

LITMUS SRep: een zinsherhaaltaak voor het Nederlands

Jan de Jong¹, Elma Blom², Chantal van Dijk³

¹*Institute for Biological and Medical Psychology, Universitetet i Bergen, Noorwegen*

²*Sociale Wetenschappen, Universiteit Utrecht*

³*Centre for Language Studies, Radboud Universiteit Nijmegen*

Samenvatting

In dit artikel wordt de Nederlandse versie beschreven van een zinsherhaaltaak die ontwikkeld is binnen de Europese COST Action IS0804. De benaming van de Nederlandse taak is LITMUS SRep NL. De taak stelt de gebruiker in staat om zinsherhaling bij meertalige kinderen te meten, door parallelle afname in de verschillende talen van het kind.

In het artikel wordt de Nederlandse taak verantwoord en van een wetenschappelijke achtergrond voorzien. Daarnaast worden de resultaten van de afname van de SRep in het Nederlands en in enkele andere talen gepresenteerd. Aan de hand van die resultaten kunnen enkele vragen over de validiteit van de taak worden beantwoord. Tevens wordt het effect van verschillende scoringsmethoden van de SRep besproken met in het bijzonder aandacht voor een scoringswijze die specifieke zinsstructuren centraal stelt.

Summary

In this article, we describe the Dutch version of a sentence repetition task developed within the European COST Action IS0804. The name of the Dutch task is LITMUS SRep NL. The task allows the user to measure sentence repetition in multilingual children, through parallel administration in the different languages of the child. The article accounts for the Dutch task and provides a scientific background. In addition, results are presented concerning administration of the SRep in Dutch and in some other languages. Based on these results, some questions about the validity of the task can be answered. The effect of different scoring methods of the SRep is also discussed, with specific attention for a scoring method that focuses on the target structure of the test items.

Correspondentieadres:

Dr. Jan de Jong

Institute for Biological and Medical Psychology
Universitetet i Bergen

Jonas Lies vei 91, 5020 Bergen, Noorwegen

E-mail: jan.jong@uib.no

Dit artikel is gelicentieerd onder de Creative Commons CC BY-NC-ND 4.0 (Naamsvermelding-NietCommercieel-GeenAfgeleideWerken) Internationale Licentie. Gebruik en distributie voor commerciële doeleinden en elke distributie van aangepast materiaal vereist schriftelijke toestemming.

Inleiding

Zinsherhaaltaken worden reeds lang gebruikt voor het vaststellen van taalproblemen bij kinderen. In het Nederlandse taalgebied vinden we ze bijvoorbeeld in de CELF 4, de CELF Pre-school en in de TAK (Verhoeven & Vermeer, 2006; Kort et al., 2008; Wiig et al., 2012). De herhaaltaak die in dit artikel wordt beschreven deelt belangrijke eigenschappen met deze tests, maar heeft een opzet die het tevens mogelijk maakt de uitkomsten voor tweetalige kinderen te vergelijken met de uitkomsten van een vergelijkbare test in hun moedertaal.

Herhaaltaken hebben een lange geschiedenis in wetenschappelijk onderzoek. In onderzoek naar verwerving vanuit een generatief taalkundig perspectief werd gezocht naar manieren om kindertaal te bestuderen zonder hinder van 'performance-factoren' - zoals geheugen, aandacht, gespreksomstandigheden - die de taal kunnen beïnvloeden. Spontane taal werd gezien als al te zeer gekleurd door zulke factoren. Grammaticaliteitsoordelen werden gezien als de meest zuivere ingang tot de *competence*, het taalvermogen. Zinsherhaling leek echter ook een goede kandidaat. De oorspronkelijke rationale van de herhaaltaak (Slobin & Welsh, 1973) luidde immers dat 'in repeating a sentence, one must filter it through one's own productive system'. Daarmee wordt zinsherhaling een meer directe maat voor de taalproductie en voor het eigen taalsysteem (de grammatica) van het kind.

Gewoonlijk wordt bij de validiteit van de stimuli in een herhaaltaak een voorbehoud geformuleerd: de te herhalen zin moet lang genoeg zijn. Zo niet, dan volstaat het kortetermijngeheugen voor de correcte herhaling van een zin en wordt het genoemde filter niet per se aangesproken. Klem et al. (2015) stellen in de titel van hun artikel vast: 'Sentence repetition is a measure of children's language skills rather than working memory limitations'. Die titel geeft de consensus over de taak adequaat weer. Bij een langere zin hebben kinderen de steun van de grammatica nodig. De aanname is dat kinderen dan geen zinnen kunnen herhalen waarvan ze de grammaticale structuur niet hebben verworven (Marinis & Armon-Lotem, 2015:97). De verwijzing naar specifieke structuren is hier van belang: de taak stelt de testontwerper in staat om items te kiezen die klinisch relevant zijn (omdat ze een taalontwikkelingsstoornis kunnen markeren) of die voorwerp van onderzoek zijn.

Niet alleen een adequate lengte is overigens een vereiste voor de items in een herhaaltaak. Zinnen verschillen ook in hoe complex hun structuur is. Als de zinnen in de taak onvoldoende complex zijn, ontstaat in een niet-taalgestoorde groep gemakkelijk een plafond-effect (Redmond, 2005).

Klinisch blijken herhaaltaken uiterst waardevol. Ze zijn een goede graadmeter voor het vaststellen van aan- of afwezigheid van een taalontwikkelingsstoornis. In dit verband wordt vaak een studie aangehaald van Conti-Ramsden, Botting en Faragher (2001), die vonden dat een lage score op de CELF-4 *Recalling Sentences* subtest (Semel, Wiig & Secord, 1992) een mogelijke markeerder is van taalproblemen bij kinderen. Om de diagnostische kwaliteit van testen te bepalen worden gewoonlijk de sensitiviteit (het vermogen van de test om

een stoornis te detecteren) en de specificiteit (het vermogen om de afwezigheid van een stoornis vast te stellen) berekend. Voor de zinsherhaaltaak vonden Conti-Ramsden et al. bij een groep eentalige Engels-sprekende kinderen een hoge sensitiviteit en specificiteit (van respectievelijk 90% en 85%).

In dit artikel wordt de Nederlandse versie van de LITMUS *Sentence repetition task* (hierna SRep NL) geïntroduceerd. De uitgangspunten en de ontwikkeling van de test worden beschreven. De test heeft verschillende scoringsprocedures die hieronder worden gespecificeerd. Aan de hand van eerste resultaten van afname wordt de betrouwbaarheid geïllustreerd en ook de samenhang tussen score en chronologische leeftijd. Tenslotte worden resultaten besproken van de afname van de LITMUS SRep bij meertalige kinderen. Daarmee wordt ook de gebruikswaarde van de test in een meertalige context geëvalueerd.

LITMUS SRep

In het kader van het Europese samenwerkingsverband COST Action IS0804 (*Language impairment in a multilingual society: Linguistic patterns and the road to assessment*) is een nieuwe zinsherhaaltaak ontwikkeld, die de eerdergenoemde rationale deelt: de vooronderstelling dat een herhaaltaak de grammaticale kennis aanspreekt en de eis dat de testitems voldoende lengte en complexiteit vertegenwoordigen. De gehele testbatterij die ontwikkeld is in de COST Action heet LITMUS, de afkorting van *Language Impairment Testing in Multilingual Settings*.

Zoals de meeste LITMUS-deeltests is de SRep (Marinis & Armon Lotem, 2015) specifiek bedoeld om het mogelijk te maken een kind in twee verschillende talen te testen. Het uitgangspunt van de LITMUS-batterij is dat afname van ieder testonderdeel mogelijk moet zijn in beide talen van een tweetalig kind. Als het kind in beide talen laag scoort op een (deel)test, dan is het vermoeden van een taalontwikkelingsstoornis gewettigd. Als het kind niet-afwijkend scoort in de ene taal, maar achterblijft in de andere, is er per definitie geen sprake van een stoornis: een taalontwikkelingsstoornis is altijd zichtbaar in beide talen (o.a. Kohnert, 2013). Bij successief-tweetalige kinderen bij wie de eerste taal dominant is, is de afname in die taal cruciaal: een niet-afwijkende score voor die taal spreekt het vermoeden van een taalstoornis tegen. Een LITMUS-taak, afgenomen in beide talen, stelt de gebruiker in staat dit belangrijke onderscheid te maken. Het is overigens niet altijd duidelijk welke taal de dominante is - dat kan variëren in de loop van de ontwikkeling. De mogelijkheid om parallelle tests voor beide talen af te nemen is daarom van groot belang.

Omdat de herhaalttest geschikt moest zijn voor parallelle afname in verschillende talen, moest bij de testconstructie een vergelijkbare mate van complexiteit in de items worden werkstelligd, ongeacht de testtaal. Om aan deze eis te voldoen werden twee uitgangspunten gekozen voor de constructie van testitems (Marinis & Armon-Lotem, 2015): alle versies van

de test moesten items bevatten die taalkundig complex zijn (ongeacht de taal) en daarnaast elementen die specifiek problematisch zijn voor (taalgestoorde) kinderen die de betreffende testtaal spreken. De herhaaltaak kan zo inderdaad dienen als een maat voor (zie boven) ‘specific structures that are being elicited’.

Taalkundige complexiteit wordt in de LITMUS SRep geoperationaliseerd door zinsstructuren te selecteren die twee taalkundige processen laten zien, afzonderlijk of in combinatie:

(1) verplaatsing (*movement*), waarbij een zinselement naar een andere positie in de zin wordt verplaatst - bijvoorbeeld bij vraagwoordzinnen. In de zin ‘Welke foto heeft hij gisteren gemaakt?’ staat het lijdend voorwerp ‘welke foto’ vooraan in de zin, in de mededelende zin ‘Hij heeft gisteren een foto gemaakt’ staat het onderwerp voorop en het lijdend voorwerp verderop in de zin. Er wordt aangenomen dat, om zo’n vraagzin te maken, het lijdend voorwerp moet worden verplaatst.

(2) inbedding (*embedding*), waarbij een ondergeschikte zin een zinsdeel is in de hoofdzin of waarbij een bijzin een zinsdeel specificiert. In de zin ‘De jongen die de postbode hielp is verdwaald’ is de betreffende bijzin ‘die de postbode hielp’ ingebed in de hoofdzin ‘De jongen is verdwaald’.

De aanname is dat deze twee taalkundige processen een zin complexer maken - en daarmee moeilijker correct te herhalen. Een combinatie van verplaatsing en inbedding vergroot de complexiteit nog meer. Marinis en Armon-Lotem (2015:101) specificeren de resulterende niveaus van complexiteit als volgt:

‘The language independent structures (are) monoclausal SVO structures (-embedding, -movement), object wh-questions (-embedding, +movement), bi-clausal sentences with coordination or subordination (+/-embedding, -movement), object relative clauses (+embedding, +movement), and conditionals (+embedding, -movement).’

De in het citaat aangehaalde taalkundige structuren en processen zijn verwerkt in de eerste, Engelstalige versie van de SRep. Bij de creatie van SRep-taken voor andere talen werden deze structuren zo veel mogelijk eveneens in de test opgenomen, maar de typologische contrasten leidden noodzakelijkerwijs tot verschillen. M. Paradis schreef ooit, in een verantwoording van zijn *Bilingual Aphasia Test*, die eveneens tracht talen te vergelijken (in dat geval de twee talen van een patiënt met afasie): “a sentence with structural complexity equivalent to an English passive (...) may require an altogether different construction in another language (e.g., a cleft object construction). In other words, if the equivalence criterion is syntactic complexity, a structure of similar complexity (quite possible not the passive) must be selected” (M. Paradis, 2004:74). Ook voor de SRep waren soms alternatieve structuren nodig als een structuur uit het (Engelse) origineel van de SRep in de doeltaal ontbrak. Door de zinnen voor alle testtalen te onderwerpen aan dezelfde eisen van complexiteit wordt een vergelijking tussen de uitkomsten van de testafname in twee talen betekenisvol. Op basis

van de genoemde variabelen zijn in de taak drie complexiteitsniveaus onderscheiden (zie ook Tabel 2 verderop).

Voor het vaststellen van taalspecifieke elementen is de belangrijkste bron van informatie de symptomatologie van taalontwikkelingsstoornissen (TOS). De elementen van de taal waarmee taalgestoorde kinderen moeilijkheden hebben, moeten aanwezig zijn in de herhaalzinnen. Symptomen verschillen al naar gelang de typologie van de taal (Leonard, 2014). Zo vormt werkwoordsmorfologie (congruentie, tijdsmarkering) een obstakel voor Nederlandse kinderen, terwijl in het Turks naamwoordsmorfologie (naamval) tot problemen leidt. Dat contrast leidt tot verschillen tussen de Nederlandse en Turkse versies van de SRep. Deze voorbeelden illustreren ook dat, terwijl de SRep complexiteitsniveaus heeft die worden bepaald door de syntaxis, de taalspecifieke elementen dikwijls (maar niet uitsluitend) de morfologie betreffen. Voor iedere versie van de LITMUS SRep worden zulke taalspecifieke onderdelen geïnccludeerd.

Niet voor alle talen is informatie beschikbaar over de symptomen van een TOS. In zulke gevallen zijn de elementen die door niet-taalgestoorde kinderen laat worden verworven bruikbare ingrediënten voor een herhaaltaak. Leonard (2014) benadrukt immers dat er een grote gelijkenis is tussen de grammaticale fouten die niet-taalgestoorde en taalgestoorde kinderen maken. Bij een taalontwikkelingsstoornis zijn de fouten echter persistent. Een andere informatiebron wordt gevormd door talen die typologisch verwant zijn. Zo is bij de totstandkoming van de Noorse SRep (Bome & Kongtorp Vangen, 2015) gebruik gemaakt van studies over TOS in het Zweeds.

De ambitie van de LITMUS SRep taak is dat, mits aan bovenstaande eisen is voldaan, er een valide herhaaltaak beschikbaar is voor beide talen van een tweetalig kind. Vooralsnog is er een groot aantal versies van de SRep, maar nog niet voor alle eerste (en tweede) talen van tweetalige kinderen is er een taak beschikbaar. Er zijn nu SRep-taken voor de volgende talen (Marinis & Armon-Lotem, 2015, hier aangevuld): Albaans, Libanees-Arabisch, Palestijns-Arabisch, Saoedi-Arabisch, Syrisch-Arabisch, Catalaans, Duits, Engels, Farsi, Frans, Iers (Gaelic), Grieks (en Cypriotisch Grieks), Hebreeuws, Iers, Italiaans, Kroatisch, Litouws, Malay, Nederlands, Noors, Pools, (Europees) Portugees, Russisch, Servisch, Turks en Welsh. SRep-taken voor andere talen zijn in voorbereiding (zie voor een actueel overzicht <https://www.bi-sli.org/sentence-repetition>).

Hoewel de test diagnostisch bruikbaar is voor een vergelijking tussen beide talen van het kind, is een vergelijking met niet-taalgestoorde leeftijdgenoten niet mogelijk: voor geen van de versies van de SRep zijn op dit moment normatieve data beschikbaar. Normatieve data zouden bovendien (voor het testen van een tweetalige groep) idealiter afkomstig moeten zijn van kinderen die tweetalig zijn. Een dergelijke normering is overigens geen eenvoudige exercitie. Tweetalige kinderen verschillen onderling in de mate waarin een van de talen dominant is en er zijn vele mogelijke taalcombinaties. De taak leent zich vooralsnog beter voor het beschrijven van de aard van de problemen op zinsniveau dan voor het vaststellen van

de stoornis.

Intussen zijn recente publicaties over de *diagnostic accuracy* van de test in andere talen hoopgevend. In een studie over de Russische en Hebreeuwse versies van de SRep bleken de sensitiviteit en specificiteit hoog (Armon-Lotem & Meir, 2016). Dat gold ook voor de Duitse SRep (Ibrahim & Fekete, 2019). Soortgelijke verkenningen vinden momenteel ook plaats voor de Nederlandse SRep.

Scoringsprocedures voor de SRep

De SRep taak kent diverse scoringsmethoden (Marinis & Armon-Lotem, 2015). De keuze van de scoringsmethode hangt af van het doel van de afname: wil de testgebruiker algemene informatie (over het vermogen om zinnen te herhalen) of informatie over specifieke elementen van de stimuluszinnen (grammaticale structuur, functiewoorden)? We beschrijven hier kort de scoringsprocedures die zijn gebruikt in studies die we verderop zullen aanhalen.

1. De *S(entence)-score*. De score is 1 als de zin woordelijk (*verbatim*) wordt herhaald, 0 als de uiting afwijkt van de modelzin.
2. De *Content word en Function word score*. Voor de *Content word score* wordt het aantal correct herhaalde *inhoudswoorden* bepaald en worden weglatingen, substituties en toevoegingen van inhoudswoorden geteld. Voor de *Function word score* wordt dezelfde procedure toegepast op *functiewoorden*. Deze twee scores werden ontleend aan een eerder ontwikkelde test, de *Sentence Imitation Test* (SIT) (Seeff-Gabriel, Chiat, & Roy, 2008).
3. De *G(rammaticaliteits)-score*. Een score van 1 wanneer de uiting van het kind grammaticaal correct is (ook als de zin niet woordelijk wordt herhaald); 0 punten als de zin ongrammaticaal is.
4. De *Doelstructuurscore*: als de doelstructuur van het item (zoals omschreven in SRep) correct wordt gebruikt 1 punt; zo niet, 0 punt.

De SRep is niet de enige herhaaltaak die verschillende scoringsprocedures kent. Zo vinden we in de verantwoording van de Nederlandse TAK Zinsvorming (Verhoeven & Vermeer, 2006) deze omschrijving: 'In de zinnen wordt nagegaan in hoeverre de leerlingen twintig syntactische elementen of functiewoorden correct kunnen gebruiken (...) In elke zin zitten ook (...) zinspatronen die gescoord worden.' Bij scoring van de test 'worden alleen de bedoelde functiewoorden of patronen goed of fout gerekend.' In de SRep kunnen functiewoorden apart worden geanalyseerd (scoringsmethode 2). De grammaticale doelstructuren die in de test zijn opgenomen kunnen eveneens afzonderlijk worden gescoord (scoringsmethode 4; scoringsmethode 3 meet de grammaticaliteit van de gehele geproduceerde zin, niet louter van de doelstructuur).

Wanneer de scores in verschillende talen worden vergeleken, is vooral de doelstructuur-score relevant: daarin is idealiter vertegenwoordigd wat de talen gemeen hebben. Eerder werd gerefereerd aan taalspecifieke ingredienten van de SRep. Die beïnvloeden de verbatim (S-)score. Op de structuurscore hebben ze echter idealiter geen effect. Zoals verderop wordt geïllustreerd, vraagt de structuurscore echter om goede analyse-instructies. Het is immers niet altijd duidelijk wanneer een doelstructuur als correct (herhaald) mag worden beschouwd.

Totstandkoming van LITMUS SRep NL: SRep met 60 items

De Nederlandse versie van de LITMUS SRep (SRep-NL) vindt haar oorsprong in een master-scriptie (Van de Scheur, 2012; de toen gebruikte benaming van de test was overigens SASIT: *School-Age Sentence Imitation Test*; Marinis et al., 2010). In deze studie werd aan de hand van een Nederlandse bewerking van de test een bescheiden pilotonderzoek uitgevoerd met 10 kinderen tussen 7 en 8 jaar oud. De betreffende test bevatte, net als het Engelse origineel, 60 items. Doel van het pilotonderzoek was primair om vast te stellen of de items adequaat waren - een evaluatie van de *face validity*.

Deze versie van de herhaaltaak werd door Van Assenbergh et al. (2015) afgenomen bij eentalige Nederlandse kinderen zonder TOS. De auteurs gebruikten de test naast de TAK Zinsvorming, eveneens een zinsherhaaltaak. De TAK heeft een score voor functiewoorden en zinspatronen. De LITMUS-taak werd op diverse wijzen gescoord. Van de eerder genoemde scores werden de S- en G-scores gebruikt en ook de scores voor inhouds- en functiewoorden. Tabel 1 geeft de resultaten van deze en andere studies weer.

Tabel 1: Resultaten van afname van de Nederlandse SRep in enkele studies.

Auteur, jaartal	Participanten	Aantal items	Wijze van scoring	Percentage correct
Van Assenbergh et al, 2015	74 eentalige kinderen van 5 en 6 jaar zonder TOS	60	S-score	52.7
			G-score	68.6
			Inhoudswoorden	86.2
			Functiewoorden	85.7
Van den Brand, 2015	29 eentalige kinderen zonder TOS, 7;0 - 8;0	60	S-score	61.5
Van der Kist, 2015	Meertalige kinderen met TOS 16 kinderen tussen 4;7 en 8;3	60	S-score	14.0
			Functiewoorden	57.0
Van den Boer, 2016	Eentalige kinderen zonder TOS 37 4-jarigen 51 5-jarigen	30	S-score	39.7

Van den Brand (2015) rapporteerde S-scores voor een groep van 29 eentalige kinderen

tussen 7 en 8 jaar zonder taalontwikkelingsstoornis (TOS). Het groepsgemiddelde was 61.5% (iets hoger dan van de jongere kinderen die participeerden in de studie van Van Assenbergh et al.). Van der Kist (2015) gebruikte dezelfde test bij 16 meertalige kinderen met een TOS, tussen 4;7 en 8;3 jaar oud. Het percentage correct herhaalde functiewoorden was hier gemiddeld 57%. De studie van Van der Kist toetste ook de validiteit van de SRep, door een vergelijking met, opnieuw, de TAK Zinsvorming. De correlatie was hoog ($r=.898, \leq .01$; de resultaten van Van Assenbergh et al., die overigens niet primair deze correlatie bestudeerden, komen overeen met deze bevinding). Zowel Van Assenbergh et al. als van der Kist berekenden een score voor functiewoorden. Die lag bij de onderzoeksgroep van Van der Kist (die overigens een grotere leeftijdsrange had) beduidend lager. De kinderen in haar studie hadden een taalontwikkelingsstoornis *en* waren meertalig. De invloed van beide factoren kan daarom niet afzonderlijk worden bepaald.

Totstandkoming van LITMUS SRep NL: SRep met 30 items

De omvang van de herhaaltaak werd in een volgende fase teruggebracht van 60 tot 30 items, als onderdeel van het CoDEmBi-onderzoeksproject (<https://www.uu.nl/organisatie/faculteit-sociale-wetenschappen/codembi>). Om het aantal items verantwoord te verlagen werd een betrouwbaarheidsanalyse uitgevoerd. Het materiaal daarvoor bestond uit afnames van de test verzameld in het kader van enkele bachelorscripties (in het bijzonder die van van Assenbergh et al., 2015). Items die de waarde van Cronbach's alpha, een maat voor de onderlinge samenhang van de items, het minst beïnvloedden en daarmee de interne consistentie van de test verminderden werden in de herziene versie uitgesloten. In deze procedure werden ook data van kinderen met een TOS gebruikt (deze data werden verzameld door Van den Brand, 2015). Items die voor deze kinderen relatief eenvoudig waren, werden eveneens uitgesloten, omdat ze de sensitiviteit van de test zouden verminderen. Bij deze bewerking van de test werd er gelijktijdig naar gestreefd de structuren van de oorspronkelijke (Engelse) test zo veel mogelijk te behouden (zoals beschreven in Marinis & Armon-Lotem, 2015). Die waren immers gekozen om de crosslinguïstische validiteit van de test veilig te stellen.

Tabel 2 geeft voorbeelden van items. De tabel laat zien dat de resulterende test voor iedere doelstructuur minimaal drie items telt (zes items voor drie verschillende typen vraagwoordzin). De twee processen die het taalkundige fundament vormen voor de complexiteitsniveaus van de SRep laten zich illustreren aan de hand van de vraagzinnen en passieve zinnen (*verplaatsing*) en de voorwaardelijke en betrekkelijke bijzinnen (*inbedding*). In de tabel zijn de drie complexiteitsniveaus weergegeven die in de SRep worden onderscheiden.

Tabel 2: Doelstructuren in de Srep NL met 30 items (tussen haakjes het aantal items in de test per doelstructuur).

Doelstructuur	Voorbeeld
Niveau 1	
Enkelvoudige zin met modaal hulpwerkwoord (3)	<i>Zij kan het meisje zien door het raam</i>
Korte passieve zin (3)	<i>Bij de grote rode lampen werd ze gestopt</i>
Vraagzinnen (6):	
Wie-vragen (2)	<i>Wie heeft de aap natgespetterd bij de zee?</i>
Wat-vragen (2)	<i>Wat heeft de prinses vorige maand gekocht?</i>
Welk(e)-vragen (2)	<i>Welke foto heeft hij gisteren gemaakt?</i>
Niveau 2	
Lange passieve zin (3)	<i>Ze werd 's ochtends door de dokter bezocht</i>
Hoofdzin met hulpwerkwoord van tijd en modaal hulpwerkwoord (3)	<i>De agent had ons een bekeuring kunnen geven</i>
Complexe ontkenningen (3) (Negatie met complex gezegde en/of met tangconstructie)	<i>Zij heeft de schapen op de boerderij niet kunnen voeren</i>
Niveau 3	
Voorwaardelijke bijzinnen (3):	
2 waar	<i>De mensen krijgen een cadeau als ze het huis schoonmaken</i>
1 onwaar	<i>Hij had zijn vriendin niet meegenomen als ze boos was</i>
Betrekkelijke bijzin bij onderwerp (3)	<i>De jongen die de postbode hielp was verdwaald</i>
Betrekkelijke bijzin bij lijdend voorwerp (3)	<i>De kinderen proefden de snoepjes die ze kregen</i>

Resultaten van de afname van LITMUS SRep NL met 30 items

Hieronder worden de resultaten besproken van de afname van de korte versie (met 30 items) van de LITMUS SRep NL bij eentalige en meertalige kinderen. Speciale aandacht verdienen hier enkele vragen over de SRep die primair de validiteit van de taak betreffen. De eerste vraag is of de score op de SRep toeneemt met de *leeftijd*. Wanneer de SRep een valide maat voor grammaticale ontwikkeling is, valt te verwachten dat er bij eentalige kinderen een positieve relatie is tussen de score en chronologische leeftijd. Vinden we die terug in de afnames? Bij meertalige kinderen wordt - in hun tweede taal - een verband verwacht met de *input*, de omvang of duur van blootstelling aan die taal (en ook met de kwaliteit van het aanbod, die moeilijker meetbaar is). De tweede vraag is of de score op de SRep afneemt met *complexiteit*. Zoals eerder omschreven kent de SRep drie complexiteitsniveaus. Die zijn primair geoperationaliseerd met behulp van de taalkundige variabelen verplaatsing en inbedding. Op basis van de toename in complexiteit tussen niveaus 1 en 3 is de verwachting dat het aantal fouten toeneemt met de complexiteit van de items (niveau 1 < niveau 2 < niveau 3). Omdat de complexiteit is vormgegeven door de selectie van syntactische structuren is de doelstructuurscore de zuiverste maat voor het bepalen van de invloed van de complexiteit. Wanneer deze score sterk correleert met de S-score, kunnen we echter aannemen dat de in-

vloed van complexiteit ook zichtbaar wordt wanneer een grovere scoringsprocedure wordt gehanteerd.

Onderzoek naar eentalige kinderen met de SRep NL

De testafnames in de eerder aangehaalde studie waarin de SRep met 30 items werd gebruikt, lieten al zien dat de scores toenemen met de chronologische leeftijd (in de studie van Van Assenbergh et al. (2015), waarin de versie met 60 zinnen werd gebruikt, werd deze relatie expliciet bestudeerd en bleek de toename significant). Van den Boer (2016) geeft resultaten van de afname van de test met 30 items bij 88 vier- en vijfjarige kinderen. De gemiddelde S-score was 11.9, omgerekend naar een percentage correct (en aldus vergelijkbaar met de scores op de versie met 60 items) 39.7. Bij Van Assenbergh et al. (2015), (vijf- en zesjarige kinderen) was, zoals gezegd, het gemiddelde 52.7, bij van den Brand (2015) (zeven- en achtjarige kinderen) 61.5. Deze studies tonen dus een gestage toename van de S-score naarmate de kinderen ouder zijn. Deze conclusie wordt bevestigd in een recente studie van van Dijk, Dijkstra en Unsworth (*under review*), die we verderop zullen bespreken, waarin een positieve correlatie werd gevonden tussen leeftijd en S-score ($r = .632$; $p < .001$).

Onderzoek naar meertalige kinderen met de SRep

In het vervolg zal worden gerefereerd aan kinderen uit twee taalvergelijkende onderzoeksprojecten. Het project van van Dijk, Dijkstra en Unsworth (*Cross-linguistic influence during real-time language processing in bilingual children*¹; van Dijk, 2021) omvat Engels-Nederlandse en Duits-Nederlandse tweetalige kinderen (naast eentalig Nederlandse kinderen). Hun onderzoeksgroep bestaat uit simultaan en vroeg-succesief tweetaligen. Het andere project (van Blom en Attar) heet '*Taalvaardigheid van kinderen die uit Syrië gevlucht zijn*'² en betreft (Syrisch)-Arabisch als eerste en Nederlands als tweede taal. In alle gevallen betreft het hier kinderen zonder een gediagnosticeerde taalontwikkelingsstoornis.

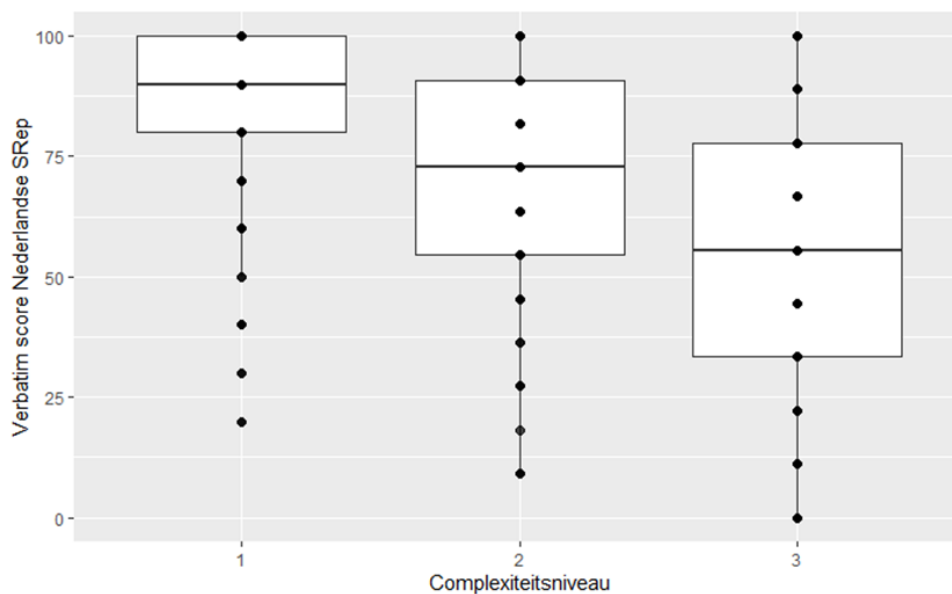
De eerste validiteitsvraag die hierboven werd gesteld betreft de *leeftijd*. Bij eentaligen wordt verwacht dat de scores op de SRep toenemen met de leeftijd. Dat geldt ook voor de thuistaal van tweetalige kinderen, mits de blootstelling aan de thuistaal in voldoende mate continueert. Voor de tweetalige groepen die we hier bespreken is de situatie echter verschillend. De Syrische kinderen zijn succesief tweetalig. Voor deze groep is er inderdaad een correlatie tussen de leeftijd en de SRep-score in de moedertaal ($r = .536$; $p < .001$). De andere twee groepen bestaan uit zowel simultaan- als succesief-tweetalige kinderen. Bovendien zijn er in die groepen aanzienlijke verschillen in de leeftijd waarop de kinderen voor het eerst blootgesteld waren aan het Nederlands. We hebben daarom een correlatie berekend tussen leeftijd en SRep-score voor het Duits, Engels en Nederlands van alleen de

¹NWO-project 23000408 NWO 322-70-010.

²NWO Aspasia beurs.

simultaan-tweetalige kinderen. Daartoe werden de kinderen geselecteerd die vanaf hun geboorte werden blootgesteld aan beide talen. Daarbij vinden we de volgende waarden: bij de Duits-Nederlandse kinderen ($n = 32$), voor de Duitse SRep: $r = .045$; $p = .805$ en voor de Nederlandse SRep $r = .562$; $p = .001$; bij de Engels-Nederlandse kinderen ($n = 26$), voor de Engelse SRep $r = .506$; $p = .008$ en voor de Nederlandse SRep $r = .387$ en $p = .051$.

Bij de successief-tweetalige kinderen is voor de score in het Nederlands de periode van blootstelling aan die taal belangrijker dan chronologische leeftijd. De Syrisch-Nederlandse groep ($n = 52$) is exclusief successief-tweetalig. Er blijkt inderdaad een positief verband te zijn tussen de periode van blootstelling aan het Nederlands en de score op de Nederlandse SRep: $r = .413$; $p = .002$. Er blijkt ook een positief verband te zijn tussen *leeftijd* en dezelfde score ($r = .420$; $p = .002$). Er is geen correlatie tussen leeftijd en lengte van blootstelling ($r = .113$; $p = .425$). Niettemin kan worden aangenomen dat, *naast* de lengte van blootstelling, ook de leeftijd de score beïnvloedt, bij voorbeeld doordat de capaciteit van het verbale-kortetermijngeheugen toeneemt.

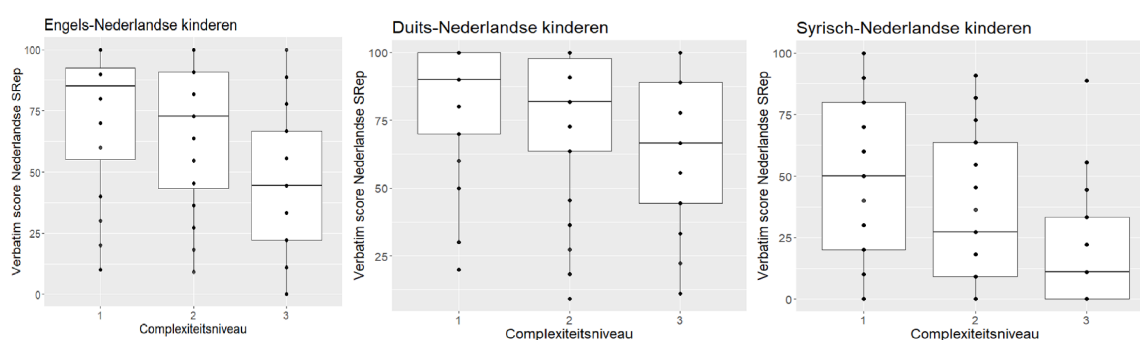


Figuur 1: Vergelijking tussen de scores op zinnen met drie complexiteitsniveaus. S-scores (verbatim). Eentalige Nederlandse kinderen.

De tweede validiteitsvraag over de SRep betreft de aanname dat de drie complexiteitsniveaus zichtbaar zullen zijn in de test scores. Ter oriëntatie toont Figuur 1 een *box plot* voor de scores op de zinnen per complexiteitsniveau, bij de *eentalige* Nederlandse kinderen uit het onderzoek van van Dijk et al. Het trapsgewijze patroon in de figuur laat een duidelijke invloed van de complexiteit zien: hoe complexer de structuren hoe meer fouten de kinderen maken en hoe lager hun score. Het bleek dat de kinderen een significant hogere score

hadden op het laagste complexiteitsniveau ten opzichte van het tweede en derde complexiteitsniveau en dat ze een hogere score hadden op het tweede in vergelijking met het derde complexiteitsniveau³.

In Figuur 2 wordt een vergelijking gemaakt tussen de complexiteitsniveaus bij de drie groepen *tweetaligen*, met respectievelijk Engels, Duits en Syrisch-Arabisch als eerste taal (in deze vergelijking zijn zowel de simultaan- als successief-tweetalige kinderen betrokken). De figuren betreffen de scores op de SRep NL. Belangrijk hier is dat de Syrisch-Arabisch groep minder lang was blootgesteld aan het Nederlands. De figuur beeldt de rol van complexiteit af. In alle groepen zien we de trapsgewijze afname van scores bij toenemende complexiteit die we ook bij eentalige kinderen zagen (Figuur 1). Statistische modellen laten zien dat deze observatie ook significant is, hoewel niet universeel: de scores op de verschillende complexiteitsniveaus verschillen niet noodzakelijkerwijze in alle groepen significant van elkaar.



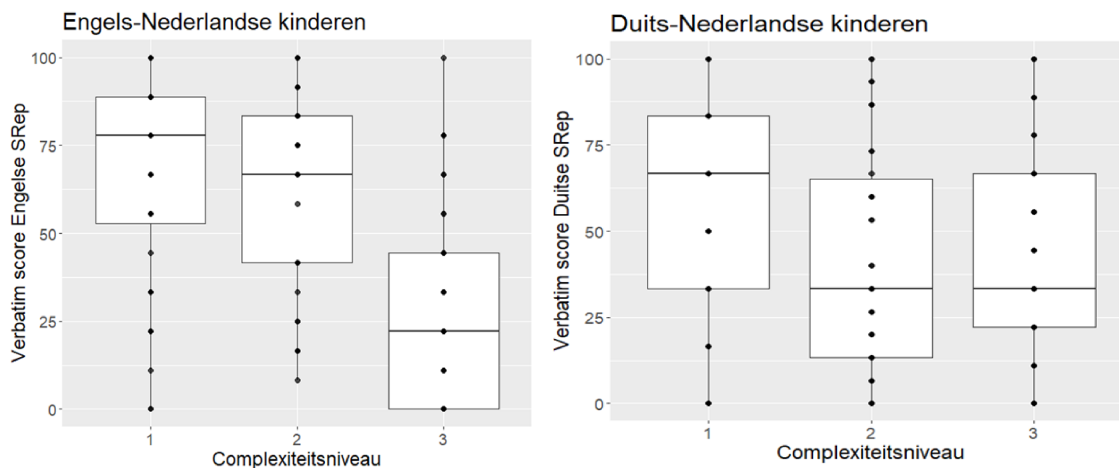
Figuur 2: Vergelijking tussen de S-scores op zinnen van de Nederlandse SRep op drie complexiteitsniveaus bij drie groepen tweetalige kinderen.

Figuur 3 laat voor de Engels-Nederlandse kinderen zien dat de trapsgewijze afname in *beide* talen zichtbaar is (vergelijk daartoe Figuur 2 en 3). De validiteit van de complexiteitsniveaus is dus voor die groep niet taalafhankelijk. In de volgende paragrafen wordt het - iets afwijkende - patroon bij de Duits-Nederlandse kinderen besproken (in de figuur ontbreekt de Syrisch-Nederlandse groep, omdat een representatie van de complexiteitsniveaus in het Arabisch niet mogelijk was).

Zoals beschreven in het begin van dit artikel wordt SRep idealiter afgenomen in beide (of alle) talen van het kind. Voor de groepen die hier besproken worden (met kinderen die

³Ter verantwoording: Linear mixed effects models laten zien dat het patroon significant is. Volgens een log-likelihood test is een model met complexiteitsniveau als fixed effect beter in het voorspellen van de resultaten van kinderen dan een model zonder complexiteitsniveau ($\chi^2 = 21.2$; $\Delta df = 2$; $p < .001$). Post-hoc tests (Tukey-methode) tonen dat kinderen een significant hogere score hadden op het laagste complexiteitsniveau ten opzichte van het tweede ($\beta = 1.15$; $SE = 0.43$; $z = 2.66$; $p = .021$) en derde complexiteitsniveau ($\beta = 2.48$; $SE = 0.45$; $z = 5.46$; $p < .001$) en dat ze een hogere score hadden op het tweede ten opzichte van het derde complexiteitsniveau ($\beta = 1.33$; $SE = 0.44$; $z = 3.05$; $p = .006$).

geen aanwijsbare taalstoornis hebben) is het relevant om de correlatie tussen de scores in de eerste en tweede taal vast te stellen. Ook dat is een maat voor de validiteit van de taak. In het geval van een kind met TOS zou immers een lage score op beide versies van de SRep verwacht worden. In de Syrische groep is er een correlatie ($r = .617$; $p < .001$) evenals in de Engelse groep ($r = .463$; $p = .003$). In de Duitse groep vinden we echter geen correlatie ($r = .271$; $p = .083$).



Figuur 3: Vergelijking tussen de scores op zinnen met drie complexiteitsniveaus. S-scores (verbatim). Duits-Nederlandse en Engels-Nederlandse kinderen. Scores op de Duitse en Engelse SRep.

Structuurscore en structuurbehoud

Figuur 3 illustreert een verschijnsel dat om een toelichting vraagt. De afname van scores met niveau van complexiteit kent hier een opvallende uitzondering: bij de Duits-Nederlandse groep zijn bij afname in het Duits de medianen voor niveau 2 en 3 gelijk. Sterker nog, er zijn geen significante verschillen tussen de scores op de verschillende niveaus⁴. De waarde van de S-score bleek in dit geval sterk beïnvloed door een eigenschap van de doeltaal. Naamvalsmorfologie is rijk in het Duits en juist daarin maakten de kinderen veel fouten - voor het verwerven van naamval is veel taalaanbod nodig. Deze taalspecifieke eigenschap verklaart wellicht ook dat in de eerder gemaakte berekeningen de scores op de Duitse SRep niet correleerden met leeftijd en dat de scores op de Duitse SRep niet correleerden met die op de Nederlandse SRep bij dezelfde kinderen. Omdat de vergelijking tussen talen hierdoor wordt vertroebeld, kozen van Dijk et al. (*under review*) voor een soepeler score (dan die wordt weergegeven in Figuur 3) waarbij de naamvalfouten niet werden ingecalculerd. Deze keus was ingegeven door de intentie om een vergelijking tussen talen mogelijk te maken, maar raakt tegelijk aan een principiële kwestie. Wanneer de SRep beoogt om specifieke structuren te meten, is de S-score dan wel het beste instrument? Bij het herhalen van de zin kan

⁴ $\chi^2 = 5.6$; $\Delta df = 2$; $p = .061$

het kind immers ook fouten maken die niet de syntactische (doel)structuur zelf betreffen. Daarvoor dient primair de structuurscore. Om die goed te hanteren zijn echter duidelijke regels nodig.

Een studie van Armon-Lotem en Meir (2016) hanteerde de doelstructuurscore van de SRep voor (onder andere) een vergelijking tussen Russisch en Hebreeuws bij dezelfde tweetalige kinderen. De aanname is dat deze structuurscore minder onderhevig is aan taalspecifieke factoren dan de andere wijzen van scores en zich daarom beter leent voor crosslinguïstische vergelijking. De structuren komen in iedere versie van de SRep - waar mogelijk - overeen met de structuren zoals omschreven door Marinis en Armon-Lotem (2015) voor de Engelstalige versie van de test (Marinis en Armon-Lotem laten daarnaast zien dat er aanpassingen van de structuren nodig waren, die zij in hun hoofdstuk illustreren voor het Russisch, Hebreeuws en Frans). Het toepassen van de structuurscore roept ook vragen op. Wanneer een woordelijke herhaling van de stimulus niet vereist is, welk deel van de uiting moet dan minimaal herhaald worden om vast te stellen dat het kind de doelstructuur heeft geproduceerd? Welke afwijking van de doeluiting is toegestaan? Hieronder worden, ter illustratie, criteria voorgesteld voor het scoren van behoud van twee structuren: de passiefzin (zonder doorbepaling) en de betrekkelijke bijzin die begint met een lijdend voorwerp (de analyse-instructies zijn zoals voorgesteld in Blom et al., 2021). De vraag hierbij is: wanneer beschouwen we de doelstructuur als (niet) behouden?

(1) Korte passieve zin

Voorbeelditems:

Bij de grote rode lampen werd ze gestopt

Ze werd hard tegen de grond geduwd

De jongen in de snoepwinkel werd geholpen

(a) De structuur is *niet* behouden bij:

- Verandering naar een actieve zin door weglating van het hulpwerkwoord 'werd'
- Weglating van het prefix (ge-), suffix (-en, -d, -t) of van prefix en suffix
- Verandering naar een actieve zin door gebruik van een verkeerd hulpwerkwoord (bijvoorbeeld 'is' in plaats van 'werd')
- Toevoeging van een bijwoordelijke bepaling met 'door...' om de handelende persoon uit te drukken
- Weglating van het onderwerp of verplaatsing van het onderwerp naar de positie achter het werkwoord

(b) De structuur is *wel* behouden bij:

- Weglaten van een bijwoordelijke bepaling (bij voorbeeld van 'in de snoepwinkel')
- Een fout in de bijwoordelijke bepaling
- Verandering van werkwoordstijd ('wordt' in plaats van 'werd')
- Verandering van 'aantal' in het onderwerp ('de jongens')
- Fout in congruentie (enkelvoud/meervoud) tussen onderwerp en persoonsvorm
- Vervanging van individuele woorden door soortgelijke woorden, of woorden die niets met het doelwoord te maken hebben

(2) Betrekkelijke bijzin bij lijdend voorwerp

Voorbeelditems:

De kinderen proefden de snoepjes die ze kregen

De aap aaide de koe die de worm bang had gemaakt

Hij moet de knuffel wassen waarmee het kind slaapt

(a) De structuur is *niet* behouden als:

- Een van de twee zinnen (hoofdzin of bijzin) wordt weggelaten
- De bijzin bij het lijdend voorwerp wordt veranderd in een betrekkelijke bijzin bij het onderwerp
- Een (persoonlijk) voornaamwoord wordt toegevoegd om terug te verwijzen naar het lijdend voorwerp
- De ondergeschikte bijzin wordt vervangen door een nevengeschikte zin
- Het onderwerp in de bijzin wordt weggelaten
- De volgorde van het onderwerp en het werkwoord in de bijzin wordt veranderd
- 'Die' of 'waarmee' wordt weggelaten (bij voorbeeld 'Hij moet de knuffel wassen het kind slaapt')

(b) De structuur is *wel* behouden bij:

- Vervanging van 'die' met 'welke'
- Vervanging van woorden door soortgelijke woorden, of woorden die niets met het doelwoord te maken hebben

Deze lijst van criteria is niet uitputtend, maar kan dienen als een richtlijn voor het vaststellen van structuurbehoud. De rationale is telkens: de structuurscore is 1 als de doelstructuur wordt gerealiseerd (behouden). Fouten die niet de doelstructuur betreffen tellen niet mee (beïnvloeden de score niet), anders dan bij de S-score.

De instructies zoals ze hier voor het Nederlands zijn aangehaald, moeten voor iedere taal afzonderlijk worden gespecificeerd. Er zijn immers taalspecifieke verschillen in de vormgeving van de doelstructuren. Een voorbeeld kan dit verduidelijken. In het Engels contrasteert *the cow that scared the worm* met *the cow that the worm scared*. In de tweede zinsnede representeert *that* het lijdend voorwerp, in de eerste het onderwerp van de zin - de woordvolgorde is hier de bepalende factor voor het identificeren van het lijdend voorwerp. In het Nederlands is *de koe die de worm bang maakte* tweeduidig: *die* kan zowel onderwerp als lijdend voorwerp zijn. Een Duits voorbeeld van een bijzin die begint met een lijdend voorwerp is *Die Frau sieht den Pinguin, den sie streicheln will*. Hier wordt duidelijk wat het lijdend voorwerp is doordat het voornaamwoord is gemarkeerd voor naamval (*den*). Instructies voor het Engels en het Duits zullen dus respectievelijk moeten refereren aan woordvolgorde en naamval.

Vooralsnog wordt bij afname van de SRep het vaakst gebruik gemaakt van de S-score. Er zijn echter aanwijzingen dat S-score en structuurscore *binnen een taal* correleren. Dat zou suggereren dat de moeilijkheidsgraad van de SRep-structuren ook in de - gemakkelijker te berekenen S-score wordt gereflecteerd. Blom et al. (2021) stelden dat vast voor de Syrisch-Nederlandse groep. De waarden voor de correlatie tussen beide scores waren hoog: voor het Nederlands $r = .87, p < .001$, voor het Arabisch $r = .90, p < .001$.

De LITMUS SRep en meertaligheid

Zoals eerder benadrukt, is de kern van de LITMUS-testbatterij dat de instrumenten parallel kunnen worden afgenomen, in de eerste en tweede taal. De SRep is daarbij een diagnostisch hulpmiddel. Afname van een test in de eerste taal wordt gewoonlijk gezien als de lakmoesproef voor het vaststellen van de aanwezigheid van een TOS. Een lage |score in beide talen bevestigt de aanname van een taalontwikkelingsstoornis. Een dergelijke observatie laat zien dat de SRep ook zonder normen waardevol kan zijn.

Er is echter ook onderzoek waarin de sensitiviteit van de test in de tweede taal wordt onderzocht. In zulk onderzoek zijn eentalige en tweetalige kinderen met en zonder TOS betrokken. Een voorbeeld is de studie van Tuller et al., (2018), waarin Duitse en Franse versies van de SRep worden gebruikt bij kinderen met Arabisch, Portugees en Turks als moedertaal. De identificatie van TOS bij de tweetalige kinderen op basis van de tests was redelijk tot goed (bij de eentalige kinderen goed tot uitstekend). Dat betekent dat ze zich in de onderzoeksgroep onderscheidden van de kinderen met TOS - in hun *tweede* taal. Ibrahim en Fekete (2019) kwamen tot een vergelijkbare conclusie over de diagnostische waarde van de test

wanneer ze de Duitse SRep combineerden met de non-woordherhaalttest uit de LITMUS-testbatterij (Chiat, 2015). Een vergelijkbare exercitie in de diagnostiek (waarbij een controlegroep immers niet beschikbaar is) vereist echter de beschikbaarheid van normatieve data.

Onderzoek dat dichterbij de LITMUS-rationale komt (afname van de SRep in beide talen) is uitgevoerd door Armon-Lotem en Meir (2016). Daarin participeerden eentalige en tweetalige kinderen met en zonder TOS. De onderzochte talen waren Hebreeuws en Russisch. De correctheid van de diagnostiek bleek het hoogst wanneer de taken in de eerste *en* de tweede taal werden geïncorporeerd, maar zelfs testafname in de tweede taal alleen toonde een hoge sensitiviteit en specificiteit.

Eerder in dit artikel werden de verschillende scoringsmethoden onderscheiden. Veel studies (zoals die van Tuller e.a. 2018) maakten gebruik van de S-score (letterlijke herhaling). Onder anderen Armon-Lotem en Meir (2016) gebruikten daarentegen de doelstructuurscore. De doelstructuurscore is nog onderwerp van verder onderzoek. Daarbij is het noodzakelijk voor iedere taal instructies te formuleren voor het bepalen van de structuurscore. Dat bevordert ook de vergelijkbaarheid van de score tussen talen. Eerder werd gesuggereerd dat de structuurscore correleert met de S-score. Dat zou in de praktijk kunnen leiden tot het exclusief gebruik van de S-score (die eenvoudiger te bepalen is). De structuurscore is echter ook intrinsiek waardevol omdat daarmee de validiteit van elk van de gekozen structuren kan worden getoetst (Zie Fleckstein et al., 2018, die de structuren - ook - afzonderlijk bestudeerden).

Conclusie

In dit artikel werd de Nederlandse versie van de LITMUS Sentence Repetition task (SRep) beschreven. Uit de afnames van de taak met een- en tweetalige kinderen blijkt dat de testcores correleren met chronologische leeftijd en/of met de lengte van blootstelling aan de tweede taal. De opbouw van de test - met items die toenemen in moeilijkheidsgraad - laat zien dat toenemende complexiteit van de items resulteert in een toename van het aantal fouten. Dat stemt overeen met de intentie van de test.

In de huidige opzet is de test een belangrijk hulpmiddel in de tweetalige diagnostiek. Bij afname in de eerste en tweede taal kunnen hypothesen worden geformuleerd over de aan- of afwezigheid van een taalontwikkelingsstoornis: een discrepantie tussen de afnames waarbij de eerste taal intact lijkt, weersprekt het vermoeden van een TOS; een lage score bij beide afnames wettigt een dergelijk vermoeden en moet aanleiding geven tot nader onderzoek. In dit artikel hebben we daarnaast gedemonstreerd dat het ontwerp van de taak informatie kan geven over de mate waarin afzonderlijke taalstructuren een probleem vormen voor het geteste kind. Dat kan helpen bij het formuleren van specifieke doelen in taaltherapie.

Literatuur

- Armon-Lotem, S. & N. Meir (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language and Communication Disorders*, 51, 715-731.
- Assenbergh, D. van, L. Hassing, T. Kool & A. Trouwborst (2015). *De Invloed van Individuele Verschilfactoren op Zinsverwerking bij Eentalige Kinderen van 5 en 6 Jaar*. Bachelor thesis. Universiteit Utrecht.
- Blom, E., A. Soto-Corominas, Z. Attar, L. Daskalaki, & J. Paradis (2021). Interdependence between L1 and L2: The case of Syrian children with refugee backgrounds in Canada and the Netherlands. *Applied Psycholinguistics*, 136.
- Boer, A.J.C. van den (2016). *De Invloed van Ouderlijke Taalstimulering op de Receptieve Woordenschat, Narratieve Vaardigheden en Zinsverwerking bij 4- en 5-jarige Kinderen*. Master thesis. Universiteit Utrecht.
- Bome C. & I. Kongtorp Vangen (2015). *The Development and Pilot of a Norwegian Adaption to the Language Impairment Testing in a Multilingual Setting Sentence Repetition Task*. Master thesis. Universitetet i Bergen.
- Brand, C. van den (2015). *Verdere ontwikkeling van de SASIT-NL: De moeilijkheidsgraad*. Bachelor thesis. Hogeschool Windesheim, Zwolle.
- Chiat, S. (2015). Non-word repetition. In: S. Armon-Lotem, J. de Jong & N. Meir (eds.). *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters.
- Conti-Ramsden, G., N. Botting & B. Faragher (2001). Psycholinguistic Markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry*, 42, 741-748.
- Dijk, C. van (2021). *Cross-linguistic influence during real-time sentence processing in bilingual children and adults*. Doctoral dissertation. Radboud Universiteit Nijmegen. <https://dx.medra.org/10.48273/LOT0598>.
- Dijk, C. van, T. Dijkstra & S. Unsworth (under review). Cross-linguistic influence during online sentence processing in bilingual children. *Bilingualism: language and cognition*.
- Fleckstein, A., P. Prévost, L. Tuller, E. Sizaret, & R. Zebib (2018). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, 25, 85101.
- Ibrahim, L.A. & I. Fekete (2019). What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks. *Frontiers in psychology*, 9, 27-57.
- Kist, E. van der (2015). *SASIT: Validiteitsstudie naar het nazeggen van functiewoorden in een zinsherhaaltaak voor het identificeren van een taalprobleem bij sequentieel meertalige kinderen*. Master thesis. Universiteit Utrecht.
- Klem, M., M. Melby-Lervag, B. Hagtvet, S.-A. Halaas Lyster, J.-E. Gustafsson & C. Hulme (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, 18, 146-154.
- Kohnert, K. (2013). *Language disorders in bilingual children and adults*. Second edition.

- San Diego: Plural Publishing.
- Kort, W., M. Schittekatte & E. Compaan (2008). *CELF-4-NL (Clinical Evaluation of Language Fundamentals-4-NL)*. Amsterdam: Pearson.
- Leonard, L.B. (2014). *Children with Specific Language impairment. Second edition*. Cambridge, Mass.: MIT Press.
- Marinis, T., S. Chiat, S. Armon-Lotem, D. Gibbons, & E. Gipps (2010). *School-Age Sentence Imitation Test (SASIT)*. Reading: University of Reading.
- Marinis, T., & S. Armon-Lotem (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol, United Kingdom: Multilingual Matters.
- Paradis, M. (2004). *A neurolinguistics theory of bilingualism*. Amsterdam: John Benjamins.
- Redmond, S. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics*, 19, 109127.
- Scheur, L. van de (2012). *Development of the SASIT-NL*. Master thesis. Universiteit van Amsterdam.
- Seeff-Gabriel, B., S. Chiat & P. Roy (2008). *Early Repetition Battery*. London: Pearson Assessment.
- Slobin, D. I., & C.A. Welsh (1973). Elicited imitation as a research tool in developmental psycholinguistics. In C. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart & Winston.
- Tuller, L., C. Hamann, S. Chilla, S. Ferré, E. Morin, P. Prevost & C. Dos Santos (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language and Communication Disorders*, 53, 888-904.
- Verhoeven L. & A. Vermeer (2006). *Verantwoording Taaltoets Alle Kinderen (TAK)*. Arnhem: CITO.
- Wiig, E., Secord, W., & E. Semel (1992). *Clinical Evaluation of Language Fundamentals-Preschool*. San Antonio, TX: The Psychological Corporation.
- Wiig, E. H., W.A. Secord, E. Semel & J. de Jong (2012). *CELF preschool-2-NL: Clinical Evaluation of Language Fundamentals: Preschool - Nederlandstalige versie*. Amsterdam: Pearson.

Appendix. SRep NL, Korte versie, 30 items

Oefenitems:

- De fiets werd gemaakt door de vader.
- De vrouw heeft ons snel kunnen vinden.

Testitems:

1. Wat heeft de prinses vorige maand gekocht?
2. Bij de grote rode lampen werd ze gestopt.
3. Zij kan het meisje zien door het raam.
4. Ze werd hard tegen de grond geduwd.
5. Wie hebben zij bij de trap gezien?
6. Het meisje moet het vliegtuig op het bord tekenen.
7. Wie heeft de aap natgespetterd bij de zee?
8. De jongen in de snoepwinkel werd geholpen.
9. De jongen moet de vloer in de keuken vegen.
10. Wat hebben zij gisteren in de sneeuw gevonden?
11. De man had niet met de auto naar huis mogen rijden.
12. Ze werd 's ochtends door de dokter bezocht.
13. Welke foto heeft hij gisteren gemaakt?
14. De agent had ons een bekeuring kunnen geven.
15. De jongen werd bang gemaakt door de clown.
16. De muis ziet de kaas in de schaal niet.
17. Hij had naar de boot moeten zwemmen.
18. Welke fles liet de jongen op de grond vallen?
19. Zij heeft de schapen op de boerderij niet kunnen voeren.
20. De kat had de bal van de trap af kunnen gooien.

21. De lunch werd door de timmerman gegeten.
22. De kinderen proefden de snoepjes die ze kregen.
23. De mensen krijgen een cadeau als ze het huis schoonmaken.
24. De bij die de man inslikte had hem pijn gedaan.
25. De aap aaide de koe die de worm bang had gemaakt.
26. Als de kinderen zich gedragen gaan we de tuin in.
27. De jongen die de postbode hielp was verdwaald.
28. Hij moet de knuffel wassen waarmee het kind slaapt.
29. Hij had zijn vriendin niet meegenomen als ze boos was.
30. De pony die de boer bekeek duwde hem opzij.