripsvaliditeit (Swanborn, 1984) en inhoudsvaliditeit (Drenth & Sijtsma, 1990).

Binnen dit onderzoek is sprake van inhoudsvaliditeit. We zoeken antwoorden op de vraag 'Meet het beoordelingsinstrument met de geformuleerde criteria en niveauomschrijvingen wat het bedoelt te meten? Daarbij kijken we naar zaken als meetbaarheid, volledigheid, vakinhoud en woordkeuze.

We hebben niet alleen gekeken naar een representatieve afspiegeling van de inhoud van de dramalessen, maar ook naar de operationalisering van deze inhoud. Met operationalisering bedoelen we de manier waarop het instrument de verschillende kenmerken van de dramalessen meet. We kijken daarbij naar zaken als overlappingsen tussen verschillende criteria, de onderlinge verhoudingen tussen de niveauomschrijvingen, de mate waarin criteria de inhoud dekken en het loskoppelen of uitsplitsen van criteria. We zoeken antwoorden op de vraag of het te meten begrip inhoudelijk voldoende is uitgewerkt in zijn verschillende aspecten.

Naast inhoud en operationalisering hebben we ook aandacht besteed aan het digitale aspect van het beoordelingsmodel (Google Docs), de praktische toepasbaarheid en de overzichtelijkheid ofwel de vormgeving van het model. Bovendien hebben we het model vergeleken met de eigen beoordelingsmethode van de docent.

Gezien de beperkte omvang van het onderzoek was het niet mogelijk om alle criteria en subcriteria te onderzoeken. Besloten is daarom ons te richten op een selectie van acht criteria.

Omdat Lok het vooronderzoek heeft gedaan, het instrument heeft ontwikkeld en zelf een drama-achtergrond heeft, hebben we ervoor gekozen haar de selectie te laten maken. We vroegen haar een selectie te maken op basis van de volgende voorwaarden:

- Zowel algemenere criteria als criteria voor specifieke lessituaties
- Zowel criteria waar ze positief over is als waarover ze nog twijfels heeft wat betreft representatie van de te beoordelen vaardigheid

Tabel 1 toont de gekozen acht criteria met bijbehorende niveauomschrijvingen.

Tabel 1: De acht door ons onderzochte rubrics uit het beoordelingsinstrument van Lok (2013).

| Hoofdcriteria | Subcriteria | Niveau omschrijvingen | | | |
|-------------------------|--|--|---|---|---|
| 1A Basisvaardigheden | 1B Creëren veilige spelomgeving | 1.1 Leerling is altijd actief bezig met creëren veilige leeromgeving | 1.2 Leerling is meestal actief bezig met creëren veilige leeromgeving | 1.3 Leerling is soms actief bezig met creëren van een veilige leeromgeving | 1.4 Leerling is te weinig actief in het creëren van een veilige leeromgeving |
| 2A Basisvaardigheden | 2B Verbeelding / fantasie | 2.1 Leerling beschikt over grote verbeelding en fantasie en zet deze altijd in | 2.2 Leerling beschikt over grote verbeelding en fantasie en zet deze in | 2.3 Leerling beschikt over verbeelding en fantasie en zet deze soms in | 2.4 Leerling vindt het moeilijk om zijn verbeelding en fantasie te gebruiken |
| 3A Spelvaardigheden | 3B Geloofwaardigheid | 3.1 De leerling speelt emoties zeer geloofwaardig | 3.2 De leerling speelt emotie geloofwaardig | 3.3 De leerling speelt emotie soms geloofwaardig | 3.4 De speler vindt het moeilijk om emoties geloofwaardig te spelen |
| 4A Spelvaardigheden | 4B Fysiek spel/ transformatie | 4.1 Het personage is heel duidelijk fysiek vormgegeven | 4.2 Het personage is duidelijk fysiek vormgegeven | 4.3 Het personage is in delen van het spel fysiek vormgegeven | 4.4 Het personage is nauwelijks fysiek vormgegeven |
| 5A Spelvaardigheden | 5B Tekstbehandeling | 5.1 Leerling is altijd waarachtig in zijn tekstbehandeling | 5.2 Leerling is meestal waarachtig in zijn tekstbehandeling | 5.3 Leerling is waarachtig in zijn tekstbehandeling | 5.4 Leerling is onvoldoende waarachtig in de tekstbehandeling |
| 6A Maken | 6B Regisseren van een scène | 6.1 Leerling heeft altijd een beeld bij een scène en kan vanuit dat beeld goed regisseren | 6.2 Leerling heeft meestal een beeld bij een scène en kan vanuit dat beeld goed regisseren | 6.3 Leerling heeft soms een beeld bij een scène en kan vanuit dat beeld regisseren | 6.4 Leerling vindt het moeilijk om te regisseren vanuit een beeld |
| 7A Maken | 7B Verbeelden | 7.1 Leerling kan heel goed verbeelding inzetten bij het maken van scènes | 7.2 Leerling kan goed verbeelding inzetten bij het maken van scènes | 7.3 Leerling kan soms verbeelding inzetten bij het maken van scènes | 7.4 Leerling vindt het moeilijk om verbeelding in te zetten bij het maken van scènes |
| 8A Reflecteren | 8B Reflecteren op eigen werk en dat van anderen | 8.1 Leerling reflecteert zeer goed op eigen werk en op het werk van anderen | 8.2 Leerling reflecteert goed op eigen werk en op het werk van anderen | 8.3 Leerling kan soms reflecteren op eigen werk en op werk van anderen | 8.4 Leerling vindt het moeilijk om te reflecteren op eigen werk en op het werk van anderen |

De onderzoeksvraag luidt:

In hoeverre is het beoordelingsinstrument voor dramalessen in het voortgezet onderwijs van Lok (2013) praktisch toepasbaar en inhoudsvalide volgens docent en leerling?

De deelvragen die hieruit voortkomen zijn:

- Is het beoordelingsmodel praktisch toepasbaar binnen een lessituatie in het voortgezet onderwijs?
- Is de inhoud van de acht geselecteerde criteria valide volgens docent en leerling?
- Is de inhoud van de niveauomschrijvingen valide en kloppen de onderlinge verhoudingen volgens docent en leerling?
- Wat is er op basis van dit onderzoek te zeggen over het totale beoordelingsmodel van Lok en is de inhoud van de overige negenentwintig criteria en niveauomschrijvingen valide volgens de docent?

8.2 De opzet van het onderzoek

We hebben de inhoudsvaliditeit onderzocht met een kwalitatieve survey. Daarbij kozen we voor semigestructureerde groepsinterviews met leerlingen en semigestructureerde interviews (Baarda, 2009) met betrokken dramadocenten nadat zij met het beoordelingsinstrument hebben gewerkt. Bij de interviews gebruikten we een topiclist.

Het verzamelen van de data

We hebben scholen geselecteerd die nog niet bekend zijn met het instrument. In totaal hebben twee dramadocenten van eenzelfde school en één dramadocent van een andere school meegewerkt. Per school hebben we leerlingen uit twee klassen bevraagd, in totaal dus vier klassen.

De docenten kregen vooraf een degelijke instructie over de 8 criteria van het beoordelingsinstrument en gingen hiermee vervolgens met hun leerlingen in drie dramalessen aan de slag. Leerlingen hadden inzage in de rubrics, omdat ze aan het einde van de derde les beoordeeld werden door de docent en medeleerlingen aan de hand van de acht rubrics. Zo hadden de leerlingen en de docenten voldoende kennis van het beoordelingsinstrument.

We vroegen leerlingen en docenten naar hun mening over de inhoudsvaliditeit en de begrijpelijkheid van de formulering, per hoofdcriterium, subcriterium en de bijbehorende niveauomschrijvingen van elk van de acht rubrics. We vroegen hen ook wat er verbeterd zou kunnen worden. Vervolgens hebben we gekeken naar de toepasbaarheid en bruikbaarheid van het gehele instrument en naar de huidige beoordelingsmethode binnen de dramalessen op de betreffende school.

De analyse

Alle interviewverslagen zijn gecodeerd en de codes zijn nader geanalyseerd. Omdat de bevindingen van individuele respondenten over specifieke rubrics erg uiteen liepen, zijn bovendien de acht rubrics per interview herschreven naar de inzichten van de betreffende respondent. Een voorbeeld hiervan is te zien in tabel 2. Enige interpretatie van ons als onderzoekers om een volledige versie van de rubrics te maken, was niet altijd te voorkomen.

Tabel 2: Voorbeeld van de aangepaste rubrics naar aanleiding van interview met docent 1, met de aanpassingen in rood.

| | | | | | |
|-------------------|---|--|--|---|--|
| Basisvaardigheden | Bijdrage aan een veilige les | Leerting draagt niet bij aan een veilige les | - | - | Leerting draagt niet bij aan een veilige les |
| Speelvaardigheden | Verbeelding / fantasie | Leerting beschikt over grote verbeelding en fantasie en zet deze altijd in. | Leerting beschikt over gemiddelde verbeelding en fantasie en zet deze in. | Leerting beschikt over verbeelding en fantasie en zet deze soms in. | Leerting vindt het moeilijk om zijn verbeelding en fantasie te gebruiken |
| Speelvaardigheden | Gelooftwaardigheid: -leerting incasseert -leerting reageert -leerting maakt zijn belang groot, genoeg | Leerting speelt alle drie de onderdelen goed waardoor spel zeer geloofwaardig is | Leerting speelt minstens twee onderdelen goed waardoor spel geloofwaardig is | Leerting speelt één of twee onderdelen goed waardoor spel redelijk geloofwaardig is | Leerting speelt geen enkel onderdeel goed waardoor spel ongelooftwaardig is |
| Speelvaardigheden | Fysiek spel: Fysieke vormgeving in: -knien -heupen -schouders -hoofd | Leerting laat zijn fysieke spel gedurende de hele scène in al zijn lichaamsdelen in houding en beweging zien | Leerting laat zijn fysieke spel gedurende een deel van de scène in veel van zijn lichaamsdelen in houding en beweging zien | Leerting laat zijn fysieke spel gedurende een klein deel van de scène in een aantal lichaamsdelen in houding en beweging zien | Leerting laat zijn fysieke spel gedurende een korte tijd in weinig lichaamsdelen in houding en beweging zien |
| Speelvaardigheden | Transformatie: -fysiek -stem | Goed | Ruim voldoende | Matig | Onvoldoende |
| Speelvaardigheden | Tekstbehandeling: -goed tekstbegrip -goede intonatie -goed tempo/ritme -goede timing | Leerting voldoet aan alle vier de subcriteria | Leerting voldoet aan twee van de vier subcriteria | Leerting voldoet aan één van de vier subcriteria | Leerting voldoet aan geen van de subcriteria |
| Maken | Regisseren van een scène: -wanuit een beeld -op spel -op samenspel -op fysieke transformatie -op tekstbehandeling | Leerting kan goed regisseren vanuit een van te voren opgegeven lesdoel | Leerting kan redelijk goed regisseren vanuit een van te voren opgegeven lesdoel | Leerting kan matig regisseren vanuit een van te voren opgegeven lesdoel | Leerting kan onvoldoende regisseren vanuit een van te voren opgegeven lesdoel |
| Maken | Leerting is creatief in het omzetten van iets (muziek, idee, thema etc) naar theater | Leerting kan heel goed verbeelding inzetten bij het maken van scènes | Leerting kan goed verbeelding inzetten bij het maken van scènes | Leerting kan redelijk goed verbeelding inzetten bij het maken van scènes | Leerting vindt het moeilijk om verbeelding in te zetten bij het maken van scènes |
| Reflecteren | Reflecteren op eigen werk -leerting kan benoemen op welk niveau hij speelt -leerting kan benoemen of er vooruitgang in spel zit -leerting kan naar zichzelf kijken op video -leerting kan zichzelf als speler vergelijken met andere spelers -leerting is zich bewust van zijn spel, lichaam, stem -leerting is zich bewust van zijn creativiteit | Leerting reflecteert zeer goed op alle subcriteria van het eigen werk | Leerting reflecteert voldoende op de meeste subcriteria van het eigen werk | Leerting reflecteert matig op een aantal subcriteria van het eigen werk | Leerting reflecteert niet of nauwelijks op eigen werk |
| Reflecteren | Reflecteren op het werk van anderen | Leerting geeft elke les feedback aan anderen | Leerting geeft elke les feedback aan anderen | Leerting geeft elke les feedback aan anderen | Leerting geeft nooit feedback aan anderen |

De bevindingen uit de interviews zijn opgedeeld naar de twee hoofdthema's: de inhoudsvaliditeit en de praktische toepasbaarheid van het beoordelingsmodel in de lespraktijk. Inhoudsvaliditeit is onderverdeeld in inhoud en operationalisering. Voorstellen voor aanpassingen en aanbevelingen die hier worden besproken zijn van de respondenten en niet van de onderzoekers.

8.3 De inhoudsvaliditeit van het model

Inhoudsvaliditeit omvat de subthema's inhoud en operationalisering.

Inhoud

Binnen inhoud hebben we gekeken of het instrument precies meet wat het bedoelt te meten. Bijvoorbeeld: kent iedereen dezelfde betekenis toe aan het begrip 'geloofwaardigheid'? Hiervoor zijn we ingegaan op de bruikbaarheid, volledigheid, meetbaarheid, woordkeuze en de taak van docent.

Bruikbaarheid

Docenten en leerlingen waren positief over de bruikbaarheid van de meeste criteria. Zo waren ze direct positief over het belang van 'de veilige spelomgeving', 'het reflecteren' en 'geloofwaardigheid'.

Leerlingen kunnen bij een aantal criteria heel helder vertellen wat zij daaronder verstaan. Ook verwijst een docenten een aantal keer naar de exameneisen voor en de overlap met het vak CKV. Hierdoor hecht hij ook extra belang aan bepaalde criteria en de toepasselijkheid ervan.

Docenten zeggen verder dat er wel veel rollen toetsbaar zijn, maar dat ze specifieke vaardigheden binnen de criteria missen:

'Ja, daar ontbreken er echt substantieel veel. En juist op die subcriteria zijn vaak lesdoelen geënt.'

Als voorbeeld noemen ze het criterium 'tekstbehandeling'. Daar passen volgens hen onderdelen als verstaanbaarheid, tekstbegrip, timing, intonatie en ritme bij. Als dit er niet bij is gevoegd, dan weten leerlingen nog niet goed wát van het brede spectrum 'tekstbehandeling' beoordeeld wordt:

'Hoe concreter het is, hoe minder discutabel het wordt voor leerlingen ook.'

De criteria zijn in het beoordelingsinstrument veelal gericht op lesprocessen en lesonderdelen en niet zozeer op het spelen van een voorstelling. Ook vragen docenten zich af in hoeverre er rekening gehouden is met de exameneisen voor het vak drama.

Verskillende criteria (zoals 'ruimtelijk bewustzijn', 'reflecteren op eigen werk en dat van anderen' en 'regievaardigheden') zijn volgens docenten pas te gebruiken bij een bepaalde leeftijd, een bepaald niveau van de leerlingen of als leerlingen drama als eindexamenvak hebben gekozen en docenten hogere eisen aan hen kunnen stellen:

'Reflecteren op eigen werk en dat van anderen: Ja, dat vind ik ingewikkeld (...) omdat je je moet afvragen of kinderen, of sommige leeftijden daar wel toe in staat zijn. Zichzelf los te koppelen van de (...) Of je dat wel kan vragen.'

Ook een leerling sluit hierop aan door te zeggen dat hij zichzelf op bepaalde criteria lastig kan beoordelen:

'De laatste drie ['regisseren van een scène', 'verbeelden', 'reflecteren op eigen werk en dat van anderen'] zijn eigenlijk voornamelijk voor de lerares bedoelt, en niet zozeer, denk ik voor de leerling (...) reflecteren zou op zich nog wel kunnen, maar je merkt, je bent vooral heel erg bezig met hoe iemand speelt.'

Docenten zeggen de hoofdcriteria 'schrijven', 'reflecteren' en 'theorie' niet te gaan

gebruiken, omdat die aspecten binnen de lessen te weinig aan bod komt (zoals schrijven) of ze leerlingen er niet op beoordelen (zoals reflecteren). Bij reflecteren kunnen volgens hen subcriteria worden samengevoegd en dat de bruikbaarheid zou verbeteren met een mondelinge uitleg. Hierdoor kan de docent zelf beter invulling geven aan het criterium:

'Elkaars presentaties beoordelen mag weg, die valt onder opbouwende feedback geven. Ik vind niet dat je de leerling op de stoel van de docent mag zetten.'

Volledigheid

Het gevaar van het samenstellen van eigen rubrics uit de zevenendertig rubrics is dat er een onvolledige versie ontstaat. Zo noemen docenten verschillende malen dat het criterium 'regisseren van een scène' in de nu onderzochte acht rubrics veel te weinig omvat van waar regisseren eigenlijk over gaat. Elementen als leiding geven, samenwerken en vormgeven mist men:

'Voor mijzelf bestaat regisseren voor ongeveer zeker 60 procent uit sociale cohesie creëren en zeg maar leiding geven en allemaal dat soort dingen en misschien 40 procent uit creativiteit.'

Wel ervaren docenten de zevenendertig rubrics als behoorlijk volledig en kunnen ze niet specifiek iets noemen wat ze missen. Doordat de niveaus met woorden zijn omschreven, dekken zij niet altijd wat het criterium wil beoordelen. Dit is bijvoorbeeld terug te zien bij het subcriterium 'geloofwaardigheid'. Binnen de niveauomschrijvingen is ervoor gekozen om emoties aan geloofwaardigheid te koppelen, bijvoorbeeld: 'De leerling speelt emoties zeer geloofwaardig'. Docenten zeggen hierover:

'Ik zou emoties eruit halen, omdat, het niet altijd met emoties van doen heeft of iets geloofwaardig is. Want als ik een fysieke opdracht geef (...) dan gaat het gewoon puur om: geloof ik dat die persoon inderdaad in een kathedraal staat?'

Ditzelfde geldt voor het subcriterium 'fysiek spel/transformatie', waaraan binnen de niveauomschrijvingen het woord 'personage' gekoppeld is, bijvoorbeeld: 'Het personage is heel duidelijk fysiek vormgegeven'. Volgens de respondenten is fysiek spel meer dan alleen het vormgeven van een personage.

Als subcriteria daarentegen onvoldoende zijn uitgewerkt, missen de respondenten ook iets. Een leerling zegt:

'Ja, dan denk ik: tekstbehandeling. Bedoel je dan dat je tekst makkelijk eigen kan maken? Of dat je je tekst goed, weet ik veel, ja, kan ontleden, of dat je hem uit je hoofd kan leren?'

Naast dat de rubrics niet altijd volledig zijn, vinden de respondenten ook dat veel dingen in de klas besproken horen te worden. Zo wordt mondeling duidelijk wat er met een bepaald criterium bedoeld wordt.

Meetbaarheid

Over de meetbaarheid van de criteria met bijbehorende niveauomschrijvingen zijn veel uiteenlopende en bruikbare opmerkingen gemaakt tijdens de interviews. Zo hechten respondenten bij meetbaarheid veel belang aan de zichtbaarheid. Kun je bijvoorbeeld als leerling en docent bepalen of iemand fantasie heeft? Volgens de respondenten het onzichtbare niet te meten en daarmee niet of lastig te beoordelen:

'Maar ze moeten het wel kunnen laten zien, snap je? Je kan in je eigen fantasiewereld leven en totaal in de verbeelding zijn, maar dat absoluut niet uitdragen op het podium.'

Of een criterium zich voordoet, dus te meten is, is soms afhankelijk van leeftijd, geslacht en aanleg. Zo zegt een docent dat het criterium 'reflecteren op eigen werk en dat van anderen' pas op latere leeftijd toepasbaar en dus meetbaar is.

Respondenten vinden zichtbare omschreven aspecten zoals emoties een goede uitwerking van een criterium. De vraag is echter of, zoals we hierboven al meldden, emoties altijd een goede vertolking van een criterium zijn. De meeste respondenten vinden 'geloofwaardigheid', 'waarachtigheid' en 'tekstbehandeling' subjectieve grootheden en dat bemoeilijkt de meetbaarheid:

'Ja, ik vind het ook zo subjectief, ik kan soms iets heel erg geloofwaardig vinden en dan vindt iemand anders het heel lelijk.'

Tekstbehandeling is volgens een aantal respondenten objectief te meten en te beoordelen als de objectieve aspecten benoemd worden. Binnen de huidige niveauomschrijving wordt dit allemaal samengevat in 'waarachtig', wat volgens de respondenten niet meetbaar is. Bij het criterium over reflectie vinden de leerlingen het onduidelijk hoe de mate van reflectie wordt beoordeeld en vraagt een leerling zich af wat hij doet als hij goed reflecteert. Verder wijzen de respondenten erop dat woorden als 'origineel' en 'sterk' subjectieve waarderings zijn die niet meetbaar zijn en daarom vermeden moeten worden.

Binnen bepaalde niveauomschrijvingen (bijvoorbeeld bij het criterium 'fysiek spel/transformatie') is het niet duidelijk of het om een momentopname van beoordeling gaat of over een gemiddelde van meer momentopnames. Het is volgens de respondenten belangrijk dat hier duidelijkheid over is, omdat dit invloed heeft op de meetbaarheid en validiteit van het criterium.

Regisseren zien leerlingen als een groepsproces en het is niet altijd te bepalen wie het idee heeft bedacht. Dit maakt het moeilijk om de individuele regievaardigheden, zoals die nu in de rubrics staan, te meten. Binnen hetzelfde hoofdcriterium 'maken' zou het de mate van dekking en meetbaarheid verbeteren als hier het subcriterium 'vormgeven' aan toegevoegd zou worden, wat volgens een docent een goed meetbaar aspect is van het maken van een scène.

Woordkeuze

Omdat de Master Rubric een uitgeschreven beoordelingsinstrument is, hebben de formuleringen grote invloed op de bruikbaarheid en validiteit. We hebben gekeken of de woordkeuze volgens respondenten aansluit op de vakinhoudelijke aspecten die beoordeeld moeten worden en of de criteria en niveauomschrijvingen begrijpelijk en eenduidig zijn. Het is voor een valide beoordeling van groot belang dat de beoordelaar en de beoordeelde een tekst op dezelfde manier verstaan en interpreteren.

Het is belangrijk dat vakinhoudelijke termen consequent worden gebruikt. Zo staat binnen eenzelfde criterium zowel 'spelomgeving' als 'leeromgeving'. Dat leidt tot verwarring en dat staat goed beoordelen in de weg. Deze inconsequentie is ook terug te vinden in de aanduiding van de beoordeelde: meestal staat er 'de leerling', maar soms ook 'de speler' en 'het personage'. Dit wekt onterecht de indruk dat het hier over iemand anders gaat. De respondenten geven de voorkeur aan 'de leerling'.

Uit de diversiteit aan opvattingen over inhoudelijke woordkeuze kunnen we vooral concluderen dat inhoudelijke woordkeuze zeer persoonlijk is.

Het is voor de respondenten van belang dat er geen twijfel of discussie kan ontstaan over wat er specifiek bedoeld wordt met een bepaald woord. Het huidige model bevat woorden als 'fysiek spel', 'fysiek vormgeven', 'waarachtig' en 'transformatie'. Voor docenten zijn deze begrippen helder, maar de leerlingen weten niet precies waar ze nu op beoordeeld worden:

'Fysiek spel is toch dat je heel erg (...) Als je boos bent, dat je dan echt op de tafel gaat slaan.'

Uit de interviews is gebleken dat sommige woorden vervangen kunnen worden door begrijpelijker en eenduidiger woorden, zoals 'geloofwaardig' in plaats van 'waarachtig'. In andere gevallen is er een voorkeur om een begrip te specificeren binnen het subcriterium van de rubrics.

Taak van docent

Bij een aantal criteria zeggen docenten dat deze in principe de taak van de docent zijn en niet van de leerlingen. Criteria zoals 'lef en durf' en 'verbeelding en fantasie', worden gemeten bij de leerling, maar voor een groot deel beïnvloed door de docent. Ook het creëren van een veilige spelomgeving zien de docenten als hun taak:

'Nou ja, als een leerling altijd actief bezig is met creëren vanuit een veilige leeromgeving, dan is hij niet met drama bezig. Want dan is hij alleen maar met de groep bezig en veiligheid. Dan neemt hij de rol van de docent over en dan is hij niet vrij, is hij niet aan het spelen.'

De niveauomschrijvingen bij dit criterium zouden volgens de meeste respondenten, zowel leerlingen als docenten, enkel moeten bestaan uit twee niveaus, namelijk het wel of niet in stand houden van de door de docent gecreëerde veilige spelomgeving.

De operationalisering

Binnen operationalisering kijken we naar de manier waarop verschillende criteria worden gemeten. Bijvoorbeeld: valt het subcriterium 'geloofwaardigheid' onder het hoofdcriterium 'spelvaardigheid' en kloppen de afstanden tussen de verschillende niveaus? Hiervoor zijn we ingegaan op de niveauomschrijvingen, de indeling en de mate waarin de criteria de inhoud dekken.

Niveauomschrijvingen

Docenten hebben verschillende visies over de vier niveauomschrijvingen binnen de rubrics. Zo wil een docent graag een vijfde niveau toevoegen, zodat leerlingen ook de veilige middenweg kunnen kiezen. Een andere docent wil juist liever terug naar drie niveaus vanuit het idee: 'Als je iets goed doet, doe je het goed, dan kun je het niet zéér goed doen.'

Voor de normering van de criteria is het volgens de respondenten wel van belang dat er consequent gebruik gemaakt wordt van kwalificatie en kwantificatie. In de huidige rubrics wordt dit regelmatig binnen eenzelfde criterium door elkaar gebruikt (bijvoorbeeld: zeer goed, goed, soms). Hiernaast moeten aanduidingen zoals 'in delen van het spel' voorkomen worden, omdat respondenten dit subjectief vinden. Ook subjectieve begrippen als 'geloofwaardigheid' zorgen voor grote diversiteit in interpretaties.

De veel terugkerende laagste normering 'vindt het moeilijk', noemen leerlingen 'lief' en 'minder lullig'. Er is geen overeenstemming of dit positief of negatief is, maar stelt wel vraagtekens bij de vraag of deze formulering de juiste is voor een laagste normering. Het niet consequent gebruiken van dezelfde normering veroorzaakt een discussie over deze laagste niveauomschrijvingen. Er kan volgens de respondenten vaker gebruik gemaakt worden van de gradatie: goed, ruim voldoende, voldoende, onvoldoende. Hiermee kun je subjectieve moeilijke begrippen zoals: 'waarachtig' vermijden. Specifiek over het hoofdcriterium 'theorie' zeggen ze dat deze gradatie de enige woorden binnen de niveauomschrijvingen zouden moeten zijn, zonder extra toelichting. Bij sommige aspecten zoals 'geloofwaardigheid' is het tonen van geloofwaardig spel hiervan het hoogst haalbare, waardoor zeer geloofwaardig spel in de niveauomschrijvingen weg valt.

Indeling

Binnen indeling hebben we gekeken naar overlap binnen of juist het liever uitsplitsen van (sub)criteria. Wanneer de subcriteria niet specifiek zijn en meer onderdelen dekken, gaat dit volgens de respondenten ten koste van de toetsbaarheid. Door de combinatie van bijvoorbeeld 'verbeelding/ fantasie', 'fysiek spel/ transformatie' en 'reflecteren op eigen werk en dat van anderen', wil het subcriterium te veel beoordelen. In deze gevallen moeten de subcriteria losgekoppeld worden, met ieder eigen bijbehorende niveauomschrijvingen:

'Maar ik vind dat reflecteren op eigen werk en dat van anderen, dat zijn eigenlijk twee verschillende dingen.'

Een aantal respondenten vindt dat veel subcriteria uitgesplitst moeten worden omwille van de helderheid en eenduidigheid. Deze uitsplitsing betekent geen opsplitsing, waar-

door er meer criteria ontstaan. Het gaat om een specificering van begrippen en vaardigheden die het criterium verhelderen, ofwel een toevoeging binnen de ruimte voor het subcriterium. Dit kan tot gevolg hebben dat andere subcriteria komen te vervallen. In plaats van de subcriteria in het model te specificeren kan dat ook in een mondelinge uitleg aan de leerlingen:

‘Tekstbehandeling gaat voor mij ook over verstaanbaarheid. Over intonatie, over ritmiek, over spanningsopbouw. Hoe behandel je überhaupt een tekst? Kun je die voor jezelf zo analyseren en weer opnieuw levend maken dat hij ook zo overkomt. Daar komt meer bij kijken. Daar komen eigenlijk allerlei stemvaardigheden ook nog eens een keer bij kijken.’

Wanneer binnen de niveauomschrijvingen twee of meer verschillende vaardigheden aan bod komen, vinden de meeste respondenten dat het de kwaliteit van de beoordeling ten goede zou komen als deze losgekoppeld worden. Zo worden het meer subcriteria, ieder met eigen niveauomschrijvingen. Een voorbeeld zien we in het subcriterium ‘regisseren van een scène’. Hier is het criterium eenduidig, maar reppen de niveauomschrijvingen over het hebben van een beeld en het overbrengen (ofwel regisseren) van het beeld. De meeste respondenten vinden het beter als beide vaardigheden apart beoordeeld worden met een eigen reeks niveauomschrijvingen. Dat heeft ook te maken met het geven van goede feedback: zo kan de leerling exacter bepalen welke vaardigheden hij nog moet verbeteren.

Naast het mogelijk opsplitsen en loskoppelen van criteria en niveauomschrijvingen zeggen veel respondenten iets over het verwarrende effect van overlappingen. Zo staat er meermalen ‘verbeelding’ in de rubrics. Zodra een vaardigheid als basisvaardigheid is beoordeeld, is het voor de leerlingen verwarrend als het binnen een ander hoofdcriterium nogmaals wordt beoordeeld. Deze verwarring ontstaat omdat er binnen het subcriterium niet is beschreven wat hier specifiek beoordeeld wordt:

‘Nou ja, wat is het verschil tussen basisvaardigheid verbeelding en maken van de verbeelding?’

De eerder genoemde specificatie is een mogelijkheid om het onderlinge verschil tussen eenzelfde vaardigheid binnen verschillende hoofdcriteria te verhelderen.

Mate waarin criteria de inhoud dekken

Dit aspect hangt nauw samen met de normering, niveauomschrijvingen en meetbaarheid. Zo hebben de respondenten het tijdens de interviews veelvuldig over wat goed en wat slecht is. Binnen de huidige rubrics zijn er volgens de leerlingen en docenten een aantal aanpassingen gewenst. Vaak genoemd is het veranderen van ‘vindt het moeilijk’ naar een niveau dat ondermaats presteren beter verwoordt zoals: ‘onvoldoende’ of ‘niet’:

‘Als je het moeilijk vindt om het in te zetten, betekent dat nog niet dat je het niet doet.’

Bij veel criteria ontbreekt dit niveau, terwijl dit wel aanwezig kan zijn binnen een groep en dus ook als zodanig beoordeeld moet kunnen worden. Volgens de leerlingen wordt er in de huidige Rubrics uitgegaan van inzet van de leerling en er is binnen de niveauperichillen geen mogelijkheid om deze inzet te meten en te beoordelen.

Ten slotte dekt de kwantitatieve aanduiding ‘altijd’ zelden tot nooit de werkelijkheid, omdat het onmogelijk is om een vaardigheid te allen tijde te tonen en juist in te zetten.

8.4 De praktische toepasbaarheid van het model

Hieronder gaan we in op de algemene reactie van de respondenten op het beoordelingsmodel, het inzetten van het instrument in de eigen lespraktijk, het digitale aspect van het instrument en de vergelijking met de eigen beoordelingsmethode van de docent.

Algemene reactie op het beoordelingsmodel

Alle docenten vinden het waardevol dat er aandacht besteed wordt aan een overkoepelend beoordelingsmodel en dat er een breder draagvlak gevonden wordt voor één beoordelingsmodel:

'Ik vind het heel erg nodig dat zo'n instrument wordt ontwikkeld, het zou heel mooi zijn als er iets is waar een veel breder draagvlak voor zou zijn, zodat je ook met meer collega's tegelijkertijd eigenlijk een soort gereedschap in handen hebt voor die beoordeling.'

Ze spreken van 'een dappere poging' en 'een uitvoerig, goed beschrijvend instrument'. Het feit dat je het model zelf kan samenstellen helpt een docent ook om te focussen en zorgt voor een positieve kijk op het instrument. Wel vragen ze zich af of je verschillende subjectieve vaardigheden moet proberen meetbaar te maken en te vangen in hokjes:

'Je kunt dat niet concreet maken, maar je leert wel vaardigheden die van essentieel belang zijn om goed te kunnen functioneren in de maatschappij.'

De leerlingen vinden het prettig om te weten waar zij op worden beoordeeld. Ze geven de voorkeur aan dit model boven de huidige manier van beoordelen, zodat ze tijdens de lessen weten waar ze op moeten letten. Een goede uitleg is volgens hen wel nodig om het model te begrijpen.

Inzetten van het beoordelingsmodel

De docenten zouden de rubrics ieder op andere momenten in willen zetten. De ene docent zou ervoor kiezen om het alleen in te zetten voor de eindbeoordeling, omdat het invullen van de rubrics veel tijd vergt en het voor een les tussendoor te weinig op zal leveren in verhouding tot de tijdsinvestering. Een andere docent vindt het een goed instrument voor gebruik in de lessen, om het gesprek aan te kunnen gaan met leerlingen, waarbij hij het vooral als reflectie-instrument of een goed naslagwerk zou gebruiken. Deze docent hecht wel meer waarde aan een uitgeschreven beoordeling en wil deze beoordeling 'niet in hokjes' (de rubrics) stoppen.

De hokjes ervaren de docenten als niet breed genoeg, ze vinden dat deze het vak drama tekort doen. Volgens hen zijn de vaardigheden die zij de leerlingen leren breder dan in rubrics gevat kan worden. De rubrics moeten in ieder geval niet gezien worden als heilige graal:

'Ik heb er een ambivalent gevoel over. (...) ik vind het ook goed dat er onderzoek naar wordt gedaan. Ik vind tegelijkertijd dat dit niet een meetlat moet worden, want daarvoor vind ik het niet valide genoeg (...) Ik gebruik het ook als instrument om leerlingen zelfbewust te maken en ze zelf te laten kijken naar het spel, maar ik vind het geen eindbeoordelingsinstrument.'

Leerlingen zeggen dat het invullen van de rubrics tijdens de les wel mogelijk is, maar dat daar wel veel tijd voor gemaakt moet worden, omdat zij goed na moeten denken over de verschillende punten:

'Maar ik vind wel dat je nogmaals heel goed naar de punten eerst moet kijken, echt, wat staat er precies? Want nu we het zo aan het behandelen zijn, snap ik het meer dan dat je het de eerste keer gaf. Toen dacht ik van, oké, blablabla.'

Dit geldt voor zowel het beoordelen van zichzelf als medeleerlingen. Zelfbeoordeling vinden ze moeilijk omdat ze zichzelf niet zien spelen. Ook vragen leerlingen zich af in hoeverre hun spel objectief te meten is, aangezien ze altijd samenwerken en afhankelijk zijn van anderen tijdens het maken of spelen van een scène.

Digitale aspect

Docenten vinden het positief dat zij zelf de rubrics samen kunnen stellen. Ze ondersteunen het belang van een open instrument, met de mogelijkheid om zelf subcriteria toe te kunnen voegen of de niveauomschrijvingen aan te kunnen passen.

Docenten willen of kunnen de chatsessies niet toepassen, omdat zij teveel leerlingen lesgeven en de leeropbrengst niet opweegt tegen de tijd die het kost met al die leerlingen te chatten. Het digitale aspect en het feit dat het rubrics zijn maken dat docenten het model als 'te afgebakend' ervaren. Ze geven de voorkeur aan een persoonlijk gesprek boven de chatfunctie. Ze zouden dit wel in overweging willen nemen als het bijvoorbeeld gaat om een specifieke (kleine) klas, omdat de leerlingen dan nog meer inzicht krijgen in hun beoordeling en leerproces:

'Ik vind dat op zich heel goed en in een klas met twaalf leerlingen, 5 vwo, zou ik dat gebruiken. Maar ja, als je... Ik heb vrijdag negen uur les gegeven in klassen met dertig leerlingen... dan ben ik het hele weekend bezig om te chatten met een leerling waar die staat (...). De tijdsinvestering weegt niet op tegen het resultaat.'

Deze rubrics werken met Google Docs, maar veel scholen hebben afspraken over de digitale leeromgeving waarmee gewerkt wordt (bijvoorbeeld Magister). Google Docs zou dus omgezet moeten kunnen worden naar een ander digitaal programma, om het binnen elke school toe te kunnen passen.

De eigen beoordelingsmethode van de docent

Docenten vinden de rubrics uitvoeriger en uitgebreider dan de eigen beoordelingsmethode. Maar ze zien ook veel onderdelen uit hun eigen methode terug in de rubrics. Hen vallen daarbij verschillende zaken op. Eén docent beoordeelt inzet en motivatie bewust niet: 'Ook bij wiskunde ga je niet toetsen of iemand een goede werkhouding heeft'. Binnen het instrument van Lok komen deze onderdelen wel aan bod.

De wijze waarop docenten beoordelen verschilt. Soms gebruiken ze eigen rubrics, soms gebeurt dit meer intuïtief, soms beoordelen leerlingen zichzelf en elkaar en vaak krijgen ze persoonlijke feedback én een cijfer. Leerlingen vertellen ook dat ze tips en tops krijgen plus een cijfer. Proces en product worden vaak los beoordeeld.

8.5 Aanbevelingen

Dit onderzoek heeft tot doel antwoord te geven op de onderzoeksvraag: In hoeverre is het beoordelingsinstrument voor dramalessen in het voortgezet onderwijs van Lok (2013) toepasbaar en inhoudsvalide volgens docent en leerling? Omdat we vanuit de acht geselecteerde criteria aanbevelingen willen doen voor het gehele instrument, hebben we waar mogelijk de criteria specifieke aanbevelingen herschreven, zodat ze toepasbaar zijn op het gehele instrument.

De (digitale) toepasbaarheid

Zoals meer docenten hebben opgemerkt, kunnen we stellen dat de Master Rubric een 'dappere poging' is om het vak drama voor havo en vwo inhoudelijk dekkend, valide en betrouwbaar te beoordelen. Het beoordelingsinstrument is een goede, uitvoerige en volledige basis en docenten kunnen niet iets specifiek noemen wat ze er niet in terug kunnen vinden. Verder sluit het volgens de docenten goed aan op de exameneisen voor het vak drama. Deze volledigheid gaat echter verloren als een docent voor een project een selectie maakt uit de zevenendertig criteria.

De praktische toepasbaarheid is van grote invloed op de functionaliteit van het instrument. De digitale omgeving van het instrument is minder praktisch dan op het eerste gezicht lijkt. Zo vinden docenten het samenstellen van een beoordelingsformulier per leerling en de bijbehorende chatfunctie te arbeidsintensief. Het huidige instrument is ontwikkeld voor Google Docs, maar veel scholen hebben niet de mogelijkheid en de vrijheid om binnen hun onderwijs voor een andere digitale leeromgeving te kiezen dan het platform waar ze al mee werken. Het dynamische karakter van de digitale vorm, waarbij de docent de criteria zelf kan samenstellen, functioneert wel goed. Het belang van de mogelijkheid om eigen subcriteria of niveaumschrijvingen toe te voegen is groot en zal, ook als het instrument op andere digitale platforms kan werken, aanpasbaar moeten blijven.

De rubrics vinden respondenten beperkend als het gaat om het dekken van het gehele vak drama. Daarom zou dit instrument niet moeten dienen als enige beoordelingsmethode. Het heeft de voorkeur om hier een uitgeschreven beoordeling van de docent aan toe te voegen, die niet wordt beperkt door de hokjes van een rubric.

Inzet van het instrument tijdens de les vergt veel tijd, zodat leerlingen de tijd en aandacht hebben om de veelomvattende mogelijkheden van het instrument te benutten. Omdat deze tijd vaak niet beschikbaar is, is het de vraag of het instrument deze directe toepassing in de les moet willen nastreven.

Binnen de criteria moet uitvoeriger uiteengezet en beschreven worden welke punten er specifiek onder vallen, zodat de rubrics concreter en minder discutabel worden. Ter verbetering van het instrument moet er nader onderzocht worden, in hoeverre het de exameneisen voor het vak drama in het voortgezet onderwijs toetst.

Inhoudelijk

Voor de bruikbaarheid en een correcte inhoudsvaliditeit zijn verbeteringen noodzakelijk. Deze verbeteringen zijn op het gehele model toepasbaar en op specifiek niveau per criterium. Een veel terugkerende kritiek is dat begrippen en kwantitatieve en kwalitatieve aanduidingen inconsequent worden toegepast binnen het instrument.

Het is helder dat er, indien mogelijk, één lijn getrokken zou moeten worden in de aanduiding van de niveaoverschillen. Een valkuil van de rubrics is wel dat er uiteindelijk overall 'goed, ruim voldoende, voldoende, matig en onvoldoende' staat. In dat geval is het de vraag wat de meerwaarde van de niveauomschrijvingen is en of er dan niet beter gewerkt kan worden met deze reguliere termen of met een lijn of hokjes waarbij leerlingen aan kunnen kruisen op welk niveau ze zich bevinden binnen een specifieke rubric.

Docenten werken binnen de dramalessen niet of nauwelijks aan het schrijven van scènes. Daarom moet het criterium hiervoor teruggebracht worden tot één subcriterium.

Binnen de niveauomschrijvingen moeten specifieke woorden die zich richten op een enkel aspect van een criterium worden voorkomen.

Het is van belang dat subcriteria voldoende uiteen worden gezet zodat iedereen precies weet wat er wordt beoordeeld. Daarnaast is het van belang dat de docent het instrument uitvoerig bespreekt met de leerlingen.

Wanneer een te beoordelen criterium een groepsproces is (zoals 'regisseren van een scène') moet de niveauomschrijving zich richten op het zichtbare en meetbare individuele aandeel.

Termen moeten binnen eenzelfde rubric consequent gebruikt worden (bijvoorbeeld 'spelomgeving' of 'leeromgeving'). Vaktermen zoals 'fysiek spel', 'fysiek vormgeven', 'waarachtig' en 'transformatie' moeten vermeden worden of duidelijk gespecificeerd en waar nodig opgesplitst in meer criteria, zodat er geen onduidelijkheid over de betekenis kan ontstaan.

Wanneer het de taak van een docent is om een juiste voorwaarde voor een criterium te creëren, zoals 'creëren veilige spelomgeving', moet de niveauomschrijving enkel beoordelen of de leerling deze veiligheid behoudt of niet, wat betekent dat dit maar twee niveaus worden.

Operationalisering

Het onvoldoende zichtbaar zijn van een criterium bij een leerling moet ook als 'onvoldoende' beoordeeld kunnen worden. De niveauomschrijving 'vindt het moeilijk' is dan niet toereikend.

Wanneer de aanwezigheid van een vaardigheid het hoogst haalbare is, moet dit ook de hoogste niveauomschrijving zijn. Zo is 'gelofwaardig' al het hoogst haalbare en kan 'zeer gelofwaardig' dus vervallen. Bij het criterium 'theorie' is een niveauomschrijving van goed, voldoende, matig en onvoldoende toereikend. Wanneer bij een criterium slechts sprake is van voldoende en onvoldoende niveau, dan volstaan twee niveauomschrijvingen.

De onderlinge verhoudingen tussen de niveaus van het beste tot en met het slechtste, moeten consequent gebruikt worden. De omschrijvingen van de niveaoverschillen moeten binnen het grote aantal criteria zo min mogelijk variëren.

Wanneer er twee vaardigheden worden benoemd binnen één subcriterium of niveauomschrijving moeten hier twee subcriteria met eigen niveauomschrijvingen

van worden gemaakt. Subcriteria die bestaan uit een grote hoeveelheid vaardigheden, moeten gespecificeerd worden. Dat komt de begrijpelijkheid van het criterium ten goede. Wanneer binnen de Master Rubric onder verschillende hoofdcriteria dezelfde subcriteria te vinden zijn, moet er een specificatie binnen het subcriterium komen. Kwantitatieve aanduidingen zoals 'altijd' moeten vermeden worden binnen de niveaumschrijvingen, omdat het onmogelijk is om een vaardigheid altijd in te zetten dan wel te tonen.

8.6 Discussie

We gaan hier in op een aantal punten die uit de interviews naar voren zijn gekomen. Dit zijn punten waarover helemaal geen overeenstemming tussen de respondenten bestond of waar wij als onderzoekers onze vragen bij hebben.

Toepasbaarheid

Het instrument is ontwikkeld om leerlingen meer inzicht en houvast te geven bij de eigen ontwikkeling. Het kan toegepast worden als reflectie-instrument en als beoordelingsinstrument. Voor beide vormen zijn vóórs en tegens te benoemen, mede afhankelijk van het niveau van de leerlingen, de leerdoelen van de lessen en de docent. We denken dat het van belang is dat de docent zelf bepaalt wanneer en hoe hij met welke rubrics werkt. Daarbij moet bekeken worden of het gebruik van andere digitale varianten in plaats van Google Docs meer voordelen en gebruikersgemak oplevert en welke varianten dit dan zijn. De rubrics bieden goede handvatten voor gesprekken met leerlingen over hun ontwikkeling, daarnaast zal een uitgeschreven beoordeling of reflectie vaak noodzakelijk zijn.

Meetbaarheid

Vaardigheden zoals 'gelofwaardigheid', 'waarachtigheid' en 'tekstbehandeling', maar ook 'origineel' en 'sterk' of de kwantitatieve aanduiding 'in delen', zijn moeilijk meetbaar. De beoordeling hiervan is snel subjectief en daarmee minder betrouwbaar. De vraag is of deze woorden zoveel mogelijk vermeden of vervangen moeten worden binnen de rubrics door objectief meetbare begrippen. Wij vragen ons echter af ook of dit mogelijk is en of je daarmee bepaalde criteria niet tekort doet. Een beter alternatief is om een subjectieve vaardigheid te specificeren. In de rubrics zouden de verschillende specificaties benoemd kunnen worden, zodat de docent kan kiezen wat hij op dat moment wil beoordelen, bijvoorbeeld dat het bij het beoordelen van 'gelofwaardigheid' om het uitspelen van emoties gaat. Binnen een veelomvattend begrip als tekstbehandeling zouden specificaties als verstaanbaarheid, intonatie, spanningsopbouw, tekstbegrip, overbrengen van tekst op publiek benoemd kunnen worden.

Mate waarin de criteria de inhoud dekken

De vraag is of ook de inzet van leerlingen bij elke vaardigheid gemeten moet worden. Dit betekent een toevoeging op de niveaumschrijvingen. Hiermee kan de docent een beter beeld krijgen van het leerproces van de leerling. Aan de andere kant worden leerlingen bij andere vakken ook niet beoordeeld op inzet, dus waarom bij drama wel?

Wij zijn van mening dat drama wel degelijk iets anders vraagt van leerlingen dan andere vakken. Inzet is wel degelijk van belang. Ook hier is een specificering zeer gewenst: Inzet binnen drama gaat namelijk om zaken als: Zet de leerling zich in om de verschillende rollen die een drama's van hem vraagt goed uit te voeren (zoals samenwerker, publiek, regisseur, speler)? Zet de leerling zich tijdens alle dramaopdrachten in? Zet de leerling zich tijdens samenwerking in verschillende groepsamenstellingen in? Zet de leerling zich in om de verschillende leerdoelen te bereiken? Zet de leerling zich in om bij te dragen aan een veilige spelomgeving?

Reflectie op het onderzoek

Het voordeel van het toetsen van een specifiek beoordelingsinstrument is dat de resultaten zeer concreet en makkelijk toe te passen zijn op dit beoordelingsinstrument. We hebben ervoor gekozen om alle interviews in nieuwe rubrics uit te werken. Dit levert

zeven aangepaste rubrics op. Als onderzoekers vonden wij dit de meest zorgvuldige vorm. Dit betekent wel dat er veel in details getreden wordt en dat er gevaar van overlap bestaat.

We hebben gemerkt dat we met zeven verschillende interviews ruim voldoende informatie hebben om een vertaalslag te kunnen maken naar het gehele instrument. Alle aanbevelingen zouden hierin verwerkt kunnen worden. Het zou daarna wel van belang zijn om nogmaals de cyclus van interviews te doorlopen binnen nieuwe scholen en met andere docenten, om te kijken of de inhoudsvaliditeit van het instrument voldoende verbeterd is.

Literatuur

Baarda D.B., Goede M.P.M. de & Teunissen J. (2009). *Basisboek Kwalitatief Onderzoek. Handleiding voor het opzetten en uitvoeren van kwalitatief onderzoek*. Groningen/ Houten: Noordhoff Uitgevers.

Drenth, P. J. D. & Sijsma. K. (1990). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen*. Houten: Bohn Stafleu Van Loghum

Klarenbeek, D. & Van der Meulen, B. (2014). *Past drama in een hokje? Een praktijkonderzoek naar de validiteit en toepasbaarheid van een beoordelingsinstrument van Lok (2013), ontwikkeld bij het lectoraat Kunst en cultuureducatie van de AHK, voor dramalessen in het voortgezet onderwijs*. Amsterdam: Praktijkonderzoek Master Kunsteducatie.

Lok, C. (2013). *Beoordeling van theaterlessen in het Voortgezet Onderwijs*. Amsterdam AHK Theaterdocentopleiding

Swanborn, P.G. (1984). *Methoden van sociaal-wetenschappelijk onderzoek. Inleiding in ontwerpstrategieën*. Meppel: Boompers drukkerijen bv

