

De psychometrische kwaliteiten van het STIP instrument

differentiatie in het basisonderwijs

Auteur: Marlou van Nus

Studentnummer: s1290843

Eerste begeleider: M. Hulsbeek MSc.

Tweede begeleider: Dr. T.H.S. Eijssink

Master Psychologie Instructie, Leren en Ontwikkeling

Universiteit Twente

Voorwoord

Voor u ligt de masterthesis waarmee ik mijn studie Psychologie, Instructie, Leren en Ontwikkeling afrond. Ik heb veel plezier beleefd aan het onderzoek doen naar de psychometrische kwaliteiten van een observatie-instrument dat differentiatie meet. Ook waren de colleges van enthousiaste docenten leerzaam voor mij. Allereerst zou ik graag mijn begeleiders Manon Hulsbeek en Tessa Eijnsink willen bedanken voor hun inzet en de vele uren die ze aan mij hebben besteed. Ook bedank ik hen voor het vertrouwen hebben in mij voor het voltooien van dit onderzoek en de opbouwende feedback. Daarnaast wil ik Pascal Wilhelm bedanken voor zijn hulp met betrekking op de statische methoden. Tevens bedank ik alle leerkrachten die mij hun vertrouwen gaven om bij hen lesobservaties af te nemen en de onderwijsexperts voor hun expertise. Ten slotte bedank ik mijn vriend Stijn, mijn ouders, mijn zus Charlotte en zwager Felix voor de interesse en steun die zij mij boden.

Marlou van Nus

Augustus, 2015

Samenvatting

Rekening houden met verschillen tussen leerlingen is kenmerkend voor goed onderwijs en wordt aangeduid als differentiatie. Sinds de invoering van het passend onderwijs is differentiatie nog noodzakelijker dan voorheen. Leerkrachten zijn in veel gevallen (nog) niet in staat om het onderwijs dat zij geven optimaal op de verschillende leerstijlen en behoeften van de leerlingen af te stemmen. Om te kunnen meten in hoeverre leerkrachten differentiëren in hun lessen is het STIP instrument ontwikkeld. Dit observatie-instrument meet in welke mate de leerkracht differentieert tussen zowel de hoog en de laag presterende leerlingen. Het doel van dit onderzoek is om de interbeoordelaarsbetrouwbaarheid, de externe- en de inhoudsvaliditeit van het STIP instrument te bepalen. Om deze psychometrische kwaliteiten te onderzoeken hebben onderwijsexperts het STIP instrument beoordeeld. Er zijn twintig lessen rekenen/wiskunde geobserveerd op zes verschillende scholen. Om de interbeoordelaarsbetrouwbaarheid van het instrument te bepalen beoordeelde de tweede beoordelaar tien lessen met behulp van het STIP instrument. Om de externe validiteit van het instrument te bepalen is het STIP instrument vergeleken met een soortgelijk instrument (ICALT). Door de derde beoordelaar tien lessen te laten beoordelen aan de hand van de ICALT is de externe validiteit bepaald. Ten slotte is aan drie onderwijsexperts gevraagd om hun oordeel te geven over de inhoud van het STIP instrument. Uit de resultaten is naar voren gekomen dat 1) de interbeoordelaarsbetrouwbaarheid hoog was, (overall score van $k = .763$), 2) een lage externe validiteit had vanwege de negatieve, niet-significante correlaties tussen de ICALT en het STIP instrument ($R_s = -.346$) en 3) een hoge inhoudsvaliditeit toonde vanwege de gemiddeld hoge beoordelingen en overeenstemming van de onderwijsexperts, op basis van het belang van differentiatie ($ICC = .517$ en $M = 4.34$) en het taalgebruik van de items ($ICC = .167$ en $M = 5.79$). Op basis van de resultaten kunnen aanbevelingen worden gedaan ter verbetering voor betrouwbaarheid en validiteit van het STIP instrument. Een aanbeveling is om een aantal items aan te passen zodat deze items meer rekening houden met de hoog en de laag presterende leerlingen. Tevens is een aanbeveling om het taalgebruik van een aantal items aan te passen waardoor het duidelijker is voor de beoordelaar om het construct 'differentiatie' te bepalen met het STIP instrument.

Keywords: differentiatie, betrouwbaarheid, validiteit, STIP instrument, ICALT, observatie-instrument.

Abstract

Every student is different in terms of their ability, learning and achievement. Nowadays it is considered a key skill for teachers to adapt their teaching to these differences, so that all students have the best possible chance of learning. However, most teachers find it difficult to adapt their lessons in such a way that they meet the diverse learning needs of their students. In order to find out to what degree teachers of primary schools create a differentiated instruction the STIP instrument has been developed. The STIP instrument is an observation instrument that measures the extent to which teachers are using differentiated instruction. The aim of this study is to measure the psychometric properties –the *reliability* and *validity*- of the STIP instrument. At first, the STIP instrument has been used in twenty different classes at six different schools. In order to measure the interrater reliability, the instrument has been used ten times by a second rater. The external validity has been tested by comparing the STIP instrument with items of a similar instrument (ICALT). A third rater has used the ICALT ten times. Lastly, the content validity has been tested by three experts in the field of education. The results show 1) a high interrater reliability (an overall score of $k = .763$), 2) a low external validity between the ICALT and the STIP instrument ($R_s = -.346$) and 3) a high content validity between the experts based on two dimensions: the importance of the item measuring differentiation ($ICC = .517$ and $M = 4.34$) and the accuracy of the language used to describe differentiation ($ICC = .167$ and $M = 5.79$). Based on the results there are a few recommendations to improve the validity and reliability of the STIP instrument. For example: a couple of items should be reconsidered as they did not seem relevant in relation to the construct ‘differentiation’, also a few items should be described more precise in order to define the construct ‘differentiation’ better. In conclusion, by measuring the implementation of the STIP instrument a better understanding of how and why the STIP instrument works and to what extent it can be improved has been gained.

Keywords: differentiation, reliability, validation, STIP instrument, ICALT, observation-instrument.

Inhoudsopgave

1. Inleiding	1
1.1 Aanleiding	1
2. Theoretisch kader	1
2.1 Differentiatie	1
2.2 Differentiatie in het basisonderwijs	5
2.3 Het STIP instrument	6
2.4 Onderzoeksvraag en hypothesen	7
3. Methode	9
3.1 Participanten	9
3.2 Domein	9
3.3 Meetinstrumenten	9
3.4 Procedure	10
3.5 Data-analyse	10
3.5.1 Interbeoordelaarsbetrouwbaarheid	10
3.5.2 Externe validiteit	11
3.5.3 Inhoudsvaliditeit	11
4. Resultaten	13
4.1 Betrouwbaarheid	13
4.1.1 Interbeoordelaarsbetrouwbaarheid	13
4.2 Validiteit	14
4.2.1 Externe validiteit	14
4.2.2 Inhoudsvaliditeit	14
4.2.2.1 Vraag a	15
4.2.2.2 Vraag b	17
5. Discussie	19
5.1 Opzet van het onderzoek	19
5.2 Psychometrische kwaliteiten van het STIP instrument	19
5.2.1 Interbeoordelaarsbetrouwbaarheid	19
5.2.2 Externe validiteit	20
5.2.3 Inhoudsvaliditeit	20
5.2.3.1 Vraag a	21
5.2.3.2 Vraag b	23

5.2.4 Aanbevelingen met betrekking op de schaalmethode	24
5.2.5 Conclusie	24
5.3 Vervolgonderzoek	24
Literatuur	26
Bijlage I. Het STIP instrument	33
Bijlage II. Items van het STIP instrument per schaal	43
Bijlage III. Schema van de analyse ICALT - STIP instrument	45

1. Inleiding

1.1 Aanleiding

Sinds 1 augustus 2014 is door de Nederlandse overheid het passend onderwijs ingevoerd. Het uitgangspunt van passend onderwijs is dat alle kinderen een plek moeten krijgen op een school die past bij hun kwaliteiten en mogelijkheden, ook als zij extra ondersteuning nodig hebben (Ministerie OCW, 2014). Kort gezegd moeten scholen hun onderwijs afstemmen op de verschillen die er bestaan tussen leerlingen. Iedere leerling verschilt van de ander in mogelijkheden, kennis, vaardigheden en motivatie. Om rekening te houden met al deze factoren is een oproep gedaan aan alle leerkrachten om hun onderwijs aan te bieden op het niveau van de leerling zodat elke leerling de kans krijgt om zo goed mogelijk te leren (Darling-Hammond, Wise & Klein, 1999; Ducette, Sewell, & Shapiro, 1996; Gamoran & Weinstein, 1995; Schoenfeld, 1999). Het omgaan met verschillen tussen leerlingen is geen nieuwe ontwikkeling voor scholen. Eerder al hebben zij te maken gehad met vernieuwingen zoals onderwijs op maat en omgaan met verschillen tussen leerlingen. Nieuw is echter het feit dat scholen vanaf de invoering van passend onderwijs nu ook een zorgplicht hebben gekregen. Dat betekent dat scholen ervoor verantwoordelijk zijn om alle leerlingen (dus ook de leerlingen die extra ondersteuning nodig hebben) een passende plek te bieden. Dat geldt zowel voor leerlingen die nieuw worden aangemeld als voor leerlingen die nu al op school zitten. Dat kan op de eigen school zijn, of, als de school niet in staat is om passende begeleiding te bieden, op een andere reguliere of speciale school. Door deze zorgplicht worden scholen gedwongen om na te denken over hun ondersteuningsprofiel: welke leerlingen kunnen zij nog wel een passende plek bieden en welke leerlingen niet? En wat betekent dat voor de leerkrachten van de school? Om ervoor te zorgen dat iedere leerling het beste uit zichzelf kan halen moeten leerkrachten nog beter in staat zijn om met de verschillende (leer)behoeften van leerlingen om te gaan. Het afstemmen van het onderwijs op de verschillen tussen leerlingen valt onder de term differentiatie (Inspectie van het onderwijs, 2014).

2. Theoretisch kader

2.1 Differentiatie

Differentiatie wordt gedefinieerd als een pedagogische lesgeefbenadering waarbij de leerkracht met verschillende lesgeefmethoden, leeractiviteiten en leermiddelen aansluit op de (leer)behoeften van individuele leerlingen zodat het beste uit de leerling wordt gehaald (Bearne,

1996; Levy, 2010; Stradling & Saunders, 1993; Tomlinson, 1999). Volgens de Inspectie van het onderwijs (2014) differentiëren leerkrachten voornamelijk door uitleg en opdrachten af te stemmen op het niveau van de leerling, gebruik te maken van de mogelijkheden die leermiddelen bieden en de beschikbare lestijd voor de individuele leerling te benutten. Dit betekent dat de rol van de leerkracht met name ligt bij het ondersteunen van de leerling (goede lesinstructie en passende opdrachten).

Vanuit de literatuur wordt de nadruk gelegd op differentiatie tijdens de instructie. De leerkracht is bij een gedifferentieerde instructie continu bezig om het niveau van de instructie aan te passen aan de behoeften, leerstijlen en/of interesses van de individuele leerlingen (Hall, 2002; Heacox, 2002; Sternberg, Torff & Grigorenko, 1998; Watts-Taffe et al., 2012). Tomlinson (2000) stelt dat leerkrachten bij deze vorm van instructie differentiëren op basis van vier elementen: inhoud, taak, proces en leeromgeving.

1) Het eerste element *inhoud* refereert aan de leerinhoud, die kan verschillen in niveau. Differentiatie in inhoud zorgt voor goed onderwijs als leerlingen van verschillende competentieniveaus werken aan verschillende leerinhouden. Volgens Van der Valk (2014) krijgen alle leerlingen de basisinhoud aangeboden en kan optioneel verdieping, verbreding of verlenging van de lesinstructie worden toegepast. De hoog presterende leerlingen kunnen de leerinhoud uitbreiden naar een hoger niveau (verdieping) of naar meer onderwerpen (verbreding). Bij verdieping van de opdracht kunnen bijvoorbeeld nieuwe begrippen aan de leerinhoud worden toegevoegd die aansluiten bij de basisinhoud. Bij het bieden van verbredende leerinhoud wordt de leerinhoud aangeboden op hetzelfde niveau maar wordt het uitgebreid naar een ander onderwerp. Voor de laag presterende leerlingen kan een verlengde instructie plaatsvinden waarbij de leerinhoud wordt herhaald. Door meer verdieping en/of verbreding aan de leerinhoud toe te voegen en/of een langere instructietijd te bieden kan een leerkracht in zijn of haar les inhoudelijk differentiëren naar zowel de hoog als laag presterende leerlingen. Uit onderzoek is gebleken dat het bieden van uitdaging op het niveau van de leerling, de leerling in de zogenoemde ‘zone van naaste ontwikkeling’ komt (Vygotsky, 1986). Door een leerling te begeleiden bij de onderdelen van de leerinhoud die hij/zij nog net niet zelfstandig kan uitvoeren, kan een leereffect ontstaan op het moment dat ze de inhoud zelf gaan begrijpen en beheersen. Als de leerling wordt begeleid om in de zone van naaste ontwikkeling te zitten zal de leerling succesvoller en gemotiveerder zijn. Ook uit neurologisch onderzoek is gebleken dat het leren wordt gestimuleerd als de leerling op zijn eigen niveau werkt en uitgedaagd wordt (Howard, 1994; Jensen, 1998; Sousa, 2001; Wolfe, 2001). Ook zal een gedifferentieerde instructie de leerling moeten helpen om relevante informatie te selecteren, organiseren en integreren waarmee

uiteindelijk betekenisvol leren wordt gestimuleerd (Mayer, 2005).

2) Het tweede element van Tomlinson (2000), *taak*, heeft voornamelijk betrekking op het eindproduct. Het aanbieden van verschillende verwerkingsopdrachten is essentieel om te kunnen differentiëren in taak (Tomlinson et al., 2003). Differentiatie in taak zorgt voor goed onderwijs wanneer de leerlingen verschillende opdrachten maken op hun eigen competentieniveau. De hoog presterende leerlingen hebben voorkeur voor ongestructureerde en flexibele leertaken (Griggs & Dunn, 1984). De leerkracht kan deze leerlingen als werkopdracht geven om de leerinhoud te presenteren, op deze manier worden de leerlingen gestimuleerd om hun eigen inbreng en geleerde kennis te implementeren in de taak. De laag presterende leerlingen zullen een duidelijke en gestructureerde opdracht moeten krijgen van de leerkracht, zoals het schrijven van een brief of een verslag. Bouwman, Brouwer, Jansen en Loman (2014) zeggen dat de leerkracht stapsgewijs de taak uit moet leggen zodat de laag presterende leerlingen de taak goed begrijpen en kunnen voltooien. In het onderzoek van Chamorro-Premuzic, Furnham en Lewis (2007) wordt een sterk verband gevonden tussen het aanbieden van meerdere werkopdrachten en het aansluiten op verschillende competentieniveaus van de leerlingen. Als de werkopdracht aansluit op het competentieniveau van de leerling zal het leereffect groter zijn. Om dit proces te bevorderen zal de leerkracht hogere en lagere orde denkvaardigheden moeten stimuleren bij de leerlingen. Het onthouden en ophalen van informatie worden geclassificeerd bij lagere orde denkvaardigheden en het analyseren, synthetiseren, evalueren en kritisch denken zijn hogere orde denkvaardigheden (Bloom, 1956; Mayer, 2005; Zohar & Dori, 2003). Bij de laag presterende leerling worden de lagere orde denkvaardigheden gestimuleerd en bij de hoog presterende leerling hogere orde denkvaardigheden. Er wordt verondersteld dat de hoog presterende leerlingen de lagere orde denkvaardigheden al beheersen en uitgedaagd kunnen worden voor moeilijke leertaken (Shepard, 1991). De meeste leerkrachten leggen bij gedifferentieerde instructie vaak de nadruk op het onthouden van de kennisstof (Mayer, 2005) omdat de aandacht meestal uitgaat naar de laag presterende leerlingen (Zohar, Degani en Vaaknin, 2001). Differentiëren naar taak stimuleert ook het creatief denken doordat verschillende taken uitgevoerd kunnen worden die op verschillende manieren aansluiten op de niveaus van de leerlingen (Awang & Ramly, 2008). Er wordt beargumenteerd dat op deze manier leerlingen de leertaak beter zullen begrijpen, analyseren en toepassen.

3) Het derde element van differentiatie volgens Tomlinson (2000), *proces*, verwijst naar de manier waarop de leerkracht de leerinhoud aanbiedt en hiermee omgaat: het stimuleren van verschillende leerstijlen en de vraag of leerlingen in groepjes of individueel aan het werk gaan. Leerlingen kunnen namelijk verschillende voorkeuren hebben voor de manier waarop ze kennis

tot zich nemen (Kolb, 1984; Honey & Mumford, 1992). De hoog presterende leerlingen hebben met vaak voorkeur voor het ontdekkend leren, waarbij ze zelf willen ontdekken en ervaren (Chan, 2001) en de leerkracht een begeleidende rol heeft. De laag presterende leerlingen hebben daarentegen vaak een voorkeur voor directe instructie. De leerkracht draagt bij deze instructiebenadering de informatie over aan de laag presterende leerlingen (Klahr, Triona, & Williams, 2007) en heeft een actieve rol in het proces om de leerinhoud zo goed mogelijk aan te bieden (Chan, 2001). Als de leerkracht rekening houdt met de verschillende leerstijlen zullen de leerprestaties van de leerling verbeteren (Dunn, Beaudry, & Klavas, 1989; Griggs, 1984; Griggs & Dunn, 1984; Grigorenko & Sternberg, 1997). Ook de leerlingen in groepjes laten werken is effectief als de instructie aansluit op de interesses en leerstijlen van de leerling (McLaughlin & Talbert, 1993). Voor de hoog presterende leerlingen is het effectief om in homogene groepen te werken omdat gelijkgestemden elkaar uitdagen en hierdoor de kans verlaagd voor onderpresteren. Voor de laag presterende leerlingen is het effectief om in heterogene groepen te werken, zodat zij ook door de hoog presterende leerlingen worden uitgedaagd (Van der Valk, 2014). Het is van belang dat de leerkracht afwisselend werkt met homogene en heterogene groepen om te voldoen aan beide behoeften. Differentiatie in proces zorgt voor goed onderwijs wanneer de laag presterende leerling directieve procesondersteuning van de leerkracht krijgt en de hoog presterende leerling een coachende procesondersteuning krijgt. Om aan te sluiten op de hogere en de lagere orde denkvaardigheden zal volgens Good (1988) de leerkracht directe, gesloten vragen moeten stellen aan de laag presterende leerlingen en moeilijkere, open vragen aan de hoog presterende leerlingen. Bij de vragen voor de hoog presterende leerlingen wordt het plannen, monitoren van het geleerde gestimuleerd (stimulatie van het zelfregulerend leren). De leerkracht zal met name de hoog presterende leerlingen stimuleren tot zelfregulerend leren omdat deze leerlingen de taken op basis van metacognitieve vaardigheden kunnen voltooien (Zimmerman & Martinez-Pons, 1990). Het beheersen van metacognitieve vaardigheden is een vereiste voor het zelfregulerend leren (Zimmerman, 1986).

4) Het vierde element volgens Tomlinson (2000), *leeromgeving*, heeft betrekking op de setting waar het leren plaatsvindt: de indeling van de klas (fysieke component) en de werksfeer (psychologische component). Differentiatie in leeromgeving zorgt ervoor dat de hoog presterende leerlingen in een 'plusklas' aan het werk kunnen en de laag presterende leerlingen in het huidige klaslokaal (Van der Valk, 2014). Ook kan de leerkracht differentiëren in leeromgeving door verschillende materialen aan te bieden. De leerkracht kan concreet materiaal aanbieden voor de laag presterende leerlingen zodat deze leerlingen de lesinhoud beter zullen begrijpen (lagere orde denkvaardigheid). Daarnaast kan de leerkracht een evaluatieplan aan de

hoog presterende leerlingen geven zodat deze leerlingen worden gestimuleerd om de lesinhoud te evalueren (hogere orde denkvaardigheid). Naast deze fysieke componenten kan de leerkracht ook rekening houden met de psychologische component. De interactie tussen de leerkracht en de leerlingen is van belang voor een positieve werksfeer in de klas (Taylor, Fraser & Fisher, 1997). Voor een positieve werksfeer zal de leerkracht zowel de hoog als de laag presterende leerlingen positief gaan benaderen. Als de leerlingen de leeromgeving als positief en aangenaam ervaren zijn ze beter bereid te leren en zullen ze hogere prestaties bereiken dan als ze de leeromgeving minder positief en onaangenaam ervaren (Pimparon, Roff, McAlleer, Poonchai & Pemba, 2000). Ook zal de leerkracht aan moeten geven wat de hoog presterende leerlingen kunnen doen tijdens de uitgestelde aandacht als de leerkracht de laag presterende leerlingen begeleidt. Door duidelijkheid te creëren weten de hoog presterende leerlingen wat ze kunnen verwachten, dit draagt bij aan een prettige leeromgeving. Concluderend zal de leerkracht met de fysieke en psychologische componenten van de leeromgeving rekening moeten houden om aan te sluiten op de onderwijsbehoefte van de leerling (Clijsen, 2007; Pimparon et al., 2000).

2.2 Differentiatie in het basisonderwijs

Hoe om te gaan met verschillen tussen leerlingen? Dat is een belangrijke vraag die het onderwijs zichzelf stelt (Bosker, 2005). Volgens Doolaard en Harms (2013) staat of valt differentiatie met basisvaardigheden van de leerkracht: klassenmanagement en goede instructie. In de praktijk wordt het IGDI-model hiervoor vaak toegepast. Dit model is leerkracht-gestuurd en gericht op het bieden van interactie. Binnen het IGDI-model zal de leerkracht met name extra aandacht geven aan de laag presterende leerlingen (Ahlers, 2009). Ondanks dat de meeste basisscholen met het IGDI-model werken, heeft een groot deel van de scholen moeite om hun lessen voldoende op verschillende onderwijsbehoeften van leerlingen af te stemmen (Inspectie van het onderwijs, 2014). Vooral op het integreren van gedifferentieerde instructie scoren leerkrachten onvoldoende; 40% van de leerkrachten past dit toe. Dit betekent dat de leerkracht in het merendeel van de gevallen (nog) niet in staat is om zijn of haar onderwijs optimaal op de verschillende leerstijlen en leerbehoeften voor de leerlingen af te stemmen.

Volgens de Inspectie van het onderwijs (2014) zijn er drie redenen waarom veel leerkrachten te weinig differentiëren in het basisonderwijs. Ten eerste ervaren leerkrachten vaak een hoge werkdruk en een gebrek aan faciliteiten waardoor ze niet altijd differentiatie toepassen. Een leerkracht is in veel gevallen een duizendpoot die veel ballen hoog moet houden. Leerkrachten die zich willen bekwamen in differentiatievaardigheden zien niet hoe ze dat kunnen doen zonder dat hun werkdruk verhoogt. Door de hoge werkdruk wordt differentiatie

door de leerkrachten gezien als struikelblok. Ten tweede worden niet alle leerkrachten tijdens hun lerarenopleiding voldoende voorbereid op differentiatie. Ongeveer de helft van de beginnende leerkrachten geeft aan te weinig te hebben geleerd vanuit de leerkrachtenopleiding om individuele lesprogramma's toe te passen aan de laag presterende leerlingen (43%) en de hoog presterende leerlingen (53%) (Onderwijsinspectie, 2015). Ten slotte hebben de meeste leerkrachten een stereotype beeld van een hoog presterende leerling: dat ze overtuigend en zelfverzekerd zijn (Gross, 1997). Dit beeld kan leiden tot een negatieve houding waarbij de leerkrachten zich weerhouden van het differentiëren bij de hoog presterende leerling (Eyre & Geake, 2002). Op deze manier kan het leerproces van de hoog presterende leerling zich niet optimaal voortzetten omdat het onderwijs niet is afgestemd op de leerbehoefte van alle leerlingen (Tschannen-Morana & Woolfolk Hoyb, 2001).

2.3 Het STIP instrument

Naast de argumentatie van de leerkrachten waarom ze zelf vinden dat ze nog niet genoeg differentiëren, is het belangrijk dat er ook een objectief beeld wordt geschetst over de mate waarin differentiatie wordt toegepast in het basisonderwijs. Om dit beeld in kaart te brengen kan differentiatie gemeten worden door middel van een observatie-instrument. Op dit moment is er op de Nederlandse markt geen observatie-instrument dat differentiatie meet. Om deze reden is een nieuw observatie-instrument ontwikkeld genaamd het *STIP instrument*. Het STIP instrument meet in welke mate de leerkracht differentieert tussen zowel de hoog en de laag presterende leerlingen. Het STIP instrument bestaat uit drie onderdelen: een lesbeschrijving, een leerkrachtobservatie en een interview (zie Bijlage I). 1) Bij het eerste onderdeel, lesbeschrijving, worden de methode en de organisatie van de les genoteerd. Ook kunnen eventuele uitspraken of vragen van de leerkracht worden opgeschreven. Tevens kunnen eventuele uitspraken of vragen van leerlingen of onverwacht gedrag van een leerling worden genoteerd; 2) Het tweede onderdeel is de leerkrachtobservatie. De leerkracht wordt op basis van vier schalen geobserveerd. Deze schalen zijn ontleend uit Tomlinson (2000): differentiatie in *inhoud*, *taak*, *proces* en *leeromgeving*. Elke schaal bevat een aantal STIP categorieën met bijbehorende items. 1) *Inhoud* bevat zeven items van gedragsverwachtingen van de leerkracht. Deze items vallen onder de volgende STIP categorieën: het aanbieden van de leerinhoud, relaties leggen met andere leerinhouden en het gebruik maken van verschillende bronnen. 2) *Taak* bevat zes items met gedragsverwachtingen van de leerkracht onder de volgende STIP categorieën: het afstemmen van de opdrachten op leerbehoeften van de leerlingen en het stimuleren van creatief denken. 3) Daarnaast omvat *proces* twaalf items met gedragsverwachtingen van de leerkracht.

Deze items vallen onder de volgende STIP categorieën: de leerlingen groeperen op basis van hun mogelijkheden of interesses, de leerinhoud op verschillende manieren uitleggen, de leerlingen leerinhoud laten ontdekken, verschillende manieren van feedback geven en het stimuleren van verschillende manieren om leerlingen kritisch na te laten denken. 4) Ten slotte omvat *leeromgeving* acht items met gedragsverwachtingen van de leerkracht. Deze items vallen onder de volgende STIP categorieën: het lokaal inrichten zodat het uitnodigt tot samenwerken, zorgen voor goede voorbereiding van de les, de leerlingen aanmoedigen en accepteren van hun gevoelens en de ideeën van leerlingen accepteren of gebruiken. Hoge verwachtingen hebben van de leerlingen en ze betrekken bij het plannen, monitoren en vaststellen van hun leren of geleerde en reflecteren op het geleerde. Deze STIP categorieën geven informatie over de setting waarin het leren van de leerlingen plaatsvindt (een fysieke en psychologische component); 3) Het derde onderdeel van het STIP instrument zijn interviewvragen. Deze interviewvragen zijn geïmplementeerd ter bevestiging of tegenspraak van het geobserveerde gedrag van de leerkracht. Vanzelfsprekend worden de interviewvragen na de leerkrachtobservatie gesteld aan de leerkracht. De inhoud van de vragen gaan over het activeren van voorkennis, het doel van de les, de koppeling met de vorige en volgende les, hoe de leerkracht rekening houdt met verschillen tussen leerlingen en of leerlingen ander werk mee krijgen naar aanleiding van de les.

2.4 Onderzoeksvraag en hypotheses

Als differentiatie in kaart wordt gebracht met behulp van het STIP instrument, is het noodzakelijk om te weten of dit instrument betrouwbaar en valide is. In dit onderzoek staat daarom de volgende onderzoeksvraag centraal:

In welke mate is het STIP instrument een betrouwbaar en valide observatie-instrument voor het meten van differentiatie bij leerkrachten in het basisonderwijs?

Om deze psychometrische kwaliteiten van het STIP instrument te onderzoeken worden verschillende elementen onderzocht: de interbeoordelaarsbetrouwbaarheid, de externe validiteit en de inhoudsvaliditeit. 1) Het eerste element dat onderzocht wordt is het bepalen van de interbeoordelaarsbetrouwbaarheid. Om de interbeoordelaarsbetrouwbaarheid te bepalen zullen twee beoordelaars differentiatie beoordelen met behulp van het STIP instrument. De eerste beoordelaar zal in de les differentiatie meten aan de hand van het STIP instrument. De tweede beoordelaar zal aan de hand van tien audio-opnames de mate van differentiatie bepalen met het STIP instrument. Volgens Van Yperen en Veerman (2008) betekent een hoge interbeoordelaarsbetrouwbaarheid dat het niet uitmaakt welke persoon het instrument gebruikt,

zodat degene die het instrument hanteert geen invloed hoort te hebben op de onderzoeksresultaten. Er wordt verwacht dat de interbeoordelaarsbetrouwbaarheid hoog is omdat de ontwikkelaar van het STIP instrument de ontwikkeling van het instrument heeft besproken met twee collega's. Daarnaast is dit instrument tijdens twee lessen op twee verschillende scholen getest. 2) Ten tweede zal de externe validiteit van het STIP instrument gemeten worden. Aan de hand van de externe validiteit kan bepaald worden in hoeverre de onderzoeksresultaten te generaliseren zijn naar andere instrumenten of experimenten (Bracht & Glass, 1968). Een onderzoek uitvoeren in een realistische setting is essentieel voor het meten van externe validiteit van een observatie-instrument volgens Durlak en Dupre (2008). Om de externe validiteit te kunnen bepalen moet het STIP instrument vergeleken worden met een ander instrument dat een soortgelijk construct meet. Het instrument dat hiervoor uitgekozen is, is de International Comparative Analysis of Learning and Teaching (ICALT). Dit observatie-instrument meet de kwaliteit van onderwijs, waarbij een aantal items gerelateerd zijn aan differentiatie. Als twee observatie-instrumenten met elkaar vergeleken worden is het van belang dat de items zoveel mogelijk overlap hebben om hetzelfde domein te meten (Baker, Gersten, Haager & Dingle, 2006). Om deze reden is een analyse van deze itemsselectie gehanteerd waarbij de items over differentiatie worden gekozen. Deze analyse staat weergegeven in Bijlage III. Er wordt een hoge externe validiteit verwacht omdat de ICALT is getest en goedgekeurd door de onderwijsinspectie (Van de Grift, 2007) en de items van het STIP instrument en de ICALT beiden differentiatie meten. 3) Ten slotte wordt de inhoudsvaliditeit van het STIP instrument bepaald door beoordelingen van drie onderwijsexperts. Deze onderwijsexperts beoordelen elk item op basis van het belang voor het meten van differentiatie en de duidelijkheid van het taalgebruik. Volgens Cohen en Swerdlick (2005) toont een hoge inhoudsvaliditeit aan dat het instrument een goede afspiegeling is van het domein. Er wordt een hoge overeenstemming verwacht omdat het STIP instrument is getest tijdens twee lessen op twee verschillende scholen en de ontwikkelaar de ontwikkeling van het instrument heeft besproken met twee collega's. Er wordt een gemiddelde score van vier punten per item verwacht.

3. Methode

3.1 Participanten

In dit onderzoek werden acht scholen van Stichting Katholiek Onderwijs Enschede (SKOE) benaderd. Hiervan gaven zes scholen toestemming en twee scholen gaven geen toestemming voor deelname. De leerkrachten werden geselecteerd in overleg met directeuren van de betreffende basisscholen op basis van de volgende criteria: de leerkrachten dienden rekenwiskundelessen te geven en lesgevend in groep 5, 7 en 8. Groep 6 was achterwege gelaten omdat de leerkrachten van deze groep al aan het STIP-project meewerkten waarbij ze al geobserveerd zijn met het STIP instrument. De hoeveelheid groepen voor dit onderzoek waren in totaal zeven groepen 5, zes groepen 7, zes groepen 8 en één combinatiegroep 7/8. De eerste beoordelaar nam twintig lesobservaties af met behulp van het STIP instrument bij twintig leerkrachten (N = 20). Bij iedere leerkracht werd één lesobservatie afgenomen. De uiteindelijke onderzoeksgroep bestond uit 20 leerkrachten, waarvan 2 mannelijke en 18 vrouwelijke leerkrachten (10% mannen; 90% vrouwen).

3.2 Domein

Er is gekozen voor rekenwiskundelessen omdat verwacht werd dat leerkrachten binnen dit domein zouden differentiëren. Volgens Hill, Schilling en Ball (2004) bieden leerkrachten tijdens rekenwiskundelessen verschillende leerstrategieën aan om aan te sluiten op het niveau van de leerling. Daarnaast wordt in de praktijk tijdens rekenwiskundelessen relatief vaak het IGDI-model gehanteerd. Volgens Vernooy (2001) is het toepassen van het IGDI-model met name effectief als basisvaardigheden op het gebied van lezen en rekenen worden aangeleerd door de leerkracht. Ook is er voor de rekenwiskundelessen gekozen om lessen binnen hetzelfde domein te vergelijken. Om deze redenen is gekozen voor het domein rekenwiskunde.

3.3 Meetinstrumenten

De meetinstrumenten bij dit onderzoek waren het STIP instrument en de ICALT. Het STIP instrument bevat 33 dichotome items, die worden gescoord door 1/0 (1 = het gedrag komt wel voor; 0 = het gedrag komt niet voor), (zie Bijlage I). De lesomschrijving en interviewvragen zijn afgenomen, maar bij de analyse zijn deze twee onderdelen buiten beschouwing gelaten omdat dit onderzoek zich beperkte tot de items van het STIP instrument. De ICALT omvatte de 17 geselecteerde items die betrekking hebben op differentiatie. De analyse staat weergegeven in Bijlage II.

3.4 Procedure

Per leerkracht kwam de observator op de afgesproken tijd naar het klaslokaal. De audioapparatuur werd voorafgaand de rekenwiskundeles op het bureau van de leerkracht gelegd zodat de audioapparaat zo dicht mogelijk bij het stemgeluid van de leerkracht was. Tijdens de leerkrachtobservatie zat de observator achterin de klas te observeren aan de hand van het STIP instrument. De lessen duurden gemiddeld 50 minuten. Na de leerkrachtobservaties stelde de onderzoeker interviewvragen aan de leerkracht. Ten slotte kreeg de leerkracht een presentje mee naar huis als bedankje voor de medewerking.

3.5 Data-analyse

Tabel 1 laat het overzicht zien van de drie manieren van de data-analyse.

Tabel 1. Overzicht data-analyse

Type validiteit / betrouwbaarheid	Beoordelaars	Dataverzameling	Meeteenheid
Interbeoordelaars-betrouwbaarheid	Tweede beoordelaar	Kwantitatieve gegevens: beoordelingen met behulp van het STIP instrument	Cohen's Kappa
Externe validiteit	Derde beoordelaar	Kwantitatieve gegevens: beoordelingen met behulp van de ICALT	Spearman Brown
Inhoudsvaliditeit	Drie onderwijsexperts	Kwantitatieve gegevens: beoordelingen met behulp van het STIP instrument	Intraclasscoëfficiënt

3.5.1 Interbeoordelaarsbetrouwbaarheid

Om de interbeoordelaarsbetrouwbaarheid te bepalen heeft een tweede beoordelaar de lessen beoordeeld aan de hand van het STIP instrument. Deze beoordelaar was een leerkracht van 23 jaar met twee jaar ervaring in het regulier basisonderwijs. De derde beoordelaar is geselecteerd op basis van praktijkervaring als leerkracht binnen het regulier basisonderwijs. De beoordelaar

kreeg een training om de betrouwbaarheid te maximaliseren. Een training kan belangrijke bedreigingen van betrouwbaarheid reduceren (Danielson, 2007). Na de training beoordeelde de tweede beoordelaar tien rekenwiskundelessen (50%). Er is gekozen voor tien rekenwiskundelessen (50%) omdat dit aansloot bij de richtlijn om de interbeoordelaarsbetrouwbaarheid te bepalen. Volgens Koopmans (2006) zal minimaal 10% van de data geanalyseerd moeten worden om de interbeoordelaarsbetrouwbaarheid te bepalen. Aangezien in dit onderzoek 10% maar twee rekenwiskundelessen omvatte, is er gekozen om tien rekenwiskundelessen (50%) te beoordelen. Om tien rekenwiskundelessen te beoordelen is er meer data beschikbaar om de interbeoordelaarsbetrouwbaarheid te bepalen. De codering omvatte de 33 items van het STIP instrument waarbij de lesbeschrijving en het interview buiten beschouwing werden gelaten. De overeenstemming tussen de eerste en tweede beoordelaar werd onderzocht door het bepalen van de Cohen's Kappa.

3.5.2 Externe validiteit

Om de externe validiteit te bepalen heeft een derde beoordelaar differentiatie beoordeeld aan de hand van de 17 geselecteerde items van de ICALT. De derde beoordelaar was een leerkracht van 27 jaar die vier jaar ervaring heeft met lesgeven in het regulier en speciaal basisonderwijs. De beoordelaar is geselecteerd voor dit onderzoek vanwege de praktijkervaring als leerkracht binnen het regulier basisonderwijs. Deze beoordelaar kreeg een training en beoordeelde vervolgens tien rekenwiskundelessen (50%). De overeenstemming tussen de twee observatie-instrumenten is onderzocht door het bepalen van de Spearman Brown. Deze statistische methode is uitgevoerd omdat het STIP instrument en de ICALT dichotome items op ordinaal meetniveau bevatten.

3.5.3 Inhoudsvaliditeit

Om de inhoudsvaliditeit te bepalen beoordeelden drie onderwijsexperts het STIP instrument. De eerste onderwijsexpert (O1) was een docent die 25 jaar werkzaam is op een middelbare school. Deze leerkracht geeft les aan leerlingen van basisberoepsgerichte en de kaderberoepsgerichte leerweg van het VMBO. De tweede onderwijsexperts (O2) was een junior onderzoeker bij het *Institute for Teacher Education, Science Communication & School Practices (ELAN)* aan Universiteit Twente. Ook heeft deze onderwijsexpert ervaring met lesgeven. De derde onderwijsexpert (O3) was een promovendus *Social Intervention at the Centre for Evidence-Based Intervention* aan de Universiteit Oxford. Ook heeft deze onderwijsexpert ervaring met lesgeven aan studenten van een masteropleiding en het observeren van leerkrachten in Groot-Brittannië. In dit onderzoek zijn drie onderwijsexperts geraadpleegd omdat dit voldoende data

zou geven om de inhoudsvaliditeit te bepalen. De onderwijsexperts beoordeelden elk item van het STIP instrument met behulp van twee vragen: in welke mate het item belangrijk is om differentiatie te meten (vraag a) en in welke mate het taalgebruik duidelijk is van het item (vraag b). De onderwijsexperts gaven hun beoordelingen aan op een zevenpunt Likert schaal. Voor vraag a gaven de onderwijsexperts antwoordt (1 = helemaal niet belangrijk; 7 = heel erg belangrijk). Bij vraag b gaven de onderwijsexperts antwoordt (1 = helemaal niet duidelijk; 7 = heel erg duidelijk). De maatstaf voor een goede beoordeling is een minimale vier score op de zevenpunt Likert schaal. Het observatie-instrument Classroom Observation Scale-Revised (COS-R) hanteerde dezelfde wijze om de inhoudsvaliditeit te bepalen (VanTassel-Baska et al., 2005). Zie Figuur 1 voor een voorbeeld van een item met vraag a en b. Om de inhoudsvaliditeit te onderzoeken is de *intraclass coëfficiënt* (ICC) uitgevoerd. Er is gekozen voor de ICC omdat deze statistische maat de mate van overeenstemming van drie beoordelaars van de schaal kan bepalen (Bartko, 1976).

Item 1 is een uitwerking van STIP categorie: de leerkracht bepaalt wat leerlingen al over de te behandelen leerinhoud weten.							
1. De leerkracht vraagt expliciet naar de voorkennis over dit onderwerp: wat weet je er al van? Helpt voorkennis activeren.							
A. In welke mate is dit item belangrijk bij differentiatie?	1	2	3	4	5	6	7
B. In welke mate is het taalgebruik duidelijk?	1	2	3	4	5	6	7

Figuur 1. Item 1 met vraag a en b.

4. Resultaten

4.1 Betrouwbaarheid

4.1.1 Interbeoordelaarsbetrouwbaarheid

Om inzicht te krijgen in de overeenstemming tussen de eerste en tweede beoordelaar is de Cohen's Kappa berekend. Daarnaast is de classificatie van Dooley (2001) gehanteerd voor het beoordelen in hoeverre er sprake is van overeenstemming. Tabel 1 laat de resultaten zien van de overeenstemming tussen de eerste en de tweede beoordelaar voor de beoordelingen van tien rekenwiskundelessen met behulp van het STIP instrument.

Tabel 1. *Interbeoordelaarsbetrouwbaarheid*

Les	K	Dooley
Les 1	.604	Redelijke overeenstemming
Les 2	.763	Goede overeenstemming
Les 3	.809	Goede overeenstemming
Les 4	.752	Goede overeenstemming
Les 5	.511	Redelijke overeenstemming
Les 6	.809	Goede overeenstemming
Les 7	.746	Redelijke overeenstemming
Les 8	.879	Goede overeenstemming
Les 9	.818	Goede overeenstemming
Les 10	.694	Redelijke overeenstemming
Les 1 tot en met 10	.763	Goede overeenstemming

Uit Tabel 1 valt af te lezen dat de interbeoordelaarsbetrouwbaarheid van tien lessen een goede overeenstemming heeft tussen de eerste en de tweede beoordelaar. De overallscore ($k = .763$) is de kappa voor les één tot en met tien. Volgens Dooley (2001) is een waarde van $k > .750$ een goede overeenstemming. Dit veronderstelt dat de overallscore ($k = .763$) een goede overeenstemming is. Ook laat deze analyse een goede overeenstemming zien van les twee ($k = .763$), les drie ($k = .809$) les vier ($k = .752$), les zes ($k = .809$), les acht ($k = .879$) en les negen ($k = .818$). Daarnaast was een redelijke overeenkomst tussen de eerste en tweede beoordelaar van les één ($k = .604$), les vijf ($k = .511$), les zeven ($k = .746$) en les tien ($k = .694$). Volgens Dooley (2001) betekent een waarde van $k > .400 - .750$ een redelijke overeenstemming. Er zijn geen zwakke overeenstemmingen tussen de twee beoordelaars, want een waarde van $k < .40$ is een zwakke overeenstemming volgens Dooley (2001). Concluderend is de overeenstemming tussen

de eerste en tweede beoordelaars goed waarmee veronderstelt kan worden dat de interbeoordelaarsbetrouwbaarheid van het STIP instrument als geheel hoog is.

4.2 Validiteit

4.2.1 Externe validiteit

Om inzicht te krijgen in het verband tussen het STIP instrument en de ICALT is de Spearman Brown uitgevoerd. Er is hiervoor gekozen omdat beide observatie-instrumenten dichotome items bevatten op ordinaal meetniveau. Tabel 2 laat de resultaten zien van de overeenstemming tussen de eerste beoordelaar met items van het STIP instrument en de derde beoordelaar met de 17 geselecteerde items van de ICALT voor tien rekenwiskundelessen.

Tabel 2. *Spearman Brown*

Les	Rs	p
Les 1	-.346	.174
Les 2	-.270	.295
Les 3	-.523	.031*
Les 4	-.022	.935
Les 5	-.299	.244
Les 6	-.015	.953
Les 7	-.311	.225
Les 8	-.118	.653
Les 9	-.118	.653
Les 10	-.278	.280
Les 1 tot en met 10	-.346	.174

Noot. * $p < .05$

Alleen les 3 heeft een significante correlatie. De andere lessen hebben een negatieve, niet-significante correlatie van de scores op de ICALT en de scores op het STIP instrument.

4.2.2 Inhoudsvaliditeit

In deze paragraaf worden de beoordelingen van de onderwijsexperts over het STIP instrument toegelicht. De drie onderwijsexperts zijn afgekort O1, O2 en O3: O1 is docent, O2 is wetenschapper aan Universiteit Twente en O3 is wetenschapper aan Universiteit Oxford. Ten eerste wordt het belang van de items besproken (vraag a). Ten tweede wordt de duidelijkheid van het taalgebruik van de items toegelicht (vraag b).

4.2.2.1 Vraag a: in welke mate de items belangrijk zijn om differentiatie te meten?

Tabel 3 laat de beoordelingen zien van de drie onderwijsexperts aan de hand van een zevenpunt Likert schaal op vraag a per item van het STIP instrument. Deze beoordelingen zijn gesorteerd in de vier schalen van het STIP instrument: differentiatie in inhoud, taak, proces en leeromgeving.

Tabel 3. Gemiddelde van de drie onderwijsexperts per item op vraag a (in hoeverre de items differentiatie meten)

Inhoud		Taak		Proces		Leeromgeving	
M		M		M		M	
1	5.33	8	5.33	14	5.67	26	3.00
2	5.67	9	5.33	15	5.67	27	2.67
3	6.00	10	2.67	16	5.67	28	3.00
4	3.33	11	3.00	17	5.33	29	3.00
5	3.33	12	3.00	18	6.00	30	3.67
6	6.00	13	3.33	19	6.33	31	3.33
7	6.00			20	5.67	32	5.67
				21	6.00	33	5.33
				22	3.00		
				23	3.00		
				24	3.00		
				25	3.00		

Uit Tabel 3 is af te lezen dat items 1, 2, 3, 6 en 7 bij de schaal *inhoud* gemiddeld hoog werden beoordeeld. Ondanks de hoge beoordelingen bij inhoud, misten de wetenschappers de relatie met differentiatie en het koppelen van de leerinhoud met andere leerinhouden bij items 4 en 5. O3 voegde schriftelijk toe dat deze items meer gerelateerd zijn aan het meten van context overstijgend doceren dan differentiatie. Om deze redenen werden deze items lager gecodeerd door de wetenschappers. Bij de schaal *taak* is af te lezen dat gemiddeld alle onderwijsexperts items 8 en 9 belangrijk vonden om differentiatie te meten. Echter beoordeelden de wetenschappers items 10, 11, 12 en 13 lager, terwijl de docent deze items hoger beoordeelde. De wetenschappers misten bij deze items het verband met differentiatie en het stimuleren van creatief denken bij leerlingen op hun eigen niveau. O2 schreef dat deze items zeker gericht zijn op het stimuleren van creatief denken, maar het aanpassen op het niveau van de leerlingen in de items ontbrak. Daarnaast schreef O3 dat deze items met name meet hoe de leraar omgaat met verschillen in leren bij leerlingen. Bij de schaal *proces* is het groeperen van de leerlingen op basis van hun mogelijkheden of interesses en de leerlingen leerinhouden laten ontdekken op een wijze die past bij hun voorkennis zijn van belang om differentiatie te meten (items 14 tot en met

17). Ondanks het hoge gemiddelde voor alle items, misten de wetenschappers de samenhang met differentiatie en het stimuleren van kritisch nadenken op verschillende manieren (items 22 tot en met 25). Volgens de wetenschappers kan het kritisch denken eerder geclassificeerd worden bij het algemeen onderwijskwaliteit dan bij differentiatie. Ten slotte is de schaal *leeromgeving* gemiddeld laag gescoord door de onderwijsexperts. De wetenschappers meldden dat deze items goed zijn om klassenmanagement te meten, maar niet voor het bepalen van differentiatie. Ondanks het lage gemiddelde voor alle items bij leeromgeving, beoordeelden de wetenschappers het item over het stimuleren van het plannen, monitoren en vaststellen van het geleerde (item 32) en het reflecteren van het geleerde (item 33) hoog om differentiatie in leeromgeving te meten. O2 meldde dat item 33 de leerkracht inzicht geeft of de leerling de leerinhoud wel of niet heeft begrepen. Dit inzicht kan de basis vormen voor extra uitleg of verdieping voor de leerlingen.

Om inzicht te krijgen in de mate van overeenstemming tussen deze beoordelingen is de ICC uitgevoerd. Tabel 4 toont de gemiddelden, standaarddeviaties en de ICC van de beoordelingen van de onderwijsexperts per schaal op vraag a van het STIP instrument.

Tabel 4. Gemiddelden, SD en ICC per schaal voor vraag a (in hoeverre de items differentiatie meten)

	Onderwijs- expert		Onderwijs- expert		Onderwijs- expert		Onderwijsexperts			Onder wijs- experts	Onder wijs- experts	Onder wijs- experts
	O1		O2		O3		O1 – O2 – O3			O1 –O2	O1 –O3	O2 –O3
	M	SD	M	SD	M	SD	M	SD	ICC	ICC	ICC	ICC
Inhoud (N = 7)	4.86	1.38	4.86	1.95	5.57	1.90	5.06	1.23	.396	-.196	-.127	.923
Taak (N = 6)	4.83	1.41	3.50	1.97	3.77	1.22	3.78	1.22	.541	.148	.156	.952
Proces (N = 12)	4.83	1.84	4.42	1.98	5.33	1.76	4.86	1.39	.631	.389	.551	.834
Leeromgeving (N = 8)	5.50	1.76	2.75	1.39	2.75	1.67	3.67	1.08	.500	.343	.170	.818

Uit Tabel 4 is af te lezen dat de drie onderwijsexperts bij de schaal *inhoud* een zwakke overeenstemming hebben (.396). Daarnaast is het gemiddelde tussen de drie onderwijsexperts hoog (5.06). Opvallend is de hoge overeenstemming tussen de twee wetenschappers (.923). Een hoge overeenstemming betekent dat de wetenschappers met elkaar eens waren dat het item wel of niet van belang is om differentiatie te meten. Daarnaast is de overeenstemming tussen de drie onderwijsexperts bij de schaal *taak* hoger dan bij inhoud (.541), maar is het gemiddelde van de

beoordelingen lager (3.78). Ook laten de resultaten zien dat de wetenschappers een goede overeenstemming hebben in hun beoordelingen (.952) en de overeenstemming tussen de docent en de wetenschappers lager (.148; .156). Bij de schaal *proces* is de overeenstemming tussen de drie onderwijsexperts (.631) met een gemiddelde beoordeling van 4.86. Daarnaast is de overeenstemming tussen de wetenschappers hoger dan de overeenstemming tussen docent en wetenschappers. Ten slotte is een matige overeenstemming gevonden voor de schaal *leeromgeving* (.500), met een gemiddelde beoordeling van 3.67. De hoogste overeenstemming is wederom tussen de wetenschappers (.818).

4.2.2.2 Vraag b: in welke mate is het taalgebruik duidelijk van de items?

Tabel 5 laat per item van het STIP instrument de beoordelingen zien van de drie onderwijsexperts op vraag b.

Tabel 5. Gemiddelde over de drie onderwijsexperts per item op vraag b (in hoeverre het taalgebruik van de items duidelijk is)

Inhoud		Taak		Proces		Leeromgeving	
M		M		M		M	
1	3.67	8	3.67	14	6.67	26	5.33
2	6.00	9	6.00	15	6.33	27	6.00
3	6.00	10	6.00	16	5.33	28	6.33
4	5.33	11	5.33	17	6.67	29	6.33
5	5.33	12	5.33	18	5.67	30	5.00
6	6.67	13	6.67	19	6.33	31	6.33
7	6.33			20	5.67	32	6.33
				21	5.33	33	6.67
				22	6.00		
				23	6.00		
				24	6.33		
				25	5.67		

Uit Tabel 5 is af te lezen dat de items van de schaal *inhoud* gemiddeld goed werden beoordeeld. Opvallend is het lage gemiddelde voor item 1. Om het activeren van voorkennis bij de leerlingen te beoordelen meldden de wetenschappers dat dit item onduidelijk is geformuleerd. Bij de schaal *taak* vond de docent alle items duidelijk geformuleerd. Ondanks dat de wetenschappers de items ook goed geformuleerd vonden, meldden ze bij item 8 dat het woord ‘bijvoorbeeld’ een irrelevant woord is en dat er te veel woorden staan waardoor het niet goed leesbaar is. Bij de schaal *proces* waren de drie onderwijsexperts erover eens dat de items duidelijk beschreven

waren. Ondanks dat de onderwijsexperts hierover eens waren, vonden de wetenschappers bij item 18, 20 en 21 dat te veel informatie was weergegeven over het ontdekken op een wijze die past bij de voorkennis van de leerlingen. Ten slotte werden gemiddeld de items bij de schaal *leeromgeving* goed beoordeeld. Om inzicht te krijgen in de mate van overeenstemming van deze beoordelingen is de ICC uitgevoerd. Tabel 6 toont de gemiddelden, standaarddeviaties en de ICC van de beoordelingen van de onderwijsexperts per schaal op vraag b van het STIP instrument.

Tabel 6. *Gemiddelden, SD en ICC per schaal voor vraag b (in hoeverre het taalgebruik van de items duidelijk is)*

	Onderwijs- expert		Onderwijs- expert		Onderwijs- expert		Onderwijsexperts			Onder wijs- experts	Onder wijs- experts	Onder wijs- experts
	O1		O2		O3		O1 – O2 – O3			O1–O2	O1–O3	O2–O3
	M	SD	M	SD	M	SD	M	SD	ICC	ICC	ICC	ICC
Inhoud (N = 7)	6.86	1.38	5.43	1.52	4.56	1.99	5.62	1.99	.190	-.078	-.198	.550
Taak (N = 6)	6.50	1.55	5.00	1.10	5.00	1.55	5.50	1.81	.256	-.267	.000	.667
Proces (N = 12)	6.42	1.51	5.25	1.22	6.33	1.89	6.00	1.47	-.105	-.130	.230	-.241
Leer- omgeving (N = 8)	6.63	1.74	5.88	1.83	5.63	1.74	6.04	1.58	.327	-.086	.226	.829

Ten eerste is uit Tabel 6 af te lezen dat de overeenstemming tussen de drie onderwijsexperts bij de schaal *inhoud* zwak is (.190), terwijl gemiddeld de items hoog werden beoordeeld (5.62). O1 vond deze items duidelijk beschreven, maar de wetenschappers waren het hier deels mee eens. Ten tweede komt de overeenstemming bij de schaal *taak* matig overeen (.256), waarbij een gemiddelde score op de items (5.50). Ten derde laten de resultaten een zeer zwakke overeenstemming zien tussen de drie onderwijsexperts bij de schaal *proces* (-.105), maar wel een hoog gemiddelde op de items (6.00). Opvallend is de zwakke overeenstemming tussen de wetenschappers (-.241), waarbij O1 en O3 het meer met elkaar eens waren dan de twee wetenschappers. Ten slotte zijn de items gemiddeld hoog beoordeeld (6.04) maar de overeenstemming tussen de drie onderwijsexperts is matig voor de schaal *leeromgeving* (.327). De overeenstemming bij leeromgeving is tussen de wetenschappers hoog (.829), maar de overeenstemming tussen de docent en de wetenschappers is laag (-.086; .226).

5. Discussie

5.1 Opzet van het onderzoek

Het doel van dit onderzoek was om de betrouwbaarheid en validiteit van het STIP instrument te onderzoeken. In dit onderzoek zijn drie psychometrische kwaliteiten van het STIP instrument onderzocht: 1) de interbeoordelaarsbetrouwbaarheid, 2) de externe validiteit en 3) de inhoudsvaliditeit. Ten eerste werd een hoge interbeoordelaarsbetrouwbaarheid verwacht. Er werd verwacht dat de interbeoordelaarsbetrouwbaarheid hoog was omdat de ontwikkelaar van het STIP instrument de ontwikkeling heeft besproken met twee collega's. Ook is het STIP instrument tijdens twee lessen op twee verschillende scholen getest, dat geleid heeft tot de definitieve versie van het STIP instrument. Naast de interbeoordelaarsbetrouwbaarheid werd een hoge externe validiteit verwacht. Deze verwachting was geschept omdat de ICALT is goedgekeurd en getest door de onderwijsinspectie (Van de Grift, 2007). Ook bevatten de ICALT en het STIP instrument beiden items over differentiatie. Ten slotte werd een hoge inhoudsvaliditeit van het STIP instrument verondersteld omdat het instrument was getest tijdens twee lessen op twee verschillende scholen en de ontwikkelaar de ontwikkeling van het instrument heeft besproken met twee collega's. Ook werd verwacht dat de overeenstemming van de beoordelingen van de drie onderwijsexperts overeenkwamen met een minimaal gemiddelde van vier punten per item. Door deze drie kwaliteiten te onderzoeken werd een beeld geschetst over de psychometrische kwaliteiten van het STIP instrument.

5.2 Psychometrische kwaliteiten van het STIP instrument

5.2.1 Interbeoordelaarsbetrouwbaarheid

De resultaten in dit onderzoek lieten zien dat de overeenstemming tussen de eerste en de tweede beoordelaar goed was. Dit betekent dat de twee beoordelaars overeenstemden in hun waarnemingen over het STIP instrument en dat het voor de kwaliteit van het instrument weinig uit zou maken welke persoon beoordeelde. Opvallend was de lage score bij les 5 ($k = .511$). Een verklaring hiervoor kan zijn dat de audio-opname vijftien minuten minder goed hoorbaar was omdat de leerlingen door de klassikale instructie van de leerkracht praatten. Door deze ruis heeft de tweede beoordelaar deze vijftien minuten niet gescoord. Dit kan de scores hebben beïnvloed wat mogelijk de lage Cohen's kappa kan verklaren.

5.2.2 Externe validiteit

De resultaten lieten zien dat er een negatief, niet-significante correlatie was tussen de ICALT en het STIP instrument. Dit betekent dat een hoge score op het STIP instrument leidde tot een lage score bij de ICALT en andersom. Om dit te verklaren is er gekeken naar de inhoud van de items van beide observatie-instrumenten. Ondanks de analyse tussen de ICALT en het STIP instrument, die voorafgaand is gehanteerd, waren de items op basis van inhoud niet exact gelijk. Het STIP instrument bevatte items over differentiatie voor de hoog en de laag presterende leerlingen en de geselecteerde items van de ICALT waren vooral gericht op differentiatie voor de laag presterende leerlingen. Op basis hiervan kan het aannemelijk zijn dat dit verschil een verklaring zou kunnen zijn voor de lage externe validiteit. Verder gaf de derde beoordelaar aan dat ze het prettiger vond als ze videobeeld van de leerkracht had. Met de audio-opnames kon de beoordelaar een klein gedeelte van de rekenwiskundeles minder goed beoordelen. De verlengde instructie, waarbij de leerkracht bijvoorbeeld hulprondes liep, was soms minder goed te beoordelen doordat de leerkracht zich te ver van de microfoon bevond. De leerkracht van les 3 ging na de klassikale instructie verder met een klassikale, verlengde instructie voor de laag presterende leerlingen. Dit had tot gevolg dat de stem van de leerkracht tijdens de verlengde instructie goed hoorbaar was. Dit kan het significante verschil voor les 3 verklaren omdat de derde beoordelaar de items beoordeelde. Ondanks deze verklaringen, voegde het weinig toe om de externe validiteit te meten, omdat de p-waarden niet significant waren.

5.2.3 Inhoudsvaliditeit

De resultaten lieten zien dat de drie onderwijsexperts de inhoud van het STIP instrument redelijk tot goed beoordeelden. Opvallend was het verschil van beoordelingen tussen de wetenschappers en de docent. Om dit te verklaren werd eerst gekeken naar de opvattingen van differentiatie van de drie onderwijsexperts. O1 omschrijft differentiatie als de manier waarop de docent omgaat met verschillen tussen de leerlingen. O2 definieert differentiatie als de benadering waarbij de onderwijsinhoud aansluit op de behoeften van de leerlingen; als de ene leerling meer behoefte heeft aan extra ondersteuning moet de leerkracht dit herkennen en zijn onderwijs daarop aanpassen. O3 omschrijft differentiatie als het erkennen van verschillen in leerstijl, niveau en tempo tussen leerlingen en het integreren van deze verschillen in een op het individu gerichte onderwijsaanpak. Alle drie onderwijsexperts gaven in hun omschrijvingen aan dat ze differentiatie zien als rekening houden met verschillen, waarbij de wetenschappers specifiek en gedetailleerder de verschillen tussen de leerlingen omschreven. Naast de beschrijvingen van de

onderwijsexperts voegde de auteur van dit onderzoek toe dat differentiatie een lesgeefbenadering is waarbij de leerkracht onderwijs aanbiedt op de (leer)behoefte, leerstijlen, competentieniveaus en interesses van de individuele leerlingen. De leerkracht heeft een belangrijke sleutelrol, waarbij het geven van goede lesinstructie en het bieden van passende opdrachten centraal staan. Ook zal de leerkracht een begeleidende en coachende rol aannemen om aan te sluiten op de (leer)behoeftes van de leerlingen.

Het verschil in beoordelingen van de wetenschappers en de docent kan verklaard worden door het verschil van perspectief met betrekking tot differentiatie. Verondersteld kan worden dat de wetenschappers de items beoordeelden vanuit hun wetenschappelijke kennis van metingen over psychometrische kwaliteiten van een onderzoek en instrumenten. De docent beoordeelde de items vanuit zijn praktijkgerichte kennis en lesgeefervaringen in het basisberoepsgerichte en kaderberoepsgerichte leerweg voor leerlingen van 14 tot en met 16 jaar.

Hieronder worden de resultaten voor vraag a (de mate waarin de items belangrijk zijn om differentiatie te meten) verklaard en aanbevelingen voorgesteld. Vervolgens worden de resultaten voor vraag b (de mate waarin het taalgebruik van de items duidelijk zijn) verklaard en aanbevelingen weergegeven. Vraag a en b worden toegelicht per schaal van het STIP instrument (differentiatie in inhoud, taak, proces en leeromgeving).

5.2.3.1 Vraag a: in welke mate zijn de items van belang om differentiatie te meten?

De drie onderwijsexperts vonden het van belang om de items betreft *inhoud* te meten en beoordeelden ze gemiddeld goed. De gemiddelde hoge beoordelingen veronderstelt dat de items van deze schaal het construct trachten te meten. Volgens Haynes, Richards en Kubany (1995) is die veronderstelling een voorwaarde om inhoudsvaliditeit te meten. Aangezien item 4 en 5 niet specifiek aangaven in welke mate de leerkracht omgaat met verschillen tussen leerlingen en een lager gemiddelde hebben van 3,3, dan de minimale maatstaf van vier punten, is de aanbeveling om deze twee items weg te laten uit het STIP instrument.

De items bij de schaal *taak* zijn grotendeels lager beoordeeld door de wetenschappers dan door de docent. Onlangs de lage beoordelingen van de items over creatief denken zijn deze items wel van belang om differentiatie te meten. Het stimuleren van creatief denken is bevorderend voor het uitvoeren van een taak als de leerkracht aansluit op de verschillende niveaus van de leerlingen (Awang & Ramly, 2008). Een aanbeveling is om bij items 10 tot en met 13 explicieter in te gaan op differentiatie en aan te geven hoe de leerkracht omgaat met verschillen tussen de hoog en de laag presterende leerlingen met betrekking op het stimuleren van creatief denken. De items zouden moeten meten dat de leerkracht de hoog presterende leerlingen stimuleert voor het

creatief denken door vragen te stellen die beroep doen op het analyseren en toepassen van de leertaak (hogere orde denkvaardigheden). Daarnaast zouden de items moeten meten dat de leerkracht de laag presterende leerlingen stimuleert voor het creatief denken door vragen te stellen die een beroep doen op het begrijpen van de leertaak (lagere orde denkvaardigheid).

Bij de schaal differentiatie naar *proces* zijn de items grotendeels hoog beoordeeld. De twee items die er met een hoge beoordeling uitsprongen zijn item 14 en 15, omdat expliciet staat vermeld hoe de leerkracht differentieert in het groeperen op basis van mogelijkheden en interesses. Volgens Van der Valk (2014) werken homogene groepjes effectiever voor het leereffect bij hoog presterende leerlingen en heterogene groepjes voor het leereffect bij laag presterende leerlingen. Ondanks de hoge beoordelingen gaven de wetenschappers aan dat ze de samenhang misten tussen differentiatie en het stimuleren van verschillende manieren om de leerlingen kritisch na te laten denken (items 22 tot en met 25). Het stimuleren van kritisch denken is wel van belang voor differentiatie. Kritisch denken is een hogere orde denkvaardigheid en kan gestimuleerd worden voor de hoog presterende leerlingen (Zohar & Dori, 2003). Aanbevolen wordt om de items over kritisch denken aan te vullen. De items moeten worden aangevuld met vragen over het al dan niet stimuleren van hoog presterende leerlingen om kritisch na te laten denken.

Ten slotte was de overeenstemming voor de schaal *leeromgeving* het laagst. Opvallend was het verschil tussen de docent en de wetenschappers; de docent gaf hoge beoordelingen en de wetenschappers lage beoordelingen. Doordat slechts twee van de acht items hoog zijn beoordeeld, impliceren de resultaten dat de inhoudsvaliditeit van differentiatie in leeromgeving laag is. Item 28 (de leerkracht is positief naar alle leerlingen) meet klassenmanagement in plaats van differentiatie. Aangezien de interactie van belang is voor een positieve werksfeer, zal de leerkracht alle leerlingen positief moeten benaderen. Om deze reden is de aanbeveling om item 28 weg te laten. Daarnaast zullen de overige items beter moeten meten of de leerkracht aansluit bij de onderwijsbehoeften van de hoog en de laag presterende leerlingen. Voor item 26 is de aanbeveling om twee items toe te voegen zodat deze twee items meten hoe de leerkracht rekening houdt met verschillen tussen leerlingen. Het eerste item zal moeten meten of de hoog presterende leerlingen samenwerken in een andere ruimte, bijvoorbeeld in een 'plusklas'. Het tweede item zal moeten meten of de laag presterende leerlingen in het huidige klaslokaal blijven om samen aan de taak te werken. Daarnaast kan item 32 worden weggelaten en twee nieuwe items worden toegevoegd om het plannen, monitoren en vaststellen van het leren of geleerde te meten. Het eerste item zou moeten meten of de leerkracht vragen stelt aan de hoog presterende leerlingen die een beroep doen op het evalueren en monitoren van het geleerde. Het tweede item

zal moeten meten of de leerkracht vragen stelt aan de laag presterende leerlingen die en beroep doen op het begrijpen van de leerinhoud. Door deze aanpassingen kan de mate van differentiatie in leeromgeving beoordeeld worden voor zowel de hoog als de laag presterende leerlingen.

5.2.3.2 Vraag b: in welke mate is het taalgebruik duidelijk van de items?

Het taalgebruik van alle items van het STIP instrument is goed beoordeeld door de drie onderwijsexperts. Opvallend is dat de docent over het algemeen het taalgebruik van een aantal items hoger beoordeelde dan de wetenschappers. De docent beoordeelde de items hoog omdat hij de woordkeuzes van de items begreep. De wetenschappers beoordeelden de items vanuit een methodologisch oogpunt waarbij ze waarschijnlijk de woordkeuzes kritisch beoordeelden.

Bij de schaal over differentiatie naar *inhoud* zijn de items hoog beoordeeld door de onderwijsexperts, met uitzondering van item 1. Item 1 zal aangepast moeten worden om duidelijker te beschrijven hoe de leerkracht rekening houdt tussen de hoog en de laag presterende leerlingen. Als aanbeveling zal een eerste item toegevoegd kunnen worden waarin de leerkracht de hoog presterende leerlingen vragen stelt zodat ze de leerinhoud gaan analyseren om hun voorkennis te activeren. Ook zal een tweede item toegevoegd kunnen worden waarin de leerkracht de voorkennis activeert voor de laag presterende leerlingen door vragen te stellen wat ze hebben onthouden van de desbetreffende leerinhoud.

Bij de schaal over differentiatie naar *taak* vond de docent alle items duidelijk geformuleerd. Ondanks dat de wetenschappers de items ook goed geformuleerd vonden, meldden ze dat het woord 'bijvoorbeeld' een irrelevant woord was en dat bij item 8 en 9 te veel woorden worden gebruikt waardoor het minder goed leesbaar werd. Als aanbeveling voor deze twee items kan het woord 'bijvoorbeeld' worden weggelaten en kan één voorbeeld per item worden genoemd.

Bij de schaal over differentiatie naar *proces* waren de meeste items duidelijk beschreven. Ondanks dat de onderwijsexperts vonden dat de items goed waren geformuleerd, meldden de wetenschappers dat bij items 18, 20 en 21 te veel informatie was weergegeven over het ontdekken op een wijze die past bij de voorkennis van de leerlingen. Als aanbeveling kan bijvoorbeeld elk item maximaal één voorbeeld bevatten. Eén voorbeeld per item is duidelijker voor de beoordelaar om differentiatie te beoordelen.

Bij de schaal over differentiatie naar *leeromgeving* zijn alle items duidelijk beschreven volgens de onderwijsexperts. Ondanks deze hoge beoordelingen adviseerde O2 om item 30 op te splitsen in twee items. De aanbeveling is om het advies van O2 op te volgen. Er zal een item

moeten komen die meet of de leerkracht de ideeën die de leerlingen inbrengen accepteert en een tweede item die meet of de leerkracht de ideeën die de leerlingen inbrengen gebruikt.

5.2.4 Aanbevelingen met betrekking op de schaalmethode

Het gebruik maken van dichotome items bracht ook gevolgen met zich mee. De antwoordmogelijkheden (1 = komt wel voor; 0 = komt niet voor) zijn kwetsbaar voor meetfouten vanwege alternatieven die uit de weg werden gehouden. Dit betekent dat meerdere antwoordmogelijkheden tussen deze twee antwoordmogelijkheden zijn weggelaten. In de praktijk differentieert de ene leerkracht beter dan de ander. Dit verschil is lastiger te beoordelen aan de hand van dichotome items aangezien er maar twee antwoordmogelijkheden zijn. Om differentiatie nog beter in kaart te brengen is de aanbeveling om in het STIP instrument te scoren aan de hand van een zevenpunt Likert schaal. Met een zevenpunt Likert schaal heeft de beoordelaar meer spreiding om het gedrag te beoordelen. Hiermee wordt de kwantitatieve data verzameld zodat de beoordelaar meerdere alternatieven heeft om differentiatie te beoordelen. Op deze manier wordt een genuanceerder beeld en een betere afspiegeling van de realiteit bereikt in het oordeel dat op basis van het STIP instrument wordt gedaan; in de mate waarin een leerkracht differentiatie toepast. Als er voor gekozen wordt om de Likert schaal te gebruiken is het van belang dat de beoordelaar getraind wordt in het toepassen van het STIP instrument om meetfouten te verminderen.

5.2.5 Conclusie

De belangrijkste conclusie van dit onderzoek is dat het STIP instrument 1) betrouwbaar is, omdat het een hoge Cohen's Kappa heeft, 2) een lage externe validiteit vertoont vanwege de negatieve, niet-significante correlaties tussen de ICALT en het STIP instrument en 3) de inhoudsvaliditeit hoog is vanwege de hoge beoordelingen en overeenstemming van de onderwijsexperts.

5.3 Vervolgonderzoek

Er is in dit onderzoek gekozen voor het afnemen van audio-opnames om consistentie in de data te behouden en validiteitbedreigingen te reduceren. Ondanks deze aanpak zal het beoordelen van lesobservaties door twee beoordelaars 'live' in de klas een goede mogelijkheid zijn. Als meerdere observatoren tegelijk de leerkracht observeren in de klas zal de kans op een hoge betrouwbaarheid groot zijn (Ho & Kane, 2013). Als dit niet mogelijk is vanwege gebrek aan tijd of middelen dan zal een video-opname een goed alternatief zijn. Een video-opname geeft meer

informatie dan een audio-opname voor de beoordelaar om het gedrag te beoordelen van een leraar. Bij een video-opname komt de informatie in het werkgeheugen via zowel het auditieve als het visuele kanaal. Doordat de informatie tegelijk in deze twee kanalen samenkomen, kan ook meer informatie worden verwerkt (Mayer, 2001). De leerkracht kan met non-verbale communicatie de leerlingen ook pedagogisch benaderen, dit is met een audio-opname niet zichtbaar. Volgens Babad, Bernieri en Rosenthal (1991) kan de leerkracht met non-verbale communicatievaardigheden rekening houden met verschillen tussen leerlingen. Dit kan in de praktijk worden gerealiseerd door het beoordelen van oogcontact, gezichtsuitdrukking en lichaamshouding. Mogelijk zet een leerkracht bewust non-verbale communicatie in voor hoog of laag presterende leerlingen om ze te stimuleren in hun leerproces.

Er zijn in dit onderzoek twintig lessen geobserveerd, echter vanwege de kleine steekproef moeten de resultaten voorzichtig worden geïnterpreteerd. De twintig lesobservaties die de onderzoeker zelf heeft afgenomen zijn geselecteerd door de directeurs van de desbetreffende scholen, met als enige criteria dat het rekenwiskundelessen moesten zijn in groep 5, 7 en 8. De onderzoeker had geen invloed welke leerkrachten uiteindelijk meewerkten aan dit onderzoek. Mogelijk koos de directeur voor leerkrachten die in zijn ogen van hoge kwaliteit zijn waardoor de steekproef een selectief karakter kreeg. Een aanbeveling voor vervolgonderzoek kan zijn dat de onderzoeker zelf de leerkrachten selecteert. De steekproef kan bestaan uit leerkrachten die goed differentiëren en een groep leerkrachten die minder goed differentiëren. Verwacht wordt dat het STIP instrument goed en minder goede differentiërende leerkrachten kan meten. Op deze wijze kan ook het verschil tussen leerkrachten die lesgeven in het speciaal basisonderwijs en het reguliere basisonderwijs worden onderzocht. Verder zou vervolgonderzoek zich kunnen richten op het herhalen van dit onderzoek zodat tweemaal dezelfde metingen worden afgenomen bij één leerkracht. Hiermee wordt de test-retest betrouwbaarheid van het STIP instrument onderzocht. De metingen worden uitgevoerd door één onderzoeker dat tweemaal plaatsvindt en zal onder dezelfde condities plaats moeten vinden (Dooley, 2001).

Door de implementatie van het STIP instrument te meten in termen van de psychometrische kwaliteiten van het instrument -betrouwbaarheid en validiteit- zijn inzichten opgedaan over de vraag hoe en waarom het STIP instrument werkt en op welke manier het kan worden verbeterd.

Literatuur

- Ahlers, L. (2009). Voortgezet technisch lezen. Omgaan met verschillen ter voorkoming van leesproblemen. Amersfoort: CPS.
- Awang, H. & Ramly, I. (2008). Creative thinking skill approach through problem-based learning: Pedagogy and practice in the engineering classroom. *International Journal of Human and Social Sciences*, 3(1), pp. 18-23.
- Babad, E., Bernieri, F. & Rosenthal, R. (1991). Students as judges of teachers' verbal and nonverbal behavior. *American Educational Research Journal*, 28(1), pp. 211-234.
- Baker, S. K., Gersten, R., Haager, D. & Dingle, M. (2006). Teaching practice and the reading growth of first-grade English learners: Validation of an observation instrument. *The Elementary School Journal*, 107(2), pp. 199-221.
- Bartko, J. (1976). On various intraclass correlation reliability coefficient. *Psychological Bulletin*, 83(5), pp. 762-765.
- Bearne, E. (1996). *Differentiation and diversity in the primary school*. Abingdon: Routledge.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1; Cognitive domain*. New York: David McKay.
- Bosker, R. J. (2005). *De grenzen van gedifferentieerd onderwijs*. (Rede uitgesproken bij de aanvaarding van het ambt van hoogleerkracht).
- Bouwman, A. Brouwer, G., Jansen, L. & Loman, E. (2014). *Differentiëren is te leren!* Amersfoort: CPS.
- Bracht, G. H. & Glass, G. V. (1968). The external validity of experiments. *American Educational Research Journal*, 5(4), pp. 437-474.

- Chamorro-Premuzic, T., Furnham, A. & Lewis, M. (2007). Personality and approaches to learning predict preference for different teaching methods. *Learning and Individual Differences*, 17(3), pp. 241-250.
- Chan, D. W. (2001). Learning styles of gifted and nongifted secondary students in Hong Kong. *Gifted Child Quarterly*, 45(1), pp. 35-44.
- Clijisen, A. (2007). *Handreiking 1-zorgroute voor leraren en intern begeleiders in het primair onderwijs*. 's Hertogenbosch: WSNS-Plus en KPC-Groep.
- Cohen, R. J. & Swerdlick, M. E. (2005). *Psychological testing and measurement: An introduction to tests and measurement*. New York: McGraw-Hill.
- Danielson, C. (2007). *Enhancing professional practice: a framework for teaching*. 2nd edition. New Jersey: ASCD.
- Darling-Hammond, L., Wise, A. & Klein, S. (1999). *A license to teach: Raising standards for teaching*. San Francisco: Jossey-Bass.
- Doolaard, S. & Harms, T. (2013). *Omgaan met excellente leerlingen in de dagelijkse onderwijspraktijk*. Groningen: GION/RUG.
- Dooley, D. (2001). *Social research methods*. Dorchester: Henry Ling Limited.
- Ducette, J., Sewell, T. & Shapiro, J. (1996). *Diversity in education: Problems and possibilities*. San Francisco: Jossey-Bass.
- Dunn, R., Beaudry, J. & Klavas, A. (1989). Survey of research in learning styles. *Educational Researcher*, 70(6), pp. 50-58.

- Durlak, J. A. & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychology*, 41(3/4), pp. 327-350.
- Eyre, D. & Geake, J. G. (2002). Trends in research into gifted and talented education in England. *Gifted and Talented International*, 17(1), pp. 13-21.
- Gamoran, A. & Weinstein, M. (1995). Differentiation and opportunity in restructured schools. *American Journal of Education*, 106(3), pp. 385-415.
- Good, T. L. (1988). *Teacher expectations*. In D. Berliner and B. Rodenshine (eds), talks to teachers. New York: Random house.
- Griggs, S. (1984). Counseling the gifted and talented based on learning style. *Exceptional Children*, 50(5), pp. 429-432.
- Griggs, S. & Dunn, R. (1984). Selected case studies of the learning style preferences of gifted students. *Gifted Child Quarterly*, 28(3), pp. 115-129.
- Grigorenko, E. & Sternberg, R. J. (1997). Styles of thinking, abilities, and academic performance. *Exceptional Children*, 63(3), pp. 295-312.
- Gross, M. U. M. (1997). Changing teacher attitudes towards gifted children: An early but essential step. *Hong Kong: World Council for Gifted and Talented Children*. pp. 3-22.
- Haynes, S. N., Richards, D. C. S. & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7(3), pp. 238-247.
- Hall, T. (2002) *Differentiated instruction*. Wakefield, MA: National Center on, 2002 – principals.in.

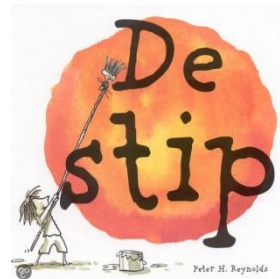
- Heacox, D. (2002). *Differentiating instruction in the regular classroom. How to reach and teach all learners, grades 3-12*. Minneapolis, US: Free Spirit Publishing.
- Hill, H. C., Schilling, S. G. & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105(1), pp. 11-30.
- Ho, A. D. & Kane, T. J. (2013). The reliability of classroom observations by school personnel. *Bill & Melinda Gates Foundation* (pp. 1-33). Harvard: Graduate School of Education.
- Honey, P. & Mumford, A. (1992). *The manual of learning styles*. Maidenhead: Peter Honey.
- Howard, P. (1994). *An owner's manual for the brain*. Austin, TX: Leornian Press.
- Inspectie van het Onderwijs (2014). De staat van het onderwijs. *Onderwijsverslag 2012/2013*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2015). Beginnen leerkrachten kijken terug. *Onderzoek onder afgestudeerden. Deel 1: de pabo*. Utrecht: Inspectie van het Onderwijs.
- Jensen, E. (1998). *Teaching with the brain in mind*. Alexandria, VA: Association for supervision and curriculum development.
- Klahr, D., Triona, L.M. & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, 44(1), pp. 183-203.
- Kolb, D. A. (1984). *Experiential learning: Experience as a source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Koopmans, H. J. M. (2006). *Professionals organiseren informeel leren*. Delft: Eburon.

- Levy, H. M. (2010). Meeting the needs of all students through differentiated instruction: Helping every child reach and exceed standards. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 81(4), pp. 161-164.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E. (2005). *Learning and instruction*. California: Pearson Education.
- McLaughlin, M. & Talbert, J. (1993). Contexts that matter for teaching and learning: Strategic opportunities for meeting the nation's educational goals. *Center for Research on the Context of Secondary School Teaching*. 143, pp. 5-32.
- Pimparon, P., Roff, S., McAlleer, S., Poonchai, B. & Pemba., S. (2000). Educational environment, student approaches to learning and academic achievement in a Thai nursing school. *Medical Teacher*, 22(4), pp. 359-364.
- Schoenfeld, A. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher*, 28(7), pp. 4-14.
- Shepard, L. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20(7), pp. 2-9.
- Sousa, D. (2001). *How the brain learns* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Stradling, B. & Saunders, L. (1993). Differentiation in practice: Responding to the needs of all pupils. *Educational Research*, 35(2), pp. 127-137.
- Sternberg, R. J., Torff, B. & Grigorenko, E. L. (1998). Teaching triarchically improves student achievement. *Journal of Educational Psychology*, 90(3), pp. 374-384.
- Taylor, P. C., Fraser, B. J. & Fisher, D. L. (1997) Monitoring constructivist classroom learning environments. *International Journal of Educational Research*, 27(4), pp. 293-302.

- Tomlinson, C., A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tomlinson, C., A. (2000). Differentiation of instruction in the elementary grades. *ERIC Clearinghouse on Elementary and Early Childhood Education*. pp. 1-7.
- Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A. & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, 27(2/3), pp. 119-145.
- Tschannen-Morana, M. & Woolfolk Hoyb, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7). pp. 783-805.
- Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, 49(2), pp. 127-152.
- Van der Valk, T. (2014). *Excellentie en differentiatie*. Den Haag: School aan zet.
- Van Yperen, T. & Veerman, JW. (2008). *Zicht op effectiviteit. Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg*. Delft: Eburon.
- VanTassel-Baska, J., Avery, L., Struck, J., Feng, A., Bracken, B. A., Drummond, D. & Stambaugh, T. (2005). Classroom observation scale-revised. *Center for Gifted Education*. pp. 1-19.
- Vernooy, K. (2001). *Het belang van effectieve instructie*. Amersfoort: 2001.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT Press.

- Watts-Taffe, S., Laster, B., Broach, L., Marinak, B., McDONALD Connor, C. & Walker-Dalhouse, D. (2012). Differentiated instruction. Making informed teacher decisions. *The Reading Teacher*, 66(4), pp. 303-314.
- Wolfe, P. (2001). *Brain matters: Translating research into classroom practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, 11(2), pp. 307-313.
- Zimmerman, B. J. & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82(3), pp. 51-59.
- Zohar, A. & Dori, Y. J. (2003) Higher order thinking skills and low-achieving students: Are they mutually exclusive? *Journal of the Learning Sciences*, 12(2), pp. 145-181.
- Zohar, A., Degani, A. & Vaaknin, E. (2001). Teachers' beliefs about low achieving students and higher order thinking. *Teaching and Teachers' Education*, 17(4), pp. 469-485.

Bijlage I. Het STIP instrument



STIP

Observatielijst voor in de klas

Observator		Datum	
School		Geobserveerde les	
Leerkracht		Starttijd les	
		Eindtijd les	
Aantal leerlingen Groep		Aantal jongens	
		Aantal meisjes	
Gebruikte methode		Onderwerp van de les	
Opstelling in de groep	0 Tafels los van elkaar in rijen en kolommen 0 Tafels in groepjes 0 Tafels in tweetallen naast elkaar 0 Anders, namelijk..... Maak een foto!		

Lees voorafgaand aan de observatie dit hele instrument goed door, zodat je weet wat er van je verwacht wordt, waar je op moet letten en hoe je dit moet scoren.

Zorg dat je voldoende vragenlijsten voor leerling en leerkracht mee hebt.

Vraag of je in de les een geluidsopname mag maken. Dan kun je later dat wat gezegd wordt nogmaals beluisteren.

Als je op school aankomt...

- Vraag naar de betreffende leerkracht van groep.
- Stel jezelf voor aan de leerkracht van de groep en vertel dat je komt voor de observatie van de afgesproken Natuur en techniekles (hetzij een reguliere, hetzij een STIP).
- Bespreek met de leerkracht waar je het beste kunt gaan zitten in de groep. Je wilt een plek waar je de hele ruimte goed kunt overzien en waar je de leerkracht goed kunt verstaan. Overleg of het goed is dat je af en toe ook rond loopt.
- Vraag of het goed is dat je een geluidsopname maakt. Dictafoon-app in je telefoon.
- Vraag of de leerkracht na afloop van de les de vragenlijsten voor de leerlingen wil laten invullen en ook de vragenlijst voor de leerkracht. Geef de vragenlijsten gelijk aan de leerkracht.
- Vul vooraf zoveel mogelijk de algemene informatie op het voorblad in. Maak ook een foto van het lokaal met je mobiele telefoon.
- Vul tijdens de observatie eerst de lesbeschrijving in. Noteer wat er gebeurt en welke interacties plaatsvinden.
- Vul na de observatie, bijvoorbeeld als de leerlingen en leerkracht de vragenlijst invullen (max. 15 minuten) op basis van je lesbeschrijving en de audio-opname de leerkrachtobservatie in.
- Neem de ingevulde vragenlijsten, samen met het ingevulde observatieformulier mee.

De lesbeschrijving

<i>Wat gebeurt er in de les? (methoden/organisatie)</i>	<i>Minuten</i>	<i>Vragen of opmerkingen (specifieke quotes)</i>		<i>Andere observaties (noteer hier ook het aantal leerlingen dat antwoord mag geven of dat betrokken is bij de vraag).</i>
		<i>Leerkracht</i>	<i>Leerling</i>	
<i>Onverwacht leerlinggedrag dat geobserveerd is.</i>				

De lesbeschrijving (vervolg)

<i>Wat gebeurt er in de les? (methoden/organisatie)</i>	<i>Minuten</i>	<i>Vragen of opmerkingen (specifieke quotes)</i>		<i>Andere observaties (noteer hier ook het aantal leerlingen dat antwoord mag geven of dat betrokken is bij de vraag).</i>
		<i>Leerkracht</i>	<i>Leerling</i>	
<i>Onverwacht leerlinggedrag dat geobserveerd is.</i>				

Vragen voor leerkracht na observatie

Bespreek de volgende vragen met de leerkracht NADAT je de observatie hebt afgerond. Dit duurt ongeveer 15 minuten.

<i>Vraag</i>	<i>Antwoord leerkracht</i>
1. Hoe heeft u besloten wat u in deze les ging doen? Welke beslissingen heeft u daarin gemaakt? Heeft u gebruik gemaakt van een uitgeschreven lesplan voor deze les? zo ja, graag kopie meenemen	
2. Wat was het doel van de les? Wat moesten de kinderen leren?	
3. Had u de voorkennis van de leerlingen in kaart gebracht? Zo ja, op welke manier?	
4. Op welke manier heeft u rekening gehouden met verschillen tussen leerlingen?	
5. Wat waren de doelen van de voorgaande les bij deze groep? Waar sloot deze les op aan?	
6. Welke inhoud gaat u in de volgende les behandelen? Welke doelen streeft u dan na?	
7. Gaat u de kinderen huiswerk of extra of ander werk geven naar aanleiding van deze les?	
8. Hoe bepaalt u of de leerlingen de doelen van deze les bereikt hebben?	
9. Zijn er aspecten van de les die u nog wilt verduidelijken voordat we deze observatie afronden?	

Leerkrachtobservatie

Differentiatie in inhoud	Score
<i>Leerlingen van verschillende competentieniveaus werken aan verschillende inhoud (met inherente complexiteit).</i>	<i>0 = komt niet voor 1 = komt voor</i>
De leerkracht:	
1. Bepaalt wat leerlingen al over de te behandelen leerinhoud weten. <i>Vraagt expliciet naar de voorkennis over dit onderwerp: wat weet je er al van? Helpt voorkennis activeren.</i>	
2. Biedt leerlingen leerinhouden aan die zijn afgestemd op hun leerbehoeften. - Concrete leerinhouden voor de moeilijk lerende leerlingen: bijvoorbeeld de schijf van 5, of de limonadefabriek. - Meer abstracte leerinhouden voor de slimmere leerlingen: bijvoorbeeld fotosynthese of het ontstaan van dag en nacht.	
3. Laat leerlingen relaties tussen behandelde leerinhoud en andere leergebieden zien of ontdekken al naar gelang hun leerbehoeften. - De leerkracht geeft aan waar de leerinhoud nog meer mee te maken kan hebben of aan welke andere onderwerpen/ leerinhouden deze leerinhoud gekoppeld kan worden. - De leerkracht biedt ruimte aan leerlingen om te ontdekken waar de behandelde leerinhoud nog meer mee te maken kan hebben of aan welke andere onderwerpen/leerinhouden deze leerinhoud gekoppeld kan worden ('zijpaden' uitdiepen).	
4. Maakt gebruik van verschillende bronnen die zijn afgestemd op de leerbehoeften van de leerlingen. - De leerkracht heeft voor moeilijk lezende of lerende leerlingen makkelijkere teksten ingezet. - De leerkracht heeft voor de makkelijk lezende of lerende leerling moeilijkere teksten ingezet.	
Opmerkingen:	

Differentiatie in taak	Score
<i>Leerlingen van verschillende competentieniveaus werken aan andere typen taken (van meer lagere naar meer hogere orde denkvaardigheden stimulerend)</i>	<i>0 = komt niet voor 1 = komt voor</i>
De leerkracht:	
5. Zorgt voor opdrachten die afgestemd zijn op de leerbehoeften van de leerlingen. - Hij stelt moeilijk lerende leerlingen bijvoorbeeld vooral vragen die een beroep doen op herinneren, begrijpen en toepassen. Stelt vragen die ingaan op feitenkennis (wat is....?), oorzaak-gevolg relaties (wat was eerst....?) of het toepassen van het geleerde (laat eens zien hoe je dit uitrekent....?). - Hij stelt makkelijk lerende leerlingen bijvoorbeeld vooral vragen die een beroep doen op het analyseren, evalueren en creëren bij leerlingen. Stelt vragen waarbij leerlingen zaken met elkaar moeten vergelijken, waarbij ze hun mening moeten geven en onderbouwen of waarbij ze gevraagd wordt iets nieuws te bedenken.	
6. Stimuleert het creatief denken bij leerlingen op hun eigen niveau. - Neemt geen genoegen met de eerst genoemde ingevingen, vraagt meer van de leerlingen. - Stelt vragen als: zou je dat ook anders kunnen bekijken? Hoe zou die of die daar tegenaan kijken? Denkt iedereen daar hetzelfde over? - Zegt expliciet dat gekke oplossingen soms helemaal zo gek nog niet zijn... bijv. de aarde is rond was een heel gekke uitspraak toen men dacht dat de aarde plat was... totdat het tegendeel werd bewezen. - Zegt bijvoorbeeld: zoek dat maar eens op of zoek dat maar eens uit... komen we er morgen op terug...	
Opmerkingen:	

Differentiatie in proces	Score
<i>Verschillende competentieniveaus krijgen verschillende procesondersteuning (meer directief, versus meer coachend).</i>	<i>0 = komt niet voor 1 = komt voor</i>
De leerkracht	
7. Groepeer leerlingen op basis van hun mogelijkheden of interesses • <i>Laat leerlingen die nog niet veel over het onderwerp weten samenwerken in een heterogene groep (niveau of interesse)</i> • <i>Laat leerlingen die al veel over het onderwerp weten samenwerken in een homogene groep (niveau of interesse)</i>	
8. Laat leerlingen de leerinhoud ontdekken op een wijze die past bij hun voorkennis. • <i>Biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.</i> • <i>Laat leerlingen die al veel van het onderwerp weten de ruimte om de leerstof zelf te ontdekken met behulp van een korte instructie en gestructureerde opdrachten.</i>	
9. Legt leerinhoud op verschillende manieren uit zodat centrale begrippen goed worden overgebracht. • <i>Werkt bij leerlingen die nog niet veel van het onderwerp weten of het moeilijk vinden met veel voorbeelden, vergelijkingen en kleine stapjes.</i> • <i>Biedt leerlingen die al veel weten of makkelijk leren meer ruimte voor verbinding, neemt grotere stappen, laat sneller los.</i>	
10. Gebruikt verschillende manieren van feedback geven al naar gelang wat de leerling nodig heeft. • <i>Geeft bij de moeilijk lerende leerlingen meer directieve feedback, zoekt meer naar voorbeelden, geeft misschien ook de oplossing.</i> • <i>Geeft de makkelijk lerende leerlingen meer procesmatige feedback: laat leerlingen zaken zelf uitzoeken, geeft hints, maar geen oplossing.</i>	
11. Stimuleert op verschillende manieren de leerlingen om kritisch te denken. • <i>Stelt vragen als: wat vind je daar van? Hoe zou jij dat aanpakken? Hoe zou jij dat oplossen? Is dat voor iedereen een probleem denk je?</i> • <i>Stelt vragen als: als je ... en ... met elkaar vergelijkt... wat zou dan ... zijn? Welke oplossing zou meer mensen helpen?</i> • <i>Stelt vragen waarin leerlingen moeten synthetiseren, hun gevonden resultaten omzetten naar een theorie. Wat betekent dat voor.... ? Kun je nu ook iets concluderen over..... ?</i> • <i>Stelt vragen als: kun je in het kort aangeven wat nu precies het probleem is en welke oplossing je uiteindelijk het beste vindt?</i>	
Opmerkingen:	

Leeromgeving	Score
<i>De leeromgeving is de setting waar het leren van de leerlingen plaatsvindt. Deze setting heeft een fysieke en psychologische component.</i>	<i>0 = komt niet voor 1 = komt voor</i>
De leerkracht	
12. Heeft het lokaal zo ingericht dat het uitnodigt tot samenwerken. <i>De tafels staan in groepjes opgesteld, in de klas staan en hangen diverse uitnodigende materialen, er is ruimte om te onderzoeken/ontdekken.</i>	
13. Zorgt voor een goede voorbereiding van de les waardoor er weinig tijd verloren gaat aan organisatie en klassenmanagement. <i>De les is goed voorbereid, materialen zijn al beschikbaar, leerlingen weten wat er van hen verwacht wordt.</i>	
14. Accepteert de gevoelens van leerlingen op een positieve manier. <i>Is positief naar alle leerlingen</i>	
15. Moedigt leerlingen aan. <i>Neemt geen genoegen met eerste antwoord, vraagt door/verder, geeft zelf ook aanvullende voorbeelden of interpretaties.</i>	
16. Accepteert of gebruikt de ideeën die leerlingen inbrengen. <i>Waardeert de inbreng van de verschillende leerlingen</i>	
17. Heeft hoge verwachtingen van de leerlingen. <i>Laat de leerlingen op een positieve manier merken dat hij verwacht dat de leerlingen ver zullen komen bij een taak of opdracht.</i>	
18. Betrekt leerlingen bij het plannen, monitoren en vaststellen van hun leren of het geleerde. <i>Stelt leerlingen vragen die erop gericht zijn om hen zelf te laten nadenken over planning of volgen van het leren.</i>	
19. Reflecteert met de leerlingen op het geleerde. <i>Vraagt leerlingen wat het betekent wat ze geleerd hebben, vraagt of ze kunnen reproduceren wat er geleerd is, vraagt of het doel bereikt is van de les, vraagt wat ze nog niet weten, waar nog vragen liggen.</i>	
Opmerkingen:	

Bijlage II. Items van het STIP instrument per schaal

Differentiatie in inhoud	
1.	De leerkracht vraagt expliciet naar de voorkennis over dit onderwerp: wat weet je er al van? Helpt voorkennis activeren.
2.	De leerkracht biedt concrete leerinhouden voor de moeilijk lerende leerlingen: bijvoorbeeld de schijf van 5, of de limonadefabriek.
3.	De leerkracht biedt meer abstracte leerinhouden voor de slimmere leerlingen: bijvoorbeeld fotosynthese of het ontstaan van dag en nacht.
4.	De leerkracht geeft aan waar de leerinhoud nog meer mee te maken kan hebben of aan welke andere onderwerpen/ leerinhouden deze leerinhoud gekoppeld kan worden.
5.	De leerkracht biedt ruimte aan leerlingen om te ontdekken waar de behandelde leerinhoud nog meer mee te maken kan hebben of aan welke andere onderwerpen/leerinhouden deze leerinhoud gekoppeld kan worden ('zijpaden' uitdiepen).
6.	De leerkracht heeft voor moeilijk lezende of lerende leerlingen makkelijkere teksten ingezet.
7.	De leerkracht heeft voor de makkelijk lezende of lerende leerling moeilijker teksten ingezet.
Differentiatie in taak	
8.	De leerkracht stelt moeilijk lerende leerlingen bijvoorbeeld vooral vragen die een beroep doen op herinneren, begrijpen en toepassen. Stelt vragen die ingaan op feitenkennis (wat is....?), oorzaak-gevolg relaties (wat was eerst....?) of het toepassen van het geleerde (laat eens zien hoe je dit uitrekt....?)
9.	De leerkracht stelt makkelijk lerende leerlingen bijvoorbeeld vooral vragen die een beroep doen op het analyseren, evalueren en creëren bij leerlingen. Stelt vragen waarbij leerlingen zaken met elkaar moeten vergelijken, waarbij ze hun mening moeten geven en onderbouwen of waarbij ze gevraagd wordt iets nieuws te bedenken.
10.	De leerkracht neemt geen genoegen met de eerst genoemde ingevingen, vraagt meer van de leerlingen.
11.	De leerkracht stelt vragen als: zou je dat ook anders kunnen bekijken? Hoe zou die of die daar tegenaan kijken? Denkt iedereen daar hetzelfde over?
12.	De leerkracht zegt expliciet dat gekke oplossingen soms helemaal zo gek nog niet zijn... bijv. de aarde is rond was een heel gekke uitspraak toen men dacht dat de aarde plat was... totdat het tegendeel werd bewezen.
13.	De leerkracht zegt bijvoorbeeld: zoek dat maar eens op of zoek dat maar eens uit... komen we er morgen op terug...
Differentiatie in proces	
14.	De leerkracht laat leerlingen die nog niet veel over het onderwerp weten samenwerken in een heterogene groep (niveau of interesse).
15.	De leerkracht laat leerlingen die al veel over het onderwerp weten samenwerken in een homogene groep (niveau of interesse).
16.	De leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.
17.	De leerkracht laat leerlingen die al veel van het onderwerp weten de ruimte om de leerstof zelf te ontdekken met behulp van een korte instructie en gestructureerde opdrachten.

18. De leerkracht werkt bij leerlingen die nog niet veel van het onderwerp weten of het moeilijk vinden met veel voorbeelden, vergelijkingen en kleine stapjes.
19. De leerkracht biedt leerlingen die al veel weten of makkelijk leren meer ruimte voor verbinding, neemt grotere stappen, laat sneller los.
20. De leerkracht geeft bij de moeilijk lerende leerlingen meer directieve feedback, zoekt meer naar voorbeelden, geeft misschien ook de oplossing.
21. De leerkracht geeft de makkelijk lerende leerlingen meer procesmatige feedback: laat leerlingen zaken zelf uitzoeken, geeft hints, maar geen oplossing.
22. De leerkracht stelt vragen als: wat vind je daar van? Hoe zou jij dat aanpakken? Hoe zou jij dat oplossen? Is dat voor iedereen een probleem denk je?
23. De leerkracht stelt vragen als: als je ... en ... met elkaar vergelijkt... wat zou dan ... zijn? Welke oplossing zou meer mensen helpen?
24. De leerkracht stelt vragen waarin leerlingen moeten synthetiseren, hun gevonden resultaten omzetten naar een theorie. Wat betekent dat voor....? Kun je nu ook iets concluderen over.....?
25. De leerkracht stelt vragen als: kun je in het kort aangeven wat nu precies het probleem is en welke oplossing je uiteindelijk het beste vindt?
Differentiatie in leeromgeving
26. De tafels staan in groepjes opgesteld, in de klas staan en hangen diverse uitnodigende materialen, er is ruimte om te onderzoeken/ontdekken.
27. De les is goed voorbereid, materialen zijn al beschikbaar, leerlingen weten wat er van hen verwacht wordt.
28. De leerkracht is positief naar alle leerlingen.
29. De leerkracht neemt geen genoegen met eerste antwoord, vraagt door/verder, geeft zelf ook aanvullende voorbeelden of interpretaties.
30. De leerkracht waardeert de inbreng van de verschillende leerlingen.
31. De leerkracht laat de leerlingen op een positieve manier merken dat hij verwacht dat de leerlingen ver zullen komen bij een taak of opdracht.
32. De leerkracht stelt leerlingen vragen die erop gericht zijn om hen zelf te laten nadenken over planning of volgen van het leren.
33. De leerkracht vraagt leerlingen wat het betekent wat ze geleerd hebben, vraagt of ze kunnen reproduceren wat er geleerd is, vraagt of het doel bereikt is van de les, vraagt wat ze nog niet weten, waar nog vragen liggen.

Bijlage III. Schema van de analyse ICALT – STIP instrument

ICALT	Uitwerking	Verdieping	STIP -item	STIP-categorie
Efficiënte lesorganisatie	Zorgt voor een doelmatig klassenmanagement	1 de lesmaterialen zijn afgestemd op het niveau en ontwikkeling van leerlingen	De leerkracht heeft voor moeilijk lerende of lezende leerlingen makkelijker teksten ingezet.	Maakt gebruik van verschillende bronnen die zijn afgestemd op de leerbehoeften van de leerlingen.
			De leerkracht heeft voor makkelijk lezende of lerende leerlingen moeilijkere teksten ingezet.	Maakt gebruik van verschillende bronnen die zijn afgestemd op de leerbehoeften van de leerlingen.
Duidelijke en gestructureerde instructie	geeft duidelijke uitleg van de leerstof	2 activeert de voorkennis van de leerlingen	De leerkracht vraagt expliciet naar de voorkennis over dit onderwerp: wat weet je er al van? Helpt voorkennis activeren.	De leerkracht bepaalt wat leerlingen al over de te behandelen leerinhoud weten.
Intensieve en activerende les	Hanteert werkvormen die leerlingen activeren	3 gebruikt een variëteit aan instructiestrategieën	De leerkracht werkt bij leerlingen die nog niet veel van het onderwerp weten of het moeilijk vinden met veel voorbeelden, vergelijkingen en kleine stapjes.	Legt leerinhoud op verschillende manieren uit zodat centrale begrippen goed worden overgebracht.
			De leerkracht biedt leerlingen die al veel weten of makkelijk leren meer ruimte voor	Legt leerinhoud op verschillende manieren uit zodat centrale begrippen goed worden

			verbinding, neemt grotere stappen, laat sneller los.	overgebracht.
	stimuleert het zelfvertrouwen van zwakke leerlingen	6 uit bij zwakke leerlingen positieve verwachtingen over wat ze gaan doen	De leerkracht laat de leerlingen op een positieve manier merken dat hij verwacht dat de leerlingen ver zullen komen bij een taak of opdracht.	De leerkracht heeft hoge verwachtingen van de leerlingen.
		7 geeft zwakke leerlingen complimenten over hun werk	De leerkracht is positief naar alle leerlingen.	De leerkracht accepteert de gevoelens van leerlingen op een positieve manier.
Afstemmen van instructie en verwerking op verschillen	gaat na of de lesdoelen werden bereikt	8 gaat na of de doelen van de les zijn bereikt	De leerkracht vraagt leerlingen wat het betekent wat ze geleerd hebben, vraagt of ze kunnen reproduceren wat er geleerd is, vraagt of het doel bereikt is van de les, vraagt wat ze nog niet weten, waar nog vragen liggen.	De leerkracht reflecteert met de leerlingen op het geleerde.
		9 gaat na wat de prestaties van de leerlingen zijn	De leerkracht vraagt leerlingen wat het betekent wat ze geleerd hebben, vraagt of ze kunnen reproduceren wat er geleerd is, vraagt of het doel	De leerkracht reflecteert met de leerlingen op het geleerde.

			bereikt is van de les, vraagt wat ze nog niet weten, waar nog vragen liggen.	
	biedt zwakke leerlingen extra leer- en instructietijd	10 geeft zwakke leerlingen extra leertijd	<i>Er is geen item dat overeenkomt met item 10 van de ICALT. Item van STIP die hieronder staat komt het dichtst in de buurt.</i>	
		11 geeft zwakke leerlingen extra instructietijd	Leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.	Laat leerlingen de leerinhoud ontdekken op een wijze die past bij hun voorkennis.
		12 geeft zwakke leerlingen extra oefeningen	Leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.	Leerkracht laat leerlingen de leerinhoud ontdekken op een wijze die past bij hun voorkennis.
		13 geeft zwakke leerlingen 'voor'- of 'na'-instructie	Leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid	Leerkracht laat leerlingen de leerinhoud ontdekken op een wijze die past bij

			instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.	hun voorkennis.
stemt de instructie af op relevante verschillen tussen leerlingen		14 zet leerlingen die minder instructie nodig hebben (alvast) aan het werk	Leerkracht biedt leerlingen die al veel weten of makkelijk leren meer ruimte voor verbinding, neemt grotere stappen, laat sneller los.	Leerkracht legt leerinhoud op verschillende manieren uit zodat centrale begrippen goed worden overgebracht.
		15 geeft aanvullende instructie aan groepjes of individuele leerlingen	De leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.	De leerkracht laat leerlingen de leerinhoud ontdekken op een wijze die past bij hun voorkennis.
		16 richt zich niet alleen op de middenmoot	De leerkracht werkt bij leerlingen die nog niet veel van het onderwerp weten of het moeilijk vinden met veel voorbeelden, vergelijkingen en kleine stapjes.	De leerkracht legt leerinhoud op verschillende manieren uit zodat centrale begrippen goed worden overgebracht.
			De leerkracht biedt leerlingen die al veel weten of makkelijk leren	De leerkracht legt leerinhoud op verschillende manieren uit zodat

			meer ruimte voor verbinding, neemt grotere stappen, laat sneller los.	centrale begrippen goed worden overgebracht.
	stemt de verwerking van de leerstof af op relevante verschillen tussen lln.	17 maakt tussen leerlingen verschil in de omvang van opdrachten	Leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.	Laat leerlingen de leerinhoud ontdekken op een wijze die past bij hun voorkennis.
		18 geeft niet alle leerlingen dezelfde tijd voor de opdracht	Leerkracht biedt leerlingen die nog onbekend zijn met het onderwerp uitgebreid instructie bij het onderwerp en opdrachten die de leerling helpen zich de stof eigen te maken.	Laat leerlingen de leerinhoud ontdekken op een wijze die past bij hun voorkennis.
		19 laat sommige leerlingen gebruik maken van hulpmaterialen	De leerkracht heeft voor moeilijk lerende of lezende leerlingen makkelijker teksten ingezet.	Maakt gebruik van verschillende bronnen die zijn afgestemd op de leerbehoeften van de leerlingen.
			De leerkracht heeft voor makkelijk lezende of lerende leerlingen moeilijkere teksten ingezet.	

			De leerkracht stelt moeilijk lerende leerlingen bijvoorbeeld vooral vragen die....	Zorgt voor opdrachten die afgestemd zijn op de leerbehoeften van leerlingen.
			De leerkracht stelt makkelijk lerende leerlingen bijvoorbeeld vooral vragen die..	