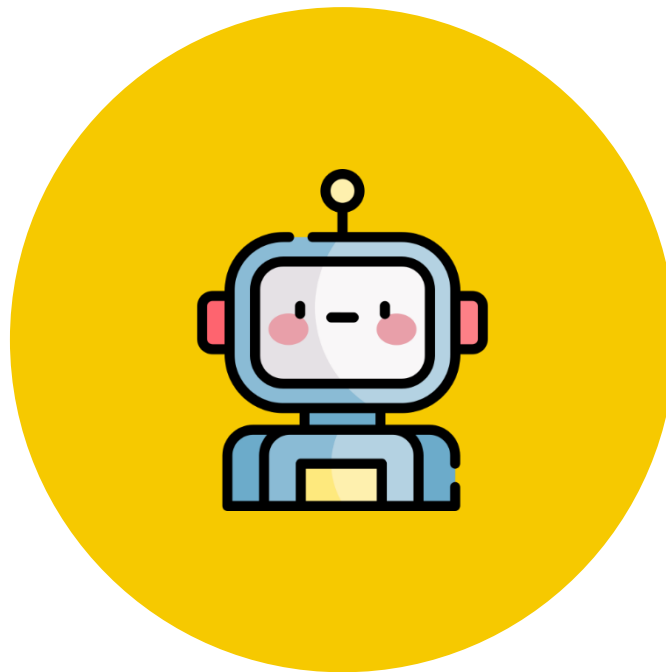


Rapport

# SchrijfBlik

Een onderzoek naar het gebruik van AI in  
schrijfopdrachten in het voortgezet onderwijs



**Auteurs:**

Marjolein Jongbloed  
Aranka Bijl  
Jasmijn Mulder  
Max van Haastrecht  
Tessa de Vries  
Levi Blom  
Iris Kamphuis

# Inhoud

	2
<b>1. Inleiding</b>	<b>3</b>
<b>2. Het SchrijfBlik prototype</b>	<b>3</b>
2.1 Het SchrijfBlik prototype	3
2.2 Docentomgeving	3
2.3 Leerlingomgeving	4
<b>3. Methode</b>	<b>6</b>
3.1 Werving	6
3.2 Deelnemers	6
3.3 Instrumenten Vragenlijst vooraf docenten	6
3.4 Procedure	9
3.5 Analyses	9
<b>4. Resultaten</b>	<b>12</b>
4.1 Docenten - competentie AI	12
4.2 Leerlingen - schrijfproces en samenwerking met Blikkie	16
4.3 Ervaringen leerlingen	21
4.4 Beoordeling docenten	26
4.5 Waarde voor beoordeling en feedback	27
<b>5. Discussie en Conclusie</b>	<b>29</b>
<b>6. Dankbetuiging</b>	<b>33</b>
<b>7. Bronnen</b>	<b>33</b>
<b>8. Bijlagen</b>	<b>34</b>
8.1 Bijlage A. Vragenlijst achteraf voor docenten	34
8.2 Bijlage B. Nadere uitwerking van leerlinginteracties met Blikkie	36

# 1. Inleiding

Generatieve kunstmatige intelligentie (genAI) neemt een steeds belangrijker plek in onze samenleving in, en dus ook in het onderwijs. Dat leerlingen veel gebruik maken van genAI is duidelijk: uit cijfers van het CBS blijkt dat 34% van de 12- tot 18-jarigen programma's zoals ChatGPT gebruikt (CBS, 2024). Sinds 2024 is het aanbod van AI-tools verder toegenomen en zijn AI-functionaliteiten steeds meer geïntegreerd in bestaande applicaties. Docenten geven aan dat zij denken dat leerlingen AI ook gebruiken in het onderwijs (Poell et al., 2025). In het hoger onderwijs lijkt gebruik wijdverbreid: een inventarisatie bij een Nederlandse instelling voor hoger onderwijs laat zien dat ruim 70% van de studenten AI-hulpmiddelen gebruikt voor hun studie (Bos et al., 2025). Docenten kijken zowel met enthousiasme als ook met bezorgdheid naar de implicaties van AI in de onderwijssector (Batista et al., 2024).

Leerlingen zetten genAI, zoals ChatGPT, vaak in om (onderdelen van) opdrachten te maken. Het is echter vaak onduidelijk hoe leerlingen het precies gebruiken (Poell et al., 2025). Voor docenten is lang niet altijd te onderscheiden welk deel van het werk authentiek is en welk deel afkomstig van genAI. Tegelijkertijd blijkt uit onderzoek dat docenten genAI niet per se willen weren uit het schrijfproces, het is vooral acceptabel in het verkennen van een onderwerp en het ontwerpen van een structuur (Barrett & Pack, 2023).

Validiteit zegt iets over de kwaliteit van zowel de opdracht en de beoordeling; meten we wat we willen meten? Het begrip validiteit omvat verschillende criteria, waaronder betrouwbaarheid, transparantie en uitvoerbaarheid. Validiteit gaat niet alleen over de inhoud en uitvoering van een opdracht, maar ook over de manier waarop resultaten geïnterpreteerd en toegepast worden (Messick, 1989; Kane, 2013).

Schrijfopdrachten hebben doorgaans als doel om schrijfvaardigheid te meten. Wanneer het bij schrijfopdrachten onduidelijk is of en hoe leerlingen genAI gebruiken, is er met name een gebrek aan transparantie hoe het werk tot stand is gekomen en welk aandeel de leerling zelf heeft in het schrijfproces. Het wordt lastig om te bepalen of je als docent de schrijfvaardigheid beoordeelt of het vermogen van leerlingen om effectief met genAI te werken. Hierdoor staat de validiteit van de beoordeling van het leerlingwerk onder druk.

Wanneer leerlingen een AI-tool gebruiken in het onderwijs (al dan niet met permissie van de docent) is het belangrijk om te weten hoe deze wordt gebruikt en ook wat de tool bijdraagt aan de onderwijskundige doelen (van Haastrecht et al., 2025). Specifiek voor het schrijven met AI spelen vragen als: *Wie (leerling of AI) heeft welk aandeel van een tekst geproduceerd? Welke prompts zijn gebruikt? Welke bewerkingen heeft de leerling zelf gedaan?*

Om docenten meer inzicht te geven in het schrijfproces, waarbij leerlingen op een verantwoorde manier genAI kunnen gebruiken, is het prototype SchrijfBlik ontwikkeld (Poell et al., 2025). Het uitgangspunt was om een applicatie te ontwikkelen waarbij de docent zowel zicht krijgt in het eindproduct als ook in het proces. Dit zou moeten leiden tot een eerlijkere beoordeling. In SchrijfBlik kan een docent een schrijfopdracht aanmaken en toewijzen. Leerlingen kunnen deze opdracht vervolgens maken met hulp van een geïntegreerde chatbot Blikkie. Blikkie geeft niet alleen antwoord op vragen van leerlingen, maar kan ook kritische

reflectievragen stellen. Na het inleveren van de opdracht kan de docent zowel het eindproduct als het schrijfproces inzien.

In het huidige onderzoek is in het voortgezet onderwijs SchrijfBlik onderzocht. Daarbij wordt onderzocht in welke mate SchrijfBlik ondersteuning kan bieden aan docenten om de validiteit van schrijfopdrachten te waarborgen. Daarbij wordt in kaart gebracht hoe zowel leerlingen als docenten het werken met SchrijfBlik ervaren, als ook hoe leerlingen hebben gewerkt samen met Blikkie.

De hoofdvraag in het onderzoek luidt: *Hoe kunnen docenten in de bovenbouw van het VO procesdata gebruiken om de validiteit van schrijfopdrachten met AI te waarborgen?*

Daarbij zijn verschillende subvragen opgesteld:

1. *Hoe competent zijn docenten op het gebied van AI? Speelt dit mogelijk een rol in het gebruik van de procesdata uit SchrijfBlik?*
2. *Hoe werken leerlingen in SchrijfBlik, en hoe werken ze samen met Blikkie?*
3. *Hoe ervaren leerlingen het werken in SchrijfBlik en het gebruik van procesdata voor hun beoordeling?*
4. *Heeft het gebruik van SchrijfBlik door leerlingen invloed op de beoordeling van de schrijfvaardigheid door docenten?*
5. *Welke elementen van het schrijfproces waar SchrijfBlik inzicht in geeft leveren waardevolle informatie voor de beoordeling en feedback en gebruiken docenten? En wat zouden ze verder nog willen zien?*

## 2. Het SchrijfBlik prototype

### 2.1 Het SchrijfBlik prototype

In het huidige onderzoek wordt het SchrijfBlik prototype (Poell et al., 2025) getest in het voortgezet onderwijs. Docenten maken een opdracht in de docentomgeving van SchrijfBlik en specificeren hoe de chatbot Blikkie gaat communiceren met leerlingen en de leerlingen ondersteunt. Leerlingen kunnen in de leerlingomgeving de opdracht inzien, deze maken en zij worden door Blikkie ondersteund in het schrijfproces. Na afronding van de opdracht kan de docent het gemaakte werk inzien en daarbij ook de interacties die de leerling met Blikkie heeft gehad.

SchrijfBlik is ontwikkeld in React (JavaScript). Voor de chatbot Blikkie werd het model GPT-4o ingezet, binnen de IT-omgeving van Cito en volgens de Algemene Verordening Gegevensbescherming (AVG). Hosting en opslag van SchrijfBlik verlopen via Google Firebase met locaties binnen Europa. De administrator kan in de database docenten toevoegen aan de SchrijfBlik-omgeving met hun e-mailadres.

### 2.2 Docentomgeving

Docenten ontvangen van de beheerder een inlogcode die is gekoppeld aan hun e-mailadres. In de SchrijfBlik docentomgeving (zie Figuur 1) kunnen docenten leerlingen toevoegen. Voor iedere leerling wordt een 5 letterige code aangemaakt. De docent kan na de code de naam van de leerling invoegen. Deze naam wordt enkel lokaal in de browser opgeslagen, de beheerder kan dit niet inzien. De docent kan voor eigen gebruik een export maken van leerlingcodes en namen. In het leerlingoverzicht kan de docent zien of de leerling nog niet is gestart met de opdracht, werkt aan de opdracht of de opdracht heeft afgerond.

Figuur 1 Docentomgeving SchrijfBlik



Wanneer de docent op instellingen klikt kan deze een opdracht klaarzetten en het systeem instellen en afstemmen op de opdracht (zie Figuur 2). De opdracht moest in deze versie van SchrijfBlik worden klaargezet in html code. Verder kan de docent de systeemprompt van Blikkie

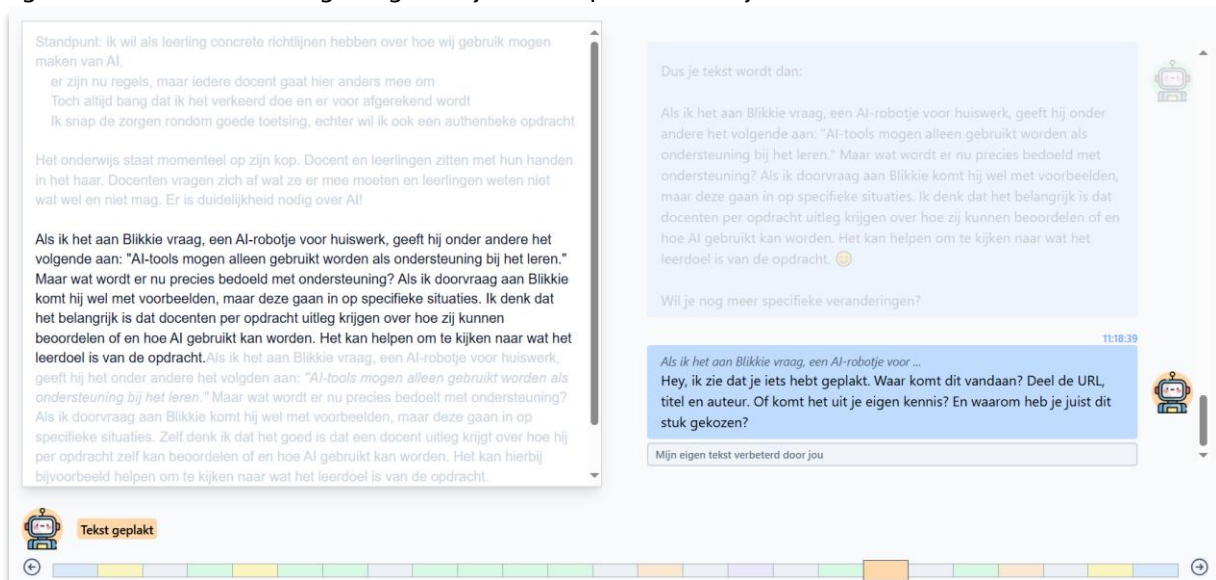
naar wens aanpassen: er kan worden ingesteld wat Blikkie doet bij aanvang en afsluiting van de opdracht, wanneer de leerling tijdelijk afwezig is geweest, welke vragen worden gesteld door Blikkie wanneer de leerling tekst knipt en plakt en welke reflectievragen worden gesteld met welk interval.

Figuur 2 Docentomgeving SchrijfBlik - Instellingen



Wanneer een leerling de opdracht heeft ingeleverd, kan de docent de opdracht inzien en ook de stappen in het schrijven van de opdracht zien, waaronder het AI gebruik van de leerling (zie Figuur 3). Dit kan de docent meenemen in de beoordeling van de opdracht.

Figuur 3 Docentomgeving SchrijfBlik – opdracht nakijken

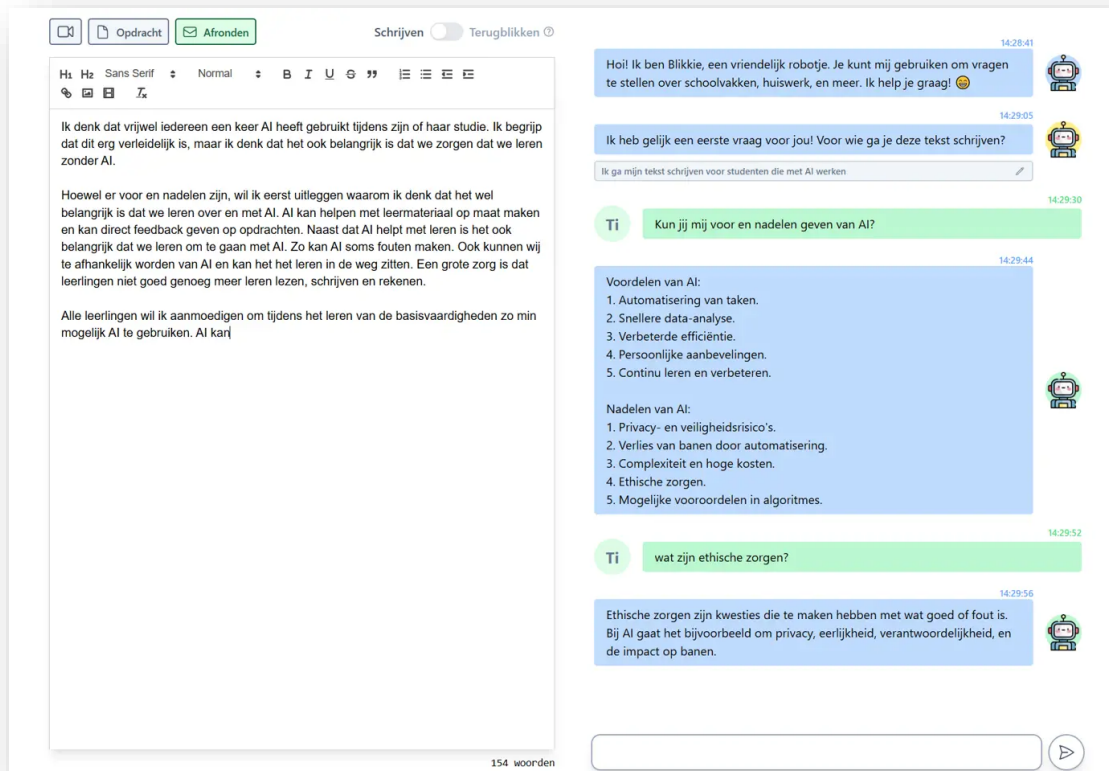


### 2.3 Leerlingomgeving

Leerlingen ontvangen van de docent hun unieke inlogcode. Figuur 4 laat zien hoe de leerlingomgeving van SchrijfBlik eruitziet. Na het inloggen worden leerlingen welkom geheten

door Blikkie. De leerlingen kunnen hun opdracht inzien. Leerlingen kunnen werken aan hun tekst in de tekstverwerker aan de linkerkant van het scherm. Zij kunnen ook een overzicht van acties inzien, zoals de docent deze ziet na het inleveren. Aan de rechterkant van het scherm kunnen leerlingen interacteren met Blikkie op eigen verzoek. Daarbij is ingesteld dat zij geen vragen kunnen stellen over gevoelige onderwerpen, zoals alcohol, drugs en geweld. Daarnaast stelt Blikkie op gezette tijden reflectievragen gerelateerd aan de opdracht, waarop de leerling kan reageren. Wanneer de leerling de opdracht heeft afgerond, kan de opdracht worden ingeleverd door op de 'afronden' knop te drukken. Ook daarna geeft Blikkie een reactie.

Figuur 4 Leerlingomgeving SchrijfBlik



## 3. Methode

### 3.1 Werving

In het huidige onderzoek is het prototype SchrijfBlik getest in het voortgezet onderwijs. Daarbij zijn docenten op verschillende manieren gevraagd om deel te nemen aan het onderzoek. Er is een LinkedIn-post geplaatst vanuit CitoLab, het onderzoek is via toetsdeskundigen bij Cito onder de aandacht gebracht bij docenten en via het Cito bureau externe medewerkers. Daarnaast heeft één van de onderzoekers het binnen het eigen netwerk onder de aandacht gebracht. Ook heeft één van de deelnemende docenten contact met de onderzoekers opgenomen op aanraden van een collega docent.

### 3.2 Deelnemers

Achttien docenten hebben zich aangemeld om deel te nemen aan het onderzoek. Zij waren werkzaam op 7 verschillende scholen. Na aanmelding en invullen van de vragenlijst vooraf heeft 1 docent (en daarmee 1 school) zich teruggetrokken omdat de school vanwege een onverwachte lage bezetting toch niet kon deelnemen. Ook heeft van een van de 6 deelnemende scholen een docent zich teruggetrokken vanwege persoonlijke omstandigheden. Uiteindelijk hebben 16 docenten verspreid over 6 scholen met een of meerdere klassen deelgenomen aan het onderzoek. De deelnemende docenten gaven les in de vakken Nederlands, Engels, Frans, Duits en Spaans. De 6 scholen waren over Nederland verspreid, in de provincies Overijssel, Noord-Brabant, Utrecht, Zeeland en Zuid-Holland. In totaal namen 24 klassen deel, met gezamenlijk 480 leerlingen. Dit waren klassen vmbo basis, vmbo kader, vmbo tl, havo en vwo, in leerjaren 4, 5 en 6, zie Tabel 1.

### 3.3 Instrumenten

#### Vragenlijst vooraf docenten

Alle deelnemende docenten hebben bij aanvang van het onderzoek een vragenlijst ingevuld, waarbij zij vragen hebben beantwoord over hun kennis over AI, vaardigheden in het omgaan met AI en hun houding ten opzichte van AI. De vragenlijst bestond uit 5 open vragen, 1 meerkeuzevraag en 23 stellingen. Daarbij is bij de stellingen gebruikt gemaakt van een 7 punts-Likertschaal (1: helemaal oneens – 7: helemaal eens). De stellingen zijn gebaseerd op een vertaalde versie van de Intelligent-TPACK-vragenlijst (Celik, 2023). De betrouwbaarheid van deze vragenlijst is onderzocht onder een grote groep Turkse docenten. De verschillende schalen vertoonden een goede interne consistentie, met Cronbach's alpha-waarden tussen 0.86 en 0.90 (Celik, 2023).

De TPACK-vragenlijst meet kennis en opvattingen van leraren met betrekking tot AI tools in het onderwijs. De TPACK-vragenlijst bestaat oorspronkelijk uit 5 schalen, in dit onderzoek zijn de schalen *Intelligent-TK*, *Intelligent-TPK*, *Intelligent-TPACK* en *Ethics* gebruikt. De schaal *Intelligent-TCK* is niet gebruikt, om de omvang van de vragenlijst te beperken en omdat deze als minst relevant werd bevonden voor het huidige onderzoek.

- *Intelligent-TK* betreft de technologische kennis van AI-tools, waaronder het kunnen gebruiken van basisfunctionaliteiten.
- *Intelligent-TPK* betreft de technologisch-pedagogische kennis, zoals het benutten van AI voor feedback, monitoring en het interpreteren van systeemmeldingen.
- *Intelligent-TPACK* betreft de kennis om AI-tools doelgericht te kiezen en in te zetten voor het behalen van leerdoelen.

- De schaal *Ethiek* meet de mate waarin leraren bij het gebruik van AI rekening houden met principes als transparantie, eerlijkheid, verantwoordingsplicht en inclusiviteit.

**Tabel 1**

*Overzicht van Deelnemende Klassen en Kenmerken*

Docent	Klassen	Niveau en jaar	Vak	Aantal leerlingen	Feedback sessie	Provincie
1	1 en 2 *	Vwo 4	Nederlands	44	Ja (2x)	Overijssel
2	3	Havo 5	Nederlands	22	-	Zeeland
3	4	Havo 5	Duits	22	Ja	Zuid-Holland
3	5	Vwo 6	Duits	22	Ja	Zuid-Holland
4	6	Vwo 6	Duits	25	Ja	Zuid-Holland
5	7	Havo 5	Duits	17	Ja	Zuid-Holland
6	8	Vwo 5	Duits	25	-	Zeeland
7	9	Vwo 6	Duits	22	Ja	Noord-Brabant
8	10	Havo 5	Frans	15	Ja	Noord-Brabant
9	11	Havo 5	Frans	17	-	Zeeland
9	12	Vwo 5	Frans	17	Ja	Zeeland
10	13	Havo 4	Frans	16	Ja	Zeeland
10	14	Vwo 6	Frans	16		Zeeland
11	15	Havo 5	Spaans	18	Ja	Noord-Brabant
11	16	Vwo 5	Spaans	17	Ja	Noord-Brabant
12	17 en 18*	Vmbo-TL 4	Engels	40	Ja	Utrecht
13	19 en 20*	Vmbo-KB 4	Engels	38	Ja	Utrecht
14	21	Vmbo-KB4	Engels	23	-	Utrecht
14	22	Vmbo-TL4		25	-	
15	23	Vmbo-BB 4	Engels	10	-	Utrecht
16	24	Vmbo-TL 4	Engels	29	-	Utrecht
Totaal				480		

\* Bij docenten die deelnamen met 2 klassen op hetzelfde niveau en in hetzelfde leerjaar, is de data niet verder per klas gespecificeerd.

### Feedbacksessies in de klas met leerlingen

Op iedere school, maar niet in iedere klas, zijn feedbacksessies uitgevoerd tijdens een lesuur. In totaal hebben 16 klassen deelgenomen aan een feedbacksessie. In totaal hebben ongeveer 310 leerlingen aan deze klassensessies deelgenomen. Daarbij is in de klas feedback opgehaald met diverse vragen. Het aantal antwoorden per vraag verschilt echter enigszins, omdat niet alle leerlingen elke vraag hebben beantwoord. Het aantal reacties per vraag ligt daardoor tussen de 99 en 110.

In de klas zijn enkele stellingen voorgelegd zoals: *Ik maak vaak gebruik van AI voor mijn schoolwerk*. Deze dienden als opwarmer en deze zijn niet meegenomen in de analyses. Vervolgens gingen de leerlingen in groepjes van maximaal 5 leerlingen uiteen om de postervragen te beantwoorden. Leerlingen schreven hun antwoorden op de vraag op een post-it, die zij vervolgens op de poster plakten. Daarna konden medeleerlingen door middel van een stipsticker aangeven dat zij het eens waren met wat een andere leerling had geschreven. Voor nieuwe antwoorden schreven zij een nieuwe post-it. Er waren in totaal 5 posters. De groepjes hadden ca. 5 minuten de tijd om een vraag te beantwoorden en draaiden daarna door naar een andere poster, totdat alle leerlingen de mogelijkheid hadden gehad om iedere vraag te beantwoorden. De vragen waren:

- Hoe verschilde jouw gebruik van SchrijfBlik van hoe je normaal AI-tools zoals ChatGPT gebruikt voor schrijfp opdrachten?
- Heb je dingen anders gedaan tijdens het schrijven omdat je wist dat de docent inzicht had in wat je deed? Waarom wel of waarom niet?
- Voor wat voor soort taken vind jij dat je geen AI mag gebruiken in je schoolwerk?
- Wat vind je ervan dat bij het nakijken niet alleen je eindresultaat beoordeeld wordt, maar ook de manier waarop je de opdracht hebt aangepakt?
- Heeft SchrijfBlik je docent geholpen om te begrijpen hoe jij AI gebruikt? Waarom wel of waarom niet?

De klassensessie werd gezamenlijk afgesloten.

Hoewel leerlingen stickers konden plakken bij antwoorden waar zij het mee eens waren, gingen niet alle klassen hier even serieus mee om. Uit de stickers bleek wel dat veelgenoemde antwoorden doorgaans ook de meeste stickers ontvingen. Het aantal stickers gaf echter een vertekend beeld, omdat er soms meer stickers bij een opmerking werden geplakt dan er leerlingen aanwezig waren. Daarom is er in de uitwerking voor gekozen om te focussen op de inhoud van de post-its en het aantal stickers verder buiten beschouwing te laten.

### Vragenlijst achteraf met leerlingen

In de klassen waar geen feedbacksessie is uitgevoerd is een online vragenlijst aangeboden. Deze vragenlijst is ingevuld door leerlingen uit klassen waarvoor geen klassensessie kon worden ingepland (8 klassen, ca. 170 leerlingen). In totaal hebben 57 leerlingen de vragenlijst ingevuld. De vragen in de enquête waren gebaseerd op de postervragen uit de feedbacksessie. Daarbij werden ook de opwarmstellingen voorgelegd. De vragenlijst werd afgesloten met de mogelijkheid om opmerkingen en tips te geven over SchrijfBlik.

### Vragenlijst achteraf docenten

Hierin werd op basis van onderzoeksvragen 1b en 2 gevraagd naar hoe het werken met SchrijfBlik werd ervaren, welke inzichten SchrijfBlik gaf, hoe de opdracht werd beoordeeld en in hoeverre de inzichten vanuit SchrijfBlik zijn gebruikt in de beoordeling. Dit werd uitgevraagd aan de hand van 2 meerkeuzevragen, 17 stellingen, 8 open vragen, een classificatievraag en een beoordelingsvraag. De vragenlijst is ingevuld door 11 docenten. Dit betekent dat 5 docenten de vragenlijst achteraf niet meer hebben ingevuld. De vragenlijst is toegevoegd in Bijlage A.

### 3.4 Procedure

Docenten hebben informatie ontvangen over het onderzoek en er is een toestemmingsverklaring getekend door de docent en een vertegenwoordiger van de school. De docenten hebben een informatiebrief ontvangen om leerlingen en hun ouders te informeren over het onderzoek.

Docenten zijn gestart met het invullen van de vragenlijst vooraf. De link naar de vragenlijst is per mail verstuurd naar de docenten. Vervolgens mochten docenten zelf een opdracht ontwerpen voor hun eigen vak om te gebruiken in het onderzoek. Op basis van deze opdracht hebben de docenten besloten of leerlingen de SchrijfBlik opdracht in één lesuur of over meerdere lessen mochten maken. Sommige collega docenten hebben dezelfde opdracht gebruikt binnen hun sectie. Docenten hebben fysiek of online een demonstratie van SchrijfBlik gezien, die gegeven werd door de onderzoekers. Aan docenten is nadrukkelijk gevraagd om aan leerlingen te communiceren geen persoonsgegevens in SchrijfBlik in te vullen, en dit ook mee te nemen in de te maken schrijfoopdracht. Wanneer aanvullende hulp nodig was bij het inrichten van de SchrijfBlik omgeving, werd hiervoor een aanvullende afspraak ingepland.

Daarna hebben de docenten de opdracht in hun klas uitgezet. Er is docenten gevraagd om leerlingen te instrueren om geen persoonsgegevens te delen in de SchrijfBlik-omgeving. Tijdens het schrijfproces is het werk van leerlingen gelogd. Na het afronden van de klassenopdracht, zijn deze logs waar nodig geanonimiseerd en onderzocht (zie 3.5 analyses). De verschillende interacties met Blikkie zijn gecodeerd, zowel de interacties die door Blikkie zijn gestart als ook de interacties die de leerling initieerde.

Na afronding en eventuele beoordeling van de opdrachten zijn de onderzoekers op bezoek gegaan bij de scholen voor een feedbacksessie in de klas. Dit vond plaats binnen een regulier lesuur (45-50 minuten) en bij voorkeur binnen twee weken na afronding. In de klassen waar geen feedbacksessie is uitgevoerd is een vragenlijst aangeboden. Er zijn briefjes uitgedeeld met een link en QR-code naar een enquête en de link en QR-code zijn ook verstuurd naar de mailadressen van de docenten om verder te delen met hun leerlingen. Docenten is gevraagd om de vragenlijst achteraf in te vullen. De link naar de vragenlijst is wederom per mail verstuurd naar de docenten. De vragenlijst vooraf, achteraf en de vragenlijst voor leerlingen waren allen Microsoft Forms enquêtes.

### 3.5 Analyses

#### Vragenlijsten docenten en leerlingen

De vragenlijsten zijn zowel kwantitatief als kwalitatief geanalyseerd. Voor de kwantitatieve analyse zijn de gegevens uit Google Forms gebruikt, waarbij alleen de vragen die verdere analyse vereisten in Excel zijn uitgewerkt, bijvoorbeeld voor het omrekenen naar percentages van positieve, neutrale en negatieve antwoorden.

De kwalitatieve analyse richtte zich met name op open vragen. De gegeven antwoorden zijn verzameld en, waar mogelijk, geclusterd op basis van overkoepelende thema's, zoals functionaliteiten van Blikkie. Daarnaast zijn de reacties op de open vragen, waar mogelijk, gecategoriseerd als (overwegend) positief of negatief om meer inzicht te krijgen in de algemene tendensen binnen de antwoorden. Representatieve uitspraken zijn geselecteerd om de belangrijkste bevindingen te illustreren en te verduidelijken.

## Feedbacksessie leerlingen

De posters met post-its en stickers zijn gedigitaliseerd. Per klas zijn de antwoorden op de verschillende vragen verzameld, waarbij de antwoorden op de post-its zijn overgenomen in een Excel bestand en daarbij is genoteerd of en hoeveel stickers zijn geplakt bij ieder antwoord. Vervolgens is dit van alle klassen samengevoegd en verder geanalyseerd in Excel.

## Logs schrijfproces

Per klas is voor elke leerling een log geëxporteerd. Voor de eerste school (2 klassen) zijn er aan de hand van de logs door twee onderzoekers gezamenlijk codes opgesteld waarmee ieder log is geanalyseerd in Excel. Daarin werden drie hoofdcategorieën onderscheiden: 1) Reactie op voorgeprogrammeerde Blikkie vraag/opmerking, 2) Eigen initiatief chat of reactie op vorige chat van Blikkie en 3) Kopiëren/plakken. Vervolgens is door twee andere onderzoekers deze code ook voor de twee daaropvolgende scholen (6 klassen) gebruikt. Er bleek tijdens deze analyse dat de codes binnen de hoofdcategorieën verdere uitwerking nodig hadden en ook enige uitbreiding. Daarop zijn de eerder geanalyseerde logs opnieuw doorgenomen en waar nodig is de codering aangepast. Met deze versie van de codes zijn de overige logs geanalyseerd (zie resultatensectie 4.2).

Voor de laatst deelnemende school (8 klassen) is geprobeerd het coderingsproces te vereenvoudigen door deze automatisch te laten coderen. Hiervoor is op basis van de eerder geanalyseerde logs en bijbehorende codes een AI-model gebouwd. Voor deze automatische codering van de interacties tussen leerlingen en Blikkie is een Python-script ontwikkeld dat gebruikmaakt van een large language model (LLM; GPT-4o). Allereerst werden de ruwe logbestanden (JSON-bestanden) met leerling-systeeminteracties ingelezen en omgezet naar een gestandaardiseerde tabelvorm, waarin per regel is vastgelegd wat voor soort gebeurtenis het is (bijvoorbeeld chatbericht, tekstbewerking, reflectievraag), wie de 'afzender' is (leerling of systeem) en wat de bijbehorende tekst is. Systematische gebeurtenissen van het systeem (Blikkie) werden hierbij gescheiden van leerlingacties, en speciale gevallen zoals geen reactie van de leerling of knip- en plakacties werden expliciet als aparte type gebeurtenissen gemarkeerd.

Daarnaast werd een vooraf opgesteld coderingsschema ingelezen vanuit een Excelbestand, waarin per code een nummer, type actie, specificatie, uitleg en voorbeelden zijn vastgelegd. Dit coderingsschema en de voorbereikte logregels werden vervolgens als input aangeboden aan de LLM. Daarnaast krijgt het model een instructie mee om uitsluitend leerlingregels te coderen, per leerlingregel precies één code toe te kennen en zelf geen nieuwe codes te bedenken maar alleen te kiezen uit de meegegeven lijst met codes. De output van het model werd vervolgens gestructureerd en opgeslagen in een Excelbestand, waarbij per regel onder meer de klascode, leerlingcode, het betreffende log-ID, de toegewezen code en een kort citaat werden vastgelegd. Het automatisch coderen is eerst getest als concept. In een tweede versie werd deze methode verder verfijnd. De meeste codes werden correct herkend, maar enkele codes werden minder accuraat herkend (bijvoorbeeld: *Doel van de prompt is onduidelijk* of *Geen respons*). Deze automatisch gecodeerde logs zijn daarna gecontroleerd en aangevuld door twee van de onderzoekers, waarbij extra aandacht was voor deze specifieke codes.

Na het coderen zijn de resultaten van alle klassen samengevoegd in één databestand in Excel. Voor de kwantitatieve analyse is per code onderzocht hoeveel leerlingen deze hebben toegepast. Op basis hiervan zijn gemiddelden en standaarddeviaties berekend om inzicht te krijgen in de spreiding en mate van voorkomen van de verschillende codes in de antwoorden

van leerlingen. Daarnaast is een kwalitatieve analyse uitgevoerd. Per code zijn relevante fragmenten uit de loggegevens geselecteerd. Deze citaten dienen ter illustratie van de manier waarop leerlingen de betreffende aspecten toepasten en geven meer context aan de kwantitatieve resultaten.

### Ingeleverde opdrachten

Daarnaast zijn er analyses uitgevoerd op de ingeleverde opdrachten, waaronder hoeveel karakters zijn gebruikt en welk percentage van de opdracht geknipt en geplakt is vanuit Blikkie of vanuit andere bronnen. Hiervoor zijn de logs met behulp van R (in VS Code) opgeschoond en samengevoegd tot één groot databestand. Op basis hiervan zijn de logs per leerling geanalyseerd op onder andere het aantal acties van de leerling, het percentage van de tekst dat door de leerling zelf is geschreven, en het aantal interacties met de chatbot Blikkie. Voor deze analyses zijn per klas de gemiddelden en standaarddeviaties berekend, waarbij alleen teksten van meer dan 450 karakters zijn meegenomen. Deze drempel is gekozen om zeer korte, waarschijnlijk onvolledige teksten uit te sluiten. Bij het analyseren van de eigen bijdrage van leerlingen, is uitgegaan van een overeenkomende minimale reeks van 15 karakters, om te voorkomen dat veelvoorkomende korte woorden ook worden gezien als tekst afkomstig van Blikkie. Als laatste zijn de hoeveelheid activiteiten door leerlingen geanalyseerd. Zoals gezegd kan de docent de stappen van schrijven inzien; deze stappen zijn namelijk opgedeeld in verschillende activiteitsblokken. Wanneer leerlingen bijvoorbeeld een langere periode aan het schrijven zijn, wordt dit samengevat in één activiteitsblok. Het aantal activiteitsblokken is daarmee ook geen reflectie van hoeveel tijd de leerling aan de schrijfoopdracht heeft gespendeerd.

Daarnaast zijn er twee aanvullende tekstanalyse-methoden toegepast om de ingeleverde opdrachten onderling te vergelijken (binnen een klas). De eerste analyse was de *trigram-based similarity index*. Deze methode kwantificeert de gelijkheid tussen teksten op basis van reeksen van drie opeenvolgende woorden. Er is gekozen voor *trigrams* in plaats van bijvoorbeeld *unigrams* (enkele woorden), omdat teksten die op basis van dezelfde opdracht zijn geschreven vaak vergelijkbare woordenschat bevatten. Analyse op enkel woordniveau zou daardoor te gevoelig zijn voor overeenkomsten. Reeksen van drie woorden geven een sterkere indicatie van overeenkomstige zinsdelen tussen teksten. Hiermee kan onderzocht worden of (delen van) teksten afkomstig van Blikkie een hogere onderlinge gelijkheid vertonen dan (delen van) teksten die leerlingen zelf hebben geschreven.

De tweede analyse is de *Coleman-Liau grade level index* (Coleman & Liau, 1975), een maat voor tekstcomplexiteit die aangeeft welk (Amerikaans school-)niveau nodig is om een tekst te begrijpen. Deze index is gebaseerd op het aantal letters per 100 woorden en het aantal zinnen per 100 woorden. De hypothese hierbij is dat teksten die (gedeeltelijk) door GenAI zijn gegenereerd, complexer zijn dan teksten die door leerlingen zelf zijn geschreven. De verwachting is daarom dat leerlingen die veel tekst van Blikkie hebben overgenomen een hogere tekstcomplexiteit laten zien dan leerlingen die hun tekst grotendeels zelf hebben geschreven.

## 4. Resultaten

De resultaten worden besproken aan de hand van het onderzoeksproces. De verschillende onderzoeksvragen komen hierbij aan bod.

### 4.1 Docenten - competentie AI

#### **Onderzoeksvraag 1: Hoe competent zijn docenten op het gebied van AI? Speelt dit mogelijk een rol in het gebruik van de procesdata uit SchrijfBlik?**

##### **Beleving van AI**

Op de vraag ‘Wat is uw definitie van AI?’ liepen de antwoorden van docenten uiteen van algemene beschrijvingen, zoals *“Een hulpmiddel dat je kunt inzetten bij diverse werkzaamheden”*, tot meer specifieke omschrijvingen, zoals *“Het nabootsen van menselijke intelligentie op het gebied van redeneren, het bedenken van oplossingen en creatief produceren”*. Sommige docenten vertaalden de term letterlijk als ‘*kunstmatige intelligentie*’, terwijl anderen AI vooral benaderden vanuit hun eigen ervaringen, bijvoorbeeld als *“een brainstormpartner die kan ondersteunen bij het zoeken naar en ontwikkelen van producten”* en *“een handig hulpmiddel om dingen op te zoeken”*.

Docenten passen AI toe in hun dagelijks leven. Zo geven zij aan ChatGPT te gebruiken, en ook toepassingen als Google (Assistent) en Gemini. Daarnaast verwezen enkele docenten niet zozeer naar concrete tools, maar naar hun eigen gebruik van AI, bijvoorbeeld bij het schrijven van e-mails en het ontwikkelen van lesideeën, of het gebruik van AI door hun leerlingen.

De gevoelens die docenten met AI associëren zijn erg wisselend. Acht docenten benoemden zowel positieve als negatieve gevoelens, zoals *“onzekerheid, ongemak, kansrijk”* en *“handig, riskant, oppervlakkig”*. Vier docenten waren overwegend positief en gebruikten termen als *“enthousiast, gemotiveerd”* en *“interessant, boeiend, nieuwsgierig, leuk”*. Zes docenten uitten daarentegen uitgesproken negatieve gevoelens, waaronder *“onzekerheid, frustratie, wantrouwen”* en *“ongrijpbaar, overweldigend, angst”*.

In een nadere toelichting van deze gevoelens komt vooral een gevoel van onwetendheid naar voren, zowel meer in het algemeen als bij leerlingen in het bijzonder, over hoe AI op een verantwoorde manier ingezet kan worden. Dit blijkt onder andere uit uitspraken als *“Mensen realiseren zich niet hoe ver het kan gaan”* en *“Gefrustreerd omdat kinderen het gebruiken in plaats van te leren, onwennig omdat ik denk te weinig kennis te hebben”*. Ook andere docenten benoemen expliciet hun eigen gebrek aan kennis over AI, bijvoorbeeld vanwege *“het gigantische aanbod en het feit dat ik er nog zo weinig van weet”*. Een ander veelgenoemd risico is dat het gebruik van AI kan leiden tot minder kritisch denken. Zo uit een docent de angst dat *“mensen beslissingen nemen op grond van verkeerde feiten”* en dat informatie sneller wordt aangenomen zonder kritische reflectie. Tegelijkertijd zien verschillende docenten ook de kansen die AI biedt en beschouwen zij AI als een onvermijdelijk onderdeel van de toekomstige samenleving. Zij benadrukken het belang van onderwijs dat hierop aansluit: *“We moeten onze leerlingen opleiden voor hun toekomst, daar hoort AI nu eenmaal bij. Daarom is het belangrijk dat we de leerlingen leren hoe ze ermee moeten omgaan.”*

## Zelfinschatting kennis over AI

De docenten hebben verschillende stellingen beantwoord over hun kennis over AI. De stellingen zijn onderverdeeld in meerdere thema's rondom AI, namelijk 1) algemene omgang met AI-toepassingen, 2) pedagogische mogelijkheden, 3) onderwijsstrategieën en -doelen, en 4) ethische aspecten. Deze stellingen zijn beantwoord op een 7 punts-Likertschaal (1: helemaal oneens – 7: helemaal eens).

### Algemene omgang en toepassingen van AI

De resultaten laten overwegend licht positieve scores zien op algemene kennis en omgang met AI-toepassingen, zoals weergegeven in Tabel 2. Respondenten geven met name aan goed te weten hoe zij taken met AI moeten uitvoeren (M = 4,9), waarbij 83% positief scoort. Ook het dagelijks omgaan met AI (M = 4,5) en het beschikken over voldoende kennis om AI te gebruiken (M = 4,4) worden door meer dan de helft van de docenten positief beoordeeld. Het initiëren van een taak via tekst of spraak (M = 4,1) en bekendheid met AI-toepassingen en hun technische capaciteiten (M = 4,1) scoren eveneens boven de neutrale middenwaarde, al is hier meer spreiding zichtbaar en een relatief grotere groep die neutraal of negatief antwoordt. Samengevat duiden de resultaten op een redelijk vertrouwen in de eigen AI-basisvaardigheden en praktische omgang met AI, met meer variatie in antwoorden waar het gaat om diepere technische bekendheid.

**Tabel 2**

*Beschrijvende statistieken van zelfgerapporteerde kennis en vertrouwdheid met AI-toepassingen (n=18).*

Stelling	Gemiddelde (M)	SD	Percentage positief (≥5)	Percentage neutraal (=4)	Percentage negatief (≤3)
Ik weet hoe ik in het dagelijks leven moet omgaan met AI-toepassingen	4,5	1,47	56%	28%	16%
Ik weet van een aantal taken hoe ik ze moet uitvoeren met AI-toepassingen	4,9	1,26	83%	11%	6%
Ik weet hoe ik een taak voor AI-toepassingen moet initialiseren via tekst of spraak	4,1	1,86	50%	17%	33%
Ik heb voldoende kennis om AI-toepassingen te gebruiken	4,4	1,82	56%	22%	22%
Ik ben bekend met AI-toepassingen en hun technische capaciteiten	4,1	1,73	39%	33%	28%

### AI en pedagogische mogelijkheden

De resultaten laten een uiteenlopend beeld zien op het gebied van kennis over de pedagogische mogelijkheden van AI, zie Tabel 3. Docenten scoren zichzelf relatief hoog op het begrijpen van de pedagogische bijdrage van AI aan hun onderwijspraktijk (M = 4,6) en op het evalueren van AI-gegenereerde feedback (M = 4,3). Meer dan de helft van de respondenten beoordeelt zichzelf op deze punten positief, wat wijst op vertrouwen in de meerwaarde van AI. De scores liggen lager bij meer toepassingsgerichte en analytische vaardigheden. Het selecteren van passende AI-toepassingen blijft rond de neutrale middenwaarde (M = 3,7), terwijl het inzetten van AI om leren te volgen (M = 2,8), het interpreteren van realtime signalen (M = 2,8) en het ondersteunen van motivatie (M = 3,1) overwegend negatief worden beoordeeld. Samengevat voelen docenten zich vooral bekwaam in het beoordelen van de pedagogische

waarde van AI, maar minder in het concreet en datagedreven toepassen ervan in het leerproces van leerlingen.

**Tabel 3**

*Beschrijvende statistieken van zelfgerapporteerde kennis over het gebruik van AI voor pedagogische doeleinden (n=18).*

Stelling	Gemiddelde (M)	SD	Percentage positief ( $\geq 5$ )	Percentage neutraal (=4)	Percentage negatief ( $\leq 3$ )
Ik kan de pedagogische bijdrage van AI-toepassingen aan mijn onderwijsveld begrijpen	4,6	1,58	61%	22%	17%
Ik kan de waarde van feedback van AI- toepassingen evalueren voor onderwijzen en leren	4,3	1,78	56%	11%	33%
Ik kan AI-toepassingen selecteren die leerlingen in staat stellen hun kennis toe te passen	3,7	1,67	33%	28%	39%
Ik weet hoe ik AI-toepassingen kan inzetten om het leren van leerlingen te volgen	2,8	1,40	11%	22%	67%
Ik kan signalen van AI-toepassingen interpreteren om realtime (d.w.z. directe) feedback te geven	2,8	1,42	11%	28%	61%
Ik kan signalen van AI-toepassingen interpreteren om het leren van leerlingen te ondersteunen	3,2	1,47	17%	28%	56%
Ik kan AI-toepassingen inzetten om de motivatie van leerlingen te ondersteunen	3,1	1,49	22%	22%	56%

### AI en onderwijsstrategieën en -doelen

De gemiddelde scores rondom vragen over AI en onderwijsstrategieën en doelen vallen overwegend laag tot rond de neutrale middenwaarde uit (Tabel 4). Met name het inzetten van AI voor adaptieve feedback (M = 3,1), gepersonaliseerd leren (M = 3,2) en realtime feedback (M = 2,9) wordt door een meerderheid van de respondenten negatief beoordeeld (steeds 56%). Dit wijst op beperkte ervaren bekwaamheid om AI vakinhoudelijk en differentieel in te zetten. Ook het volgen van het leerproces van leerlingen met behulp van AI scoort laag (M = 2,4), met 72% negatieve antwoorden en geen positieve scores. Daarnaast voelt slechts een kleine groep zich in staat een leiderschapsrol te vervullen bij AI-integratie (M = 2,9). De meer algemene didactische inzet van AI laat een iets gematigder beeld zien. Het onderwijzen van een vak met behulp van AI voor diverse didactische strategieën (M = 3,9) en het combineren van AI met didactische strategieën (M = 3,7) scoren rond de neutrale grens, met een verdeeld antwoordpatroon. Samengevat ervaren docenten vooral beperkingen in het vakdidactisch, adaptief en datagedreven inzetten van AI, terwijl zij zich iets zekerder voelen over de meer algemene didactische toepassing ervan.

**Tabel 4**

*Beschrijvende statistieken van zelfgerapporteerde kennis over het gebruik van AI voor onderwijsstrategieën en onderwijsdoelen (n=18).*

Stelling	Gemiddelde (M)	SD	Percentage positief (≥5)	Percentage neutraal (=4)	Percentage negatief (≤3)
In mijn vakgebied weet ik hoe ik verschillende AI-toepassingen kan inzetten voor adaptieve feedback	3,1	1,73	28%	17%	56%
In mijn vakgebied weet ik hoe ik verschillende AI-toepassingen kan inzetten voor gepersonaliseerd leren	3,2	1,80	22%	22%	56%
In mijn vakgebied weet ik hoe ik verschillende AI-toepassingen kan inzetten voor realtime (d.w.z. directe) feedback	2,9	1,63	17%	28%	56%
Ik kan een vak onderwijzen met behulp van AI-toepassingen voor diverse didactische strategieën	3,9	1,78	39%	28%	33%
Ik kan lessen geven waarin ik mijn onderwijsinhoud, AI-toepassingen en didactische strategieën op de juiste manier combineer	3,7	1,87	50%	6%	44%
Ik kan een leiderschapsrol vervullen bij mijn collega's wat betreft de integratie van AI-toepassingen in onze onderwijspraktijk	2,9	1,49	11%	39%	50%
Ik kan verschillende AI-toepassingen inzetten om het leren van mijn leerlingen in het onderwijsproces te volgen	2,4	1,25	0%	28%	72%

### AI en ethische aspecten

Docenten hebben verschillende vragen beantwoord over ethische aspecten van AI. Binnen deze categorie vallen de scores laag uit, zie Tabel 5. Alle gemiddelden liggen ruim onder de neutrale middenwaarde (tussen  $M = 2,3$  en  $M = 2,7$ ), wat wijst op een overwegend negatieve zelfinschatting van de eigen bekwaamheid op het gebied van het beoordelen van AI-toepassingen. Met name het kunnen beoordelen of AI-toepassingen rekening houden met verschillen tussen leerlingen ( $M = 2,3$ ; 78% negatief) en het kunnen identificeren van verantwoordelijke ontwikkelaars ( $M = 2,4$ ; 78% negatief) worden door een grote meerderheid negatief beoordeeld. Ook het inschatten van eerlijkheid tegenover alle leerlingen ( $M = 2,3$ ; 72% negatief) en het begrijpen van de verantwoording achter AI-beslissingen ( $M = 2,7$ ; 72% negatief) scoren zeer laag. Samengevat ervaren docenten duidelijke beperkingen in hun kennis en vaardigheden met betrekking tot ethische aspecten, zoals bias, transparantie en verantwoordingsstructuren van AI-toepassingen.

**Tabel 5**

*Beschrijvende statistieken van zelfgerapporteerde kennis over ethische aspecten van AI (n=18).*

Stelling	Gemiddelde (M)	SD	Percentage positief ( $\geq 5$ )	Percentage neutraal (=4)	Percentage negatief ( $\leq 3$ )
Ik kan beoordelen in hoeverre AI-toepassingen rekening houden met verschillen (bv. geslacht) van	2,3	1,57	11%	11%	78%
Ik kan beoordelen in hoeverre AI-toepassingen zich eerlijk gedragen tegenover alle leerlingen in mijn onderwijspraktijk	2,3	1,41	6%	22%	72%
Ik kan de verantwoording begrijpen van elke beslissing die een AI-toepassing neemt	2,7	1,71	11%	22%	72%
Ik kan begrijpen wie de verantwoordelijke ontwikkelaars zijn in het ontwerp en de besluitvorming van AI-toepassingen	2,4	1,65	6%	17%	78%

## 4.2 Leerlingen - schrijfproces en samenwerking met Blikkie

### Onderzoeksvraag 2: Hoe werken leerlingen in SchrijfBlik, en hoe werken ze samen met Blikkie?

#### Ingeleverde teksten

De gemiddelde lengte van de uiteindelijk ingeleverde teksten (m.u.v. teksten onder de 450 karakters) was 1181,04 karakters. Er is veel spreiding tussen de leerlingen ( $sd = 750,49$ ) met een lengte tussen de 458 en 4395 karakters. Dit beeld is hetzelfde tussen de klassen en is waarschijnlijk toe te schrijven aan de verschillende schrijfopdrachten die leerlingen hebben gemaakt. Zo is het laagste gemiddelde 763,52 karakters ( $sd = 161,75$ ,  $min = 568$ ,  $max = 1122$ ) en het hoogste gemiddelde 2767,55 karakters ( $sd = 714,61$ ,  $min = 1009$ ,  $max = 4264$ ). Over alle klassen heen was de eigen bijdrage van leerlingen in de schrijfproducten gemiddeld 79,75 % ( $sd = 34,89$ ). De klas met de hoogste gemiddelde eigen bijdrage ( $M = 94,07$  %;  $sd = 19,15$ ) maakte een persoonlijke opdracht over hun toekomst, en was daardoor ook geen opdracht die een leerling volledig aan Blikkie kan uitbesteden. De klas met de laagste eigen bijdrage van de leerlingen ( $M = 47,21$  %;  $sd = 45,45$ ), schreef een forumbijdrage waarin leerlingen hun mening geven over een onderwerp, een schrijfopdracht die ook goed door een genAI zoals Blikkie kan worden geschreven.

Van de teksten die leerlingen hebben ingeleverd, was gemiddeld 15,15 % van de tekst afkomstig van Blikkie ( $sd = 30,96$ ). Gemiddeld hebben leerlingen 5,10 % van de tekst overgenomen uit andere, externe bronnen ( $sd = 19,49$ ). Beide statistieken hebben wederom een grote spreiding en variëren tussen 0 en 100 %. Leerlingen maakten dus gemiddeld gezien meer gebruik van Blikkie dan van bronnen buiten de SchrijfBlik-omgeving. We kunnen echter niet vaststellen of teksten van buiten de omgeving genAI teksten zijn. De klas met het hoogste percentage tekst afkomstig van Blikkie had een gemiddelde van 48,26 % ( $sd = 45,51$ ). Van de 16 deelnemende docenten, heeft bij één docent geen enkele leerling tekst van Blikkie (van meer

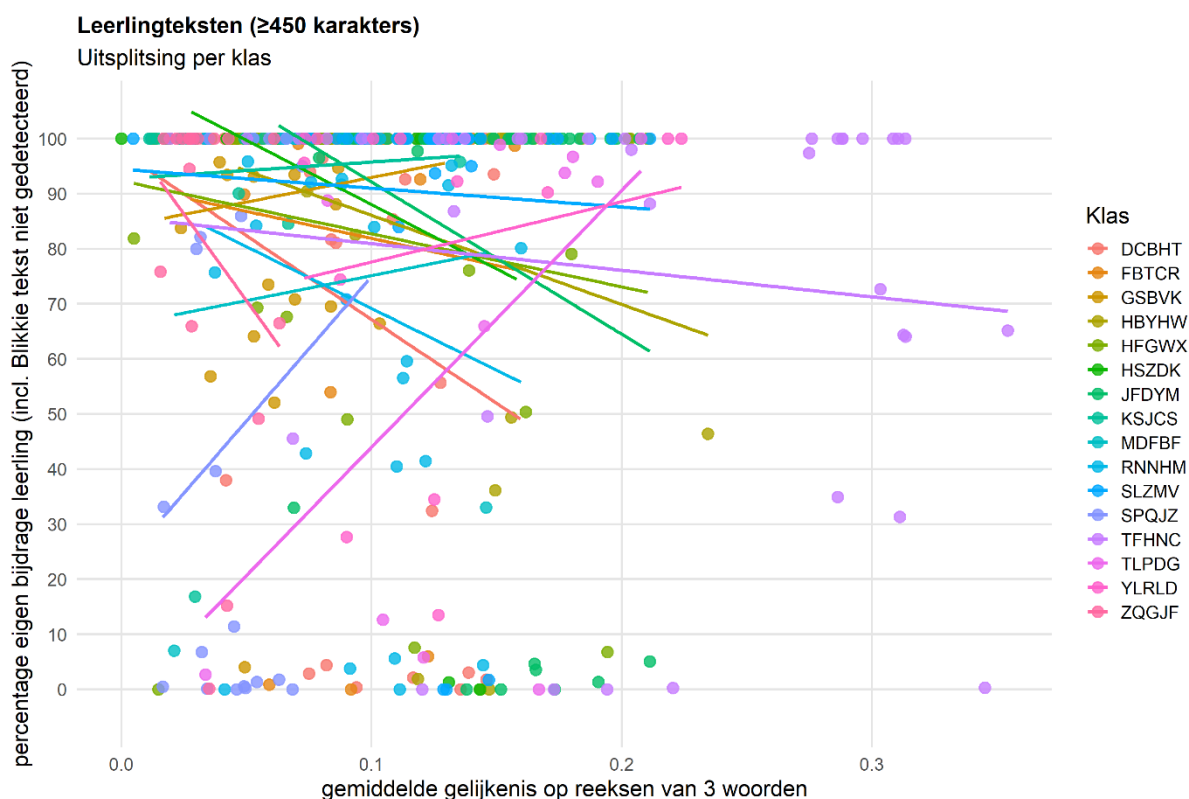
dan 15 karakters) overgenomen in hun uiteindelijke tekst. Dit betekent niet dat zij niks hebben geplakt, maar wel dat overgenomen tekst in de uiteindelijke tekst niet meer te herleiden was naar tekst gegenereerd door Blikkie. In deze klas is tekst geplakt vanuit externe bronnen wel in de uiteindelijke teksten terecht gekomen ( $M = 9,94$ ,  $sd = 30,59$ ).

### Aantal acties en interacties

Leerlingen hadden gemiddeld 39,46 actieblokken ( $sd = 23,94$ ); ook hier is de spreiding groot tussen leerlingen met een variatie tussen de 5 en 143 actieblokken. De klas met het hoogste gemiddelde aantal acties had gemiddeld 73,13 actieblokken ( $sd = 27,02$ ,  $min = 10$ ,  $max = 139$ ); de klas met het laagste gemiddelde had gemiddeld 20,85 actieblokken ( $sd = 9,38$ ,  $min = 6$ ,  $max = 40$ ). Dit betekent niet dat de ene groep leerlingen langer in de SchrijfBlik-omgeving heeft gewerkt, alleen dat de totale tijd in minder verschillende actieblokken is opgedeeld. Leerlingen waren gemiddeld 0,66 momenten ( $sd = 0,97$ ) voor langere tijd (60 seconden) weg uit de SchrijfBlik omgeving tussen het starten en inleveren van de schrijfpdracht. De leerlingen hadden gemiddeld 11,15 interacties met de chatbot Blikkie ( $sd = 8,61$ ); de reflectievragen vallen hier buiten. De klas met gemiddeld de meeste interacties met Blikkie had gemiddeld 23,93 interacties ( $sd = 11,30$ ,  $min = 2$ ,  $max = 51$ ). De klas met de minste interacties met Blikkie had gemiddeld 4,88 interacties ( $sd = 2,74$ ,  $min = 1$ ,  $max = 11$ ).

### Tekstgelijkenissen en -complexiteit

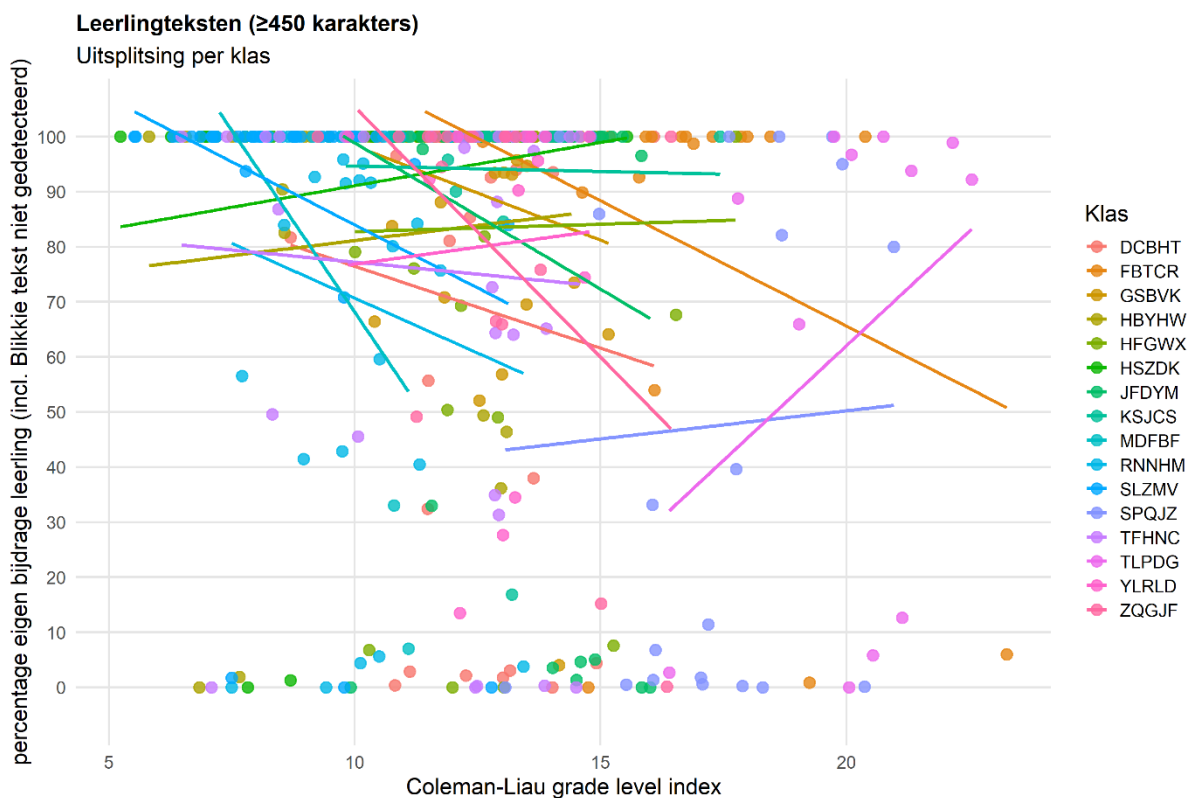
**Figuur 5** Percentage eigen bijdrage leerling afgezet tegen de gemiddelde gelijkenis op reeksen van 3 woorden (geanalyseerd per klas)



Figuur 5 geeft een visuele weergave van de relatie tussen het percentage eigen bijdrage van de leerling aan de uiteindelijk ingeleverde tekst en de gelijkenis van de tekst van de leerling met

andere teksten in dezelfde klas. Op basis van de figuur lijkt er geen eenduidige relatie te zijn tussen het percentage eigen bijdrage van de leerling en de gelijkenis met andere teksten. Voor de meeste klassen lijkt een kleinere bijdrage van de leerling zelf te leiden tot een hogere gelijkenis met andere teksten in de klas, maar bij vier klassen lijkt de relatie andersom te zijn. Vooral de klas TFHNC (zie Figuur 5) laat een hoge gelijkenis zien tussen de verschillende teksten, ook tussen teksten waarvan leerlingen een hoge eigen bijdrage hebben geleverd. Leerlingen in deze klas kregen de opdracht een specifieke klacht te schrijven over een specifiek product. Het is dus niet ondenkbaar dat ook teksten die volledig door de leerling zelf zijn geschreven veel van dezelfde woorden bevatten.

**Figuur 6** Percentage eigen bijdrage leerling afgezet tegen de *Coleman-Liau grade level* index (geanalyseerd per klas)



Figuur 6 geeft een visuele weergave van de tekstcomplexiteit afgezet tegen de eigen bijdrage van de leerling. Uit de grafiek is te zien dat teksten die (bijna) volledig zijn geschreven door Blikkie (lage eigen bijdrage) brede verschillen hebben in tekstcomplexiteit, net als de teksten die volledig door de leerlingen zijn geschreven. Ook hier is er per klas een verschil in de relatie die deze twee variabelen laten zien. Voor sommige klassen (zoals RNNHM en FBTCR) is de tekstcomplexiteit hoger naarmate de eigen bijdrage van de leerling lager is. In andere klassen is de tekstcomplexiteit (zoals TLPDG) juist hoger als leerlingen meer van de tekst zelf hebben geschreven. Voor het grootste deel van de klassen is er echter geen (significante) relatie tussen percentage eigen bijdrage en tekstcomplexiteit.

### Gebruik SchrijfBlik - logs

#### Interacties met Blikkie

De logdata laten zien dat leerlingen Blikkie op verschillende manieren gebruikten tijdens het schrijffproces (zie Tabel 6 en 7). De meest voorkomende interactie bestond uit kopiëren en

plakken. Daarnaast werden vertaalacties veel gebruikt, met name het vertalen van individuele woorden. Ook gebruikten leerlingen Blikkie regelmatig voor het stellen van inhoudelijke, taalkundige en technische hulpvragen. Bovendien werden interacties geregistreerd waarbij leerlingen vroegen om herformulering van tekst of om inspiratie voor de inhoud van hun opdracht.

Tabel 6 laat zien dat het vertalen van enkele woorden veruit de meest gebruikte functie was. Gemiddeld gebruikten leerlingen die dit type interactie minimaal één keer gebruikten deze 11,2 keer. Het vertalen van volledige zinnen of meerdere zinnen tegelijk kwam duidelijk minder vaak voor. De relatief hoge standaarddeviatie bij woordvertalingen wijst bovendien op grote verschillen tussen de leerlingen in de intensiteit van het gebruik.

Ook hulpvragen vormden een belangrijk deel van de interacties. Inhoudelijke hulpvragen kwamen het vaakst voor wanneer alle leerlingen worden meegeteld, terwijl taalkundige hulpvragen relatief intensief werden gebruikt door de leerlingen die deze functie benutten. Technische hulpvragen kwamen beduidend minder vaak voor. Verder blijkt dat leerlingen Blikkie regelmatig inzetten voor het herformuleren of aanpassen van tekst. Verzoeken om inspiratie of ideeën voor de inhoud van de tekst kwamen daarentegen slechts beperkt voor.

De kwalitatieve analyse van de interacties laat zien dat leerlingen Blikkie vooral gebruikten als hulpmiddel binnen het schrijfproces. De specifieke vormen van interactie en voorbeelden hiervan zijn opgenomen in Bijlage A.

**Tabel 6**

*Aantal interacties, gemiddelde per leerling en standaarddeviatie voor verschillende manieren van gebruik van Blikkie.*

Type input	Aantal	Gemiddelde per leerling totaal (M)	Gemiddelde per leerling die actie daadwerkelijk gebruikte	Standaarddeviatie (SD)
Vragen om vertaling van enkele woorden	1305	3,1	11,2	5,7
Vragen om vertaling van één zin	503	1,2	7,4	3,0
Vragen om vertaling van meerdere zinnen	175	0,4	3,4	1,2
Inhoudelijke hulpvraag	495	1,0	2,1	1,4
Taalkundige hulpvraag	388	0,8	4,6	1,9
Technische hulpvraag	100	0,2	2,8	0,7
Herformulering of verbetering	191	0,4	3,5	1,2
Inspiratie	44	0,09	2,4	0,4

Daarnaast zetten leerlingen vaak de strategie in om te kopiëren en plakken vanuit Blikkie. In tegenstelling tot het vertalen, worden bij kopiëren en plakken vooral meerdere zinnen (tekstblokken) gekopieerd. Het kopiëren en plakken van één zin of enkele woorden komt veel minder voor (zie Tabel 7). Leerlingen die van deze functie minimaal één keer gebruikmaakten, kopieerden gemiddeld 3,7 keer meerdere zinnen vanuit Blikkie.

Het gebruik van externe bronnen kwam aanzienlijk minder vaak voor dan het kopiëren vanuit Blikkie. Ook hierbij werden vooral meerdere zinnen gekopieerd, maar de gemiddelde aantallen interacties per leerling lagen duidelijk lager. Daarnaast wijzen de relatief lage standaarddeviaties erop dat verschillen tussen leerlingen in het gebruik van externe kopieeracties beperkter waren.

De kwalitatieve analyse en voorbeeld van deze interacties zijn ook opgenomen in Bijlage B.

**Tabel 7**

*Aantal interacties, gemiddelde per leerling en standaarddeviatie voor verschillende typen kopieeracties vanuit Blikkie en externe bron(nen).*

Type input	Aantal	Gemiddelde per leerling totaal (M)	Gemiddelde per leerling die actie daadwerkelijk gebruikte	Standaarddeviatie (SD)
Kopiëren van enkele woorden vanuit Blikkie	56	0,1	1,8	1,4
Kopiëren van één zin vanuit Blikkie	148	0,3	2,1	1,5
Kopiëren van meerdere zinnen vanuit Blikkie	310	0,6	3,7	1,3
Kopiëren van enkele woorden vanuit andere bron	42	0,09	1,4	0,5
Kopiëren van één zin vanuit andere bron	42	0,09	1,2	0,5
Kopiëren van meerdere zinnen vanuit andere bron	94	0,2	1,5	1,0

Tijdens het schrijfproces stelt Blikkie tussentijds vragen of opmerkingen aan de leerling. De mate waarin leerlingen hierop reageren verschilt sterk. Zoals weergegeven in Tabel 8 werd in 1439 gevallen geen reactie gegeven op een vraag of opmerking van Blikkie.

Wanneer leerlingen wel reageerden, gebeurde dit meestal met een kort of onvolledig antwoord. In totaal kwamen 709 van dergelijke reacties voor. Leerlingen die dit type interactie gebruikten, deden dit gemiddeld 2,3 keer per leerling. De toon en inhoud van deze reacties varieerden sterk. Sommige reacties waren zeer kort, zoals “ja!”, “nee” of “ik kijk wel”, terwijl andere iets uitgebreider maar nog steeds beperkt waren, zoals “Ik denk dat de tekst wel oké zou zijn”. In enkele gevallen reageerden leerlingen ook afwijzend, bijvoorbeeld met opmerkingen als “IK WEET HET ZELF” of “maak jij maar”.

Uitgebreide en inhoudelijke reacties op vragen van Blikkie kwamen duidelijk minder vaak voor. In totaal werden 69 van dergelijke reacties geregistreerd. Bij de leerlingen die wel uitgebreider reageerden, gebeurde dit gemiddeld 1,2 keer per leerling. Deze reacties kwamen bijvoorbeeld voor bij reflectievragen over de tekst, waarbij leerlingen opmerkingen gaven zoals: “Leuke opbouw, pakkende titel, aanspreekbaar onderwerp en interessante woordspeling/keuzes” of “De tekst komt denk ik overzichtelijk over. Vooral de conclusie op het einde.”

**Tabel 8**

*Aantal interacties, gemiddelde per leerling en standaarddeviatie van type reacties op vragen of opmerkingen van Blikkie.*

Type input	Aantal	Gemiddelde per leerling totaal (M)	Gemiddelde per leerling waarbij interactie daadwerkelijk plaatsvond	Standaarddeviatie (SD)
Geen response	1439	3,0	3,4	2,2
Inhoudelijk, maar kort of onvolledig antwoord op de vraag van Blikkie	709	1,5	2,3	1,7
Inhoudelijk, uitgebreid antwoord op de vraag van Blikkie	69	0,1	1,2	0,5

### 4.3 Ervaringen leerlingen

#### *Onderzoeksvraag 3: Hoe ervaren leerlingen het werken in SchrijfBlik en het gebruik van procesdata voor hun beoordeling?*

Om inzicht te krijgen in de ervaringen en opvattingen van leerlingen over het gebruik van AI bij schrijfopdrachten en het prototype SchrijfBlik, is feedback op twee manieren verzameld: via klassensessies en via een vragenlijst.

In het volgende gedeelte worden de resultaten uit de klassensessies en de vragenlijst eerst afzonderlijk besproken. Op deze manier kan duidelijk worden gemaakt welke inzichten uit beide vormen van dataverzameling naar voren komen. Vervolgens wordt per onderwerp een algemeen beeld geschetst op basis van de gecombineerde resultaten.

#### **Algemeen AI-gebruik**

Om meer inzicht te krijgen in het algemene AI-gebruik van leerlingen, bevatte de vragenlijst enkele stellingen over hoe vaak leerlingen AI gebruiken, hoe zij de invloed ervan op hun schrijven ervaren en of zij behoefte hebben aan meer onderwijs over het gebruik van AI. Deze stellingen zijn ook besproken in de klassensessies, maar hiervan zijn de responses niet bijgehouden.

De meerderheid van de leerlingen (66,7%) maakt regelmatig gebruik van AI voor het maken van schoolwerk. In totaal gaven 45 leerlingen een nadere toelichting op deze stelling. Slechts 3 van hen gaven aan nooit AI te gebruiken voor aan school gerelateerde taken. Andere termen waarmee leerlingen hun AI-gebruik omschrijven zijn onder meer “*vaak*”, “*dagelijks*” en “*elke week*”. Daarnaast specificerden sommige leerlingen de context van hun gebruik, bijvoorbeeld wanneer zij iets niet begrijpen, bij het maken van huiswerk of wanneer zij weinig motivatie hebben om een opdracht te maken die wel afgerond moet worden.

Een groot deel van de leerlingen (70,2%) geeft aan beter te schrijven met behulp van AI. In totaal gaven 32 leerlingen een nadere toelichting op deze stelling. Van deze groep gaven 25 leerlingen een positieve toelichting. Zij stelden onder andere dat AI over meer kennis beschikt en betere grammatica hanteert dan zichzelf. Daarnaast werd meermaals genoemd dat leerlingen zelf nog veel fouten maken bij het schrijven van teksten, terwijl AI hun zinsconstructies verbetert en fouten voorkomt. 7 leerlingen zijn het hier niet mee eens. Zij geven aan dat schrijven in eigen woorden beter is en dat het gebruik van AI de creativiteit vermindert.

Ondanks dat een groot deel van de leerlingen regelmatig gebruikmaakt van AI, vond de meerderheid (64,9%) dat zij meer zouden moeten leren over het verantwoord en effectief inzetten van AI. In totaal gaven 27 leerlingen een nadere toelichting op wat zij hierover graag zouden willen leren. Zo werd onder andere genoemd: *“hoe je kunt controleren of AI-informatie klopt en welke bronnen betrouwbaarder zijn”, “hoe je vragen zo kunt formuleren dat AI je in de juiste richting helpt, met een meer ondersteunende rol”* en *“hoe je AI kunt gebruiken als hulpmiddel zonder dat dit ten koste gaat van de eigen leerprestaties”*. Daarentegen gaf 35,1% van de leerlingen aan geen behoefte te hebben aan verdere instructie over het gebruik van AI.

In grote lijn komen de resultaten van de vragenlijst overeen met wat er in de klassensessies aan bod kwam. Daarbij is er wel een verschil bij de laatste stelling. Waar in de vragenlijst het merendeel aangeeft dat zij meer zouden moeten leren over het gebruiken van AI, waren veel leerlingen in de klassensessies het oneens met deze stelling. Zij gaven veelal aan dat zij al veel wisten over AI en het gebruiken ervan, en dat zij bovendien veel meer wisten dan hun docenten.

### Beoordeling

Tijdens de klassensessies werd leerlingen gevraagd hoe zij het vinden dat bij de beoordeling van een opdracht niet alleen naar het eindresultaat wordt gekeken, maar ook naar het proces waarmee de opdracht tot stand is gekomen. De antwoorden laten zien dat leerlingen hier verschillend over denken. Een meerderheid van de leerlingen staat hier positief tegenover, terwijl een aanzienlijk deel ook kritisch is. Daarnaast geven sommige leerlingen een meer genuanceerd of contextafhankelijk antwoord.

De grootste aantal post-its (53,7%) is positief ingevuld over het meenemen van het proces in de beoordeling. Volgens deze leerlingen kan inzicht in het proces helpen om beter te begrijpen hoe een leerling tot een antwoord is gekomen. Ook zien zij voordelen in het ontvangen van feedback op hun werkwijze. Zo schrijft een leerling: *“Ja goed, want dan wordt er ook gekeken naar of je er en hoe je erover na hebt gedacht. En dan kun je beoordeeld worden op je proces, wat ook fijn kan zijn als het eindresultaat tegenvalt.”* Een andere leerling benadrukt het belang van feedback: *“Het is handig als de docent het proces kan zien en feedback kan geven aan leerlingen.”*

Tegelijkertijd geeft een groep leerlingen met 38 post-its (35,2%) aan negatief tegenover deze vorm van beoordelen te staan. In hun antwoorden komt naar voren dat zij vinden dat vooral het eindresultaat zou moeten meetellen. Sommige leerlingen geven aan dat er verschillende manieren zijn om tot een goed antwoord te komen en dat het proces daarom minder relevant is voor de beoordeling. Zo schrijft een leerling: *“Ik vind dat het eindantwoord het meeste moet meetellen, want er zijn meerdere manieren om een antwoord te kunnen bereiken.”* Anderen trekken een vergelijking met toekomstige werksituaties, waarin volgens hen vooral het eindproduct belangrijk is: *“Best wel dom, want later op werk maakt het ook niemand uit hoe je het doet, als het maar af is.”* Daarnaast noemen sommige leerlingen dat het zichtbaar maken van het proces hun

mogelijkheden om AI te gebruiken kan beperken. Een leerling schrijft bijvoorbeeld: *“Het is niet goed, want als je ooit AI wil gebruiken dan krijg je ineens een lager cijfer.”*

Tot slot geven 12 post-its (11,1%) een gemengd of contextafhankelijk antwoord weer. Deze leerlingen zien zowel voordelen als nadelen van het beoordelen van het proces. Zo schrijft een leerling: *“Ik ben er deels mee eens, als er een moment is dat je eindresultaat beter is dan de manier hoe je het hebt aangepakt, dat dat er toch toe doet.”* Een andere leerling merkt op dat het inzicht in het proces nuttig kan zijn, maar tegelijkertijd ook als controlerend kan voelen: *“Het is goed, maar ook vervelend, omdat je gecontroleerd wordt op wat je doet en hierdoor kan je AI niet meer voor alles gebruiken.”*

De resultaten uit de vragenlijst laten een vergelijkbaar beeld zien. Op de vraag wat leerlingen ervan vinden wanneer naast het eindresultaat ook het proces wordt meegenomen bij de beoordeling van een opdracht, liepen de antwoorden eveneens uiteen. Een meerderheid van de leerlingen (66,6%) gaf aan dit een goede ontwikkeling te vinden. Sommige leerlingen plaatsen daarbij wel een kanttekening: zij geven aan dat zij dit persoonlijk niet altijd prettig vinden, maar het wel noodzakelijk achten. Zo schrijft een leerling: *“Als je dan toch al AI mag gebruiken vind ik dat goed, anders kan je hem gewoon de hele opdracht laten schrijven en is er geen eigen inbreng meer.”* Een andere leerling merkt op: *“Soms goed maar soms heel vervelend.”*

Tegelijkertijd gaf 22,8% van de leerlingen aan het hier niet mee eens te zijn. Deze leerlingen vinden dat de beoordeling zich vooral moet richten op het eindresultaat en niet op de manier waarop dit tot stand is gekomen. Daarnaast merkten enkele leerlingen op dat in hun latere leven waarschijnlijk ook niet gecontroleerd zal worden hoe bijvoorbeeld een opdracht, mail of ander product tot stand is gekomen. 10,5% van de leerlingen heeft deze vraag niet beantwoord.

Gezamenlijk laten de resultaten uit zowel de klassensessies als de vragenlijst zien dat een meerderheid van de leerlingen positief staat tegenover het meenemen van het proces in de beoordeling, maar dat er tegelijkertijd ook zorgen bestaan over de mate van controle en de rol die het eindresultaat daarbij nog speelt.

### **Gebruik SchrijfBlik**

Tijdens de klassensessies werd leerlingen gevraagd of zij hun werkwijze tijdens het schrijven hadden aangepast omdat de docent inzicht had in hun schrijfproces. De antwoorden laten zien dat een groot deel van de leerlingen aangeeft geen veranderingen te hebben aangebracht, terwijl een kleinere groep hun werkwijze wel heeft aangepast. Daarnaast was er een groep leerlingen die niet wist dat de docent kon meekijken.

Op een groot deel van de post-its (49,1%) wordt aangegeven dat de leerlingen niets anders te hebben gedaan tijdens het schrijven. Sommige leerlingen geven aan dat het feit dat de docent kon meekijken voor hen geen verschil maakte. Zo schrijft een leerling: *“Nee. Anders moeten ze het me niet laten gebruiken.”* Een andere leerling geeft aan dat het hen weinig uitmaakte: *“Nee, het boeide me niet zoveel.”*

Daarnaast is op 35 post-its (31,8%) geschreven dat leerlingen hun aanpak wel hebben aangepast omdat zij wisten dat de docent kon meekijken. Uit hun toelichtingen blijkt dat dit vooral te maken had met het beperken van hun gebruik van AI. Zo schrijft een leerling: *“Ja, want anders had ik hem meer laten schrijven.”* Een andere leerling koppelt dit aan de beoordeling van

de opdracht: *“Ja, omdat het meetelde voor je taalportfolio waar je wel een goed cijfer voor wil hebben en als de docent dan ziet dat je het niet zelf hebt gedaan krijg je een slecht cijfer.”*

Tot slot is op 19,1% van de post-its aangegeven dat het onbekend was dat de docent kon meekijken tijdens het schrijfproces. Hierdoor heeft deze mogelijkheid hun gedrag niet beïnvloed. Een leerling licht dit bijvoorbeeld toe: *“Niet, omdat ik niet wist dat ze mee kon kijken. Maar als ik het wist zou ik het niet veranderen.”*

De resultaten uit de vragenlijst laten een vergelijkbaar beeld zien. De meeste leerlingen (78,9%) geven aan geen andere aanpak te hebben gehanteerd tijdens het schrijven, ondanks het feit dat de docent inzicht had in hun werkwijze. In hun toelichting geven leerlingen vooral aan dat het hen weinig uitmaakte dat de docent kon meekijken, zoals blijkt uit uitspraken als *“omdat het me niets boeit”* en *“omdat het niet zo veel uitmaakt.”* Daarnaast geven sommige leerlingen aan dat zij niet wisten dat de docent inzicht had in hun werkwijze, bijvoorbeeld in reacties als *“ik had in het begin geen idee dat de docent mee kon kijken met wat we schreven”* en *“ik wist toen nog niet dat zij inzicht had.”*

Een kleinere groep leerlingen (21,1%) geeft aan bewust een andere werkwijze te hebben gehanteerd. Deze leerlingen noemen onder andere dat zij wilden laten zien dat zij een deel van de opdracht zelfstandig konden uitvoeren, of dat zij rekening hielden met het feit dat de docent precies kon zien wat zij typten en welke vragen zij aan AI stelden.

Gezamenlijk laten de resultaten zien dat het inzicht van de docent in het schrijfproces voor een deel van de leerlingen invloed heeft op hun werkwijze, maar dat een meerderheid aangeeft dat dit hun gedrag tijdens het schrijven niet heeft veranderd.

### SchrijfBlik en andere AI-tools

Tijdens de klassensessies werd leerlingen gevraagd hoe hun gebruik van SchrijfBlik verschilde van hun gebruik van andere AI-tools bij schrijfopdrachten. Op veel post-its (54,2%) maken leerlingen expliciet een vergelijking tussen SchrijfBlik en ChatGPT of een andere AI-tool. Een veelgenoemd verschil is dat leerlingen SchrijfBlik als persoonlijker en meer gespreksgericht ervaren. Leerlingen geven aan dat de interactie met het systeem meer aanvoelde als een dialoog. Ook wordt genoemd dat het systeem vriendelijker overkomt. Daarnaast valt het leerlingen op dat SchrijfBlik zelf vragen stelt, wat zij minder gewend zijn van andere AI-tools. Tegelijkertijd geven verschillende leerlingen aan dat het prototype minder goed aansluit bij de manier waarop zij AI normaal gebruiken voor schoolwerk. Zij ervaren praktische beperkingen en sommigen gaven aan dat zij Blikkie kinderachtiger of minder professioneel vonden.

Daarnaast laten enkele antwoorden zien dat SchrijfBlik het gebruiksgedrag van leerlingen kan veranderen. Op 26 post-its (19,8%) wordt beschreven dat zij het anders inzetten dan gebruikelijke AI-toepassingen. Waar leerlingen bij andere AI-tools vaak direct om een volledig antwoord vragen, wordt SchrijfBlik vaker gebruikt om ondersteuning te krijgen bij het zelf vinden van een antwoord. Op 30 post-its (22,9%) geven leerlingen aan dat hun gebruik van Blikkie vergelijkbaar is met dat van andere AI-tools.

De resultaten uit de vragenlijst laten vergelijkbare ervaringen zien. Gezamenlijk laten de resultaten uit zowel de klassensessies als de vragenlijst zien dat veel leerlingen duidelijke verschillen ervaren tussen SchrijfBlik en andere AI-tools, waarbij vooral de meer gespreksgerichte aanpak van SchrijfBlik wordt genoemd en door een deel van de leerlingen als

positief wordt ervaren. Tegelijkertijd benoemen leerlingen verschillende beperkingen in het gebruik ervan.

### **Gebruik AI**

Tijdens de klassensessies werd leerlingen gevraagd bij welke soorten taken zij vinden dat AI niet gebruikt zou mogen worden binnen schoolwerk. Uit de antwoorden blijkt dat veel leerlingen vinden dat er bepaalde situaties zijn waarin het gebruik van AI minder passend is, al verschillen de redenen en grenzen die zij daarbij stellen.

Op het grootste aandeel post-its (41,4%) staat dat AI niet gebruikt zou mogen worden bij toetsen en examens. Volgens deze leerlingen moet een toets laten zien wat een leerling zelf weet en kan. Zij vinden dat het gebruik van AI tijdens toetsmomenten het leerproces kan ondermijnen. Daarnaast staat op 25 post-its (21,6%) geschreven dat AI niet gebruikt zou moeten worden voor het volledig laten maken van opdrachten of verslagen. Vooral het genereren en kopiëren van complete teksten wordt hierbij als problematisch gezien. Een kleiner aantal post-its (12,9%) koppelt beperkingen aan specifieke vakken of opleidingen. Hierbij wordt genoemd dat AI minder passend is in vakgebieden waarbij kennis, verantwoordelijkheid en creativiteit een belangrijke rol spelen. Tegelijkertijd geven sommige leerlingen aan dat AI in principe overall gebruikt zou mogen worden. In totaal staat op 12 post-its (10,3%) dat er geen duidelijke beperkingen voor het gebruik van AI binnen schoolwerk zouden moeten zijn. Tot slot staan op 16 post-its (13,8%) andere specifieke situaties beschreven waarin AI volgens de leerlingen minder geschikt is. Dit betreft bijvoorbeeld opdrachten waarbij eigen inzicht of schrijfvaardigheid centraal staat.

De resultaten uit de vragenlijst laten een vergelijkbaar beeld zien. Gezamenlijk laten de resultaten zien dat veel leerlingen het gebruik van AI binnen schoolwerk accepteren, maar daarbij wel grenzen stellen, met name bij taken waarbij het belangrijk wordt gevonden dat zichtbaar is wat een leerling zelf weet of kan.

### **Inzicht in AI gebruik met SchrijfBlik**

Tijdens de klassensessies en in de vragenlijst werd leerlingen gevraagd of zij denken dat SchrijfBlik docenten helpt om beter te begrijpen hoe leerlingen AI gebruiken bij het maken van een opdracht. De antwoorden laten zien dat de meningen hierover verdeeld zijn.

De helft van de leerlingen geeft aan dat zij denken dat SchrijfBlik wel kan helpen om docenten meer inzicht te geven in het AI-gebruik van leerlingen. Zo kan een docent via het systeem kan welke vragen leerlingen stellen en hoe zij AI inzetten tijdens het maken van een opdracht. Volgens hen kan een docent daardoor beter inzicht krijgen in het proces en de manier waarop AI wordt gebruikt om tot een antwoord te komen.

De andere helft van de leerlingen geeft aan dat het prototype docenten niet helpt om beter te begrijpen hoe zij AI gebruiken. Sommige leerlingen geven aan dat docenten volgens hen al voldoende kennis hebben van hoe AI werkt, waardoor het systeem weinig extra inzicht zou bieden. Andere leerlingen geven aan dat het gebruik van AI niet altijd iets zegt over de leerling zelf of over hoe hij of zij leert.

Gezamenlijk laten de resultaten zien dat leerlingen verschillend denken over de mate waarin een systeem als SchrijfBlik docenten helpt om inzicht te krijgen in het AI-gebruik van leerlingen. Terwijl sommige leerlingen verwachten dat het systeem meer transparantie biedt in het gebruik van AI, twijfelen anderen of dit daadwerkelijk een realistisch beeld geeft van hoe zij AI normaal inzetten.

## 4.4 Beoordeling docenten

### **Onderzoeksvraag 4: Heeft het gebruik van SchrijfBlik door leerlingen invloed op de beoordeling van de schrijfvaardigheid door docenten?**

De docenten kregen na de afname een vragenlijst toegestuurd, met als doel inzicht te krijgen in hoe zij de SchrijfBlik omgeving hadden gebruikt in de beoordeling van de opdrachten en hoe zij het werken met SchrijfBlik hadden ervaren.

Over het algemeen vonden de docenten SchrijfBlik nuttig om inzicht te verkrijgen in het schrijfproces van hun leerlingen. Zo gaven 8 van hen aan dat SchrijfBlik an sich inzichtelijk is. Daarnaast gaven 9 docenten aan dat het goede inkijk biedt in het schrijfproces van leerlingen, en 10 docenten dat het inzicht geeft in hun AI-gebruik.

Ondanks deze positieve beoordeling gebruikten slechts 5 docenten de informatie uit SchrijfBlik 'een beetje' of 'aardig veel' bij de beoordeling van de schrijfopdracht. De docenten gaven verschillende redenen waarom zij de informatie uit SchrijfBlik wel of niet hebben gebruikt bij de beoordeling. Vijf docenten hebben de opdracht überhaupt niet beoordeeld, zo meldde een docent: *“Niet gebruikt om te beoordelen. Leerlingen knipten veel van Blikkie. Ze gaven aan dat feedback niet nodig was, omdat ze grotendeels alles hadden geknipt en geplakt.”* Een andere docent gaf aan: *“Ik heb het niet beoordeeld, omdat ik merkte dat veel leerlingen door de onbekendheid met het programma het niet zo serieus hebben gemaakt zoals ik wilde.”*

Drie docenten hebben de schrijfopdracht op een formatieve manier beoordeeld. Zij hebben de informatie uit SchrijfBlik gebruikt om inzicht te verkrijgen in het schrijfproces van leerlingen, om daar vervolgens feedback op te geven. Bijvoorbeeld: *“Ik kijk of de leerlingen zelf tot taalproductie zijn gekomen en niet de gehele brief door Blikkie hebben laten schrijven.”* en *“Niet zozeer voor de beoordeling zelf, maar meer om inzicht te verkrijgen in welke kennis de leerlingen blijkbaar toch nog niet paraat hebben.”*

Slechts één docent heeft een cijfer gegeven voor de schrijfopdracht. Hierbij heeft de docent specifiek gekeken naar wat de leerling door Blikkie heeft laten uitvoeren: *“Hierdoor kon ik zien wat de leerling zelf heeft gedaan en wat hij door AI heeft laten doen. Als er een goede tekst uitgekomen is, is dat prima, maar de leerlingen die dit zelf gedaan hebben met ondersteuning van AI krijgen een hogere waardering dan leerlingen die AI hebben gevraagd iets te schrijven en dit overgenomen hebben.”*

## 4.5 Waarde voor beoordeling en feedback

### **Onderzoeksvraag 5: Welke elementen van het schrijfproces waar SchrijfBlik inzicht in geeft leveren waardevolle informatie voor de beoordeling en feedback en gebruiken docenten? En wat zouden ze verder nog willen zien?**

#### **Functionaliteiten van SchrijfBlik en Blikkie**

De docenten gaven aan dat de plakdetectie de meest waardevolle functie in SchrijfBlik is. Meer dan de helft beoordeelde deze functie als “redelijk waardevol” of “heel waardevol”. Deze functie gaf de docenten inzicht in het werkgedrag van hun leerlingen: *“Sommige leerlingen maken te veel gebruik van de “gemakkelijke” weg (lees: laat Blikkie met directe antwoorden komen)”* en *“Ze knippen en plakken alles helaas.”*

De inhoudelijke vragen die Blikkie stelde aan de leerlingen werden daarentegen als minder nuttig ervaren. 8 docenten gaven aan deze functie als “niet waardevol” of “enigszins waardevol” te zien. Docenten merkten dat leerlingen onvoldoende gestimuleerd werden om de vragen van Blikkie te beantwoorden: *“Ik zie dat ze niet kritisch genoeg zijn in hun vragen en de reacties van Blikkie negeren”, “Veel leerlingen negeerden de vragen van Blikkie.”* en *“Mijn leerlingen werden gek van Blikkie, omdat hij ‘zich constant bemoeide’ met het werk van de leerlingen. Hij bleef input leveren, ook als ze hier niet om vroegen.”*

Interessant genoeg verschilt dit van hoe docenten de vragen die leerlingen zelf aan Blikkie stelden beoordeelden. Deze werden over het algemeen als waardevol ervaren, van “enigszins waardevol” tot “heel waardevol”. Ook de verwerking van de antwoorden van Blikkie in de teksten kreeg positieve beoordelingen, vrijwel alle docenten vonden dit van “enigszins waardevol” tot “heel waardevol”. Aanpassingen die leerlingen tijdens de schrijfopdracht zelf in de tekst aanbrachten, werden daarentegen minder nuttig gevonden. Meer dan de helft van de docenten beoordeelde deze als “niet waardevol” of “enigszins waardevol”, wat volgens de docenten vooral kwam doordat leerlingen veel kopieerden en plakten en weinig daadwerkelijk aanpasten.

#### **Beoordeling inzet SchrijfBlik in het onderwijs**

Ondanks dat er meerdere positieve kanten benoemd zijn van SchrijfBlik, gaven 8 van de docenten aan dat het onwaarschijnlijk is dat zij SchrijfBlik opnieuw zouden inzetten bij het afnemen van een schrijfopdracht. Slechts 3 docenten beoordeelden de waarschijnlijkheid op hergebruik met een 6 of hoger. Zo benoemt een docent dat leerlingen het gebruik van SchrijfBlik zagen als toestemming om AI in te zetten: *“Het lokte eigenlijk juist nog meer uit om AI te gebruiken. [...] Voor mijn gevoel was het eigenlijk een ‘loze’ les, ook al vind ik het concept wel heel leuk.. Het is niet zo goed uit de verf gekomen hier. Sorry!”,* en vervolgt met: *“Dat komt dus doordat leerlingen het als ‘vrijbrief’ zagen om alleen nog maar AI te gebruiken. AI zit natuurlijk in het programma zelf, dus dat nodigt uit. Ze weten dan zelf niet goed welke info ze moeten vragen aan Blikkie naast ‘schrijf deze brief’. Het in eigen woorden zetten is ook geen succes helaas.”*

Daarnaast benoemden docenten verschillende zorgen rondom de beoordeling: *“Door het gebruik van Blikkie lijkt het me veel moeilijker om objectief te kunnen beoordelen hoe leerlingen schrijven.”* en *“Na een eerste blik te hebben geworpen op de producten van enkele leerlingen vrees ik dat de gebruikelijke criteria bij het beoordelen van een schrijfvaardigheidstoets door de toepassing van*

*Blikkie nauwelijks van toepassing kunnen zijn: de gemiddelde leerling genereert zelf vrijwel niets aan schrijfproduct. Wel laat hij/zij zien hoe hij/zij met hulp van tekstgenerator Blikkie tot een tekst kan komen. Dat lijkt me eerder een andere vaardigheid dan wat men doorgaans onder schrijfvaardigheid verstaat.”*

### **Suggesties voor verbetering**

Docenten deden meerdere suggesties voor verbeteringen. Suggesties die bij drie docenten naar voren kwamen, hadden betrekking op het gebruik van Blikkie. Zo werd aangegeven: *“Blikkie moet vooral minder zeggen. Als lln. de prompt: 'wil je alleen wat zeggen als ik wat vraag' gaven, bleef hij zich 'bemoeien' met de lln, die 'gewoon aan het werk wilde'.”* en *“Blikkie hoeft niet altijd met tekst te komen en zou "overbodige" vragen niet moeten beantwoorden! (bijv. "Wat wordt zondag de uitslag bij Ajax en PSV?")”*. Daarnaast benoemden twee docenten dat ze graag in één oogopslag zouden willen zien wat leerlingen hebben geplakt, bijvoorbeeld door dit met kleur te markeren. Ook werden andere wensen genoemd, zoals het kunnen exporteren van de teksten en de optie om mee te kunnen kijken met de leerlingen tijdens het maken van de schrijfofdracht.

## 5. Discussie en Conclusie

In dit onderzoek is het SchrijfBlik prototype getest en geëvalueerd bij verschillende scholen in het voortgezet onderwijs. De deelnemende docenten gaven les in de vakken Nederlands, Engels, Frans, Duits en Spaans. Docenten mochten voor hun klas zelf een opdracht ontwerpen om te gebruiken in het onderzoek. In totaal namen 24 klassen deel (vmbo basis, vmbo kader, vmbo tl, havo en vwo, in leerjaren 4, 5 en 6), met gezamenlijk 480 leerlingen. In dit onderzoek werden verschillende onderzoeksvragen beantwoord, waarbij de ervaringen van het werken met SchrijfBlik van zowel leerlingen als docenten centraal stonden.

Als eerste werden in het onderzoek de competentie van docenten in AI en hun gevoelens over AI in kaart gebracht. Gevoelens over AI lopen uiteen: sommige docenten zien zowel kansen als risico's, anderen zijn enthousiast en nieuwsgierig, terwijl weer anderen onzekerheid, frustratie of wantrouwen uiten. Docenten hebben redelijk vertrouwen in basisvaardigheden en praktische omgang met AI, maar wisselende didactische en technische kennis. Ze voelen zich relatief bekwaam om de pedagogische meerwaarde van AI te beoordelen, maar minder om AI concreet en datagedreven in leerprocessen toe te passen. Docenten hebben op dit moment weinig kennis over ethische aspecten zoals bias, transparantie en verantwoording. De scores van de docenten waren vergelijkbaar met de scores uit het onderzoek van Celik (2023) voor de basisvaardigheden en technisch/pedagogische kennis. Met name voor ethiek waren de scores van de docenten in dit onderzoek lager, gemiddeld 1 punt op de 5-puntsschaal.

De integratie van AI in het onderwijs lijkt voor veel docenten nog in ontwikkeling te zijn. Dit heeft implicaties voor het gebruik van een tool als SchrijfBlik. Wanneer docenten zich nog onvoldoende zeker voelen in hun kennis en vaardigheden rondom AI, zou dit het ontwerpen, begeleiden en beoordelen van schrijfopdrachten waarbij AI-gebruik is toegestaan kunnen bemoeilijken. Zo kan het lastig zijn om onderscheid te maken tussen zinvol en minder zinvol AI-gebruik, of om het leerproces van leerlingen adequaat te monitoren en te sturen. Dit wijst erop dat succesvolle implementatie van dergelijke tools mogelijk niet alleen afhankelijk is van de technologie zelf, maar ook van de mate waarin docenten ondersteund worden in het ontwikkelen van AI-gerelateerde kennis en vaardigheden (AI-geletterdheid).

In het gebruiken van de procesdata voor het evalueren van de schrijfopdrachten zagen we dat docenten vooral de plakdetectie hebben bekeken. Een deel van de docenten heeft in meer detail gekeken naar welke vragen leerlingen hebben gesteld aan Blikkie, om inzicht te krijgen in waar leerlingen staan in het leerproces. Het merendeel van de docenten heeft de opdracht uiteindelijk niet beoordeeld. Verschillende docenten gaven aan dat de inzet van SchrijfBlik juist aanleiding leek te zijn voor leerlingen om meer AI te gebruiken, ook buiten de beoogde toepassing voor de opdracht zelf. Dit riep teleurstelling op bij de docenten en leidde tot vragen over hoe zij de uiteindelijke producten van leerlingen nog moeten beoordelen. Met name het gevoel dat leerlingen zelf weinig tot niets hebben toegevoegd aan hun schrijfopdracht speelt hierin een belangrijke rol.

Tegelijkertijd laten de leerlingreacties een ander beeld zien over de mate van AI-gebruik in SchrijfBlik. De meerderheid van de leerlingen geeft aan dat zij Blikkie niet wezenlijk anders hebben gebruikt dan andere AI-tools die zij normaal gesproken tot hun beschikking hebben. Vanuit hun perspectief is er dus geen sprake van een duidelijke toename of verandering in

gebruik, maar eerder van een voortzetting van bestaand gedrag, waar de docent nu inzicht in krijgt.

Deze discrepantie tussen de percepties van docenten en leerlingen roept vragen op over verwachtingen, transparantie en de rol van AI in het schrijfonderwijs. Het lijkt erop dat niet alleen het gebruik van SchrijfBlik zelf, maar vooral de interpretatie ervan door docenten en leerlingen uiteenloopt, wat vraagt om verdere reflectie op hoe AI-gebruik bespreekbaar en begeleid kan worden in de klas.

De analyse van de ingeleverde schrijfoopdrachten en het proces laten zien dat er veel onderlinge verschillen zijn, zowel tussen klassen als tussen leerlingen. Deze verschillen waren te verwachten, aangezien docenten hun eigen schrijfoopdracht mochten vormgeven. Deze verschillen waren in die zin ook toe te schrijven aan bepaalde kenmerken van de opdracht en docent, al is dit niet systematisch onderzocht. Het is bijvoorbeeld interessant dat er in bij één klas geen tekst afkomstig van Blikkie in de uiteindelijke teksten terugkwam. Mogelijk heeft de instructie van de docent hierin een rol ingespeeld. Het is interessant om verder onderzoek te weiden aan de verschillende kenmerken van schrijfoopdrachten en startinstructies, die op hun beurt invloed kunnen hebben op hoe een schrijfomgeving als SchrijfBlik wordt gebruikt.

Uit de tekstanalyses voor tekstgelijkenissen en -complexiteit kwam naar voren dat de relatie tussen de eigen bijdrage van de leerling en deze variabelen verschillend was per klas. Dit kan veel verschillende oorzaken hebben, zoals de kenmerken van de schrijfoopdracht maar ook verschillen in AI-geletterdheid van leerlingen die beïnvloed hoe zij samen met Blikkie de schrijfoopdracht maken. Wat wel duidelijk uit de analyses naar voren komt is dat, op basis van deze twee tekstanalyses – de *trigram-based similarity index* en de *Coleman-Liau grade level index* – teksten die (grotendeels) met genAI zijn geschreven niet sterk te onderscheiden zijn van teksten die door de leerlingen zelf zijn geschreven. Dit onderschrijft de moeite die docenten in het algemeen hebben met het herkennen van AI-gegenereerde teksten.

Belangrijke vraagstukken die volgen uit dit onderzoek zijn: welke opdrachten zijn geschikt voor een omgeving zoals SchrijfBlik, aan welke voorwaarden moeten deze opdrachten voldoen, welke instructies zou een leerling hierbij moeten krijgen en op welke wijze kunnen deze beoordeeld worden? Docenten geven aan dat zij het lastig vinden om te bepalen welke opdrachten zich het best lenen voor inzet in een AI-ondersteunde leeromgeving en in hoeverre deze formatief dan wel summatief ingezet kunnen worden. Formatieve opdrachten lijken vooralsnog het meest geschikt, omdat deze ruimte bieden voor experimenteren, feedback en procesgericht leren, zonder directe consequenties voor de eindbeoordeling. Tegelijkertijd roept dit de vraag op in hoeverre summatieve toetsing aangepast moet worden wanneer leerlingen gebruikmaken van ondersteuning van Blikkie.

In dit onderzoek zijn enkel taalvakken meegenomen. Binnen taalvakken wordt Blikkie vooral gebruikt voor vertaal- en herschrijfoopdrachten. In andere vakken, zoals de exacte vakken en maatschappijvakken, zijn andere toepassingen waarschijnlijk passender, zoals het geven van uitleg, het ordenen van informatie en het helpen bij probleemoplossing. In een dergelijke context is het schrijfproduct niet het doel maar een middel, en staat het denkproces en kennisopbouw van leerlingen over een onderwerp juist centraal. Het is echter nog onduidelijk hoe dit er in de praktijk uitziet. Het is aannemelijk dat de geschiktheid van SchrijfBlik verschilt per vak en context, maar hier is op basis van dit onderzoek nog geen eenduidig beeld van.

Een ander knelpunt betreft de instructie van docent naar leerling. Effectieve inzet van SchrijfBlik lijkt te vragen om duidelijke richtlijnen over hoe Blikkie gebruikt kan worden, evenals over de mate van zelfstandigheid die van leerlingen verwacht wordt. Dit kan door docenten als uitdagend worden ervaren, met name wanneer de AI-geletterdheid nog beperkt is, zoals ook naar voren komt uit de resultaten van de TPACK-vragenlijst. Dit kan er mede aan hebben bijgedragen dat het voor leerlingen niet altijd duidelijk was hoe zij Blikkie konden inzetten, hoe zij effectieve prompts moesten formuleren of in hoeverre de docent inzicht had in het schrijfproces in SchrijfBlik. Daarnaast lijkt er nog weinig zicht te zijn op hoe docenten gerichte feedback kunnen geven op het AI-gebruik van leerlingen. Dit wijst erop dat zowel expliciete instructie als begeleiding en feedback op het gebruik van AI aandacht vragen in de onderwijspraktijk, zowel voor docent als leerling.

Docenten mochten aangeven welke aanpassingen zij zelf nog graag zouden zien in SchrijfBlik. Deze suggesties waren vooral gericht op het aanpassen van instellingen van Blikkie. Daarnaast zijn er docenten die de plakdetectie nog inzichtelijker zouden willen hebben. Leerlingen gaven aan dat zij met name de reflectievragen van Blikkie vervelend vonden. Zij vergeleken Blikkie vaak met ChatGPT, en vonden ChatGPT gemakkelijker werken. Blikkie was dusdanig ingesteld dat leerlingen niet op al hun vragen antwoord kregen, wanneer Blikkie het idee had dat het onderwerp van de vraag ongepast was. We zagen dat een aantal leerlingen probeerden om Blikkie dit soort vragen te stellen. Hierin is ook zichtbaar dat zowel docenten als leerlingen verdere instructies nodig hebben om te leren werken met AI-tools als Blikkie.

Deze studie nodigt uit tot reflectie op hoe leerlingen en docenten SchrijfBlik beschouwen vanuit het perspectief van het sociaal contract (UNESCO, 2021). Het AI-systeem mengt zich in het contract tussen leerling en docent, waarbij de leerling belooft eerlijk werk af te leveren en de docent dat werk vervolgens eerlijk beoordeelt. Leerlingen kunnen het gevoel krijgen dat ze worden gecontroleerd, doordat de docent het proces kan inzien (van Haastrecht et al., 2025). De meeste leerlingen gaven aan dat zij niet anders hebben gewerkt in SchrijfBlik omdat de docent inzicht had in het schrijfproces; voor veel leerlingen maakte dit geen verschil. Zij zagen dit kennelijk niet als een inbreuk op het sociaal contract. Ongeveer een derde van de leerlingen gaf aan dat zij erop hadden gelet dat zij niet te veel vragen stelden aan Blikkie, dat ze wilden laten zien aan de docent dat ze zoveel mogelijk zelf hadden geschreven. Deze groep voelde zich kennelijk wel gecontroleerd. Zo'n 20 procent van de leerlingen gaf aan dat ze zich er niet bewust van waren geweest dat de docent het proces kon inzien. Ondanks dat docenten aangaven dit te hebben verteld, blijkt dat niet alle leerlingen deze informatie hebben opgenomen.

Tot slot roept dit onderzoek de vraag op in hoeverre er nog sprake is van 'schrijfvaardigheid' wanneer AI-gebruik is toegestaan, zeker wanneer leerlingen de tool inzetten zoals in dit onderzoek werd waargenomen. Leerlingen maakten bijvoorbeeld gebruik van Blikkie voor het vertalen van teksten, het genereren van volledige zinnen en in sommige gevallen zelfs voor grotere delen van de opdracht. Hierbij kan je je afvragen in hoeverre de uiteindelijke tekst nog een representatie is van de individuele schrijfvaardigheid van de leerling.

Tegelijkertijd lijkt het te eenvoudig om te concluderen dat schrijfvaardigheid hiermee volledig waarde verliest. Mogelijk verschuift de aard van de vaardigheid: naast traditionele aspecten van schrijven, zoals formuleren en structureren, worden ook andere vaardigheden relevanter. Te denken valt aan digitale en AI-geletterdheid, zoals het effectief formuleren van prompts, het kritisch beoordelen van gegenereerde output en het doelgericht inzetten van AI als

ondersteuning bij het schrijfproces. In dat opzicht kan schrijven met AI ook worden gezien als een vorm van samenwerken met technologie. Daarbij lijkt het schrijfproces minder te draaien om enkel tekstproductie, en meer om het selecteren, evalueren en bewerken van (al dan niet door AI-gegenereerde) tekst, waarbij leerlingen actief de regie voeren over het eindproduct (Woo et al., 2025).

Dit roept bredere vragen op over wat precies gemeten wordt wanneer schrijfvaardigheid wordt beoordeeld in een context waarin AI-gebruik is toegestaan, en welke vaardigheden daarbij centraal zouden moeten staan. Verdere uitwerking en afbakening van deze concepten lijken noodzakelijk om tot passende beoordelingscriteria te komen.

Een belangrijke beperking van dit onderzoek is dat de toepasbaarheid van SchrijfBlik in de onderwijspraktijk nog niet volledig kan worden vastgesteld. Dit hangt samen met een aantal eerder benoemde, nog onbeantwoorde vraagstukken rondom de inrichting van opdrachten, instructie en beoordeling binnen een AI-ondersteunde leeromgeving. Zolang hierover geen duidelijke kaders bestaan, blijft onduidelijk hoe SchrijfBlik effectief kan worden ingezet. Dit vormt daarmee een belangrijke randvoorwaarde voor zowel de verdere ontwikkeling als de succesvolle implementatie van het systeem.

Daarnaast blijkt uit de resultaten dat leerlingen niet altijd serieus gebruik hebben gemaakt van Blikkie. Hiervoor zijn verschillende mogelijke verklaringen. Ten eerste kan de wijze waarop de opdracht in de klas is geïntroduceerd en begeleid van invloed zijn geweest op de mate van betrokkenheid van leerlingen. Ten tweede hebben docenten in dit onderzoek vrijwel geen gebruik gemaakt van een beoordeling met een cijfer, wat mogelijk de ervaren urgentie en inzet van leerlingen heeft verminderd. Tot slot gaf circa 20% van de leerlingen aan zich er niet van bewust te zijn dat hun schrijfproces in SchrijfBlik inzichtelijk was voor de docent, wat mogelijk heeft bijgedragen aan minder doelgericht gebruik.

Deze bevindingen en beperkingen benadrukken het belang van duidelijke didactische kaders voor de inzet van AI in het onderwijs. Vervolgonderzoek zou daarom moeten uitwijzen hoe opdrachten, instructie en beoordeling effectief kunnen worden ingericht binnen een AI-ondersteunde leeromgeving. Daarbij is het ook van belang om te onderzoeken welke rol schrijfvaardigheid en AI-geletterdheid spelen, en hoe deze vaardigheden zich tot elkaar verhouden binnen het leerproces.

## 6. Dankbetuiging

Wij willen allereerst de leerlingen, docenten en scholen bedanken voor hun deelname! Dit onderzoek is tot stand gekomen met de hulp van vele CitoLab collega's. Onze dank gaat uit naar Sander Looise voor het beheren van SchrijfBlik, zijn hulp bij het exporteren van data en zijn hulp bij vragen. Tijmen Poell was tijdens dit onderzoek en bij het schrijven van het rapport beschikbaar voor vragen, dank daarvoor! Franka Buytenhuijs willen we bedanken voor haar hulp bij het automatisch coderen.

## 7. Bronnen

- Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20. <https://doi.org/10.1186/s41239-023-00427-0>
- Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, 107468. <https://doi.org/10.1016/j.chb.2022.107468>
- Centraal Bureau voor de Statistiek. (2024). *Bijna kwart Nederlanders gebruikt kunstmatige intelligentie zoals ChatGPT*. <https://www.cbs.nl/nl-nl/nieuws/2024/36/bijna-kwart-nederlanders-gebruikt-kunstmatige-intelligentie-zoals-chatgpt>
- Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2), 283-284. <https://doi.org/10.1037/h0076540>
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement* 50, 1–73. <https://doi.org/10.1111/jedm.12000384>
- Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher* 18, 5–11. <https://doi.org/10.3102/0013189X018002005>
- Poell, T., Maas, L., Balk, M., & Van Haastrecht, M. (2025). SchrijfBlik: Safeguarding the validity of writing assessment in the age of AI. In *Artificial Intelligence in Education: Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium, Blue Sky, and WideAIED (AIED 2025)*. Springer. [https://doi.org/10.1007/978-3-031-99261-2\\_6](https://doi.org/10.1007/978-3-031-99261-2_6)
- UNESCO (2021). *Reimagining our futures together: a new social contract for education*. Tech. rep.,405 Educational and Cultural Organization of the United Nations, Paris, France
- Van Haastrecht, M., de Groot, L., Jongbloed-Pereboom, M, Buytenhuijs, F. & Kruis J. (2025). Artificial intelligence for educational measurement: Where is the value for education? *Frontiers in Education*, 10. <https://doi.org/10.3389/educ.2025.1677255>
- Woo, D. J., Yu, Y., & Guo, K. (2025). *Exploring EFL secondary students' AI-generated text editing while composition writing*. arXiv. <https://doi.org/10.48550/arXiv.2505.17041>

# 8. Bijlagen

## 8.1 Bijlage A.

### Vragenlijst achteraf voor docenten

Heeft u de schrijfpdracht in SchrijfBlik beoordeeld?

- Ja, om een cijfer te kunnen geven (summatief)
- Ja, om feedback te kunnen geven (formatief)
- Nee
- Anders, ....

In welke mate heeft u de extra informatie die SchrijfBlik biedt over het schrijfproces gebruikt in de beoordeling van de schrijfpdracht?

- Helemaal niet
- Een beetje
- Aardig veel
- Heel veel

Kunt u een korte omschrijving geven van de schrijfpdracht die u heeft afgenomen?

In hoeverre heeft u gebruik gemaakt van de volgende aspecten van het schrijfproces waar SchrijfBlik inzicht in geeft bij het beoordelen van de schrijfpdracht? (4 puntsschaal; niet gebruikt, enigszins gebruikt, redelijk gebruikt, veel gebruikt)

- Aanpassingen in de tekst gedurende de schrijfpdracht
- De vragen die gesteld zijn aan Blikkie
- De verwerking van de antwoorden van Blikkie in de tekst
- De antwoorden op de inhoudelijke vragen van Blikkie
- De plakdetectie van Blikkie

Kunt u vertellen hoe u de extra informatie uit SchrijfBlik heeft gebruikt tijdens het beoordelen van de schrijfpdracht?

In hoeverre vond u de volgende aspecten van het schrijfproces waar SchrijfBlik inzicht in geeft waardevol bij het beoordelen van de schrijfpdracht? (4 puntsschaal; niet waardevol, enigszins waardevol, redelijk waardevol, heel waardevol)

- Aanpassingen in de tekst gedurende de schrijfpdracht
- De vragen die gesteld zijn aan Blikkie
- De verwerking van de antwoorden van Blikkie in de tekst
- De antwoorden op de inhoudelijke vragen van Blikkie
- De plakdetectie van Blikkie

Heeft het gebruik van SchrijfBlik geleid tot een andere benadering van het beoordelen van schrijfpdrachten? Zo ja, wat was daar anders aan?

Geef aan in welke mate u het eens bent met de volgende stellingen (5 puntsschaal: helemaal oneens, oneens, niet eens/niet oneens, eens, helemaal eens)

- Het beoordelen met SchrijfBlik kostte mij meer tijd dan ik zou willen.
- Het beoordelen met SchrijfBlik kostte meer moeite dan ik zou willen.
- Het beoordelen met SchrijfBlik gaf mij een goede inkijk in het schrijfproces van leerlingen.
- Het beoordelen met SchrijfBlik gaf mij een goede inkijk in het AI-gebruik van leerlingen.

Denkt u dat door gebruik te maken van de extra informatie die SchrijfBlik biedt tijdens het nakijken, u een objectievere beoordeling van de schrijfvaardigheid van leerlingen kunt geven? En waarom denkt u dat?

Kunt u de onderstaande aspecten sorteren op volgorde van hoe belangrijk u dit vond in de beoordeling van de schrijfoopdrachten met SchrijfBlik? Sorteert ze van meest belangrijk naar minst belangrijk.

- Hoe vaak leerlingen gebruikmaakten van AI
- Hoe effectief leerlingen gebruikmaakten van AI
- De kwaliteit van het eindproduct

Kunt u door gebruik te maken van SchrijfBlik beter feedback geven aan leerlingen in hoe ze hun schrijfvaardigheid kunnen verbeteren? En waarom denkt u dat?

Waren er verrassende uitkomsten bij het gebruik van SchrijfBlik in uw klas? Zo ja, welke?

Geef aan in welke mate u het eens bent met de volgende stellingen (5 puntsschaal: helemaal oneens, oneens, niet eens/niet oneens, eens, helemaal eens)

- SchrijfBlik is inzichtelijk
- SchrijfBlik is te complex
- SchrijfBlik heeft mij geholpen om een eerlijke beoordeling te geven

Bent u obstakels tegengekomen in het gebruik van SchrijfBlik? Zo ja, welke waren dat?

Hoe waarschijnlijk is het dat u nogmaals een schrijfoopdracht zal afnemen met SchrijfBlik indien u daar de mogelijkheid voor heeft? (score van 0, zeer onwaarschijnlijk tot 10, zeer waarschijnlijk)

Wat kan er aan SchrijfBlik worden toegevoegd (of verwijderd) om de schrijfvaardigheid van leerlingen verder inzichtelijk te maken?

## 8.2 Bijlage B.

### Nadere uitwerking van leerlinginteracties met Blikkie

#### **Vertalingen**

De logdata laten zien dat leerlingen Blikkie vooral gebruikten om individuele woorden te vertalen. Dit type interactie kwam met 1.305 geregistreerde acties beduidend vaker voor dan het vertalen van één volledige zin (503 interacties) of meerdere zinnen tegelijk (175 interacties).

Binnen de vertaling van enkele woorden zijn verschillende strategieën zichtbaar. Sommige leerlingen verfijnen hun prompt stapsgewijs: ze starten met een paar woorden zoals “ *vorige week*”, voegen vervolgens een context toe (“ *vorige week in Frans*”) en eindigen met een goed geformuleerde vertaalopdracht (“ *vertaal ik hoor naar het Frans*”). Andere leerlingen formuleren direct een volledige en duidelijke prompt, zoals “ *Wat is ‘van lezen word je slimmer’ in het Frans?*” of “ *Wat is ‘zouden moeten’ in het Frans?*”.

Zoals weergegeven in Tabel 6 blijkt dat het gemiddelde aantal interacties per leerling voor het vertalen van enkele woorden 3,1 is wanneer alle leerlingen worden meegeteld (klassen Nederlands geëxcludeerd). Daarnaast is voor elk interactietype een gemiddeld aantal interacties berekend voor uitsluitend de leerlingen die het betreffende interactietype minimaal één keer gebruikten. Voor het vertalen van enkele woorden ligt dit gemiddelde op 11,2 interacties per leerling. De standaarddeviatie van 5,7 geeft aan dat er grote verschillen tussen leerlingen zijn in hoe vaak ze dit type interactie inzetten. Voor het vertalen van één zin of meerdere zinnen liggen de gemiddelden beduidend lager, zowel voor het totaal als voor de actieve gebruikers.

#### **Stellen van vragen**

Naast vertaal- en kopieeracties gebruikten leerlingen Blikkie ook om inhoudelijke, taalkundige en technische hulpvragen te stellen. In totaal werden 495 inhoudelijke, 388 taalkundige, en 100 technische hulpvragen geregistreerd (zie Tabel 6).

Inhoudelijke hulpvragen kwamen daarmee het vaakst voor. Wanneer alle leerlingen worden meegeteld, ligt het gemiddelde aantal inhoudelijke hulpvragen op 1,0 per leerling, tegenover 0,8 taalkundige en 0,2 technische hulpvragen per leerling. Wanneer alleen leerlingen worden meegenomen die het betreffende interactietype minimaal één keer gebruikten, verschuift dit beeld. Taalkundige hulpvragen komen dan gemiddeld 4,6 keer per leerling voor en technische hulpvragen 2,8 keer, terwijl inhoudelijke hulpvragen gemiddeld 2,1 keer per leerling voorkomen. Deze aantallen laten zien dat inhoudelijke hulpvragen door meer leerlingen incidenteel worden gesteld, terwijl taalkundige hulpvragen door een kleinere groep leerlingen intensiever worden gebruikt.

Inhoudelijke hulpvragen hadden betrekking op de inhoud en opbouw van de tekst. Leerlingen vroegen om argumenten om in hun tekst te gebruiken, zoals in vragen naar ondersteuning van hun standpunt. Daarnaast werd er regelmatig gevraagd om bronnen of bronvermelding, bijvoorbeeld: “ *gebruik hier minstens 3 kwalitatief goede bronnen*” en “ *kun je mij een goeie bron geven*”. Ook vroegen sommige leerlingen om inhoudelijke verduidelijking, bijvoorbeeld: “ *wat is een stelling-argumentenstructuur?*” of “ *is betoog in ik vorm*”.

Taalkundige hulpvragen hadden vaak betrekking op grammaticale of lexicale aspecten. Leerlingen vroegen bijvoorbeeld naar het geslacht van woorden, zoals in vragen als *“Welk geslacht is het woord Faktor in het Duits?”* of *“Is Zukunft mannelijk of vrouwelijk?”*. Ook vroegen leerlingen regelmatig om feedback op zinnen die zij zelf hadden geschreven, bijvoorbeeld: *“Is deze zin correct: Menschen sagen fach, dass die Zukunft seht wichtig ist für jonge Leute.”*

Technische hulpvragen hadden daarentegen vooral betrekking op de werking van het systeem of praktische aspecten van de opdracht. Voorbeelden hiervan zijn vragen als *“Hoeveel woorden heeft dit?”* en vragen over de functionaliteit van Blikkie zelf, zoals *“Hoe kan ik jou inschakelen als hulpmiddel?”* en *“Jij kan dingen voor mij vertalen?”*.

### **Herformuleren en inspiratie**

Leerlingen gebruikten Blikkie ook om zinnen te herformuleren of hun tekst te verbeteren. In totaal werden 191 interacties geregistreerd waarbij leerlingen om een herformulering of aanpassing van hun tekst vroegen (zie Tabel 6). Wanneer alle leerlingen worden meegeteld, komt dit neer op een gemiddelde van 0,4 interacties per leerling. Wanneer alleen de leerlingen worden meegenomen die deze actie minimaal één keer gebruikten, ligt het gemiddelde op 3,5 interacties per leerling, wat erop wijst dat leerlingen die deze functie gebruiken dit vaak meerdere keren tijdens het schrijfproces doen.

De verzoeken voor herformulering hadden vaak betrekking op het vereenvoudigen, inkorten, verlengen of aanpassen van de formulering van een tekst. Voorbeelden hiervan zijn opmerkingen als *“kan je wat makkelijkere woorden gebruiken”*, *“kort dit in tot 120 woorden”* of *“zonder gekke tekens”*. In sommige gevallen waren de reacties van leerlingen minder specifiek of moeilijker te interpreteren, bijvoorbeeld opmerkingen als *“nee Blikkie, ik ben niet blij”* of *“klinkt niet”*.

Daarnaast gebruikten leerlingen Blikkie soms als bron van inspiratie voor hun tekst, al kwam dit relatief weinig voor. In totaal werden 44 interacties geregistreerd waarbij leerlingen expliciet om inspiratie vroegen. Wanneer alle leerlingen worden meegeteld, ligt het gemiddelde op 0,09 interacties per leerling, en 2,4 interacties per leerling wanneer alleen de leerlingen worden meegenomen die deze actie minimaal één keer gebruikten. De standaarddeviatie van 0,4 wijst erop dat het gebruik van deze strategie relatief consistent maar beperkt is onder de leerlingen die deze functie benutten.

Inspiratievragen hadden vaak betrekking op ideeën of voorbeelden voor de inhoud van de tekst. Leerlingen vroegen bijvoorbeeld om voorbeeldzinnen, zoals in de vraag *“geef voorbeelden van zinnen die ik kan zeggen in het Duits hierover”*. Ook vroegen leerlingen soms om hulp bij het uitbreiden van hun tekst, bijvoorbeeld: *“wat kan ik nog meer in de tweede alinea zetten”* of *“kan je inspiratie geven wat ik kan doen met de opdracht”*.

### **Kopiëren en plakken vanuit Blikkie**

Daarnaast zetten leerlingen vaak de strategie in om te kopiëren en plakken vanuit Blikkie. In tegenstelling tot het vertalen, worden bij kopiëren en plakken vooral meerdere zinnen (tekstblokken) gekopieerd. Het kopiëren en plakken van één zin of enkele woorden komt veel minder voor (zie Tabel 7).

Het gemiddelde aantal interacties per leerling voor het kopiëren van meerdere zinnen is 0,6 wanneer alle leerlingen worden meegeteld, en 3,7 wanneer alleen de leerlingen die deze actie

minimaal één keer gebruiken worden meegenomen. Voor het kopiëren van één zin liggen de gemiddelden lager: respectievelijk 0,3 en 2,1. Voor het kopiëren van enkele woorden is het nog lager: respectievelijk 0,1 en 1,8.

### ***Kopiëren en plakken vanuit andere bron(nen)***

Naast het kopiëren en plakken vanuit Blikkie maakten leerlingen ook gebruik van externe bronnen buiten de SchrijfBlik-omgeving. In totaal werden 178 interacties geregistreerd waarbij leerlingen woorden, één zin of meerdere zinnen kopieerden en plakten vanuit een externe bron. Dit aantal ligt aanzienlijk lager dan het aantal vergelijkbare interacties vanuit Blikkie, waar 514 interacties werden geregistreerd.

Zoals weergegeven in Tabel 7, gebruikten leerlingen externe bronnen vooral voor het kopiëren en plakken van meerdere zinnen (94 interacties). Het kopiëren van enkele woorden of van één zin kwam minder vaak voor (beide 42 interacties). Wanneer alle leerlingen worden meegeteld, liggen de gemiddelde aantallen interacties per leerling relatief laag: 0,09 voor het kopiëren van enkele woorden, 0,09 voor het kopiëren van één zin en 0,2 voor het kopiëren van meerdere zinnen. Wanneer alleen de leerlingen worden meegenomen die deze actie minimaal één keer gebruikten, liggen de gemiddelden hoger, maar nog steeds relatief beperkt: 1,4 voor enkele woorden, 1,2 voor één zin en 1,5 voor meerdere zinnen. De standaarddeviaties zijn relatief laag (0,5–1,0), wat erop wijst dat er minder grote verschillen tussen leerlingen zijn in het gebruik van externe kopieeracties. Dit suggereert dat het kopiëren en plakken vanuit externe bronnen minder intensief en minder frequent voorkomt dan kopiëren en plakken vanuit Blikkie.