

Assessment in kunsteducatie

Talita Groenendijk
Marie-Louise Damen
Folkert Haanstra
Carla van Boxtel

Eindrapport NWO Review studie 411-12-228

Universiteit van Amsterdam
Research Institute of Child Development and Education
Nieuwe Achtergracht 127
1018 WS Amsterdam

Contact:

C.A.M.vanBoxtel@uva.nl

Maart 2015

Inhoudsopgave

1 Looptijd project	3
2 Wetenschappelijke output en activiteiten	3
3 Samenvatting	4
4 Opzet en resultaten van de review studie	6
4.1 Inleiding	6
4.1.1 <i>Afbakening van de studie</i>	6
4.2 Methode	7
4.2.1 <i>De zoekopdracht</i>	7
4.2.2 <i>Selectiecriteria</i>	8
4.2.3 <i>Analyse</i>	9
4.2.4 <i>Overzicht van de instrumenten per kunstdiscipline en schoolniveau</i>	9
4.3 Resultaten	10
4.3.1 <i>Soorten instrumenten</i>	10
4.3.2 <i>Taken</i>	10
4.3.3 <i>Scoringsinstrumenten</i>	12
4.3.4 <i>Formatief / summatief toetsen</i>	13
4.3.5 <i>Soorten competenties</i>	14
4.3.6 <i>Beoordelingscriteria</i>	14
4.3.7 <i>Kwaliteit</i>	20
4.4 Samenvatting en conclusie	22
Bijlage Zoekwoorden review	25

1 Looptijd project

Het project startte op 1 juni 2013 en is afgerond op 31 mei 2014.

2 Wetenschappelijke output en activiteiten

Wetenschappelijk artikel

Groenendijk, T., Damen, M., Haanstra, F., & van Boxtel, C. (submitted).
Beoordelingsinstrumenten in de kunstvakken - een review.

Bijdrage aan boek

Haanstra, F., Damen, M., Groenendijk, T., & van Boxtel, C. (in press, 2015). A review of assessment instruments in arts education. In S. Schonmann, *The wisdom of the many. Key issues in arts education. International Yearbook for Research in Arts Education*. New York : Waxman.

Vakpublicatie

Groenendijk, T., Damen, M., Haanstra, F., & van Boxtel, C. (2014).
Beoordelingsinstrumenten in danseducatie. In F. Haanstra (Red.), *Beoordelen in de kunstvakken: instrumenten en onderzoek* (pp. 55-67). Amsterdam: Lectoraat Kunst- en Cultuureducatie Amsterdamse Hogeschool voor de Kunsten.

Presentaties op congressen

Groenendijk, T. (2015). *Beoordelen in het kunstonderwijs*. Presentatie op het VELON Congres van Lerarenopleiders, Arnhem, 27 maart 2015.

Damen, M.L. (2014). *Review van beoordelingsinstrumenten in de kunstvakken: context, opzet en globale resultaten*. Paper gepresenteerd op de Onderzoeksconferentie Cultuureducatie en Cultuurparticipatie, 24 november 2014.

Haanstra, F. (2014). *Review van beoordelingsinstrumenten in de kunstvakken: gemeten competenties en criteria*. Paper gepresenteerd op de Onderzoeksconferentie Cultuureducatie en Cultuurparticipatie, 24 november 2014.

Groenendijk, T. (2014). *Review van beoordelingsinstrumenten in de kunstvakken: de kwaliteit van de instrumenten*. Paper gepresenteerd op de Onderzoeksconferentie Cultuureducatie en Cultuurparticipatie, 24 november 2014.

Haanstra, F., Groenendijk, T., Damen, M., & van Boxtel, C. (2014).
Beoordelingsinstrumenten in de kunstvakken. Symposium op de Onderzoeksconferentie Cultuureducatie en Cultuurparticipatie, november 2014, Utrecht.

Groenendijk, T., Damen, M., Haanstra, F., & van Boxtel, C. (2014). *Een reviewstudie naar beoordelingsinstrumenten in de kunstvakken*. Paper gepresenteerd op de Onderwijs Research Dagen juni 2014, Groningen.

Groenendijk, T., & Damen, M. (2013). *Assessment in Arts Education*. Paper gepresenteerd op de Onderzoeksconferentie Cultuureducatie, november 2013, Tilburg.

Website

De resultaten van het onderzoek zijn toegankelijk via de informatiebank van het Landelijk

Kennisinstituut Cultuureducatie en Amateurkunst.

<http://www.lkca.nl/kennisdossiers/beoordelen/reviewstudie-assessment-in-kunsteducatie>

Het gaat om een samenvatting van het onderzoek, de complete literatuurlijst en een overzicht van de kenmerken van de beoordelingsinstrumenten die zijn opgenomen in de review.

In voorbereiding

-Een artikel voor een Engelstalig peer reviewed tijdschrift op het domein van de kunstvakken.

3 Samenvatting

Met het programma 'Cultuureducatie met kwaliteit' tracht de Nederlandse overheid de ontwikkeling van doorgaande leerlijnen en van een beoordelingsinstrumentarium te stimuleren. Voor leerlijnen zijn landelijke leerplankaders ontwikkeld die als richtsnoer kunnen dienen. Voor het ontwikkelen van een beoordelingsinstrumentarium ontbreekt zo'n richtsnoer. Er bestaat wel literatuur op dit gebied. Deze beschrijft een grote variëteit aan beoordelingsinstrumenten en vormen. Naast het onderscheid in kunstdisciplines en het onderscheid in de belangrijkste kunstdomeinen (actief en receptief), wordt onderscheid gemaakt tussen de verschillende functies van de beoordelingsvormen: formatief, summatief en evaluatief. De verschillende domeinen en functies resulteren in verschillende beoordelingsvormen. Wat echter ontbreekt is een overzicht van beoordelingsinstrumenten waarin niet alleen wordt aangegeven wat en hoe beoordeeld wordt maar ook gegevens worden verzameld over kwaliteitscriteria zoals validiteit, betrouwbaarheid, bruikbaarheid en eerlijkheid.

In deze review werden de volgende onderzoeksvragen beantwoord:

- a) Welke beoordelingsinstrumenten voor de kunstvakken in het basis- en voortgezet onderwijs zijn in de literatuur sinds 2000 beschreven?,
- b) Wat wordt er met deze instrumenten beoordeeld?
- c) Wat wordt gerapporteerd over de kwaliteit van deze instrumenten?

De review omvatte vier fasen. De eerste fase betrof de zoekopdracht zelf. In databases als Eric, Psychinfo en kunstvakspecifieke databases is met een zoekprofiel een search gedaan naar literatuur van de afgelopen dertien jaar (2000-2013). Dit resulteerde in 8336 titels.

In de tweede fase is in enkele stappen een selectie van de publicaties gemaakt. De eerste selectie betrof publicaties met een beschrijving van een instrument of procedure van beoordeling om het leren van leerlingen in de kunstvakken basis- of voortgezet onderwijs te monitoren of beoordelen. Om tot een verdere inperking van resultaten te komen, werden alleen peer reviewed journal artikelen opgenomen. Tevens werd besloten dat het artikel iets moest zeggen over de kwaliteit van het instrument of de procedure, zoals bijvoorbeeld de validiteit en betrouwbaarheid. Dit resulteerde in 119 artikelen uit 44 peer reviewed Engelstalige tijdschriften (allen uitgegeven in de Verenigde Staten of Groot-Brittannië), waarin 153 beoordelingsinstrumenten zijn beschreven

In de derde fase zijn de geselecteerde instrumenten gecodeerd en beschreven. Codes betreffen bibliografische gegevens, kunstdiscipline, onderwijsniveau, toetstaak (artistieke uiting, pen-en-papier opdracht, portfolio, etc.); type scoringsinstrument (rubric, criterialijst, observatielijst etc.); object van meting (create, perform, respond), functie (summatief of formatief); proces of product beoordeling; beoordelingscriteria en kwaliteitskenmerken.

In de vierde fase vond een kwalitatieve synthese plaats en werden de onderzoeksvragen beantwoord.

De review studie laat zien dat de beoordelingstaken en scoringsinstrumenten voor de kunstvakken divers zijn. Er is een vergelijkbaar aantal instrumenten gericht op primair onderwijs en voortgezet onderwijs. De meeste beoordelingsinstrumenten die in de wetenschappelijke literatuur worden beschreven hebben betrekking op muziek. Voor drama is slechts een gering aantal instrumenten gevonden. Meestal gaat het om de beoordeling van een uiting in het artistieke medium (zoals een muziekkuitvoering of een beeldend product) aan de hand van criterialijsten of rubrics. De meeste instrumenten richten zich op competenties die vallen onder 'perform' en 'create'. Elke kunstdiscipline heeft zijn eigen

terminologie en vakspecifieke competenties. Beoordelingscriteria die te maken hebben met het beschouwen van kunst of met reflecteren komen relatief weinig voor in de review. De instrumenten beperken zich niet tot makkelijk te meten kennis of technische vaardigheden, maar hebben ook betrekking op kwalitatieve aspecten, zoals bijvoorbeeld expressie, toonkwaliteit, sfeer en originaliteit. De kwaliteitsmetingen zijn erg heterogeen. Er konden twintig verschillende kwaliteitsoordelen onderscheiden worden. De instrumenten die in de review zijn opgenomen scoorden in bijna alle gevallen een voldoende tot goed op de gekozen kwaliteitsmetingen. In een aantal artikelen werd de kwaliteit van verschillende beoordelingsinstrumenten met elkaar vergeleken.

De review beperkte zich tot publicaties in Engelstalige peer-reviewed tijdschriften waarin iets wordt gezegd over de kwaliteit van de beschreven instrumenten. In vervolgonderzoek zou ook gekeken kunnen worden naar handboeken voor de praktijk en niet-Engelstalige literatuur. Vervolgonderzoek kan zich tevens richten op instrumenten voor beoordeling binnen de discipline theater, de beoordeling van samenwerking bij uitingen in het artistieke medium en van kunstbeschouwing en reflectie. Over deze aspecten van beoordelen in de kunstvakken wordt nog nauwelijks gerapporteerd in wetenschappelijke peer-reviewed tijdschriften.

Ook in de kunstvakken worden leerlingen steeds vaker beoordeeld. Deze beoordeling kan – passend bij de specifieke kunstdiscipline en de visie op beoordelen – op verschillende manieren. De review biedt leerkrachten, vakexperts en onderzoekers een overzicht van mogelijke beoordelingsinstrumenten (taken en scoringsinstrumenten) voor de beoordeling van het leren binnen de kunstvakken en van mogelijke beoordelingscriteria voor verschillende competenties.

4 Opzet en resultaten van de review study

4.1 Inleiding

Volgens de Onderwijsraad en de Raad voor Cultuur (2012) ontbreekt het bij kunstvakken in het onderwijs (zoals beeldende vakken, muziek, dans en theater) aan opbrengstgericht werken. Er is volgens de raden vooral in het primair onderwijs behoefte aan leerlijnen waarin de gewenste ontwikkeling van kennis, vaardigheden en houding beschreven worden. Dit vraagt om een uitbreiding van het monitor- en beoordelingsinstrumentarium, waarmee docenten een beter inzicht krijgen in de culturele ontwikkeling van hun leerlingen. Het daarop door OCW geïnitieerde programma 'Cultuureducatie met kwaliteit' tracht de ontwikkeling van doorgaande leerlijnen en van een beoordelingsinstrumentarium te stimuleren. Voor leerlijnen zijn landelijke leerplankaders ontwikkeld die als richtsnoer kunnen dienen, zoals het leerplankader kunstzinnige oriëntatie voor het primair onderwijs van de SLO (2014) en het leerplankader Cultuur in de Spiegel voor primair en voortgezet onderwijs (Hoeven e.a., 2014). Voor het ontwikkelen van een beoordelingsinstrumentarium ontbreekt zo'n richtsnoer. Er bestaat uiteraard wel literatuur op dit gebied. Eind jaren negentig zijn in de Verenigde Staten verschillende overzichten van beoordelingsprocedures en instrumenten gepubliceerd (Beattie, 1997; Gardner, 1996). Recenter verschenen review artikelen over beoordeling in de kunsten in handboeken zoals het "New Handbook of Research in Music Teaching and Learning" (Colwell & Richardson, 2002), in het "International Handbook of Research in Arts Education" (Bresler, 2007); in artikelen over historische en huidige beoordelingspraktijken in de kunsteducatie (Taylor, 2006), en in een boek over formatieve beoordeling in muziekonderwijs (Fautley, 2010). Deze literatuur beschrijft een grote variëteit aan beoordelingsinstrumenten en vormen. Naast het onderscheid in verschillende kunstdisciplines en het onderscheid in de belangrijkste kunst domeinen (actief en receptief), wordt er in de literatuur onderscheid gemaakt tussen de verschillende functies van de beoordelingsvormen: formatief, summatief en evaluatief. De verschillende domeinen en functies resulteren in verschillende beoordelingsvormen variërend van gestandaardiseerde toetsen over de kennis en waardering voor de kunsten tot het creëren van uitingen in het kunstzinnige medium, en van portfolio's tot kritieken.

Herpin, Washington en Li (2012) verzamelden informatie over de huidige beoordelingspraktijken in kunstvakken in de Verenigde Staten en de wensen in het veld daaromtrent. Ze concludeerden dat er een tekort is aan openbaar toegankelijke, kwalitatief hoogwaardige beoordelingsinstrumenten. Hetzelfde geldt voor Nederland. Er is daarom behoefte aan een literatuuroverzicht van beoordelingsinstrumenten, waarin niet alleen wordt aangegeven wat en hoe beoordeeld wordt maar ook gegevens worden verzameld over kwaliteitscriteria zoals validiteit, betrouwbaarheid, bruikbaarheid en eerlijkheid.

4.1.1 Afbakening van de studie

Bij leerplannen van kunstvakken is de gangbare hoofdindeling die tussen (re)productie en receptie, dat wil zeggen tussen het zelf doen en het kijken of luisteren naar. Die indeling geldt voor alle kunstdisciplines. Verdere onderverdelingen in dimensies zijn vaak opsplitsingen van de genoemde tweedeling. In de Amerikaanse nationale assessments (NAEP) wordt in de kunstvakken een onderscheid gemaakt in drie verschillende competenties: Creating (het bedenken en ontwikkelen van nieuwe artistieke ideeën en werk), Performing (realiseren van artistieke ideeën door interpretatie en presentatie), en Responding (begrijpen en evalueren hoe de kunsten betekenis verlenen). Er zijn ook leerplannen waarbij er naast de hoofddimensies productie en receptie een derde dimensie bestaat die op beide andere dimensies betrekking heeft, bijvoorbeeld reflectie. De driedeling productie, receptie en reflectie is ook in Nederland gangbaar (Oostwoud Wijdenes en

Haanstra, 1997). Reflectie is een metacognitieve vaardigheid en wordt in de leerplannen vaak opgevat als het denken over het maakproces en over het kijk- en beoordelingsproces, hetzij tijdens het proces zelf hetzij achteraf. In deze studie focussen we op actieve competenties ((re)creatie en uitvoering), receptieve competenties om kunst te ervaren en te interpreteren en reflectieve competenties in zowel actieve als receptieve kunsteducatie.

Het bevorderen van creativiteit en creatief denken is in kunstvakken een vaak genoemde doelstelling. Mede om die reden wil men niet alleen het beeldende eindproduct of de eindvoorstelling beoordelen, maar ook het artistieke werkproces dat heeft geleid tot het product. Hierbij speelt het bevorderen van creatief denken een rol. In het SLO leerplankader kunstzinnige oriëntatie vormt het creatieve proces (verdeeld in de fasen oriënteren, onderzoeken, uitvoeren en evalueren) het uitgangspunt om de competenties in te delen. Beoordeling van proces en product worden in deze studie onderscheiden.

Ook bij de beoordeling van het leren in de kunstvakken loopt men tegen klassieke problemen aan als validiteit, betrouwbaarheid, haalbaarheid en eerlijkheid. Kunstzinnige processen en producten zijn verbonden met persoonlijke expressie en individuele kwaliteiten, terwijl beoordelingen objectiviteit en vergelijkbaarheid veronderstellen. Deze spanning heeft ertoe geleid dat de houding van docenten in de kunstvakken ten aanzien van beoordeling zelfs is omschreven als vijandelijk (Soep, 2004). Nationale standaarden en gestandaardiseerde examens bestaan in de kunstvakken, maar worden ook bekritiseerd om hun sturende invloed. Analytische beoordeling waarbij duidelijke criteria ten aanzien van proces en product zijn geformuleerd, verhogen de betrouwbaarheid en transparantie. Maar zulke criteria worden ook gewantwoord omdat ze de voorspelbaarheid van uitkomsten vastleggen. In kunstvakken streeft men zelden naar uniforme, voorspelbare leerresultaten. Voor sommigen is daarom een holistische beoordeling meer geëigend. En ook in de kunstvakken hoort men de roep om authentieke beoordeling. Gulikers (2011) definieert authentieke beoordeling als de beoordeling van de uitvoering van een taak die een combinatie van kennis, vaardigheden en houding vraagt en vergelijkbaar is met dergelijke taken in de professionele kunstpraktijk.

Tot nu toe combineren bestaande reviewstudies geen informatie over de verschillende soorten beoordelingsinstrumenten met informatie over de bovengenoemde aspecten van kwaliteitscriteria en authenticiteit. Wij streven ernaar om dit wel te doen. De vragen in deze reviewstudie zijn:

- a) Welke beoordelingsinstrumenten voor de kunstvakken in het basis- en voortgezet onderwijs zijn in de literatuur sinds 2000 beschreven?
- b) Wat wordt er met deze instrumenten beoordeeld?
- c) Wat wordt gerapporteerd over de kwaliteit van deze instrumenten?

4.2 Methode

4.2.1 De zoekopdracht

We startten de zoekopdracht in vijf (sociaal wetenschappelijke en kunstdiscipline specifieke) databases: ERIC, Psycinfo, Arts & Architecture Index, RILM Abstracts of Music Literature en International Index to Performing Arts. Deze databases zijn geselecteerd op basis van relevantie voor de reviewstudie. Gezocht werd naar literatuur vanaf het jaar 2000 tot 2 oktober 2013.

We gebruikten zoektermen zoals gedefinieerd in de thesauri van de verschillende databases. Daarnaast werden onze eigen zoektermen ingevoerd. Gezocht werd in titel, trefwoorden en abstract van de publicatie. Zoekwoorden werden opgesteld in drie velden die gecombineerd werden: 'kunsten', 'beoordeling' en 'leerlingen'. Tevens werden velden aangemaakt om 'language arts' en 'therapy' uit sluiten omdat deze woorden veel irrelevantie

publicaties opleverden. We kwamen uit op 155 zoektermen voor de verschillende databases (zie Bijlage). De zoektocht leverde 7909 unieke publicaties op. Tenslotte werd nog gezocht in de database van het Landelijk Kennisinstituut voor Cultuureducatie en Amateurkunst (LKCA). Hier kon eenvoudigweg gebruik gemaakt worden van een enkele zoekterm 'toetsing'. Dit leverde nog eens 427 publicaties op.

4.2.2 Selectiecriteria

De selectie van de publicaties heeft plaatsgevonden in verschillende stappen. We begonnen met brede en algemene selectiecriteria. In de vervolgstappen zijn de criteria steeds verder ingeperkt. De publicaties zijn eerst bekeken op relevantie. Op basis van titel en abstract werd besloten of een artikel werd opgenomen in de studie. Criteria bij de eerste selectie van de artikelen waren:

- 1) het artikel betreft één of meerdere kunstdisciplines;
- 2) er wordt een instrument of procedure voor beoordeling besproken en
- 3) het gaat om primair of secundair onderwijs.

Er was sprake van een instrument of procedure voor beoordeling zodra de auteurs één van de volgende zaken beschreven:

- a) een specifiek instrument dat in het artikel is opgenomen (of met voorbeelden van items) of dat ontwikkeld en gepubliceerd is door een professionele organisatie.
- b) formatieve of summatieve beoordelingsprocedures die uitgebreid beschreven worden in beoordelingsterminologie.

Veel publicaties hadden geen betrekking op kunstonderwijs. Het zoekwoord 'art' resulteerde in veel irrelevante 'hits', maar kon uiteraard niet uit de zoekopdracht verwijderd worden. Tevens hadden veel publicaties betrekking op hoger onderwijs. Ook dezen werden niet in de studie opgenomen. Eveneens uitgesloten werden hele specifieke beoordelingsinstrumenten die alleen voor één specifieke les bruikbaar zijn, bijvoorbeeld Weprin (2003) die aanbevelingen doet voor kunstlessen over het ontwerp van auto's. Tevens werden docentenvragenlijsten waarbij gebruikte beoordelingsinstrumenten slechts geïnventariseerd werden niet opgenomen. Deze instrumenten werden te beperkt beschreven. Als het een artikel betrof over aanleg als aangeboren eigenschap, dan werd de publicatie niet opgenomen wegens geringe relevantie voor het onderwijs. Artikelen die twijfel opriepen werden in een aparte map bewaard en door de tweede onderzoeker nogmaals beoordeeld op relevantie.

Deze eerste selectieronde op basis van titel en abstract resulteerde in 624 relevante publicaties. Onder de publicaties waren zowel artikelen waarbij het beoordelingsinstrument of de beoordelingsprocedure onderwerp van de publicatie was, als artikelen waarbij het ging om de meting van een afhankelijke variabele waarbij een artistiek product beoordeeld werd om de effectiviteit van een interventie te bepalen.

Om tot een verdere inperking van het aantal resultaten te komen werden de publicaties in een volgende ronde ingedeeld in 'type' publicatie. We onderscheidden boeken, dissertaties, collected work, conference proceedings, rapporten en journal artikelen. Er werd besloten om enkel peer reviewed journal artikelen op te nemen in de definitieve studie., als ware een extra kwaliteitsindicator. Dit betrof 202 publicaties. De 16 relevante peer reviewed artikelen die we selecteerden uit de LKCA search zaten allen reeds binnen de 202 artikelen die we vonden in de overige databases.

Tijdens de analyse werden de 202 artikelen voor het eerst in full text gelezen. Dit betekende dat er wederom artikelen afvielen wegens irrelevantie. Zo bleken veel artikelen beschouwende artikelen te zijn. Tevens werd besloten dat er ook iets over de kwaliteit van het instrument of de procedure gezegd moest worden. Uiteindelijk werden 119 publicaties

geanalyseerd en opgenomen in de studie. De 119 artikelen zijn afkomstig uit 44 peer reviewed tijdschriften en ze beschrijven gezamenlijk 153 beoordelingsinstrumenten. De tijdschriften waren zonder uitzondering Engelstalig, ze werden uitgegeven in de Verenigde Staten of Groot-Brittannië.

4.2.3 Analyse

Vervolgens is een codeersysteem ontwikkeld om de gevonden artikelen te coderen. Codeercategorieën betroffen: bibliografische gegevens, kunstdiscipline, onderwijsniveau, toetstaken, type beoordelingsinstrument, object van meting, beoordelingscriteria, benadering en informatie over kwaliteit van het instrument. In sommige gevallen ging een artikel over meerdere instrumenten. In dat geval werd de publicatie vaker ingevoerd. Uiteindelijk werden 153 beoordelingsinstrumenten ingevoerd.

Bij het coderen is uitgegaan van de woorden van de auteur(s). Beoordelingscriteria werden in letterlijke bewoordingen overgenomen uit de artikelen. Tevens werden de onderzoeksgegevens uit de geselecteerde publicaties aangehouden. Als een instrument in het betreffende artikel ingezet en getest werd in het primair onderwijs, werd dat onderwijsniveau ingevoerd. Mogelijkerwijs heeft een instrument een bredere toepasbaarheid, maar werd dat niet als zodanig gecodeerd. Ook voor de kwaliteit van het instrument werden de woorden van de auteur aangehouden en de betreffende passage werd als quote ingevoerd.

Het coderen van de artikelen werd uitgevoerd door beide onderzoekers (de eerste twee auteurs). De eerste vijf artikelen zijn gezamenlijk gecodeerd. Daarna werden de artikelen verdeeld op basis van alfabetische volgorde van auteur. Tijdens het coderen werden logboeken bijgehouden met twijfelpunten. De logboeken werden regelmatig doorgesproken om een gelijke aanpak door beide onderzoekers te garanderen.

Tenslotte werd een secundaire analyse uitgevoerd op de beoordelingscriteria. De beoordelingscriteria waren letterlijk overgenomen uit de artikelen. Experts uit de verschillende disciplines hebben de beoordelingscriteria vervolgens gecodeerd. Codeercategorieën ontstonden daarbij bottom up vanuit de data.

4.2.4 Overzicht van de instrumenten per kunstdiscipline en schoolniveau

We hebben beoordelingsinstrumenten geselecteerd die geschikt, ontworpen, gebruikt of getoetst zijn in het primair en secundair onderwijs. Enkele instrumenten zijn gebruikt bij jongere kinderen (vanaf 3 jaar). Omdat de scheidslijn tussen voorschools - en primair onderwijs internationaal niet eenduidig is, hebben we ervoor gekozen die instrumenten ook onder het primair onderwijs te scharen. We hebben instrumenten geselecteerd in de kunstdisciplines beeldend, muziek, dans en drama. De verdeling van de instrumenten over de schoolniveaus en kunstdisciplines is als volgt. Twaalf procent van de instrumenten is beschreven voor zowel primair als secundair onderwijs, 43% is alleen beschreven voor secundair onderwijs en 40% is alleen beschreven voor primair onderwijs. Bij 5% (N=7) van de instrumenten is het onderwijsniveau niet genoemd in het artikel. De verdeling van de disciplines over de schoolniveaus is min of meer gelijk.

Opvallend is dat een meerderheid (61%) van de instrumenten beschreven is voor de discipline muziek (N=94), 22% (N=33) voor de discipline beeldend, 11% (N=16) voor de discipline dans, 5 instrumenten (3%) voor de discipline drama en 5 instrumenten (3%) zijn beschreven voor alle disciplines samen. Een aantal van deze instrumenten is specifiek gericht op één type muziekinstrument (bv. snaar- of blaasinstrumenten), genre (bv. ballet) of medium (bv. tekenen). In het secundair onderwijs zijn vaker specifieke instrumenten beschreven dan in het primair onderwijs.

Tabel 1. Verdeling van de instrumenten over kunstdisciplines en schoolniveaus (aantallen en percentages)

	Beeldend	muziek	dans	drama	instrumenten voor meerdere disciplines	totaal
primair	11 33%	38 40%	4 25%	3 60%	5 100%	61 40%
secundair	9 27%	44 47%	11 69%	2 40%	-	66 43%
primair én secundair	10 30%	8 9%	1 6%	-	-	19 12%
schoolniveau niet benoemd	3 9%	4 4%	-	-	-	7 5%
totaal	33 100%	94 100%	16 100%	5 100%	5 100%	153 100%

4.3 Resultaten

Vraag 1: *Welke beoordelingsinstrumenten voor de kunstvakken in het basis- en voortgezet onderwijs zijn in de literatuur sinds 2000 beschreven?*

4.3.1 Soorten instrumenten

We maken in de beschrijving van de beoordelingsinstrumenten een onderscheid tussen de taak die een leerling moet doen (bv. een beeldend werkstuk maken) en het scoringsinstrument om die taak te beoordelen. Deze twee onderdelen samen vormen het beoordelingsinstrument. Het scoringsinstrument kan door de docent, maar ook door de leerling zelf of door een leeftijdsgenoot gebruikt worden om de taak te beoordelen. In sommige gevallen is de taak heel specifiek beschreven (kies drie ideeën, beelden of gevoelens op basis van een gegeven beeldend werk om daarmee een dans van max. 1 minuut samen te stellen) (Englebright & Mahoney, 2012). In dat geval is het scoringsinstrument vaak ook taakspecifiek. In andere gevallen is de taak heel open (bv. 'dansen in de klas tijdens de les') en is het scoringsmodel algemener toepasbaar (bv. een observatielijst op basis van criteria). In sommige gevallen is alleen de taak beschreven en niet het scoringsmodel (N=8). Ook maken we een onderscheid tussen instrumenten bedoeld voor formatieve of summatieve beoordeling.

4.3.2 Taken

De meeste taken in onze review betreffen 'uitingen in het artistieke medium'. Het gaat om het toepassen of gebruiken van kennis en vaardigheden waarbij iets gemaakt wordt, zoals een dans, muziek, theater of beeldend werk. Oreck (2005) en Oreck, Owen en Baum (2003) ontwikkelden een performance based assessment voor muziek, dans en drama om potentieel talent te identificeren in heterogene groepen leerlingen in het primair onderwijs. Leerlingen krijgen vijf lessen in de betreffende discipline. Bij dans komen verschillende stijlen

aan bod. Drama gaat vooral over improvisatieopdrachten en bij muziek komen vooral percussie en zang aan de orde. De leerlingen worden tijdens deze lessen geobserveerd door onafhankelijke observanten. Kaufman, Evans en Baer (2010) vergeleken de zelfbeoordeling door leerlingen met de beoordeling door onafhankelijke juryleden. De auteurs lieten leerlingen een vragenlijst invullen waarin ze hun eigen creativiteit moesten inschatten. In de daarop volgende dagen moesten de leerlingen verschillende creatieve opdrachten uitvoeren. Voor het vak beeldend, moesten leerlingen bijvoorbeeld een 'interessant en gek ontwerp' aanleveren. Ze kregen daarvoor een vel wit papier, lijm, potloden, en gekleurd papier. Juryleden beoordeelden deze beeldende uitingen.

Een andere categorie taken is het beantwoorden van schriftelijke of mondelinge vragen. In Miller (2009) krijgen leerlingen bijvoorbeeld contrast- en vergelijkingsopgaven. Leerlingen krijgen hetzelfde muziekstuk in verschillende versies te horen en worden gevraagd om de stukken te vergelijken op punten als melodie, harmonie, ritme en vorm. Schmid (2003) gebruikt de zogenoemde Visual Thinking Strategies (VTS) en laat leerlingen vragen beantwoorden als: 'Wat gebeurt er op dit plaatje?', 'Wat zie je nog meer?' en 'Wat zie je dat je dat zegt?'. Deze vragen zijn complementair aan de artistieke uitingen die de leerlingen ook maken.

Er zijn ook cases waarin de leerlingen geen specifieke taak krijgen, maar waarin ze als het ware worden 'geobserveerd' in hun bestaande gedrag of houding. Dunbar (2012) heeft een methode ontwikkeld om leerlingen in 30 seconden of minder te beoordelen. Deze methode past ze elke muziekles toe, waardoor de gezamenlijke observaties een compleet beeld van de leerling geven. Per les bepaalt ze welke competenties, kennis of vaardigheden ze wil beoordelen. Vervolgens scoort ze op een namenlijst van de klas of een leerling de betreffende vaardigheid goed, voldoende of onvoldoende laat zien. Ze observeert de leerlingen dus tijdens de les, zonder dat ze specifieke taken hebben gekregen.

Vooral in de beeldende vakken is het samenstellen van een portfolio een veel gebruikte taak. Een portfolio bevat vaak verschillende soorten ingrediënten. Zo kan het werkstuk (beeldend) opgenomen worden, maar daarnaast ook bewijsmateriaal ten aanzien van het proces of geschreven reflecties. In dat geval bevat het portfolio zowel performance taken als andere taken. In het geval van Lindström (2006) bestaat het portfolio uit schetsen, ontwerpen en uiteindelijke producten, uit logboeken, inspiratiebronnen en videoverslagen van opgenomen gesprekken met leerlingen.

Een logboek, leerrapport of journal kan aan een portfolio toegevoegd worden of op zichzelf staan. Abril (2006) onderzoekt de leeruitkomsten van leerlingen in multiculturele muziekeducatie aan de hand van een leerrapport. Leerlingen worden gevraagd te reageren op twee aanwijzingen: 'Ik heb geleerd dat ...' en 'Ik heb geleerd hoe ...'. Ook Gower (2012) gebruikt een leerrapport door aan de leerlingen te vragen waarom ze doen wat ze doen.

Ten slotte zijn er nog taken of opdrachten waarin verschillende componenten gecombineerd worden.

In tabel 2 staat beschreven hoe de verschillende taken verdeeld zijn over de disciplines. Zoals gezegd geldt voor alle disciplines dat de 'uiting in het artistieke medium' de meest voorkomende taak is. Voor drama geldt zelfs dat dit het enige type taak is in de artikelen die zijn opgenomen in deze review. In de discipline beeldend is het samenstellen van een portfolio een belangrijke taak, deze komt minder voor bij de andere disciplines. Bij muziek wordt relatief vaker bestaand gedrag of een houding bekeken. Ook het beantwoorden van mondelinge of schriftelijke vragen komen relatief vaker voor bij muziek.

Tabel 2. Taken en disciplines (aantallen)

	muziek	beeldend	dans	Drama	instrumenten voor meerdere disciplines	totaal
uiting in het artistieke medium	65	18	10	5	0	98
mondelinge/schriftelijke vragen	13	3	1	0	0	17
houding of gedrag; geen specifieke taak	9	1	2	0	3	15
portfolio	3	6	1	0	1	11
combinatietaak	1	5	2	0	1	9
logboek	3	0	0	0	0	3
	94	33	16	5	5	153

4.3.3 Scoringsinstrumenten

In tabel 3 staan de scoringsinstrumenten genoemd en hun verdeling over de disciplines. We zien dat behalve voor dans geldt dat het gebruik van een criterialijst het meest voorkomende scoringsinstrument is. Op de tweede plaats komen de rubrics, die in de dansvakken vaker gebruikt worden dan criteria. Bij de podiumkunsten wordt gebruik gemaakt van zelfrapportages, maar dit geldt niet voor de beeldende vakken. Behalve bij drama zijn we scoringsinstrumenten tegengekomen in de vorm van commentaar of opmerkingen. Zowel in de disciplines muziek en beeldend zijn we jurybeoordelingen tegengekomen, maar niet bij dans of drama. Het gebruik van software als scoringsinstrument zijn we alleen tegengekomen bij muziek. Het gebruik van ankers zijn we één keer tegengekomen, bij beeldend. De scoringsinstrumenten werden ingevuld door de leerkracht (22%), door de leerling zelf (9%), door peers (1%), door een onafhankelijke beoordelaar (bv de onderzoeker, 34%) of door meerdere van bovenstaande (34%).

Tabel 3. Scoringsinstrumenten en disciplines (aantallen)

	Muziek	beeldend	dans	Drama	instrumenten voor meerdere disciplines	totaal
criteria	55	11	4	3	4	77
rubric	17	10	8	1	0	36
zelfrapportage	8	0	2	1	0	11
commentaar	4	3	2	0	0	9
jury	5	3	0	0	0	8
onbekend	2	5	0	0	1	8
software	3	0	0	0	0	3
ankers	0	1	0	0	0	1
	94	33	16	5	5	153

Oreck (2005; Oreck et al., 2003) gebruikt bij de ontwikkelde instrumenten voor muziek, dans

en theater een criterialijst voor vaardigheden, motivatie en creativiteit. Voorbeelden van criteria in de categorie vaardigheden voor dans zijn: 'fysieke controle', 'coördinatie en behendigheid' en 'ruimtelijk bewustzijn'. Bij muziek wordt motivatie bijvoorbeeld gemeten aan de hand van categorieën als: 'enthousiasme' en 'doorzettingsvermogen'. En een criterium voor creativiteit is bij drama bijvoorbeeld: 'verbeeldingskracht'. Later in dit artikel gaan we dieper in op de verschillende criteria die beschreven worden in de artikelen. Als ook competentieniveaus van verschillende criteria beschreven worden, dan is het scoringsinstrument een rubric. In het geval van Lindström (2006) zijn expliciet zowel processen als producten beoordeeld.

In 11 instrumenten is sprake van zelfrapportage als scoringsinstrument. Een voorbeeld van die zelfbeoordeling door leerlingen is SEMA – Self Esteem of Music Assessment, gebruikt in Randles (2010). Het betreft een vragenlijst van 43 items. Een voorbeeld van een item is: 'Ik voel me goed als mijn ouders mijn vooruitgang in de muziek opmerken' (p. 11). De meeste zelfrapportages maken gebruik van vragenlijsten.

Een andere vorm scoringsinstrument is het geven van commentaar of opmerkingen door docenten bij de taken die de leerlingen hebben uitgevoerd. Dit is in alle gevallen een kwalitatieve vorm van feedback geven. Watts (2010) beschrijft bijvoorbeeld dat veel groepsleerkrachten in het primair onderwijs de tekeningen van leerlingen niet beoordelen, omdat ze niet goed weten hoe. Watts heeft een lijst met vragen opgesteld die een docent kan gebruiken om met een leerling over een tekening in gesprek te gaan en zodoende de leerling te stimuleren om zich verder te ontwikkelen. Een voorbeeldvraag is: "*How are the people in your picture feeling? How might you show in a drawing the way one person feels about another?*" (p. 149).

Acht instrumenten maakten expliciet gebruik van een jury als scoringsinstrument. Chan, Chan, en Chau (2009) lieten de tekeningen van leerlingen bijvoorbeeld beoordelen door onafhankelijke juryleden als onder gemiddeld, gemiddeld of boven gemiddeld. Bergee (2007) liet muziekkuitvoeringen van leerlingen door juryleden beoordelen en deed dit op een schaal van 0 (slecht) tot 100 (excellent). Opgemerkt moet worden dat ook juryleden altijd minstens één beoordelingscriterium hanteren om de leerlingen te beoordelen. Bij jurybeoordelingen wordt in de helft van de instrumenten (N=4) gebruik gemaakt van een holistisch oordeel (op basis van 1 criterium), bij 3 instrumenten is sprake van een analytisch oordeel (meerdere criteria), bij 1 instrument is dit niet beschreven. In plaats van onafhankelijke juryleden wordt er ook gebruik gemaakt van software als scoringsinstrument. Dat komt echter relatief weinig voor. In deze review hebben we 3 instrumenten gevonden die gebruik maken van software, allemaal in de discipline muziek. De software meet dan bijvoorbeeld of de leerlingen de juiste toonhoogte spelen (Trollinger, 2003; Leighton & Lamont, 2006) of hun stembereik (Schneider, Zumtobel, Prettenhofer, Aichstill, & Jocher, 2010).

Een laatste categorie kwam slechts eenmaal voor in deze review: het gebruik van ankers. Myford (2002) ontwikkelde de 'descriptive graphic' (een soort rubric, maar dan een continue schaal). De ankers (voorbeelden van leerlingwerkstukken geplaatst in het continuüm van de beoordelingschaal) dienden als referentiepunten in de training van de beoordelaars.

4.3.4 *Formatief/summatief toetsen*

In principe bepaalt de docent of de beoordelaar of een instrument formatief of summatief ingezet wordt in het onderwijs; de docent kiest namelijk of een instrument wordt ingezet om vast te stellen of een niveau gehaald is en op basis daarvan een cijfer te geven (summatief gebruik) of om een instrument in te zetten voor feedback waarbij vooral aandacht is voor de volgende stap (formatief gebruik). Uiteraard zijn veel instrumenten zowel formatief als

summatief te gebruiken. Niet alle instrumenten lenen zich echter voor summatief gebruik, zoals bijvoorbeeld het instrument van Watts (2010).

In de beschreven artikelen vinden we voor 50% summatieve beoordelingsinstrumenten en voor 11% formatieve beoordelingsinstrumenten. Zeventien procent van de instrumenten zijn in de beschreven artikelen voor beide type beoordelingen gebruikt. Voor 34 beoordelingsinstrumenten was niet duidelijk welke vorm van beoordeling het betrof.

Vraag 2: Wat wordt er met deze instrumenten beoordeeld?

In de beschreven artikelen wordt er in 71% van de gevallen naar een product gekeken (zoals een tekening of een muziekkuitvoering) , in vijf procent van de gevallen naar een proces (bijvoorbeeld het maakproces van een tekening) en in 13% van de gevallen naar beiden. Voor 17 beoordelingsinstrumenten kon op basis van het artikel niet vastgesteld worden of het om een product- of procesbeoordeling ging. De verhoudingen verschillen enigszins over de disciplines. Bij muziek, dans en drama gaat het hoofdzakelijk om productbeoordelingen, bij beeldend is er in een derde van de instrumenten sprake van zowel een proces- als een productbeoordeling.

4.3.5 Soorten competenties

Zoals in de inleiding beschreven, maken we in navolging van de NAEP onderscheid in de competenties: perform, create, respond en reflect. In tabel 4 staat een overzicht van het voorkomen van de competenties in de beschreven artikelen over de verschillende disciplines.

Tabel 4. Soorten competenties (aantallen)

	Muziek	beeldend	dans	drama	instrumenten voor meerdere disciplines	totaal
Perform	55	0	6	2	0	63
Create	3	17	0	0	3	23
Respond	15	2	1	0	0	18
Reflect	6	1	1	0	0	8
Alles	2	0	1	0	0	3
perform+create	8	1	3	2	0	14
overige combinaties	5	12	4	1	2	24
	94	33	16	5	5	153

De meeste instrumenten toetsen competenties in de domeinen perform en/of create. In de podiumkunsten (muziek, dans en drama) ligt de nadruk op perform, terwijl de beeldende vakken zich uitsluitend richten op create. Respond en reflect worden in mindere mate getoetst in de beschreven instrumenten.

4.3.6 Beoordelingscriteria

Vervolgens bekijken we de criteria die gehanteerd worden bij de beoordeling. Niet bij elk instrument is overigens sprake van criteria. Sommige artikelen beschrijven holistische beoordelingsprocedures, waarbij de beoordelaar een algemeen oordeel over een geheel geeft, gebaseerd op ervaring, gevoel of intuïtie. Ook bij formatieve procedures waarbij vooral

kwalitatieve feedback wordt gegeven is niet altijd sprake van criteria. De criteria die wel genoemd zijn, zijn door ons per discipline opgesplitst en voorgelegd aan vier vakexperts en vakdidactici met de vraag om de criteria te groeperen naar deelcompetenties of dimensies. In tabel 5 is de indeling in dimensies zoals gemaakt door de experts weergegeven in kolom 1. In kolom 2 zijn de criteria die daar onder vallen, eventueel in iets ingekorte vorm, weergegeven.

Bij muziek meten de beoordelingsinstrumenten vooral aspecten van het uitvoeren: de interpretatie en mate van expressiviteit, de technische en ritmische beheersing. Ook veel criteria op het gebied van houding (zoals inzet, doorzettingsvermogen, nauwkeurigheid) hebben betrekking op het uitvoeren van muziek. Andere dimensies betreffen het gehoor/klankvoorstelling en muzikaal vocabulaire. Hoewel muziek maken meestal een groepsgebeuren is, zijn er weinig instrumenten die aspecten van samenspel beoordelen. Ook weinig beoordelingsinstrumenten hebben betrekking op muzieknnotatie en op het scheppen van muziek: componeren. Bij de beeldende vakken staat het beoordelen van creëren voorop, niet alleen de beoordeling van het beeldende eindproduct, maar ook van het beeldende proces. Bij beoordeling van het product gaat het om technische aspecten, vormaspecten en inhoudelijke aspecten (het concept). Bij procesbeoordeling gaat het om onderzoek en experiment, zelfreflectie, nieuwsgierigheid en inzet. De belangrijke categorie creativiteit heeft betrekking op zowel proces- als productbeoordeling. Weinig instrumenten hebben betrekking op kunstbeschuwing, hoewel dat zeker in het voortgezet onderwijs een belangrijk onderdeel van het curriculum is.

Beoordelingsinstrumenten bij dans hebben net als bij muziek vooral betrekking op het uitvoeren: op technische aspecten, muzikaal ritmische aspecten, expressiviteit, ruimtegebruik, improvisatie en fysieke aspecten. Bij alle kunstvakken worden houdingsaspecten beoordeeld, maar ze komen in de instrumenten voor dans verhoudingsgewijs veel voor. Het gaat dan om doorzettingsvermogen, toewijding, correcties kunnen ontvangen en dergelijke. De weinige instrumenten op het gebied van theater hebben vooral betrekking op spel, zoals bijvoorbeeld tekstbeheersing, expressie, geloofwaardigheid en samenwerking. Er zijn geen instrumenten gevonden die zich richten op het schrijven van theaterteksten en op regie.

De te meten dimensies en de gebruikte criteria zijn in meerderheid disciplinespecifiek, maar er zijn op een hoger abstractieniveau wel overeenkomsten aan te wijzen. In alle kunstdisciplines richten instrumenten zich op technische beheersing, originaliteit/creativiteit, zelfreflectie, samenwerking en inzet/houding. De instrumenten beperken zich niet tot makkelijk te meten kennis of technische vaardigheden, maar hebben ook betrekking op kwalitatieve aspecten die typerend zijn voor de kunstvakken, zoals bijvoorbeeld expressie, bewegingskwaliteit, toonkwaliteit, sfeer, originaliteit, interpretatie en visie.

Tabel 5a. Beoordelingscriteria muziek

Dimensie	Criteria	% van totaal aantal criteria (N=269)
Expressie/ interpretatie	articulatie; articulatie en dynamiek; articulaties; consequent fraseren; dictie; dynamiek; een idee of gevoel uitdrukken; esthetische aantrekking; expressie; expressiviteit; expressiviteit (mate waarin de compositie muzikaal expressief is en esthetische gevoeligheid laat zien); frase; geen geheel/ vloeiend; geschreven interpretatie; hoe muzikaal een stuk wordt opgevoerd; individuele toon; interpretatie; interpretatie fermaten; interpretatie muzikaal effect; melodie; melodie/ ritmische ontwikkeling; melodisch motief; melodische inconsequentie; muzikale interpretatie; muzikale vloeiendheid; niet aantrekkelijk/ aantrekkelijk; niet memorabel/ memorabel; niet met sfeer/met sfeer; onbekend/ bekend; ongestructureerd/ gestructureerd; ontwikkeld melodisch motief; ontwikkeld ritmisch motief; ritme/ gevoel voor timing; ritmische inconsequentie; saai/ levendig; stijl; tempo; vibrato; vloeiende lijn	26%
Technische beheersing	adem controle; ademhaling; ademondersteuning; akkoorden correct wisselen; akkoorden stevig indrukken; articulatie; articulatie/toon; articulaties; beperkt bereik/ breed bereik; boog; coördinatie; dictie; dynamiek; etude specifieke criteria (bij. Een crescendo in 8 maten vertrekt van een werkelijk piano; gebruik van het totale bereik van de kopstem(hoogte verkenning); handen correct plaatsen; houding; instrument vloeiendheid; linkerhand; luchtstroom; melodische nauwkeurigheid; moeilijkheid van uitgevoerde repertoire; nauwkeurigheid; noot nauwkeurigheid; noot nauwkeurigheid en intonatie; noten; noten/ritme; positie; ritme- en nootnauwkeurigheid; slag over alle snaren; stabiel tempo volhouden; stembereik; techniek; techniek/articulatie; techniek/dictie; techniek- articulatie; technisch gemak; technisch onbekwaam/ technisch bekwaam; vibrato; vingerzetting; zangkwaliteit	19%
Gehoor/ klankvoorstelling	aandachtig luisteren; accuraat identificeren van overeenkomsten en verschillen (melodie, harmonie, timbre, ritme, vorm en expressiemiddelen); akkoord herkenning; balans; correctheid van toonhoogten; geluidperceptie; harmonie; II-V-I herkenning; intonatie; klankmenging; melodie; nauwkeurigheid en precisie toonhoogte; nauwkeurigheid in toon patroon reproductie; nauwkeurigheid intonatie; niet klankrijk/ klankrijk; toon; toon overname; toonhoogte; toonhoogte en ritme; toon-intonatie; toonkwaliteit; zuiverheid	19%
Ritmische beheersing	maat/ritme; metronomisch tempo; nauwkeurigheid ritme; niet ritmisch/ritmisch; ontregeling van de slag in waarneembaar veranderde structuren; puls; ritme; ritme/ gevoel voor timing; ritme/ tempo; ritmes; ritmische complexiteit; ritmische nauwkeurigheid; stabiel tempo volhouden; tempo; toonhoogte en ritme	16%
Muzikaal vocabulaire/ grammatica	accuraat identificeren van overeenkomsten en verschillen (melodie, harmonie, timbre, ritme, vorm en expressiemiddelen); akkoord herkenning; compositie en improvisatie; consequent fraseren; correct gebruik van muzikale terminologie; harmonie; II-V-I herkenning; melodie/ ritmische ontwikkeling; melodisch motief; muzikale syntax; nauwkeurigheid in muziek beschrijven; onbekend/ bekend; ongestructureerd/ gestructureerd; ontwikkeld melodisch motief; ontwikkeld ritmisch motief; ritmisch motief; simpel/ complex; stijl; technische kwaliteit (mate waarin tonische en ritmische elementen technische competentie laten zien m.b.t. tonaal center en ritmische regulariteit); toegenomen verticale opbouw; toonladders samenstellen; type kennis: feiten, conceptueel, procedureel, metacognitief; cognitief proces: onthouden, begrijpen, toepassen, analyseren, evalueren en creëren; uitwerking in muziek beschrijven; vakmanschap (vorm, technische kwaliteit, detail,	11%

Attitude	complexiteit, organisatie); verwachte cadensformules en doorbreken van het stereotype; vorm basisvaardigheden presenteren; betrokken zijn; doorzetten bij moeilijkheid; doorzettingsvermogen; enthousiasme; houding; interesse hebben; muziek vergeleken met andere activiteiten; nauwkeurigheid; nauwkeurigheid van de uitvoering in de categorieën van zuiverheid, ritmiek, bogen/ articulatie, tempo, uitvoering van de genoteerde expressie tekens, pauzes/ fermates, herhalingen; netheid; opdracht voltooiën; overtuiging; persoonlijke toewijding; presentatie; vaardigheid om te focussen; verantwoordelijkheid nemen; zekerheid	9%
Creativiteit	creativiteit; creativiteit (mate van originaliteit en coherentie van de compositie); creativiteit (originaliteit, nieuw gebruik van timbre, nieuwe muzikale ideeën en variatie); innovatief gebruik van ritmische elementen; niet creatief/ creatief; onorigineel/ origineel; originaliteit; verwachte cadensformules en doorbreken van het stereotype	6%
Samenspel	balans; harmonie; klankmenging; samenspel	4%
Muzieknotatie	muzieknotatie; netheid; partituur; schrijftechniek; strategieën voor het van het blad lezen (bestuderen van de eerste maat, identificeren toonsoort, identificatie maatsort, vaststellen van een passend tempo alvorens te beginnen met uitvoeren, partituur scannen op eventuele obstakels	2%
Zelfreflectie	kwaliteit van de inhoud (reflectief denken over, accurate beoordeling van de individuele kennis, vaardigheden en werk); mogelijkheden en zelfbeeld; organisatie strategieën (bijhouden wat geleerd wordt); verbeterstrategieën (oefenen om te verbeteren, zelfcorrectie); voltooiing inhoud	2%
Rest/ onbekend	anders; beslissingstijd; consistentie; contour nauwkeurigheid; goede antwoorden; muziektechnologie; niet aantrekkelijk/ aantrekkelijk; onaf/ af; schoolmuziek; strategieën voor het uit het hoofd spelen, (conceptueel, kinesthetisch, muzikaal) en strategieën voor het op gehoor spelen (conceptueel, kinesthetisch, muzikaal)	4%

Voetnoot: Criteria muziek tellen op tot meer dan 100% omdat enkele criteria in meer dan één dimensie werden geplaatst

Tabel 5b. Beoordelingscriteria beeldend

Dimensie	Criteria	% van totaal aantal criteria (N=130)
Originaliteit/ creativiteit (proces en product)	creativiteit; divergent denken; flexibel denken; flexibiliteit; identificatie van verre associaties; in welke mate zijn de ideeën origineel (inventief, ingenieus, niet-standaard) in concept en vorm; inventiviteit; originaliteit; risico nemen; speelsheid; uitwerking; vloeïendheid; vloeïendheid van ideeën; voorkeur voor nieuwheid	18%
Beeldaspecten (product)	compositie; contourversmelting; detail; esthetische organisatie; gebruik van compositie; hoe effectief gaat de leerling om met verre, nabije vormen en overlap?; hoe effectief integreert de gehele compositie de manier waarop het hier geplaatst is?; in welke mate is het hier in de omgeving geplaatst?; kleur, vorm en compositie; lijnvoering; ontwerpruimte gebruik; organisatie (totale compositie); platte of ruimtelijke weergave; realistische plaatsing van elementen; representatie; sfeer; veel details in het gezicht geplaatst; verhoudingen die kloppen met de werkelijkheid; vormdifferentiatie	17%
Onderzoek/ experiment (proces)	aantal keer gummen; betrokkenheid bij experiment; dagdromen als incubatie activiteit; exploratietijd; kritisch bronnen uit visuele cultuur analyseren en demonstreer begrip van doel, betekenis en context; onderzoekswerk; ontwikkel	14%

	ideeën door experiment, exploratie en evaluatie; problemen formuleren en conceptualiseren; proportie tijd probleem vinden; totale tekentijd; vaardigheid om vastleggen van artistiek probleem uit te stellen; zoekt stimulans en kennis; vaardigheid om voorbeelden te gebruiken;	
Techniek (product)	audio; belichting; camera; continuïteit; materiaalgebruik; mediagebruik; techniek; tekenvaardigheden; vakmanschap; video effectiviteit	9%
Concept/idee (product)	combineren of herordenen van concepten; complexiteit van visuele of conceptuele ideeën; duidelijke intentie; hoe geschikt zijn de vormen voor de uitgedrukte ideeën? drukken de vormen de bedoelde betekenis uit?; persoonlijke ideeën, intenties, ervaringen, informatie en meningen in visuele en andere vorm opslaan; toepassen van abstracte relaties om nieuw begrip te genereren; verbeelding; visie; zichtbaarheid van intentie	8%
Zelfreflectie (proces)	accurate en coherente reactie (op eigen kunst); capaciteit voor zelfbeoordeling; evalueren en verfijnen; focus in reageren op eigen kunst; kwaliteiten van het werk evalueren; persoonlijke heuristiek; presenteren van een coherente en georganiseerde selectie van werken en eindproducten die een persoonlijke en doordachte keuze laat zien en overeenkomt met intenties; reflectie en evaluatie; reflectiviteit	7%
Kunstbeschouwing	beschrijving van kunst; heldere, vloeiende uiteenzetting; interpretatie van kunst in termen van kleur, toon, stijl, techniek, onderwerp, identificeren van focus/ gevoel/expressie van de kunstenaar; kunstperceptie	3%
Nieuwsgierigheid/ openheid (proces)	nieuwsgierigheid; openheid voor ervaring; tolerantie voor onbekende sensorische stimulatie; vragen stellend	4%
Inzet/houding (proces)	Doorzettingsvermogen; toewijding; vertrouwen, onafhankelijkheid en plezier; zelfmotivatie	3%
Expressie	expressie; expressiviteit	2%
Samenwerking (proces)	samenwerking en communicatie; teamwork	2%
Overig	bewustzijn van gaten of missende delen; elegante oplossing; excellente artistieke vaardigheden; feedback gebruiken; impact van het product; inspiratie; kennis en begrip; meervoudige intelligentie vaardigheden; productiviteit; schoon en netjes; sensitiviteit; strategieën en vaardigheden; synthese; vaardigheid om dingen te transformeren; vaardigheid om te abstraheren en van geheel naar delen te gaan; voorkeur voor cognitieve complexiteit	14%

Tabel 5c. Beoordelingscriteria dans

Dimensies	Criteria	% van totaal aantal criteria (N=93)
Dans techniek/ bewegingsuitvoering/actie	(bewegings)actie; ademgebruik; articulatie van de voet; bewegingsnauwkeurigheid; coördinatie; coördinatie en lenigheid; coördinatie en verbinding; fysieke controle; gebruik van gewicht; gebruik van plié; gebruik van port de bras; gebruik van rotatie in benen en bekken; nauwkeurigheid; nauwkeurigheid/afstemming; pirouettes; placering van de ruggengraat; sprongen en lifts; stijl consistentie; technische bekwaamheid; voet- en armposities	26%
Muzikaliteit/ritme/ bewegingsdynamiek	correcte dynamiek; dynamisch bereik; frasering en gebruik van overgangen; muzikaliteit; ritme; ritmische scherpeheid; vermogen tot fraseren	15%

Kracht/ intentie/expressie	artistieke intentie uitstralen; bewegingskwaliteit; bewegingskwaliteiten; expressiviteit; gebruik van focus en energie; helderheid; kracht; richten van de focus	14%
Persoonlijke attitude	doorzettingsvermogen; toewijding t.a.v. training en leerproces; vaardigheid om correcties te ontvangen, toe te passen en te onthouden; vaardigheid om doelen te stellen, na te volgen en te bereiken; vaardigheid om te focussen; vertrouwen, onafhankelijkheid en plezier; wil en vaardigheid om risico te nemen; zelfregulatie tijdens creatief proces; zelfbewustzijn/ metacognitie tijdens creatief proces	12%
Ruimtegebruik	correcte ruimtelijke oriëntatie; driedimensionaliteit van het lichaam; gebruik van ruimte; ruimtegebruik; ruimtelijk bewustzijn; ruimtelijke nauwkeurigheid	10%
Samenwerking	betrokkenheid in creatief proces; positieve en open houding; gezamenlijkheid; samenwerking; samenwerking en communicatie, samenwerking/ruimtelijk bewustzijn; toewijding aan gemeenschap van de klas	7%
Vermogen tot improviseren, creativiteit	creativiteit; identificatie en implementatie van creatieve mogelijkheden; improvisatie; innovatie; strategieën en vaardigheden	7%
Bewegingsgeheugen	observatie en herinnering van ritme; observatie en geheugen; vermogen om combinaties te onthouden en omkeren	3%
Uithoudingsvermogen/ fysieke fitheid/ flexibiliteit	bereik in beweging en extensie; fitheid; kwalitatief bewegingsbereik	3%
Overig	kennis en begrip; reflectie en evaluatie; zelfevaluatie tijdens creatief proces	3%

Tabel 5d. Beoordelingscriteria theater

Criteria		% van totaal aantal criteria (N=22)
Tekstregels onthouden	tekstregels onthouden; script kennen	14%
Bewustzijn	fysiek bewustzijn; zintuiglijk bewustzijn	9%
Decorstukken gebruiken		9%
Expressie	lichaamsexpressie & beweging; verbale & nonverbale expressie	9%
Geloofwaardige karakters neerzetten		9%
Samenwerking	samenwerking; samenwerking & communicatie	9%
Verbeelding		9%
Focus/ toewijding		5%
Vertrouwen, onafhankelijkheid & plezier		5%
Creativiteit		5%
Strategieën & vaardigheden		5%
Kennis & begrip		5%
Reflectie & evaluatie		5%

4.3.7 Kwaliteit

Vraag 3: Wat wordt gerapporteerd over de kwaliteit van deze instrumenten?

In tabel 6 is een onderscheid gemaakt tussen verschillende soorten betrouwbaarheid en validiteit waarover in de artikelen gerapporteerd werd.

Tabel 6. Kwaliteitsmetingen van de instrumenten (aantallen)

	muziek	beeldend	dans	drama	instrumenten voor meerdere disciplines
1. Betrouwbaarheid	59	17	5	5	3
Interbeoordelaar	42	17	2	5	3
Interne consistentie	23	2	2	0	2
Stabiliteit	2	0	1	1	2
Anders over betrouwbaarheid	2	0	1	0	0
2. Empirisch geteste validiteit	23	6	3	1	3
Inhoudsvaliditeit (factoranalyse)	7	3	2	1	0
Concurrent	7	2	2	1	1
Discriminant	1	0	1	1	1
Voorspellend	4	0	1	1	1
Externe (geen onbedoelde verschillen tussen groepen)	5	1	1	1	3
Andere test validiteit	5	1	0	0	0
3. Overig validiteit	68	27	13	3	5
Pilot test	41	11	7	2	5
Theorie-gebaseerd	41	19	4	1	0
Experts geraadpleegd in constructiefase/ achteraf face validiteit	7	2	2	1	4
Geplaatst in discussie over authenticiteit van beoordeling	9	8	5	3	1
Validiteit gewoonweg genoemd (zonder uitleg)	16	4	4	2	2
4. Haalbaarheid (letterlijk genoemd)	8	2	2	3	3
5. Eerlijkheid (letterlijk genoemd)	5	1	3	2	0
6. Transparantie (letterlijk genoemd of constructie instrument met leerlingen)	4	0	5	1	0
	94	33	16	5	5

In ongeveer de helft van de gevallen was sprake van een empirische test van betrouwbaarheid; dat wil zeggen interbeoordelaarsbetrouwbaarheid, interne consistentie van de schaal of stabiliteit van de meting over tijd/plaats. Interbeoordelaarsbetrouwbaarheid werd

in de meeste gevallen gerapporteerd.

In ongeveer een kwart van de artikelen was sprake van een empirische test voor het vaststellen van validiteit. Het ging dan om factoranalyses om constructen vast te stellen, concurrent validiteit waarbij gekeken wordt of de beoordeling overeenkomt met een andere meting van hetzelfde of gelijkend concept of discriminant validity waarbij juist gekeken wordt of het construct wel echt afwijkt van een andere meting (bv. IQ). Tenslotte werd onderzocht of verschillende groepen (type onderwijs/ etniciteit) niet significant verschillend scoren op het instrument, dus of een instrument niet onbedoeld groepen uitsluit. Deze metingen kwamen relatief weinig voor.

Onder 'overig validiteit' werd verstaan het rapporteren over pilots in de fase van instrumentconstructie of afname, beschrijving van de theoretische basis van het instrument, betrokkenheid van experts in de constructiefase of ten behoeve van het vaststellen van 'face validiteit' en het gewoonweg vermelden van authenticiteit of validiteit (zonder verdere toelichting).

Er werd relatief weinig gerapporteerd over de praktische haalbaarheid, eerlijkheid en transparantie van instrumenten. Er was dan ook nauwelijks sprake van systematisch kwalitatief onderzoek in de klassen waarbij gebruik van instrumenten geobserveerd werd. Althans, rapportage over dergelijk onderzoek werd niet gevonden.

Met zoveel verschillende publicaties en zoveel verschillende typen kwaliteitsmetingen is het lastig om een vergelijking te maken tussen de publicaties op basis van de gerapporteerde kwaliteit. Enerzijds zijn er instrumenten die op veel verschillende typen kwaliteit zijn getoetst. Anderzijds zijn er instrumenten die weliswaar op slechts één kwaliteitscriterium zijn bekeken, maar daarop wel goed scoorden. Opvallend, doch niet per se verrassend, is dat alle beschreven instrumenten minimaal voldoende en vaker zelfs ruim voldoende of goed scoorden op de gekozen kwaliteitscriteria. Worden onvoldoende resultaten niet beschreven of zijn ze niet voorgekomen in het onderzoek? Waarschijnlijk is dat hier sprake is van selectieve rapportages. Extra informatief zijn daarom die publicaties waarin twee (of meer) instrumenten met elkaar vergeleken zijn (zie bijvoorbeeld Dorn en Sabol (2006) of Myford (2002)).

Aangezien in vrijwel alle publicaties in deze review positief gerapporteerd is over de beschreven kwaliteitscriteria (met uitzondering van Thorpe (2012)), kunnen we aannemen dat publicaties waarin meer verschillende kwaliteitscriteria onderzocht zijn, hoger scoren op algehele kwaliteit. Ook kunnen we een onderscheid maken tussen de empirisch getoetste beoordelingsinstrumenten en instrumenten die anderszins beoordeeld zijn. We onderscheidden 18 verschillende kwaliteitsaspecten (tabel 9). Er zijn 59 instrumenten (39%) getest op één van deze aspecten, de meerderheid (52%, N=79) is getest op twee tot zes criteria, de rest (10%, N=15) op zes of meer criteria.

In de publicatie van Oreck et al. (2003) worden elf verschillende kwaliteitscriteria getoetst, waarvan zeven empirisch. Daarmee is dit het meest uitgebreid getoetste instrument uit deze review. In dit artikel wordt een instrument voor talent identificatie (D/M/T TAP; dans, theater en muziek) systematisch geëvalueerd op validiteit en betrouwbaarheid. Ten eerste werd een factoranalyse uitgevoerd om construct validiteit vast te stellen. Vervolgens werd gekeken of de score verschilde van andere metingen zoals een wiskunde test (discriminant validity). Ook werd gekeken of gender of etniciteit de score beïnvloedde. Er werd onderzocht of er een correlatie bestond met docentoordelen (concurrent validity). Tenslotte werd gekeken of als talentvol geselecteerde leerlingen anders scoorden op het instrument dan niet als talentvol geselecteerde leerlingen (contrasting groups evidence) en of door het instrument geselecteerde leerlingen later succesvol waren in vervolgonderwijs. Met betrekking tot betrouwbaarheid werd onderzocht of de beoordelaars het met elkaar eens waren (interrater reliability) en of de meting stabiel was (op verschillende momenten

vergelijkbare scores gemeten).

Een andere manier om naar kwaliteit te kijken is om verschillende instrumenten (of verschillende versies van een instrument) met elkaar te vergelijken. Myford (2002) onderzocht op een systematische en empirische wijze de betrouwbaarheid van de 'descriptive graphic' (een soort rubric, maar dan een continue schaal). Ze maakte daarbij gebruik van leerlingwerk en beoordelingsschalen uit de NAEP assessment, beeldend. Myford onderzocht verschillende varianten van die descriptive graphic: met/zonder aangegeven 'middens' (midpoints), afstandstreepjes (hatchmarks), beschrijvingen bij begin-eind en/of midden van de schaal en met/zonder gebruik van aanvullende ankers voor training van beoordelaars. Onderzocht werd of de vormgeving van beoordelingsschalen de betrouwbaarheid beïnvloedt en wat een optimaal aantal schaalpunten is. Beoordelaars waren gebaat bij het bespreken van ankerproducten voor de begin- en eindpunten van de schaal. Ankers voor de hele schaalrange leverde geen winst op voor betrouwbaarheid.

4.4 Samenvatting en conclusie

Dit artikel geeft een overzicht van 153 beoordelingsinstrumenten voor de kunstvakken (muziek, beeldend, drama en dans) in het primair en secundair onderwijs, zoals beschreven in de internationale literatuur tussen 2003 en oktober 2013. De huidige reviews richtten zich dikwijls op één kunstdiscipline of beperkten zich tot instrumenten voor ofwel het primair ofwel het secundair onderwijs, maar bovenal ontbrak er vaak een bespreking van de kwaliteit van de instrumenten en hoe die kwaliteit gemeten is. Deze review heeft geprobeerd in te gaan op deze lacune.

Ook in de kunstvakken worden leerlingen steeds vaker beoordeeld. De taken en scoringsinstrumenten zijn echter divers en er is veel discussie over het nut en de noodzaak van het beoordelen. De verschillende benaderingen in de verschillende kunstvakken vragen om verschillende beoordelingsinstrumenten. In deze review selecteerden we 119 peer reviewed artikelen van de afgelopen 10 jaar uit 44 internationale tijdschriften die de kwaliteit van in totaal 153 beoordelingsinstrumenten beschreven. Hiermee pretenderen we niet een volledig en up to date overzicht te hebben van de beoordelingsinstrumenten zoals ze in de praktijk toegepast worden, maar laten we wel zien wat er in de wetenschappelijke literatuur verschenen is over beoordelingsinstrumenten in de kunstvakken. De vaak genoemde diversiteit vinden we ook in deze literatuur terug. Beoordelingsinstrumenten voor muziek zijn duidelijk in de meerderheid, voor drama vinden we maar heel weinig instrumenten beschreven in de literatuur. Voor beeldend vinden we meer instrumenten dan voor dans. Er worden ongeveer evenveel instrumenten beschreven voor het primair onderwijs als voor het secundair onderwijs.

In deze selectie van beoordelingsinstrumenten maakten we een onderscheid in verschillende taken en diverse scoringsinstrumenten. Het meest voorkomende instrumentarium is een uiting in het artistieke medium (zoals een muziekkuitvoering of een beeldend product) dat beoordeeld wordt aan de hand van criterialijsten of rubrics. Dit geldt zonder uitzondering voor alle kunstdisciplines.

Om de vraag te beantwoorden wat er met de instrumenten wordt beoordeeld hebben we een overzicht gemaakt van de belangrijkste competenties die de instrumenten bestrijken: perform, create, respond en reflect. Het blijkt dat de meeste instrumenten beschreven worden voor competenties die vallen binnen het spectrum van perform (podiumkunsten) en create (beeldend). Beoordelingscriteria die te maken hebben met het beschouwen van kunst en vooral met reflecteren komen verhoudingsgewijs weinig voor. De beoordelingscriteria die worden gehanteerd om deze competenties (en bijbehorende deelcompetenties) te beoordelen laten een zeer uiteenlopend en breed spectrum zien. Het is

duidelijk dat elke kunstdiscipline zijn eigen terminologie en eigen vakspecifieke competenties beoordeeld in de verschillende instrumenten. Toch vinden we op een hoger abstractieniveau wel degelijk overeenkomsten tussen de verschillende disciplines. Algemene competenties die we bij meerdere disciplines tegenkomen zijn: technische beheersing, originaliteit/creativiteit, zelfreflectie, samenwerking, inzet/houding en interpretatie (perform).

Voor alle beschreven instrumenten in deze review geldt dat ze zijn beschreven in een peer reviewed tijdschrift en dat er in het betreffende artikel iets gezegd wordt over de kwaliteit van het instrument. De kwaliteitsmetingen zijn echter minstens zo heterogeen als de instrumenten zelf. We konden op basis van de beschrijvingen twintig verschillende kwaliteitsoordelen onderscheiden. Elf hiervan gaan over kwaliteitsoordelen waarvoor empirisch onderzoek is uitgevoerd, de negen andere oordelen worden 'face value' gemaakt. Empirisch onderzoek wordt er gedaan naar de betrouwbaarheid en validiteit van de instrumenten. De betrouwbaarheid van bijvoorbeeld verschillende criteria wordt gemeten aan de hand van interne consistentie (Cronbachs alpha) of een interbeoordelaarsbetrouwbaarheid. Ook worden er verschillende typen validiteit onderscheiden. Bijna alle beschreven instrumenten scoren minimaal een voldoende en vaker ruim voldoende of goed op de gekozen kwaliteitsmetingen. Slechte instrumenten komen in de opgenomen artikelen nauwelijks voor. Een vergelijking van de beoordelingsinstrumenten op basis van de kwaliteitsmetingen is lastig vanwege de grote heterogeniteit van de kwaliteitsmetingen. Interessant in termen van kwaliteit zijn die artikelen waarin verschillende beoordelingsinstrumenten of verschillende versies van hetzelfde instrument met elkaar vergeleken worden (op basis van dezelfde kwaliteitsmetingen). Deze artikelen komen echter maar sporadisch voor in deze review.

De tijdschriften waren zonder uitzondering afkomstig uit de Verenigde Staten of Groot-Brittannië. Veel artikelen werden gevonden via de Amerikaanse database ERIC. Dit betekent dat er een bias in onze selectie zit. Uit een proefsearch is gebleken dat in de Duitstalige literatuur diverse relevante publicaties op het gebied van muziek en beeldende kunst zijn verschenen. We weten uit de praktijk dat er veel meer beoordelingsinstrumenten beschikbaar zijn en gebruikt worden, maar deze instrumenten vinden kennelijk niet hun weg naar de wetenschappelijke literatuur. Wel wordt er over gepubliceerd in vaktijdschriften voor docenten. Daarbij wordt aandacht besteed aan het kwaliteitsaspect praktische haalbaarheid, maar niet aan validiteit en betrouwbaarheid.

Vervolgonderzoek zou zich enerzijds kunnen richten op een uitbreiding van deze review in de richting van anderstalige literatuur en handboeken voor de praktijk. Anderzijds zou vervolgonderzoek zich kunnen richten op empirisch onderzoek waarin verschillende beoordelingsinstrumenten of verschillende typen van hetzelfde instrument systematisch met elkaar vergeleken worden. Er wordt in peer review artikelen nog weinig gerapporteerd over instrumenten die gericht zijn op beoordelen bij theater, het beoordelen van samenwerking bij uitingen in het artistieke medium, en van kunstbeschouwing en reflectie. Meer onderzoek op deze terreinen is gewenst, zodat er meer inzicht kan worden gegeven in beoordelingsinstrumenten die zich hierop richten en de kwaliteit daarvan.

De lijst met artikelen en de metagegevens over deze artikelen zoals we die voor deze review verzameld hebben worden toegankelijk gemaakt via de website van het Landelijk Kennisinstituut voor Cultuureducatie en Amateurkunst (LKCA)¹. We hopen daarmee voor zowel leerkrachten, onderzoekers als vakexperts een bron van mogelijke beoordelingsinstrumenten te ontsluiten die kan dienen ter inspiratie voor praktijk en verder onderzoek naar beoordelingen in de kunstvakken.

¹ www.lkca.nl/publicaties/informatiebank#

Literatuur

- Abril, C. R. (2006). Learning outcomes of two approaches to multicultural music education. *International Journal of Music Education* 24(1), 30-42.
- Beattie, D. K (1997) *Assessment in art education*. Worcester, MA: Davis Publications.
- Bergee, M. J. (2007). Performer, rater, occasion, and sequence as sources of variability in music performance assessment. *Journal of research in music education*, 55(4), 344-358.
- Bresler, L. (Ed) (2007). *International handbook of research in arts education*. Dordrecht: Springer Verlag.
- Chan, D. W., Chan, L., & Chau, A. (2009). Judging Drawing Abilities of Hong Kong Chinese Gifted Students: Could Nonexperts Make Expert-Like Judgments? *Gifted Child Quarterly* 53(1), 15-24.
- Colwell, R. & Richardson, C. (Eds) (2002). *The New Handbook of Research in Music Teaching and Learning*. Oxford: Oxford University Press.
- Dorn, C. M., & Sabol, F. R. (2006). The Effectiveness and Use of Digital Portfolios for the Assessment of Art Performances in Selected Secondary Schools. *Studies in Art Education*, 47(4), 344-362.
- Dunbar, L. (2012). Performance assessment of the masses in 30 seconds or less. *General music today*, 25(2), 31-35.
- Englebright, K., & Mahoney, M. R. (2012). Assessment in elementary dance education. *Journal of dance education* 12(3), 87-92.
- Fautley, M (2010). *Assessment in music education* . Oxford music education series.
- Gardner, H. (1996). The assessment of student learning in the arts. In D. Boughton, E. Eisner & J. Ligtvoet (Eds) *Evaluating and assessing the visual arts in education* (pp 131-155). New York: Teachers College Press.
- Gower, A. (2012). Integrating informal learning approaches into the formal learning environment of mainstream secondary schools in England. *British journal of music education*, 29(1), 13-18.
- Gullikers, J. (2011). Authentiek beoordelen: kansen voor het kunstonderwijs. *Cultuur + Educatie* 31 (pp. 38- 61). Utrecht: Cultuurnetwerk.
- Herpin, S., Washington, A.Q., & Li, J. (2012). *Improving the assessment of student learning in the arts- State of the field and recommendations*. Washington: National Endowments for the Arts.
- Hoeven, M. van der, e.a. (2014). *Cultuur in de Spiegel in de praktijk. Een leerplankader voor cultuuronderwijs*. Enschede: SLO.
- Kaufman, J. C., Evans, M. L., & Baer, J. (2010). The American Idol effect: Are students good judges of their creativity across the domains? *Empirical studies of the arts*, 28(1), 3-18.
- Leighton, G., & Lamont, A. (2006). Exploring children's singing development: Do experiences in early schooling help or hinder? *Music education research*, 8(3), 311-330.
- Lindström, L. (2006). Creativity:What Is It? Can You Assess It? Can It Be Taught? *International Journal of Art & Design Education*, 25(1), 53-66.
- Miller, B. A. (2009). Mom and Little Richard: The benefits of contrast-and-compare activities. *General music today*, 23(1), 4-11.
- Myford, C. M. (2002). Investigating design features of descriptive graphic rating scales. *Applied Measurement in Education* 15(2), 187-215.
- Onderwijsraad & Raad voor Cultuur (2012). *Cultuureducatie: leren, creëren, inspireren!* Den Haag: Onderwijsraad.
- Oostwoud Wijdenes, J. & Haanstra, F. (1997). *Over actief, receptief en reflectief*. Utrecht:

LOKV.

- Oreck, B. A. (2005). A Powerful Conversation: Teachers and Artists Collaborate in Performance-Based Assessment. *Teaching Artist Journal*, 3(4), 220-227.
- Oreck, B. A., Owen, S. V., & Baum, S. M. (2003). Validity, Reliability, and Equity Issues in an Observational Talent Assessment Process in the Performing Arts. *Journal for the Education of the Gifted* 27(1), 62-94.
- Randles, C. A. (2010). The Relationship of Compositional Experiences of High School Instrumentalists to Music Self-Concept. *Bulletin of the Council for Research in Music Education*, 184, 9-20.
- Schmid, D. (2003). Authentic assessment in the arts: Empowering students and teachers. *Journal of dance education*, 3(2), 65-73.
- Schneider, B., Zumtobel, M., Prettenhofer, W., Aichstill, B., & Jocher, W. (2010). Normative voice range profiles in vocally trained and untrained children aged between 7 and 10 years. *Journal of voice*, 24(2), 153-160.
- SLO (2014) Leerplankader kunstzinnige oriëntatie. Geraadpleegd op 6 augustus 2014 van <http://www.kunstzinnigeoriëntatie.slo.nl>
- Soep, E. (2004). Assessment and visual arts education In E. W. Eisner & M.D.Day (Eds.), *Handbook of research and policy in art education*. (pp.579 - 583). New Jersey, London: Lawrence Erlbaum Publishers.
- Taylor, P. (Ed) (2006). *Assessment in arts education*. Porthmouth: Heinemann.
- Thorpe, V. (2012). Assessment rocks? The assessment of group composing for qualification. *Music education research*, 14(4), 417-429.
- Trollinger, V. L. (2003). Relationships between pitch-matching accuracy, speech fundamental frequency, speech range, age, and gender in American English-speaking preschool children. *Journal of research in music education*, 51(1), 78-94.
- Watts, R. (2010). Responding to Children's Drawings. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 38(2), 137-153.
- Weprin, K. L. (2003). Spin Your Wheels. *School Arts*, 102(8), 54.

Bijlage Zoekwoorden Review

ERIC

Educational Resources Information Center (OvidSP)

#1 Arts

art education/ OR art/ OR fine arts/ OR visual arts/ OR studio art/ OR art teachers/ OR childrens art/
OR art history/ OR art.ti,ab,id. OR arts.ti,ab,id. OR artistic*.ti,ab,id. OR aesthetic education/ OR
aesthetic*.ti,ab,id. OR theater arts/ OR drama/ OR theater*.ti,ab,id. OR theatr*.ti,ab,id. OR
drama.ti,ab,id. OR dance education/ OR dance/ OR danc*.ti,ab,id. OR ballet.ti,ab,id. OR
choreograph*.ti,ab,id. OR music education/ OR music/ OR music appreciation/ OR music teachers/
OR music*.ti,ab,id. OR audiovisual instruction/ OR audiovisual instruction*.ti,ab,id. OR audiovisual
education*.ti,ab,id. OR film production/ OR film study/ OR critical viewing/ OR design crafts/ OR
design education.ti,ab,id. OR design & technology education.ti,ab,id. OR "design and technology
education".ti,ab,id.

Results: 85.802 (October 2, 2013)

#2 Assessment

educational assessment/ OR alternative assessment/ OR informal assessment/ OR assessment*.ti,ab.
OR evaluation/ OR evaluation methods/ OR formative evaluation/ OR formative.ti,ab. OR summative
evaluation/ OR summative.ti,ab. OR holistic evaluation/ OR peer evaluation/ OR "self evaluation
(individuals)"/ OR student evaluation/ OR evaluation criteria/ OR evaluation*.ti,ab. OR measurement/
OR measurement*.ti,ab. OR testing/ OR tests/ OR test.ti,ab. OR tests.ti,ab. OR performance tests/ OR
standardized tests/ OR test reliability/ OR test validity/ OR standards/ OR exit examinations/ OR exit
examination*.ti,ab. OR achievement rating/ OR reflection/ OR reflection.ti,ab. OR observation/ OR
portfolio assessment/ OR portfolio*.ti,ab. OR process folio*.ti,ab. OR product folio*.ti,ab. OR journal
writing/ OR journal?ing.ti,ab. OR scoring rubrics/ OR rubric*.ti,ab.

Results: 386.016 (October 2, 2013)

#3 Primary and secondary school students

(students/ OR student*.ti,ab. OR (early childhood education or preschool education OR elementary
education OR elementary secondary education OR grade 1 OR grade 2 OR grade 3 OR grade 4 OR
grade 5 OR grade 6 OR grade 7 OR grade 8 OR grade 9 OR grade 10 OR grade 11 OR grade 12 OR
intermediate grades OR junior high schools OR middle schools OR primary education OR secondary
education OR high schools).el. OR nursery schools/ OR kindergarten/ OR kindergarten*.ti,ab. OR
early childhood education/ OR early childhood education.ti,ab. OR preschool children/ OR preschool
education/ OR preschool*.ti,ab. OR grade 1/ OR grade 2/ OR grade 3/ OR grade 4/ OR grade 5/ OR
grade 6/ OR grade 7/ OR grade 8/ OR grade 9/ OR grade 10/ OR grade 11/ OR grade 12/ OR K-
12*.ti,ab. OR K12.ti,ab. OR elementary secondary education/ OR elementary schools/ OR elementary
education/ OR elementary school students/ OR elementary education.ti,ab. OR primary education/ OR
primary education.ti,ab. OR public schools/ OR secondary schools/ OR secondary school students/
OR secondary education/ OR secondary education.ti,ab. OR junior high schools/ OR junior high
school students/ OR junior high.ti,ab. OR high schools/ OR high school students/ OR school*.ti,ab. OR
classroom*.ti,ab.) OR ((children/ OR young children/ OR child*.ti,ab. OR kid.ti,ab. OR kids.ti,ab. OR
preadolescents/ OR prepubescen*.ti,ab. OR prepuberty*.ti,ab. OR puberty/ OR puberty.ti,ab. OR
pubescen*.ti,ab. OR teen*.ti,ab. OR young*.ti,ab. OR youth/ OR youth*.ti,ab. OR girl*.ti,ab. OR
boy*.ti,ab. OR youth/ OR youth*.ti,ab. OR adolescents/ OR early adolescents/ OR adolescen*.ti,ab.)
AND (educat*.ti,ab. OR teach*.ti,ab. OR learn*.ti,ab. OR grade.ti,ab.))

Results: 954.447 (October 2, 2013)

#4 Therapy and language arts

exp therapy/ OR therap*.ti,ab. OR exp language arts OR language arts.ti,ab.

Results: 122.218 (October 2, 2013)

**(1 AND 2 AND 3) NOT 4
Limit 1993 - current**

**11.115 results (October 2, 2013)
4.882 results (October 2, 2013)**

PsycINFO

OvidSP

#1 Arts

art education/ OR art/ OR art.ti,ab,id. OR arts.ti,ab,id. OR artistic*.ti,ab,id. OR aesthetics/ OR aesthetic*.ti,ab,id. OR drama/ OR theater*.ti,ab,id. OR theatr*.ti,ab,id. OR drama.ti,ab,id. OR dance/ OR danc*.ti,ab,id. OR ballet.ti,ab,id. OR choreograph*.ti,ab,id. OR improvisation/ OR music/ OR music education/ OR music perception/ OR music*.ti,ab,id. OR audiovisual instruction/ OR audiovisual instruction*.ti,ab,id. OR audiovisual education*.ti,ab,id. OR film study/ OR design education.ti,ab,id. OR design & technology education.ti,ab,id. OR "design and technology education".ti,ab,id.

Results: 82.029 (October 2, 2013)

#2 Assessment

educational measurement/ OR assessment*.ti,ab,id. OR evaluation/ OR formative.ti,ab,id. OR summative.ti,ab,id. OR peer evaluation/ OR self evaluation/ OR evaluation criteria/ OR evaluation*.ti,ab,id. OR measurement*.ti,ab,id. OR test.ti,ab,id. OR tests.ti,ab,id. OR performance tests/ OR standardized tests/ OR test reliability/ OR test validity/ OR educational standards/ OR reflection.ti,ab,id. OR portfolio*.ti,ab,id. OR process folio*.ti,ab,id. OR journal writing/ OR journal?ing.ti,ab,id. OR rubric*.ti,ab,id.

Results: 814.676 (October 2, 2013)

#3 Primary and secondary school students

students/ OR student*.ti,ab,id. OR nursery schools/ OR nursery school students/ OR kindergartens/ OR kindergarten students/ OR kindergarten*.ti,ab,id. OR early childhood education.ti,ab,id. OR preschool education/ OR preschool students/ OR preschool*.ti,ab,id. OR elementary schools/ OR elementary education/ OR elementary school students/ OR elementary education.ti,ab,id. OR primary school students/ OR primary education.ti,ab,id. OR K-12*.ti,ab,id. OR K12.ti,ab,id. OR public school education/ OR middle schools/ OR middle school education/ OR middle school students/ OR secondary education/ OR secondary education.ti,ab,id. OR junior high schools/ OR junior high school students/ OR junior high.ti,ab,id. OR high schools/ OR high school students/ OR high school education/ OR school*.ti,ab,id. OR classroom*.ti,ab,id. OR (((preschool age 2 5 yrs OR school age 6 12 yrs OR adolescence 13 17 yrs).ag. OR child*.ti,ab,id. OR kid.ti,ab,id. OR kids.ti,ab,id. OR prepubescen*.ti,ab,id. OR prepuberty*.ti,ab,id. OR teen*.ti,ab,id. OR young*.ti,ab,id. OR youth*.ti,ab,id. OR girl*.ti,ab,id. OR boy*.ti,ab,id. OR adolescen*.ti,ab,id.)) AND (educat*.ti,ab,id. OR teach*.ti,ab,id. OR learn*.ti,ab,id. OR grade.ti,ab,id.))

Results: 758.299 (October 2, 2013)

#4 Therapy or language arts

exp treatment/ OR therap*.ti,ab,id. OR exp language arts education/ OR language arts.ti,ab,id.

Results: 665.025 (October 2, 2013)

(1 AND 2 AND 3) NOT 4

Limit 1993 - current

5.306 results (October 2, 2013)

2.884 results (October 2, 2013)

Art & Architecture Index
EBSCO

#1 Assessment

DE ("evaluation" OR "GRADING & marking (Students)" OR "EDUCATIONAL evaluation -- Research" OR "EDUCATIONAL evaluation" OR "SELF-evaluation") OR TI ("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR AB ("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR KW ("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*")

Results: 32.268 (October 2, 2013)

Verwijderd:

DE "ARTS -- Study & teaching -- Evaluation"

DE "STUDENTS -- Self-rating of"

#2 Primary and secondary school students

DE ("LEARNING" OR "STUDENTS" OR "ART -- Study & teaching") OR TI (("student*" OR "kindergarten" OR "early childhood education" OR "preschool" OR "school" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade"))) OR AB (("student*" OR "kindergarten" OR "early childhood education" OR "preschool" OR "school" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade"))) OR KW(("student*" OR "kindergarten" OR "early childhood education" OR "preschool" OR "school" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade")))

Results: 52.125 (October 2, 2013)

#3 Therapy and language arts

TI ("therap*" OR "language arts") OR AB ("therap*" OR "language arts") OR KW ("therap*" OR "language arts")

Results: 2.353 (October 2, 2013)

(1 AND 2) NOT 3

2.083 results (October 2, 2013)

Limit 1993 - current

1.609 results (October 2, 2013)

RILM Abstracts of Music Literature

Repertoire International de Litterature Musicale (EBSCO)

#1 Assessment

TI ("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR AB ("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR SU ("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*")

Results: 17.142 (October 2, 2013)

#2 Primary and secondary school students

TI (("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*"))) OR AB (("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*"))) OR SU (("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*")))

Results: 47.351 (October 2, 2013)

#3 Therapy and language arts

TI ("therap*" OR "language arts") OR AB ("therap*" OR "language arts") OR SU ("therap*" OR "language arts")

Results: 9.843 (October 2, 2013)

(1 AND 2) NOT 3

3.889 results (October 2, 2013)

Limit 1993 - current,

Language: Dutch, English, German 2.372 results (October 2, 2013)

International Index to Performing Arts

Proquest

#1 Assessment

TI("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR AB("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR BSU("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*") OR NSU("assessment*" OR "formative" OR "summative" OR "evaluation*" OR "measurement*" OR "test" OR "tests" OR "reflection" OR "portfolio*" OR "process folio*" OR "journal?ing" OR "rubric*")

Results: 4.388 (October 2, 2013)

#2 Primary and secondary school students

TI(("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*")) OR AB(("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*")) OR BSU(("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*")) OR NSU(("student*" OR "kindergarten*" OR "early childhood education" OR "preschool*" OR "school*" OR "K-12*" OR "K12" OR "elementary education" OR "primary education" OR "secondary education" OR "junior high" OR "classroom*") OR (("child*" OR "kid" OR "kids" OR "prepubescen*" OR "prepuberty*" OR "puberty" OR "pubescen*" OR "teen*" OR "young*" OR "youth*" OR "girl*" OR "boy*" OR "adolescen*") AND ("educat*" OR "teach*" OR "learn*" OR "grade*"))

Results: 24.593 (October 2, 2013)

#3 Therapy and language arts

TI("therap*" OR "language arts") OR AB("therap*" OR "language arts") OR BSU("therap*" OR "language arts") OR NSU("therap*" OR "language arts")

Results: 1.161 (October 2, 2013)

(1 AND 2) NOT 3

392 results (October 2, 2013)

Limit 1993 - current

391 results (October 2, 2013)

Assessment in kunsteducatie

Samenvatting - Eindrapport NWO Review studie 411-12-228

2014

Universiteit van Amsterdam

Talita Groenendijk

*Universiteit van Amsterdam, Research Institute of Child Development and Education/
Amsterdamse Hogeschool voor de Kunsten*

Marie-Louise Damen

Universiteit van Amsterdam, Research Institute of Child Development and Education

Folkert Haanstra

Amsterdamse Hogeschool voor de Kunsten

Carla van Boxtel

Universiteit van Amsterdam, Research Institute of Child Development and Education

Met het programma 'Cultuureducatie met kwaliteit' tracht de Nederlandse overheid de ontwikkeling van doorgaande leerlijnen en van een beoordelingsinstrumentarium te stimuleren. Voor leerlijnen zijn landelijke leerplankaders ontwikkeld die als richtsnoer kunnen dienen. Voor het ontwikkelen van een beoordelingsinstrumentarium ontbreekt zo'n richtsnoer. Er bestaat wel literatuur op dit gebied. Deze beschrijft een grote variëteit aan beoordelingsinstrumenten en vormen. Naast het onderscheid in kunstdisciplines en het onderscheid in de belangrijkste kunst domeinen (actief en receptief), wordt onderscheid gemaakt tussen de verschillende functies van de beoordelingsvormen: formatief, summatief en evaluatief. De verschillende domeinen en functies resulteren in verschillende beoordelingsvormen. Wat echter ontbreekt is een overzicht van beoordelingsinstrumenten waarin niet alleen wordt aangegeven wat en hoe beoordeeld wordt, maar ook gegevens worden verzameld over kwaliteitscriteria zoals validiteit, betrouwbaarheid, bruikbaarheid en eerlijkheid.

In deze review werden de volgende onderzoeksvragen beantwoord:

- a) Welke beoordelingsinstrumenten voor de kunstvakken in het basis- en voortgezet onderwijs zijn in de literatuur sinds 2000 beschreven?
- b) Wat wordt er met deze instrumenten beoordeeld?
- c) Wat wordt gerapporteerd over de kwaliteit van deze instrumenten?

De review omvatte vier fasen. De eerste fase betrof de zoekopdracht zelf. In databases als Eric, Psychinfo en kunstvakspecifieke databases is met een zoekprofiel een search gedaan naar literatuur van de afgelopen dertien jaar (2000-2013). Dit resulteerde in 8336 titels.

In de tweede fase is in enkele stappen een selectie van de publicaties gemaakt. De eerste selectie betrof publicaties met een beschrijving van een instrument of procedure van beoordeling om het leren van leerlingen in de kunstvakken basis- of voortgezet onderwijs te monitoren of beoordelen. Om tot een verdere inperking van resultaten te komen, werden alleen peer reviewed journal artikelen opgenomen. Tevens werd besloten dat het artikel iets moest zeggen over de kwaliteit van het instrument of de procedure, zoals bijvoorbeeld de validiteit en betrouwbaarheid. Dit resulteerde in 119 artikelen uit 44 peer reviewed Engelstalige tijdschriften (allen uitgegeven in de Verenigde Staten of Groot-Brittannië), waarin 153 beoordelingsinstrumenten zijn beschreven

In de derde fase zijn de geselecteerde instrumenten gecodeerd en beschreven. Codes betreffen bibliografische gegevens, kunstdiscipline, onderwijsniveau, toetstaak (artistieke uiting, pen-en-papier opdracht, portfolio, etc.); type scoringsinstrument (rubric, criterialijst, observatielijst etc.); object van meting (create, perform, respond), functie (summatief of formatief); proces of product beoordeling; beoordelingscriteria en kwaliteitskenmerken.

In de vierde fase vond een kwalitatieve synthese plaats en werden de onderzoeksvragen beantwoord.

De review studie laat zien dat de beoordelingstaken en scoringsinstrumenten voor de kunstvakken divers zijn. Er is een vergelijkbaar aantal instrumenten gericht op primair onderwijs en voortgezet onderwijs. De meeste beoordelingsinstrumenten die in de wetenschappelijke literatuur worden beschreven hebben betrekking op muziek. Voor drama is slechts een gering aantal instrumenten gevonden. Meestal gaat het om de beoordeling van een uiting in het artistieke medium (zoals een muziekuivoering of een beeldend product) aan de hand van criterialijsten of rubrics. De meeste instrumenten richten zich op competenties die vallen onder 'perform' en 'create'. Elke kunstdiscipline heeft zijn eigen terminologie en vakspecifieke competenties. Beoordelingscriteria die te maken hebben met het beschouwen van kunst of met reflecteren komen relatief weinig voor in de review. De instrumenten beperken zich niet tot makkelijk te meten kennis of technische vaardigheden, maar hebben ook betrekking op kwalitatieve aspecten, zoals bijvoorbeeld expressie, toonkwaliteit, sfeer en originaliteit. De kwaliteitsmetingen zijn erg heterogeen. Er konden twintig verschillende kwaliteitsoordelen onderscheiden worden. De instrumenten die in de review zijn opgenomen scoorden in bijna alle gevallen een voldoende tot goed op de gekozen kwaliteitsmetingen. In een aantal artikelen werd de kwaliteit van verschillende beoordelingsinstrumenten met elkaar vergeleken.

De review beperkte zich tot publicaties in Engelstalige peer-reviewed tijdschriften waarin iets word gezegd over de kwaliteit van de beschreven instrumenten. In vervolgonderzoek zou ook gekeken kunnen worden naar handboeken voor de praktijk en niet-Engelstalige literatuur. Vervolgonderzoek kan zich tevens richten op instrumenten voor beoordeling binnen de discipline theater, de beoordeling van samenwerking bij uitingen in het artistieke medium en van kunstbeschouwing en reflectie. Over deze aspecten van beoordelen in de kunstvakken wordt nog nauwelijks gerapporteerd in wetenschappelijke peer-reviewed tijdschriften.

Ook in de kunstvakken worden leerlingen steeds vaker beoordeeld. Deze beoordeling kan – passend bij de specifieke kunstdiscipline en de visie op beoordelen – op verschillende manieren. De review biedt leerkrachten, vakexperts en onderzoekers een overzicht van mogelijke beoordelingsinstrumenten (taken en scoringsinstrumenten) voor de beoordeling van het leren binnen de kunstvakken en van mogelijke beoordelingscriteria voor verschillende competenties.