

DAT MOET TOCH ANDERS KUNNEN?

Op zoek naar een betrouwbare én efficiënte beoordelingsmethode voor schoolexamens schrijfvaardigheid

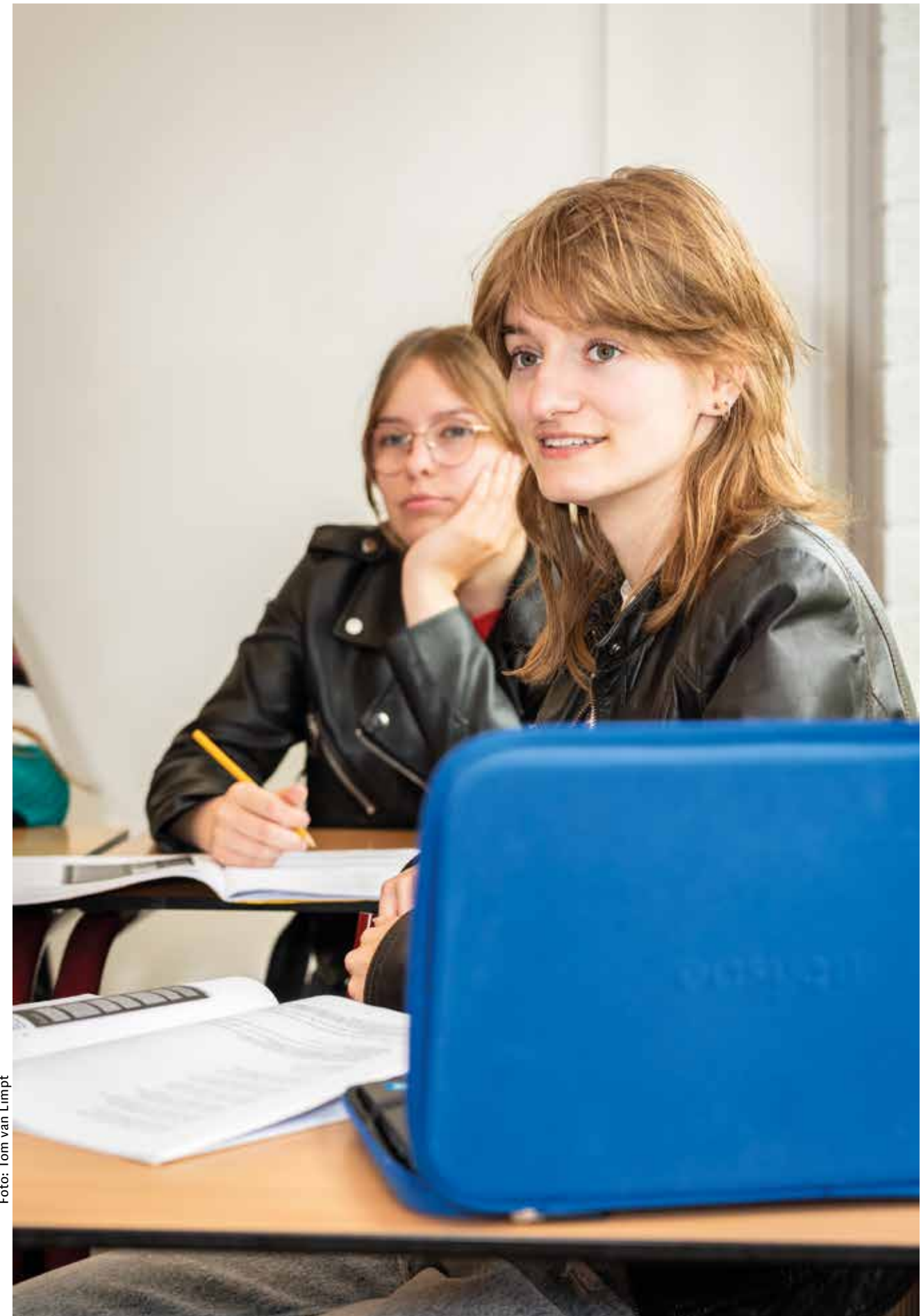
Leraren Nederlands zijn vaak niet tevreden over het beoordelen van schrijfproducten. Ze willen een hoge betrouwbaarheid, maar ook een acceptabele tijdsinvestering. Welke beoordelingsmethode kan dit bieden? Aan de hand van recente literatuur worden in dit artikel mogelijke benaderingen voor het beoordelen van schoolexamens schrijfvaardigheid besproken.

JEFF VAN DER LINDEN

Het beoordelen van schrijfproducten bij het vak Nederlands in de bovenbouw van het voortgezet onderwijs is een uitdagende opgave die menig docent tot wanhoop kan drijven. Vorig jaar kostte het me bijna twee maanden om de 96 betogende schrijfopdrachten van mijn vierdeklassers na te kijken. Naast het beoordelen van deze schoolexamens gingen mijn reguliere taken, zoals het voorbereiden van lessen en het nakijken van toetsen, gewoon door. Maar de grote boosdoener achter deze tijdrovende taak bleek vooral het beoordelingsformulier dat we in onze vakgroep gebruikten: het duurde ongeveer een halfuur per betoog voordat ik alle deelaspecten op een vijfpuntsschaal had beoordeeld. Dat moet toch anders kunnen?

Bouwer et al. (2022) onderscheiden verschillende soorten beoordelingswijzen, zoals een analytische, absolute benadering op basis van beoordelingsformulieren en rubrics én een holistische, comparatieve benadering, die uitgaat van paarsgewijze vergelijkingen of ankerschalen. Ze stellen vervolgens terecht dat docenten in het voortgezet onderwijs zich moeten afvragen hoe tekstkwaliteit en de schrijfvaardigheid van leerlingen het best kunnen worden beoordeeld. De beoordeling moet namelijk niet alleen aansluiten bij de leerdoelen van de opdracht, maar docenten moeten ook oog hebben voor de verschillen in betrouwbaarheid tussen beoordelingsmethoden. Daarnaast is het minstens zo belangrijk om de tijdsinvestering in beschouwing te nemen. Het moet immers wel te doen zijn. De vraag die docenten zouden

Foto: Tom van Limpt



moeten stellen is dan ook: hoe kunnen schoolexamens schrijfvaardigheid (betogen, beschouwingen, essays) betrouwbaar worden beoordeeld, zonder dat de tijdsinvestering onredelijk wordt?

Verschillende beoordelingswijzen

Om die vraag te kunnen beantwoorden, moet eerst duidelijk zijn wat die verschillende beoordelingswijzen nou precies inhouden. De bekendste en meest gebruikte methode is het analytisch beoordelen van een schrijfproduct op basis van een beoordelingsformulier: het werk van de leerling wordt gescoord op vooraf opgestelde deelonderwerpen, zoals: *inhoud, structuur en stijl, of specifiekere criteria als pakkende titel of overtuigingskracht van de argumentatie*. Afhankelijk van een toegekende weging vormt de optelsom van de deelscores vervolgens het eindcijfer. Een alternatief voor het beoordelingsformulier is de rubric. Hierbij wordt ieder criterium begeleid door een gedetailleerde beschrijving per prestatieniveau voor dat specifieke onderdeel. Daarmee biedt dit beoordelingsinstrument leerlingen een duidelijk beeld van de verwachte competenties.

Daartegenover staat een holistische, comparatieve beoordeling, waarbij beoordelaars zich niet richten op de eerdergenoemde deelaspecten, maar een oordeel vormen over de totaalvaardigheid en het schrijfproduct als geheel. De manier om tot een oordeel te komen, is het vergelijken van teksten: ‘afhankelijk van het communicatieve doel van de tekst bepaal je welke tekst je informatiever, overtuigender of spannender vindt’ (Bouwer & Gerits, 2022, p. 12). Voor een dergelijke comparatieve beoordeling kan worden gekozen uit twee verschillende beoordelingswijzen: de paarsgewijze vergelijking of een beoordelingsschaal met ankerteksten.

Bij het beoordelen met een beoordelingsschaal op basis van ankerteksten worden representatieve teksten geselecteerd die verschillende kwaliteitsniveaus illustreren. De beoordelaar bepaalt vooraf welke

teksten dienen als vergelijking voor bijvoorbeeld een zwakke, matige, voldoende of goede tekst. Vervolgens vergelijkt de beoordelaar ieder schrijfproduct met de geselecteerde ankerteksten op de beoordelingsschaal om de tekstkwaliteit te bepalen.

In het geval van een paarsgewijze vergelijking kiest de beoordelaar telkens de beste van twee teksten. Deze vergelijking geschiedt intuïtief op basis van een algehele indruk zonder te beoordelen op expliciete criteria; de algemene kwaliteit en de impact van de tekst als geheel zijn leidend. Hierna wordt weer een nieuw willekeurig paar vergeleken, totdat alle teksten meerdere malen zijn beoordeeld. De tekst die de meeste vergelijkingen ‘wint’, wordt vervolgens beschouwd als de beste tekst.

Een paarsgewijze beoordeling kan handmatig worden uitgevoerd (waarbij alle teksten worden geprint), maar met een groter aantal teksten is het een erg tijdsintensieve en complexe klus om de benodigde willekeurige vergelijkingen te maken. In de literatuur wordt daarom meermaals verwezen naar online tools als D-PAC en Comproved, die veel voordelen met zich mee zouden brengen. Nadat leerlingen hun werk in het programma hebben geüpload, neemt de software de logistiek uit handen door telkens willekeurige tekstparen samen te stellen. De beoordelaar voert op basis hiervan meerdere paarsgewijze vergelijkingen uit, waarna de teksten automatisch worden gerangschikt van goed naar minder goed.

Een gedeelde frustratie

Volgens Bouwer en Gerits (2022) heerst er steeds meer een gedeelde frustratie over het gebruik van beoordelingsformulieren en rubrics. Zo is de tijdsinvestering bij het beoordelen op basis van dit soort analytische beoordelingsmodellen te groot, maar lijkt ook de betrouwbaarheid én de kwaliteit van de beoordeling te wensen over te laten.

Dat zou in eerste instantie komen doordat analytische beoordelingsmodellen, die juist worden ingezet om op uniforme wijze na te kijken, geen garantie bieden voor consistente en betrouwbare oordelen.

Ondanks de duidelijkheid van de criteria en puntenverdeling blijkt dat beoordelaars aanzienlijk kunnen verschillen in hun beoordelingen van tekstkwaliteit. Zelfs wanneer er specifieke criteria zijn, bestaat er ruimte voor subjectieve interpretatie en variatie tussen beoordelaars (Bouwer, 2020b).

Verder constateren Lesterhuis et al. (2015) dat beoordelaars tijdens een analytische beoordeling ook een holistisch oordeel vellen. Dit kan leiden tot een gebrek aan transparantie, doordat de beoordeling – al dan niet bewust – wordt aangepast aan de holistische inschatting: ‘Analytische schema’s zorgen daarmee eerder voor schijnobjectiviteit dan dat ze consistente beoordelingen waarborgen’ (Bouwer & Gerits, 2022, p. 11).

Daarnaast lijkt de complexiteit van schrijven of de kwaliteit van een tekst niet adequaat te kunnen worden beoordeeld door de totaalscore van deelaspecten. Lesterhuis et al. (2015) stellen namelijk dat een selectie van deelpunten nooit de gehele totaalvaardigheid kan omvatten, en Echten et al. (2022) beargumenteren dat schrijfvaardigheid überhaupt moeilijk op te delen is in losse criteria met het bijkomend risico dat niet alle relevante aspecten worden meegenomen. De focus lijkt in ieder geval te veel op details te liggen, zoals de inzet van signaalwoorden of het correct gebruik van interpunctie, waardoor de aandacht voor de communicatieve effectiviteit van de tekst als geheel wordt weggenomen (Bouwer & Gerits, 2022).

Hoge kwaliteit en betrouwbaarheid

Waar de kwaliteit en betrouwbaarheid van de beoordeling bij analytische modellen voornamelijk problematisch lijken te zijn, lijkt een holistische en comparatieve beoordeling juist te zorgen voor een hogere kwaliteit en betrouwbaarheid. Deze benadering legt namelijk de nadruk op de expertise van de docent en richt zich op de algehele kwaliteit van de tekst in plaats van losse deelpunten. Pas als twee teksten in bepaalde opzichten sterk op elkaar lijken, zal een beoordelaar meer letten op

bepaalde deelaspecten als taalgebruik of spelling (Bouwer, 2020a).

Het vergelijken van teksten is daarnaast eenvoudiger voor beoordelaars, en daarmee ook betrouwbaarder, doordat het aansluit bij onze natuurlijke manier van beoordelen: het maken van vergelijkingen gaat de mens beter af dan het scoren van een op zichzelf staande prestatie (Bouwer, 2020b). De betrouwbaarheid is ook hoger, doordat er bijvoorbeeld geen punten toegekend hoeven te worden, en beoordelaars niet worden beïnvloed door de kwaliteit van eerder beoordeelde teksten (Lesterhuis et al., 2015; Bouwer, 2020b).

Bouwer (2020a) merkt overigens op dat bij zowel paarsgewijze vergelijking als bij een beoordeling op basis van ankerteksten de uitkomst pas betrouwbaar is als iedere tekst meerdere keren wordt vergeleken. Dit is een uitstekende aanleiding om het nakijkwerk te verdelen onder verschillende beoordelaars, wat de kwaliteit ook weer ten goede komt. Om de betrouwbaarheid van beoordelingen te waarborgen moeten we dus ‘toewerken naar een cultuur van samen beoordelen, waarbij leraren onderling teksten uitwisselen en samen tot een beoordeling komen’ (Bouwer & Gerits, 2022, p. 13). Via de eerdergenoemde software kunnen tekstparen automatisch worden gedistribueerd onder de gekoppelde beoordelaars. Bij een handmatige beoordeling op basis van ankerteksten zouden verschillende beoordelaars bijvoorbeeld hetzelfde proces kunnen doorlopen om vervolgens nog eens samen te kijken naar afwijkende uitkomsten.

Tijdwinst binnen handbereik

Ondanks dat een docent niet alleen schrijfopdrachten van zijn eigen leerlingen moet beoordelen maar ook nog de werken van collega’s moet lezen, benoemen Bouwer en Gerits (2022) dat zowel het gebruiken van een beoordelingsschaal op basis van ankerteksten als het uitvoeren van een paarsgewijze vergelijking minder tijd kost dan de analytische beoordelingswijzen. Met name het automatisch

Analytische beoordelingsmodellen,
die juist worden ingezet om
op uniforme wijze na te kijken,
bieden geen garantie voor
consistente en
betrouwbare oordelen

Zowel het gebruiken van een
beoordelingsschaal op basis
van ankerteksten als het uitvoeren
van een paarsgewijze vergelijking
kost minder tijd dan de
analytische beoordelingswijzen

Ondanks alle tekortkomingen lijken analytische beoordelingsinstrumenten toch de boventoon te voeren

genereren van paren, toewijzen aan beoordelaars en rangschikken van teksten met behulp van vergelijkingssoftware zou voor een aanzienlijke besparing zorgen (Bouwer, 2020b).

Maar ook zonder digitale hulpmiddelen lijkt de beoogde tijdwinst binnen handbereik te liggen. Een holistische lezing kost namelijk veel minder tijd dan een intensieve beoordeling van allerlei deelaspecten. Als een docent bijvoorbeeld 60 schrijfproducten van alleen zijn eigen leerlingen op analytische wijze moet becijferen, kost dit ongeveer 30 uur (60 x 30 minuten = 1.800 minuten), terwijl een globale beoordeling van laten we zeggen 120 werken uitkomt op slechts 10 uur (120 x 5 minuten = 600 minuten). Dit biedt meer dan genoeg ruimte om de werken te bespreken met collega's én te vergelijken op basis van een beoordelingsschaal – zelfs als er twee keer zoveel tijd voor de globale beoordeling zou worden uitgetrokken. Door eerst onafhankelijk holistisch te waarderen en daarna samen een comparatieve taxatie toe te passen, ontstaat tevens een deskundige beoordeling zonder dat ieder punt moet worden verantwoord op een formulier.

Bij schrijfopdrachten zijn docenten niet echt bekend met een dergelijke werkwijze, maar bij de beoordeling van bijvoorbeeld debatten des te meer. De beperkte tijd dwingt namelijk om op een meer globale wijze tot een vonnis te komen – en dat is al redelijk ingeburgerd: meerdere collega's wonen het debat bij, noteren voor zichzelf algemene opmerkingen, kennen een voorlopig cijfer toe en gaan na afloop met elkaar in beraad. Vervolgens is het dus een zinvolle toevoeging om leerlingen met elkaar te vergelijken door een ranking te maken en op basis daarvan de cijfers vast te stellen. Omdat experts met elkaar discussiëren tot er overeenstemming is bereikt, voelt dit proces betrouwbaar.

Efficiënte en rechtvaardige beoordeling

Ondanks alle tekortkomingen lijken analytische beoordelingsinstrumenten toch de boventoon te voeren. Zo brachten studenten tijdens een PTA-markt van de masteropleiding Leraar Nederlands in Sittard de schrijfexamens met bijbehorende beoordelingsmodellen van hun eigen

school mee. Opvallend genoeg maakte vrijwel iedere school gebruik van de traditionele beoordelingsformulieren met lange criterialijsten. In de literatuur is er dus een beweging ontstaan richting een holistische en comparatieve beoordelingswijze, maar in de praktijk is dit alternatief blijkbaar nog niet omarmd (Bouwer, 2020a).

De holistische en comparatieve benadering, ondersteund door tools als D-PAC en Comproved, lijkt dus uitkomst te bieden voor het betrouwbaar beoordelen van schrijfproducten in een acceptabel tijdsbestek. Toch roept ook deze methode nog enkele vragen op. Natuurlijk is het goed dat er oog is voor de kwaliteit en de communicatieve effectiviteit van de tekst als geheel, maar zijn 'oppervlakkige aspecten, zoals lay-out of taalfouten' zoals Bouwer (2020b, p. 4) benoemt dan niet belangrijk?

Desalniettemin wordt het tijd dat docenten hun conventionele beroepspraktijken heroverwegen en nadenken over alternatieven voor een efficiënte en rechtvaardige beoordeling van schrijfproducten. De huidige praktijk schiet immers tekort: het nauwkeurig beoordelen van pakweg honderd schoolexamens schrijfvaardigheid op basis van analytische methoden is simpelweg onhaalbaar. ■

LITERATUUR

- Bouwer, R. (2020a). *Comparatief beoordelen van schrijfproducten voor waardevolle feedback*. Didactiek Nederlands. <https://didactiek-nederlands.nl/zo-kan-het-ook/2020/11/comparatief-beoordelen-van-schrijfproducten-voor-waardevolle-feedback/>
- Bouwer, R. (2020b). *Op naar een beter schoolexamen schrijfvaardigheid*. Didactiek Nederlands. <https://ltn.publicatieserver.nl/download/41606>
- Bouwer, R., & Gerits, H. (2022). Aan de slag met het schoolexamen schrijfvaardigheid. *Levende Talen Magazine*, 109(2), 10–15.
- Bouwer, R., Ockenburg, L. van, Loo, J. van der, & Weijen, D. van. (2022). *Literatuurstudie naar de kenmerken van effectief schrijfonderwijs in het voortgezet onderwijs*. Universiteit Utrecht.
- Echten, K., Linnert, A., & Ockenburg, L. van. (2022). *Wie schrijft die blijft... afstemmen: Methode voor een betere beoordelaarsafstemming in één lesuur*. Didactiek Nederlands. <https://ltn.publicatieserver.nl/download/41653>
- Lesterhuis, M., Donche, V., De Maeyer, S., Daal, T. van, Van Gasse, R., Coertjens, L., Verhavert, S., Mortier, A., Coenen, T., Vlerick, P., Vanhoof, J., & Van Petegem, P. (2015). *Competenties kwaliteitsvol beoordelen: Brengt een comparatieve aanpak soelaas?* *Tijdschrift voor Hoger Onderwijs*, 33(2), 55–67.

binnenkort

20 maart 2025, Dag van de Franse Taal

20 maart 2025, Dag van de Literatuur, Rotterdam, <dagvandeliteratuur.nl>

20 – 22 maart 2025, Nascholing Duits: Recent verschenen werken, Nordhorn, <vllt.nl/rvw>

3 – 4 april 2025, Neerlandistiekdagen, Nijmegen, <vllt.nl/neerlandistiek2025>

12 april 2025, Algemene Ledenvergadering Levende Talen, Utrecht, <www.levendetalen.nl>

12 april 2025, Soapbox Challenge, Amsterdam, <www.soapboxchallenge.nl>

16 mei 2025, Flash Fiction Challenge, <vllt.nl/flashfiction>

25 mei 2025, Dag van de Jonge Jury, Rotterdam, <jongejury.nl>

20 – 21 juni 2025, Literaire Vertaaldagen, Amsterdam, <literairvertalen.org>

25 – 29 augustus 2025, Colloquium Neerlandicum, Brussel, <colloquium.ivn.nu>

14 november 2025, Congres Levende Talen, Ede, <lewendetalen.nl>

6 – 7 maart 2026, Nationaal Congres Duits, Noordwijkerhout, <nationaalcongresduits.nl>

20 maart 2026, Nationaal Congres Engels, Ede, <nationaalcongresengels.nl>

Hét Talencongres

Op 10 april organiseert ThiemeMeulenhoff Hét Talencongres 2025 in Utrecht. In de twee plenaire sessies bepleit Rob Martens, hoogleraar Onderwijswetenschappen, voor meer speelruimte en minder druk in het onderwijs, en gaat SLO in op de actualisatie van de kerndoelen en het examenprogramma voor Nederlands en de moderne vreemde talen. Verder zijn er sessies over speels leren, rijke teksten en onderwijsinnovaties zoals virtual reality in de klas.

Meer informatie is beschikbaar op <vllt.nl/hettalencongres25>.

Avondcursus gender in Duitse literatuur en taal

Jaarlijks organiseert de Vereniging van Germanisten aan Nederlandse Universiteiten (VGNU) een avondcursus. In dat kader vindt vanaf deze maand een cursus over het thema 'Gender in Literatur und Sprache' plaats. Tot midden juni zijn er in Utrecht tien lezingen in de avonduren door verschillende sprekers. De lezingen kunnen ook los worden gevolgd.

Zie voor details en aanmelding <vllt.nl/genderde>.



Lezen Centraal 2025

Op woensdag 2 april 2025 vindt in het Beatrix Theater Utrecht Lezen Centraal plaats, het jaarlijkse congres van Stichting Lezen. Het thema is 'De kracht van meertaligheid'. Ruim een kwart van de kinderen in Nederland groeit op met een andere thuistaal dan het Nederlands. Het erkennen en benutten van deze talen in de kinderopvang en het onderwijs draagt bij aan een pedagogisch klimaat waarin een kind zich gezien en gewaardeerd voelt. Goede beheersing van de eerste taal helpt bovendien bij de ontwikkeling van de tweede taal. Lezen Centraal belicht de laatste ontwikkelingen en biedt inspiratie en informatie voor kinderopvangorganisaties, scholen, opleidingen en bibliotheken.

In de ochtend is Jim Cummins de eerste keynotespreker; deze Canadese wetenschapper is een van de grondleggers van de huidige kennis over meertaligheid. Hij zal ingaan op het belang van de inzet van thuistalen voor de identiteit en het leerproces van meertalige kinderen. Vervolgens gaan Orhan Agirdag (KU Leuven) en Joana Duarte (Rijksuniversiteit Groningen) in op meertaligheid in het Nederlandse en Vlaamse onderwijs en de praktische toepassing van recente wetenschappelijke inzichten. Ellen-Rose Kambel geeft de derde keynote. Zij is oprichter van het internationale netwerk van Taalvriendelijke Scholen. Met voorbeelden vanuit deze scholen vertelt ze hoe andere landen omgaan met meertaligheid in het (lees)onderwijs. Het middagprogramma bestaat uit sessies van professionals die op verschillende praktische aspecten van meertaligheid en leesbevordering ingaan en er zijn presentaties van auteurs en illustratoren.

Voor meer informatie zie <vllt.nl/lezencentraal25>.

Dag van de Literatuur

Op donderdag 20 maart 2025 vindt in de Doelen te Rotterdam de Dag van de Literatuur plaats, een tweejaarlijks literatuurfestival voor leerlingen uit de bovenbouw havo/vwo en voor mbo-studenten. Onder anderen Splinter Chabot, Adriaan van Dis, Lize Spit, Elfie Tromp en Maartje Wortel gaan in gesprek met de deelnemers. De programmering bestaat verder uit voorleessessies, (interactieve) interviews en talkshows, waarbij de talkshowhosts de leerlingen zelf zijn. Literatuur presenteert zich multimediaal, daarom zijn filmvertoningen, muziekvertolkingen, dans en kleinkunst nadrukkelijk ook in het programma opgenomen.

Zie <dagvandeliteratuur.nl> voor meer informatie.