



Universiteit Utrecht
Opleiding MSc Logopediewetenschap
Clinical Language, Speech, and Hearing Sciences

Master's Thesis

**De kwaliteit van leerkrachtoordelen over de
aanwezigheid van taalbegripsproblemen bij
kleuters**

Aysa Dawson
3449688

Supervisie:

Judith Pijnacker, eerste begeleider
Nina Davids, eerste begeleider
Frank Wijnen, tweede begeleider
Rianneke Crielaard, opdrachtgever

12 augustus 2013

Samenvatting

Door het steeds vaker schrappen van de preventieve logopedie bij 5-jarige kinderen, is een belangrijke rol ontstaan voor leerkrachten bij de signalering van taalontwikkelingsstoornissen. In deze studie is onderzocht in hoeverre leerkrachten goed kunnen inschatten of een kind taalbegripsproblemen heeft. Daarnaast is onderzocht of leerkrachten hier meer problemen mee hebben als een kind ook andere problemen heeft, zoals een beneden gemiddelde intelligentie, gedragsproblemen en/of aandachtsproblemen. Hiertoe zijn leerkrachtoordelen over het taalbegripsniveau van 94 kinderen in de kleuterleeftijd, bij wie er een vermoeden was van spraak-taalproblematiek, vergeleken met uitkomsten op formele taalbegripstesten. De resultaten tonen aan dat er een positieve relatie is tussen het leerkrachtoordeel en de scores op taalbegripstesten. Echter, de leerkrachtoordelen wijken toch regelmatig af van de formele testuitkomsten. 80% van de kinderen met taalbegripsproblemen wordt als zodanig door de leerkracht herkend. Van de kinderen zonder taalbegripsproblemen wordt slechts 50% als zodanig herkend door de leerkracht. Er zijn in dit onderzoek geen aanwijzingen gevonden dat leerkrachten meer moeite hebben met het inschatten van taalbegripsproblemen als een kind andere stoornissen heeft zoals een beneden gemiddelde intelligentie, aandachtsproblemen en gedragsproblemen. Het huidige onderzoek vormt een startpunt voor de validering van leerkrachtoordelen over de aanwezigheid taalontwikkelingsstoornissen bij kleuters. Toekomstig onderzoek zal zich vooral moeten richten op de ontwikkeling van een gevalideerde en genormeerde leerkrachtvragenlijst die in de toekomst gebruikt kan gaan worden als screeningsinstrument waarmee taalontwikkelingsproblemen gesignaleerd kunnen worden bij Nederlandse kinderen.

Introductie

Een specifieke taalontwikkelingsstoornis is één van de meest voorkomende ontwikkelingsstoornissen bij kinderen. Ongeveer 7% van de 5- en 6-jarige kinderen in de Verenigde Staten hebben specifieke problemen in de taalontwikkeling

(Tomblin et al., 1997). Getallen voor de Nederlandse situatie zijn onbekend. De meest gangbare Nederlandse definitie van een specifieke taalontwikkelingsstoornis is een stoornis in de opbouw van het taalsysteem die niet verklaarbaar is door gehoorproblemen, intelligentieproblemen, motorische problemen, neurologische stoornissen of door emotionele en gedragsmatige problemen (Jong, 1999). Door de gestoorde opbouw van het taalsysteem heeft het kind problemen met het praten en/of met het begrijpen van taal. Uit onderzoek van Conti-Ramsden en Botting (2004) blijkt dat kinderen met taalontwikkelingsstoornissen een aanzienlijk groter risico lopen om later sociale en gedragsmatige problemen te ontwikkelen. Daarnaast hebben kinderen met taalproblemen een verhoogde kans op leesproblemen op latere leeftijd (Catts, Fey, Tomblin, & Zhang, 2002). Vroegtijdige signalering van taalontwikkelingsstoornissen is dus van groot belang om het kind tijdig adequate zorg te kunnen verlenen.

Taalontwikkelingsstoornissen hebben verschillende verschijningsvormen. Sommige kinderen hebben alleen problemen in de taalproductie. Dat wil zeggen dat ze alleen problemen hebben met het praten. Andere kinderen hebben daarnaast ook problemen in het taalbegrip. De DSM-IV (American Psychiatric Association, 1994) maakt onderscheid tussen expressieve en gemengd receptief-expressieve taalstoornissen. Bij een expressieve taalstoornis heeft het kind problemen in taalproductie en een leeftijdsadequaat taalbegripsniveau. Bij een receptief-expressieve taalstoornis zijn er zowel taalproductie- als taalbegripsproblemen. Geschat wordt dat 5% van kleuters een gemengd receptief-expressieve taalstoornis heeft en 3% tot 7% van de schoolgaande kinderen hebben alleen expressieve taalproblemen (American Psychiatric Association, 1994).

Expressieve taalproblemen, zoals problemen in de articulatie of zinsbouw, zijn voor de omgeving meestal direct waarneembaar. Taalbegripsproblemen zijn daarentegen moeilijker waarneembaar en kunnen alleen worden aangetoond met een genormeerde taalbegripstest. Taalbegripsproblemen komen meestal tot uiting aan de hand van secundair gedrag: het kind reageert niet op gesproken taal, het kind volgt instructies verkeerd op, het kind luistert niet of is snel afgeleid. Ten onrechte zou het kind dan door de omgeving bestempeld worden met 'slechte luisterhouding', 'aandachtsproblemen' of 'gedragsproblemen', terwijl een taalbegripsprobleem de oorzaak is van het zichtbare gedrag. Het herkennen van taalbegripsproblemen door de omgeving van het kind is dus lastig.

Naast kinderen met specifieke taalontwikkelingsstoornissen, zijn er veel kinderen die een taalontwikkelingsstoornis hebben in combinatie met nevenstoornissen (Goorhuis-Brouwer & Wijnberg-Williams, 1996). Zo worden taalproblemen vaak gezien bij kinderen met autisme (Kjelgaard & Tager-Flusberg, 2001; Mitchell et al., 2006), ADHD (Bruce, Thernlund, & Nettelblatt, 2006) en bij kinderen met algemene leerproblemen (Tomblin & Zhang, 1999). Taalontwikkelingsstoornissen en in het bijzonder taalbegripsproblemen zijn bij deze kinderen soms extra lastig te herkennen, omdat de nevenstoornis het meest op de voorgrond treedt waardoor het taalprobleem over het hoofd gezien wordt. Bij kinderen met bijzonderheden in het gedrag (bijvoorbeeld het niet goed opvolgen van instructies of het snel afgeleid zijn) is het vaak niet duidelijk wat de oorzaak (een taalbegripsprobleem of een andere stoornis) van het gedrag is.

Op een audiologisch centrum, zoals het Kentalis Audiologisch Diagnostisch Centrum in Sint-Michielsgestel, is men gespecialiseerd in het uitvoeren van spraak-taaldiagnostiek. Als een kind aangemeld wordt voor spraak-taalonderzoek, vindt er eerst een uitgebreide anamnese plaats, waarbij met vragenlijsten voor zowel ouders als leerkrachten en een intakegesprek zoveel mogelijk informatie over het kind verzameld wordt. Daarna vindt een multidisciplinair onderzoek plaats, waarbij de volgende disciplines betrokken kunnen zijn: logopedie, gedragswetenschappen, maatschappelijk werk, audiologie en KNO-geneeskunde en eventueel de jeugdpsychiatrie. Op deze manier wordt het kind breed onderzocht. Taalontwikkelingsstoornissen worden vastgesteld door logopedisch onderzoek. De logopedist gebruikt gestandaardiseerde en genormeerde taaltesten die het taalniveau van een kind in kaart brengen in vergelijking met dat van leeftijdgenootjes. Er zijn in Nederland zowel taalproductie- als taalbegripstesten beschikbaar. In de klinische praktijk wordt aangenomen dat een kind een taalprobleem heeft als het minimaal één standaarddeviatie onder het populatiegemiddelde haalt. De aanwezigheid van (signalen voor) intelligentieproblemen, aandachtsproblemen en gedragsproblemen wordt vastgesteld door een gedragswetenschapper. Voordat een kind wordt aangemeld bij het audiologisch centrum voor uitgebreide spraak-taaldiagnostiek, moeten er eerst serieuze signalen voor taalproblemen zijn herkend door iemand uit de omgeving van het kind.

Pas als een mogelijk probleem in de taalontwikkeling van het kind gesignaleerd is, komt het kind in aanmerking voor spraak-taaldiagnostiek bij een Audiologisch Centrum. Signalering van taalontwikkelingsstoornissen kan plaats vinden met behulp van spraak-taalscreening. Zo een screening is een verkorte vorm van diagnostiek bij er een inschatting gemaakt of het kind een normale spraak-taalontwikkeling heeft of dat er aanwijzingen zijn voor spraak-taalproblematiek. Spraak-taalscreening is minder tijdrovend dan diagnostiek en kost aanzienlijk minder geld. Bij spraak-taalscreeningen en de signalering van spraak-taalproblemen is het gebruikelijk de uitkomst uit te drukken in de termen 'pluis' en 'niet-pluis'. Als de uitkomst 'pluis' is, zijn er geen aanwijzingen voor spraak-taalproblemen. Als de uitkomst 'niet pluis' is, zijn er wel aanwijzingen voor problemen en is uitgebreidere diagnostiek door bijvoorbeeld een audiologisch centrum gewenst.

In het Basistakenpakket Jeugdgezondheidszorg staat vermeld dat de gemeenten verantwoordelijk zijn voor de spraak-taalscreening bij 5-jarige kinderen (NVLF, 2010). Bij deze screening wordt beoordeeld of het spraak-taalniveau 'pluis' is en dus leeftijdsadequaat, of dat het spraak-taalniveau 'niet-pluis' is en het kind een verhoogd risico heeft op lees- en spellingsproblemen. Naar aanleiding van de uitkomst van deze screening kunnen kinderen met spraak-taalproblemen nog vóór groep 3, waarin het lees- en spellingsonderwijs start, nader onderzocht en eventueel behandeld worden.

In het verleden werd de 5-jarigenscreening uitgevoerd door logopedisten die gespecialiseerd zijn in spraak-taalproblematiek. De afgelopen jaren snijden veel gemeenten in deze vorm van preventieve logopedie als onderdeel van bezuinigingsplannen. Steeds meer gemeenten kiezen ervoor om de spraak-taalscreening bij 5-jarigen niet meer door de logopedist uit te laten voeren. In plaats daarvan wordt de screening afgenomen door een schoolarts. Het is onbekend in hoeverre schoolartsen geschikt zijn om spraak-taalproblemen te signaleren. Het is mogelijk dat taalontwikkelingsproblemen, behalve door schoolartsen, ook door ouders en leerkrachten gesignaleerd worden. Bekend is dat schoolartsen, leerkrachten en ouders over het algemeen niet gespecialiseerd zijn in de diagnostiek van spraak-taalproblemen.

Leerkrachten zouden in de nieuwe situatie die ontstaat door het schrappen van de preventieve logopedie bij 5-jarigen, een belangrijke rol kunnen gaan spelen in de vroege signalering van spraak-taalproblemen. Leerkrachten hebben dagelijks

contact met kinderen. Daarnaast zijn zij, mede door het routinematig bijhouden van de voortgang van leerlingen, goed in staat om vergelijkingen te maken tussen leerlingen. Als leerkrachten in de toekomst een sleutelfunctie moeten gaan vervullen in de signalering van taalproblemen, is het van groot belang te weten of leerkrachten goed het taalniveau van kinderen in kunnen schatten. Zoals eerder beschreven, is dat makkelijker voor expressieve taalproblemen dan voor taalbegripsproblemen. Het is daarom in het bijzonder van belang uit te zoeken of leerkrachtoordelen over de aanwezigheid van taalbegripsproblemen bij een kind valide zijn.

Bij de validering van meetinstrumenten wordt het te valideren instrument vaak vergeleken met een 'gouden standaard' die hetzelfde construct meet. Bij de validering van leerkrachtoordelen over de aanwezigheid van taalbegripsproblemen, kunnen genormeerde en gevalideerde taalbegripstesten gebruikt worden als gouden standaard. Als er overeenstemming is tussen de gouden standaard het leerkrachtoordeel, kan men spreken van een valide leerkrachtoordeel.

In Nederland zijn er verschillende taalbegripstesten beschikbaar die het taalbegripsniveau van een kind meten. De bekendste zijn de Reynell Test voor Taalbegrip (RTTB; Eldik et al., 1995), de Schlichting Test voor Taalbegrip (STTB; Schlichting & Lutje Spelberg, 2010), de subtesten Begrijpen en Aanwijzingen Volgen (BAV) en Zinsbegrip van de Clinical Evaluation of Language Fundamentals 4 Nederlandse versie (CELF-4-NL; Semel, Wiig, & Secord, 2010) en de subtesten Zinsbegrip 1 (ZB1) en Zinsbegrip 2 (ZB2) van de Taaltoets Alle Kinderen (TAK; Verhoeven & Vermeer, 2001). Deze testen hebben globaal dezelfde opzet: al deze testen meten het taalbegrip op zinsniveau. Het kind krijgt steeds mondeling een zin, vraag of opdracht aangeboden, waarna het vervolgens het juiste plaatje moet kiezen of een adequate handeling moet verrichten. De COTAN-beoordeling van deze testen is op de meeste onderdelen 'voldoende' of 'goed'¹. Tabel 1 beschrijft de kenmerken van de afzonderlijke taalbegripstesten.

¹ De criteriumvaliditeit van de RTTB, STTB en CELF-4-NL is beoordeeld als 'onvoldoende' omdat hiernaar geen onderzoek verricht is. De begripsvaliditeit van de STTB is beoordeeld als 'onvoldoende' omdat gegevens over de interne structuur ontbreken. De betrouwbaarheid van de subtest ZB van de CELF-4-NL is beoordeeld als 'onvoldoende'. Alle overige onderdelen zijn als 'voldoende' of 'goed' beoordeeld.

Test	Omschrijving
Reynell test voor taalbegrip (Eldik et al., 1995), Schlichting test voor taalbegrip (Schlichting & Lutje Spelberg, 2010)	<p>De Reynell en Schlichting test voor taalbegrip hebben veel overeenkomsten. De Schlichting is een vernieuwde versie van de Reynell test voor taalbegrip. De Reynell test is bestemd voor kinderen van 1;2 tot 6;3 jaar. De Schlichting test is voor kinderen van 2;0 tot 7;0 jaar. Bij beide testen krijgt het kind op gestructureerde wijze bij speelgoed of bij prenten steeds een mondelinge opdracht die het kind moet uitvoeren (bijvoorbeeld "Zet de dieren hier op een rij" en "Maak nu een torentje van drie zwarte stenen"). De opdrachten zijn bij aanvang van de test eenvoudig en worden naarmate de test vordert steeds moeilijker. Op basis van de behaalde score en de leeftijd van het kind, wordt het taalbegripsquotiënt (TBQ) berekend. Men spreekt van een gemiddeld taalbegripsniveau als het TBQ tussen de 85 en 115 ligt.</p>
Clinical Evaluation of Language Fundamentals 4 – Nederlandse versie (Semel, Wiig, & Secord, 2010)	<p>Testbatterij voor kinderen van 5 tot 15 jaar die zowel de algemene taalvaardigheid als specifieke vaardigheden met betrekking tot taal meet. De batterij omvat 18 subtesten. Per subtest wordt de totaalscore vergeleken met die van leeftijdsgenootjes en omgerekend tot een normscore. Een normscore tussen 7 en 13 wordt beschouwd als gemiddeld. Een lagere score wijst op een probleem in de betreffende vaardigheid.</p> <ul style="list-style-type: none"> <li data-bbox="236 965 496 1144">- Begrippen en aanwijzingen volgen Subtest voor kinderen van 5 tot 12 jaar met 49 items waarbij het kind een boek met steeds meerdere plaatjes voor zich krijgt. De onderzoeker noemt een aantal van de plaatjes, waarna het kind de plaatjes in de juiste volgorde moet aanwijzen, bijvoorbeeld "Wijs het huis aan nadat je de auto's hebt aangewezen" en "Het eerste huis en daarna de laatste schoen". <li data-bbox="236 1178 496 1301">- Zinnen begrijpen Subtest voor kinderen van 5 tot 8 jaar bestaande uit 27 items. Het kind krijgt steeds auditief een zin aangeboden, waarbij het kind uit vier plaatjes het juiste plaatje aan moet wijzen. Voorbeeld van aangeboden zinnen: "Het meisje heeft haar nieuwe regenjas aan, hoewel het niet nodig is".
Taaltoets Alle Kinderen (Verhoeven & Vermeer, 2001)	<p>Test voor kinderen van groep 1 tot en met groep 4. De behaalde scores door het kind worden vergeleken met één van de vijf groepen afhankelijk van de leeftijd en het leerjaar van het kind: groep 1 (gemiddelde leeftijd 4;7 jaar), begin groep 2 (5;6 jaar), eind groep 2 (6;3 jaar), eind groep 3 (7;3 jaar) en eind groep 4 (8;3 jaar). De score wordt omgerekend naar standaarddeviaties.</p> <ul style="list-style-type: none"> <li data-bbox="236 1581 496 1704">- Zinsbegrip 1: Functiewoorden Subtest die de kennis van functiewoorden toetst. Het kind krijgt steeds een zin auditief aangeboden waarbij het steeds uit drie plaatjes het juiste moet kiezen. Voorbeelden van opgaven: "Hij duwt haar" en "Hoewel het regende, ging hij fietsen". <li data-bbox="236 1738 496 1850">- Zinsbegrip 2: Zinspatronen Subtest met dezelfde opbouw als Zinsbegrip 1, alleen wordt bij deze subtest vooral kennis getoetst van patronen die grammaticale relaties tussen woorden in zinnen aangeven. Voorbeelden van opgaven: "De hond is niet groter" en "Met deze kar kan je niet meer rijden".

Tabel 1: Overzicht van de meest gebruikte Nederlandstalige taalbegripstesten op zinsniveau voor kleuters.

In Tabel 2 wordt een overzicht weergegeven hoe de leerkrachtoordelen geclassificeerd kunnen worden in relatie tot de uitkomsten de taalbegripstesten. Sensitiviteit en specificiteit zijn vaak gebruikte termen in valideringsonderzoek. Met sensitiviteit (te berekenen met de formule $A/(A+C)$) wordt de proportie bedoeld van de kinderen met een taalbegripsprobleem die als 'niet pluis' beoordeeld is. Hierbij gaat het dus om de nauwkeurigheid waarmee een leerkrachtoordeel de aangedane gevallen identificeert. Met specificiteit (te berekenen met de formule $D/(D+B)$) wordt aangeduid welk percentage van de kinderen zonder taalbegripsprobleem terecht als zodanig (pluis) beoordeeld is. Specificiteit en sensitiviteit betreffen dus de gevallen waarbij de leerkracht een juiste inschatting heeft gemaakt. Leerkrachtoordelen zijn valide, als zowel specificiteit als sensitiviteit hoog is.

		Aanwezigheid taalbegripsproblemen	
		Positief $Sd < -1$	Negatief $Sd \geq -1$
Leerkrachtoordeel	Positief 'niet pluis'	A (echt positief, terecht alarm)	B (fout positieven, loos alarm)
	Negatief 'pluis'	C (fout negatieven, gemiste gevallen)	D (echt negatief, terecht negatief)

Tabel 2: Classificatie van leerkrachtoordeel in relatie tot het behaalde testresultaat. De groengekleurde cellen betreffen de gevallen waarbij het leerkrachtoordeel overeenkomt met de uitkomst van de taalbegripstest en de roodgekleurde cellen betreffen de gevallen waarbij het leerkrachtoordeel en de testuitkomst niet overeenkomen.

Er is geen consensus over de drempelwaarden van specificiteit en sensitiviteit. In het geval van de validiteit van leerkrachtoordelen is het voornamelijk belangrijk dat sensitiviteit hoog is. Bij een sensitiviteit van 90%, wordt 90% van alle kinderen met een taalbegripsprobleem gesignaleerd. De overige 10% van de kinderen met een taalbegripsprobleem wordt door de leerkracht gemist en zou daardoor niet op tijd de juiste zorg kunnen krijgen. Anderzijds is het ook wenselijk dat de specificiteit zo hoog mogelijk is. Bij een specificiteit van 50% wordt de helft van de kinderen zonder taalbegripsproblemen onterecht als 'niet pluis' beoordeeld. Deze kinderen zouden dan onterecht doorgestuurd worden voor diagnostisch onderzoek dat (onnodig) veel tijd en geld kost. Het is dus van belang dat zowel de sensitiviteit als specificiteit zo hoog mogelijk zijn.

In de diagnostiek en behandeling van taalontwikkelingsstoornissen is de er de laatste jaren steeds meer aandacht ontstaan voor het dagelijks functioneren van het kind, naast de prestaties van het kind op formele taaltesten. Leerkrachten wordt steeds vaker gevraagd om informatie te verschaffen over de taalvaardigheden en -beperkingen van het kind tijdens verschillende klassikale situaties waarin taal gebruikt wordt, zoals kringgesprekken, voorlezen en in contact met andere kinderen. Ook door verschillende onderzoekers wordt bepleit dat het dagelijkse functioneren van een kind essentieel is bij onderzoek naar taalproblemen (Dockrell, George, Lindsay, & Roux, 1997; Huerta-Marcías, 1995; McNamara & Roever, 2006; Teasdale & Leung, 2000). Bij het beschrijven van het dagelijks functioneren, speelt het oordeel van de leerkracht een cruciale rol, omdat de leerkracht het kind dagelijks in verschillende situaties ziet. Verschillende onderzoekers zijn van mening dat uitsluitend psychometrische testgegevens geen volledig beeld geven van de vaardigheden van het kind (Teasdale & Leung, 2000). In de wetenschappelijke praktijk is er tot nu toe echter vooral veel aandacht besteed aan de ontwikkeling van validiteit van psychometrische maten, terwijl er weinig aandacht is besteed aan de ontwikkeling van validiteit van leerkrachtoordelen (Teasdale & Leung, 2000).

In het verleden hebben slechts enkele studies onderzocht hoe goed leerkrachten verschillende taalvaardigheden kunnen beoordelen bij hun leerlingen. In deze studies werden leerkrachtoordelen vergeleken met genormeerde psychometrische testgegevens. De resultaten zijn niet eenduidig. Uit de studie van Gilmore en Vance (2007) bleek dat leerkrachten uit Groot-Brittannië goed het taalbegripsniveau en het niveau van de auditieve aandacht in konden schatten bij 4- en 5-jarige kleuters. In deze studie werden de leerkrachtoordelen over luister- en taalbegripsvaardigheden van 56 kinderen vergeleken met genormeerde testcores. De deelnemende kinderen aan dit onderzoek vertoonden aanwijzingen voor auditieve verwerkings- en/of taalbegripsproblematiek. Bij beide problematieken heeft een kind problemen met het begrijpen van gesproken taal. Het verschil is dat er bij auditieve verwerkingsproblemen het verwerken van een auditief signaal problemen oplevert, terwijl bij taalbegripsproblemen het koppelen van een betekenis aan een geluidssignaal niet goed lukt. Leerkrachtoordelen werden in dit onderzoek gemeten aan de hand van een lijst met vragen die met een Likertschaal van één tot vijf beantwoord moesten worden. De leerkrachtoordelen over auditieve aandacht en taalbegripsvaardigheden correleerden positief met de genormeerde taalbegrips-

scores. Oordelen over auditieve verwerkingsvaardigheden zoals auditieve discriminatie correleerden echter niet met de testcores. Het gebrek aan correlatie hier, is mogelijk te verklaren doordat de validiteit van de auditieve discriminatietest zwak was. Deze studie laat dus zien dat leerkrachten in ieder geval goed kunnen inschatten of een kind problemen heeft in auditieve aandacht en taalbegrip.

Williams (2006) vergeleek leerkrachtoordelen over algemene taalvaardigheid (taalbegrip en taalproductie) met formele testresultaten van 60 kinderen van 3;6 tot 6;6 jaar oud van vijf verschillende scholen in West-Australië. De leerkrachten werd gevraagd om in totaal 40 kinderen met een zwakke taalvaardigheid en 20 kinderen met een normale taalontwikkeling te selecteren. De leerkrachten moesten zelf een inschatting maken of het kind zwakke taalvaardigheid had of niet. Deze inschattingen van de leerkrachten werden vergeleken met formele testuitkomsten. De resultaten toonden aan dat leerkrachten in staat zijn om kinderen met een zwakke taalvaardigheid te identificeren op basis van expressieve taalvaardigheden, met een sensitiviteit van 86% en een specificiteit van 68,2%. Dit betekent dat 86% van de kinderen met een zwakke taalvaardigheid correct gesignaleerd werd door de leerkracht. Van de kinderen met een normale taalontwikkeling, werd 68,2% correct beoordeeld door de leerkracht. Wanneer de jongste groep kinderen van 3;6 tot 4;6 jaar geëxcludeerd werden, steeg de sensitiviteit tot 92% en de specificiteit tot 85,7%. De resultaten suggereren dat leerkrachten een betere inschatting van het taalniveau kunnen maken naarmate de kinderen ouder zijn.

De resultaten van Antoniazzi, Snow en Dickson-Swift (2010) suggereren het tegenovergestelde: zij concludeerden dat leerkrachten uit Australië niet goed kunnen inschatten of kinderen taalproductieproblemen hebben. De onderzoekers vergeleken uitkomsten van de *Clinical Evaluation of Language Fundamentals* screeningtest, een verkorte versie van het gelijknamige onderzoeksinstrument (Semel, Wiig, & Secord, 2004), met uitkomsten van de door leerkrachten ingevulde Children's Communication Checklist-2 (Bishop, 2003) van in totaal 149 kinderen van 5;1 tot 6;6 jaar. Van alle kinderen met taalproductieproblemen werd slechts 40,6% herkend door de leerkracht. Dit betekent dat 59,4% van de kinderen met productieproblemen niet door de leerkrachten gesignaleerd werd.

In tegenstelling tot de beperkte hoeveelheid studies naar de validiteit van leerkrachtoordelen over taalvaardigheden van kleuters, zijn er in het verleden meerdere studies uitgevoerd waarin werd nagegaan in hoeverre leerkrachten

kunnen inschatten of een kind risico heeft op lees- en spellingsproblemen. Ook deze studies vertonen eveneens wisselende resultaten (Martin, 2005; Gijsel, Bosman, & Verhoeven, 2006). Omdat lees- en spellingsproblematiek weliswaar raakvlakken heeft met een taalontwikkelingsstoornis, maar niet dezelfde stoornis betreft, zeggen deze onderzoeksbevindingen weinig over de validiteit van leerkrachtoordelen over taalvaardigheden van kleuters.

Een mogelijke verklaring voor de inconsistente onderzoeksbevindingen zijn de verschillende onderzoeksdesigns. De leerkrachtoordelen worden in de verschillende studies op verschillende manieren verkregen. Er bestaat geen standaard leerkrachtvragenlijst waarmee leerkrachten de taalvaardigheden van kleuters kunnen beoordelen. Daarnaast verschillen de studies van elkaar in populatie. Zo ligt de focus bij Gilmore en Vance (2007) op kinderen waarbij al een vermoeden is van auditieve verwerkings- en/of taalbegripsproblematiek. In andere studies is een steekproef uit de normale populatie genomen. Verder is het aannemelijk dat leerkrachten in Australië verschillen van leerkrachten in de Verenigde Staten wat betreft kennis, opleiding, ervaring, werkwijze en taken.

Er is nog weinig bekend over de validiteit van Nederlandse leerkrachtoordelen over taalvaardigheden van kinderen. Het is de vraag of bevindingen uit buitenlandse studies ook geldig zijn voor de Nederlandse situatie, omdat bijvoorbeeld het onderwijssysteem, de scholing van leerkrachten en het screeningsbeleid van Nederland niet overeenkomt met dat van landen als Groot-Brittannië, de Verenigde Staten of Australië. Om een beter beeld te krijgen van de validiteit van leerkrachtoordelen met betrekking tot de taalvaardigheid (taalproductie en taalbegrip) van kinderen in de onderbouw van het Nederlandse basisonderwijs is verder onderzoek een vereiste. Er zijn internationaal nog maar weinig studies verricht naar de leerkrachtoordelen over zowel taalproductie als taalbegrip. De meeste studies naar de kwaliteit van leerkrachtoordelen richtten zich op lees- en spellingsproblemen. Voor zover bekend hebben alleen Gilmore en Vance (2007) onderzoek gedaan naar de validiteit van leerkrachtoordelen over taalbegripsproblemen, terwijl taalbegripsproblemen juist zo lastig waarneembaar zijn. Het huidige onderzoek vergelijkt leerkrachtoordelen over taalbegripsniveau van Nederlandse kleuters in het reguliere basisonderwijs met de uitkomsten van genormeerde taalbegripstesten.

Vraagstelling

Het primaire doel van deze studie is om te onderzoeken of leerkrachtoordelen over taalbegrip bruikbaar zijn bij de signalering van receptieve taalontwikkelingsstoornissen bij kleuters. Met andere woorden: er wordt nagegaan wat de relatie is tussen leerkrachtoordelen over het taalbegripsniveau van een kind en de uitkomsten op genormeerde taalbegripstesten. De eerste hoofdvraag in het huidige onderzoek luidt: *Is er overeenstemming tussen het oordeel van de leerkracht over het taalbegrip van een kind en de uitkomst op een taalbegripstest?*

Het is aannemelijk dat het leerkrachtoordeel over taalbegrip niet altijd overeenstemt met de uitkomsten op taalbegripstesten. Indien dit het geval is, zal nagegaan worden waardoor deze afwezigheid van overeenstemming veroorzaakt wordt. Een secundair doel van dit onderzoek is te inventariseren of leerkrachten meer moeite hebben met het inschatten van het taalbegripsniveau als er specifieke nevenstoornissen aanwezig zijn bij het kind, zoals een beneden gemiddelde intelligentie, concentratie- en aandachtsproblemen of gedragsproblemen. De tweede hoofdvraag in dit onderzoek is daarom: *Zijn nevenstoornissen van invloed op de relatie tussen leerkrachtoordeel en de uitkomst op een taalbegripstest?* De tweede hoofdvraag is exploratief van aard en geeft mogelijk richting aan vervolgonderzoek.

Methode

Deelnemers

In dit onderzoek is gebruik gemaakt van data van 94 kleuters die tussen 2009 en 2013 het Kentalis Audiologisch Diagnostisch Centrum in Sint-Michielsgestel hebben bezocht. Alle kinderen waren op het moment van onderzoek tussen 4;0 en 7;0 jaar oud. Alleen kinderen uit kleuterklassen in het Nederlandse reguliere basisonderwijs zijn geïnccludeerd. De gemiddelde leeftijd van de geselecteerde kinderen is 68 maanden (standaarddeviatie 8,6 maanden). Exclusiecriteria in dit onderzoek zijn: aanwezigheid van een perceptief gehoorverlies, syndromale aandoeningen (zoals Downsyndroom en velocardiofaciaal syndroom), neurologische afwijkingen en meertaligheid.

Alle data zijn verkregen uit dossiers van het Audiologisch Diagnostisch Centrum Kentalis in Sint-Michielsgestel. Alleen dossiers uit 2009 tot 2013 van kinderen die aan de hierboven genoemde eisen voldoen en bij wie ten minste één taalbegripstest (STTB, RTTB, BAV, ZB, ZB1 of ZB2) is afgenomen, zijn meegenomen. Een leerkrachtvragenlijst behoort tot de standaardprocedure van spraak-taaldiagnostiek bij dit audiologisch diagnostisch centrum. In deze leerkrachtvragenlijst staan vier vragen die verband houden met de taalbegripsvaardigheden van het kind.

In dit onderzoek is er geen controlegroep geïnccludeerd, aangezien het binnen het tijdsbestek van dit project praktisch niet haalbaar is geweest om voldoende zich normaal ontwikkelende kinderen te werven en te onderzoeken.

Dataverzameling

Van de Kentalis leerkrachtvragenlijst zijn vier meerkeuzevragen geselecteerd die betrekking hebben tot het taalbegrip en die daarmee relevant zijn voor dit onderzoek. Deze meerkeuzevragen, de antwoordmogelijkheden en de bijbehorende scores zijn weergegeven in Tabel 3. Verder zijn van elk kind de testuitslagen op de taalbegripstesten geregistreerd.

<p>a. Hoe is het begripsniveau van de leerling in vergelijking met het gemiddelde begripsniveau van zijn of haar klasgenoten?</p> <ol style="list-style-type: none"> 1. Lager 2. Beneden gemiddeld 3. Gemiddeld 4. Iets beter 5. Hoger
<p>b. Wanneer het kind een aantal opeenvolgende instructie krijgt, moeten deze nog eens herhaald worden.</p> <ol style="list-style-type: none"> 1. Vaak 2. Soms 3. Nooit
<p>c. Hoe vaak aarzelt de leerling of raakt de leerling in de war wanneer hij of zij mondelinge instructies opvolgt?</p> <ol style="list-style-type: none"> 1. Altijd 2. Vaak 3. Gemiddeld 4. Af en toe 5. Nooit
<p>d. Hoe volgt het kind kringgesprekken of verhaaltjes?</p> <ol style="list-style-type: none"> 1. Is niet bij het gesprek of verhaal betrokken 2. Volgt het gesprek of verhaal matig 3. Verliest wel eens de aandacht bij het gesprek of verhaal 4. Volgt zonder problemen het gesprek of verhaal

Tabel 3: De meerkeuzevragen en de bijbehorende antwoordmogelijkheden uit de Kentalis leerkrachtvragenlijst die relevant zijn in het huidige onderzoek. De nummers voor de antwoorden refereren aan het gewicht van het betreffende antwoord.

Naast het leerkrachtoordeel en de testuitkomsten, is bijgehouden of de gedragswetenschapper van het Audiologisch Diagnostisch Centrum sterke signalen heeft gezien die wijzen op de volgende nevenstoornissen: een beneden gemiddelde intelligentie (onderbouwd met een diagnostische test), aandachtsproblemen en gedragsproblemen.

Analyse

Als eerste stap van de analyse, zijn de uitkomsten op de taalbegripstesten vertaald naar een binaire variabele: positief (een taalbegripsprobleem) of negatief (geen taalbegripsprobleem). Als het kind op ten minste één van de taalbegripstesten meer dan één standaarddeviatie onder het gemiddelde gescoord heeft, is de uitkomst positief. De uitkomst is negatief in alle overige gevallen.

In dit onderzoek wordt het leerkrachtoordeel gemeten aan de hand van de vier vragen uit Tabel 3. Elk mogelijk antwoord is gekoppeld aan een score¹. Om de antwoorden op deze vragen te kunnen vergelijken met de uitkomsten op de taalbegripstesten, zijn de scores op de vier vragen omgerekend tot één variabele, namelijk een composietscore. Een composietscore is een variabele samenvoeging van meerdere variabelen. Bij de totstandkoming van deze score zijn de vragen gewogen op basis van het aantal antwoordmogelijkheden, zodat elke vraag even zwaar meetelt. Deze composietscore is een ratiovariabele met een bereik van 0 tot 10. Hoe lager de behaalde score, des te meer aanwijzingen zijn er voor taalbegripsproblemen. De berekening van deze score is als volgt^{2,3}:

$$\text{Composietscore} = 3,31 * \left(\frac{1}{5}a + \frac{1}{3}b + \frac{1}{5}c + \frac{1}{4}d - 0,98 \right)$$

Nadat de composietscore voor elk kind is berekend, is vervolgens de optimale grenswaarde bepaald die de 'pluis'-oordelen van de 'niet-pluis'-oordelen onderscheidt. Zoals besproken in de introductie, heeft een valide leerkrachtoordeel zowel een hoge sensitiviteit als specificiteit. Dus positieve ('pluis') leerkrachtoordelen zijn terecht positief en negatieve ('niet pluis') leerkrachtoordelen zijn terecht negatief. De optimale afkapwaarde wordt gevonden met behulp van een ROC-curve in het computerprogramma SPSS 21. Deze curve geeft de relatie weer tussen sensitiviteit en specificiteit voor verschillende waarden van de composietscore. De optimale afkapwaarde geeft een zo hoog mogelijke sensitiviteit bij een zo hoog mogelijke specificiteit.

Vervolgens zijn op basis van de gekozen grenswaarde de kinderen ingedeeld in twee groepen: de kinderen waarbij het leerkrachtoordeel 'pluis' (negatief) is en de kinderen bij wie het leerkrachtoordeel 'niet pluis' (positief) is. Samen met de binaire uitkomst op de taalbegripstesten zijn de kinderen hierna ingedeeld in een 2x2 kruistabel met vier classificatiegroepen: 'Terecht alarm', 'Loos alarm', 'Gemist' en 'Terecht negatief'. Het verband tussen de variabelen leerkrachtoordeel en de uitkomsten op de taalbegripstesten is statistisch getoetst met de X^2 -toets. De

¹ In enkele gevallen heeft de leerkracht meer dan één antwoord ingevuld. Indien dit het geval is, is in dit onderzoek de gemiddelde score op de vraag geregistreerd.

² De getallen 3,31 en 0,98 zijn dusdanig gekozen zodat het bereik van de composietscore van 0 tot 10 is.

³ a, b, c en d refereren aan de vragen in Tabel 3.

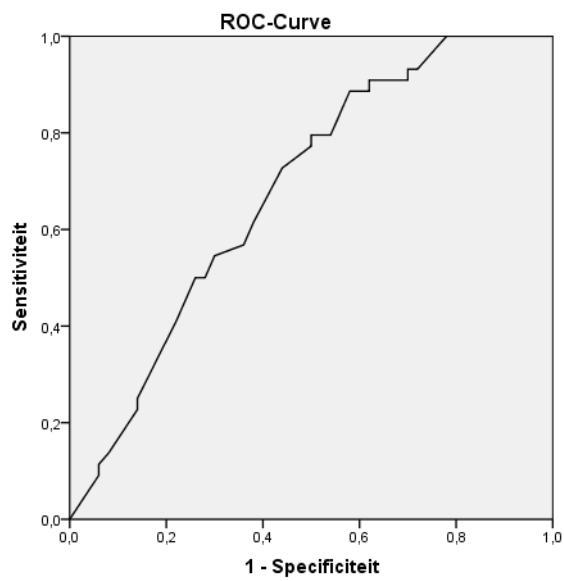
verwachting is dat er een relatie is tussen het leerkrachtoordeel en de uitkomst op de taalbegripstest en dat de X^2 -toets dus een significant verband als uitkomst heeft.

Indien het leerkrachtoordeel niet overeenstemt met de uitkomsten op de taalbegripstesten, is de mogelijke invloed van nevenstoornissen onderzocht door per classificatiegroep ('Terecht alarm', 'Loos alarm', 'Gemist' en 'Terecht negatief') de groepssamenstelling in kaart te brengen met een cirkeldiagram. Zo is te zien hoeveel procent van de kinderen in elke groep een beneden gemiddelde intelligentie, aandachtsproblemen, gedragsproblemen of een combinatie van deze nevenstoornissen. Ook hier is de X^2 -toets gebruikt om te kijken of nevenstoornissen significant vaker voorkomen in bepaalde groepen. De verwachting is dat nevenstoornissen vaker voorkomen in de groepen 'Loos alarm' en 'Gemist' dan in de andere twee groepen.

Resultaten

Bij alle 94 deelnemers is ten minste één van de taalbegripstesten uit Tabel 1 afgenomen. Bij 49 kinderen was de Schlichting of Reynell test voor taalbegrip afgenomen, bij 22 kinderen de subtest Begrippen en Aanwijzingen Volgen (CELF-4-NL), bij 12 kinderen de subtest Zinsbegrip (CELF-4-NL) en bij 22 kinderen zijn de zinsbegripstaken van de Taaltoets Alle Kinderen afgenomen.

Als eerste is bepaald welke kinderen een taalbegripsprobleem hebben. Van de in totaal 94 deelnemers, is gebleken dat 44 kinderen (46,8%) een taalbegripsprobleem heeft en dus meer dan één standaarddeviatie beneden het gemiddelde gescoord heeft op tenminste één van de taalbegripstesten. Vervolgens zijn voor elk kind de antwoorden op vier vragen uit de leerkrachtvragenlijst omgerekend naar een composietscore (gemiddelde: 5,19; standaarddeviatie: 1,55). De optimale grenswaarde is bepaald met behulp van een ROC-curve. De sensitiviteit en specificiteit van de leerkrachtoordelen is afhankelijk van deze grenswaarde. Figuur 1 toont de ROC-curve en Tabel 4 geeft de bijbehorende coördinaten weer.



Figuur 1: ROC-curve van het leerkrachtoordeel in verhouding met de aanwezigheid van taalbegripsproblemen. De ROC-curve laat de verhouding zien tussen sensitiviteit en 1-specificiteit naarmate de grenswaarde van de composietscore toeneemt. Bij een lage afkapwaarde is de sensitiviteit laag en de specificiteit hoog; bij een hoge afkapwaarde is dit omgekeerd.

Afkapwaarde	Sensitiviteit	1 – Specificiteit
0,0	0,0%	0
0,9	9%	6%
1,3	11%	6%
1,4	14%	8%
1,7	23%	14%
2,0	25%	14%
2,3	41%	22%
2,5	50%	26%
2,7	54%	30%
2,9	57%	36%
3,2	61%	38%
3,6	73%	44%
3,8	77%	50%
4,0	80%	50%
4,2	80%	54%
4,5	84%	56%
4,7	89%	58%
4,9	89%	62%
5,0	91%	62%
5,3	91%	64%
5,5	91%	68%
5,7	93%	70%
6,0	93%	72%
6,1	100%	76%

Tabel 4: Coördinaten van de ROC-curve met bijbehorende grenswaarden.

Uit Figuur 1 en Tabel 4 blijkt dat bij het verhogen van de grenswaarde de sensitiviteit afneemt en de specificiteit toeneemt. Er is hier geen grenswaarde waarbij zowel sensitiviteit als specificiteit hoog is. In het huidige onderzoek is een hoge sensitiviteit te verkiezen boven een hoge specificiteit, omdat een hoog aantal vals negatieven (kinderen met taalbegripsproblemen die niet door de leerkracht als zodanig herkend worden) vervelende gevolgen zoals het mislopen van uitgebreide diagnostiek of inadequate behandeling met zich mee kan brengen. Daarom wordt hier gekozen voor de grenswaarde 4,0 en waarbij de sensitiviteit 80% en de specificiteit 50% bedraagt.

Met de gekozen grenswaarde van 4,0 wordt 80% van alle kinderen met taalbegripsproblemen gesignaleerd door de leerkracht. De specificiteit van 50% is laag. Slechts de helft van alle kinderen zonder taalbegripsproblemen wordt door de

leerkracht terecht beoordeeld als 'pluis'; de andere helft krijgt onterecht een 'niet-pluis'-oordeel.

Nu de optimale grenswaarde van de composietscore bekend is, kunnen de leerkrachtoordelen over het taalbegripsniveau worden ingedeeld in twee categorieën: 'pluis' en 'niet pluis'. Als de composietscore gelijk is aan of lager is dan 4,0 is het leerkrachtoordeel 'niet pluis'. In alle andere gevallen is het leerkrachtoordeel 'pluis'. Van de 94 kinderen hadden 60 kinderen een 'niet pluis'-oordeel en 34 kinderen een 'pluis'-oordeel.

Op basis van de leerkrachtoordelen en de uitkomsten op de taalbegripstesten zijn alle kinderen in vier groepen geclassificeerd (Tabel 5). Het verband tussen leerkrachtoordelen en uitkomsten op alle taalbegripstesten is significant ($X^2 = 8,849$, $p = 0,003$). Dat betekent dat er een betekenisvol verband is tussen leerkrachtoordeel en de uitkomst op de taalbegripstesten.

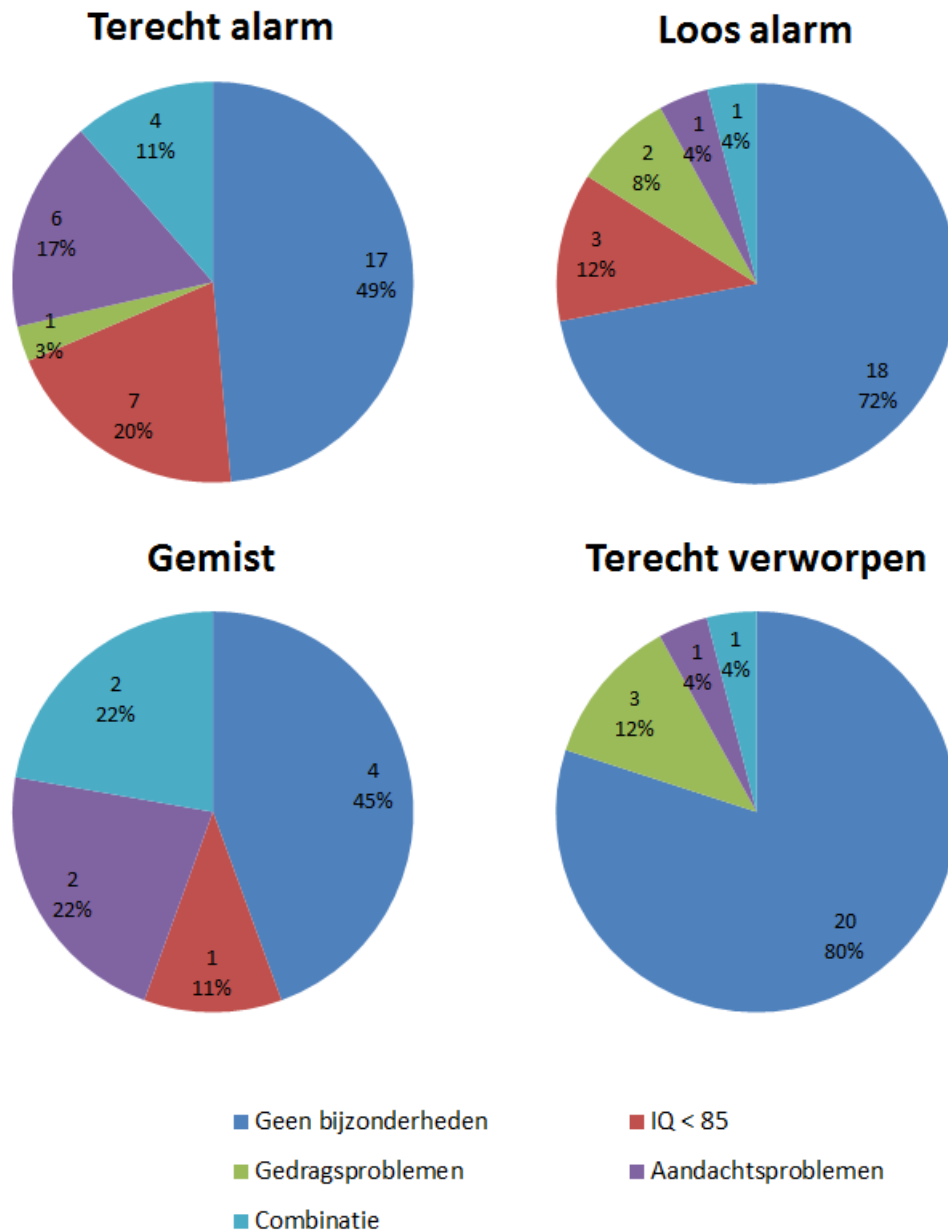
		Aanwezigheid taalbegripsproblemen		Totaal
		Positief Sd < -1	Negatief Sd ≥ -1	
Leerkracht- oordeel	Positief 'niet pluis'	Terecht alarm: 35 (37,2%)	Loos alarm: 25 (26,6%)	60
	Negatief 'pluis'	Gemist: 9 (9,6%)	Terecht negatief: 25 (26,6%)	34
Totaal		44	50	94

Tabel 5: Frequenties van de vier verschillende classificatiegroepen.

34 van de 94 leerkrachtoordelen (36,2%) komt niet overeen met de uitkomst op de taalbegripstesten. 26,6% van alle kinderen is onterecht als 'niet pluis' beoordeeld en 9,6% van alle kinderen had volgens de formele taalbegripstesten, de gouden standaard, een taalbegripsprobleem dat niet als zodanig is herkend door de

leerkracht. De vraag die nu rijst is: waarom is er bij deze kinderen een discrepantie tussen het leerkrachtoordeel en testuitkomsten? Zoals beschreven in de introductie, komen bij taalontwikkelingsstoornissen vaak nevenstoornissen voor. Het is denkbaar dat de aanwezigheid van nevenstoornissen het leerkrachtoordeel beïnvloedt.

Om de mogelijke invloed van nevenstoornissen in kaart te brengen, is er per classificatiegroep een cirkeldiagram van de groepssamenstelling gemaakt. Figuur 2 toont per groep aan bij welk percentage de nevenstoornissen beneden gemiddelde intelligentie, (sterke signalen voor) aandachtsproblemen, (sterke signalen voor) gedragsproblemen of een combinatie van twee of drie deze factoren, aanwezig is.



Figuur 2: Cirkeldiagrammen van de samenstelling van elke classificatiegroep. Donkerblauw vertegenwoordigt de kinderen bij wie geen van de onderzochte nevenstoornissen gevonden werden. Rood, groen en paars vertegenwoordigen respectievelijk de kinderen met een beneden gemiddelde intelligentie, (sterke signalen voor) voor gedragsproblemen en (sterke signalen voor) aandachtsproblemen. Lichtblauw staat voor de kinderen bij wie een combinatie van deze nevenstoornissen gevonden werd. De losse getallen in de diagrammen staan voor het aantal kinderen dat de conditie heeft binnen de groep. De percentages in de diagrammen geven de frequentie aan waarin de nevenstoornis binnen een classificatiegroep vertegenwoordigd is.

Ten eerste is het de vraag of in de groep ‘Gemist’ meer kinderen zitten met een nevenstoornis dan in de groep ‘Terecht alarm’. Dit blijkt niet het geval te zijn, want in de groep ‘Gemist’ heeft 55% een nevenstoornis en in de groep ‘Terecht alarm’ heeft 51% een nevenstoornis. Dit verschil is niet significant ($X^2 = 0,439$; $p = 0,508$). Ten tweede is het vraag of de aanwezigheid van een nevenstoornis ertoe leidt dat een leerkracht vaker ten onrechte denkt dat de leerling een taalbegripsprobleem heeft. In de groep ‘Loos alarm’ heeft 18% een nevenstoornis,

tegenover 20% in de groep 'Terecht negatief'. Ook dit verschil is niet significant (Fisher's Exact test¹: $p= 1,0$). Het lijkt er dus niet op dat leerkrachten in war gebracht worden in hun oordeel over het taalbegripsniveau als er nevenstoornissen aanwezig zijn.

Discussie

Omdat door de recente bezuinigingen van gemeenten op preventieve logopedie een belangrijke rol is ontstaan voor leerkrachten bij de signalering van taalontwikkelingsstoornissen bij kinderen, is in dit onderzoek nagegaan of leerkrachten in staat zijn in te schatten of een kind taalbegripsproblemen heeft. Doel van dit onderzoek is geweest te na te gaan of leerkrachtoordelen bruikbaar zijn in de signalering van taalbegripsproblemen. Hiertoe zijn leerkrachtoordelen over de aanwezigheid van taalbegripsproblemen bij kleuters vergeleken met formele testresultaten. Vervolgens is gekeken of nevenstoornissen als een beneden gemiddelde intelligentie, aandachtsproblemen en gedragsproblemen mogelijk van invloed zijn op de kwaliteit van leerkrachtoordelen.

In dit onderzoek is er een significant verband gevonden tussen de leerkrachtoordelen en de uitkomsten op de taalbegripstesten. Dit betekent dat er dus een positieve relatie is tussen het oordeel dat een leraar geeft over het taalbegripsniveau van het kind en de prestaties van het kind op formele taalbegripstesten.

Echter, in dit onderzoek zijn lage waarden voor sensitiviteit en specificiteit gevonden. De sensitiviteit van 80% is in deze situatie aan de lage kant, omdat 20% van de kinderen met taalbegripsproblemen niet als zodanig herkend wordt door de leerkracht. Deze kinderen zouden als niemand anders hun taalbegripsprobleem herkent, adequate behandeling op jonge leeftijd mislopen. De specificiteit met 50% is eveneens laag en houdt in dat de helft van de kinderen zonder taalbegripsproblemen terecht als zodanig beoordeeld wordt. Kinderen zonder taalbegripsproblemen hebben dus gelijke kansen op 'pluis'- of 'niet pluis'-oordeel te krijgen. De lage

¹ Hier is de Fisher's Exact test gebruikt omdat het aantal gegevens in de groep 'Gemist' te laag is om de X^2 -toets te kunnen toepassen.

specificiteit heeft eveneens ongewenste gevolgen: teveel kinderen zouden onterecht tijdrovende en kostbare diagnostiek krijgen.

In dit onderzoek is eveneens nagegaan of de lage specificiteit en sensitiviteit te verklaren zijn door de volgende co-morbide nevenstoornissen: een beneden gemiddelde intelligentie, aandachtsproblemen en gedragsproblemen. Er is gekeken of er in de groep met kinderen die onterecht door de leerkracht als 'pluis' zijn beoordeeld ('Vals alarm') vaker nevenstoornissen komen dan in de groep met kinderen die terecht als 'pluis' beoordeeld zijn ('Terecht negatief'). Ook is er gekeken of er in de groep met kinderen die onterecht als 'niet pluis' beoordeeld zijn ('Gemist'), vaker nevenstoornissen voorkomen dan in de groep met kinderen terecht als 'niet-pluis' beoordeeld zijn ('Terecht alarm'). Er zijn geen significante verschillen gevonden tussen de groepen wat betreft de prevalentie van nevenstoornissen. De resultaten ondersteunen dus niet de hypothese dat de lage specificiteit en sensitiviteit te verklaren zijn door nevenstoornissen.

Een andere mogelijke verklaring voor de beperkte overeenstemming tussen de leerkrachtoordelen en de uitkomsten formele testresultaten, kunnen omgevingsfactoren zijn. Omgevingsfactoren kunnen zowel de prestaties in de klas beïnvloeden als de prestaties tijdens de afname van de taalbegripstest. Voorbeelden van omgevingsfactoren zijn: een drukke klas, veel achtergrondlawaai in de klas, het kind voelt zich niet op zijn/haar gemak in de klas of het kind is zenuwachtig tijdens de onderzoekssituatie. De aanwezigheid van deze factoren is niet gemakkelijk te manipuleren en te meten, waardoor het lastig te bepalen is in hoeverre deze factoren van invloed zijn geweest.

Een derde mogelijke verklaring voor de lage sensitiviteit en specificiteit, is de beperkingen van het instrument waarmee de leerkrachtoordelen. De Kentalis leerkrachtvragenlijst is niet gevalideerd en uit de reacties van leerkrachten komt regelmatig naar voren dat de vragen niet altijd duidelijk geformuleerd zijn. Meerdere leerkrachten zetten een vraagteken bij de vraag of vulden meer dan twee antwoorden in, vooral bij de vraag "Hoe is het begripsniveau van de leerling in vergelijking met het gemiddelde begripsniveau van zijn of haar klasgenoten?". Bij deze vraag zorgden de antwoordmogelijkheden ("lager", "beneden gemiddeld", "gemiddeld", "iets beter" of "hoger") regelmatig voor verwarring. Uit de vragenlijst wordt niet duidelijk wat "lager", "beneden gemiddeld" en "gemiddeld" inhoudt. Ook is het aannemelijk dat leerkrachten niet altijd weten wat er "begripsniveau" bedoeld

wordt. Wellicht zouden de termen “taalbegripsniveau” of “het begrijpen van taal” hier duidelijker zijn geweest. Tevens een beperking aan de Kentalis leerkrachtvragenlijst als onderzoeksinstrument is de totstandkoming van de composietscore. In de berekening van de composietscore woog elke vraag even zwaar mee. Er is niet nagegaan of één vraag sterker correleerde met de testuitkomsten dan de andere. Verder is de composietscore uiteraard niet genormeerd en is optimale grenswaarde gekozen op basis van de uitkomsten van uitsluitend de deelnemers aan dit onderzoek. Een genormeerde, gevalideerde, helder geformuleerde vragenlijst zou waarschijnlijk een betere overeenstemming hebben opgeleverd.

De resultaten van dit onderzoek zijn in lijn met die van buitenlandse studies van Williams (2006) en Gilmore en Vance (2007). Uit het onderzoek van Williams (2006) kwam een sensitiviteit van 86% en een specificiteit van 68% naar voren voor leerkrachtoordelen over algemene taalvaardigheid (zowel taalproductie als taalbegrip). Deze waarden liggen dicht bij de verkregen waarden in het huidige onderzoek. Een belangrijke conclusie van Williams was, dat bij jonge kinderen (tussen 3;6 en 4;6) leerkrachten meer moeite hebben met het inschatten van de taalvaardigheden. Dit zou in het huidige onderzoek ook het geval kunnen zijn. Uit de studie van Gilmore en Vance (2007) bleek dat leerkrachten goed het taalbegripsniveau in konden schatten bij 4- en 5-jarige kleuters. Uit het onderzoek Antoniazzi, Snow en Dickson-Swift (2010) gericht op leerkrachtoordelen over taalproductie kwamen tegenovergestelde resultaten naar voren: zij concludeerden dat leerkrachten niet goed kunnen inschatten of een kind taalproductieproblemen heeft. De conclusie dat leerkrachten beter taalbegripsproblemen kunnen signaleren dan taalproductieproblemen is hoogstwaarschijnlijk te voorbarig. Mogelijk is, op de resultaten van Antoniazzi, Snow en Dickson-Swift (2010) van invloed geweest, het gekozen screeningsinstrument om de leerkrachtoordelen te meten (Children’s Communication Checklist-2; Bishop (2003)). Deze vragenlijst is namelijk niet primair voor leerkrachten bedoeld, maar voor de ouders van kinderen.

Een voorname beperking van dit onderzoek is de afwezigheid van een controlegroep. Bij de deelnemende kinderen aan deze studie was bij voorbaat al een vermoeden van taalontwikkelingsproblemen, omdat deze kinderen niet zonder reden aangemeld waren voor spraak-taalonderzoek bij het Kentalis Audiologisch Diagnostisch Centrum. Het is goed denkbaar dat dit het oordeel van de leerkracht

sterk beïnvloed heeft. Men moet dus voorzichtig zijn de resultaten van dit onderzoek te generaliseren naar zich normaal ontwikkelende kinderen.

Concluderend draagt dit onderzoek bij aan het inzetten van leerkrachten bij de signalering van taalontwikkelingsstoornissen bij de kleuters in Nederland. Door in de toekomst een heldere leerkrachtvragenlijst/ -screeningsinstrument te ontwikkelen met duidelijke instructies, is het aannemelijk dat zal blijken dat leerkrachten goed kunnen inschatten of een kleuter taalbegripsproblemen heeft. Verder zou specifieke training van leerkrachten in het signaleren van taalbegripsproblemen ook kunnen bijdragen aan een betrouwbaar leerkrachtoordeel.

Verder onderzoek zal zich voornamelijk moeten richten op de ontwikkeling van gevalideerde en genormeerde leerkrachtvragenlijst die in de toekomst mogelijk gebruikt kan gaan worden als screeningsinstrument. Ook zal er verder onderzoek verricht moeten worden naar zich normaal ontwikkelende kinderen, om te kijken of leerkrachten bij deze groep taalbegripsproblemen goed kunnen herkennen.

Literatuurlijst

- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)*. Washington, DC: American Psychiatric Association.
- Antoniazzi, D., Snow, P., & Dickson-Swift, V. (2010). Teacher identification of children at risk for language impairment in the first year of school. *International Journal of Speech-Language Pathology, 12*(3), 244-252.
- Bishop, D. (2003). *The Children's Communication Checklist CCC-2*. (2 ed.). London: Harcourt Assessment.
- Botting, N., Conti-Ramsden, G., & Crutchley, A. (1997). Concordance between teacher/therapist opinion and formal language assessment scores in children with language impairment. *European Journal of Disorders of Communication, 32*, 317-327.
- Bruce, B., Thernlund, G., & Nettelbladt, U. (2006). ADHD and language impairment. *European Child & Adolescent Psychiatry, 15*(1), 52-60.

- Catts, H., Fey, M., Tomblin, J., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*, 1142-1157.
- Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research, 47*, 145-161.
- Dockrell, J., George, R., Lindsay, G., & Roux, J. (2007). Problems in the identification and assessment of children with specific speech and language difficulties. *Educational psychology in practice, 13*(1), 29-38.
- Eldik, M. van, Schlichting, J., Lutje Spelberg, H., Meulen, B. van, & Meulen, S. van (1995). *Reynell Test voor Taalbegrip*. Nijmegen: Berkhout.
- Gilmore, J., & Vance, M. (2007). Teacher ratings of children's listening difficulties. *Child Language Teaching and Therapy, 23*(2), 133-156.
- Gijssel, M., Bosman, A., & Verhoeven, L. (2006). Kindergarten risk factors, cognitive factors and teacher judgements as predictors of early reading in Dutch. *Journal of Learning Disabilities, 39*, pp. 558-772.
- Goorhuis-Brouwer, S., & Wijnberg-Williams, B. (1996). Specificity of specific language impairment. *Folia Phoniatica et Logopaedica, 48*(6), 269-274.
- Huerta-Marcías, A. (1995). Alternative assessment: responses to commonly asked questions. *TESOL Journal, 5*, 8-11.
- Jong, J. de (1999). *Specific language impairment in Dutch: Inflectional morphology and argument structure*. Groningen: Grodil.
- Kjelgaard, M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16*, 287-308.
- Martin, S. (2005). *Examining the accuracy of teacher judgements of students' early literacy skills*. Bethlehem, PA: Lehigh University.
- Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., Smith, I., & Bryson, S. (2006). Early language and communication development of infants later diagnosed with autism spectrum disorder. *Journal of Developmental & Behavioral Pediatrics, 27*(2), 69-78.
- McNamara, T., & Roever, C. (2006). *Language Testing: The social dimension*. Malden, USA: Blackwell Publishing.

- NVLF (2011). Position statement Preventieve logopedie. Woerden, Nederland
Retrieved september 25, 2012, from:
http://www.logopedie.nl/bestanden/ikbenlogopedist/documentatiecentrum/Dossiers/Dossier_Preventieve_logopedie/
- Schlichting, L., & Lutje Spelberg, H. (2010). *Schlichting Test voor Taalproductie*. Houten: Bohn Stafleu van Loghum.
- Semel, E., Wiig, E., & Secord, W. (2004). *Clinical Evaluation of Language Fundamentals, fourth edition - Screening Test*. Toronto: The Psychological Corporation.
- Semel, E., Wiig, E.H., & Secord, W.A. (2010). *Clinical Evaluation of Language Fundamentals Nederlandse Versie (CELF-4-NL). Handleiding*. (Nederlandse bewerking van W. Kort, E. Compaan, M. Schittekatte en P. Dekker). Amsterdam: Pearson Assessment and Information.
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: a case of paradigm crossing? *Language testing*, 17(2), 163-184.
- Tomblin, J., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40(6), 1245-1260.
- Tomblin, J., & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In: Tager-Flusberg, H. (Ed), *Neurodevelopmental Disorders. Developmental Cognitive Neuroscience*, (361-382). Cambridge, MA, US: The MIT Press.
- Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen*. Arnhem: Citogroep Arnhem.
- Williams, C. (2006). Teacher judgements of the language skills of children in the early years of schooling. *Child Language Teaching and Therapy*, 22(2), 135-154.