

Is de MILK een betrouwbaar instrument voor het meten van de kwaliteit van leermiddelen?

Naam en studentnummer: Janneke de Jong, 3136884

Naam begeleider: Dr. Gijsbert Erkens

Naam tweede beoordelaar: Dr. Bert Versloot

Kwalitatief goede leermiddelen zijn onmisbaar voor het onderwijs. Het CLU heeft een basisinstrument met stellingen ontwikkeld om de kwaliteit van leermiddelen te meten, genaamd de MILK. Dit instrument bevat tien verschillende leerfuncties van leermiddelen, waarvan er drie zijn gelegitimeerd. De drie gelegitimeerde functies bestaan ieder uit vijf of zes onderwerpen, welke weer bestaan uit verschillende items. In dit onderzoek is gekeken in hoeverre de gelegitimeerde leerfuncties van de MILK betrouwbaar en valide zijn. Om de betrouwbaarheid te meten is aan zes middelbare school docenten en vijf onderwijskundige experts gevraagd om vijf leermiddelen te beoordelen aan de hand van de MILK. De deelnemers hebben eerst uitleg gekregen over de MILK waarna ze zelfstandig de MILK hebben ingevuld. Uit de analyse van de beoordelingscores is gebleken dat de MILK niet bij alle leerfuncties en onderwerpen even betrouwbaar is. Als er echter wordt gekeken naar de MILK in het geheel dan is de betrouwbaarheid van de MILK redelijk goed, maar de betrouwbaarheid van de verschillende leerfuncties en onderwerpen kan en moet echter wel vergroot worden. Het is lastig om de onderzoeksresultaten te generaliseren vanwege het lage aantal beoordelaars en leermiddelen.

Probleemstelling en onderzoeksvragen

Probleemstelling

Definitie leermiddelen

Leermiddelen zijn onmisbaar voor het onderwijs. Leermiddelen in bijvoorbeeld de vorm van tekstboeken, werkboeken en films spelen een belangrijke rol in het onderwijsproces, omdat ze informatie geven die de basis is voor het leren in een onderwijssetting of een aanvulling op de onderwijssetting (Saracho, 1985). Vaak wordt de onderwijssetting aangepast aan het gebruikte leermiddel, maar wat voor middelen zijn eigenlijk leermiddelen? Volgens Min (2000) kan materiaal in het onderwijs verdeeld worden in vier verschillende soorten materiaal. Het eerste type onderwijsmateriaal is het instructiemiddel, waarbij de vakinhoud door auteurs en/of docenten wordt aangeboden en zijn sturend van aard. Voorbeelden van instructiemiddelen zijn instructieboeken en instructiedvd's. Een tweede type onderwijsmateriaal is materiaal waarbij de leerling het initiatief moet nemen, ofwel leerlinggestuurd materiaal. Dit materiaal bevat minder vakinhoud dan een instructiemiddel, omdat de leerlingen gestimuleerd worden om zelf aan de slag te gaan. Hierbij valt bijvoorbeeld te denken aan een computersimulatie. Volgens Min (2000) zijn er ook mengvormen van leerlinggestuurde leermiddelen en instructiemiddelen mogelijk. Hierbij valt te denken aan een leerboek met concrete opdrachten, het internet of een interactieve cd-rom. Een derde type materiaal in het onderwijs zijn communicatiemiddelen, zoals internetfora. Als laatste noemt Min (2000) hulpmiddelen. Dit zijn middelen als een tekstverwerker en een rekenmachine. Een belangrijk kenmerk van communicatie- en hulpmiddelen is dat ze geen leerinhoud van zichzelf hebben. Onderwijsmateriaal als communicatie- en hulpmiddelen zijn daarom geen leermiddelen. Volgens Reints (2008) kan onderwijsmateriaal pas leermiddelen genoemd worden als ze het leren faciliteren. Daarnaast geeft van de Grift (1989) aan dat leermiddelen bepaalde vaardigheden, houdingen en prestaties bij de leerlingen moeten oproepen. Het is daarbij belangrijk om het doel van het leerproces in de gaten te houden. Leermiddelen kunnen dan als volgt gedefinieerd worden:

“Leermiddelen zijn teksten en/of beelden die bedoeld zijn om het leren te faciliteren, in de zin dat de leerinhouden geselecteerd en geordend zijn met het oog op te bereiken leerdoelen, en die aanzetten tot het uitvoeren van leeractiviteiten en van activiteiten die het leerproces reguleren” (Reints, 2008, p. 30).

In deze definitie wordt duidelijk dat leermiddelen verschillende functies kunnen vervullen. Leerfuncties zijn functies van leermiddelen die noodzakelijk zijn om leren te faciliteren (Reints & Wilkens, 2010). De volgende drie functies van leermiddelen zijn uit de definitie van Reints (2008) te halen:

1. het aanbieden van leerinhouden. Het is van belang dat leerinhouden op een goede manier worden geselecteerd en geordend, zodat leren gefaciliteerd wordt. Dit wordt gedaan met het oog op de te bereiken leerdoelen;
2. het aanzetten van cognitief, affectief en/of psychomotorisch handelen, en het reguleren van het leerproces. Dit is noodzakelijk om leren plaats te laten vinden;
3. het representeren van inhoud en handelen in (digitale) teksten en beelden. De inhoud moet op een manier gepresenteerd worden die het leren optimaal wordt bevordert (Reints & Wilkens, 2010).

De functies die voortkomen uit de beschreven definitie zijn geschikt voor digitale leermiddelen en folio leermiddelen, zoals lesboeken. Onderwijsmateriaal moet deze functies vervullen voordat het een leermiddel genoemd mag worden (Reints & Wilkens, 2010). Pennings (2006) heeft gekeken naar een definitie voor alleen digitaal leermateriaal. Er is namelijk veel digitaal materiaal te vinden dat niet direct voor educatieve doelen is ontwikkeld, maar wel in het onderwijs gebruikt wordt. Kan men dit digitaal leermateriaal noemen? Digitaal materiaal is volgens Pennings (2006) pas leermateriaal als de informatie, die wordt aangeboden via het internet of andere technologieën, ook voor het onderwijs is bedoeld. Hierbij kan men denken aan les- en toetsmateriaal dat via het internet wordt aangeboden. Als de informatie niet voor het onderwijs is bedoeld, dan is er geen sprake van leermateriaal. Dit materiaal kan wel in het onderwijs gebruikt worden als ondersteuning. Leerlingen kunnen namelijk gemotiveerd worden door materiaal dat digitaal wordt aangeboden, aangezien leerlingen in het dagelijks leven thuis veel te maken hebben met de computer. Leermiddelen hebben echter niet alleen betrekking op leerlingen, maar ook andere groepen hebben bepaalde belangen bij leermiddelen. Voorbeelden van deze groepen zijn docenten, onderwijskundige experts en uitgevers. Docenten willen bijvoorbeeld dat leermiddelen aansluiten bij de planning en evaluatie van het onderwijs, terwijl bij uitgeverijen commercie een belangrijke rol speelt (Reints, 2000). Leermiddelen worden vaak gemaakt met het oog op de leerlingen, maar volgens Grossman en Thompson (2008) moeten leermiddelen ten eerste de docenten aanspreken. Als een leermiddel de docenten niet goed informeert en stimuleert, dan krijgen de leerlingen niet met het materiaal te maken. Het gebruik van een leermiddel is dus afhankelijk van de mening van de docenten over dat leermiddel. Het kan daarom belangrijk zijn om het doel van het leermiddel duidelijk te vermelden in de tekst.

Evalueren van leermiddelen

Leermiddelen in het onderwijs moeten geëvalueerd kunnen worden op kwaliteit. Veel docenten vinden dat een kwalitatief goed leermiddel van belang is voor goed onderwijs,

omdat in veel scholen leermiddelen de kern zijn waar leeractiviteiten om heen worden gemaakt (Lester, Cheek & Earl, 1998). Met behulp van een evaluatie kunnen gebruikers duidelijk krijgen wat de sterke en zwakke punten zijn van een product (Saracho, 1985). Bij het ontwerpen van leermiddelen is het daarom goed om een evaluatie van een leermiddel niet uit te stellen tot het laatste moment, maar om in de eerste fase van ontwerpen al te evalueren (Hoobroeckx & Haak, 2002). De evaluatie van leermiddelen tijdens het ontwerpproces wordt ook wel formatieve evaluatie genoemd. Volgens Brakenhoff en Homminga (1995) heeft een formatieve evaluatie het doel om het didactisch handelen of het didactisch materiaal zo veel mogelijk bij te sturen en te verbeteren. Volgens Medley-Mark en Weston (1988) gaat het hierbij voornamelijk om het verbeteren van aspecten in leermiddelen die voor problemen kunnen zorgen. Naast formatieve evaluaties zijn er ook summatieve evaluaties. Bij een summatieve evaluatie wordt het 'definitieve' leermiddel beoordeeld op effectiviteit (Krus, Thurlow, Turnure & Taylor, 1976). Evaluaties van leermiddelen worden meestal uitgevoerd door educatieve auteurs en docenten. In de vorige paragraaf is duidelijk geworden dat leermiddelen door verschillende doelgroepen gebruikt worden, zo ook door leerlingen. Leerlingen evalueren echter bijna nooit leermiddelen. Als ze leermiddelen evalueren, dan evalueren ze deze leermiddelen anders dan docenten dat doen. Leerlingen bekijken leermiddelen namelijk met een ander perspectief dan docenten (Price, 2007).

Over het algemeen wordt er in het onderwijs weinig aandacht gegeven aan het evalueren van de kwaliteit van leermiddelen. Dit kan komen doordat er geen goed evaluatie-instrument is om de kwaliteit van leermiddelen te beoordelen (Reints, 2000). Er zijn echter ook andere redenen om de kwaliteit van leermiddelen niet te evalueren:

- het is moeilijk om een instrument te maken dat bruikbaar is voor alle leermiddelen in alle situaties (Reints, 2000);
- het is moeilijk om een instrument te maken dat de verschillende visies op leren en de rol van leren in acht neemt. Bij het beoordelen van leermiddelen kan de visie op leren een belangrijke rol spelen. (Reints, 2000; Sambrook, Geertshuis & Cheseldine, 2001);
- bij een eventuele negatieve evaluatie zijn er veel kosten (Hoobroeckx & Haak, 2002);
- er zijn verschillende factoren die het gebruik van leermiddelen kunnen beïnvloeden, zoals de kennis, opvattingen, vaardigheden en ervaring van de docent (de Vos, 1998).

Toch is het van belang dat docenten leermiddelen evalueren op kwaliteit, omdat docenten bij een evaluatie van een leermiddel kunnen aangeven op welke punten toekomstige leermiddelen verbeterd kunnen worden (Hoobroeckx & Haak, 2002; Krus et al., 1976). Docenten moeten een leermiddel niet accepteren als ze niet eerst hebben gekeken naar de

kwaliteit van een leermiddel. Naast het aangeven van verbeterpunten voor toekomstige leermiddelen is er nog een aantal redenen om de kwaliteit van leermiddelen te evalueren:

- de kwaliteit van leermiddelen heeft effect op kwaliteit onderwijs (Lester et al., 1998). De evaluatie van de kwaliteit van leermiddelen kan gedaan worden door docenten, maar ook door onderwijskundige experts die door scholen worden ingeschakeld;
- de leermiddelen moeten passen bij de leerdoelen van een docent of school. Door de kwaliteit van een leermiddel te evalueren kan een docent kijken of een leermiddel past bij de onderwijssetting (Krus et al., 1976). Als een onderwijskundig expert wordt ingeschakeld voor de evaluatie, dan moet de expert ook rekening houden met de leerdoelen van de school;
- met de komst van Wikiwijs worden docenten gestimuleerd om de mogelijkheden van ICT beter te benutten. Er wordt gratis en vrij toegankelijk digitaal lesmateriaal aangeboden dat docenten kunnen gebruiken en bewerken (Ministerie OC&W, 2009). De docenten krijgen met Wikiwijs de ruimte om de leermiddelen aan te passen aan hun eigen doel en het doel van de school. Aangezien er vele verschillende leermiddelen via Wikiwijs worden aangeboden is het van belang om de kwaliteit van deze leermiddelen te evalueren;
- er zijn (vooral) beginnende docenten die een leermiddel als 'expert' zien op een bepaald onderwerp (Lester et al., 1998). Leermiddelen bevatten volgens hen geen of weinig fouten. Daarnaast zien beginnende docenten een leermiddel als een handig hulpmiddel voor instructie (Grossman & Thompson, 2008). Deze docenten steunen daarom in grote mate op de inhoud van het leermiddel. Het is dan belangrijk dat de kwaliteit van het leermiddel goed is, ook al is geen leermiddel perfect (Grossman & Thompson, 2008). Docenten moeten dus kritisch zijn ten opzichte van leermiddelen. Een evaluatie-instrument zou hierbij kunnen helpen.

Evalueren met een instrument

Het maken van een evaluatie-instrument voor het beoordelen van de kwaliteit van leermiddelen is niet gemakkelijk, maar wel belangrijk. Daarom heeft het Centrum voor Leermiddelenstudie in Utrecht [CLU] het Meet Instrument Leermiddelen Kwaliteit ofwel de MILK ontwikkeld. Met het evaluatie-instrument de MILK kunnen verschillende soorten leermiddelen geëvalueerd worden, van lesboeken tot digitale leermiddelen. De MILK kan gebruikt worden door docenten die nieuwe leermiddelen in de klas willen gaan gebruiken en willen weten of het leermiddel dat ze willen gebruiken kwalitatief goed is. Daarnaast kan de MILK gebruikt worden door onderwijskundige experts. Experts kunnen door scholen worden ingeschakeld om een leermiddel te beoordelen,

omdat scholen zelf geen tijd hebben of omdat scholen weinig ervaring hebben met het beoordelen van leermiddelen.

De MILK is gemaakt aan de hand van de drie functies van leermiddelen die zijn beschreven onder *definitie leermiddelen*. Deze drie functies zijn omgezet in de volgende drie thema's: kwaliteit van de content; pedagogische benadering; en design en presentatie. Deze thema's kunnen uitgewerkt worden in verschillende leerfuncties. Leerfuncties zijn de functies van leermiddelen die het leerproces optimaliseren (Reints & Wilkens, 2010). De kwaliteit van een leermiddel kan geëvalueerd worden aan de hand van leerfuncties. Bij de MILK komen in totaal negen leerfuncties aan bod bij de drie verschillende thema's. Bij het eerste thema (kwaliteit van de content) gaat het om de leerfuncties *selectie van de content*, *ordening van de content* en *keuze van modaliteiten*. Bij het tweede thema (pedagogische benadering) gaat het om de leerfuncties *keuze van leeractiviteiten*, *keuze van de instructiestrategie* en *regulatie leerproces*. Bij het derde thema (design en presentatie) gaat het om leerfuncties als *begrijpelijke teksten*, *representatie beelden*, *aantrekkelijke vormgeving*. Daarnaast wordt er, buiten de drie thema's om, gekeken naar het gebruiksgemak van het leermiddel. Bij gebruiksgemak komen items aan bod als: 'Het is niet nodig om horizontaal te scrollen'. Het is bij dit onderdeel dan ook de vraag of het leermiddel gemakkelijk in gebruik is (Reints & Wilkens, 2010).

De verschillende leerfuncties zijn opgedeeld in onderwerpen, welke weer bestaan uit verschillende items waarmee de kwaliteit van leermiddelen beoordeeld kan worden. De items, in de vorm van stellingen, worden beoordeeld aan de hand van beoordelingsscores. Naast het geven van beoordelingsscores aan verschillende items kunnen deelnemers aan verschillende onderwerpen binnen een leerfunctie een wegingsfactor geven. De leerfunctie *keuze van modaliteiten* bestaat bijvoorbeeld onder andere uit de onderwerpen teksten, beelden, geluid en authentiek materiaal. Deze onderwerpen bevatten weer verschillende items. Met een wegingsfactor kan een deelnemer aangeven hoe belangrijk hij of zij een bepaald onderwerp vindt voor het beoordelen van de kwaliteit van leermiddelen. Een hoge wegingsfactor betekent dat een deelnemer een bepaald onderwerp, zoals teksten, belangrijk vindt voor de kwaliteit van een leermiddel.

De MILK is nog in ontwikkeling en de items die gemaakt zijn bij de verschillende leerfuncties moeten nog worden aangepast. Er zijn tot nu toe drie leerfuncties gelegitimeerd aan de hand van literatuur. Per leerfunctie is gekeken naar passende (wetenschappelijke) literatuur, waarna de items en de onderwerpen van de leerfuncties in de MILK aan de hand van de literatuur zijn aangepast. De items komen dus voort uit een literatuurstudie. Dit wordt ook wel legitimeren genoemd en bij de MILK zijn tot nu toe drie leerfuncties gelegitimeerd. Deze drie leerfuncties zijn: *keuze*

van modaliteiten, keuze van de instructiestrategie en begrijpelijke teksten. Uit ieder thema is dus één leerfunctie gelegitimeerd. De andere leerfuncties moeten nog gelegitimeerd worden aan de hand van wetenschappelijke literatuur.

Al is de MILK nog in ontwikkeling, het is wel belangrijk dat de betrouwbaarheid en de validiteit van de MILK getoetst wordt. Steeds meer onderwijskundige experts en docenten krijgen namelijk te maken met de MILK en om de MILK te kunnen gebruiken moet het wel een betrouwbaar instrument zijn. Er is echter nog niet uitgebreid gekeken naar de betrouwbaarheid van de gelegitimeerde leerfuncties MILK. Daarom wordt er in dit onderzoeksartikel aandacht gegeven aan de betrouwbaarheid van gelegitimeerde leerfuncties van de MILK, maar voordat de onderzoeksvragen aan bod komen wordt er eerst kort ingegaan op de drie verschillende gelegitimeerde leerfuncties uit de drie verschillende thema's.

Kwaliteit van content

Keuze van modaliteiten

De eerste leerfunctie die hier wordt besproken is *keuze van modaliteiten*. Leerstof in leermiddelen kan op verschillende manieren aangeboden ofwel 'verpakt' worden. Dit wordt ook wel het representatieformat genoemd (de Jong et al., 1998). In leermiddelen worden de verschillende manieren van aanbieden ook wel modaliteiten genoemd. Leerstof in leermiddelen kan via de volgende modaliteiten aangeboden worden: teksten, beelden en geluid. Daarnaast kan leerstof aangeboden worden via authentiek materiaal. Leerstof in teksten en beelden komen de hersenen van de leerling binnen door de ogen, het visuele kanaal. Terwijl leerstof die aangeboden wordt via de modaliteit geluid binnenkomt via de oren, ofwel het auditieve kanaal (Lujan & DiCarlo, 2006). Bij authentiek materiaal kan de leerstof via verschillende zintuigen opgenomen worden (neus, huid, mond, ogen en oren), dit is verschillend voor ieder object. De genoemde modaliteiten zijn weer onder te verdelen in verschillende soorten tekst, beeld, geluid en authentiek materiaal.

Een leermiddel kan alleen teksten, alleen beelden, alleen geluid of alleen authentiek materiaal bevatten. Een leermiddel kan echter ook verschillende modaliteiten bevatten. Door het combineren van modaliteiten, bijvoorbeeld beeld en tekst, kan diep en betekenisvol leren bij leerlingen plaatsvinden (Mayer, 2003). Het gebruik maken van meerdere modaliteiten wordt ook wel multimodaliteit genoemd. Volgens Mayer (2003) zijn er een aantal principes die belangrijk zijn bij het gebruik van multimedia en multimodaliteit in een leermiddel, namelijk: het multimediaprincipe (het gebruik van meer modaliteiten is beter dan het gebruik van één modaliteit), het modaliteitsprincipe ('tekst en geluid' of 'beeld en geluid' is beter dan 'tekst en beeld'), het coherentieprincipe (toevoegen van irrelevant materiaal kan leiden tot slechtere resultaten), ruimtelijke- en

tijd nabijheidprincipe (woorden, beelden en geluiden moeten zo dicht mogelijk bij elkaar of na elkaar gepresenteerd worden), personalisatieprincipe (leerlingen worden gestimuleerd door de leerstof in een informele stijl te presenteren in plaats van een formele stijl) en het redundantieprincipe (maak geen gebruik van meerdere modaliteiten als de leerstof met één of twee modaliteiten al duidelijk genoeg is).

Naast de verschillende modaliteiten en het gebruik van meerdere modaliteiten moet er bij deze leerfunctie ook gekeken worden naar de aansluiting van de modaliteiten bij de doelgroep en het leerdoel. Verschillende doelgroepen kunnen namelijk op andere manieren gemotiveerd worden en het is belangrijk om de leerstof in betekenisvolle modaliteiten aan de doelgroep te presenteren (Jonassen, Howland, Marra & Crismond, 2008). Een doelgroep kan verschillende kenmerken hebben waarop leermiddelen aan kunnen sluiten. Men kan bijvoorbeeld denken aan groepsgrootte, algemene kennis over modaliteiten en de handicaps van de doelgroep (van Merriënboer & Kirschner, 2007). Volgens Price (2007) is het voornamelijk belangrijk om te kijken naar de ervaringen van de betreffende doelgroep.

Pedagogische benadering

Keuze van de instructiestrategie

De tweede leerfunctie welke wordt gebruikt in dit onderzoek is *keuze van de instructiestrategie*. Volgens Valcke (2005) zijn instructiestrategieën concrete handelingen waardoor het leren van leerlingen wordt uitgelokt. Deze handelingen kunnen komen van een docent, maar kunnen ook gestimuleerd worden in een leermiddel. Volgens Reints (2008) moet een instructie het leergedrag van leerlingen beïnvloeden en faciliteren, zodat het leerdoel bereikt kan worden. Het leerdoel kan bereikt worden door tijdens de instructie de benodigde kennis, informatie en vaardigheden aan te bieden. Hiermee wordt het leerdoel niet direct bereikt, want de leerling zal eerst de benodigde kennis en vaardigheden eigen moeten maken (Marzano, 1998). Leerlingen kunnen vaardigheden eigen maken door middel van opdrachten.

Het is in een onderwijssituatie belangrijk dat een leermiddel leerlingen ondersteunt bij het verwerven van kennis en vaardigheden. Om een goede ondersteuning te bieden aan leerlingen zijn instructiestrategieën nodig. Volgens van Gagné (1970) zijn er zeven instructiestrategieën te onderscheiden voor leermiddelen, namelijk: voorkennis activeren, de aandacht van de lerende richten, de leerstof inoefenen, de leerstof verwerken, de lerende motiveren, feedback geven en reflectie. Deze zeven instructiestrategieën komen allen terug bij de leerfunctie *keuze van de instructiestrategie* in de MILK. Instructiestrategieën kunnen worden uitgevoerd door middel van verschillende instructietactieken (Leshin, Pollock & Reigeluth, 1992). Bij het onderwerp

aandacht van de lerende richten binnen deze leerfunctie kan men bijvoorbeeld denken aan de instructietactiek: ‘De lesdoelen zijn van te voren gegeven’.

Design en presentatie

Begrijpelijke teksten

De laatste leerfunctie welke in dit onderzoek gebruikt wordt is de leerfunctie *begrijpelijke teksten*. Volgens Klare (2000) is leesbaarheid het gemak waarmee een tekst te begrijpen is door de manier van schrijven. In deze definitie ligt de focus voornamelijk op de schrijfstijl. Het gaat bij leesbaarheid om het begrijpen van woorden en zinnen. Een veelomvattende definitie van leesbaarheid is de definitie van Dale en Chall (1949): Alle elementen die terugkomen in leermiddelen beïnvloeden over het algemeen het succes dat de gebruikers ermee hebben. Het succes is dat gebruikers kunnen aantonen dat ze het begrijpen, dat ze de tekst met optimale snelheid kunnen lezen en het interessant vinden.

Het gaat bij deze definitie dus om het begrip van de tekst, met een optimale snelheid de tekst lezen en als laatste is de interesse van de lezer belangrijk. Volgens Versnel (2009) zijn er uit de drie punten van de definitie van Dale en Chall (1949) verschillende aspecten te halen. De verschillende aspecten zijn: vocabulaire (bekende of onbekende woorden), zinsstructuur (verbindingswoorden, signaalwoorden, etc.), samenhang (relaties tussen zinnen), kwaliteit en verve (aansprekende stijl), tekststructuur (koppen en introducties), lay-out en motivatie. Volgens Versnel (2009) is het belangrijk om op deze zeven punten te letten als men kijkt naar de leesbaarheid van een leermiddel.

Onderzoeksvragen

Met dit onderzoek wordt er getracht antwoord te geven op de volgende onderzoeksvragen:

“In hoeverre zijn de gelegitimeerde leerfuncties *keuze van modaliteiten*, *keuze van de instructiestrategie* en *begrijpelijke teksten*, van het door het CLU ontwikkelde evaluatie-instrument de MILK, valide en betrouwbare onderdelen van het instrument?”

Om antwoord te kunnen geven op deze onderzoeksvraag, zijn de volgende deelvragen opgesteld:

- In hoeverre is er sprake van consistentiebetrouwbaarheid bij het gebruik van de gelegitimeerde leerfuncties van de MILK?
- In hoeverre is er sprake van een interbeoordelaarsbetrouwbaarheid bij het gebruik van de gelegitimeerde leerfuncties van de MILK?
- In hoeverre is er een verband tussen de gelegitimeerde leerfuncties van de MILK?

- In hoeverre is er een verschil tussen de beoordelingen van de verschillende leerfuncties bij het wel of niet gebruik maken van wegingsfactoren bij de gelegitimeerde onderdelen van de MILK?
- Is er een verschil in de interbeoordelaarsbetrouwbaarheid van de gelegitimeerde onderdelen van de MILK tussen onderwijskundige experts en docenten?

Methode

Opzet van het onderzoek

Dit onderzoek naar de betrouwbaarheid van de MILK is een voorbeeld van een formatief evaluatief onderzoek. Met dit onderzoek is namelijk geprobeerd om de waarde van het instrument de MILK vast te stellen (Stokking, 1998). De MILK is nog in ontwikkeling, wat betekent dat de uitslagen van dit onderzoek gebruikt kunnen worden voor het bijstellen van het instrument (Stokking, 1998). Door formatief evaluatief onderzoek kan het object van de evaluatie, de MILK, ook versterkt worden (Trochim, 2006).

De onderzoeksvragen zijn beantwoord door zes verschillende leermiddelen (drie leermiddelen van het vak Wiskunde en drie leermiddelen van het vak Nederlands) elk door drie docenten en twee experts te laten beoordelen aan de hand van de drie gelegitimeerde leerfuncties van de MILK. De docenten en experts kregen, voordat ze de leermiddelen gingen beoordelen, uitleg over de MILK. Na deze uitleg konden de docenten en experts zelf aan de slag met het beoordelen van de MILK voor twee of drie verschillende leermiddelen. De deelnemers kregen een kopie van één hoofdstuk van het originele leermiddel om te beoordelen.

Deelnemers

Om de betrouwbaarheid van de gelegitimeerde onderdelen van de MILK te toetsen moest de MILK door verschillende deelnemers voor meerdere leermiddelen ingevuld worden. Om de betrouwbaarheid van de MILK te kunnen beoordelen is de MILK door vijf onderwijskundige experts ingevuld. Onderwijskundige experts zijn mensen die zich in het dagelijks leven bezig houden met het onderwerp 'leermiddelen' en hebben een universitaire en onderwijskundige achtergrond. De vijf experts hebben ieder twee of drie leermiddelen beoordeeld. Om antwoord te kunnen geven op de deelvraag over de interbeoordelaarsbetrouwbaarheid werd ieder leermiddel, in totaal zes leermiddelen, door twee experts beoordeeld.

Daarnaast is de MILK door zes verschillende docenten ingevuld. Om antwoord te kunnen geven op de laatste deelvraag, verschil in interbeoordelaarsbetrouwbaarheid tussen docenten en experts, hebben de docenten en experts dezelfde leermiddelen beoordeeld. Er zijn dus zes leermiddelen door elf mensen beoordeeld. De zes docenten

kwamen uit twee verschillende vakgroepen, namelijk Wiskunde en Nederlands. Drie wiskundedocenten hebben ieder drie leermiddelen van het vak Wiskunde beoordeeld met behulp van de MILK. Alle drie de docenten hebben dezelfde drie leermiddelen beoordeeld aan de hand van de MILK. Één leermiddel die ze hebben beoordeeld gebruiken ze zelf in de klas. De andere twee leermiddelen waren onbekend voor de docent. Daarnaast hebben drie docenten Nederlands drie verschillende leermiddelen van het vak Nederlands beoordeeld met behulp van de MILK. Alle drie de docenten hebben dezelfde drie leermiddelen beoordeeld aan de hand van de MILK. Één leermiddel die ze hebben beoordeeld gebruiken ze zelf in de klas. De andere twee leermiddelen waren, net als bij de wiskundedocenten, onbekend voor de docent. In Tabel 1 is een kort overzicht te vinden over hoe de verdeling van deelnemers en leermiddelen is gemaakt.

Tabel 1. *Overzicht verdeling deelnemers en leermiddelen*

Aantal deelnemers	Indeling	vorm			
6 docenten	Elke docent gaat drie leermiddelen beoordelen. Één leermiddel waarmee de docent zelf werkt en twee ‘nieuwe leermiddelen’.	Docent	Leermiddel Wiskunde	Docent	Leermiddel Nederlands
		A	1	D	4
		B	2	E	5
		C	3	F	6
		Expert	Leermiddel	Expert	Leermiddel
6 experts	Elke expert gaat twee leermiddelen beoordelen (experts en docenten beoordelen dezelfde leermiddelen)	G	1 en 4	J	1 2 en 4
		H	2 en 5	K	3 5 en 6
		I	3 en 6		

Bij het verwerven van de beoordelingen van de experts is gebleken dat één leermiddel van het vak Nederlands niet hetzelfde door docenten en experts beoordeeld kon worden. De docenten hebben de bijbehorende cd-rom kunnen openen, terwijl de cd-rom bij de experts verlopen en daardoor niet te openen was. De experts misten dus een deel van het leermiddel. Hierdoor is er een verschil in betrouwbaarheid ontstaan en daarom is dit leermiddel weggelaten bij de analyse van de gegevens.

Instrumenten

In dit onderzoek is gebruik gemaakt van het instrument de MILK. De MILK is ontwikkeld door het CLU en is gemaakt om de kwaliteit van leermiddelen (digitaal en niet digitaal leermateriaal) te kunnen evalueren. Niet alleen docenten, maar ook experts kunnen gebruik maken van de MILK om de kwaliteit van een leermiddel te beoordelen.

Om de kwaliteit van leermiddelen goed te evalueren is MILK onderverdeeld in drie thema's. Deze verdeling is ontstaan uit de verschillende functies van leermiddelen, die bij de probleemstelling zijn besproken. Deze functies van leermiddelen zijn: het aanbieden van leerinhouden; het aanzetten tot cognitief, affectief en/of psychomotorisch handelen, het representeren van inhoud en handelen in (digitale) teksten en beelden. Uit deze functies zijn respectievelijk de thema's *kwaliteit van de content*, *pedagogische benadering* en *design en presentatie* ontstaan (Reints & Wilkens, 2010). De onderverdeling van thema's in leerfuncties is in de paragraaf *evalueren met een instrument* beschreven.

De verschillende leerfuncties die aan de orde komen in de MILK bestaan uit een variërend aantal onderwerpen welke weer onderverdeeld zijn in een variërend aantal items. De verdeling van de leerfuncties en onderwerpen is terug te vinden in Tabel 2. De leerfunctie *keuze van modaliteiten* bestaat uit de onderwerpen: tekst, beeld, geluid, authentiek materiaal, multimodaliteit en doelgroep/leerdoel. Een voorbeeld van een item binnen het onderwerp multimodaliteit is: 'De relatie tussen tekst en beeld is duidelijk voor de leerling'. De leerfunctie *keuze van de instructiestrategie* bestaat uit de onderwerpen: voorkennis activeren, aandacht van leerlingen wordt gericht, mogelijkheid tot oefenen, mogelijkheid voor stofverwerking, motivatie, feedback en reflectie. Als laatste bestaat de leerfunctie *begrijpelijke teksten* uit de volgende onderwerpen: vocabulaire, zinsstructuur, samenhang, kwaliteit en verve, tekststructuur, lay-out en motivatie. In de bijlage is één leerfunctie opgenomen om de onderverdeling van een leerfunctie duidelijk te maken.

De deelnemers van het onderzoek hebben antwoord gegeven op de verschillende items, van de gelegitimeerde leerfuncties, door middel van een beoordelingsscore. Schaal één betekent 'helemaal niet mee eens', schaal twee is 'een beetje mee eens', schaal drie is 'grotendeels mee eens', schaal vier is 'helemaal mee eens'. Soms was een item in een leermiddel niet te beoordelen of kwam het niet voor in een leermiddel. Daarom konden de beoordelaars ook kiezen voor de alternatieven 'niet te beoordelen' en 'niet van toepassing'. Deze twee alternatieven zijn bij de analyses als missing values opgegeven. Als een docent een item hoog scoort bij een leermiddel, dan betekent dit dat een leermiddel van goede kwaliteit is. Bij sommige items moesten de beoordelingsscores voor de analyses wel gehercodeerd worden.

Voordat de deelnemers een beoordeling konden geven aan de verschillende items moesten ze wegingsfactoren geven aan de items. Met een wegingsfactor kon een deelnemer aangeven hoe belangrijk volgens hem of haar een bepaald onderwerp binnen een leerfunctie is voor de kwaliteit van leermiddelen. Waarbij een hoge wegingsfactor betekende dat een deelnemer een item relatief belangrijk vindt. Per leerfunctie (in dit onderzoek drie) konden er 100 punten verdeeld worden.

Naast de MILK is er in dit onderzoek gebruik gemaakt van een korte vragenlijst om meer informatie te krijgen over de achtergrond van de deelnemers. De achtergrond van een deelnemer kan namelijk invloed hebben op de verschillen in de beoordeling van een leermiddel. In deze korte vragenlijst werden daarom vragen gesteld over de leeftijd, geslacht, het beroep en het aantal jaren werkervaring van de deelnemer. Aangezien het aantal beoordelaars per onderwerp gering was, was het lastig om met de verschillende factoren rekening te houden. Het was met name moeilijk om rekening te houden met de factoren leeftijd en aantal jaar werkervaring. In dit onderzoek is er daarom alleen gekeken naar de verschillen in beroep.

Procedure

Om de betrouwbaarheid van de MILK te kunnen meten is dit instrument voor verschillende leermiddelen ingevuld, zoals in de paragraaf over de deelnemers duidelijk is geworden. De leermiddelen van dit onderzoek zijn gekozen aan de hand van de docenten. Iedere docent werkt zelf met één van de leermiddelen. Er zijn voor dit onderzoek daarom docenten bij elkaar gezocht die werken met andere leermiddelen. Hierdoor hebben de docenten twee leermiddelen beoordeeld waarmee ze zelf niet werken en één leermiddel waarmee ze wel werken. De onderwijskundige experts werken met geen van de leermiddelen die ze moesten beoordelen.

Tabel 2. *Verdeling van leerfuncties en onderwerpen binnen de MILK*

Leerfunctie	Onderwerp
<i>Leerfunctie keuze van modaliteiten</i>	Teksten Beelden Geluid Authentiek materiaal Multimodaliteit Doelgroep/leerdoel
<i>Leerfunctie keuze van de instructiestrategie</i>	Voorkennis activeren Aandacht wordt gericht Oefening Stofverwerking Motivatie Feedback Reflectie
<i>Leerfunctie begrijpelijke teksten</i>	Vocabulaire Zinsstructuur Samenhang Kwaliteit en verve Tekststructuur Lay-out Motivatie

De verschillende deelnemers werden per e-mail benaderd of ze mee willen werken aan het onderzoek. Bij een bevestiging werd er een afspraak gemaakt voor een bijeenkomst. In deze bijeenkomst (eventueel met meerdere deelnemers tegelijkertijd) is uitgelegd wat de MILK precies is, uit welke leerfuncties de MILK bestaat, hoe men de MILK moet invullen en hoe de wegingsfactoren werken. Deze uitleg duurde maximaal dertig minuten. Na deze bijeenkomst konden docenten en experts zelf aan de slag met het invullen van de MILK. De docenten hoefden de MILK niet tijdens de bijeenkomst in te vullen, maar kregen de tijd buiten de bijeenkomst om de MILK voor drie leermiddelen in te vullen en terug te sturen. De onderwijskundige experts hebben de leermiddelen wel direct beoordeeld na de uitleg over de MILK. Voordat de deelnemers de MILK gingen invullen, moesten de deelnemers eerst de korte vragenlijst invullen voor de achtergrondgegevens.

Voor de deelnemers zou het veel tijd kosten om een geheel leermiddel te beoordelen en daarom hebben de deelnemers één hoofdstuk van een leermiddel beoordeeld in plaats van het gehele leermiddel. Bij een hoofdstuk van een leermiddel hoort niet alleen de leerstof in het lesboek en werkboek, maar ook eventueel aanvullend materiaal dat bij het hoofdstuk hoort. Deze verschillende aspecten (lesboek, werkboek en aanvullend materiaal) behoren tot hetzelfde leermiddel. Alle deelnemers moesten van elk leermiddel hetzelfde hoofdstuk beoordelen, zodat er geen verschil in beoordeling kon ontstaan door het beoordelen van verschillende hoofdstukken. Daarom zijn de leermiddelen in dit onderzoek allen beoordeeld op hoofdstuk drie van het leermiddel.

Analyse

De verschillende onderzoeksvragen moesten op verschillende manieren geanalyseerd worden. In deze paragraaf wordt per deelvraag besproken hoe ze geanalyseerd zijn.

1. De eerste deelvraag gaat over de consistentiebetrouwbaarheid van de MILK. Deze vraag kon worden beantwoord door een Cronbach's alpha te berekenen over de gegevens van de verschillende items bij de verschillende onderwerpen binnen een leerfunctie. Daarnaast is de Cronbach's alpha ook berekend over de gegevens van de verschillende items bij de gehele leerfuncties.
2. De tweede deelvraag ging over de interbeoordelaarsbetrouwbaarheid van de MILK. Deze vraag kon beantwoord worden door een Intraclass Correlation Coefficient [ICC] te berekenen over de gegevens van de ingevulde MILK per leermiddel. Met een ICC kan de overeenkomst tussen de beoordelingen van meer dan twee beoordelaars worden berekend (Howell, 2007). De ICC tussen beoordelaars is berekend over iedere leerfunctie en over de MILK in het geheel.
3. De derde deelvraag gaat over het verband tussen de verschillende onderdelen van de MILK. Deze vraag kon beantwoord worden door voor de drie gelegitimeerde

leerfuncties een totaalscore te maken. De totaalscore werd gemaakt door de scores van de items per onderdeel bij elkaar op te tellen. De opgetelde scores van verschillende leerfuncties zijn met behulp van een Pearson correlatie met elkaar vergeleken.

4. De vierde deelvraag gaat over het verschil in betrouwbaarheid tussen het gebruiken en weglaten van wegingsfactoren. Deze onderzoeksvraag kon worden beantwoord door een analyse uit te voeren waar rekening gehouden werd met de wegingsfactoren en een analyse uit te voeren waar de wegingsfactoren werden weggelaten. Deelnemers konden per leerfunctie 100 punten verdelen over de zes of zeven verschillende onderwerpen. De scores van de items werden vermenigvuldigd met de wegingsfactor van het onderwerp waartoe het item behoorde. Daarna werden de scores van de verschillende items binnen een onderwerp bij elkaar opgeteld, waarna het werd gedeeld door het aantal items binnen een onderwerp. Hierdoor ontstond er een totaalscore per onderwerp. De totaalscores van de zes of zeven onderwerpen werden bij elkaar opgeteld en gedeeld door 400. Er konden namelijk honderd punten gegeven worden en vier was de maximale beoordelingsscore die gegeven kon worden. Hierdoor ontstond er een score die kon variëren van 0 tot 1. Met een gepaarde t-toets is gekeken of het gebruik van wegingsfactoren zinvol is.
5. Bij de vijfde deelvraag is gekeken of er een verschil is in de interbeoordelaarsbetrouwbaarheid tussen experts en docenten. Het kijken naar verschillen tussen docenten en experts is mogelijk door dezelfde leermiddelen te laten beoordelen door docenten en experts. Deze vraag is, net als deelvraag twee, beantwoord met behulp van de Intraclass Correlation Coefficient. De ICC is voor experts en docenten apart berekend.

In Tabel 3 is een overzicht gegeven van de verschillende deelvragen, de instrumenten die bij deze deelvragen gebruikt worden en de analysetechnieken die bij de deelvragen gebruikt worden.

Tabel 3. *Overzicht gebruik van instrumenten en analysetechnieken per deelvraag*

Deelvraag	Instrument	Analyse
1. In hoeverre is er sprake van consistentiebetrouwbaarheid bij het gebruik van de gelegitimeerde leerfuncties van de MILK?	MILK Schalen: 1= helemaal niet mee eens, 2= een beetje mee eens, 3= grotendeels mee eens, 4= helemaal mee eens	Cronbach's alpha
2. In hoeverre is er sprake van een interbeoordelaarsbetrouwbaarheid bij het gebruik van de gelegitimeerde leerfuncties van de MILK?	MILK Schalen: 1= helemaal niet mee eens, 2= een beetje mee eens, 3= grotendeels mee eens, 4= helemaal mee eens	Intraclass correlatie coëfficiënt
3. In hoeverre is er een verband tussen de gelegitimeerde leerfuncties van de MILK?	MILK Schalen: 1= helemaal niet mee eens, 2= een beetje mee eens, 3= grotendeels mee eens, 4= helemaal mee eens	Pearson correlatietoets
4. In hoeverre is er een verschil tussen de beoordelingen van de verschillende leerfuncties bij het wel of niet gebruik maken van wegingsfactoren bij de gelegitimeerde onderdelen van de MILK?	MILK Schalen: 1= helemaal niet mee eens, 2= een beetje mee eens, 3= grotendeels mee eens, 4= helemaal mee eens Wegingsfactoren in de MILK	Gepaarde t-toets
5. Is er een verschil in de betrouwbaarheid van de gelegitimeerde onderdelen van de MILK tussen experts en docenten?	MILK Schalen: 1= helemaal niet mee eens, 2= een beetje mee eens, 3= grotendeels mee eens, 4= helemaal mee eens Vragenlijst Wat doet u voor werk?	Intraclass correlatie coëfficiënt

Resultaten

Deelvraag 1: Consistentiebetrouwbaarheid

Bij de eerste deelvraag is er gekeken naar de consistentiebetrouwbaarheid van de gelegitimeerde leerfuncties van de MILK en de consistentiebetrouwbaarheid van de verschillende onderwerpen binnen deze leerfuncties. Een leerfunctie bestaat namelijk uit verschillende onderwerpen welke weer bestaan uit verschillende items. In de bijlage is een deel van de MILK opgenomen, zodat deze verdeling duidelijk wordt. Bij de consistentiebetrouwbaarheid wordt gekeken in hoeverre de items binnen de verschillende onderwerpen met elkaar verbonden kunnen worden en in hoeverre de items binnen leerfuncties met elkaar verbonden kunnen worden. De consistentiebetrouwbaarheid van een onderwerp is goed als de Cronbach's alpha $> .60$. In Tabel 4 is een overzicht te vinden van de Cronbach's alpha per onderwerp binnen een leerfunctie.

Tabel 4. Cronbach's alpha per onderwerp binnen een leerfunctie

	N	Aantal items	Cronbach's alpha	Cronbach's alpha indien één item verwijderd
<i>Leerfunctie keuze van modaliteiten</i>	18	26	.86	.88
Teksten	25	5	.57	.63
Beelden	19	7	.88	.89
Geluid	4	3	.91	.91
Authentiek materiaal	10	4	.22	.73
Multimodaliteit	5	10	.91	.93
Doelgroep/leerdoel	24	6	.54	.68
<i>Leerfunctie keuze van de instructiestrategie</i>	7	38	.87	.85
Voorkennis activeren	25	6	.60	.70
Aandacht wordt gericht	24	9	.73	.77
Oefening	9	7	.00	.42
Stofverwerking	21	9	.67	.69
Motivatie	24	3	.77	.73
Feedback	19	2	.91	-
Reflectie	23	2	-.16	-
<i>Leerfunctie begrijpelijke teksten</i>	20	24	.77	.82
Vocabulaire	25	4	.70	.69
Zinsstructuur	25	5	.06	.64
Samenhang	25	3	.63	.66
Kwaliteit en verve	25	1	-	-
Tekststructuur	25	2	.55	-
Lay-out	25	5	.29	.52
Motivatie	20	4	.89	.91

Keuze van modaliteiten

Uit de analyse blijkt dat bij de leerfunctie *keuze van modaliteiten* de onderwerpen beelden, geluid en multimodaliteit een goede consistentiebetrouwbaarheid hebben, want de Cronbach's alpha >.80. De Cronbach's α ligt bij de onderwerpen tekst en doelgroep/leerdoel lager, namelijk onder de grens van $\alpha = .60$. De Cronbach's α bij het onderwerp tekst wordt significant als het volgende item wordt weggelaten: 'De gebruikte teksten moeten leerlingen vermaken'. De Cronbach's α bij het onderwerp doelgroep/leerdoel kan vergroot worden tot een $\alpha > .60$ als het volgende item wordt weggelaten: 'De gebruikte modaliteiten houden rekening met de verschillen binnen de doelgroep'. De Cronbach's α bij het onderwerp authentiek materiaal is zeer laag, maar kan vergroot worden naar een significante Cronbach's α als het volgende item wordt weggelaten: 'Geluid wordt in een authentieke vorm aangeboden'.

Als er wordt gekeken naar de consistentiebetrouwbaarheid tussen de verschillende items van de gehele leerfunctie *keuze van modaliteiten*, dus zonder onderscheid te maken tussen de verschillende onderwerpen, dan is het moeilijk om

uitspraken te doen. Dit komt doordat alle beoordelaars één van de items binnen deze leerfunctie beantwoord heeft met ‘niet te beoordelen’ of ‘niet van toepassing’. De meeste leermiddelen uit het onderzoek bevatten namelijk geen geluid en geen van de leermiddelen bevat tastbaar authentiek materiaal. Om de Cronbach's α van deze leerfunctie te berekenen zijn alle items welke geluid en authentiek materiaal bevatten weggelaten. Dit waren in totaal negen items. De Cronbach's α van de leerfunctie *keuze van modaliteiten* ligt, na het weglaten van de negen items, boven een $\alpha = .80$ wat betekent dat er een goede consistentiebetrouwbaarheid is tussen alle items van deze leerfunctie.

Keuze van de instructiestrategie

Bij de leerfunctie *keuze van de instructiestrategie* hebben de onderwerpen voorkennis activeren, aandacht, stofverwerking, motivatie en feedback een goede Cronbach's α . Het is echter wel belangrijk om rekening te houden met het lage aantal items bij het onderwerp feedback, want twee items in een onderwerp is eigenlijk te beperkt voor het berekenen van de consistentiebetrouwbaarheid. De leerfunctie *keuze van de instructiestrategie* bevat daarnaast twee onderwerpen (oefening en reflectie) bij welke de Cronbach's $\alpha < .60$. De Cronbach's α bij het onderwerp oefening is zeer laag, maar kan vergroot worden naar $\alpha = 0.42$ als het volgende item wordt verwijderd: ‘De opdrachten kunnen direct na het maken nagekeken worden’. De Cronbach's α van het onderwerp oefening kan vergroot worden naar $\alpha = .55$ als het volgende item ook wordt verwijderd: ‘De opdrachten zijn kort en duidelijk’. De Cronbach's α bij het laatste onderwerp reflectie was negatief. Aangezien dit onderwerp uit twee items bestaat is het niet mogelijk om de Cronbach's alpha te vergroten door het verwijderen van een item.

Als er wordt gekeken naar de Cronbach's α over alle items binnen de leerfunctie *keuze van de instructiestrategie*, dan is er een grote consistentie terug te vinden tussen de verschillende items. Er moet hier echter wel rekening gehouden worden met het lage aantal beoordelaars welke alle items hebben ingevuld met een beoordelingscore van één tot en met vier, dit waren namelijk maar zeven beoordelaars.

Begrijpelijke teksten

Bij de laatste leerfunctie *begrijpelijke teksten* hebben de onderwerpen vocabulaire, samenhang en motivatie een goede consistentiebetrouwbaarheid. De Cronbach's α bij het onderwerp tekststructuur lag echter onder de grens van $\alpha = .60$ en kan niet vergroot worden door het verwijderen van items, aangezien dit onderwerp uit twee items bestaat. Twee onderwerpen van de leerfunctie *begrijpelijke teksten* hebben een Cronbach's α ver onder de grens. Dit zijn de onderwerpen zinsstructuur en lay-out. De consistentiebetrouwbaarheid van het onderwerp zinsstructuur kan vergroot worden tot $\alpha = .64$ als het volgende item wordt weggelaten: ‘Er wordt geen gebruik gemaakt van

hulpwerkwoorden'. De consistentiebetrouwbaarheid van het onderwerp lay-out kan vergroot worden naar $\alpha = .52$ als het volgende item wordt verwijderd: 'Teksten worden niet in kolommen weergegeven'. De leerfunctie *begrijpelijke teksten* heeft nog één laatste onderwerp over. Dit onderwerp bestaat echter maar uit één item waardoor er geen Cronbach's α over dit onderwerp berekend kon worden. Het aantal items binnen dit onderwerp is namelijk te beperkt.

De Cronbach's alpha van de leerfunctie *begrijpelijke teksten* in het geheel, dus zonder onderscheid tussen verschillende onderwerpen, ligt boven de grens van $\alpha = .60$. De Cronbach's α van deze leerfunctie kan vergroot worden naar een $\alpha > .80$ als het item 'Er wordt gebruik gemaakt van hulpwerkwoorden' wordt weggelaten.

Deelvraag 2: Interbeoordelaarsbetrouwbaarheid

De tweede deelvraag gaat in op de interbeoordelaarsbetrouwbaarheid van de MILK. Om antwoord te kunnen geven op de tweede deelvraag over de interbeoordelaarsbetrouwbaarheid, is er per leermiddel de Intraclass Correlatie Coëfficiënt [ICC] berekend (Howell, 2007). De ICC is berekend voor de MILK in het geheel en per leerfunctie. De ICC is voldoende groot als deze groter is dan $.60$. De ICC is niet berekend over ieder onderwerp aangezien de consistentiebetrouwbaarheid bij de gehele leerfuncties voldoende groot is en er bij een aantal onderwerpen de consistentiebetrouwbaarheid erg laag is. De Intraclass Correlatie Coëfficiënten van de verschillende leermiddelen zijn terug te vinden in Tabel 5.

Uit de analyse van de tweede deelvraag blijkt dat de ICC van de gehele MILK bij al de vijf leermiddelen sterk is, oftewel een $ICC > .60$. Een voorbeeld van een sterke ICC van de gehele MILK is te vinden bij leermiddel één ($F(18, 72) = 3.61, p < .05$).

De ICC van de leerfunctie *keuze van modaliteiten* is ook bij alle vijf de leermiddelen sterk. Een voorbeeld van een sterke ICC is bij leermiddel één, waarbij de $ICC = .69$ ($F(4, 16) = 4.33, p < .05$).

Bij de leerfunctie *keuze van de instructiestrategie* tonen drie van de vijf leermiddelen een sterke ICC, namelijk bij leermiddel één, ($F(6, 24) = 8.39, p < .05$), leermiddel drie ($F(4, 16) = 3.26, p < .05$) en leermiddel vier ($F(5, 20) = 8.87, p < .05$). Leermiddel twee heeft geen sterke ICC, de ICC bij dit leermiddel is namelijk $.24$ ($F(6, 24) = 1.89, p > .05$). Leermiddel vijf heeft ook geen sterke ICC, de ICC bij dit leermiddel is ook $.24$ ($F(5, 20) = 1.67, p > .05$).

Bij de laatste leerfunctie, *begrijpelijke teksten*, hebben vier van de vijf leermiddelen een sterke ICC, namelijk leermiddel twee ($F(6, 24) = 6.36, p < .05$), leermiddel drie ($F(6, 24) = 7.34, p < .05$), leermiddel vier en leermiddel vijf. Bij

leermiddel één is geen sterke ICC gevonden, de ICC was bij dit leermiddel namelijk .15, $F(6, 24) = 1.26, p > .05$.

Tabel 5. *Intraclass Correlatie Coëfficiënt per leermiddel*

Leermiddel	Leerfunctie	ICC	F	df1	df2	p
Leermiddel 1	<i>Milk totaal</i>	.67	3.61	18	72	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.69	4.33	4	16	.02
	<i>Leerfunctie keuze van de instructiestrategie</i>	.76	8.39	6	24	.00
	<i>Leerfunctie begrijpelijke teksten</i>	.15	1.26	6	24	.31
Leermiddel 2	<i>Milk totaal</i>	.61	4.13	18	72	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.73	6.88	4	16	.00
	<i>Leerfunctie keuze van de instructiestrategie</i>	.24	1.89	6	24	.12
	<i>Leerfunctie begrijpelijke teksten</i>	.70	6.36	6	24	.00
Leermiddel 3	<i>Milk totaal</i>	.80	5.73	16	64	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.85	8.99	4	16	.00
	<i>Leerfunctie keuze van de instructiestrategie</i>	.66	3.26	4	16	.04
	<i>Leerfunctie begrijpelijke teksten</i>	.85	7.34	6	24	.00
Leermiddel 4	<i>Milk totaal</i>	.84	7.84	17	68	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.75	5.07	4	16	.01
	<i>Leerfunctie keuze van de instructiestrategie</i>	.79	8.87	5	20	.00
	<i>Leerfunctie begrijpelijke teksten</i>	.91	11.27	6	24	.00
Leermiddel 5	<i>Milk totaal</i>	.83	6.81	17	68	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.94	23.12	4	16	.00
	<i>Leerfunctie keuze van de instructiestrategie</i>	.24	1.67	5	20	.19
	<i>Leerfunctie begrijpelijke teksten</i>	.85	6.34	6	24	.00

Deelvraag 3: Verband leerfuncties

Bij de derde deelvraag is er gezocht naar een verband tussen de gelegitimeerde leerfuncties van de MILK. De verschillende leerfuncties meten allen de kwaliteit van leermiddelen, is er dan ook een verband te vinden tussen de verschillende leerfuncties? Dit verband is gezocht met behulp van een Pearson correlatietoets. De nulhypothese is dat er geen correlatie is tussen de verschillende leerfuncties, dus $H_0: \mu_x \neq \mu_y$. In Tabel 6 zijn de beschrijvende statistieken van deze toets terug te vinden.

Tabel 6. *Beschrijvende statistieken verband leerfuncties*

	N	Gemiddelde	Standaarddeviatie
Totaal leerfunctie keuze van modaliteiten	13	2.63	.42
Totaal leerfunctie keuze van de instructiestrategie	25	2.47	.43
Totaal leerfunctie begrijpelijke teksten	25	2.81	.33

Uit de Pearson correlatietoets is gebleken dat er een significante positieve correlatie is tussen de leerfuncties *keuze van modaliteiten* en *keuze van de instructiestrategie*, $r(23) = .55$, $p < .05$. Daarnaast is er een significant positieve correlatie gevonden tussen de leerfuncties *keuze van de instructiestrategie* en *begrijpelijke teksten*, $r(23) = .66$, $p < .05$. Tussen de leerfuncties *keuze van modaliteiten* en *begrijpelijke teksten*, $r(23) = .16$, $p > .05$, is echter geen positieve of negatieve correlatie te vinden.

Deelvraag 4: Wegingsfactoren

Bij de vierde deelvraag is er gekeken naar het verschil in de beoordelingen van de verschillende leerfuncties bij het wel of niet gebruik maken van wegingsfactoren bij de gelegitimeerde onderdelen van de MILK. Aangezien er sprake is van een normale verdeling is voor deze onderzoeksvraag een gepaarde t-toets gebruikt. De nulhypothese is dat de variabelen *keuze van modaliteiten zonder wegingsfactoren* en *keuze van modaliteiten met wegingsfactoren* aan elkaar gelijk zijn. Ditzelfde geldt voor de andere twee leerfuncties, dus $H_0: \mu_x = \mu_y$. In Tabel 7 zijn de beschrijvende statistieken als aantal deelnemers, gemiddelden, standaarddeviatie, standaardfout, de T-waarde en de significantie te vinden.

Bij de vergelijking van de gegevens van de leerfunctie *keuze van modaliteiten* met en zonder wegingsfactoren is een significant verschil gevonden, $t(24) = -6.15$, $p = .00$. Dit betekent dat de nulhypothese verworpen kan worden. Bij de vergelijking van de gegevens van de tweede leerfunctie *keuze van de instructiestrategie* is ook een significant gevonden, $t(24) = -2.78$, $p = .01$. Dit betekent dat de nulhypothese ook bij deze vergelijking verworpen kan worden. Als laatste is er gekeken naar het wel of niet gebruik maken van wegingsfactoren bij de leerfunctie *begrijpelijke teksten*. Bij deze leerfunctie is geen significant verschil gevonden, $t(24) = -.82$, $p = .42$. De nulhypothese kan bij deze vergelijking dus niet verworpen worden.

Tabel 7. Beschrijvende statistieken gepaarde t-toets wegingsfactoren

	Aantal deelnemers	Gemiddelde	Standaard deviatie	Standaardfout	T	P
Leerfunctie keuze van modaliteiten zonder wegingsfactoren	25	.58	.09	.02	-	-
Leerfunctie keuze van modaliteiten met wegingsfactoren	25	.66	.11	.02	-	-
Vershil leerfunctie keuze van modaliteiten met en zonder wegingsfactoren	25	-.08	.06	.01	-6.15	.00
Leerfunctie keuze van de instructiestrategie zonder wegingsfactoren	25	.58	.11	.02	-	-
Leerfunctie keuze van de instructiestrategie met wegingsfactoren	25	.60	.11	.02	-	-
Vershil leerfunctie keuze van instructiestrategie met en zonder wegingsfactoren	25	-.02	.04	.01	-2.78	.01
Leerfunctie begrijpelijke teksten zonder wegingsfactoren	25	.69	.09	.02	-	-
Leerfunctie begrijpelijke teksten met wegingsfactoren	25	.70	.10	.02	-	-
Vershil leerfunctie begrijpelijke teksten met en zonder wegingsfactoren	25	-.01	.04	.01	-.82	.42

Deelvraag 5: Experts en docenten

De laatste deelvraag gaat in op het verschil in interbeoordelaarsbetrouwbaarheid van de gelegitimeerde onderdelen van de MILK tussen onderwijskundige experts en docenten. Om antwoord te kunnen geven op deze deelvraag is per leermiddel en per groep beoordelaars (docenten en experts) de ICC berekend. De ICC is voldoende groot als deze groter is dan .60. In Tabel 8 en 9 zijn de Intraclass Correlatie Coëfficiënten gegeven van alleen de docenten en alleen de experts.

Docenten

Uit de analyse van de vijfde deelvraag blijkt dat de Intraclass Correlatie Coëfficiënt van de gehele MILK, tussen de verschillende docenten, bij drie van de vijf leermiddelen sterk is. De ICC is sterk te noemen bij leermiddel drie, leermiddel vier en leermiddel vijf. Een voorbeeld van een hoge ICC is te vinden bij leermiddel drie, $F(18,36) = 3.22$, $p < .10$. Bij leermiddel één en leermiddel twee is de ICC $< .60$ en dat betekent dat de ICC niet hoog is.

Bij de leerfunctie *keuze van modaliteiten* hebben ook de leermiddelen drie, vier en vijf een sterke ICC.

Tabel 8. *Intraclass Correlatie Coëfficiënt van alleen docenten*

Leermiddel	Leerfunctie	ICC	F	df1	df2	p
Leermiddel 1	<i>Milk totaal</i>	.34	1.75	18	36	.08
	<i>Leerfunctie keuze van modaliteiten</i>	.24	1.50	4	8	.29
	<i>Leerfunctie keuze van de instructiestrategie</i>	.54	4.27	6	12	.02
	<i>Leerfunctie begrijpelijke teksten</i>	.02	1.03	6	12	.46
Leermiddel 2	<i>Milk totaal</i>	.25	1.89	18	36	.05
	<i>Leerfunctie keuze van modaliteiten</i>	.46	4.74	4	8	.03
	<i>Leerfunctie keuze van de instructiestrategie</i>	-.21	0.22	6	12	.96
	<i>Leerfunctie begrijpelijke teksten</i>	.63	5.25	6	12	.01
Leermiddel 3	<i>Milk totaal</i>	.69	3.22	17	34	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.70	3.29	4	8	.07
	<i>Leerfunctie keuze van de instructiestrategie</i>	.52	2.32	5	10	.12
	<i>Leerfunctie begrijpelijke teksten</i>	.75	3.85	6	12	.02
Leermiddel 4	<i>Milk totaal</i>	.73	4.55	17	34	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.57	2.71	4	8	.11
	<i>Leerfunctie keuze van de instructiestrategie</i>	.66	6.64	5	10	.01
	<i>Leerfunctie begrijpelijke teksten</i>	.84	5.86	6	12	.01
Leermiddel 5	<i>Milk totaal</i>	.70	3.53	17	34	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.94	15.79	4	8	.00
	<i>Leerfunctie keuze van de instructiestrategie</i>	-.81	0.41	5	10	.83
	<i>Leerfunctie begrijpelijke teksten</i>	.69	3.06	6	12	.05

Bij de leerfunctie *keuze van de instructiestrategie* heeft één van de vijf leermiddelen een sterke ICC (ICC >.60), namelijk leermiddel vier. De andere leermiddelen hebben een ICC < .60. Leermiddel twee en leermiddel vijf tonen zelfs een negatieve ICC. De ICC van de leerfunctie *keuze van de instructiestrategie* zijn bij deze leermiddelen dan ook zeer zwak. Deze leerfunctie heeft bij leermiddel twee bijvoorbeeld een ICC van $-.21$ ($F(6,12) = .02$, $p >.05$).

De ICC van de verschillende docenten is bij de laatste leerfunctie *begrijpelijke teksten* sterk bij vier van de vijf leermiddelen. Alleen bij leermiddel één is de ICC tussen de drie docenten zwak te noemen, ICC = $.02$ ($F(6,12) = 1.03$, $p >.05$).

Tabel 9. *Intraclass Correlatie Coëfficiënt van alleen experts*

Leermiddel	Leerfunctie	ICC	F	df1	df2	p
Leermiddel 1	<i>Milk totaal</i>	.38	1.60	18	18	.17
	<i>Leerfunctie keuze van modaliteiten</i>	.65	3.27	4	4	.14
	<i>Leerfunctie keuze van de instructiestrategie</i>	.88	16.51	6	6	.00
	<i>Leerfunctie begrijpelijke teksten</i>	-.64	0.37	6	6	.87
Leermiddel 2	<i>Milk totaal</i>	.56	3.08	18	18	.01
	<i>Leerfunctie keuze van modaliteiten</i>	.77	3.67	4	4	.12
	<i>Leerfunctie keuze van de instructiestrategie</i>	.62	6.56	6	6	.02
	<i>Leerfunctie begrijpelijke teksten</i>	.27	1.96	6	6	.22
Leermiddel 3	<i>Milk totaal</i>	.67	4.16	16	16	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.87	7.68	4	4	.04
	<i>Leerfunctie keuze van de instructiestrategie</i>	.52	2.44	4	4	.20
	<i>Leerfunctie begrijpelijke teksten</i>	.68	4.79	6	6	.04
Leermiddel 4	<i>Milk totaal</i>	.73	3.92	19	19	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.68	4.31	5	5	.07
	<i>Leerfunctie keuze van de instructiestrategie</i>	.45	1.73	6	6	.26
	<i>Leerfunctie begrijpelijke teksten</i>	.91	9.74	6	6	.01
Leermiddel 5	<i>Milk totaal</i>	.73	4.77	19	19	.00
	<i>Leerfunctie keuze van modaliteiten</i>	.56	4.11	5	5	.07
	<i>Leerfunctie keuze van de instructiestrategie</i>	.19	1.23	6	6	.41
	<i>Leerfunctie begrijpelijke teksten</i>	.93	17.91	6	6	.00

Experts

Uit de analyse van de vijfde deelvraag blijkt dat de Intraclass Correlatie Coëfficiënt van de gehele MILK, tussen de verschillende experts, bij drie van de vijf leermiddelen sterk is. De ICC is sterk te noemen bij leermiddel drie, vier en vijf. Een voorbeeld van een sterke ICC is te vinden bij leermiddel vijf. De ICC bij leermiddel één is niet sterk, deze is namelijk .38 ($F(18,18) = 1.60, p > .05$).

Bij de leerfunctie *keuze van modaliteiten* is de ICC, tussen de verschillende experts, bij vier van de vijf leermiddelen sterk. Alleen leermiddel vijf heeft een iets zwakkere ICC, de ICC is hier namelijk .56 ($F(5,5) = 4.11, p < .05$).

Bij de leerfunctie *keuze van de instructiestrategie* is de ICC sterk bij twee van de vijf leermiddelen (ICC $>.60$), namelijk bij leermiddel één en twee. De drie andere

leermiddelen hebben een zwakke tot matige ICC. Deze leerfunctie heeft bij leermiddel vier bijvoorbeeld een ICC van .45 ($F(6,6) = 1.73, p > .05$).

De ICC van de verschillende experts is bij de laatste leerfunctie *begrijpelijke teksten* hoog bij drie van de vijf leermiddelen, namelijk bij leermiddel drie, vier en vijf. Leermiddel één heeft bij deze leerfunctie een negatieve ICC, namelijk $-.64$ ($F(6,6) = .04, p > .05$).

Conclusie

Naar aanleiding van de analyse en de resultaten kan er antwoorden gegeven worden op de verschillende vragen van dit onderzoek.

Deelvraag 1: Consistentiebetrouwbaarheid

De eerste deelvraag was als volgt: In hoeverre is er sprake van consistentiebetrouwbaarheid bij het gebruik van de gelegitimeerde leerfuncties van de MILK?

Uit de resultaten van deze deelvraag komen een aantal belangrijke punten naar voren. Van de twintig verschillende onderwerpen binnen de drie verschillende leerfuncties hebben elf onderwerpen een Cronbach's $\alpha > .60$. Dit betekent dat er bij de items binnen deze onderwerpen sprake is van een goede consistentie tussen items, wat betekent dat de items binnen deze onderwerpen goed met elkaar samenhangen. De consistentie van de onderwerpen binnen de leerfunctie *keuze van modaliteiten* welke onder de grens van $\alpha = .60$ lagen (teksten, authentiek materiaal en doelgroep/leerdoel), kunnen worden vergroot door één item per onderwerp weg te laten. De Cronbach's α overschrijdt dan de grens van $\alpha = .60$.

De twee onderwerpen binnen de leerfunctie *keuze van de instructiestrategie* welke een Cronbach's α hebben $< .60$, kunnen niet vergroot worden naar een $\alpha > .60$ door het weglaten van een item. Dit zijn de onderwerpen oefeningen en reflectie. Het onderwerp reflectie heeft een negatieve α wat betekent dat de interne consistentie binnen dit onderwerp slecht is. Als een beoordelaar bij één item binnen dit onderwerp laag scoort, dan scoort hij bij een ander item hoog. De items moeten dus onderverdeeld worden bij andere onderwerpen. De Cronbach's α van het onderwerp oefeningen wordt niet groot genoeg als er één item wordt weggehaald. Bij deze onderwerpen is er dus weinig samenhang tussen de verschillende items.

Bij de laatste leerfunctie *begrijpelijke teksten* lag het onderwerp tekststructuur iets onder de grens van $\alpha = .60$, namelijk $\alpha = .55$. De items van dit onderwerp hangen dus niet voldoende met elkaar samen. De onderwerpen zinsstructuur en lay-out liggen ver onder de grens van $\alpha = .60$. Wat betekent dat er weinig samenhang is. Als er bij het onderwerp zinsstructuur één item wordt weggehaald, dan is de samenhang tussen de

overgebleven items groot genoeg voor een consistentie tussen items van $\alpha > .60$. De consistentie tussen items bij het onderwerp lay-out kan vergroot worden door een item weg te halen, maar de Cronbach's α wordt dan vergroot tot een alpha van $.52$. De samenhang tussen de overgebleven items is dan nog niet voldoende.

Van de twintig verschillende onderwerpen blijven vier onderwerpen onder de grens van $\alpha = .60$ als er items worden verwijderd. Bij deze vier is de samenhang tussen items dus niet groot genoeg, ook als er één item wordt weggehaald. Als er echter wordt gekeken naar de samenhang tussen de verschillende items binnen een leerfunctie zonder onderscheid te maken tussen verschillende onderwerpen, dan hebben alle drie de leerfuncties een significante Cronbach's α . Dit betekent dat alle items binnen een leerfunctie, zonder onderscheid tussen verschillende onderwerpen, wel voldoende met elkaar samenhangen.

Deelvraag 2: Interbeoordelaarsbetrouwbaarheid

De tweede deelvraag luidde als volgt: In hoeverre is er sprake van een interbeoordelaarsbetrouwbaarheid bij het gebruik van de gelegitimeerde leerfuncties van de MILK?

Uit de analyse van deze deelvraag is gebleken dat de Intraclass Correlatie Coëfficiënt [ICC] van de MILK in het geheel sterk is bij alle vijf de leermiddelen. Hetzelfde geldt voor de leerfunctie *keuze van modaliteiten*. Dit betekent dat er een duidelijke correlatie is tussen de verschillende beoordelaars bij het beoordelen van een leermiddel. Er zijn niet veel noemenswaardige verschillen tussen de vijf beoordelaars die eenzelfde leermiddel beoordeeld hebben. Dit betekent dus dat er een goede overeenstemming is tussen de verschillende beoordelaars bij de gehele MILK en bij de leerfunctie *keuze van modaliteiten*.

Bij de leerfunctie *keuze van de instructiestrategie* hebben drie van de vijf leermiddelen een sterke ICC, namelijk leermiddel één, drie en vier. Dit betekent dat er bij drie van de vijf leermiddelen een duidelijke correlatie is tussen de beoordelingen van de verschillende beoordelaars op deze leerfunctie. Deze leerfunctie heeft dus bij drie van de vijf leermiddelen een goede overeenstemming tussen beoordelaars. De ICC van de leermiddelen twee en vijf zijn echter niet sterk. Dit betekent dat er duidelijke verschillen zijn in de beoordelingen van de verschillende beoordelaars bij deze leerfunctie. De overeenstemming tussen beoordelaars kan bij deze leerfunctie dus verbeterd worden. De verschillen in beoordelingen bij deze leerfunctie kunnen zijn ontstaan doordat de consistentie tussen items bij de onderwerpen reflectie en oefeningen niet voldoende groot was.

Bij de laatste leerfunctie, *begrijpelijke teksten*, hebben vier van de vijf leermiddelen een sterk resultaat (twee, drie vier en vijf). Dit betekent dat er bij vier van

de vijf leermiddelen een duidelijke correlatie te zien is tussen de beoordelingen van de verschillende beoordelaars. Bij deze leermiddelen is er dus sprake van een goede overeenstemming tussen de beoordelaars bij de leerfunctie *begrijpelijke teksten*. De ICC van leermiddel één is echter niet sterk. Dit betekent dat er verschillen zijn in de beoordelingen bij dit leermiddel. De overeenstemming tussen beoordelaars bij deze leerfunctie kan dus verbeterd worden. De verschillen in beoordelingen tussen de verschillende deelnemers kunnen zijn ontstaan doordat de consistentie tussen items bij de onderwerpen zinsstructuur en lay-out bij deze leerfunctie niet voldoende groot was.

Deelvraag 3: Verband leerfuncties

De derde deelvraag luidde als volgt: In hoeverre is er een verband tussen de gelegitimeerde leerfuncties van de MILK?

Uit de resultaten is gebleken dat er een significante positieve correlatie is tussen de leerfuncties *keuze van modaliteiten* en *keuze van de instructiestrategie*. Dit betekent dat een hoge score bij de totaalscore van de leerfunctie *keuze van modaliteiten* een hoge score bij de totaalscore van de leerfunctie *keuze van de instructiestrategie* is. Er is dus een verband tussen deze twee leerfuncties.

Tussen de leerfunctie *keuze van de instructiestrategie* en *begrijpelijke teksten* is ook een positieve correlatie gevonden. Dit betekent dat een hoge score bij het totaal van de leerfunctie *keuze van de instructiestrategie*, een hoge score bij het totaal van de leerfunctie *begrijpelijke teksten* is. Hier is dus ook een verband te vinden tussen de twee leerfuncties.

Er is echter geen verband gevonden tussen de leerfuncties *keuze van modaliteiten* en *begrijpelijke teksten*. De oorzaak hiervan kan zijn dat de leerfunctie *keuze van modaliteiten* onderwerpen bevat die weinig te maken heeft met leesbaarheid, namelijk de onderwerpen beelden en geluid.

Deelvraag 4: Wegingsfactoren

De vierde deelvraag was als volgt: In hoeverre is er een verschil tussen de beoordelingen van de verschillende leerfuncties bij het wel of niet gebruik maken van wegingsfactoren bij de gelegitimeerde onderdelen van de MILK?

Bij de analyse over het gebruik van wegingsfactoren, bij de drie verschillende leerfuncties, zijn verschillende resultaten naar voren gekomen. Bij de leerfuncties *keuze van modaliteiten* en *keuze van de instructiestrategie* zijn de nulhypotheses verworpen. Dit betekent dat de variabelen *keuze van modaliteiten zonder wegingsfactoren* en *keuze van modaliteiten met wegingsfactoren* niet aan elkaar gelijk zijn. Hetzelfde geldt voor de leerfunctie *keuze van de instructiestrategie*. Het gebruik van wegingsfactoren zorgt dus voor een verschil in de beoordeling van de kwaliteit van leermiddelen. De deelnemers

vinden sommige onderwerpen binnen de leerfuncties *keuze van modaliteiten* en *keuze van de instructiestrategie* dus belangrijker dan andere onderwerpen waardoor verschillen in beoordelingen zijn ontstaan. Bij de laatste leerfunctie, *begrijpelijke teksten*, is geen significant verschil gevonden tussen het wel of niet gebruiken van wegingsfactoren. Hierdoor is de nulhypothese niet verworpen. Dit betekent dat het gebruik van wegingsfactoren bij deze leerfunctie niet heeft geleid tot een verschil in beoordeling. Dit kan komen doordat de deelnemers de verschillende onderwerpen binnen deze leerfunctie even belangrijk vinden. Als ze de verschillende onderwerpen dezelfde wegingsfactoren geven, dan is het verschil tussen het wel of niet gebruiken van wegingsfactoren klein.

Deelvraag 5: Experts en docenten

De vijfde deelvraag luidde als volgt: Is er een verschil in de interbeoordelaarsbetrouwbaarheid van de gelegitimeerde onderdelen van de MILK tussen onderwijskundige experts en docenten?

Uit de resultaten is gebleken dat de ICC van de MILK in het geheel tussen de verschillende docenten groot is bij drie van de vijf leermiddelen, hetzelfde geldt voor de leerfunctie *keuze van modaliteiten*. Dit betekent dat er bij drie leermiddelen een duidelijke correlatie is tussen de verschillende docenten bij het beoordelen van een leermiddel. Bij deze drie leermiddelen zijn er geen duidelijke verschillen tussen de vijf beoordelaars die eenzelfde leermiddel beoordeeld hebben. Dit betekent dat er een goede overeenstemming is tussen beoordelaars bij de gehele MILK en bij de leerfunctie *keuze van modaliteiten* bij deze drie leermiddelen. De ICC van de MILK in het geheel is tussen de verschillende experts ook bij drie van de vijf leermiddelen sterk. Het gaat hierbij om dezelfde leermiddelen als bij de docenten (drie tot en met vijf). Er is dus geen verschil in overeenstemming tussen de groep docenten en de groep experts bij de gehele MILK en de leerfunctie *keuze van modaliteiten*. De MILK is dus voor deze twee verschillende groepen even goed te gebruiken.

Bij de leerfunctie *keuze van modaliteiten* zijn er wel verschillen te vinden tussen docenten en experts wat betreft interbeoordelaarsbetrouwbaarheid. Bij de docenten hebben drie van de vijf leermiddelen een sterke ICC, terwijl bij de experts vier van de vijf een sterke ICC hebben. De docenten hadden bij de leermiddelen drie, vier en vijf een sterke ICC. De experts hadden bij de leermiddelen één tot en met vier een sterke ICC. Hier is dus wel een verschil te vinden tussen experts en docenten bij het beoordelen van verschillende leermiddelen.

Bij de leerfunctie *keuze van de instructiestrategie* zijn er verschillen te vinden tussen docenten en experts wat betreft de overeenstemming binnen de groepen. Bij de docenten heeft maar één van de vijf leermiddelen een sterke ICC ($ICC > .60$), namelijk leermiddel vier. Bij de experts hebben twee leermiddelen een sterke ICC en dit zijn de

leermiddelen één en twee. Bij de docenten hebben twee van de vijf leermiddelen zelfs een negatieve ICC. Een negatieve ICC betekent dat de tussen-groep variantie kleiner is dan een binnen-groep variantie. Een derde variabele heeft dus effecten gehad op de verschillende beoordelaars. Aangezien dit onderzoek per leermiddel weinig beoordelaars heeft, is het moeilijk te bepalen wat deze derde variabele is. Bij deze leerfunctie hebben de meeste leermiddelen een lage ICC. Dit kan komen doordat er een aantal onderwerpen zijn binnen deze leerfunctie met een lage consistentie tussen items. Daarnaast kan een lage ICC te verklaren zijn door het kleine aantal beoordelaars per leermiddel.

Bij de laatste leerfunctie, *begrijpelijke teksten*, is er ook een verschil tussen docenten en experts te vinden. Dit verschil is echter niet zo groot als bij de leerfunctie keuze van de instructiestrategie. Bij de docenten hebben vier van de vijf leermiddelen een sterke ICC, alleen de ICC bij leermiddel één is zwak te noemen. Bij de experts hebben drie van de vijf leermiddelen een sterke ICC, alleen de leermiddelen één en twee zijn zwak. Bij leermiddel één is de ICC zelfs negatief, wat betekent dat de tussen-groep variantie kleiner is dan een binnen-groep variantie.

Conclusie onderzoeksvraag

Aan het begin van dit artikel werd de volgende vraag gesteld: “In hoeverre zijn de gelegitimeerde leerfuncties *keuze van modaliteiten*, *keuze van de instructiestrategie* en *begrijpelijke teksten*, van het door het CLU ontwikkelde evaluatie-instrument de MILK, valide en betrouwbare onderdelen van het instrument?”

Bij deelvraag één is gebleken dat elf van de twintig onderwerpen van de drie verschillende leerfuncties met elkaar samenhangen. Bij vijf van de negen overgebleven onderwerpen kan de consistentie tussen items wel vergroot worden, maar dan moeten er items worden verwijderd. Bij vier onderwerpen kan de consistentie tussen items niet vergroot worden. Dit betekent dat de leerfuncties binnen de MILK niet altijd even betrouwbaar zijn. Als er echter wordt gekeken naar de leerfuncties in het geheel, dan is tussen de verschillende items binnen een gehele leerfunctie (dus zonder onderscheid in onderwerpen) wel een goede consistentie tussen de verschillende items te vinden. De indeling van de verschillende items binnen de leerfuncties moet dus veranderd worden om de betrouwbaarheid van deze leerfuncties te vergroten.

Naast de consistentiebetrouwbaarheid is er in dit onderzoek ook gekeken naar de interbeoordelaarsbetrouwbaarheid van de MILK. Bij de totaalscore van de MILK is de overeenstemming tussen beoordelaars bij alle vijf de leermiddelen voldoende groot. Hetzelfde geldt voor de leerfunctie *keuze van modaliteiten*. Kijkt men echter naar de leerfunctie *keuze van de instructiestrategie* dan is te zien dat de overeenkomsten tussen beoordelaars bij twee van de vijf leermiddelen laag is. Dit betekent dat de interbeoordelaarsbetrouwbaarheid bij deze leerfunctie verbeterd kan worden. Bij de

leerfunctie *begrijpelijke teksten* is maar bij één van de vijf leermiddelen een lage overeenkomst tussen beoordelaars te vinden, wat betekent dat de betrouwbaarheid van deze leerfunctie redelijk is, maar nog wel verbeterd kan worden.

Om de betrouwbaarheid van de MILK te kunnen beoordelen is er ook gekeken naar het verband tussen de drie gelegitimeerde leerfuncties. Als een deelnemer een hoge score heeft op een leerfunctie, betekent dit dan ook een hoge score op de andere twee leerfuncties? Uit de resultaten is gebleken dat de meeste leerfuncties goed samenhangen, behalve de leerfuncties *keuze van modaliteiten* en *begrijpelijke teksten*. Hiervoor is echter een redelijke verklaring gevonden en dit beperkt dus niet de betrouwbaarheid van de MILK.

De vierde deelvraag, over wegingsfactoren, heeft ook betrekking op de betrouwbaarheid van de MILK. Het geven van wegingsfactoren is namelijk een onderdeel van het invullen van de MILK. Heeft het zin om docenten wegingsfactoren te laten geven of is dit onnodig extra werk? Uit de resultaten is gebleken dat alleen bij de leerfunctie *begrijpelijke teksten* geen verschil is tussen het wel of niet geven van wegingsfactoren. Bij de andere twee leerfuncties is wel een verschil gevonden in de scores en daardoor ook in de mate hoe kwalitatief een beoordelaar het leermiddel vindt. De wegingsfactoren zorgen bij deze twee leerfuncties dus voor een betrouwbaarder instrument, want voor de beoordelaars wordt meer zichtbaar welke leermiddelen zij van betere kwaliteit vinden. Daarom is het zinvol om wegingsfactoren te laten geven door beoordelaars.

Bij de laatste deelvraag is duidelijk geworden dat er bij de totaalscore van de gehele MILK en de totaalscore van de leerfunctie *keuze van modaliteiten* geen verschil is gevonden tussen de interbeoordelaarsbetrouwbaarheid van experts en docenten. Bij de leerfuncties *keuze van de instructiestrategie* en *begrijpelijke teksten* zijn wel verschillen gevonden. Dit betekent dat de MILK bij deze leerfuncties voor docenten op andere gebieden betrouwbaar is dan voor experts. De beoordelingen van experts correleren met elkaar bij bepaalde leermiddelen, terwijl de beoordelingen van de docenten meer correleren bij andere leermiddelen. De correlaties verschillen dus per leermiddel. Een verschil in interbeoordelaarsbetrouwbaarheid betekent dus niet dat de beoordelingen van experts en docenten bij een leermiddel anders zijn. In vervolgonderzoek kan hier echter wel naar gekeken worden. Een verschil in interbeoordelaarsbetrouwbaarheid betekent dat bij de ene groep beoordelaars wel een overeenstemming is binnen de groep, terwijl deze overeenstemming tussen beoordelingen binnen de andere groep niet te vinden is.

Uit de verschillende deelvragen is duidelijk geworden dat de MILK deels betrouwbaar is, maar dat er nog wel veranderingen moeten plaatsvinden binnen de MILK voordat dit instrument echt betrouwbaar genoemd kan worden.

Discussie

Bij dit onderzoek moet rekening gehouden worden met een aantal factoren ter relativering.

Een eerste punt ter relativering is dat één leermiddel niet meegenomen kon worden in de analyses, doordat de cd-rom bij het leermiddel door de experts niet te openen was. Twee van de drie docenten hadden deze cd-rom wel gebruikt bij de beoordeling. Door dit punt kunnen onbedoelde verschillen in de beoordelingen zijn ontstaan en daarom is er voor gekozen om dit leermiddel niet mee te nemen.

Een tweede punt is dat de deelnemers een kopie van de methode kregen om te beoordelen. Een kopie geeft bepaalde aspecten van de lay-out anders weer dan bij het originele leermiddel, terwijl de lay-out wel belangrijk is. Een lay-out kan namelijk helpen bij het verduidelijken van de indeling van het leermiddel. De deelnemers kregen een zwart-wit kopie, omdat een originele leermiddel lastig te verkrijgen was voor vijf verschillende beoordelaars. Bij vervolgonderzoek moeten beoordelaars de mogelijkheid krijgen om het origineel in ieder geval in te zien. Dit is mogelijk door alle docenten en experts tijdens de bijeenkomst de MILK te laten invullen in plaats van thuis.

Een derde punt is dat geen enkel leermiddel in dit onderzoek digitaal was. Bij vervolgonderzoek is het goed om te kijken naar de beoordeling van deelnemers wat betreft digitaal leermateriaal. Er is geen onderscheid gemaakt, omdat er al veel vergelijkingen zijn gemaakt binnen dit onderzoek. Een vergelijking tussen digitaal en niet-digitaal leermateriaal zou te veel zijn geweest voor één onderzoek. Bij drie van de vijf geanalyseerde leermiddelen werd er wel gebruik gemaakt van een cd-rom, maar er zat niet altijd geluid bij en nooit tastbaar authentiek materiaal. Hierdoor is het moeilijk om de betrouwbaarheid van de onderwerpen geluid en authentiek materiaal bij de leerfunctie *keuze van modaliteiten* te bepalen. De betrouwbaarheid van deze onderwerpen kan pas goed bepaald worden in vervolgonderzoek met digitaal materiaal. Onderzoek naar digitaal materiaal is ook belangrijk, omdat er steeds meer leermiddelen aangeboden worden via het internet (Ministerie OC&W, 2009).

Een vierde punt is dat van ieder leermiddel maar één hoofdstuk is beoordeeld in plaats van het gehele leermiddel. Hierdoor kunnen verschillen zijn ontstaan in de beoordelingen, aangezien iedere docent met één van de beoordeelde leermiddelen werkt. Deze docenten weten wat er in de andere hoofdstukken van het leermiddel wordt besproken, terwijl de andere beoordelaars dat niet weten aangezien zij geen beschikking hebben over de rest van het leermiddel. Bij vervolgonderzoek kan gekeken worden of er inderdaad verschillen zijn in de beoordelingen van docenten die bekend zijn met het leermiddel en docenten die niet bekend zijn met het leermiddel.

Een vijfde punt ter relativering is dat het aantal beoordelaars per leermiddel en het aantal leermiddelen bij dit onderzoek klein is. Het is daardoor moeilijk, om aan de hand van de analyses van dit onderzoek, generaliseerbare uitspraken te doen betreft de MILK. Bij de conclusies moet daarom rekening gehouden worden met het kleine aantal beoordelaars en leermiddelen en daarmee met de generaliseerbaarheid van de conclusies. Ondanks dat het moeilijk is om de resultaten te generaliseren kan er met dit onderzoek wel bepaalde uitspraken gedaan worden over de MILK. De betrouwbaarheid van de MILK kan bijvoorbeeld vergroot worden. Het is belangrijk dat de betrouwbaarheid van de MILK vergroot wordt, omdat er naast de MILK eigenlijk geen tot weinig evaluatie-instrumenten bestaan om de kwaliteit van leermiddelen te beoordelen (Reints, 2000). De MILK is in ontwikkeling en kan aan de hand van onder andere dit onderzoek aangepast worden. Het is belangrijk dat de MILK wordt aangepast, omdat docenten kwalitatief goede leermiddelen moeten gebruiken in scholen. Leermiddelen zijn namelijk de kern waar leeractiviteiten om heen worden gemaakt (Lester et al., 1998).

Een ander punt is dat bij een aantal onderwerpen binnen de MILK een lage of zelfs negatieve Cronbach's alpha gevonden is. Dit betekent dat de items binnen deze onderwerpen eigenlijk niet met elkaar samenhangen. In dit onderzoek is er bij de andere deelvragen wel gewerkt met deze indeling van de items binnen de onderwerpen. Bij vervolgonderzoek moeten de items van deze onderwerpen dus anders ingedeeld worden binnen de leerfunctie, zodat de Cronbach's alpha hoger wordt en de andere analyses van het onderzoek betrouwbaarder. Ook bij de analyse van interbeoordelaarsbetrouwbaarheid is duidelijk geworden dat de MILK verbeterd kan worden. Het verschilt namelijk per leermiddel of er overeenkomsten zijn tussen beoordelaars. Deze verschillen kunnen zijn ontstaan doordat er verschillende factoren zijn die het gebruik van leermiddelen kunnen beïnvloeden, zoals de kennis en de opvattingen van een beoordelaar (de Vos, 1998). Deze verschillen kunnen ook zijn ontstaan doordat het moeilijk is om een evaluatie-instrument te maken dat bruikbaar is voor alle leermiddelen in alle situaties (Reints, 2000). Een evaluatie-instrument om de kwaliteit van leermiddelen te beoordelen is dus niet gemakkelijk om te maken.

Referenties

- Brakenhoff, J., & Homminga, S. (1995). *Ontwikkelingspsychologie voor het onderwijs*. Groningen: Wolters-Noordhoff.
- Dale, E., & Chall, J. S. (1949). The Concept of Readability. In E. Dale (Ed.), *Readability* (pp. 1-49). Illinois: National Council of Teachers of English.
- De Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., et al. (1998). Acquiring knowledge in science and mathematics : the use of multiple

- representations in technology-based learning environments. In M.W. van Someren, P. Reimann, H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with multiple representations* (pp. 9-40). Oxford: Elsevier Science, Ltd.
- De Vos, W. A. (1998). *Het leermiddelengebruik op basisscholen*. Maastricht: Shaker Publishing B.V.
- Gagné, R. M. (1970). *The conditions of learning*. 2nd edition. New York: Holt, Rinehart and Winston.
- Grossman, P., & Thompson, C. (2008). Learning from curriculum materials: scaffolds for new teachers? *Teaching and Teacher Education*, 24, 2014-2026.
- Hoobroeckx, F., & Haak, E. (2002). *Onderwijskundig ontwerpen: Het ontwerp als basis voor leermiddelenontwikkeling*. Houten-Diegem: Bon Stafleu van Loghum.
- Howell, D. C. (2007). *Statistical methods for Psychology. Sixth edition*. Belmont: Thomson Wadsworth.
- Jonassen, D., Howland, J., Marra, R., & Crismond, D. (2008). *Meaningful learning with technology. Third edition*. New Jersey: Pearson Education, Inc.
- Klare, G. (2000). The measurement of readability: useful information for communicators. *ACM Journal of Computer Documentation*, 24 (3), 107-121.
- Krus, P. H., Thurlow, M. L., Turnure, J. E., & Taylor, A. M. (1976). Summative evaluation of instructional materials. *Studies in educational evaluation*, 2, 121-130.
- Leshin, C. B., Pollock, J., & Reigeluth, M. (1992). *Instructional design, strategies and tactics*. Educational Technology Publications: New Jersey.
- Lester, J. H., Cheek, Jr., Earl, H. (1998). The 'real' experts address textbook issues. *Journal of Adolescent & Adult Literacy*, 41, 282-291.
- Lujan, H. L., & DiCarlo, S. E. (2006). First-year medical students prefer multiple learning styles. *Advances in Physiology Education*, 30, 13-16.
- Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, CO: Mid-continent Research for Education and Learning
- Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*, 13, 125-139.
- Medley-Mark, V., & Weston, C. B. (1988). A comparison of student feedback obtained from three methods of formative evaluation of instructional materials. *Instructional Science*, 17, 3-27.
- Min, R. (2000). *Soorten leermiddelen: leermiddelen, instructiemiddelen, communicatiemiddelen en hulpmiddelen*. Gevonden op 18 november 2009, op <http://projects.edte.utwente.nl/pi/Teksten/Leermiddelen.html>
- Ministerie van Onderwijs, Cultuur & Wetenschappen (2009). *Innovatie in het onderwijs*.

- Wikiwijs. Gevonden op 4 januari 2010, op <http://www.minocw.nl/innovatieinhetonderwijs/1554/Wikiwijs.html>
- Nunally, J. C. (1978) *Psychometric Theory*. New York: McGraw-Hill.
- Pennings, L. (2006). *Aspecten van een kwaliteitssysteem voor educatieve content*. TNO Rapport. Delft: TNO.
- Price, L. (2007). Lecturers' vs. students' perceptions of the accessibility of instructional materials. *Instructional Science*, 35, 317-341.
- Reints, A. J. C. (2000). Een keurmerk voor leermiddelen: waar wachten we nog op. In K. Stokking, G. Erkens, B. Versloot & L. van Wessum (Eds.), *Van onderwijs naar leren: tussen het aanbieden van kennis en het faciliteren van leerprocessen* (pp. 143-156). Leuven-Apeldoorn: Garant.
- Reints, A. J. C. (2008). De invloed van de visie op leren op de rol van leermiddelen in een opleiding. In W. van Dijk & I. Verheul (Eds.), *De verleiding weerstaan: over de noodzaak van het doordacht ontwerpen van opleidingen* (pp. 25-41). Houten: Bohn Stafleu van Loghum.
- Reints, A. J. C., & Wilkens, H. J. (2010). *Evaluating the quality of textbooks from the perspective of the learning process*. Artikel ingediend voor publicatie.
- Sambrook, S., Geertshuis, S., & Cheseldine, D. (2001). Developing a quality Assurance system for computer-based learning materials: problems and issues. *Assessment & Evaluation in Higher Education*, 26, 417-426.
- Saracho, O.N. (1985). Evaluating instructional materials using educational products information exchange product report. *Education*, 106, 78-80.
- Stokking, K. M. (1998). *Bouwstenen voor onderzoek in onderwijs en opleiding*. Utrecht: capaciteitsgroep Onderwijskunde.
- Trochim, W. M. K. (2006). *Introduction to evaluation*. Gevonden op 9 februari 2010, op <http://www.socialresearchmethods.net/kb/intreval.htm>
- Valcke, M. (2005). *Onderwijskunde als ontwerpwetenschap*. Gent: Academia Press.
- Van de Grift, W. (1989). Evaluatie van leermiddelen. In A. J. C. Reints, R. A. de Jong, & N. A. J. Lagerweij (Eds.), *Om de kwaliteit van het leermiddel* (pp. 163-166). Tilburg: Zwijsen.
- Van Merriënboer, J.J.G., & Kirschner, P.A. (2007). *Ten steps to complex learning: a systematic approach to four-component instructional design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Versnel, S. (2009). *Legitimizing stageopdracht CLU: Het funderen van een beoordelingsinstrument op basis van wetenschappelijke literatuur*. Utrecht: CLU.

Bijlage

Naam beoordelaar:

Naam methode:

Helemaal niet mee eens	Een beetje mee eens	Grotendeels mee eens	Helemaal mee eens	Niet te beoordelen	Niet van toepassing	Wegingsfactoren
1	2	3	4	5	6	

A Kwaliteit van de content

1. Modaliteiten **Keuze van modaliteiten¹ is in overeenstemming met doelgroep/leerdoel.**

1 *Teksten*

- De gebruikte teksten bevatten informatieve feiten
- De gebruikte teksten dienen tot instructie
- De gebruikte teksten moeten leerlingen overtuigen
- De gebruikte teksten moeten leerlingen vermaken
- Verschillende soorten teksten wisselen elkaar af

2 *Statische (stilstaande) en dynamische (bewegende) beelden²*

- De gebruikte beelden zijn alleen decoratief*
- De gebruikte beelden ondersteunen de leerstof
- De gebruikte beelden leggen relaties tussen verschillende onderdelen van de leerstof
- De gebruikte beelden geven uitleg bij moeilijke begrippen in het leermiddel
- De gebruikte beelden bevatten ezelsbruggetjes
- Verschillende soorten beelden wisselen elkaar af
- Dynamische en statische beelden wisselen elkaar af

3 *Geluid*

- Teksten worden in een gesproken vorm aangeboden
- De geluidseffecten passen bij de leerstof
- Het achtergrondgeluid is relevant

4 *Authentiek materiaal*

- Teksten worden in een authentieke vorm aangeboden, zoals bijvoorbeeld krantenartikelen of een alinea uit een leesboek
- Beelden worden in een authentieke vorm aangeboden, zoals bijvoorbeeld een tv-programma of een reclameposter
- Geluid wordt in een authentieke vorm aangeboden, zoals bijvoorbeeld cd's of geluidsfragmenten van een tv-programma
- Er wordt materiaal gebruikt dat tastbaar is voor de leerlingen, zoals bijvoorbeeld een leskist met authentiek materiaal

¹ De term modaliteiten heeft betrekking op de 'verpakings'wijze van de leerstof. In welke vorm wordt leerstof aangeboden? In de vorm van tekst, beeld, geluid, authentiek materiaal of een combinatie van dit alles?

² Onder beelden verstaan we: illustraties, pictogrammen, foto's, grafieken en filmpjes

	1	2	3	4	5	6	Wegingsfactoren
	Helemaal niet mee eens	Een beetje mee eens	Grotendeels mee eens	Helemaal mee eens	Niet te beoordelen	Niet van toepassing	
5 Multimodaliteit							
- In het leermiddel worden meerdere modaliteiten aangeboden							
- Tekst en beeld komen beide voor							
- Beeld en geluid komen beide voor							
- De leerstof in tekst en beeld komen met elkaar overeen							
- De leerstof in beeld en geluid komen met elkaar overeen							
- De relatie tussen tekst en beeld is duidelijk voor de leerling							
- De relatie tussen beeld en geluid is duidelijk voor de leerling							
- Tekst en beeld worden dicht bij elkaar gepresenteerd							
- Beeld en geluid worden tegelijkertijd gepresenteerd							
- Bij de opdrachten in het leermiddel wordt gebruik gemaakt van verschillende passende modaliteiten							
6 De gebruikte modaliteiten sluiten aan bij de doelgroep en de leerdoelen							
- De gebruikte modaliteiten houden rekening met de verschillen binnen de doelgroep							
- De gebruikte modaliteiten houden rekening met de ervaringen uit het dagelijks leven van de lerenden							
- De gebruikte modaliteiten passen bij het leerniveau van de lerenden							
- De gebruikte modaliteiten passen bij de leeftijd van de lerenden							
- De gebruikte modaliteiten ondersteunen het bereiken van de leerdoelen							
- De gebruikte modaliteiten bevatten alleen relevante informatie							
Totaal							100

