Welke efficiënte en betrouwbare werkwijzen zijn er om het startniveau en de ontwikkeling van de spreekvaardigheid van volwassen NT2-leerders tijdens onderwijs Nederlands te meten?

Beantwoord door Maurice de Greef (kennismakelaar Kennisrotonde).

24 december 2019 – update 10 januari 2024

KR. 745

Kort antwoord

Om de spreekvaardigheid van volwassen NT2-leerders te meten zijn zowel een brede toetsinhoud die aansluit bij concrete taalsituaties als getrainde beoordelaars nodig. Mogelijke toetsvormen zijn de kandidaat-examinator toets, de dyadische toets (toets in tweetallen) en de groepstoets (toets in groepen van drie of vier personen), waarbij de laatste twee iets meer geschikt lijken te zijn voor het toetsen van mondelinge bekwaamheid vanwege het feit dat er meer mogelijkheid tot interactie is. Daarnaast is het mogelijk om de zinsreproductie als controlemiddel of incidenteel alternatief te gebruiken.

Antwoord

Het blijkt niet altijd eenvoudig om tijdens NT2-lessen het beheersingsniveau met betrekking tot het spreken van NT2-leerders in kaart te brengen. Volgens Van Til et al. (2011) zijn er op dit gebied twee vaardigheden, die volgens het ERK (Europees Referentie Kader) Talen opgedeeld kunnen worden opgedeeld in subvaardigheden namelijk spreken (monologen en een publiek toespreken) en gesprekken voeren (informele gesprekken, bijeenkomsten en vergaderingen, zaken regelen en informatie uitwisselen).

Als men deze (sub)vaardigheden wil toetsen moeten de opgaven onder andere uit een niveau-aanduiding, situatiebeschrijving en opdracht bestaan (Van Til et al., 2011).[1] Vraag is welke vorm van toetsing nu gebruikt kan worden om op een efficiënte en betrouwbare werkwijze het startniveau en de ontwikkeling van spreekvaardigheid van NT2-leerders te bepalen.

[1] In Nederland zijn een aantal toetsen die men hiervoor kan gebruiken, zoals de spreektoetsen van Ad Appel Taaltrainingen, de Toets Gesproken Nederlands, de Spreektoets van NT2TaalMenu, de digitale curriculumonafhankelijke NT2-niveautoetsen (TANG-toetsen), de toetsen Nederlands spreken en toetsen gesprekken voeren van TOA of de Staatsexamens NT2.

Brede en concrete inhoud die vloeiend spreken triggeren

Om het spreekniveau en het vloeiend spreken te kunnen beoordelen en vast te kunnen stellen, blijkt een brede en concrete inhoud van belang te zijn. Deygers (2012) geeft aan, dat men eerst de vraag zou moeten stellen waarom men toetst (doel) in plaats van wat men toetst (breedte en inhoud). Er zijn namelijk toetsen met een hoge en een lage impact. Het meten van de voortgang heeft bijvoorbeeld minder impact dan een taaltoets voor migratiedoeleinden, die aan striktere kwaliteitscriteria zou moeten voldoen (Deygers, 2012). Een diagnostische toets zou in principe een meer bredere inhoud moeten hebben (Deygers, 2012). Volgens Cucchiarini en Jaspaert (1996) is het gemakkelijker om over het bereikte niveau van functionele taalvaardigheid conclusies te trekken, als deze bredere inhoud van de toetsen meer aansluit bij concrete taalsituaties. Onderzoek van Baker-Smemoe (2014) naar de toetsing van taalvaardigheid onder tweede taalleerders laat zien, dat het vloeiend kunnen spreken (in deze taalsituaties) dan deels informatie geeft of men de taal wel of niet beheerst. Met name de snelheid en lengte van het spreken naast het aantal pauzes, lijken er toe te doen (vooral bij de hogere niveaus).

Vier soorten toetsing: van kandidaat-examinator toets tot groepstoets

Sandlund et al. (2016) onderscheiden op basis van een reviewstudie drie soorten valide toetsen om de spreekvaardigheid van NT2-leerders te toetsen, te weten:

- Kandidaat-examinator toets
- Dyadische toets: toets in tweetallen
- Groepstoetsen: toets in groepen van drie of vier personen

Daarnaast is volgens Decoo (2017) de zinsreproductie een mogelijke toetsvorm.

Kandidaat-examinator toets - Deze toets bestaat meestal uit drie delen, waarbij de deelnemer een afbeelding moet omschrijven, een ononderbroken rede houdt en een conversatie voert (Sandlund et al., 2016). Deze vorm van toetsing lijkt iets minder geschikt te zijn voor het toetsen van mondelinge bekwaamheid vanwege de beperkte mogelijkheid tot interactie. De deelnemer kan veel minder verschillende acties uitvoeren, het onderwerp initiëren of veranderingen tijdens het gesprek zelf in gang zetten (Sandlund et al., 2016). De toets zorgt echter wel voor minder problemen als het gaat om standaardisatie van de toetsprocedure.

Dyadische toets: toets in tweetallen - Deze toetsvorm geeft vermoedelijk een betere natuurlijke conversatie weer (Sandlund et al., 2016). Onderzoek van Davis (2009) naar toetsing onder tweede taallleerders laat zien, dat de verschillen tussen de niveaus van de twee deelnemers niet expliciet van invloed zijn op de scores van de afzonderlijke deelnemers. Een deelnemer met een hoger niveau ondervindt vrijwel geen hinder van het samenwerken met een deelnemer van een lager niveau, terwijl de deelnemer met een lager niveau geen betere score haalt door de samenwerking met de deelnemer van een hoger niveau (Davis, 2009). Toch geeft Sun (2014) aan, dat er kritisch gekeken moet worden naar de impact van de andere gesprekspartner en het perspectief van de beoordelaar.

Groepstoetsen: toets in groepen van drie of vier personen - Groepstoetsen bestaan meestal uit groepen van drie of vier deelnemers, terwijl de examinator stil blijft en er een groepsinteractie wordt uitgelokt (Sandlund et al., 2016). Uit onderzoek van Leaper en Brawn (2019) naar de inzet van een groepstoets blijkt, dat aan de hand van de mate van complexiteit, accuraatheid en vlotheid van het spreken kon worden vastgesteld, dat

deelnemers bij de tweede toets op de meeste gebieden significant beter scoorden ten opzichte van de eerste toets. Volgens onderzoek van McDonald (2018) naar toetsing onder tweede taalleerders zou men naast de elementen vlotheid, accuraatheid, complexiteit ook aan de hand van interactie en communicatieve effectiviteit iemands taalvaardigheid kunnen beoordelen, waarbij men elk onderdeel een score van 1 tot 5 kan geven. Bij de groepstoets moet men wel rekening houden met de groepsgrootte. Vergelijkbaar onderzoek van Nakatsuhara (2011) naar de inzet van een groepstoets laat namelijk zien, dat groepen van vier deelnemers niet altijd geschikt zijn om het communicatievermogen te laten zien. Bij groepen van drie wordt duidelijk, dat de interactie natuurlijker en minder 'mechanisch' is en dat de mate van extraversie van de deelnemers zelf minder van invloed is. Ook bij deze toetsvorm moet volgens Sun (2014) op basis van een reviewstudie eveneens kritisch gekeken worden naar de impact van de andere gesprekspartners en het perspectief van de beoordelaar.

Zinsreproductie - Tenslotte blijkt zinsreproductie (het laten nazeggen van tien tot vijftig zinnen met een scoring op een vierpuntsschaal) een volwaardige toetsvorm om spreekvaardigheid te kunnen toetsen (Decoo, 2017). Onderzoek van Deygers (2020) onder volwassen tweede taalleerders bevestigt dat deze toetsvorm een adequate wijze is om spreekvaardigheid te toetsen.

Volgens Decoo (2017) laat onderzoek zien, dat de scores die deelnemers op 'open' spreektoetsen halen, goed correleren met scores op zinsreproductie. Hierbij kan dan gekeken worden naar de mate van perfecte herhaling naast fouten en haperingen. Als een deelnemer de zin goed kan herhalen, beantwoordt deze aan zijn of haar taalvermogen en kan het een voorspellende waarde voor de mondelinge competentie zijn (Decoo, 2017). Decoo (2017) geeft echter wel aan, dat het het beste lijkt om deze vorm van toetsing als controlemiddel of incidenteel alternatief in te lassen, omdat een goede score op deze toetsvorm niet inhoudt dat de deelnemer in andere omstandigheden net zo vlot spreekt.

Naast de toets is ook de kwaliteit van de beoordelaar van belang

Beoordelaars die consequent toetsen zorgen voor een grotere betrouwbaarheid van de toets. Dit is naast validiteit (meten we wat we willen meten) en haalbaarheid (wat betreft werkbelasting en technische mogelijkheden) een belangrijk criterium voor toetsing (Deygers, 2012). Kang et al. (2019) tonen met hun onderzoek aan, dat in een aantal gevallen ook de achtergrond en houding en met name de eigen spreekvaardigheid van een beoordelaar van invloed kunnen zijn op het geven van beoordelingen. Een korte gebruikersvriendelijke trainingssessie voor de beoordelaars blijkt deze bias echter veel kleiner te maken (Kang et al., 2019). Het trainen van beoordelaars lijkt hierdoor een noodzakelijke randvoorwaarde om een betrouwbare beoordeling te kunnen geven.

Geraadpleegde bronnen

Baker-Smemoe, W., Dewey, D.P., Bown, J., Martinsen, R.A. (2014). Does Measuring L2 Utterance Fluency Equal Measuring Overall L2 Proficiency? Evidence From Five Languages. *Foreign Language Annals*, 47(4), 707-728.

Cucchiarini, C. & Jaspaert, K. (1996). Tien voor taal? Toetsen van taalvaardigheid. In: VONwerkgroep NT2 (red.), *Taakgericht onderwijs: een onmogelijke taak? in de reeks Taalcahiers.* Deurne: Uitgeverij Plantyn, p. 195-226.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, *26*(3), 367-396.

Decoo, W. (2017). Zeg me na! Zinsreproductie als toetsvorm voor spreekvaardigheid. *Levende Talen (7)*, 16-21.

Deygers, B. (2012). Taaltoetsen ontwikkelen: tips en valkuilen. In Mottart, A. & Vanhooren, S., *Zesentwintigste conferentie het schoolvak Nederlands, Katholieke Hogeschool Brugge-Oostende, 16 & 17 november 2012*. Gent: Academia Press, p. 316-322.

Deygers, B. (2020). Elicited Imitation: A test for all learners? Examining the EI performance of learners with diverging educational backgrounds. *Studies in Second Language Acquisition*, *42*, p. 933 – 957.

Kang, O., Rubin, D. & Kermad, A. (2019). The effect of training and rater differences on oral proficiency assessment. *Language Testing*, *36*(4), 481-504.

Leaper, D.A. & Brawn, J.R. (2019). Detecting development of speaking proficiency with a group oral test: A quantitative analysis. *Language Testing*, *36*(2), 181-206.

McDonald, K. (2018). Post hoc evaluation of analytic rating scales for improved functioning in the assessment of interactive L2 speaking ability. *Language Testing in Asia*, *8*(19), 1-23.

Nakatsuhara, F. (2011). Effects of test-taker characteristics and the number of participants in group oral tests. *Language Testing*, *28*(4), 483-508.

Sandlund, E., Sundqvist, P. & Nyroos, L. (2016). Testing L2 Talk: A Review of Empirical Studies on Second-Language Oral Proficiency Testing. *Language and Linguistics Compass*, *10*(1), 14-29.

Sun, H. (2014). Paired and Group Oral Assessment. *Working Papers in TESOL & Applied Linguistics*, *14*(2), 68-83.

Van Til, A., Beeker, A., Fasoglio, D. & Trimbos, B. (2011). *Toetsen en beoordelen met het ERK*. Cito / SLO (nationaal expertisecentrum leerplanontwikkeling: Arnhem / Enschede.

Meer weten?

Wat is er bekend over het effectief en efficiënt formatief beoordelen?

Wat is er bekend over de effecten van formatief evalueren bij leerlingen en bij docenten?

Over deze vraag

Opgesteld door:	Maurice de Greef (kennismakelaar Kennisrotonde)
Vraagsteller:	Docent volwassenonderwijs

Onderwijssector: Volwasseneducatie

Trefwoorden: Toetsing, taal leren, NT2, beoordeling, toetsvormen

Referentie: Kennisrotonde. (2019). *Welke efficiënte en betrouwbare werkwijzen zijn er om het startniveau en de ontwikkeling van de spreekvaardigheid van volwassen NT2-leerders tijdens onderwijs Nederlands te meten?* (KR. 745)

Dit antwoord is gepubliceerd op Kennisrotonde.nl. De Kennisrotonde is samen met NCO en Onderwijskennis een dienst van het NRO.