

AI-gegenereerde feedback in het onderwijs: een scoping review naar de aansluiting op effectieve feedbackstrategieën

Student: R.A. Kerkvliet (s4576446)

Begeleider en eerste beoordelaar: prof. dr. J.W. Strijbos
Tweede beoordelaar: dr. M.N. Mascareño

Rijksuniversiteit Groningen
Faculteit der Gedrags- en Maatschappijwetenschappen
Juni 2025

Aantal woorden: 6895

Abstract

In recent years, the integration of artificial intelligence (AI) in educational contexts has gained significant traction, particularly through the deployment of AI-generated feedback systems to support writing skills and language acquisition in secondary and higher education. This scoping review examines the types of feedback produced by Large Language Models (LLMs) and Natural Language Processing (NLP)-based systems, and assesses their alignment with pedagogically grounded feedback strategies as delineated by Narciss (2008) and Van der Kleij (2019). Through a systematic analysis of 29 peer-reviewed studies published between 2021 and 2025, the findings reveal that AI-generated feedback predominantly adopts formative and corrective forms, effectively fulfilling cognitive functions by identifying errors and suggesting textual improvements. However, the incorporation of metacognitive and motivational feedback components, essential for promoting self-regulation and student motivation, remains inconsistent. The review further indicates a prevalent application of AI feedback in online learning environments, with varying degrees of adaptivity to individual learner needs. These results highlight the capacity of AI systems to enhance technical aspects of writing while underscoring deficiencies in supporting deeper, reflective learning processes. Practical implications include the need for AI system enhancements to integrate metacognitive prompts and motivational elements, alongside professional development for educators to critically implement these tools. This study contributes to the educational discourse by offering a comprehensive synthesis of AI feedback applications and their pedagogical value, providing evidence-based recommendations for future development and implementation.

Keywords: AI-generated feedback, Large Language Models, Natural Language Processing, writing skills, language acquisition, pedagogical feedback strategies

AI-gegenereerde feedback in het onderwijs: een scoping review naar de aansluiting op effectieve feedbackstrategieën.

In de afgelopen jaren is de opmars van artificiële intelligentie (AI) in diverse maatschappelijke domeinen nauwelijks te negeren. Zo is ook binnen het onderwijs een toenemende integratie van AI zichtbaar. AI-technologieën gebaseerd op Large Language Models (LLM's) en Natural Language Processing (NLP) systemen worden in toenemende mate ingezet om leerprocessen te ondersteunen en gepersonaliseerde leerervaringen te verwezenlijken (Dai et al., 2024; Stahl et al., 2024). Eén van de meest veelbelovende toepassingen van AI binnen onderwijscontexten betreft het genereren van feedback voor studenten, met name bij vakken waarbij schrijfvaardigheid en taalverwerving centraal staan (Escalante et al., 2023). AI-gegenereerde feedback heeft de potentie om de beschikbaarheid en efficiëntie van feedback te vergroten, met name in onderwijssituaties met veel studenten en waar

persoonlijke feedback van docenten zeer tijdsintensief is (Demszky et al. 2024). Desondanks roept de toenemende inzet van AI-feedbacksystemen vragen op over de kwaliteit, relevantie en effectiviteit van de geboden feedback. Hoewel sommige studies aantonen dat AI in staat is om inhoudelijk correcte en gepersonaliseerde feedback te leveren (Tong et al., 2021), blijft het onduidelijk in hoeverre deze feedback aansluit bij pedagogisch-didactische principes van effectieve feedbackstrategieën, zoals het bieden van tijdige en specifieke correcties, het stimuleren van metacognitieve reflectie, en het bevorderen van motivatie en zelfregulatie (Hattie & Timperley, 2007; Narciss, 2008; Van der Kleij, 2019)

In dit onderzoek wordt onder AI-gegenereerde feedback verstaan: elke vorm van inhoudelijke feedback op studentprestaties die volledig of gedeeltelijk wordt gegenereerd door een AI-systeem, zonder dat er sprake is van directe menselijke tussenkomst. De feedback kan worden gegeven in de vorm van tekst of spraak en betrekking hebben op prestaties in schrijfvaardigheid of taalverwerving. Het onderzoek beperkt zich tot feedback gegenereerd door LLM's (zoals GPT-4) en NLP-gebaseerde systemen, inclusief Intelligent Tutoring Systems (ITS), Automated Essay Scoring (AES) tools en educatieve chatbots. Feedback binnen digitale leeromgevingen zonder AI-component wordt niet meegenomen in dit onderzoek. Verder ligt de focus op het voortgezet en hoger onderwijs (inclusief universiteit, hogeschool, beroepsonderwijs), met uitsluiting van het primair onderwijs.

Effectieve Feedbackstrategieën

Effectieve feedbackstrategieën zijn essentieel voor het bevorderen van leerprocessen en worden gekenmerkt door een samenspel van pedagogisch-didactische principes die zorgvuldig zijn onderbouwd in de literatuur. Volgens Hattie en Timperley (2007) moet feedback gericht zijn op drie kernvragen: waar staat de student nu (huidige prestaties), waar moet de student naartoe (leerdoelen), en hoe kan de student daar komen (strategieën voor verbetering). Dit impliceert dat feedback niet alleen fouten corrigeert, maar ook het leerproces ondersteunt door studenten te helpen hun eigen voortgang te begrijpen en aan te sturen. Van der Kleij (2019) stelt dat effectieve feedback tijdig, specifiek, helder en afgestemd op het niveau van de leerling moet zijn. Tijdigheid zorgt ervoor dat studenten feedback ontvangen wanneer deze nog relevant is voor hun leerproces, terwijl specificiteit en helderheid ervoor zorgen dat de feedback praktisch toepasbaar is. Adaptiviteit impliceert dat feedback rekening houdt met de individuele behoeften en vaardigheden van de student, wat essentieel is voor gepersonaliseerd leren. Narciss (2008) biedt een aanvullend kader door feedback te conceptualiseren als een proces dat drie functies vervult:

- **Cognitieve functies:** Feedback biedt informatie over fouten, ontbrekende kennis of onduidelijkheden, waardoor studenten hun kennis kunnen corrigeren, aanvullen of herstructureren. Bijvoorbeeld, bij een schrijfo opdracht kan feedback specifieke taalfouten identificeren en uitleggen hoe deze te corrigeren, zoals het juiste gebruik van werkwoordsvormen.

- **Metacognitieve functies:** Feedback ondersteunt studenten bij het monitoren, evalueren en aanpassen van hun leerstrategieën. Dit kan worden bereikt door reflectieve vragen te stellen, zoals: “Hoe kun je je argumenten beter onderbouwen met bewijs?” of door criteria te bieden voor zelfevaluatie.
- **Motivationale functies:** Feedback bevordert doorzettingsvermogen, zelfeffectiviteit en motivatie door voortgang zichtbaar te maken of succeservaringen te benadrukken. Een voorbeeld is het erkennen van verbeteringen in de 'schrijfstijl van een student, zoals: “Je gebruik van signaalwoorden is sterker geworden, wat de coherentie van je tekst verbetert.”.

Deze functies onderstrepen dat effectieve feedback verder gaat dan het louter aanwijzen van correcte of incorrecte antwoorden; het is een dynamisch proces dat informerende, begeleidende en regulerende elementen integreert, afgestemd op de leerdoelen en kenmerken van de student (Narciss, 2008). Shute (2008) benadrukt daarnaast dat feedback procesgericht moet zijn, waarbij de nadruk ligt op leerstrategieën en inspanning in plaats van enkel op het eindresultaat. Dit is met name relevant voor schrijfvaardigheid en taalverwerving, waar iteratieve verbetering en reflectie cruciale componenten zijn van het leerproces (Eckstein et al., 2011). AI-gegenereerde feedbacksystemen worden in toenemende mate geacht deze functies te kunnen vervullen, maar empirisch bewijs over de mate waarin zij daarin slagen, is schaars (Zawacki-Richter et al., 2019).

Hoewel AI-systemen steeds beter worden in het genereren van samenhangende, grammaticaal correcte en inhoudelijk aannemelijke teksten, blijft het onduidelijk of zij erin slagen om feedback te genereren die daadwerkelijk bijdraagt aan het leerproces. In het bijzonder bestaat er weinig systematisch inzicht in de mate waarin AI-gegenereerde feedbacksystemen aansluiten bij effectieve feedbackstrategieën die zijn ontwikkeld vanuit pedagogisch en onderwijskundig onderzoek zoals beschreven door Van der Kleij (2019) en Narciss (2008). Bestaand onderzoek naar AI-feedback richt zich vaak op technische aspecten, zoals nauwkeurigheid, responsiviteit of taalgebruik van AI-gegenereerde feedback (Hein et al., 2024). Daarentegen is er minder aandacht voor de pedagogische waarde van deze feedback: is de feedback inhoudelijk passend, accuraat, motiverend, afgestemd op de leerling, bevorderlijk voor zelfregulatie, en draagt het bij aan het leerproces? De vraag blijft in hoeverre AI-gegenereerde feedbacksystemen, zoals LLM's en NLP-gebaseerde tools, in staat zijn deze functies te vervullen en te voldoen aan deze effectieve principes, een vraag die in dit onderzoek centraal zal staan.

Huidig onderzoek

De maatschappelijke relevantie van dit onderzoek ligt in het toenemende gebruik van AI in het voortgezet en hoger onderwijs. Steeds meer instellingen overwegen of implementeren AI-tools om feedback te genereren, vaak zonder inzicht in de pedagogische waarde ervan. Onderwijsprofessionals, beleidsmakers en ontwikkelaars van educatieve technologieën hebben daardoor behoefte aan een helder overzicht van de voor- en nadelen van AI-gegenereerde feedback. Door pedagogisch-

didactische theorieën over effectieve feedback te verbinden aan recente ontwikkelingen in AI, biedt deze scoping review een nieuwe invalshoek. Het levert een bijdrage aan de onderbouwing van de pedagogische waarde en potentiële valkuilen van AI-feedbacksystemen, en aan het tot stand brengen van richtlijnen voor de implementatie van AI-gegenereerde feedback in het onderwijs. Bovendien ontbreekt in de literatuur een overzicht van de typen AI-gegenereerde feedbacksystemen en de contexten waarin zij worden toegepast. Hierdoor is het voor onderwijsprofessionals lastig om een goede, weloverwogen keuze te maken ten aanzien van het gebruik van dergelijke AI-systemen. Gezien de vlotte opmars van AI in het onderwijs en het grote aantal beschikbare AI-tools, is er dringend behoefte aan overzicht.

Het doel van dit onderzoek is om in kaart te brengen welke typen AI-gegenereerde feedback worden gebruikt in de verschillende lagen van het onderwijs, met uitzondering van het primaire onderwijs, en in hoeverre deze feedback aansluit bij effectieve feedbackstrategieën van Van der Kleij (2019) en Narciss (2008). Dit wordt onderzocht aan de hand van een scoping review van recente wetenschappelijke literatuur. Specifiek wordt onderzocht welke hiaten nog bestaan in de huidige kennis over AI-feedback om aanbevelingen te doen voor de verdere ontwikkeling van AI-feedbacksystemen. Gebaseerd op bovenstaande probleem- en doelstelling is de volgende hoofdvraag geformuleerd:

Welke typen feedback genereren LLM's en NLP-gebaseerde AI-systemen op schrijfvaardigheid en taalverwerving in onderwijscontexten, en in hoeverre sluit de AI-gegenereerde feedback aan bij effectieve feedbackstrategieën?

Om deze hoofdvraag te kunnen beantwoorden, zijn de volgende deelvragen opgesteld:

1. Welke typen feedback worden voornamelijk beschreven in studies naar AI-gegenereerde feedback voor schrijfvaardigheid en taalverwerving (bijvoorbeeld normatief, formatief, adaptief, correctief, evaluatief, informatief)?
2. In welke leeromgevingen (fysiek, online, hybride) kunnen of worden AI-feedbacksystemen ingezet?
3. In hoeverre houdt AI-gegenereerde feedback zich aan effectieve feedbackstrategieën afgeleid uit onderzoek van Narciss (2008) en Van der Kleij (2019)?

Methode

Onderzoeksdesign

Het uitgangspunt van deze studie was om een breed en zo volledig mogelijk inzicht te verkrijgen in de wijze waarop AI-gegenereerde feedback op het gebied van schrijfvaardigheid en taalverwerving wordt ingezet in het onderwijs, en hoe deze feedback aansluit bij effectieve feedbackprincipes. Om dit te realiseren is een scoping review uitgevoerd (Arksey & O'Malley, 2005; Levac, Colquhoun & O'Brien, 2010). Dit type review is passend gegeven de focus op een relatief

nieuw fenomeen, zoals AI-gegenereerde feedback in onderwijscontexten. Om bias te minimaliseren en transparantie te waarborgen, werden de volgende standaarden gehanteerd:

- Transparantie door het selectieproces vast te leggen via het PRISMA 2020 flowdiagram.
- Gebruik van een gestructureerd en vooraf ontworpen data-extractieformulier.

Selectieproces

Dit onderzoek richt zich specifiek op Large Language Models (LLM's) en Natural Language Processing (NLP)-gebaseerde systemen die worden ingezet voor het geven van feedback aan studenten en hoe deze aansluiten bij effectieve feedbackstrategieën. In deze studie wordt uitsluitend het gebruik van AI-gegenereerde feedback onderzocht binnen het beroepsonderwijs, voortgezet en hoger onderwijs, met een focus op schrijfvaardigheid en taalverwerving. Gegeven de recente opkomst van AI richt de scoping review zich op literatuur die niet ouder is dan 5 jaar. Verder wordt er alleen gebruikgemaakt van Engels- en Nederlandstalige literatuur. Hieruit volgt dat om in aanmerking te komen voor inclusie in dit onderzoek voldaan moet zijn aan de volgende vooropgestelde criteria:

- Gepubliceerd in een peer-reviewed tijdschrift of gepresenteerd op een peer-reviewed conferentie.
- Gepubliceerd tussen het jaar 2021 en 2025.
- Geschreven in het Engels of Nederlands.
- Specifieke focus op AI-gebaseerde feedback op schrijfvaardigheid en taalverwerving in het hoger en voortgezet onderwijs.
- Beschrijving van empirische onderzoeksresultaten.
- Specifieke focus op het beroepsonderwijs, voortgezet onderwijs en hoger onderwijs

Studies die enkel theoretische modellen, technische algoritmes zonder toepassing in het onderwijs of conceptuele kaders zonder empirisch bewijs beschrijven zijn uitgesloten.

Zoekstrategie en selectieprocedure

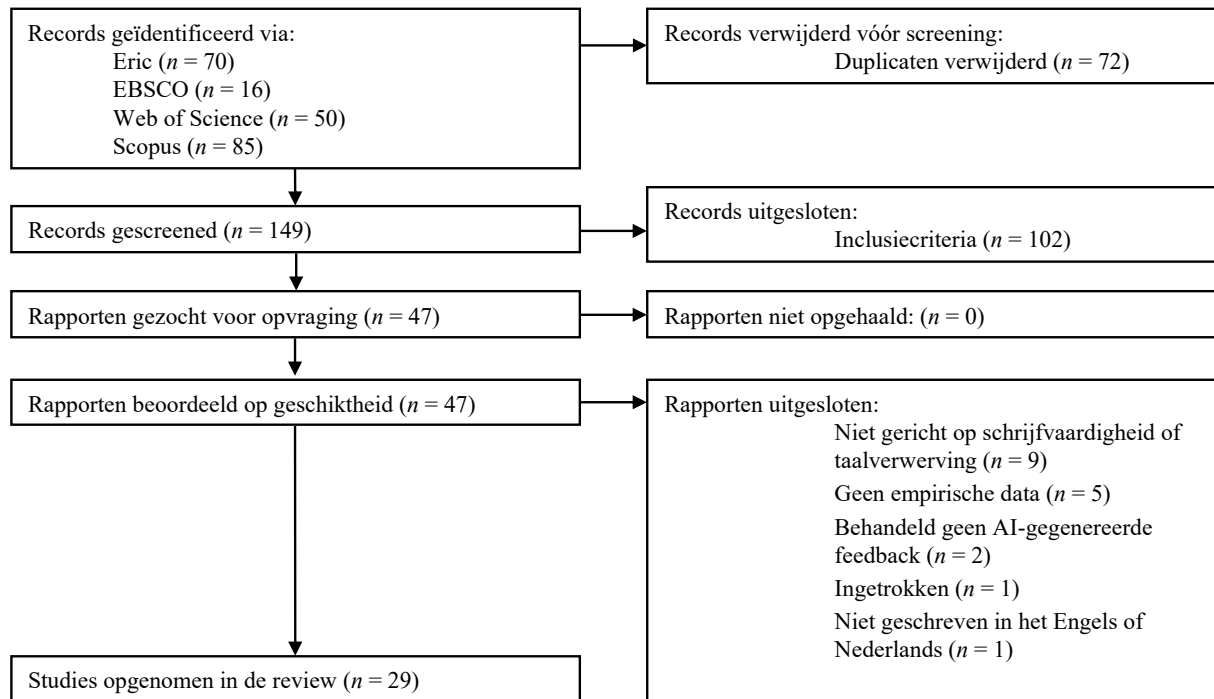
Een uitgebreide zoekslag is uitgevoerd tussen maart en april 2025, in de databases ERIC, Medline, PsycINFO, IEEE Xplore, Scopus, SocINDEX en Web of Science. Tabel 1 toont de hoofd- en zoektermen die voor de zoekstrategie zijn gebruikt.

Om een selectie van relevante literatuur te maken werd er eerst aan de hand van de titel en het abstract gekeken of een bron relevant is. De screening voor het selecteren van studies is uitgevoerd in Rayyan volgens de vastgelegde inclusie- en exclusiecriteria. Vervolgens is er een full-tekst screening gedaan door te bepalen of de bron voldeed aan de inclusiecriteria. In Figuur 1 is het aantal geïncludeerde en geëxcludeerde bronnen bijgehouden aan de hand van het PRISMA 2020 flowdiagram. Tot slot zijn andere relevante bronnen eventueel verworven door gebruik te maken van de sneeuwbalmethode.

Tabel 1*Zoektermen*

Hoofdterm	Zoektermen
AI-technologieën voor feedback	"large language model*" OR "LLM*" OR "natural language processing" OR "NLP" OR "AI-based feedback" OR "machine learning feedback" OR "automated feedback" OR "artificial intelligence feedback" OR "intelligent tutoring systems" OR "chatbot feedback" OR "AI-generated feedback" OR "groot taalmodel" OR "taalmodel" OR "kunstmatige intelligentie"
Onderwijscontexten	"higher education" OR "tertiary education" OR "college students" OR "university students" OR "post-secondary education" OR "graduate education" OR "university of applied sciences" OR "vocational education" OR "hoger onderwijs" OR "universiteit" OR "hogeschool" OR "beroepsonderwijs" OR "mbo" OR "hbo" OR "wo" OR "studenten"
Feedbacktypen	"educational feedback" OR "automated feedback" OR "computer-assisted feedback" OR "AI-generated feedback" OR "LLM feedback" OR "machine-generated feedback" OR "teacher feedback" OR "peer feedback" OR "onderwijsfeedback" OR "automatische feedback" OR "AI gegenereerde feedback" OR "docentfeedback"

De zoekstrategie leverde 221 resultaten op. Na het verwijderen van duplicaten ($n = 72$), werden 149 unieke titels en abstracts gescreend. Aan de hand van de screening op titel en abstract zijn er 102 studies geëxcludeerd, waarna 47 studies overbleven voor de screening van de full text. De screening heeft plaatsgevonden door één beoordelaar en er zijn 19 studies geëxcludeerd om uiteenlopende redenen (te vinden onder 'Rapporten uitgesloten' in Figuur 1). Uiteindelijk resulteerde dit in 29 studies die zijn geïnccludeerd voor verdere analyse. Alle records verkregen uit Medline, PsycINFO en SocINDEX zijn in Figuur 1 samengevoegd onder dezelfde zoekmachine (EBSCO). De database IEEE Xplore leverde echter geen records op.

Figuur 1*PRISMA 2020 Flow Diagram***Data-extractie**

Een deductief codeerschema is opgesteld op basis van bestaande literatuur over feedbackstrategieën (Narciss, 2008; Van der Kleij, 2019) (zie Tabel 2). Er is een verschil in de leeromgevingen tussen de verschillende studies, van face-to-face leeromgevingen tot online- en hybride leeromgevingen. Dit onderscheid is meegenomen in het codeerschema als codering. De eerste zeven categorieën van het codeerschema behandelen de studiekekenmerken zoals beschreven in het data-extractieformulier, de laatste zeven categorieën behandelen de aspecten van effectieve feedbackstrategieën.

Analyse

Beschrijvende statistieken zijn uitgevoerd om inzicht te verkrijgen in de verdeling van kenmerken binnen de geïncludeerde studies, zoals het type AI-technologie, het onderwijsniveau en de vorm van feedback. Deze beschrijvende analyses zijn bedoeld om een globale indruk te krijgen van trends binnen de literatuur, waaronder de onderwijscontexten waarin AI-gegenereerde feedback voornamelijk wordt ingezet en welke typen feedback daarin centraal staan.

Aanvullend is een kwalitatieve inhoudsanalyse toegepast volgens de methode van Flick (2013) waarbij relevante gegevens zijn gecodeerd aan de hand van een vooraf opgesteld codeerschema. Categorieën en terugkerende patronen zijn vervolgens geïdentificeerd.

Tabel 2*Codeerschema*

Categorie	Operationalisatie
Auteur(s) en publicatiejaar	Naam van auteurs en het jaar van publicatie
Type AI-gegenereerde feedback	NLP, LLM, AWE, ITS, etc.
Onderwijsniveau	Voortgezet onderwijs, beroepsonderwijs, hoger onderwijs, etc.
Kenmerken van de leeromgeving	Face-to-face leeromgeving, online leeromgeving of hybride leeromgeving?
Type feedback	Summatief, formatief, adaptief, correctief, evaluatief, informatief, diagnostisch, affirmatief, procesgericht
Gebruikte methode van het onderzoek	Kwantitatief, kwalitatief, mixed methods
Belangrijke bevindingen	Belangrijkste bevindingen uit het onderzoek.
Cognitieve component	Mate waarin de feedback inhoudelijke informatie over fouten, ontbrekende kennis of mogelijke correcties biedt (Narciss, 2008)
Metacognitieve component	Mate waarin de feedback stimuleert om te reflecteren op het leerproces en strategieën bij te sturen (Narciss, 2008)
Motiverende component	In hoeverre de feedback motivatie, zelfvertrouwen of bereidheid om door te zetten versterkt (Narciss, 2008)
Timing van feedback	Wordt de feedback direct na een taak verstrekt? (Van der Kleij, 2019)
Specifieke feedback	Bevat de feedback voldoende details en is deze praktisch toepasbaar? (Van der Kleij, 2019)
Adaptiviteit van feedback	Past de feedback zich aan het niveau en de behoeften van de leerling aan? (Van der Kleij, 2019)
Duidelijke instructies	Wordt er uitleg gegeven over de correctie en hoe deze toegepast kan worden? (Van der Kleij, 2019)

Het opgestelde codeerschema is gebaseerd op bestaande literatuur over effectieve feedbackstrategieën (Narciss, 2008; Van der Kleij, 2019). Binnen het codeerschema zijn aspecten opgenomen zoals de aanwezigheid van cognitieve, metacognitieve en motiverende feedbackfuncties, alsook de timing, specificiteit, adaptiviteit en instructieve helderheid van de feedback. De laatste zeven categorieën zijn gecodeerd op een schaal van 0 (niet uit de studie op te maken), 1 (niet aanwezig), 2 (gedeeltelijk aanwezig) tot 3 (volledig aanwezig). In Appendix A wordt per categorie de codering verder gespecificeerd in de tabellen A1 tot en met A7 aan de hand van voorbeelden uit de literatuur.

Op basis van deze coderingen zijn studies geclusterd. Deze ordening vormde de basis voor een narratieve synthese, waarin verschillen tussen typen AI-technologieën, onderwijscontexten en feedbackvormen zijn vergeleken en samengevat. Hierbij is onder andere gekeken naar methodologische verschillen (onderzoeksopzet, meetmethoden, etc.) en contextuele verschillen (leeromgeving, onderwijsniveau, etc.). Tot slot werd onderzocht in hoeverre deze AI-feedbacksystemen aansluiten bij effectieve feedbackprincipes. Opvallende observaties worden hierbij expliciet benoemd.

Alle werkzaamheden met betrekking tot dataverzameling, codering en analyse zijn digitaal uitgevoerd, met gebruik van institutionele databases van de Rijksuniversiteit Groningen en Microsoft Excel voor de verwerking van de gegevens.

Resultaten

In deze sectie worden eerst de beschrijvende analyses behandeld om een eerste indruk te geven van de trends binnen de literatuur, zoals in welke onderwijsvormen AI-gegenereerde feedback vooral toegepast wordt en in welke vorm dit voornamelijk geschiedt. Vervolgens worden de resultaten gepresenteerd door de categorieën in geclusterde tabellen weer te geven waarna de resultaten die volgen uit de tabellen beschreven zullen worden. Voor illustratieve voorbeelden van de gecodeerde data wordt verwezen naar Appendix A. De belangrijkste bevindingen uit de literatuur worden uiteengezet in Appendix B.

Categorie 1: Type AI-gegenereerde Feedback

Tabel 3 toont een overheersing van Large Language Models (LLM) 17 artikelen (59%), en Automated Writing Evaluation (AWE) (10 artikelen, 34%), met Natural Language Processing (NLP) (één artikel, 3%) en Digital Writing Assistants (DWA) (één artikel, 3%) als minder frequente typen. AWE-systemen zoals 'Write & Improve', 'Grammarly', 'Pigai', en 'E-rater' worden vaak gebruikt in online en face-to-face omgevingen (bijv. ELEbyary et al., 2024; Taskiran et al., 2024), terwijl LLM-systemen zoals 'ChatGPT' en 'GPT-4' voornamelijk in online contexten worden toegepast (bijv. Chan

et al., 2024; Dai et al., 2024). Geen van de geïncludeerde studies maakte gebruik van Intelligent Tutoring Systems (ITS).

Categorie 2: Onderwijsniveau en Leeromgeving (samenvoeging van categorieën)

Tabel 3 toont dat alle 29 artikelen zich richten op hoger onderwijs en niet op voortgezet en beroepsonderwijs, wat een voorname toepassing van AI-feedback in universitaire of hogeronderwijssetting aangeeft. AI-feedback wordt met name toegepast in online omgevingen (19 artikelen, 66%), gevolgd door face-to-face omgevingen (5 artikelen, 17%) en hybride omgevingen (drie artikelen, 10%). In twee artikelen (Chen et al., 2025; Nazaretsky et al., 2024) is de leeromgeving onduidelijk of niet beschreven. In online omgevingen worden vaak AWE-systemen zoals ‘Write & Improve’ gebruikt, die directe feedback leveren in open en afstandsonderwijs (bijv. Taskiran et al., 2024; Taskiran & Göksel, 2022). In face-to-face omgevingen worden zowel AWE (bijv. Altınay & Aydın, 2024; ElEbyary et al., 2024) als LLM’s (bijv. Chan et al., 2024) gebruikt, waarbij docentbegeleiding vaak ook een rol speelt. Hybride omgevingen (El-Garawany, 2024; Liu & Yu, 2022) combineren AI-feedback met persoonlijke interactie met de docent.

Categorie 3: Onderzoeksmethode

Uit Tabel 3 blijkt dat de onderzoeksmethoden variëren: 16 artikelen (55%) gebruiken mixed methods, 7 (24%) kwantitatieve methoden, en 6 (21%) kwalitatieve methoden. Mixed methods worden vaak toegepast in studies die zowel prestaties als percepties onderzoeken (bijv. Li & Kim, 2024; Taskiran et al., 2024), terwijl kwantitatieve studies zich richten op meetbare uitkomsten zoals schrijfscores (bijv. Altınay & Aydın, 2024; Taskiran & Göksel, 2022). Kwalitatieve studies, zoals Venter et al. (2024), analyseren feedbackkwaliteit en gebruikerservaringen en opvattingen.

Categorie 4: Type Feedback

Tabel 4 toont dat de meeste artikelen (19 van 29) rapporteren een combinatie van formatieve, correctieve, en informatieve feedback, wat wijst op een brede toepassing van deze typen in AI-systemen. Formatieve feedback ondersteunt revisie en zelfregulatie (bijv. Taskiran et al., 2024; Wambsganss et al., 2022), correctieve feedback richt zich op grammaticale fouten en taalfouten (bijv. ElEbyary et al., 2024; Geng & Razali, 2022), en informatieve feedback biedt uitleg over structuur en inhoud (bijv. Chan et al., 2025; Dai et al., 2024). Affirmatieve feedback (Ruwe & Mayweg-Paus, 2023) en procesgerichte feedback (Solopova et al., 2023) komen minder frequent voor, die zowel motiverende als strategische elementen benadrukken (affirmatief 1 van 29 en procesgericht 2 van 29).

Tabel 3*Kenmerken van de Studies*

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Onderwijsniveau ^a	Kenmerken van de leeromgeving ^b	Gebruikte methode van het onderzoek ^c
Solak (2024)	LLM (ChatGPT)	2	1	2
Hou et al. (2024)	LLM (ChatGPT, Bard, Gemini)	2	2	3
Zhang & Xu (2024)	AWE (Pigai)	2	2	3
Ruwe & Mayweg-Paus (2023)	LLM (AI-systeem onbekend)	2	2	1
Solopova et al. (2023)	Hybrid AI (RoBERTa, ELECTRA, BERT, BERTopic, rule-based)	2	3	1
Kinder et al. (2025)	LLM (ChatGPT)	2	2	1
Dai et al. (2024)	LLM (GPT-3.5, GPT-4)	2	2	1
Nazaretsky et al. (2024)	LLM (generative AI, systeem niet gespecificeerd)	2	0	1
Chen et al. (2025)	LLM (ChatGPT)	2	0	3
AlGhamdi (2024)	LLM (ChatGPT)	2	2	1
Chan, Lo & Wong (2025)	LLM (ChatGPT)	2	2	3
Banihashem et al. (2024)	LLM (ChatGPT)	2	2	1
Escalante, Pack & Barrett (2023)	LLM (ChatGPT, GPT-4)	2	2	3
Chan, Lo & Wong (2024)	LLM (GPT-3.5-turbo)	2	1	3
Venter, Coetzee & Schmulian (2024)	LLM (GPT-4)	2	2	2
Ruwe & Mayweg-Paus (2024)	LLM (vermoedelijk ChatGPT)	2	2	3
Wambsganss, Janson & Leimeister (2022)	NLP (Argumentation Mining)	2	2	3

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Onderwijsniveau ^a	Kenmerken van de leeromgeving ^b	Gebruikte methode van het onderzoek ^c
Guo & Wang (2024)	LLM (ChatGPT)	2	0	3
El-Garawany (2024)	DWA (QuillBot) Digital Writing Assistant	2	3	3
ElEbyary, Shabara & Boraie (2024)	AWE (Grammarly, E-rater)	2	1	2
Wang (2024)	AWE (Poe)	2	2	2
Akiba & Garte (2024)	LLM (ChatGPT, Google Gemini)	2	2	3
Liu & Yu (2022)	AWE (Write & Improve)	2	3	3
Li & Kim (2024)	AWE (Multiple AFSs: Grammarly, ELSA, etc.)	2	2	3
Yu (2023)	AWE (SafeAssign, Grammarly, Packback)	2	2	3
Geng & Razali (2022)	AWE (Multiple: Pigai, Grammarly, Criterion, etc.)	2	1	3
Taskiran, Yazici & Aydin (2024)	AWE (Write & Improve)	2	2	3
Altınay & Aydın (2024)	AWE (Write & Improve)	2	1	2
Taskiran & Göksel (2022)	AWE (Write & Improve)	2	2	2

^a De coderingen in deze categorie betekenen: 1 = voortgezet onderwijs, 2 = hoger onderwijs, 3 = beroepsonderwijs.

^b De coderingen in deze categorie betekenen: 1 = face-to-face leeromgeving, 2 = online leeromgeving, 3 = hybride leeromgeving.

^c De coderingen in deze categorie betekenen: 1 = kwantitatief, 2 = kwalitatief, 3 = mixed methods.

Tabel 4*Inhoud van AI-gegenereerde Feedback*

Auteur(s) en publicatiejaar	Type feedback ^a	Specifieke feedback ^b	Duidelijke instructies ^c
Solak (2024)	2, 4, 6	3	2
Hou et al. (2024)	2, 4, 6	3	3
Zhang & Xu (2024)	2, 4, 6	2	0
Ruwe & Mayweg-Paus (2023)	6, 8, 9	2	2
Solopova et al. (2023)	2, 4, 6, 7, 9	3	2
Kinder et al. (2025)	2, 4, 6	3	3
Dai et al. (2024)	2, 6	3	3
Nazaretsky et al. (2024)	2	0	0
Chen et al. (2025)	4	2	1
AlGhamdi (2024)	2, 6	3	0
Chan, Lo & Wong (2025)	2, 4, 6	3	2
Banihashem et al. (2024)	2, 4, 6	2	1
Escalante, Pack & Barrett (2023)	2, 6	3	0
Chan, Lo & Wong (2024)	2,4,6	3	3
Venter, Coetzee & Schmulian (2024)	2,4,6	3	2
Ruwe & Mayweg-Paus (2024)	2	0	0
Wambsganss, Janson & Leimeister (2022)	2,4	3	3
Guo & Wang (2024)	2,4,5	3	3
El-Garawany (2024)	2,4,6	3	3
ElEbyary, Shabara & Boraie (2024)	2,4,6	3	3
Wang (2024)	2,4	3	0
Akiba & Garte (2024)	2,4,6	3	3
Liu & Yu (2022)	2,4,6	3	3
Li & Kim (2024)	2,4,6	3	3
Yu (2023)	2,4,6	3	3
Geng & Razali (2022)	2,4,6	3	3
Taskiran, Yazici & Aydin (2024)	2,4,6	3	2
Altınay & Aydın (2024)	2,4,6	0	3
Taskiran & Göksel (2022)	2,4,6	3	0

^a De coderingen in deze categorie betekenen: 1 = summatief, 2 = formatief, 3 = adaptief, 4 = correctief, 5 = evaluatief, 6 = informatief, 7 = diagnostisch, 8 = affirmatief, 9 = procesgericht.

^b De coderingen in deze categorie betekenen: 0 = onduidelijk, 1 = zeer generiek, 2 = matig specifiek, 3 = zeer specifiek. Zie Tabel A1 in de Appendix A voor voorbeelden.

^c De coderingen in deze categorie betekenen: 0 = onduidelijk, 1 = geen instructies, 2 = basisinstructies, 3 = gedetailleerde instructies. Zie Tabel A2 in Appendix A voor voorbeelden.

Categorie 5: Specifieke Feedback

Uit tabel 4 blijkt dat specifieke feedback, die voldoende details biedt en praktisch toepasbaar is (Van der Kleij, 2019), zeer specifiek is in 22 artikelen (76%), matig specifiek in vier artikelen (14%), en onduidelijk in drie artikelen (10%). Hoge specificiteit wordt gerapporteerd in studies met AWE (bijv. Liu & Yu, 2022; Taskiran & Göksel, 2022) en LLM (bijv. Chan et al., 2025; Dai et al., 2024), waarbij feedback specifieke fouten en correcties benoemt. Matige specificiteit komt voor wanneer feedback algemener is (bijv. Banihashem et al., 2024; Zhang & Xu, 2024).

Categorie 6: Duidelijke Instructies

Tabel 4 illustreert dat duidelijke instructies, die goed onderbouwde en begrijpelijke uitleg bieden over correcties en toepassing (Van der Kleij, 2019), gedetailleerd zijn in 14 artikelen (48%), basisinstructies in zes artikelen (21%), afwezig in twee artikelen (7%) en onduidelijk in zeven artikelen (24%). Gedetailleerde instructies worden gerapporteerd in AWE-studies (bijv. ELEbyary et al., 2024; Taskiran & Göksel, 2022) en sommige LLM-studies (bijv. Dai et al., 2024), terwijl basisinstructies voorkomen bij minder expliciete feedback (bijv. Banihashem et al., 2024; Taskiran et al., 2024). Onduidelijkheid is frequent in LLM-studies (bijv. AlGhamdi, 2024; Nazaretsky et al., 2024). De overheersing van volledige aanwezigheid sluit aan bij effectieve feedback, die duidelijke instructies vereist (Van der Kleij, 2019).

Categorie 7: Cognitieve Componenten

Cognitieve componenten worden gedefinieerd als de mate waarin feedback inhoudelijke informatie biedt over fouten en correcties (Narciss, 2008). Tabel 5 toont dat deze in 19 artikelen (66%) als volledig aanwezig zijn gecodeerd. Deze studies melden verbeteringen in grammaticale nauwkeurigheid. Cognitieve componenten waren slechts gedeeltelijk aanwezig in acht artikelen, vaak vanwege een focus op te algemene of beschrijvende feedback (bijv. Banihashem et al., 2024; Escalante et al., 2023). Slechts twee artikelen gaven geen inzicht in de aan- of afwezigheid van de cognitieve component en in geen enkel artikel werd gesproken over de afwezigheid van de cognitieve component.

Categorie 8: Metacognitieve Componenten

Uit tabel 5 blijkt dat metacognitieve componenten, die reflectie en zelfregulatie stimuleren (Narciss, 2008), volledig aanwezig zijn in negen artikelen (31%), gedeeltelijk aanwezig in negen artikelen (31%), en de aanwezigheid hiervan was onduidelijk in 11 artikelen (38%). In geen enkel artikel werd afwezigheid van de metacognitieve component gerapporteerd. Studies met volledige aanwezigheid benadrukken zelfregulatie en autonomie, vaak in online omgevingen (bijv. Li & Kim, 2024; Taskiran et al., 2024; Taskiran & Göksel, 2022). Gedeeltelijke aanwezigheid werd

gerapporteerd wanneer feedback enige zelfevaluatie suggereerde, maar er geen bewijs voor actieve reflectie gevonden kon worden (bijv. Altınay & Aydın, 2024; El-Garawany, 2024).

Categorie 9: Motiverende Componenten

Tabel 5 toont dat motiverende componenten, die motivatie, zelfvertrouwen en doorzettingsvermogen versterken (Narciss, 2008), volledig aanwezig zijn in acht artikelen (28%), gedeeltelijk aanwezig in 16 artikelen (55%) en onduidelijk in vijf artikelen (17%). Volledige aanwezigheid werd gerapporteerd in studies die motivatie en zelfvertrouwen benadrukken, vaak gekoppeld aan directe feedback en autonomie (bijv. Chan et al., 2025; Taskiran et al., 2024; Taskiran & Göksel, 2022). Gedeeltelijke aanwezigheid komt voor wanneer feedback enige aanmoediging biedt, maar geen consistente motivatie verwezenlijkt bij de studenten (bijv. Akiba & Garte, 2024; Zhang & Xu, 2024). Afwezigheid of onduidelijkheid wijst op studies die niets over motiverende elementen rapporteren (bijv. Altınay & Aydın, 2024; Chen et al., 2025). In één artikel kwam naar voren dat de motivatie toenam door de directe en onbeperkte beschikbaarheid van feedback gegenereerd door AI-systemen (Taskiran et al., 2024).

Categorie 10: Timing van Feedback

Tabel 6 maakt duidelijk dat directe feedback in 17 artikelen (59%) werd gerapporteerd, terwijl 12 artikelen (41%) geen informatie boden over de timing van feedback. Geen enkel artikel rapporteert vertraagde of uitgestelde feedback. Directe feedback wordt consistent (18 maal) gerapporteerd in studies die AWE- en LLM-systemen inzetten, vooral in online omgevingen (bijv. El-Garawany, 2024; Taskiran et al., 2024). De afwezigheid van timing-informatie in 11 studies (bijv. Dai et al., 2024; Ruwe & Mayweg-Paus, 2024) kan wijzen op een gebrek aan focus op dit aspect. In drie studies werd onderzocht of feedbackontvangers AI-feedback of menselijke feedback prefereerden zonder te weten van welke bron de feedback afkomstig was (bijv. Wang, 2024), dit kan de afwezigheid van timing-informatie verklaren doordat in dergelijke onderzoeken beide feedbacksoorten tegelijk teruggekoppeld moeten worden aan de feedbackontvanger.

Categorie 11: Adaptiviteit van Feedback

Tabel 6 toont dat adaptiviteit, waarbij feedback zich aanpast aan het niveau en de behoeften van de leerling (Van der Kleij, 2019), volledig adaptief is in acht artikelen (28%), enigszins adaptief in negen artikelen (31%), afwezig in twee artikelen (7%) en onduidelijk in 10 artikelen (34%). Volledige adaptiviteit wordt gerapporteerd in studies met geavanceerde AWE- of LLM-systemen (bijv. El-Garawany, 2024; Wambsganss et al., 2022), terwijl enige adaptiviteit voorkomt bij systemen die slechts gedeeltelijk personaliseren (bijv. Chan et al., 2024; Taskiran et al., 2024). Onduidelijkheid of afwezigheid is frequent in LLM-studies (bijv. Akiba & Garte, 2024; Chen et al., 2025) waar vaste prompts worden gebruikt.

Tabel 5*Functie van AI-gegenereerde Feedback*

Auteur(s) en publicatiejaar	Cognitieve componenten ^a	Metacognitieve componenten ^b	Motiverende componenten ^c
Solak (2024)	3	2	2
Hou et al. (2024)	3	3	3
Zhang & Xu (2024)	2	3	2
Ruwe & Mayweg-Paus (2023)	2	0	2
Solopova et al. (2023)	2	2	2
Kinder et al. (2025)	3	2	2
Dai et al. (2024)	3	0	0
Nazaretsky et al. (2024)	0	0	0
Chen et al. (2025)	2	0	0
AlGhamdi (2024)	2	3	2
Chan, Lo & Wong (2025)	3	3	3
Banihashem et al. (2024)	2	0	2
Escalante, Pack & Barrett (2023)	2	0	2
Chan, Lo & Wong (2024)	3	0	2
Venter, Coetzee & Schmulian (2024)	2	2	3
Ruwe & Mayweg-Paus (2024)	0	0	0
Wambsganss, Janson & Leimeister (2022)	3	0	2
Guo & Wang (2024)	3	0	2
El-Garawany (2024)	3	2	3
ElEbyary, Shabara & Boraie (2024)	3	3	2
Wang (2024)	3	0	2
Akiba & Garte (2024)	3	3	2
Liu & Yu (2022)	3	2	2
Li & Kim (2024)	3	3	3
Yu (2023)	3	2	3
Geng & Razali (2022)	3	2	0
Taskiran, Yazici & Aydin (2024)	3	3	3
Altınay & Aydın (2024)	3	2	0
Taskiran & Göksel (2022)	3	3	3

^a De coderingen in deze categorie betekenen: 0= onduidelijk 1 = niet aanwezig, 2 = gedeeltelijk aanwezig, 3 = volledig aanwezig. Zie Tabel A3 in Appendix A voor voorbeelden.

^b De coderingen in deze categorie betekenen: 0= onduidelijk 1 = niet aanwezig, 2 = gedeeltelijk aanwezig, 3 = volledig aanwezig. Zie Tabel A4 in Appendix A voor voorbeelden.

^c De coderingen in deze categorie betekenen: 0= onduidelijk 1 = niet aanwezig, 2 = gedeeltelijk aanwezig, 3 = volledig aanwezig. Zie Tabel A5 in Appendix A voor voorbeelden.

Tabel 6*Delivery van AI-gegenereerde Feedback*

Auteur(s) en publicatiejaar	Timing van feedback ^a	Adaptiviteit van feedback ^b
Solak (2024)	3	2
Hou et al. (2024)	3	2
Zhang & Xu (2024)	3	2
Ruwe & Mayweg-Paus (2023)	0	0
Solopova et al. (2023)	3	3
Kinder et al. (2025)	3	3
Dai et al. (2024)	0	3
Nazaretsky et al. (2024)	0	0
Chen et al. (2025)	3	1
AlGhamdi (2024)	0	1
Chan, Lo & Wong (2025)	0	2
Banihashem et al. (2024)	0	0
Escalante, Pack & Barrett (2023)	3	0
Chan, Lo & Wong (2024)	0	2
Venter, Coetzee & Schmulian (2024)	0	2
Ruwe & Mayweg-Paus (2024)	0	0
Wambsganss, Janson & Leimeister (2022)	3	3
Guo & Wang (2024)	3	0
El-Garawany (2024)	3	3
ElEbyary, Shabara & Boraie (2024)	3	3
Wang (2024)	0	3
Akiba & Garte (2024)	3	0
Liu & Yu (2022)	3	0
Li & Kim (2024)	3	0
Yu (2023)	3	3
Geng & Razali (2022)	3	0
Taskiran, Yazici & Aydin (2024)	3	2
Altınay & Aydın (2024)	3	2
Taskiran & Göksel (2022)	0	2

^a De coderingen in deze categorie betekenen: 0 = onduidelijk 1 = veel vertraging, 2 = enige vertraging, 3 = directe feedback. Zie Tabel A6 in Appendix A voor voorbeelden.

^b De coderingen in deze categorie betekenen: 0 = onduidelijk 1 = geen adaptiviteit, 2 = enige adaptiviteit, 3 = volledige adaptiviteit. Zie Tabel A7 in Appendix A voor voorbeelden.

Discussie

Dit onderzoek had als doel om inzicht te verschaffen in de feedbacktypen die worden gegenereerd door Large Language Models (LLM's) en Natural Language Processing (NLP)-gebaseerde AI-systemen in de context van schrijfvaardigheid en taalverwerving binnen het voortgezet en hoger onderwijs. Daarnaast werd onderzocht in hoeverre deze AI-gegenereerde feedback aansluit bij effectieve feedbackstrategieën, zoals beschreven door Narciss (2008) en Van der Kleij (2019). De centrale onderzoeksvraag luidde: *Welke typen feedback genereren LLM's en NLP-gebaseerde AI-systemen op schrijfvaardigheid en taalverwerving in onderwijscontexten, en in hoeverre sluit de AI-gegenereerde feedback aan bij effectieve feedbackstrategieën?* Deze vraag werd beantwoord door middel van een scoping review van recente wetenschappelijke literatuur, waarbij specifiek aandacht werd besteed aan de cognitieve, metacognitieve en motivationele functies van feedback, evenals aan de timing, specificiteit en adaptiviteit ervan.

Bevindingen in het Licht van Theorie

De resultaten tonen aan dat AI-gegenereerde feedback voornamelijk formatieve en correctieve vormen aanneemt, met een nadruk op het identificeren en corrigeren van taalfouten, het verbeteren van tekststructuur en het bieden van specifieke suggesties voor verbetering. Dit sluit aan bij de cognitieve functies van feedback, zoals beschreven door Narciss (2008), waarbij feedback informatie biedt over fouten en ontbrekende kennis, waardoor studenten hun werk kunnen corrigeren en aanvullen. Studies zoals die van Solopova et al. (2023) en ElEbyary et al. (2024) illustreren dat AI-systemen, zoals Automated Writing Evaluation (AWE)-tools en LLM's, in staat zijn om gedetailleerde, taalspecifieke feedback te leveren die studenten helpt om grammaticale en stilistische aspecten van hun schrijven te verbeteren. Deze bevindingen komen overeen met het principe van Hattie en Timperley (2007) dat feedback effectief is wanneer het studenten helpt te begrijpen waar ze staan ten opzichte van hun leerdoelen en hoe ze deze kunnen bereiken.

Echter, metacognitieve en motivationele feedbackcomponenten zijn minder consistent aanwezig. Hoewel sommige studies, zoals die van Hou et al. (2024) en Taskiran et al. (2024), rapporteren dat AI-feedback reflectieve vragen of strategische suggesties kan bevatten die zelfregulatie bevorderen, blijft deze metacognitieve ondersteuning beperkt. Dit is opvallend, gezien het belang van metacognitieve feedback voor het stimuleren van zelfevaluatie en strategisch denken, zoals benadrukt door Narciss (2008). Bovendien wordt motivationele feedback, die gericht is op het versterken van zelfeffectiviteit en doorzettingsvermogen, slechts in een klein aantal studies expliciet genoemd (bijvoorbeeld Taskiran et al., 2024). Dit wijst op een kloof tussen de potentie van AI-systemen en hun huidige toepassing, met name in het ondersteunen van de motivationele aspecten die Van der Kleij (2019) essentieel acht voor effectieve feedback. Toekomstig onderzoek zou zich meer moeten richten op het versterken van de metacognitieve en motivationele aspecten van AI-gegenereerde feedback.

Een belangrijke bevinding is de dominante toepassing van AI-feedback in online leeromgevingen, zoals gerapporteerd in studies zoals Zhang en Xu (2024) en Dai et al. (2024). Deze contexten maken directe feedbacklevering mogelijk, wat aansluit bij het principe van tijdigheid (Van der Kleij, 2019). Directe feedback is met name effectief bij routinematige taken, zoals het corrigeren van taalfouten, omdat het snelle aanpassingen mogelijk maakt. Echter, zoals Shute (2008) opmerkt, kan directe feedback minder geschikt zijn voor taken die reflectie en probleemoplossing vereisen, omdat het de zelfstandige ontwikkeling van oplossingsstrategieën kan beperken. Hybride leeromgevingen, zoals beschreven door El-Garawany (2024) en Liu en Yu (2022), lijken een veelbelovende middenweg te bieden door AI-feedback te combineren met menselijke interactie, wat mogelijk beter aansluit bij effectieve strategieën door de integratie van adaptiviteit en motivationele ondersteuning. Het is van belang dat toekomstige studies zich richten op de effectiviteit van een dergelijke synergetische aanpak.

De beperkte aanwezigheid van affirmatieve en procesgerichte feedback is een andere opvallende bevinding. Slechts een klein aantal studies (bijv. Guo en Wang, 2024) benoemt affirmatieve feedback, die gericht is op het benadrukken van successen en het versterken van motivatie. Procesgerichte feedback, die strategieën en inspanning benadrukt in plaats van enkel het eindresultaat, is eveneens ondervertegenwoordigd. Dit is zorgwekkend, gezien het belang van deze feedbackvormen voor het bevorderen van zelfregulatie en langetermijnleerprocessen, zoals onderstreept door Shute (2008) en Hattie & Timperley (2007). De focus van AI-systemen op correctieve en formatieve feedback suggereert dat ze momenteel beter geschikt zijn voor technische verbeteringen dan voor het ondersteunen van diepgaande leerprocessen.

Tot slot tonen de resultaten een wisselende mate van adaptiviteit in AI-feedback. Studies zoals die van Kinder et al. (2025) en Wambsganss et al. (2022) laten zien dat sommige AI-systemen in staat zijn om feedback af te stemmen op individuele behoeften, wat aansluit bij het principe van adaptiviteit zoals beschreven door Van der Kleij, 2019. In andere studies, zoals Nazaretsky et al. (2024), ontbreekt deze adaptiviteit, wat de effectiviteit van de feedback kan beperken omdat generieke feedback mogelijk minder relevant is voor de specifieke leerbehoeften en het leerproces van de student. Deze variabiliteit onderstreept de noodzaak om AI-systemen verder te ontwikkelen om consistente, gepersonaliseerde feedback te verwezenlijken die aansluit bij de individuele leerbehoeften van studenten.

Methodologische Beperkingen

Hoewel dit onderzoek waardevolle inzichten biedt, kent het enkele methodologische beperkingen die de generaliseerbaarheid en diepgang van de bevindingen beïnvloeden. Ten eerste is de scoping review beperkt tot studies die expliciet AI-gegenereerde feedback in de context van schrijfvvaardigheid en taalverwerving behandelen. Hierdoor zijn mogelijk relevante studies over andere toepassingen van AI-feedback, zoals in andere vakgebieden of onderwijsniveaus, buiten beschouwing

gelaten. Dit kan een vertekend beeld geven van de brede toepasbaarheid van AI-feedbacksystemen. Dit benadrukt de noodzaak voor toekomstig onderzoek dat ook andere vakgebieden en onderwijscontexten meeneemt om zo een vollediger beeld te krijgen van de inzetbaarheid van AI-feedback.

Een tweede beperking betreft de heterogeniteit van de geïncludeerde studies. De variatie in onderzoeksopzet, meetmethoden en AI-technologieën (bijvoorbeeld LLM's vs. AWE-tools) bemoeilijkt directe vergelijkingen tussen studies. Bijvoorbeeld, studies zoals Solopova et al. (2023) gebruikten hybride AI-systemen, terwijl andere, zoals Zhang en Xu (2024), zich richtten op specifieke AWE-tools zoals 'Pigai'. Deze diversiteit maakt het lastig om algemene conclusies te trekken over de effectiviteit van AI-feedback. Voor vervolgonderzoek is het aan te raden om de focus te leggen op één specifieke AI-technologie in plaats van meerdere tegelijk. Daarnaast kan het gebruik van gestandaardiseerde onderzoeksopzetten en vergelijkbare meetinstrumenten helpen om eenduidigere en beter generaliseerbare conclusies te kunnen trekken over de effectiviteit van AI-feedback.

Ten derde is de codering van feedbackfuncties en -inhoud gebaseerd op de rapportage in de artikelen, wat gevoelig is voor onvolledige of onduidelijke beschrijvingen. In sommige studies, zoals die van Ruwe en Mayweg-Paus (2024), werd het type AI-systeem niet gespecificeerd, wat de analyse van technologie-specifieke effecten bemoeilijkt. Bovendien waren sommige studies, zoals Nazaretsky et al. (2024), vaag over de aanwezigheid van metacognitieve of motivationele feedbackcomponenten, wat de betrouwbaarheid van de coderingen mogelijk heeft beïnvloed.

Een vierde beperking is de focus op Engelstalige en Nederlandstalige literatuur, zoals toegankelijk via institutionele databases van de Rijksuniversiteit Groningen. Dit kan hebben geleid tot een ondervertegenwoordiging van studies uit niet-westerse contexten, waar AI-feedback mogelijk anders wordt toegepast. Deze geografische en taalkundige beperking kan de generaliseerbaarheid van de bevindingen naar wereldwijde onderwijscontexten beperken.

Ten vijfde is de kwalitatieve inhoudsanalyse, hoewel zorgvuldig uitgevoerd volgens Flick (2013), subjectief. Ondanks het gebruik van een gestructureerd codeerschema, kunnen interpretaties van feedbackfuncties variëren tussen onderzoekers. Dit benadrukt de noodzaak van toekomstig onderzoek met meer gestandaardiseerde meetmethoden, zoals vooraf vastgestelde rubrieken of gestandaardiseerde beoordelingsschalen om de vergelijkbaarheid en betrouwbaarheid van de bevindingen te vergroten.

Tot slot laten methodologische beperkingen, zoals de heterogeniteit van studies en de afhankelijkheid van gerapporteerde data, de noodzaak voor verder onderzoek zien. Praktisch gezien kunnen onderwijsinstellingen en ontwikkelaars profiteren van een integratie van AI-tools, met aandacht voor metacognitieve en motivationele feedback en de combinatie van AI met menselijke interactie. Toekomstig onderzoek zou zich moeten richten op gestandaardiseerde meetmethoden en de

toepassing van AI-feedback in diverse onderwijskundige contexten om een breder begrip van de effectiviteit ervan te verkrijgen.

Praktische Implicaties

De bevindingen van dit onderzoek bieden verschillende praktische implicaties voor onderwijsinstellingen, docenten en ontwikkelaars van AI-feedbacksystemen. Ten eerste suggereert de dominante toepassing van AI-feedback in online leeromgevingen dat onderwijsinstellingen kunnen investeren in digitale platforms die directe, formatieve feedback leveren. Deze platforms zijn met name geschikt voor grootschalige cursussen waarbij persoonlijke feedback van docenten tijdsintensief is. Docenten kunnen AI-tools zoals ‘Grammarly’ of ‘ChatGPT’ integreren in hun lespraktijk om studenten te ondersteunen bij technische aspecten van schrijven, zoals grammatica en structuur, terwijl ze zelf focussen op diepgaandere, metacognitieve begeleiding.

Ten tweede benadrukt de beperkte aanwezigheid van metacognitieve en motivationele feedback de noodzaak voor ontwikkelaars om AI-systemen te verbeteren op deze gebieden. Ontwikkelaars kunnen bijvoorbeeld reflectieve prompts of affirmatieve feedbackmechanismen integreren, zoals suggesties voor zelfevaluatie of het benadrukken van verbeteringen in studentwerk. Dit zou AI-feedback beter laten aansluiten bij de effectieve principes van Narciss (2008) en Van der Kleij (2019), die het belang van zelfregulatie en motivatie onderstrepen.

Een derde implicatie betreft het potentieel van hybride leeromgevingen. Onderwijsinstellingen kunnen experimenteren met ‘blended learning’-modellen waarin AI-feedback wordt gecombineerd met menselijke interactie. AI kan bijvoorbeeld worden gebruikt voor initiële correcties en suggesties, terwijl docenten aanvullende feedback bieden die gericht is op motivatie en strategische ontwikkeling. Dit kan de effectiviteit van feedback vergroten, vooral bij complexe taken die reflectie vereisen.

Tot slot wijzen de bevindingen op de noodzaak van training voor docenten in het gebruik van AI-feedbacktools. Meermaals (7 keer) benoemen studies de noodzaak van kwalitatief goede prompts (Zie Appendix B). Docenten moeten worden uitgerust met de kennis om AI-gegenereerde feedback kritisch te genereren, evalueren en aan te vullen, zodat deze optimaal aansluit bij de leerbehoeften van studenten. Dit vereist professionele ontwikkelingstrajecten die zowel technische als pedagogische aspecten van AI-integratie in het onderwijs behandelen.

Referenties

* Studies die zijn geïncludeerd en geanalyseerd.

*Akiba, D., & Garte, R. (2024). Leveraging AI Tools in University Writing Instruction: Enhancing Student Success While Upholding Academic Integrity. *Journal of Interactive Learning Research*, 35(4), 467-480. <https://doi.org/10.70725/355152wkijve>

* Altinay, G., & Aydin, S. (2024). Automated Feedback and Writing Achievement in a Foreign Language Context. *Online Submission*. <https://files.eric.ed.gov/fulltext/ED662856.pdf>

Arksey, H., & O'malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>

* Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback?. *International Journal of Educational Technology in Higher Education*, 21(1), 23. <https://doi.org/10.1186/s41239-024-00455-4>

*Chan, S., Lo, N., & Wong, A. (2024). Leveraging generative AI for enhancing university-level English writing: comparative insights on automated feedback and student engagement. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186x.2024.2440182>

*Chan, S. T. S., Lo, N. P. K., & Wong, A. M. H. (2024). Enhancing university level English proficiency with generative AI: Empirical insights into automated feedback and learning outcomes. *Contemporary Educational Technology*, 16(4), ep541. <https://doi.org/10.30935/cedtech/15607>

*Chen, Z., Wei, W., Zhu, X., & Yao, Y. (2025). Unpacking the rejection of L2 students toward ChatGPT-Generated Feedback: An Explanatory research. *ECNU Review of Education*. <https://doi.org/10.1177/20965311241305140>

Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2024). Can automated feedback improve teachers' uptake of student ideas? Evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation and Policy Analysis*, 46(3), 483-505. <https://doi.org/10.3102/01623737231169270>

*Dai, W., Tsai, Y., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education Artificial Intelligence*, 7, 100299. <https://doi.org/10.1016/j.caeai.2024.100299>

Eckstein, G., Chariton, J., & McCollum, R. M. (2011). Multi-draft composing: An iterative model for academic argument writing. *Journal of English for academic purposes*, 10(3), 162-172. <https://doi.org/10.1016/j.jeap.2011.05.004>

- * ElEbyary, K., Shabara, R., & Boraie, D. (2024). The differential role of AI-operated WCF in L2 students' noticing of errors and its impact on writing scores. *Language Testing in Asia*, 14(1), 1-24. <https://doi.org/10.1186/s40468-024-00312-1>
- * El-Garawany, M. S. M. (2024). The Effects of a QuillBot-Based Intervention on English Language Majors' EFL Writing Performance, Apprehension, and Self-Efficacy. *Language Teaching Research Quarterly*, 43, 167-189. <https://files.eric.ed.gov/fulltext/EJ1457355.pdf>
- * Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Flick, U. (Ed.). (2013). *The SAGE handbook of qualitative data analysis*. sage. <https://dx.doi.org/10.4135/9781446282243>
- * Geng, J., & Razali, A. B. (2022). Effectiveness of the Automated Writing Evaluation Program on Improving Undergraduates' Writing Performance. *English Language Teaching*, 15(7), 49-60. <https://doi.org/10.5539/elt.v15n7p49>
- * Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health information & libraries journal*, 26(2), 91-108. .1111/j.1471-1842.2009.00848.x
- Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2013). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology*, 34(3), 269-290. <https://doi.org/10.1080/01443410.2013.785384>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hein, I., Cecil, J., & Lerner, E. (2024). Acceptance and motivational effect of AI-driven feedback in the workplace: An experimental study with direct replication. *Frontiers in Organizational Psychology*, 2, 1468907. <https://doi.org/10.3389/forgp.2024.1468907>
- * Hou, X., He, S., & Cuigong, R. (2024, June). Learner Use of AI-Generated Feedback for Written Corrective Feedback in L2 Writing: Usefulness, User Proficiency and Attitude. In *Proceedings of the 2024 8th International Conference on Education and Multimedia Technology* (pp. 70-76). <https://doi.org/10.1145/3678726.3678767>
- * Yu, E. (2023). Intelligent Enough? Artificial Intelligence for Online Learners. *Journal of Educators Online*, 20(1), n1. <https://files.eric.ed.gov/fulltext/EJ1383757.pdf>
- * Kinder, A., Briese, F. J., Jacobs, M., Dern, N., Glodny, N., Jacobs, S., & Leßmann, S. (2025). Effects of adaptive feedback generated by a large language model: A case study in teacher

- education. *Computers and Education: Artificial Intelligence*, 8, 100349. <https://doi.org/10.1016/j.caeai.2024.100349>
- Levac, D., Colquhoun, H., & O'brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation science*, 5, 1-9. <https://doi.org/10.1186/1748-5908-5-69>
- * Li, L., & Kim, M. (2024). It is like a friend to me: Critical usage of automated feedback systems by self-regulating English learners in higher education. *Australasian Journal of Educational Technology*, 40(1), 1-18. <https://doi.org/10.14742/ajet.8821>
- * Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78-105. <https://doi.org/10125/73480>
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In *Handbook of research on educational communications and technology* (pp. 125-143). Routledge.
- * Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024, September). AI or human? Evaluating student feedback perceptions in higher education. In *European Conference on Technology Enhanced Learning* (pp. 284-298). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72315-5_20
- PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only — PRISMA statement. (z.d.). PRISMA Statement. <https://www.prisma-statement.org/prisma-2020-flow-diagram>
- * Ruwe, T., & Mayweg-Paus, E. (2023). “Your argumentation is good”, says the AI vs humans—The role of feedback providers and personalised language for feedback effectiveness. *Computers and Education: Artificial Intelligence*, 5, 100189. <https://doi.org/10.1016/j.caeai.2023.100189>
- * Ruwe, T., & Mayweg-Paus, E. (2024, October). Embracing LLM Feedback: the role of feedback providers and provider information for feedback effectiveness. In *Frontiers in Education* (Vol. 9, p. 1461362). Frontiers Media SA. <https://doi.org/10.3389/feduc.2024.1461362>
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795>
- Stahl, M., Biermann, L., Nehring, A., & Wachsmuth, H. (2024). Exploring LLM prompting strategies for joint essay scoring and feedback generation. *arXiv preprint arXiv:2404.15845*. <https://doi.org/10.48550/arXiv.2404.15845>
- * Solak, E. (2024). Examining Writing Feedback Dynamics from CHATGPT AI and Human Educators: a Comparative Study. *Педагогика*, 96(7), 955-970. <https://doi.org/10.53656/ped2024-7.05>
- * Solopova, V., Rostom, E., Cremer, F., Gruszczynski, A., Witte, S., Zhang, C., ... & Landgraf, T. (2023, September). Papagai: Automated feedback for reflective essays. In *German Conference on Artificial Intelligence (Künstliche Intelligenz)* (pp. 198-206). Cham: Springer Nature Switzerland. <https://arxiv.org/pdf/2307.07523>

- * Taskiran, A., & Goksel, N. (2022). Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal of Distance Education*, 23(2), 120-139. <https://doi.org/10.17718/tojde.1096260>
- * Taskiran, A., Yazici, M., & Erdem Aydin, I. (2024). Contribution of automated feedback to the English writing competence of distance foreign language learners. *E-Learning and Digital Media*, 21(1), 24-41. <https://doi.org/10.1177/20427530221139579>
- Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In SAGE Publications Ltd eBooks (pp. 17–36). <https://doi.org/10.4135/9781526405555.n2>
- Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), 1600-1631. <https://doi.org/10.1002/smj.3322>
- Van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education*, 85, 175-189. <https://doi.org/10.1016/j.tate.2019.06.010>
- * Venter, J., Coetzee, S. A., & Schmulian, A. (2024). Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. *Assessment & Evaluation in Higher Education*, 1-21. <https://doi.org/10.1080/02602938.2024.2415649>
- * Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, 191, 104644. <https://doi.org/10.1016/j.compedu.2022.104644>
- * Wang, D. (2024). Teacher-versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distributed Learning*, 25(3), 37-56. <https://doi.org/10.19173/irrodl.v25i3.7646>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International journal of educational technology in higher education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- * Zhang, Z., & Xu, L. (2024). Student engagement with automated feedback on academic writing: A study on Uyghur ethnic minority students in China. *Journal of Multilingual and Multicultural Development*, 45(8), 3466-3479. <https://doi.org/10.1080/01434632.2022.2102175>

Appendix A

Voorbeelden gecodeerde data

Tabel A1

Voorbeelden Specifieke Feedback

Codering	Betekenis	Voorbeeld uit literatuur
1	Zeer generiek	Niet voorgekomen in de literatuur
2	Matig specifiek	“This means that students, with their contextual knowledge and critical thinking skills, may be better equipped to identify issues within the essays that ChatGPT may overlook.” (Banihashem et al., 2024, p. 19)
3	Zeer specifiek	“Those that preferred AI-generated feedback primarily cited the clarity and specificity of the feedback as being useful for improving their writing.” (Escalante et al., 2023, p. 13)

Tabel A2

Voorbeelden Duidelijke Instructies

Codering	Betekenis	Voorbeeld uit literatuur
1	Geen instructies	“ChatGPT did not give me the exact sentences, like which sentence is unclear; it just told me to revise all the paragraphs, so I will not make any changes on this point.” (Respondent in Chen et al., 2025, p. 11)
2	Basisinstructies	“The AI’s suggestion to follow general writing rules, despite not knowing specific paragraph requirements, may be consistent with Constructivist Learning Theory.” (Solak, 2024, p. 963)
3	Gedetailleerde instructies	“The adaptive feedback followed Hattie and Timperley’s (2007) model, offering specific suggestions for improving their decision-making and line of argumentation.” (Kinder et al., 2025, p. 2)

Tabel A3*Voorbeelden Cognitieve Componenten*

Codering	Betekenis	Voorbeeld uit literatuur
1	De feedback geeft geen inhoudelijke toelichting.	Niet voorgekomen in de literatuur.
2	De feedback benoemt fouten of geeft beperkte inhoudelijke aanwijzingen.	“ChatGPT did not give me the exact sentences, like which sentence is unclear; it just told me to revise all the paragraphs, so I will not make any changes on this point.” (Respondent in Chen et al., 2025, p. 11)
3	De feedback geeft duidelijke en bruikbare inhoudelijke uitleg en/of correcties.	“The teachers also commented that Chat- GPT’s feedback was detailed and specific, as it not only pointed out students’ problems but also recommended solutions to the problems. More importantly, ChatGPT explained the motives or purposes for its feedback.” (Guo & Wang, 2024, p. 8446)

Tabel A4*Voorbeelden Metacognitieve Componenten*

Codering	Betekenis	Voorbeeld uit literatuur
1	De feedback bevat geen aanzet tot reflectie of zelfregulatie.	Niet voorgekomen in de literatuur.
2	De feedback suggereert enige zelfevaluatie of strategie-aanpassing.	“Then, they pasted these drafts to QuillBot and collaboratively reviewed the generated feedback, identified the errors, and developed strategies... revising their writing and composing their final drafts based on what they had learnt from the QuillBot feedback” (El-Garawany, 2024, p. 176)
3	De feedback moedigt actief reflectie, monitoring of strategisch bijsturen aan.	“Participants highlighted that they can also get unlimited AF and this process contributes to learning by making mistakes, learning through trial and error. They stated that by constantly trying and making mistakes in their drafts, their writing competence improved, and they were satisfied with this awareness.” (Taskiran et al., 2024, p. 33)

Tabel A5*Voorbeelden Motiverende Componenten*

Codering	Betekenis	Voorbeeld uit literatuur
1	De feedback heeft geen motiverend element.	Niet voorgekomen in de literatuur.
2	De feedback bevat beperkte aanmoediging of waardering.	“...whilst suggesting that slight improvements in motivation and engagement may partly account for this relationship.” (Chan et al., 2024, p. 14)
3	De feedback ondersteunt actief motivatie, zelfvertrouwen en doorzettingsvermogen.	“It is therefore possible that AI feedback functioned through increasing student motivation and encouraging engagement during the task” (Chan et al., 2025, p. 16)

Tabel A6*Voorbeelden Timing van Feedback*

Codering	Betekenis	Voorbeeld uit literatuur
1	Veel vertraging	Niet voorgekomen in de literatuur.
2	Enige vertraging	Niet voorgekomen in de literatuur.
3	Directe feedback	“...thanks to the instant feedback provided with AF.” (Taskiran et al., 2024, p. 36)

Noot. Met AF wordt geautomatiseerde feedback bedoeld.

Tabel A7*Voorbeelden Adaptiviteit van Feedback*

Codering	Betekenis	Voorbeeld uit literatuur
1	Geen adaptiviteit	“Specifically, ChatGPT could not provide personalized suggestions based on proficiency levels and lacked emotional support, leading to discomfort and negative behavior in some students.” (Chen et al., 2025, p. 11)
2	Enige adaptiviteit	“Indeed, the specificity of the feedback tailored to the individual was remarked upon as helpful in this regard, with one participant stating, ‘I was honestly impressed with how well it had able to identify the points within my paper, so the shock was probably an emotional reaction’.” (Chan et al., p. 12)
3	Volledige adaptiviteit	“We designed and built an adaptive learning tool that provides students with automated feedback on their argumentation skill level based on our argumentation mining model. In designing the tool, we followed the findings of recent user studies (Afrin et al., 2021; Wambsganss et al., 2020), a review of the extant literature related to the provision of feedback related to failure (e.g., Bjork et al. (2013); Metcalfe (2017); Lawrence and Reed (2019)), and self-regulated learning theory” (Wambsganss et al., 2022, p. 5)

Appendix B

Belangrijke bevindingen

Tabel B1

Belangrijke Bevindingen per Studie

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Belangrijke bevindingen	Belang goede prompts ^a	Synergetische aanpak ^b
Solak (2024)	LLM (ChatGPT)	AI-feedback was objectiever en uitgebreider dan menselijke feedback; menselijke feedback was meer empathisch en gepersonaliseerd		
Hou et al. (2024)	LLM (ChatGPT, Bard, Gemini)	Deelnemers verbeterden significant in grammatica, vocabulaire, structuur en academische stijl. Promptkwaliteit bleek cruciaal voor effectiviteit.	Ja	
Zhang & Xu (2024)	AWE (Pigai)	AWE-systeem verhoogde betrokkenheid. Reflectie, herziening en revisiestrategieën verbeterden. Promptgebruik was beperkt, maar beïnvloedde resultaat.	Ja	
Ruwe & Mayweg-Paus (2023)	LLM (AI-systeem onbekend)	AI-feedback werd als deskundig ervaren; geen significant effect op motivatie, emoties of zelfvertrouwen. AI scoorde hoog op expertise, lager op warmte.		
Solopova et al. (2023)	Hybrid AI (RoBERTa, ELECTRA, BERT, BERTopic, rule-based)	PapagAI werd als effectief ervaren; scoorde beter dan GPT-3 op specifieke taken; verwerkt emoties/motivatie; transparant; promptkwaliteit beïnvloedt feedback.	Ja	
Kinder et al. (2025)	LLM (ChatGPT)	ChatGPT-feedback leidde tot langere teksten, hogere waardering dan statische feedback.		

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Belangrijke bevindingen	Belang goede prompts ^a	Synergetische aanpak ^b
Dai et al. (2024)	LLM (GPT-3.5, GPT-4)	GPT-4 genereerde structurele, leesbare en evidence-based feedback die meerdere effectieve dimensies afdekt (feed-up, feed-forward, process, self-regulation), en scoorde hoger dan GPT-3.5 en menselijke feedback. Veel overeenkomsten met evidence-based feedbackprincipes.		
Nazaretsky et al. (2024)	LLM (generative AI, systeem niet gespecificeerd)	Studenten beoordelen AI-feedback positiever wanneer de bron onbekend is, maar prefereren menselijke feedback zodra de identiteit duidelijk is. AI-feedback wordt als minder geloofwaardig ervaren.		
Chen et al. (2025)	LLM (ChatGPT)	L2-studenten wijzen veel AI-feedback (vooral inhoudelijk) afwijzen wegens verwarring, werkdruk en mismatch met verwachtingen. Slechts 54% van de AI-feedback werd geaccepteerd. Form-focused feedback werd vaker geaccepteerd dan content-focused feedback. De AI-feedback sloot niet altijd aan bij docent- of peerfeedback, wat voor wantrouwen zorgde. Geen aanwijzingen dat studenten leerstrategieën aanpasten of motivatie ontwikkelden op basis van de feedback. De AI-feedback werd gegenereerd op basis van een vaste prompt opgesteld door docenten. Eventuele aanpassing moest volledig van de student zelf komen; het systeem zelf bood geen adaptiviteit.	Ja	

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Belangrijke bevindingen	Belang goede prompts ^a	Synergetische aanpak ^b
AlGhamdi (2024)	LLM (ChatGPT)	Studenten vonden de ChatGPT-feedback soms behulpzaam ('helped me improve my writing'), maar ook onpersoonlijk ('lacked human touch') en inconsistent. De feedback was gebaseerd op vaste output, zonder aanpassing aan individuele studenten. Adaptiviteit ontbrak. De feedback gegenereerd door ChatGPT werd als informatief ervaren en leidde bij sommige studenten tot verbetering. Tegelijkertijd gaven studenten aan dat de feedback vaak te algemeen en niet gepersonaliseerd was. Er werd daarom een synergetische aanpak aanbevolen waarin AI-feedback gecombineerd wordt met menselijke begeleiding		Ja
Chan, Lo & Wong (2025)	LLM (ChatGPT)	Studenten die AI-feedback ontvingen toonden statistisch significante verbetering in schrijfkwaliteit en hogere motivatie. De feedback werd gewaardeerd om duidelijkheid en toepasbaarheid, maar ook als emotioneel neutraal en onpersoonlijk ervaren. Daarom wordt aanbevolen AI te combineren met menselijke feedback voor betere ondersteuning.		Ja
Banihashem et al. (2024)	LLM (ChatGPT)	ChatGPT gaf vooral beschrijvende feedback (wat er is geschreven), terwijl peers vaker probleemgerichte feedback gaven (wat ontbreekt). Er was geen relatie tussen essaykwaliteit en feedbackkwaliteit. Beide vervullen verschillende feedbackdimensies		
Escalante, Pack & Barrett (2023)	LLM (ChatGPT, GPT-4)	Geen significant verschil in leeruitkomsten tussen AI- en docentfeedback. Studenten waardeerden AI om snelheid, maar ook menselijke feedback om nuance. Beide vormen worden aanbevolen in een blended aanpak.		Ja

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Belangrijke bevindingen	Belang goede prompts ^a	Synergetische aanpak ^b
Chan, Lo & Wong (2024)	LLM (GPT-3.5-turbo)	AI-feedback verbeterde essayscores met 3,113 punten, met significante toename in motivatie en betrokkenheid, maar gemengde emotionele reacties		
Venter, Coetzee & Schmulian (2024)	LLM (GPT-4)	AI-gegenereerde feedback voldeed aan principes van effectieve feedback (gemiddelde score 2,67/3), maar vertoonde variabiliteit, met zwakkere prestaties in het leveren van hoogwaardige informatie en het faciliteren van zelfevaluatie	Ja	
Ruwe & Mayweg-Paus (2024)	LLM (vermoedelijk ChatGPT)	LLM-feedback werd als betrouwbaarder beschouwd dan feedback van docenten, waarbij informatie over de aanbieder de effectiviteit beïnvloedde. feedbackgeletterdheid ondersteunt betrokkenheid		
Wambsganss, Janson & Leimeister (2022)	NLP (Argumentation Mining)	Geautomatiseerde feedback verbeterde argumentatievaardigheden; studenten meldden hoge gebruiksvriendelijkheid en zelfeffectiviteit		
Guo & Wang (2024)	LLM (ChatGPT)	ChatGPT genereerde meer feedback dan docenten, met gelijke focus op inhoud, organisatie en taal. gebruikte meer directieve en positieve feedback, terwijl docenten informatieve en vragende feedback gaven; docenten zagen ChatGPT als een nuttige samenwerkingspartner, maar merkten beperkingen op zoals off-topic feedback.	Ja	Ja
El-Garawany (2024)	DWA (QuillBot) Digital Writing Assistant	QuillBot-interventie verbeterde schrijfstijl, verminderde schrijfangst en verhoogde zelfeffectiviteit. studenten waardeerden de gebruiksvriendelijkheid, maar merkten beperkingen van de gratis versie		

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Belangrijke bevindingen	Belang goede prompts ^a	Synergetische aanpak ^b
ElEbyary, Shabara & Boraie (2024)	AWE (Grammarly, E-rater)	AI-feedback verbeterde het opmerken van fouten in vergelijking met docentenfeedback. Grammarly was het meest effectief voor grammaticale fouten, docentenfeedback voor inhoudelijke fouten. Alle feedback verbeterde de schrijfscores (dus een soort blended aanpak wordt aanbevolen).		Ja
Wang (2024)	AWE (Poe)	AI-gegenereerde feedback (Poe) presteerde beter dan docentenfeedback en geen feedback in het verminderen van schrijfangst en het verbeteren van nauwkeurigheid, vloeiendheid en complexiteit		
Akiba & Garte (2024)	LLM (ChatGPT, Google Gemini)	LLM-feedback was duidelijk, specifiek en uitvoerbaar, wat betrokkenheid en zelfstandigheid bevorderde. studenten vonden het ethisch, wat de verleiding tot spieken verminderde, hoewel sommigen de feedback als onpersoonlijk ervoeren		
Liu & Yu (2022)	AWE (Write & Improve with Cambridge)	Leerlingen besteedden meer tijd en moeite aan indirecte feedback, maar hadden een lagere opname en minder succesvolle revisies. de mate van explicietheid beïnvloedt betrokkenheid. onnauwkeurige feedback leidde tot fouten		
Li & Kim (2024)	AWE (Multiple AFSs: Grammarly, ELSA, etc.)	Studenten waardeerden geautomatiseerde feedbacksysteem (AFSs) voor nauwkeurigheid, voortgangsregistratie en vertrouwen. kritische betrokkenheid maakten de beperkingen minder prominent. begeleiding van docenten en menselijke feedback waren cruciaal		Ja

Auteur(s) en publicatiejaar	Type AI-genereerde feedback	Belangrijke bevindingen	Belang goede prompts ^a	Synergetische aanpak ^b
Yu (2023)	AWE (SafeAssign, Grammarly, Packback)	AI-feedback verbeterde schrijven, metacognitieve vaardigheden en vertrouwen, maar niet de creativiteit. onderpresterende studenten hadden meer ondersteuning nodig bij inhoudsontwikkeling. AI moet worden afgestemd op de behoeften van studenten om eigenaarschap van leren te versterken		
Geng & Razali (2022)	AWE (Multiple: Pigai, Grammarly, Criterion, etc.)	AWE-feedback verbeterde algeheel schrijven, nauwkeurigheid, revisievaardigheid en grammaticale kennis het bevordert ook autonomie, maar is beperkt in inhoud/organisatie en vereist ondersteuning van docenten		Ja
Taskiran, Yazici & Aydin (2024)	AWE (Write & Improve)	Geautomatiseerde feedback verbeterde schrijfvaardigheid, autonomie en motivatie aanzienlijk. Motivatie nam toe door de beschikbaarheid van onbeperkte feedback. Gebruiksfrequentie correleerde met prestaties. Leerlingen met lage vaardigheid hadden moeite met het begrijpen van feedback.		
Altınay & Aydın (2024)	AWE (Write & Improve)	Geautomatiseerde feedback verbeterde schrijfpresetaties meer dan docentenfeedback, met verbeteringen in nauwkeurigheid en vloeiendheid		
Taskiran & Göksel (2022)	AWE (Write & Improve)	Zowel docenten- als geautomatiseerde feedback verbeterde schrijfpresetaties aanzienlijk; docentenfeedback presteerde iets beter dan geautomatiseerde feedback		

^a Studies gemarkeerd met 'Ja' wijzen op het belang van kwalitatief goede prompts bij het genereren van AI-feedback.

^b Studies gemarkeerd met 'Ja' bevelen een synergetische aanpak aan waarbij AI-feedback en menselijke feedback gecombineerd worden.