

DE INTEGRATIE VAN TAAL EN SCIENCE ONDERWIJS:
ANALYSE- EN BEOORDELINGSINSTRUMENT VOOR SCHRIJFPRODUCTEN VAN
LEERLINGEN IN DE MIDDEN- EN BOVENBOUW VAN HET PRIMAIR ONDERWIJS

Paulien Bosch, 2007

Samenvatting

Dit artikel beschrijft een ontwerpgericht onderzoek dat ten doel heeft een betrouwbaar, valide en praktisch bruikbaar instrument te ontwikkelen om schrijfproducten van leerlingen in de midden- en bovenbouw van het primair onderwijs te analyseren en beoordelen. Het gaat in dit onderzoek specifiek om schrijfproducten, die leerlingen produceren tijdens nieuw ontwikkelde lessenseries rond het thema 'Het Weer', waarin taal en natuuronderwijs worden geïntegreerd.

De ontwikkeling en optimalisering van een betrouwbaar, valide en praktisch bruikbaar analyse- en beoordelingsinstrument heeft plaats gevonden door een aantal fasen te doorlopen. Ten eerste is een eerste versie van het instrument ontwikkeld aan de hand van een literatuurstudie en doormiddel van overleg met vijf deskundigen. Daarna is een pilottest uitgevoerd om de eerste versie van het instrument te testen. Een viertal leerkrachten en een vijftal deskundigen hebben versie 1 van het instrument geëvalueerd. Bij de leerkrachten is informatie verkregen door middel van de 'think-aloud' procedure (hardop denken) en door middel van het stellen van vragen. Bij de deskundigen is informatie verkregen door middel van een groepsdiscussie over de gegeven scores. Aan de hand van de resultaten van de pilottest is een tweede versie van het instrument ontwikkeld en zijn er een aantal discussiepunten opgesteld voor de eindtest. Ten slotte is er een eindtest uitgevoerd om de tweede versie van het instrument te testen. Hiervoor hebben drie leerkrachten versie 2 van het instrument geëvalueerd en is er informatie verkregen door middel van een groepsdiscussie. Aan de hand van de resultaten van de eindtest zijn aanbevelingen geformuleerd met het oog op de toekomst en er is een derde versie van het instrument in ontwikkeling.

Uit het onderzoek is gebleken dat versie 2 van het analyse- en beoordelingsinstrument met een aantal aanpassingen redelijk betrouwbaar, valide en praktisch bruikbaar is. Het instrument is een rubric met een onderverdeling in vier niveaus. De kenmerken die geanalyseerd en beoordeeld kunnen worden met behulp van het instrument zijn inhoud science, inhoud talig, structuur, communicatie en conventies. Deze kenmerken zijn weer onderverdeeld in subkenmerken. De gewenste aanpassingen aan versie 2 worden gegeven in de aanbevelingen en worden verwerkt in versie 3 van het instrument, die op dit moment in ontwikkeling is. Daarnaast worden er nog aanbevelingen gedaan over het gebruik van het instrument en over veranderingen die aangebracht zouden moeten worden aan het curriculummateriaal om de betrouwbaarheid, validiteit en praktische bruikbaarheid van het instrument te optimaliseren.

Inleiding¹

Taal is een belangrijk middel voor communicatie en kennisverwerving (conceptualisering) bij natuuronderwijs oftewel science onderwijs (Freeman & Taylor, 2006). In het Nederlands onderwijs wordt echter veelal onderscheid gemaakt tussen de leerlijnen van de zaakvakken, waaronder science, en de leerlijnen van taal. Dit vertaalt zich in de dagelijkse onderwijspraktijk naar aparte lessen voor taal en de zaakvakken. Hoewel al langere tijd wordt gestreefd naar meer integratie tussen beide leerlijnen, omdat dit de leerresultaten ten goede komt (Freeman & Taylor, 2006), komt er in de praktijk weinig van terecht.

Om na te gaan welke mogelijkheden ICT biedt om de integratie tussen taal en de zaakvakken te bevorderen is SLO in 2005 gestart met het Research & Development project 'Taal bij de zaakvakken en de potentie van ICT'.

In 2005-2006 is in het kader van dit project curriculummateriaal ontwikkeld voor groep 5, 6, 7 en 8 van het primair onderwijs rond het thema 'Het Weer'. Dit materiaal is formatief (tussentijds) geëvalueerd. Het curriculummateriaal ondersteunt leerkrachten bij de integratie van taal, voornamelijk schrijven, in het science onderwijs. Er is expliciet gebruik gemaakt van ICT, het programma Kidspiration, voor het maken van conceptmaps als tussenproduct bij het schrijven. In 2007 is de aandacht verlegd naar de implementatie van het materiaal bij nieuwe docenten en tevens naar de effectiviteit van het materiaal met het oog op het leren van leerlingen. Dit laatste vormt de aanleiding voor dit onderzoek.

De leerresultaten van leerlingen kunnen onder andere worden gemeten door de schrijfproducten, die de leerlingen produceren tijdens de lessen van het nieuw ontwikkelde curriculummateriaal, te analyseren en beoordelen. Hiervoor is echter nog geen geschikt instrument. Dit onderzoek heeft daarom als doel een instrument te ontwikkelen om de taalontwikkeling en de ontwikkeling van science concepten van leerlingen te kunnen analyseren en beoordelen aan de hand van schrijfproducten, zodat er op termijn iets gezegd kan worden over de effectiviteit van het ontwikkelde curriculummateriaal met het oog op het leren van leerlingen. Dit heeft geresulteerd in de volgende onderzoeksvraag:

Hoe ziet een betrouwbaar, valide en praktisch bruikbaar instrument eruit, dat geschikt is om de schrijfproducten van basisschoolleerlingen, geproduceerd tijdens de nieuw ontwikkelde lessenseries rond het thema 'Het Weer', te analyseren en beoordelen op het gebied van taalontwikkeling en op het gebied van de ontwikkeling van science concepten?

Theoretisch kader

Op het gebied van taal en science moeten leerlingen in het primair onderwijs een ontwikkeling laten zien wat betreft inhoud en proces (Paus, 2006; Smits & Voogt, 2006). Deze ontwikkeling kan deels worden gemeten door middel van het analyseren en beoordelen van schrijfproducten van de leerlingen.

De schrijfproducten kunnen geanalyseerd en beoordeeld worden door op verschillende kenmerken te letten. Deze kenmerken zijn inhoud, organisatie, communicatie, stijl en conventies (Krom et al., 2004; Miller & Calfee, 2004; Roid, 1994; Van de Gein, 2004; Wolowitsj in Wesdorp, 1978). Bij inhoud gaat het erom dat de tekst inhoudelijk correct is. Bij organisatie gaat het om de juiste ordening van de inhoudelijke informatie in de tekst en om het gebruik van de juiste structuur. Bij communicatie gaat het erom dat de schrijver doelgericht en publiekgericht schrijft. Bij stijl gaat het om het juiste woordgebruik en om het gebruik van lopende zinnen en bij het kenmerk conventies gaat het om de juiste spelling, grammatica en interpunctie. Bij het analyseren en beoordelen van schrijfproducten is het wenselijk vooraf te bepalen op welke kenmerken er beoordeeld moet gaan worden.

¹ Met dank aan dr. J.M. Voogt (Universiteit Twente), dr. S.E. Mckenney (Universiteit Twente) en drs. M.C. Smits (Stichting Leerplanontwikkeling) voor de begeleiding bij dit onderzoek.

Er zijn twee methoden mogelijk om schrijfproducten te analyseren en beoordelen, de holistische methode (één algemeen oordeel) en de analytische methode (oordeel per kenmerk). De holistische methode is minder betrouwbaar (Bauer, Veal & Hudson in Huot, 1990; Roid, 1994; Wesdorp, 1978) en heeft een lagere diagnostische functie (Roid, 1994; Wiggins, 1998) dan de analytische methode, waardoor de analytische methode een lichte voorkeur heeft. De keuze van de scoringsmethode is echter afhankelijk van de specifieke situatie waarin je hem gebruikt (Huot, 1994; Meredith & Williams, 1984; Krom et al., 2004) en er kan ook een combinatie van methoden worden gebruikt (Krom et al., 2004). Bij het analyseren en beoordelen van schrijfproducten is het goed om te bepalen welke methode het beste gebruikt kan worden.

Naast de methode moet ook worden bepaald welke vorm het analyse- en beoordelingsinstrument krijgt. Er zijn verschillende vormen mogelijk waarmee beoordeeld en geanalyseerd kan worden, bijvoorbeeld een beoordelingsschaal, een rubric of een checklist.

Ten slotte kan bij de implementatie van een nieuw instrument voor het analyseren en beoordelen van schrijfproducten rekening worden gehouden met toepasbaarheid, samenhang met de huidige situatie, investering ten opzichte van resultaat en vertrouwen van ouders en schoolbesturen.

Methode

Om de onderzoeksvraag te kunnen beantwoorden is een ontwerpgericht onderzoek uitgevoerd. Ontwerpgericht onderzoek richt zich op het ontwikkelen en optimaliseren van het analyse- en beoordelingsinstrument voor schrijfproducten door middel van evaluatie.

De leerlingen produceren tijdens de lessenseries twee soorten schrijfproducten. Het eerste soort schrijfproduct is een proevenboekje, dat tijdens de eerste vier lessen wordt gemaakt. Het proevenboekje is een tussenproduct, waarbij het schrijven vooral een manier is om te leren over science en een manier om te leren schrijven.

Aan het eind van de lessenserie wordt een tweede soort schrijfproduct gemaakt. Het verschilt per lessenserie welke vorm dit eindproduct heeft. Het eindproduct van de eerste lessenserie zal in dit onderzoek worden gebruikt. Bij deze eerste lessenserie was het eindproduct een groot proevenboek waarin leerlingen hun zelfbedachte proef hebben beschreven. Andere leerlingen van ongeveer dezelfde leeftijd vormden de doelgroep bij deze opdracht. Bij het eindproduct is het schrijven meer een vorm om te testen wat de leerlingen uiteindelijk weten (over het onderwerp) en kunnen (vaardigheden).

Het instrument om deze schrijfproducten te analyseren en beoordelen is geschikt als het betrouwbaar, valide en praktisch bruikbaar is. Bij betrouwbaarheid gaat het in dit onderzoek om de interbeoordelaarsbetrouwbaarheid van het te ontwikkelen analyse- en beoordelingsinstrument. Het instrument is betrouwbaar wanneer verschillende beoordelaars, die hetzelfde schrijfproduct beoordelen, tot hetzelfde oordeel komen.

Bij validiteit gaat het in dit onderzoek om de inhoudsvaliditeit van het te ontwikkelen analyse- en beoordelingsinstrument. Het instrument is inhoudsvalide wanneer het de taalontwikkeling en de ontwikkeling van science concepten meet.

Bij praktische bruikbaarheid is het belangrijk om rekening te houden met de leerkrachten, want dit zijn de uiteindelijke gebruikers. Er zal daarom worden gelet op de duidelijkheid van de gebruikswijze van het instrument (toepasbaarheid), op de aansluiting bij de huidige manier van analyseren en beoordelen van schrijfproducten (samenhang met de huidige situatie) en op de investering wat betreft tijd en moeilijkheid (investering ten opzichte van resultaat).

De ontwikkeling en optimalisering van een betrouwbaar, valide en praktisch bruikbaar analyse- en beoordelingsinstrument heeft plaats gevonden door een aantal fasen te doorlopen. Ten eerste is een eerste versie van het instrument ontwikkeld aan de hand van een literatuurstudie en met behulp van deskundigen. Daarna is een pilottest uitgevoerd om de eerste versie van het instrument te testen. Aan de hand van de resultaten van de pilottest is een tweede versie van het instrument ontwikkeld en zijn er een aantal discussiepunten opgesteld. Ten slotte is er een eindtest uitgevoerd om de tweede versie van het instrument te testen. Aan de hand van de resultaten van de eindtest zijn aanbevelingen geformuleerd met het oog op de toekomst. Er is een derde versie van het instrument in ontwikkeling waarin de aanbevelingen worden verwerkt. Alle volledige versies van de instrumenten zijn op te vragen bij de auteur.

Ontwikkeling versie 1 van het instrument

Procedure

Er is een eerste versie van het instrument ontworpen aan de hand van een literatuurstudie en door middel van kennisdelen met deskundigen. De inhoud van het instrument is meerdere malen besproken met deskundigen en daarna bijgesteld, totdat het instrument valide bevonden werd door de onderzoeker en de deskundigen.

Respondenten

Er zijn drie deskundigen uit de projectgroep 'Taal bij de zaakvakken en de potentie van ICT' geraadpleegd. De drie deskundigen zijn ieder deskundig op een ander gebied, er is één deskundige op het gebied van science (een natuurkundige), één deskundige op het gebied van taal (een taalkundige) en één deskundige op het gebied van primair onderwijs (een leerkracht). Daarnaast zijn nog twee externe deskundigen geraadpleegd, namelijk een taalkundige en een expert op het gebied van toetsing.

Instrumenten

De meest recente uitwerking van het instrument vormde telkens de basis aan de hand waarvan het gesprek met de deskundigen plaats heeft gevonden. Er zijn probleemgebieden geformuleerd bij het instrument om de deskundigen een leidraad te geven voor het geven van de feedback. Er zijn echter geen specifieke vragen gesteld om de blik op het instrument zo breed mogelijk te houden, zodat alle facetten van het instrument belicht zouden worden en niet alleen de facetten waarbij de onderzoeker moeilijkheden ervoer. Zodoende is er steeds een brede feedback verkregen.

Data-analyse

Het instrument is telkens gereviseerd aan de hand van de verkregen informatie van de respondenten. Niet alle feedback is meegenomen in de revisie. Alleen de feedback die door meerdere deskundigen werd gegeven en waar de onderzoeker het mee eens was is gebruikt om het instrument aan te passen. Na diverse reviseringen is versie 1 van het instrument ontstaan.

Pilottest / ontwikkeling versie 2 van het instrument

Procedure

De pilottest bestond uit twee delen. Het ene deel is uitgevoerd om de praktische bruikbaarheid van het instrument te testen, waarbij leerkrachten de respondenten vormden. Het andere deel is uitgevoerd om de betrouwbaarheid van het instrument te testen, waarbij deskundigen de respondenten vormden.

De praktische bruikbaarheid is getest door vier leerkrachten versie 1 van het instrument te laten gebruiken. Versie 1 bevat alle kenmerken die in het theoretisch kader zijn genoemd, ze zijn echter iets anders genoemd en beschreven. Ze zijn na overleg met deskundigen verwerkt in het instrument als de kenmerken inhoud science, inhoud talig, structuur, communicatie en conventies. Deze kenmerken zijn in het instrument tevens onderverdeeld in subkenmerken.

Alle leerkrachten hebben hetzelfde proevenboekje in twee gedeelten geanalyseerd en beoordeeld. Het eerste gedeelte omvatte bladzijde 3, 4, 6 en 7 van het proevenboekje. Dit waren de bladzijden die de leerlingen zelf hadden ingevuld tot en met les 2. Het tweede gedeelte omvatte bladzijde 9 en 10 van het proevenboekje. Dit waren de bladzijden die de leerlingen zelf hadden ingevuld tot en met les 4. Voor deze onderverdeling is gekozen, zodat halverwege en aan het eind van het proevenboekje naar het product gekeken zou worden en alleen naar die gedeelten, die zelfstandig zijn ingevuld door de leerlingen.

De leerkrachten moesten het proevenboekje analyseren en beoordelen op een aantal (sub)kenmerken die betrekking hadden op het tussenproduct, namelijk een subkenmerk van inhoud science (gegeven science begrippen), het kenmerk inhoud talig en een subkenmerk van structuur (innerlijke structuur). Naast het proevenboekje hebben alle leerkrachten ook dezelfde proef uit het grote proevenboek geanalyseerd en beoordeeld. Ze hebben deze proef op alle kenmerken geanalyseerd en beoordeeld.

De leerkrachten hebben het instrument onafhankelijk van elkaar geëvalueerd waarbij de onderzoeker aanwezig was. Tijdens het gebruik van het instrument is door middel van de 'think-aloud'

procedure (hardop denken) nagegaan hoe bruikbaar het instrument was. Daarnaast is na het gebruik van het instrument een gesprek gehouden over de praktische bruikbaarheid van het instrument.

De betrouwbaarheid is getest door vijf deskundigen uit de projectgroep versie 1 van het instrument te laten evalueren. Alle deskundigen hebben twee dezelfde proevenboekjes geanalyseerd en beoordeeld, waarvan één proevenboekje hetzelfde betrof als die in het andere gedeelte van de pilottest. De deskundigen hebben beide proevenboekjes in dezelfde twee gedeeltes geanalyseerd en beoordeeld als de leerkrachten in het andere deel van de pilottest en tevens op dezelfde (sub)kenmerken. Daarnaast hebben alle deskundigen dezelfde proef uit het grote proevenboek geanalyseerd en beoordeeld. Dit betrof dezelfde proef als in het andere gedeelte van de pilottest. Ze hebben deze proef ook op alle kenmerken geanalyseerd en beoordeeld.

De deskundigen hebben tegelijkertijd de pilottest uitgevoerd. Na het gebruik van het instrument is een groepsdiscussie gehouden over verschillen in de analyse en beoordeling.

Bij het selecteren van de twee proevenboekjes voor de pilottest was de leesbaarheid een belangrijk criterium. Bij het selecteren van de proef was volledigheid van het schrijfproduct een belangrijk criterium.

Respondenten

Aan het deel om de praktische bruikbaarheid van het instrument te bepalen hebben vier leerkrachten meegewerkt die werkzaam zijn, of zijn geweest, in de midden- en/of bovenbouw van het primair onderwijs. Aan het deel om de betrouwbaarheid te bepalen hebben vijf deskundigen uit de projectgroep 'Taal bij de zaakvakken en de potentie van ICT' meegewerkt. Onder deze deskundigen bevonden zich drie onderwijkskundigen, een taalkundige en een natuurkundige.

Instrumenten

Versie 1 van het instrument is geëvalueerd bij de pilottest door de respondenten.

Om het hardop denken ('think-aloud') van de leerkrachten tijdens het gebruik van het instrument te bevorderen zijn een aantal hulpvragen opgesteld (Box 1).

Na het gebruik van versie 1 van het instrument is tevens een gesprek met de leerkrachten gevoerd over de praktische bruikbaarheid. Daarvoor zijn een aantal vragen opgesteld (Box 2).

Om de groepsdiscussie met de deskundigen te leiden zijn de gegeven scores als basis gebruikt. Per schrijfproduct zijn de gegeven scores behandeld en vooral de verschillen tussen de scores. Ook hiervoor zijn hulpvragen opgesteld (Box 3).

Box 1. Hulpvragen 'think-aloud' procedure leerkrachten

- Waarom vind je het lastig om deze score te geven?
 - Snap je wat je moet doen?
 - Past het product in meerdere hokjes?
 - Is er onduidelijkheid over de beschrijvingen in de rubric?

Box 2. Vragen gesprek leerkrachten

- Was het duidelijk hoe je het instrument moest gebruiken?
- Sluit het instrument aan op de manier waarop je nu schrijfproducten analyseert in de klas? Zo nee, vind je het erg als het anders is?
- Het kost ongeveer 15 minuten per leerling voor de gehele lessenserie van vijf lessen om op deze manier te analyseren en beoordelen, vind je dat haalbaar, zou je dat er voor over hebben?
- Zou je een soortgelijk instrument gebruiken in je eigen klas, zie je er het nut van in?

Box 3. Hulpvragen groepsdiscussie deskundigen

- Waarom heb je deze score gegeven?
 - Hoe interpreteer je de beschrijving?
 - Waar heb je specifiek op gelet, waarop lag bij jou de nadruk?
 - Heb je getwijfeld bij het geven van de score?
- Hoe komt het denken jullie dat jullie scores verschillen?
 - Ligt het aan verschil in interpretatie?
 - Ligt het eraan waar een ieder de nadruk op heeft gelegd?
 - Ligt het aan twijfel bij het geven van de score?
 - Was het subkenmerk überhaupt moeilijk te scoren voor dit schrijfproduct?

Data-analyse

Versie 1 van het instrument is gereviseerd aan de hand van de verkregen resultaten. Hier is versie 2 van het analyse- en beoordelingsinstrument uit voortgekomen. Daarnaast zijn aan de hand van de verkregen resultaten discussiepunten opgesteld over het instrument. Over deze punten was nog teveel twijfel om ze al te veranderen in het instrument. Deze discussiepunten zijn aan de orde gesteld tijdens de discussie bij de eindtest.

Eindtest

Procedure

Drie leerkrachten hebben versie 2 van het analyse- en beoordelingsinstrument gebruikt. Iedere leerkracht heeft een bladzijde uit hetzelfde proevenboekje geanalyseerd en beoordeeld met behulp van het instrument. Iedere leerkracht heeft tevens dezelfde proef uit het grote proevenboek geanalyseerd en beoordeeld. Er is gebruik gemaakt van het proevenboekje en de grote proef, die ook zijn geanalyseerd en beoordeeld bij de pilottest door zowel de leerkrachten als de deskundigen. Na het gebruik van het instrument is een discussie gehouden over de betrouwbaarheid en de bruikbaarheid van het instrument.

Respondenten

Aan de eindtest hebben een drietal leerkrachten meegewerkt, die interesse hebben om lessenserie 3 uit te gaan voeren op hun school.

Instrumenten

Versie 2 van het analyse- en beoordelingsinstrument is geëvalueerd bij de eindtest door de respondenten. In een groepsdiscussie zijn een aantal discussiepunten aan de orde gesteld (Box 4).

Box 4. *Discussiepunten eindtest*

1. Tijd versus nut. Hoe kan het instrument het best worden ingezet?
2. Science begrippen apart analyseren?
3. Conventies analyseren bij tussenproducten?
4. Vaktaal en algemene taal meer uitwerken?
5. Moet een leerling altijd op het hoogste niveau gescoord kunnen worden?
6. Termen als goed, redelijk, veel en weinig gebruiken?

Data-analyse

De praktische bruikbaarheid en betrouwbaarheid zijn kwalitatief beschreven aan de hand van de informatie verkregen bij de groepsdiscussie. Aan de hand hiervan zijn aanbevelingen gedaan met betrekking tot het analyse- en beoordelingsinstrument en het gebruik ervan met het oog op de toekomst. Er is een derde versie van het instrument in ontwikkeling waarin de aanbevelingen worden verwerkt.

Resultaat

Ontwikkeling versie 1 van het instrument

Uit de literatuurstudie is gebleken dat er een bepaalde methode kan worden gebruikt, de holistische of de analytische methode. De analytische methode is meer betrouwbaar en heeft een grotere diagnostische functie, daarom heeft deze methode de voorkeur en is er een analytisch analyse- en beoordelingsinstrument ontworpen.

Uit de literatuurstudie bleek tevens dat een analyse- en beoordelingsinstrument verschillende vormen kan aannemen, dit kan bijvoorbeeld een beoordelingsschaal, rubric of checklist zijn. Voor een praktisch bruikbaar instrument moet het niet veel tijd kosten om het instrument te gebruiken voor de leerkracht, maar het instrument moet wel goede richtlijnen geven. Daarom is in dit onderzoek gekozen om het analyse- en beoordelingsinstrument de vorm van een rubric te geven, omdat deze gedetailleerdere richtlijnen geeft dan de beoordelingsschaal, maar minder omvangrijk is dan de

checklist. Er is gekozen om de rubric een onderverdeling te geven in 4 niveaus, waarbij niveau 4 het hoogste niveau is.

Aan de hand van de vertaling van de kerndoelen en de onderzoeksliteratuur is gebleken dat bij het analyseren en beoordelen van schrijfproducten kan worden gelet op verschillende kenmerken, namelijk inhoud, organisatie, communicatie, stijl en conventies. Deze kenmerken zijn daarom meegenomen bij het ontwerpen van een instrument om schrijfproducten te analyseren en beoordelen om te komen tot een zo groot mogelijke validiteit. De uitwerking van de kenmerken is echter aangepast aan de nieuw ontwikkelde lessenseries rondom het thema 'Het Weer'. Dit is gedaan in samenwerking met deskundigen. Een deel van versie 1 van het analyse- en beoordelingsinstrument is zichtbaar in Figuur 1, in de figuur is het gedeelte met niveau 2 en 3 echter weggelaten.

Figuur 1. Deel van versie 1 van het analyse- en beoordelingsinstrument

Kenmerken	Niveau 1	Niveau 2	Niveau 3	Niveau 4
Inhoud science <ul style="list-style-type: none"> ▪ Gegeven science begrippen ▪ Redeneren 	<ul style="list-style-type: none"> ▪ Er worden bijna geen gegeven science begrippen (vaktaal: bv. lucht, druk, kracht) genoemd. ▪ Proeven en verschijnselen worden niet op een correcte wijze beschreven. 			<ul style="list-style-type: none"> ▪ Bijna alle gegeven science begrippen (vaktaal: bv. lucht, druk, kracht) worden genoemd. ▪ Proeven en verschijnselen worden op een correcte wijze beschreven, geïnterpreteerd en tevens correct verklaard door middel van het leggen van relaties tussen de science begrippen.
Inhoud talig <ul style="list-style-type: none"> ▪ Woord- en zinsniveau 	<ul style="list-style-type: none"> ▪ De algemene taal is niet of verkeerd gebruikt. 			<ul style="list-style-type: none"> ▪ Er is goed gebruik gemaakt van de algemene taal in combinatie met de science begrippen.
Structuur <ul style="list-style-type: none"> ▪ Innerlijke structuur ▪ Uiterlijke structuur 	<ul style="list-style-type: none"> ▪ De redeneerlijn ontbreekt. Er is geen gebruik gemaakt van denktalwoorden (signaalwoorden). De tekst springt van de hak op de tak. ▪ Het is één lange tekst, zonder structuur. De tekst heeft hooguit een titel. 			<ul style="list-style-type: none"> ▪ De redeneerlijn is helder weergegeven met behulp van de juiste denktalwoorden (signaalwoorden). ▪ De tekst heeft een titel, bestaat uit verschillende alinea's, de alinea's hebben kopjes en er is gebruik gemaakt van lay-out kenmerken ter verduidelijking (bv. vetgedrukt, cursief, onderstreept, illustraties enz.).
Communicatie <ul style="list-style-type: none"> ▪ Doelgerichtheid ▪ Publiekgerichtheid 	<ul style="list-style-type: none"> ▪ De informatie is niet overgebracht door de schrijver. ▪ De informatie is niet aangepast aan de behoeften, attitudes en voorkennis van het publiek. 			<ul style="list-style-type: none"> ▪ De informatie is goed overgebracht door de schrijver. ▪ De informatie is goed aangepast aan de behoeften, attitudes en voorkennis van het publiek.
Conventies <ul style="list-style-type: none"> ▪ Grammatica ▪ Interpunctie ▪ Spelling 	<ul style="list-style-type: none"> ▪ Er zijn vrijwel geen zinnen correct geformuleerd op woord- en zinsniveau. ▪ Er zijn geen leestekens en hoofdletters gebruikt. (Hooguit aan het begin en eind van de hele tekst.) ▪ In de meest gangbare categorieën zijn veel spelfouten gemaakt. 			<ul style="list-style-type: none"> ▪ Bijna alle zinnen zijn correct geformuleerd op woord- en zinsniveau. ▪ De zinnen in de tekst zijn correct van elkaar gescheiden met de juiste leestekens. Alle zinnen beginnen met een hoofdletter. ▪ In de meest gangbare categorieën zijn bijna geen spelfouten gemaakt.

Het gevonden kenmerk inhoud uit de literatuur is in versie 1 van het analyse- en beoordelingsinstrument het kenmerk inhoud science geworden. Dit kenmerk omvat het subkenmerk gegeven science begrippen, waarbij het gaat om het gebruik van de gegeven science begrippen. Daarnaast omvat dit kenmerk het subkenmerk redeneren, waarbij het gaat om het correct beschrijven, interpreteren en verklaren van verschijnselen en proeven.

Het kenmerk stijl uit de literatuur is in versie 1 van het instrument het kenmerk inhoud talig geworden. Het omvat het subkenmerk woord- en zinsniveau, waarbij het gaat om het goed gebruiken van algemene taal in combinatie met de science begrippen.

Het kenmerk organisatie uit de literatuur is het kenmerk structuur geworden in versie 1 van het instrument, waarbij een onderverdeling is gemaakt in de subkenmerken innerlijke structuur en uiterlijke structuur. Bij innerlijke structuur gaat het om de talige redeneerlijn (goede volgorde van inhoud) en het gebruik van signaalwoorden om deze redeneerlijn te ondersteunen. Bij uiterlijke structuur gaat het om het aanbrengen van uiterlijke structuurkenmerken zoals titel, alinea's, kopjes en andere lay-out kenmerken.

Het kenmerk communicatie uit de literatuur heeft dezelfde term gehouden en bevat de subkenmerken doelgerichtheid en publiekgerichtheid. Bij doelgerichtheid gaat het erom dat de

informatie correct is overgebracht en bij publiekgerichtheid gaat het erom dat de tekst is aangepast aan de behoeften, attitudes en voorkennis van het publiek.

Ook het kenmerk conventies uit de literatuur heeft dezelfde term gehouden en bevat de subkenmerken grammatica, interpunctie en spelling. Bij grammatica gaat het om het correct formuleren van zinnen op woord- en zinsniveau. Bij interpunctie is het van belang dat er leestekens en hoofdletters worden gebruikt en bij spelling gaat het om een goede spelling van de gebruikte woorden.

Bij het analyseren en beoordelen van een tussenproduct gaat het vooral om het proces. De subkenmerken gegeven science begrippen, woord- en zinsniveau en innerlijke structuur zijn daarbij van belang. Deze zijn daarom grijs gearceerd in het instrument. Het subkenmerk redeneren is nog niet goed te scoren, omdat leerlingen bij het tussenproduct pas op het laatst hoeven te verklaren, dus dit kenmerk is nog niet van toepassing. Het subkenmerk uiterlijke structuur en de kenmerken communicatie en conventies zijn vooral bij het eindproduct belangrijk.

Bij het analyseren en beoordelen van een eindproduct zijn alle kenmerken (inhoud science, inhoud talig, structuur, communicatie en conventies) van belang.

Pilottest

Gegeven scores

In Tabel 1a en 1b zijn de gegeven scores per subkenmerk weergegeven van proevenboekje 1 van de twee gedeelten van het boekje die beoordeeld zijn. Alleen alle deskundigen hebben dit boekje beoordeeld en daarom is het aantal beoordelaars 5. Uit deze gegevens valt op te maken dat de deskundigen vrij eensgezind scoren. Alle scores behalve die op het subkenmerk innerlijke structuur in Tabel 1b verschillen niet meer dan één niveau.

In Tabel 2a en 2b zijn de gegeven scores per subkenmerk weergegeven van proevenboekje 2 van de twee gedeelten van het boekje die beoordeeld zijn. Zowel de deskundigen als de leerkrachten hebben dit boekje beoordeeld, er was echter bij beide gedeelten één deskundige die een deel van het proevenboekje niet heeft beoordeeld vanwege tijdgebrek en daarom is het aantal beoordelaars 8. Uit deze gegevens valt op te maken dat de deskundigen en leerkrachten niet allemaal hetzelfde scoren, maar de scores nog wel redelijk dicht bij elkaar ligt. De scores verschillen niet meer dan twee niveaus.

In Tabel 3 zijn de gegeven scores per subkenmerk weergegeven van de proef uit het grote proevenboek. Zowel de deskundigen als de leerkrachten hebben deze proef beoordeeld. Twee deskundigen hebben echter alleen op de subkenmerken beoordeeld waarop de tussenproducten beoordeeld moesten worden en niet op alle kenmerken. Deze beoordelaars waren niet volledig in hun beoordeling en zijn daarom buiten beschouwing gelaten. Het aantal beoordelaars komt daardoor op 7. Uit deze gegevens valt op te maken dat de deskundigen en de leerkrachten op de proef vrij eensgezind scoren. Behalve publiekgerichtheid en gegeven science begrippen verschillen de scores niet meer dan één niveau. Daarbij is er bij publiekgerichtheid en gegeven science begrippen maar één beoordelaar die afwijkt van de meerderheid en dit is in beide gevallen dezelfde beoordelaar.

Versie 1 van het instrument is redelijk betrouwbaar, want de respondenten beoordelen vrij eensgezind. Om het instrument meer betrouwbaar te maken moet worden gekeken naar de verklaringen voor de verschillen in gegeven scores.

Tabel 1a. *Gegeven scores bij proevenboekje 1 bladzijde 3, 4, 6 en 7 (N=5)*

	niveau 1	niveau 2	niveau 3	niveau 4
Gegeven science begrippen	3	2		
Inhoud talig		4	1	
Innerlijke structuur		1	4	

Tabel 1b. *Gegeven scores bij proevenboekje 1 bladzijde 9 en 10 (N=5)*

	niveau 1	niveau 2	niveau 3	niveau 4
Gegeven science begrippen	2	3		
Inhoud talig	1	4		
Innerlijke structuur	2	2	1	

Tabel 2a. Gegeven scores bij proevenboekje 2 bladzijde 3, 4, 6 en 7 (N=8)

	niveau 1	niveau 2	niveau 3	niveau 4
Gegeven science begrippen	2	3	3	
Inhoud talig		5	1	2
Innerlijke structuur		1	3	4

Tabel 2b. Gegeven scores bij proevenboekje 2 bladzijde 9 en 10 (N=8)

	niveau 1	niveau 2	niveau 3	niveau 4
Gegeven science begrippen		2	4	2
Inhoud talig		1	6	1
Innerlijke structuur			4	4

Tabel 3. Gegeven scores bij proef uit het grote proevenboek (N=7)

	niveau 1	niveau 2	niveau 3	niveau 4
Gegeven science begrippen		3	3	1
Redeneren	1	6		
Inhoud talig		3	4	
Innerlijke structuur			4	3
Uiterlijke structuur			4	3
Doelgerichtheid			5	2
Publiekgerichtheid		1	2	4
Grammatica			3	4
Interpunctie			1	6
Spelling			1	6

Verklaringen voor verschillen in gegeven scores

Uit de 'think-aloud' procedure en gesprekken met de leerkrachten en de groepsdiscussie bij de deskundigen blijkt dat verschillen in scores goed te verklaren zijn.

Het verschil in scores bij het subkenmerk gegeven science begrippen bij zowel het tussenproduct als het eindproduct (Tabel 1a, 1b, 2a, 2b en 3) is te verklaren doordat niet geheel duidelijk was welke woorden nou precies onder gegeven science begrippen vallen. Het verschil in scores bij het subkenmerk inhoud talig bij zowel het tussenproduct als het eindproduct (Tabel 1a, 1b, 2a, 2b en 3) is te verklaren doordat niet geheel duidelijk was wat algemene taal is. Er moet daarom duidelijker worden gemaakt wat vaktaal en algemene taal is.

Het verschil in scores bij het subkenmerk innerlijke structuur bij het tussenproduct (Tabel 1a, 1b, 2a en 2b) is te verklaren doordat de structuur in de proevenboekjes al deels gegeven was door de templates. Respondenten vonden het daarom lastig om dit subkenmerk te scoren. Het is daarom beter om het subkenmerk innerlijke structuur alleen bij het eindproduct te analyseren.

De reden waarom één respondent bij het subkenmerk redeneren in zijn score voor het eindproduct afweek van de andere respondenten (Tabel 3), was dat hij vond dat er een aantal stappen ontbraken in de redenering. Er zijn meer respondenten die dit hebben geconstateerd, maar zij hebben dit mee laten wegen in hun score voor publiekgerichtheid in plaats van in hun score voor redeneren.

Het verschil in scores bij het subkenmerk uiterlijke structuur bij het eindproduct (Tabel 3) zit hem in een interpretatieverschil van lay-out kenmerken. De één vindt een paar van deze kenmerken voldoende, de ander wil er meer zien.

Het verschil in scores bij de kenmerken communicatie en conventies bij het eindproduct (Tabel 3) komt door begrippen als goed, redelijk, weinig, veel enz. Dit geeft wat ruimte voor eigen interpretatie van de beoordelaar.

Daarnaast werd duidelijk dat het bij de tussenproducten lastig is om een aantal bladzijden samen te analyseren. Dit was ook soms de oorzaak voor verschillen in scores bij de tussenproducten (Tabel 1a, 1b, 2a en 2b). Wanneer het tussenproduct tussentijds geanalyseerd dient te worden kan er beter worden gekozen voor het analyseren van alleen de laatste bladzijde van een onderdeel.

Praktische bruikbaarheid instrument

Het is duidelijk hoe je het instrument moet gebruiken. Het instrument sluit echter niet aan op de manier waarop leerkrachten normaal analyseren en beoordelen. Meestal analyseren en beoordelen leerkrachten niet op alle kenmerken en geven ze een meer globale beoordeling. De leerkrachten vinden het echter geen probleem dat de manier van analyseren en beoordelen afwijkt van de huidige manier. Ze vinden het juist wel goed om eens op een andere manier naar schrijfproducten te kijken.

Verder vinden leerkrachten en deskundigen het instrument nuttig, maar ze geven ook aan dat het teveel tijd kost om zowel tussenproduct als eindproduct te analyseren en beoordelen. Er werden verschillende oplossingen aangedragen om de tijdsinvestering te verminderen. Een oplossing die door twee leerkrachten en een aantal deskundigen werd gegeven is om alleen het eindproduct te analyseren en beoordelen met behulp van het instrument en de tussenproducten meer globaal te analyseren. Een andere oplossing die werd gegeven is om niet meer per subkenmerk te scoren maar per kenmerk een niveau aan te geven. Het instrument zou hiervoor wat gesimplificeerd moeten worden. Een laatste oplossing was om niet altijd alle schrijfproducten te analyseren en beoordelen.

Wel vinden alle respondenten het best moeilijk om het instrument te gebruiken. De deskundigen die hun gegeven scores in een groepsdiscussie hebben besproken geven aan dat het heel goed en fijn is om er samen over te praten en te discussiëren, zo kun je een meer doordachte score geven.

Inhoud instrument

De deskundigen vonden het gevaarlijk om je bij de tussenproducten wat betreft inhoud science enkel te richten op de science begrippen. Natuurwetenschappelijk redeneren is belangrijker dan het precieze gebruik van de vaktaal. Daarom moet bij de tussenproducten ook het redeneren worden geanalyseerd.

Ook gaf men aan dat er duidelijk moet worden aangegeven in de uitleg van het instrument (of in de handleiding van het lesmateriaal) dat het logisch is dat de kinderen bij de tussenproducten nog op niveau 1 of 2 zitten. Anders denken leerkrachten dat hun leerlingen het allemaal 'slecht' doen. Pas gaandeweg de lessenseries moet het niveau hoger worden.

Verder behoeft versie 1 van het instrument nog enige aanpassing, zodat het toepasbaar is op de producten van alle lessenseries rond het thema 'Het Weer'.

Veranderingen instrument

De pilottest heeft een aantal veranderepunten opgeleverd voor versie 1 van het instrument. Deze worden beschreven bij de ontwikkeling van versie 2 van het instrument. Daarnaast heeft de pilottest een aantal discussiepunten opgeleverd voor de eindtest. Over deze punten bestaat onduidelijkheid, meer helderheid hierover is gewenst, voordat er aanpassingen aan het instrument plaats vinden. Deze discussiepunten zijn genoemd op bladzijde 8 bij de instrumenten voor de eindtest (Box 4).

Ontwikkeling versie 2 van het instrument

De pilottest heeft enkele veranderingen opgeleverd voor versie 1 van het instrument, waardoor versie 2 van het instrument is ontstaan. Een deel van versie 2 van het analyse- en beoordelingsinstrument is zichtbaar in Figuur 2, in de figuur is het gedeelte met niveau 2 en 3 echter weggelaten.

Ten eerste is in versie 2 veranderd waar het tussenproduct op beoordeeld moet worden. Het tussenproduct moet op beide subkenmerken van inhoud science worden geanalyseerd en niet meer op het subkenmerk innerlijke structuur. De grijze arcering is hierop aangepast. Zowel redeneren als het gebruik van science begrippen moet worden geanalyseerd bij het tussenproduct, omdat het niet alleen gaat om het gebruiken van science begrippen, maar vooral om de juiste natuurkundige redenering. Het subkenmerk redeneren is daarom bovenaan komen te staan bij het kenmerk inhoud science, omdat dit belangrijker is dan het gebruiken van science begrippen. Het tussenproduct hoeft niet te worden beoordeeld op innerlijke structuur, omdat de innerlijke structuur al gegeven wordt in de templates van het curriculummateriaal.

Daarnaast is in versie 2 van het instrument de tekst bij het subkenmerk science begrippen gewijzigd. Het woord 'gegeven' is weggehaald, omdat door dit woord de vrijheid van de leerlingen

wordt ingeperkt bij het gebruiken van science begrippen. Het gaat erom dat ze relevante science begrippen gebruiken, dit hoeven niet per se de gegeven science begrippen te zijn.

Verder is versie 2 van het instrument hetzelfde als versie 1 van het instrument. Er zijn niet meer wijzigingen aangebracht, omdat er over sommige punten nog teveel twijfel bestond om het al te wijzigen. Deze punten van twijfel zijn als discussiepunten ingebracht bij de eindtest.

Figuur 2. Deel van versie 2 van het analyse- en beoordelingsinstrument.

Kenmerken	Niveau 1	Niveau 2	Niveau 3	Niveau 4
Inhoud science ▪ Redeneren ▪ Science begrippen	<ul style="list-style-type: none"> ▪ Proeven en verschijnselen worden niet op een correcte wijze beschreven. ▪ Bijna geen science begrippen, die bij het onderwerp van de lessenserie horen, worden genoemd. 			<ul style="list-style-type: none"> ▪ Proeven en verschijnselen worden op een correcte wijze beschreven, geïnterpreteerd en tevens correct verklaard door middel van het leggen van relaties tussen de science begrippen. ▪ Bijna alle science begrippen, die bij het onderwerp van de lessenserie horen, worden genoemd.
Inhoud talig ▪ Woord- en zinsniveau	<ul style="list-style-type: none"> ▪ De algemene taal is niet of verkeerd gebruikt. 			<ul style="list-style-type: none"> ▪ Er is goed gebruik gemaakt van de algemene taal in combinatie met de science begrippen.
Structuur ▪ Innerlijke structuur ▪ Uiterlijke structuur	<ul style="list-style-type: none"> ▪ De redeneerlijn ontbreekt. Er is geen gebruik gemaakt van denktaalwoorden (signaalwoorden). De tekst springt van de hak op de tak. ▪ Het is één lange tekst, zonder structuur. De tekst heeft hooguit een titel. 			<ul style="list-style-type: none"> ▪ De redeneerlijn is helder weergegeven met behulp van de juiste denktaalwoorden (signaalwoorden). ▪ De tekst heeft een titel, bestaat uit verschillende alinea's, de alinea's hebben kopjes en er is gebruik gemaakt van lay-out kenmerken ter verduidelijking (bv. vetgedrukt, cursief, onderstreept, illustraties enz.).
Communicatie ▪ Doelgerichtheid ▪ Publiekgerichtheid	<ul style="list-style-type: none"> ▪ De informatie is niet overgebracht door de schrijver. ▪ De informatie is niet aangepast aan de behoeften, attitudes en voorkennis van het publiek. 			<ul style="list-style-type: none"> ▪ De informatie is goed overgebracht door de schrijver. ▪ De informatie is goed aangepast aan de behoeften, attitudes en voorkennis van het publiek.
Conventies ▪ Grammatica ▪ Interpunctie ▪ Spelling	<ul style="list-style-type: none"> ▪ Er zijn vrijwel geen zinnen correct geformuleerd op woord- en zinsniveau. ▪ Er zijn geen leestekens en hoofdletters gebruikt. (Hooguit aan het begin en eind van de hele tekst.) ▪ In de meest gangbare categorieën zijn veel spel fouten gemaakt. 			<ul style="list-style-type: none"> ▪ Bijna alle zinnen zijn correct geformuleerd op woord- en zinsniveau. ▪ De zinnen in de tekst zijn correct van elkaar gescheiden met de juiste leestekens. Alle zinnen beginnen met een hoofdletter. ▪ In de meest gangbare categorieën zijn bijna geen spel fouten gemaakt.

Eindtest

Bij de eindtest lag de focus op de zes discussiepunten die opgesteld zijn aan de hand van de resultaten van de pilottest (Box 4, blz. 8). Wat betreft het eerste discussiepunt, het gebruik van het instrument, bleek uit de groepsdiscussie dat het instrument het best alleen voor het eindproduct gebruikt kan worden, anders kost het teveel tijd. Het is wel goed om de tussenproducten ook te analyseren, maar deze kunnen ook algemeen beoordeeld worden aan de hand van aanwijzingen in het curriculummateriaal. Daarnaast vond men het het beste om het instrument te gebruiken ten behoeve van formatieve analyse, zodat er feedback aan de beoordeling wordt gekoppeld waarmee de leerling verder kan. Er zijn drie lessenseries en iedere lessenserie heeft een eindproduct. Aan de hand van de analyse van het eindproduct kan er feedback worden gegeven aan een leerling met het oog op de volgende lessenserie.

Uit de discussie over het tweede discussiepunt bleek dat men het erg goed vond om de science begrippen apart te analyseren. Deze science begrippen vormen een kapstok voor een goede natuurkundige redenering en moeten daarom mee worden genomen in de analyse en beoordeling.

Wat betreft het derde discussiepunt, het analyseren van conventies bij de tussenproducten, vond men het nog niet nodig om dit te doen. De conventies zijn nog niet zo belangrijk bij het tussenproduct, deze worden pas belangrijk bij het eindproduct.

Over het discussiepunt van de vaktaal en algemene taal was men het eens dat de vaktaal en algemene taal meer moeten worden uitgewerkt in het lesmateriaal. Er mogen meer woorden worden gegeven die in de instructie aan de orde kunnen komen. Deze woorden zijn voor de leerkrachten een steun bij het aanbieden van de juiste begrippen en ze helpen leerlingen om to the point te formuleren.

Daarnaast bleek uit de discussie over discussiepunt vijf, dat het bij het eindproduct bezwaarlijk is als je een leerling sowieso niet op het hoogste niveau kunt scoren, vanwege de gegeven opdracht. Dit is bijvoorbeeld het geval bij het analyseren van het subkenmerk redeneren bij de proef uit het grote proevenboek. De leerlingen hoeven wat betreft het subkenmerk redeneren nog niets van de hogere niveaus te laten zien, dus het geeft een vertekend beeld wanneer een leerling voor dit eindproduct op niveau 2 gescoord wordt, terwijl hij de opdracht wel prima heeft uitgevoerd. Er kan dan beter voor de betreffende schrijfproducten "n.v.t." worden aangegeven bij het betreffende subkenmerk.

Wat betreft het zesde discussiepunt stoorden de respondenten zich niet aan termen als goed, redelijk, veel of weinig. Dit zorgt volgens hen voor interpretatievrijheid als het gaat om het scoren van een leerling uit groep 5/6 ten opzichte van een leerling uit groep 7/8. Het instrument is zo flexibel inzetbaar voor meerdere groepen. De leerkracht moet naar eigen inzicht wat anders scoren bij groep 5/6 dan bij groep 7/8.

Tijdens de discussie kwamen nog een tweetal andere opmerkingen naar voren. Een groot nadeel van het instrument is volgens de respondenten dat leerlingen op het kenmerk inhoud talig niet hoger dan niveau 1 of 2 gescoord kunnen worden als ze weinig science begrippen gebruiken. Dit moet veranderd worden, want het gaat bij dit kenmerk niet alleen om het gebruik van science begrippen. Verder werd aangegeven dat er ook op andere manieren naar leerresultaten kan worden gekeken, dan door middel van het analyseren en beoordelen van schrijfproducten, bijvoorbeeld door middel van observeren. Dit is iets om rekening mee te houden bij verder onderzoek.

Conclusie en aanbevelingen

De onderzoeksvraag waarop een antwoord is gezocht luidt:

Hoe ziet een betrouwbaar, valide en praktisch bruikbaar instrument eruit, dat geschikt is om de schrijfproducten van basisschoolleerlingen, geproduceerd tijdens de nieuw ontwikkelde lessenseries rond het thema 'Het Weer', te analyseren en beoordelen op het gebied van taalontwikkeling en op het gebied van de ontwikkeling van science concepten?

Uit de resultaten blijkt dat een goed begin is gemaakt met de ontwikkeling van een instrument om de taalontwikkeling en ontwikkeling van science concepten te meten. Versie 2 van het instrument is gebaseerd op een gedegen literatuurstudie en is valide bevonden door verschillende deskundigen. Daarnaast is het instrument vrij betrouwbaar, omdat respondenten redelijk eensgezind scoren. Met de aanpassingen die worden gegeven in de aanbevelingen, die worden verwerkt in versie 3, is het instrument waarschijnlijk nog meer betrouwbaar. Verder is het instrument om drie redenen praktisch bruikbaar. Ten eerste, omdat het duidelijk is voor leerkrachten hoe ze het instrument moeten gebruiken in de praktijk, dus de toepasbaarheid is groot. Ten tweede, omdat de leerkrachten het niet erg vinden dat de manier van analyseren en beoordelen niet geheel aansluit bij de huidige manier, ze vinden de nieuwe manier juist goed. En ten slotte, omdat de tijdsinvestering lijkt op te wegen tegen het resultaat, wanneer je het instrument zou gebruiken voor de eindproducten van de lessenseries.

Versie 2 van het analyse- en beoordelingsinstrument is met enkele aanpassingen die worden verwerkt in versie 3 van het instrument, dus redelijk valide, betrouwbaar en praktisch bruikbaar binnen de kaders van dit onderzoek. Het instrument zou echter op grotere schaal toegepast moeten worden om te kunnen concluderen dat het instrument valide, betrouwbaar en praktisch bruikbaar is voor een grote groep mensen.

Aan de hand van het onderzoek zijn een aantal aanbevelingen opgesteld op het gebied van het bevorderen van de praktische bruikbaarheid van het instrument en het bevorderen van de betrouwbaarheid en validiteit van het instrument.

Bevorderen praktische bruikbaarheid

Gebruik instrument:

- Het instrument alleen voor de eindproducten van de lessenseries gebruiken, anders kost het teveel tijd.
- Het instrument gebruiken ten behoeve van formatieve analyse. Er zijn drie lessenseries en iedere lessenserie heeft een eindproduct. Je kunt aan de hand van de analyse van het eindproduct

feedback geven aan een leerling, zodat deze weet waarop hij moet letten bij de volgende lessenserie.

- Het instrument moet flexibel worden ingezet, dan is het geschikt voor zowel groep 5/6 als groep 7/8. De leerkracht moet naar eigen inzicht wat anders scoren bij groep 5/6 dan bij groep 7/8. De verwachtingen voor groep 5/6 mogen lager liggen dan bij groep 7/8.
- Het is goed om het instrument samen met collegae te gebruiken. Het is prettig om te bespreken welke scores anderen zouden geven en om er samen over te discussiëren. Zo ontstaat een eenduidig interpretatiekader en leert men het instrument op dezelfde manier te gebruiken. Zodoende kunnen meer doordachte scores worden gegeven.
- Bij de handleiding van het instrument zou het goed zijn om ook voorbeeldproducten te geven die gescoord zijn met behulp van het instrument, dit om het makkelijker te maken voor de gebruiker om te leren scoren met het instrument.

Bevorderen betrouwbaarheid en validiteit

Veranderingen aanbrengen in versie 2 van het analyse- en beoordelingsinstrument

- De tekst bij het kenmerk inhoud talig moet worden aangepast. Op dit moment kunnen leerlingen talig niet hoger scoren dan niveau 1 of 2 als ze geen science begrippen hebben gebruikt en dat is niet goed. De combinatie met science begrippen moet eruit.
- Wanneer het hoogste niveau helemaal niet bereikt hoeft te worden bij het subkenmerk redeneren, moet er "n.v.t." komen te staan bij dit subkenmerk. Dit is het geval bij het eindproduct van lessenserie 1.
- Het instrument moet alleen gebruikt worden voor de eindproducten, dus de grijze arcering voor de kenmerken waarop de tussenproducten beoordeeld zouden moeten worden kan weggehaald worden.

Veranderingen aanbrengen in het curriculummateriaal

- Tussenproducten moeten algemeen beoordeeld kunnen worden aan de hand van aanwijzingen in het curriculummateriaal. Het proces is erg belangrijk, daarom moet vooral gelet worden op inhoud science en inhoud talig, de rest is minder belangrijk.
- De vaktaal en algemene taal moeten meer worden uitgewerkt in het lesmateriaal. Er mogen meer woorden worden gegeven die in de instructie aan de orde kunnen komen.

Door dit onderzoek is veel bekend geworden over een geschikt analyse- en beoordelingsinstrument voor schrijfproducten van de lessenseries rond het thema 'Het Weer'. Er is meer duidelijkheid over hoe een dergelijk instrument er uit kan zien en hoe het gebruikt kan worden in de praktijk. Wanneer de gegeven aanbevelingen worden uitgevoerd ligt er een goede versie van het analyse- en beoordelingsinstrument om verder onderzoek mee te doen. Er is daarom een derde versie van het instrument in ontwikkeling waarin de aanbevelingen worden verwerkt (deze is op te vragen bij de auteur). Verder onderzoek is nodig om de betrouwbaarheid, validiteit en praktische bruikbaarheid van het instrument te verbeteren en de generaliseerbaarheid van de resultaten te vergroten.

Discussie

Om de onderzoeksvraag te kunnen beantwoorden is een ontwerpgericht onderzoek uitgevoerd. Dit ontwerpgerichte onderzoek richtte zich op het ontwikkelen en optimaliseren van een analyse- en beoordelingsinstrument voor schrijfproducten door middel van evaluatie. Het betrof een explorierend onderzoek, daar er nog niet eerder onderzoek was gedaan op dit gebied binnen het project.

Het onderzoek was gericht op het nieuw ontwikkelde curriculummateriaal. De vraag rijst daarmee of het instrument ook op andere schrijfproducten, dan die van het nieuw ontwikkelde curriculummateriaal, toepasbaar is. Het zou interessant zijn om verder onderzoek te doen om de vertaalslag te kunnen maken naar andere schrijfproducten.

Binnen het onderzoek is één manier onderzocht om zicht te krijgen op leerresultaten van leerlingen. Er is onderzocht hoe schrijfproducten geanalyseerd en beoordeeld kunnen worden met behulp van een instrument. Het zou een toevoeging zijn om ook andere manieren te onderzoeken die gebruikt kunnen worden om naar de leerresultaten van leerlingen te kijken, bijvoorbeeld observeren.

Dit kan een breder beeld verschaffen van de taalontwikkeling en de ontwikkeling van science concepten bij leerlingen.

Bij de tests van het instrument is gekozen voor een relatief klein aantal respondenten om diepgaande informatie te verkrijgen. Door bij de pilottest de 'think-aloud' procedure te gebruiken bij de leerkrachten en de groepsdiscussie bij de deskundigen kon diep op het instrument worden ingegaan. Ook bij de eindtest was er een relatief klein aantal respondenten waarmee de discussie is gevoerd, dit leverde ook een diepgaande discussie op. De gekozen onderzoeksaanpak heeft dus diepgaande informatie opgeleverd, maar is door het kleine aantal respondenten ook minder generaliseerbaar. Het is interessant om het instrument door een grotere groep respondenten te laten gebruiken om te kijken wat voor resultaten dit oplevert.

Het punt van tijdsinvestering ten opzichte van resultaat blijft een discussiepunt. Ook Doyle & Ponder (1977-78) geven aan dat dit lastig is. De verhouding tussen de investering die een leerkracht moet doen en het verwachte resultaat van het analyse- en beoordelingsinstrument moet in evenwicht zijn volgens hen. Het evenwicht tussen tijdsinvestering en resultaat kan bij iedere leerkracht echter anders liggen. In het begin kost het meer tijd om het instrument te gebruiken, dus dat is een nadeel. Zullen leerkrachten daardoor niet snel afhaken? In hoeverre moet je het instrument vereenvoudigen of minder laten gebruiken vanwege tijdgebrek? Als onderzoeker ben ik geneigd te zeggen dat het instrument niet verder vereenvoudigd dient te worden, want dat gaat ten koste van de validiteit. Verder ben ik als onderzoeker geneigd te zeggen dat het instrument zo vaak mogelijk gebruikt moet worden, dit is optimaal voor het geven van feedback aan de leerlingen. Als leerkracht ben ik echter geneigd te zeggen dat het instrument niet te ingewikkeld moet zijn en niet te vaak gebruikt kan worden, vanwege de tijdsinvestering die het met zich mee brengt. Het blijft constant zoeken naar een evenwicht. In mijn onderzoek ben ik er naar mijn idee redelijk in geslaagd dit evenwicht te vinden door aan te bevelen het instrument alleen te gebruiken bij het eindproduct, maar dan wel als formatieve evaluatie met het oog op de volgende lessenserie. Natuurlijk zie je in het kader van de ontwikkeling van de leerlingen het liefst dat het instrument zo vaak mogelijk gebruikt wordt, maar het moet ook haalbaar zijn en daarom moeten er concessies worden gedaan.

Daarnaast blijven termen als goed, redelijk, veel of weinig in een instrument discutabel. Het geeft ruimte aan de leerkracht om het instrument flexibel te gebruiken voor meerdere groepen, maar tegelijkertijd kan deze interpretatieruimte ook een nadeel vormen. Het hangt van de interpretatie van de gebruiker af hoe er gescoord wordt. Wiggins (1998) ziet termen als goed, redelijk, veel of weinig niet graag in een rubric. Hij pleit voor omschrijvingen die de specifieke kenmerken van ieder niveau weergeven. Toch zie je in dit onderzoek dat verschillende beoordelaars redelijk op één lijn zitten, dus de interpretatieruimte is niet erg groot. Dit komt overeen met hetgeen Wesdorp (1978) stelt, namelijk dat er minder kans is op effecten van beoordelaarvariabiliteit wanneer een analytische methode wordt gebruikt. Naast de vraag of de gekozen termen in het instrument wel of niet goed zijn, gaat het er ook om dat leerkrachten nadenken over alle kenmerken die belangrijk zijn bij het analyseren en beoordelen van een schrijfproduct en er het liefst ook met elkaar over praten en discussiëren. Er kan een grote winst behaald worden als de leerkrachten zich meer bewust worden van alle kenmerken waarop een schrijfproduct geanalyseerd en beoordeeld kan worden. Hopelijk wordt deze winst behaald door middel van dit onderzoek en eventueel verder onderzoek.

Referenties

- Doyle, W., & Ponder, G. A. (1977-78). The practicality ethic in teacher decision-making [Elektronische versie]. *Interchange*, 8 (3), 1-11.
- Freeman, G., & Taylor, V. (2006). *Integrating science and literacy instruction: A framework for bridging the gap*. Lanham, MD: Rowman & Littlefield Publishers.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends [Elektronische versie]. *Review of Educational Research*, 60 (2), 237-263.
- Krom, R., Van de Gein, J., Van der Hoeven, J., Van der Schoot, F., Verhelst, N., Veldhuijzen, N., & Hemker, B. (2004). *Balans van het schrijfonderwijs op de basisschool. Uitkomsten van de peilingen in 1999: halverwege en einde basisonderwijs en speciaal basisonderwijs*. PPON-reeks 28. Arnhem, NL: Stichting Cito Instituut voor Toetsontwikkeling.
- Meredith, V. H., & Williams, P. L. (1984). Issues in direct writing assessment: Problem identification and control. *Educational Measurement: Issues and Practice*, 3, 11-15, 35-36.
- Miller, R. G., & Calfee, R. C. (2004). Building a better reading-writing assessment: bridging cognitive theory, instruction and assessment [Elektronische versie]. *English Leadership Quarterly*, 26 (3), 6-13.
- Paus, H. (Red.). (2006). *Portaal: Praktische taaldidactiek voor het primair onderwijs*. Bussum, NL: Coutinho.
- Roid, G. H. (1994). Patterns of writing skills derived from cluster analysis of direct-writing assessments. *Applied Measurement in Education*, 7 (2), 159-170.
- Smits, M., & Voogt, J. (2006). Natuuronderwijs en schrijfonderwijs met ICT geïntegreerd. *Nieuwsbrief Taal voor Opleiders en Begeleiders*, 4 (2), 22-29.
- Van de Gein, J. (2004). *Balans van taalkwaliteit in schrijfwerk uit het primair onderwijs. Uitkomsten van de peilingen in 1999*. PPON-reeks 29. Arnhem, NL: Stichting Cito Instituut voor Toetsontwikkeling.
- Wesdorp, H. (1978). *Evaluatie-technieken voor het moedertaalonderwijs*. Amsterdam: Stichting Research Instituut voor de Toegepaste Psychologie aan de Universiteit van Amsterdam.
- Wiggins, G. (1998). *Educative assessment. Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.