https://dx.doi.org/10.20880/kler.2025.60.4.283

LLM 활용 글쓰기에서 언어적 동형화 양상
— 뇌신경 연결성 축소와 인지의 부채

장동민 전주오송초등학교 교사(제1저자)

박종호 대구교육대학교 강사(교신저자)

I. 서론

대규모 언어모델(LLM)을 활용한 글쓰기 보조는 어느새 우리의 삶에 깊이 스며들었다. 이는 글쓰기에서 문법의 정확성과 문장 유창성, 생산성을 유의하게 높여주기 때문이다 (Agarwal, Vashistha, & Naaman, 2025; Bhat, Shrivastava, & Guo, 2023; Fu, Newman, Jakesch, & Kreps, 2023). 이러한 효율성에 힘입어 학술적 글쓰기(논문, 리포트, 에세이 등)에서도 ChatGPT, 제미나이, 뤼튼 등 LLM 활용이 빠르게 보편화되고 있다.

그러나 효율성의 이면에서 필자의 실제 글쓰기 능력의 저하(Budiyono, Marzuki, Pudjaningsih, Prastio, & Maulidina, 2025; Kosmyna, Hauptmann, Yuan, Situ, Liao, Beresnitzky, & et al., 2025)와 서로 다른 필자 산출물이 유사한 어조, 구조, 어휘로 수렴하는 동형화(homogenization) 문제가 꾸준히 지적된다 (Padmakumar & He, 2024; Anderson, Shah, & Kreminski, 2024; Agarwal et al., 2025). 실제로 AI가 생성한 글을 반복적으로 채택함에 따라 필자의 어휘 다양성과 주도성, 글쓰기 능력이 저하되는 경향이 나타남이 보고되었다(Fu et al., 2023; Bhat, Shrivastava, & Guo,

2023). 또한 비서구권 필자의 글이 '자연스러운 미국식' 문체로 수렴하는 문화적 차원의 수렴 현상도 관찰된다(Agarwal et al., 2025).

인간의 쓰기는 텍스트 산출을 넘어 사고의 진전과 재구성을 동반하는 복합적 심리 인지 활동으로 언어적 외화와 반성적 내면화의 순환 속에서 사고가 고도화된다(Vygotsky, 2009). Bereiter & Scardamalia(1987)가 구분한 '지식 나열(knowledge-telling)'과 '지식 변형(knowledge-transforming)' 가운데, 능숙한 필자의 쓰기는 문제 재정의와 의미 재구성을 통해 지식 변형과 통찰의 창출에 이른다. 즉, 능숙한 필자는 한 편의 글을 위해 장기기억의 지식, 신념, 경험을 조직해 전략적으로 연결하는 복잡한 과정을 거친다.

반면, AI 활용 LLM의 텍스트 생성은 대규모 말뭉치의 통계적 패턴 예측으로 설명되며 인간의 쓰기 과정에서 개입되는 독자에 대한 인식, 목표 기반 추론, 자기 수정과 같은 주체적 사고 과정이 전제되지 않는다 (Bender, Gebru, McMillan-Major, & Shmitchell, 2021; Durak, Eğin, & Onan, 2025). 다시 말해 인간 쓰기가 사고의 확장과 비판적 성찰, 사회적 상호작용을 매개하는 과정이라면, AI 기반 쓰기는 사용자의 요청에 따라 확률적 최적화에 기반한 결과물을 생성하는 과정이라 할 수 있다.

AI 기반 쓰기의 즉시성과 편의성은 필자의 숙고 과정을 대체하거나 단축한다. 그 결과, 필자가 편의성에 기대어 충분한 내적 처리와 기억 부호화 과정을 생략할 경우, 사고 재구성이 지연되는 상태인 '인지의 부채(cognitive debt)'가 형성될 수 있다. 필자가 이처럼 형성된 부채를 안은 채 LLM이 제시하는 고확률의 무난한 표현을 비판 없이 수용하게 되면, 대안을 탐색하는 숙고 과정은 더욱 줄어들게 된다. 결국 필자의 이러한 수동적 선택이 반복됨에 따라, 서로 다른 필자들의 글이 유사한 어휘와 구조로 수렴하는 언어적 동형화를 초래한다. 이에 따라, 개별 필자의 주도성과 어휘, 구성 방식은 평준화되고 글쓰기 학습 효과와 필자 목소리(voice) 형성에 부정적 영향이 나타날 수 있다.

그동안 국어교육 분야에서 AI 관련 연구는 글쓰기 윤리, 수업 방법·모

형, 성과 지표를 정리하며 실천적 지침을 제공해 왔다 (강동훈, 2023; 곽수범, 2024; 김유민·이경화, 2025; 나은미, 2024; 손달임, 2023; 오선경, 2023; 이미나, 2024). 이 연구들은 교육과정·수업·평가 전반에서 AI 활용을 가능하게 하는 실천적 기반과 이론적 준거를 마련했다는 점에서 의의가 있다.

더 나아가, 최근에는 '사고 도구'로서 AI를 정당화하며 인지의 확장 및 표현 보완 등의 효과가 있다고 밝힌 연구(홍진옥·전기화, 2023; Rivera-Novoa & Duarte Arias, 2025)도 찾아볼 수 있다. 그러나 이러한 관점은 AI가 확장한 외부 인지를 필자 내부 인지 능력의 향상과 등치함을 전제하는 것이다. 그 결과로 수반될 수 있는 필자 내생적 숙고 약화와 전략적 멈춤 감소 같은 인지적 비용의 부정적 영향에 대해 충분히 검증하지 못했다. 또한 AI가 제공하는 어휘사전 가시화·유사어 제안 등 편의 기능이 고확률 표현으로의 수렴을 촉진해 어휘·표현의 다양성 축소와 필자 고유성 희석을 초래할 위험에 대해서도 체계적 논의가 미흡하다.

'인지의 부채'는 글쓰기 과정에서 필자의 독립적 구성 능력을 약화시켜 필자 고유의 목소리를 침식하고 다른 필자와 표면 어휘, 의미 구조가 비슷해지는 언어적 동형화를 초래한다. 쓰기 교육의 핵심 목표는 창의적 사유와 자기 목소리를 구축·표현하는 필자를 양성하는 데 있다. 따라서 쓰기 교육에서 AI 사용을 장려하기에 앞서 'LLM 보조' 글쓰기가 어떤 제약과 비용을 유발하는지 먼저 규명해야 한다.

본 연구에서는 뇌신경 신호를 직접 측정하지 않지만, 그 인과모형이 예측하는 지표적 귀결(표현·의미 수렴, 전환 시 막힘)이 텍스트·행동 수준에서 관찰되는지를 검증한다. 이를 위해 대학생 59명의 글쓰기에서 '필자 숙고' 조건의 글쓰기와 'LLM 보조' 조건에서의 글쓰기를 교차 설계한다. 그리고 글쓰기 산출문의 단어 n-gram 분포와 임베딩 공간상의 군집 구조를 비교한다. 이를 통해 'LLM 보조' 글쓰기가 표면 어휘, 의미 구조의 수렴을 강화하는지, 반대로 '필자 숙고' 조건이 필자 고유성(개별성)을 유지·확대하는지를 정량적으로 판별한다.

다음으로 59명의 대학생 가운데 표집한 6명을 대상으로 'LLM 보조'와 '필자 숙고' 조건의 글쓰기 과정을 화면 녹화기를 활용해 수집·분석한다. 이를 통해 'LLM 보조' 글쓰기에서 '필자 숙고'로 전환 시, 장시간 지연(쓰기 막힘)의 발생 가능성을 검토한다. 이러한 작업을 통해 밝힌 사실은 생성형 인공지능 시대의 글쓰기 교육에서 개별 필자의 목소리(voice)를 보존·강화하는 교수와 평가 설계의 근거를 제공하고 고전 글쓰기 이론의 변화 필요성을 제기할 것이다. 이 연구의 연구 문제는 다음과 같다.

연구 문제 1. 대학생 필자의 '필자 숙고' 조건과 'LLM 보조' 조건의 글쓰기에서 n-gram(표면 어휘) 분포는 어떻게 달라지는가?

연구 문제 2. 대학생 필자의 동일 주제의 글쓰기 과제에서 '필자 숙고' 조건 글쓰기 집단과 'LLM 보조' 글쓰기 집단의 글은 임베딩 공간(의미 구조)에서 서로 다른 군집 구조를 형성하는가?

연구 문제 3. 'LLM 보조 → 필자 숙고' 전환 시 나타나는 쓰기 막힘의 발생 맥락·빈도·지속 시간은 LLM 활용 글쓰기 강도(고·저)가 다른 필자 사이에서 어떻게 달라지는가?

II. 이론적 배경

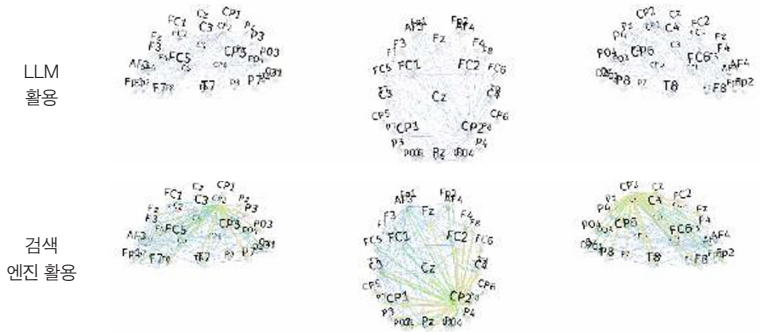
1. 뇌신경 연결성 축소와 인지의 부채

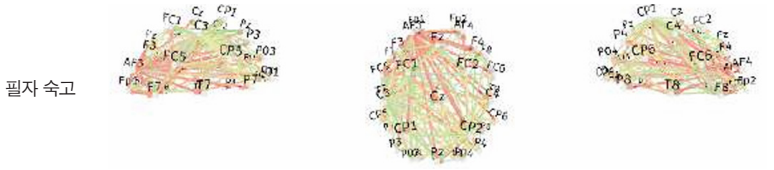
생성형 AI를 활용한 글쓰기는 표면적 유창성과 완결성을 높이는 대신, 의미 구성에 필요한 뇌신경적 조정 과정을 축소하는 경향을 보인다. Kosmyna et al. (2025)는 ① 필자 숙고에 의한 글쓰기, ② 검색 보조 글쓰기, ③

LLM(ChatGPT 기반) 보조 글쓰기를 비교한 뇌 영상 연구에서 LLM 조건에서 필자 간 뇌신경 연결성이 가장 낮다고 보고하였다.

특히 의미 통제와 작업기억을 담당하는 전두-두정 통제망(FrontoParietal Control Network, 이하: FPCN)과 내부 모델 갱신에 관여하는 기본 모드망(Default Mode Network, 이하: DMN)의 동시 활성의 약화가 나타났는데, 이는 글쓰기 맥락에서 예측 부호화와 모델 업데이트 기능의 저하를 시사한다. 이를 작문 교육 관점에서 해석하면, 필자가 '무엇을 쓸 것인가(계획)'와 '어떻게 썼는가(점검)'를 연결하는 인지적 고리가 물리적으로 느슨해졌음을 의미한다. 즉, 뇌신경 연결의 약화는 필자가 스스로 의미를 구성하고 조정하는 '주체적 숙고'의 과정이 생략되고 있음을 보여주는 신경학적 증거이다.

더불어 오류 모니터링과 자기 점검에 관여하는 배외측 전전두 피질(dorsolateral prefrontal cortex, 이하: DLPFC)-전대상 피질(Anterior Cingulate Cortex, 이하: ACC) 간 기능적 결합 저하는 메타인지적 감독 비용을 낮추지만, 필자가 외부 산출물을 비판적으로 선택·수정하는 메타인지적 자기 점검 능력이 저하됨이 관찰되었다. 다시 말해, 전두-두정 통제망 약화로 자기 점검 기능이 떨어지면 한정된 표현에 안주하기 쉬워지고, 다양한 표현을 모색하며 수정하는 표현의 탐색 폭이 줄어들게 된다. 다음 〈그림 1〉은 Kosmyna et al. (2025) 실험에서 각 조건의 글쓰기 중 뇌 영역 간 인과적 연결성(Dynamic Direct Transfer Function, 이하: dDTF)을 나타낸 것이다.





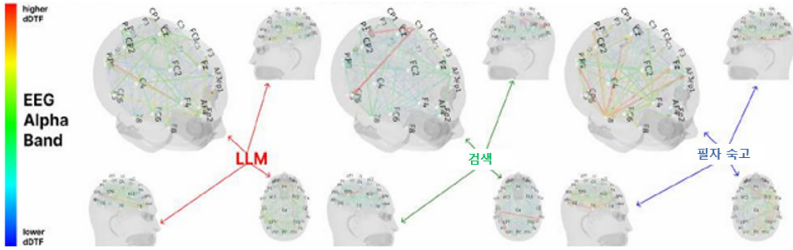
〈그림 1〉 뇌 영역 간 인과적 연결성(Kosmyna et al., 2025)

〈그림 1〉에서 dDTF의 비교 색상은 연결 강도를 나타낸다. dDTF는 시간·주파수 영역에서 한 영역의 활성 변화가 다른 영역으로 방향성 있게 전달되는 영향(유효 연결성)을 추정하는 지표로서 파란색에서 빨간색으로 갈수록 세기가 강함을 표현한다. 'LLM 보조' 조건에서는 전반적으로 연결 강도가 낮고 패턴이 단조로운 반면, '필자 숙고' 조건은 전전두-측두 및 전두-두정 축을 중심으로 높은 연결과 다채로운 경로가 관찰되었다. 이는 의미 조직 과정에서 광범위한 뇌신경 네트워크가 연결되었음을 시사한다. '검색 보조' 조건은 두 조건의 중간적 특성을 보이며 외부 정보 보조가 연결성을 일정 부분 유지하되 LLM만큼의 단조화를 유발하지는 않음을 보여준다.

또한 dDTF 기반 분석은 연결성 저하가 주파수 대역별로 다르게 나타남을 보여준다. '필자 숙고' 조건에서는 알파·베타·세타·델타[1] 전 대역에 걸쳐 전두-두정-후두 축의 유효 연결이 넓고 강하게 분포한다. 반면 'LLM 보조 조건'에서는 동일 축의 하향식 조절 신호가 체계적으로 감소한다. 특히 알파·베타 대역의 전두 → 두정 흐름 약화는 주의 배분, 목표 유지, 작업기억 갱신과 의미 통합을 매개하는 FPCN-DMN 상호작용의 수축을 나타낸다. '검색 보조' 조건은 시각 탐색·평가·통합을 반영하는 후두 → 전두 흐름이 비교적 'LLM 보조 조건'에 비해 유지되는 중간 수준을 보인다. 다음 〈그림2〉는 세 조건에서 알파 대역 dDTF 분석을 예시한다.

---

1) 알파(8-12Hz)는 차분함·집중 상태의 리듬, 베타(13-30Hz)는 생각 정리·문제 해결, 세타(4-7Hz)는 아이디어 떠올림·기억 연결, 델타(0.5-3Hz)는 가장 느린 파로, 넓은 영역을 함께 조율할 때 나타난다.



〈그림 2〉 알파 대역에 대한 dDTF EEG 분석(Kosmyna et al., 2025)

〈그림 2〉를 살펴보면, 외부의 인지 보조가 많을수록 전반 결합 강도가 단계적으로 축소되는 경향이 나타난다. 이는 표면적으로 드러나는 유창성의 이면에서 예측 → 평가 → 오차 수정 → 재표상 → 부호화로 이어지는 내부 순환과 각 절차에서의 정밀도가 저하된다는 것을 의미한다. 쉽게 말해, 능숙한 필자의 뇌는 글을 쓸 때 끊임없이 정보를 인출하고 검증하며 복잡한 '신경학적 도로망'을 구축하지만, LLM을 과도하게 의존하는 필자의 뇌는 이러한 경로를 개척하지 않는다. 필자는 결과물(텍스트)에 빠르게 도달할 수 있을지 모르나, 뇌 내부에는 '도로를 주행한 경험(신경 연결의 강화)'이 남지 않게 되는 것이다. 즉, 외부 보조가 많을수록 필자의 뇌는 더 적은 반복과 낮은 수준의 정교화를 거친 의미를 표상하게 된다. 이에 따라 쓰기 맥락과 필자의 연결이 얕아지게 되어 장기기억의 의미 결속도가 약화하게 된다.

이러한 연결성 저하는 세부 연결 수준에서도 포착되며, 특히 조건 전환 시에 민감하게 나타난다. 'LLM 보조' → '필자 숙고' 조건으로 전환할 때 통상적으로 고난도 글쓰기에서 관찰되는 전두-두정 협응 신호(Fz → P4, AF3→CP6)가 현저히 약화되었다. 이는 글을 쓸 때, 상위 통제 자원과 작업기억의 동시 가동이 충분히 유도되지 못함을 시사한다. 반대로, '필자 숙고' → 'LLM 보조' 전환 시에는 전전두-측두-후두 경로 일부에서 국소 dDTF가 증가하는 경향이 관찰되었으나, 전반 유효 연결성은 낮은 수준을 유지하였다. 이는 특정 영역 사이의 정보 흐름이 부분적으로만 잠시 활발해졌을 뿐, 뇌 전체 수준의 조율과 통합은 회복되지 않았다는 뜻이다. 다시 말해, 국소적

우회로는 생겼지만, 네트워크 전반의 흐름은 여전히 낮게 머문 셈이다.

또한 'LLM 보조' 조건에서 관찰되는 고주파(특히 베타) 조절 회로의 저활성과 FPCN-DMN 상호작용 약화는 DLPFC-ACC로 대표되는 오류 모니터링과 자기 점검 회로의 동시 활성 저하와 연결된다. 그 결과, 과제를 하며 떠올린 핵심 아이디어들이 깊이 기억으로 자리 잡지 못하고, 글이 겉보기로만 그럴듯하게 맞춰지는 수준에 머물 가능성이 커진다. 'LLM 보조' 조건에서는 이러한 신경적 배경과 함께 전략적 멈춤(계획, 점검, 수정 목적)의 빈도와 지속이 전반적으로 감소하는데 이는 상위 통제의 미세조정이 생략되거나 축소되었음을 시사한다.

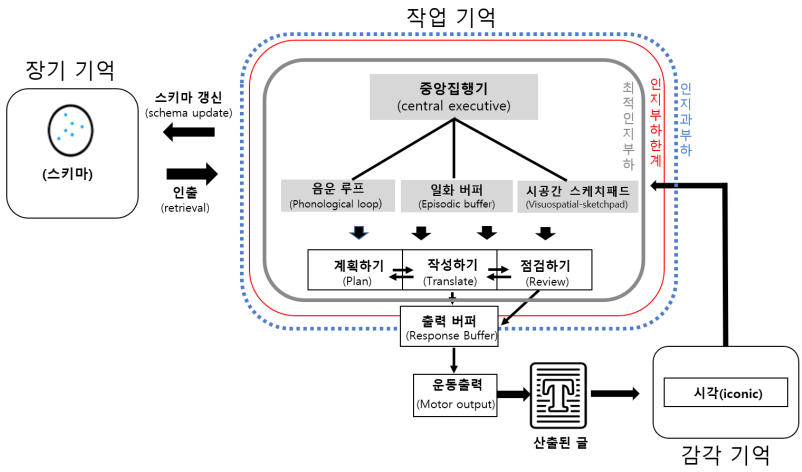
정리하면, 'LLM 보조' 조건에서 관찰되는 뇌신경 연결의 축소는 필자의 주체적인 숙고 과정이 생략되었음을 보여주는 신경학적 증거이다. 즉, 외부 도구에 대한 의존으로 인해 내부의 계획-점검 순환이 가동되지 못하고, 이러한 내적 처리의 미이행분이 학습자에게 누적된 상태가 곧, '인지의 부채'이다. 이는 겉보기 유창성을 얻는 대가로 의미 구성 능력, 어휘력, 문제 재정의 역량이 약화될 위험을 동반한다.

여기서 말하는 '인지의 부채'는 필자가 외부 산출물에 과도하게 의존함에 따라, 스스로 의미를 구성하는 뇌 내부 회로를 충분히 가동하지 않아 발생하는 인지적 미처리 분량을 뜻한다. 그 결과 전략적 멈춤(계획·점검·수정 목적)의 빈도와 지속이 감소하고, 작업기억의 갱신-검증 루프가 축소되어 장기기억으로의 정교한 부호화가 지연된다. 이러한 부채는 과업 전환 시 인출 지연, 상위 통제의 미세 조정 부족, 전개·표현의 상위 어휘·표현의 과집중으로 표면화된다. 그리고 결국 '지금 당장 그럴듯해 보이는 글' 뒤에 남아 있던 내생 처리 미수가 점점 다음 작업으로 이월되는 누적 형태로 나타난다.

2. 인지 외주화에 따른 언어적 동형화

필자의 쓰기 행위는 여러 종류의 기억이 작용하고 순환하는 복잡한 과

정을 통해 이루어진다. 다음 〈그림 3〉은 글쓰기 과정에서 작업기억, 장기기억, 감각 기억의 작용과 인지 부하이론(Cognitive Load Theory)을 결합하여 연구자가 단순화하여 나타낸 것이다.



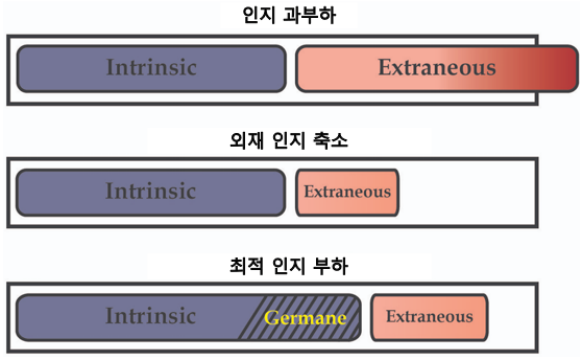
〈그림 3〉 글쓰기 과정에서 기억장치의 작용과 인지 부하

〈그림 3〉은 글쓰기 수행에서 작업기억 내부의 통제-처리-산출-환류로 이어지는 과정을 보여준다. 먼저 중앙집행기(Central executive)가 주의를 배분하고 전략을 설정하면 음운 루프(Phonological loop)·일화 버퍼(Episodic buffer)·시공간 스케치패드(visuospatial sketchpad)[2]가 호출되어 글쓰기 계획-작성-점검의 순환이 작동한다. 작성물은 작업기억 경계의 출력 버퍼(Response Buffer)를 거쳐 운동 출력(타이핑·필기)으로 전환되고 생성된 텍스트는 감각 기억을 통해 다시 작업기억으로 유입되어 시각 피드백 기반의 미세 수정이 반복된다. 동시에 글쓰기 중에는 장기기억의 스키마

2) 음운 루프는 말소리 정보를 잠시 유지·반복하고, 시공간 스케치패드는 시각·공간 정보를 머릿속에서 조작·기억한다. 일화 버퍼는 이 둘의 정보를 통합해 하나의 장면으로 묶고 장기기억과 교환한다.

가 인출되어 관련 정보를 한 덩어리로 묶어 이해와 처리를 빠르게 한다. 점검·수정에서 생긴 새 지식은 스키마로 편입·갱신되어 장기기억에 저장되고, 그 결과 다음 글쓰기에서 주체가 체감하는 내적 인지 부하가 낮아진다.

이때 각 단계는 필연적으로 인지자원을 소모한다. 따라서 인지 과부하를 방지하기 위해 중앙집행기는 과부하 위험을 탐지하면 총 인지 부하를 사전적으로 조정한다(Houichi & Sarnou, 2020). 다음 〈그림 4〉는 이를 보여준다.



〈그림 4〉 최적 인지 부하 과정(Shaffer, 2022)

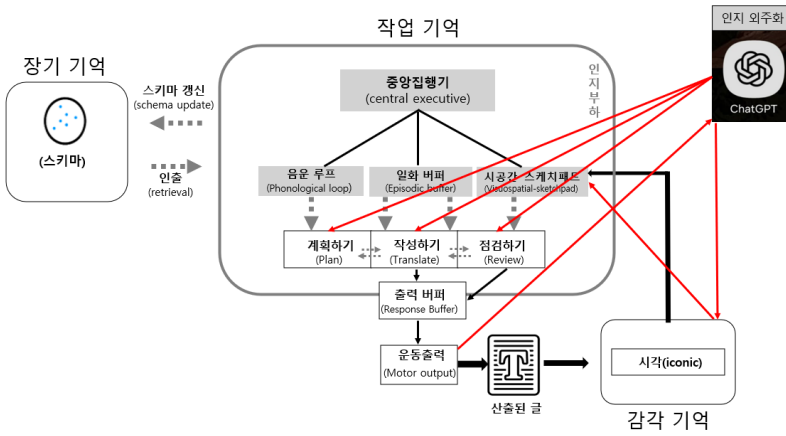
인지 부하는 쓰기 과제의 본질적 요구를 처리하는 내재 인지 부하(Intrinsic Cognitive Load), 쓰기 과제 지시문의 모호함을 처리하는 외재 인지 부하(Extraneous Cognitive Load), 의미 구성·스키마 형성에 투입되는 정교화 인지 부하(Germane Cognitive Load)로 나눌 수 있다. 그런데 인지 과부하 상태가 되면 중앙집행기는 외재 인지를 축소하고 확보한 자원을 활용하여 정교화 인지 부하에 활용한다(Sweller, van Merriënboer, & Paas, 2019). 이렇게 내재·외재·정교화 인지를 최적화한 상태를 최적 인지화라고 한다.

한편, 인지 부하 이론 관점에서 동일한 과제라도 필자별 산출물이 각기 다른 이유는 가용 인지자원의 크기와 배분 방식이 함께 달라지기 때문으로 설명된다. 작업기억 용량의 개인차는 같은 정보량을 처리할 때 요구되는 동시 처리 요소 수를 상이하게 한다. 따라서 작업기억 용량이 작은 필자는 계

획·아이디어 생성·문장화가 직렬적이고 단편적으로 진행되어 아이디어 밀도와 전체 글의 응집성이 떨어지게 된다. 반대로 용량이 큰 필자는 더 많은 요소를 동시에 관리하고 조작하여 주장-근거-설명의 정렬, 반론 예견, 단락 간 논리적 전이를 포함하는 거시 구조를 설계할 여유를 갖는다.

또한 필자 글의 개별성은 스키마의 차이로도 설명된다. 고숙련 필자는 주제·장르·절차 지식을 절차적 스키마로 통합하여 과제의 핵심 요소를 의미 단위로 처리함으로써 내재 인지 부하를 완화한다. 또한 표준화된 계획·검토 루틴을 통해 지시문 해석·자료 탐색에서 발생하는 외재 인지 부하를 최소화한다. 이로써 확보된 인지자원은 정교화 처리로 재배분되어 주장 명료성, 근거의 적합성과 다양성, 논증의 일관성, 어휘·구문 다양도 그리고 수정의 정밀성을 체계적으로 향상시킨다.

반면 'LLM 보조' 글쓰기에서는 자료 요약·구조화·문장화 같은 전처리가 필자의 요구에 따라 즉각적으로 제공된다. 이에 과제의 표면적 난점이 빠르게 해소되면서 외재 인지 처리의 요구가 감소되고 이에 따라 내재 인지 부하 체감이 완화된다. 다음 〈그림 5〉는 글쓰기 과정에서 인지 외주화의 영향 범위와 작용을 연구자가 단순화하여 나타낸 것이다.



〈그림 5〉 인지 외주화의 영향 범위와 작용(빨간선=외부 입력, 회색 점선=순환 약화)

〈그림 5〉를 살펴보면, LLM 보조 글쓰기는 작업기억에 외부 인지 대리인이 병렬 접속되어 계획–작성–점검의 일부 연산을 우회·대체하는 구성으로 이해된다. 중앙집행기의 질문은 운동 출력을 통해 LLM으로 전달되고 응답은 감각 기억을 거쳐 작업기억에 재유입되어 음운 루프·시공간 스케치패드·일화 버퍼의 하위 조작을 부분 대체한다. 이로써 지시문 해석·자료 탐색·문장화의 전환 비용이 줄어 외재 인지 부하와 내재 인지 부하가 낮게 체감(본질 난이도 변화 아님)되고 단기적으로 문장 유창성과 표현 정밀도가 개선된다.

그러나 LLM의 이러한 편의 제공 이면에는, 글쓰기 작업기억 과 장기기억 간 지식 인출과 스키마 갱신 순환의 빈도와 깊이가 낮아지는(그림 5의 회색 점선) 부작용 가능성이 지적된다. LLM의 제안을 수용하여 글을 완성하는 과정이 반복되면 필자가 자기 기억에서 아이디어를 끌어내고 축적된 지식을 재구성하는 기회가 줄어들 수 있다. 실제로 Kosmyna et al. (2025) 실험에서도 LLM 도움을 받은 학생들은 본인이 쓴 문장의 내용을 제대로 기억하지 못하는 경향을 보였다. 이 연구에서 글쓰기 참여 대학생 중 80%에 이르는 참가자가 자신이 작성한 글에서 한 문장도 다시 기억해서 인용하지 못했다.

LLM이 제시하는 확률적으로 평이한 표현을 필자가 별다른 숙고 과정 없이 즉각적으로 채택할 때, 뇌신경 수준의 연결 강화 기회는 상실되고 대안적 표현을 탐색하는 인지적 과정은 축소된다. 필자의 이러한 선택이 누적된 결과, 서로 다른 필자들의 문장이 유사한 어휘와 의미 구조로 수렴하는 언어적 동형화 현상이 나타나게 된다.

그런데 언어적 동형화는 단순히 표현의 유사성을 넘어 필자 고유의 언어적 개성과 정체성을 약화시키는 방향으로 작동한다. LLM이 제공하는 평균화된 표현과 전형적 문체는 개별 필자의 어조·리듬·문장 구조적 선호를 점차 희석시켜, 결과적으로 ‘누구의 글인지 구별되지 않는 글’을 양산한다. 이는 표현의 다양성과 언어 창의성의 기반을 잠식하고, 집단 수준에서는 담화 양식의 단조화와 서술 패턴의 표준화를 초래한다.

더 나아가 창의적 표현을 시도할 때 필자가 느끼는 심리적 부담 즉, ‘LLM이 제시한 방식이 더 정확할 것’이라는 자기검열이 강화되어, 창의적 어휘 선택과 비정형적 문장 구성의 시도가 체계적으로 위축되는 경향이 있다. 결국 언어적 동형화는 필자의 내면적 음성과 글쓰기 세계의 스펙트럼을 좁혀, 인간 필자가 지닌 고유한 사고의 흔적과 표현의 결을 점차 소거시킬 가능성이 있다.

3. 쓰기 막힘

지금까지 작문 연구에서는 쓰기 과정 중, ‘휴지(休止)’를 멈춤과 막힘을 아우르는 중립적 상위 범주로 상정한다. 그리고 그 안에서 멈춤은 의도적·긍정적 기능으로 막힘은 비자발적이고 부정적(서종훈, 2019; 이승주, 2019) 결과를 유도하는 하위 현상으로 구분한 바 있다. 한편, 고리사(2025)는 멈춤을 “필자가 의도적으로 쓰기를 잠시 지연시키며 더 양질의 텍스트 산출을 위해 고민하는 현상(인지·정의적으로 긍정적으로 인식)”으로 막힘을 “연속적 생산이 불가능하고 필자에게 부정적으로 인식되는 현상”으로 대비해 설명하였다. 또한 강동훈(2016)은 학습자로부터 산출된 자료(글자 입력기록·시선추적)를 분석하면서 ‘쓰기 휴지’를 쓰기를 일시적으로 멈추거나 머뭇거리는 현상으로 개념화하되 그 안에 유용한 지연(멈춤)과 어려움으로 작용하는 지연(막힘)이 함께 존재함을 지적한다.

막힘(writer’s block)은 “기본 기능 결여 외의 이유로 쓰기를 시작·지속하지 못하는 상태(Rose, 1984)”, “유창성 결여와 기한 내 집필 실패(Boice, 1990)”, “능력이 있음에도 과제를 완성하지 못하는 상태(Hjortshoj, 2001)”로 기술된다. 특히 Hjortshoj(2001)는 막힘이 특정 지점과 특정 장르에서 선택적으로 발생할 수 있음을 지적하며 그 위치와 맥락을 미시적으로 추적할 것을 제안한다.

이는 과정 모형의 조율 관점에서 더욱 분명해진다. Flower & Hayes

(1981)에 따르면 계획-전사-검토는 상위 모니터의 통제 아래 회귀적 상호작용을 이룬다. 생산적인 멈춤은 바로 이 조율 고리에 의도적으로 투입되어 담화 단위의 재구성, 근거 점검과 국면 전환을 준비하지만, 막힘은 조율 체계가 비동조화되면서 다시 읽기, 삭제의 반복, 시작 지연, 단계 간 단절의 형태로 가시화된다(Matsuhashi, 1981; Olive, 2014; Schilperoord, 1996).

이때, 쓰기 막힘의 주요 원인은 인지적 요인에 해당하는 표현 및 문법 사용의 어려움, 논리 전개의 문제, 아이디어 생성 곤란 내용 조직 등으로 파악되었다(고리사, 2025). 아울러 글쓰기 과제 자체의 난이도나 주제 이해의 정도, 논증 구조와 같은 과제적 요인, 그리고 제한된 시간과 같은 외적 환경적 요인도 쓰기 막힘에 영향을 미치는 것으로 논의되었다. 이러한 다양한 맥락적 요인들은 필자가 유보한 인지적 부담, 즉 '인지의 부채'와도 연결된다.

'LLM 보조'는 문장 생성, 표현 수정, 형식 점검의 일부를 외주화하여 단기적 유창성을 높인다. 그 대가로 필자 내부의 탐색-평가-표상 변환 순환이 얕아지고 표상 재구성을 통한 장기기억 부호화로 이어지는 내생 처리의 누적이 감소할 수 있다. 'LLM 보조'에 의존하며 필자 내부에 누적된 미이행 처리분('인지의 부채')은 필자가 외부 도움 없이 스스로 글을 전개하고 통제해야 하는 '필자 숙고' 환경으로 전환되는 순간 상환 압박으로 작용한다. 이때 필자는 미처 조직하지 못한 사고와 논리를 다시 연결하는 데, 어려움을 겪으며 이는 곧 '쓰기 막힘'으로 표면화된다.

특히, 'LLM 보조' 강도가 높은 일부 필자에게서는 지속·반복 지연의 전형을 보이는 쓰기 막힘으로 이행할 위험이 커진다. 이런 전이가 반복되면 자기 점검·재구성·표상 전환과 같은 핵심 자기조절 과정이 점차 약화되어 'LLM 보조'가 없는 상태에서의 기본 수행 능력이 서서히 저하될 수 있다. 결국 장기적으로는 'LLM 보조'가 없을 때 글쓰기 역량의 하향 안정화가 나타날 가능성을 배제하기 어렵다.

III. 연구 방법

1. 연구 대상자

본 연구의 연구 대상자는 대구광역시 D 교대에서 개설된 「화법교육론」과 「의사소통의 이해」 과목을 수강하는 국어교육학과 학생과 8개 연합 학과 학부생 59명이다.[3] 이들 모두는 연구의 목적과 방법에 관해 사전에 고지받았고 실험 참여에 대해 서면으로 동의하였다. 다음 〈표 2〉는 학부생 59명에 대한 정보이다. 이들은 모두 대학 교양 및 전공 수업에서 다양한 글쓰기 과제(보고서, 학술 에세이, 발표문 작성 등)를 수행한 경험이 있다.

〈표 2〉 연구 대상자 정보

학년	학과	학생 수	ChatGPT 활용 비율
3	국어	23명(남: 9명 여: 14명)	90.5%
3	연합	36명(남: 13명 여: 23명)	88.1%

선행 조사 결과 59명의 학생 중 89.3%(약 53명)가 ChatGPT, 뤼튼 등

3) 선행 뇌신경 연구(Kosmyna et al., 2025)는 자기표현적 에세이 작성 시 활성화되는 뇌신경망 연결이 LLM 보조 시 약화됨을 보고하였다. 본 연구의 과제인 '교사와 학생의 관계 및 의사소통'에 대한 글쓰기 역시 예비교사로서 개인의 가치관과 신념을 정립해야 하는 성찰적 성격을 띠므로, 정보 나열보다는 내면의 숙고와 재구성이 필수적이라는 점에서 Kosmyna 연구의 과제 맥락과 맞닿아 있다.
또한 고리사(2025)는 한국어 학습자(L2)의 쓰기 막힘을 분석하였으나, 본 연구는 '인지의 부채'가 예비교사와 같은 모국어(L1) 숙련 필자에게도 일시적인 인지적 병목을 유발하여 마치 미숙련 필자와 유사한 쓰기 막힘 현상을 재현할 수 있다는 가설에 기반한다. 즉, 글쓰기 교육 친연성이 높은 집단이라 할지라도 도구 의존도가 높아질 때 발생하는 절차적 지식의 비활성화가 실제 수행에서 어떤 막힘 양상으로 전이되는지를 규명하고자 한다.

을 비롯한 대규모 언어모델(LLM)을 글쓰기에 활용하고 있는 것으로 나타났다. 학생들은 주제 탐색, 관련 자료 조사, 문법 및 표현 교정, 글의 구조·전개 방식 추천, 재맥락화, 자신의 글에 대한 평가 등에 LLM을 활용한다고 밝혔다. 따라서 이들은 본 연구의 목적에 부합하는 적절한 연구 대상이라 할 수 있다. 이어 쓰기 막힘을 확인하기 위해 59명 중 6명을 층화된 목적표집(stratified purposive sampling)[4]한다. 우선 사전 설문을 통해 59명의 LLM 활용 양상을 '전면적 활용군(34명)', '제한적 활용군(23명)', '비활용군(2명)'으로 범주화한다.

'전면적 활용군'은 과제 이해–아이디어 발산–구성 설계–문장 생성–수정에 이르는 글쓰기 전 과정에서 LLM을 적극 활용하는 것으로 보고한 학생들로 규정한다. '제한적 활용군'은 LLM을 주제 탐색, 맞춤법 검사, 자신의 글에 대한 평가 등 국소적 보조에 국한하고 논리 전개, 어휘 선택 및 구조 설계는 자기 숙고에 의존한다고 응답한 학생들로 정의한다. '비활용군'은 글쓰기 과정에서 ChatGPT, 뤼튼 등 'LLM 보조'를 전혀 받지 않는다고 응답한 집단으로 정의한다. 이들 세 집단에서 각각 2명의 학생 총 6명을 표집한다. 그리고 글쓰기 과정을 녹화한 후, 분석하여 이들을 '전면적 활용'과 '제한적 활용'으로 구분한다. 이 절차는 2인의 연구자에 의해 수행한다.

2. 연구 절차

본 연구의 실험 절차는 3개의 연구 문제에 따라 세 단계로 구성한다. 실험 ①은 '필자 숙고'와 'LLM 보조' 조건에서 산출된 텍스트의 n-gram 패턴을 비교·분석하여 표면 어휘의 동형화 현상을 비교한다. 실험 ②는 동일 주제 글쓰기에서 두 조건 간 임베딩 공간의 군집 구조와 집단 내·집단 간 평

4) 층화된 목적표집은 사전 기준으로 모집단을 유의미한 하위집단(층)으로 나눈 뒤, 각 층에서 연구 목적에 부합하는 사례를 의도적으로 선별하는 표집 전략이다.

균 코사인 유사도(히트맵)를 비교하여 의미의 수렴을 평가한다. 본 논문에서 '언어적 동형화'는 표면 어휘·구문의 분포가 고확률 표현으로 수렴하는 현상과 의미적 수렴을 뜻하며 두 지표는 상관될 수 있으나 동일 개념이 아니므로 해석 범위는 각각의 측정수준에 맞게 지표적(표면 vs 의미 공간)으로 제한한다. ③은 'LLM 보조'에서 '필자 숙고'로 전환할 때 나타나는 쓰기 막힘을 심층 분석한다.

참여자는 '국어교육과' 23명과 '연합과' 36명(계 59명)으로 구성되었으며, '국어교육과'는 세션 1 → 2 → 3, '연합과'는 세션 2 → 1 → 3 순으로 과제를 수행한다. 과제 순서를 집단 간, 다르게 배치한 이유는 동일 주제의 과제를 '필자 숙고'와 'LLM 보조' 조건에서 수행할 때 임베딩 공간 구조가 어떻게 달라지는지(연구문제 2) 교차 비교하기 위함이다. 과제는 각 과제마다 글자 크기 11, A4 1쪽 분량(지침 준수)으로 작성하도록 하였으며 시간은 전체 과제 작성을 위해 100분이 주어졌다. 과제는 다음과 같으며 각 과제 1, 2, 3의 내용을 포함하여 세션 3에서는 한 편의 서론·본론·결론을 갖춘 글을 작성하도록 하였다.

과제 1: "교사와 학생의 좋은 관계란 무엇인가?"에 대한 자신의 생각을 작성하세요.

과제 2: "교사와 학생이 좋은 관계를 형성해야 하는 이유는 무엇인가?"에 대한 자신의 생각을 작성하세요.

과제 3: "교사와 학생이 좋은 관계를 형성하기 위해 교사가 갖추어야 할 의사소통은 무엇인가?"에 대해 작성하세요.

세션 1과 세션 3에서는 어떠한 보조 도구도 사용하지 않는 '필자 숙고' 조건으로 세션 2에서는 ChatGPT를 활용한 'LLM 보조' 조건으로 동일한 유형의 과제를 수행한다. 〈표 4〉는 세션별 글쓰기 조건과 과제이다.

〈표 4〉 세션별 글쓰기 조건과 과제

구분	세션 1	세션 2	세션 3
조건	필자 숙고	LLM 보조(ChatGPT)	필자 숙고
과제 배치(국어)	과제 1	과제 2	과제 3
과제 배치(연합)	과제 2	과제 1	과제 3

실험 ①인 n-gram 패턴 분석은 다음과 같은 절차를 따른다. 먼저, 각 세션 산출물을 전처리한다. 전처리는 과제 지시어, 형식표지 등 최소 불용어만 제거해 내용어 손실을 억제하고, 합성어와 고유명은 원표기를 유지하였다. 전처리 규칙은 엑셀 자료에서 임의 추출한 5개 문서로 사전 검증한 뒤 전체 코퍼스에 일괄 적용하였으며, 세션별로 단어 n-gram(1-4-gram)[5] 빈도를 집계해 출현 빈도 ≥3 항목만 분석에 포함하였다. 세션 간 n-gram 분포를 비교하고 '필자 숙고'(세션 1)에서의 다양성과 'LLM 보조'(세션 2)에서의 수렴성을 '국어교육과'와 '연합과' 별로 분석하여 'LLM 보조' 글쓰기의 언어적 동형화(표면 어휘 수렴) 효과를 탐색한다.

실험 ②인 임베딩 군집 구조 분석은 다음과 같은 절차를 따른다. 전처리는 실험 ①에서 수행한 표기 통일·기호 및 공백 정규화 절차를 공유하되, 불용어 제거와 n-gram 필터링은 임베딩 품질 저하를 방지하기 위해 적용하지 않는다. 이어 세션 1, 2의 과제 1, 2 산출물을 문장 단위로 분할하고 Sentence-BERT(SBERT)[6]로 문장 임베딩을 산출한다. 문장 임베딩은 길이/TF-IDF[7] 가중 평균으로 문서 임베딩으로 집계한 뒤 L2 norm으로 정규화한다. 다음으로, 정규화된 임베딩을 PaCMAP으로 2차원에 투영하여 '국어교

5) '1-4-gram'은 텍스트에서 연속된 단어 n개 묶음(n-gram) 중 n=1~4를 뜻한다(1=유니그램, 2=바이그램, 3=트라이그램, 4=쿼드그램).

6) 문장을 벡터로 변환해 의미 유사도를 계산하는 문장 임베딩 모델이다.

7) 단어의 빈도와 희귀성을 함께 고려해 텍스트의 핵심어를 가중치로 표현하는 통계적 방법이다.

육과'의 '필자 숙고'(과제 1) · 'LLM 보조'(과제 2)와 '연합과'의 'LLM 보조'(과제 1) · '필자 숙고'(과제 2)를 대비시켜 임베딩 공간의 점군 구조를 비교하여 산출 텍스트의 의미 응집 · 분리를 구조적으로 살펴본다.

또한 조건별로 문서 임베딩 간 코사인 유사도를 계산하여 필자 숙고'와 'LLM 보조' 글쓰기에서 집단 내에서 유사도 행렬을 구성하고 이를 히트맵으로 제시한다. 이를 통해 앞서 살펴본 응집 · 분리의 구조분석을 행렬 수준의 정량 근거로 보완한다.

실험 ③인 쓰기 막힘 심층 분석은 다음과 같은 절차를 따른다. 참여자 6명의 세션 3 글쓰기 전 과정을 ShareX로 화면 녹화한다. 다음으로 연구자-참여자가 함께 영상을 보며 회상적 사고구술과 회상 자극 면담을 실시하여 쓰기 막힘 빈도와 지속 시간을 분석한다. 마지막으로 녹음된 모든 음성을 네이버 클로바 노트로 전사하고 NVivo 15로 코딩하여 쓰기 막힘의 발생 맥락을 체계적으로 수집하고 분석한다. 본 연구의 실험 과정을 정리하면, 다음 〈표 5〉와 같다.

〈표 5〉 글쓰기 실험 과정

실험	핵심 질문	분석 세션	비교	
① n-gram 분석	LLM 보조가 언어적 동형화(수렴)를 유발하는가?	세션 1, 2	과제 1, 2	국어과 '숙고' vs 'LLM' 연합과 '숙고' vs 'LLM
② 임베딩 군집 구조 분석	LLM 보조 글쓰기에서 표현의 군집화 양상은?	세션 1, 2	과제 1	국어과 '숙고' vs 연합과 'LLM'
			과제 2	국어과 'LLM' vs 연합과 '숙고'
③ 쓰기 막힘	LLM 보조 → 숙고에서 '막힘'의 맥락은?	세션 3	과제 3	전면적 활용 3인 vs 제한적 활용 3인

3. 분석 방법

실험 ①에서는 연구 대상자가 생산한 텍스트의 표면적 어휘 유사성을

분석한다. 세션 1, 2에서 산출된 텍스트의 단어 n-gram 분포에 근거하여 언어 동형화 지표를 산출한다. 동형화의 구분은 (a) 상위 k개 빈출 n-gram 중복률(Top-k, k=50), (b) 상위 r개 항목의 누적 사용 비중을 추적하는 누적 커버리지 곡선 비교로 구성한다.

상위 50개 빈출 n-gram 중복률은 같은 세션에서 참가자 *i*, *j*의 상위 50개 n-gram 집합 T_i, T_j에 대해 $|T_i \cap T_j|/50$으로 모든 쌍에 평균해 추정하며, 값이 클수록 표현 선택의 수렴(동형화)이 강함을 의미하는 것으로 판단한다. 누적 커버리지 곡선은 세션별로 모든 문서의 빈도를 합산해 내림차순 정렬한 분포{f(m)}서 상위 *r*개 항목의 누적 비중 $C(r)=\Sigma^{r}_{m=1}f(m)/\Sigma_{m}f(m)$(r=1...50)을 추적하여 상위 표현에 대한 집중도를 시각화하는 것이다.

세션 간 비교는 문서 단위 부트스트랩(1,000회)로 95% 신뢰구간을 추정하고, 필요시 문서 길이와 주제 고정효과를 공변량으로 포함해 편향을 점검한다. 마지막으로 'LLM 보조' 조건에서 Top-k 중복률이 상승하고 누적 커버리지 곡선의 기울기가 증가할수록 언어적 동형화가 심화된 것으로 판정한다(Gries, 2024).

실험 ②에서는 연구 대상자가 생산한 텍스트의 의미적 수렴 현상을 살펴본다. 먼저, 문장 임베딩을 문서 단위로 집계하여 L2 정규화한 뒤 코사인 거리를 정의하고, PaCMAP 2차원 투영으로 시각화한다. 분리와 응집은 코사인 거리 기반 Silhouette 계수와 참가자 단위를 보존한 PERMANOVA 퍼뮤테이션 검정(α=.05)으로 평가한다. Silhouette이 유의하게 상승하고 PERMANOVA에서 조건 효과가 확인되면, 두 조건이 임베딩 공간에서 상이한 군집 구조를 형성한 것으로 해석한다(Anderson, 2001).

또한 조건별('LLM 보조', '필자 숙고') 문서쌍 코사인 유사도로 집단 내부 유사도 행렬을 구성하여 히트맵으로 제시한다. 동형화(수렴) 정도는 다음 아래 수식으로 판단한다.

(i) 집단 평균 유사도

$$\bar{s}_c = \frac{2}{n_c(n_c-1)} \sum_{i<j} \frac{\mathbf{e}_i \cdot \mathbf{e}_j}{|\mathbf{e}_i||\mathbf{e}_j|}$$

(ii) 집단 내부 유사도의 표준편차(분산 지표)

(iii) 임계 초과 비율(유사도 ≥ .80 문서쌍 비율)

결과 해석 시, 집단 평균 유사도가 크고 표준편차가 작으며 임계 초과 비율이 높으면 언어적 동형화가 강화된 것으로 해석한다.그런데 임베딩 산점도는 고차원 의미 공간의 형태적 윤곽을 직관적으로 제시하지만, 2차원 투영 특성상 쌍별 관계의 손실과 투영 왜곡이 불가피하다. 이에 본 연구에서는 산점도와 함께 코사인 유사도 히트맵을 병기하여 텍스트 사이 관계(유사·비유사)를 입증한다.

실험 ③에서는 세션 3에서의 녹화 화면을 연구자와 연구 대상 학생이 함께 보며 회상적 사고구술 및 회상 자극 면담을 수행한다. 녹화 화면에서 키보드 입력이 없고 커서가 5회 깜빡임이 발생할 때마다 연구자가 핵심 질문을 제시하는 방법[8]을 활용하여 쓰기 막힘을 판정한다. 이어 회상적 사고구술과 회상 자극 면담 전사본을 Nvivo 15를 활용하여 질적으로 코딩하여 발생 맥락을 도출한다. 코딩 신뢰도는 두 연구자의 독립 코딩 결과에 대한 Cohen's κ로 검증한다. 분석의 초점은 LLM 보조 활용'LLM 보조'에서 '필자 숙고'로의 전환 국면에 나타나는 인지적 휴지 양상을 규명하는 데 둔다.

본 연구의 모든 텍스트 전처리, 통계, 시각화는 Python 3.13에서 구현하였으며, 한국어 형태소 분석과 토크나이징은 연구 설계에 따라 형태소/어절 단위를 병행하였다.

---

8) 고리사(2025)에서는 3회의 깜빡임이 발생 시, 핵심 질문을 제시하였지만 본 연구에서는 사전 텍스트 결과를 바탕으로 5회로 정하였다.

IV. 연구 결과

1. 표면적 어휘 유사성 분석

n-gram 분석은 언어 동형화 중 텍스트의 표면적 어휘 유사성을 살펴볼 수 있다. 이를 위해 세션 1과 2에서 대학생 필자가 작성한 글을(a) 상위 50개 빈출 n-gram 중복률, (b) 상위 r개 항목의 누적 사용 비중을 추적하는 누적 커버리지 곡선으로 살펴보았다.

1) n-gram 중복률

국어과(23명)와 연합과(36명) 각각에 대해 세션 1·2별로 해당 학과 학생의 텍스트에서 각자 최빈출 단어 50개(unigram Top-50)를 추출하였다. 다음으로, 같은 학과 학생 쌍의 Top-50 교집합 비율($|A \cap B| / 50$)을 모두 계산하여 중복률을 도출하였다(〈표 5〉).

〈표 5〉 n-gram 중복률 평균

학과	필자 숙고(세션 1)	LLM 보조(세션2)	세션1 → 2변화
	중복률(SD)	중복률(SD)	중복률(SD)
국어과	0.04(0.01)	0.13(0.02)	-0.09**
연합과	0.05(0.02)	0.16(0.03)	-0.11**

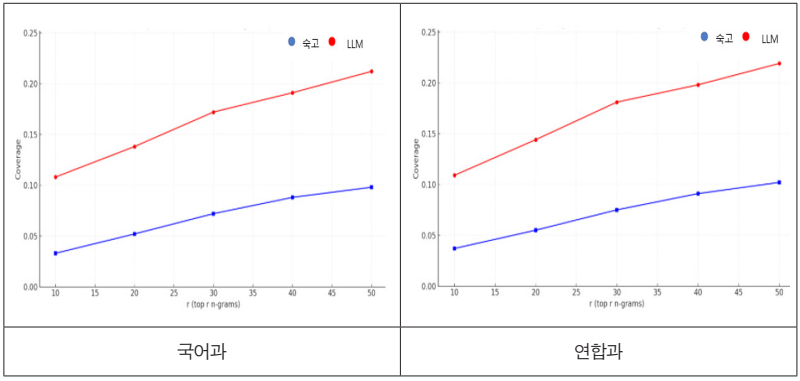
〈표 5〉를 살펴보면, 국어과와 연합과 모두 세션별 학생 간 Top-50 n-gram 중복률이 'LLM 보조' 글쓰기에서 유의하게 높은 것으로 나타났다. 이는 'LLM 보조' 조건의 글쓰기에서 집단 내부의 상위 표현(어휘) 선택이 현저히 수렴함을 보여준다.

2) 누적 커버리지 곡선 비교

누적 커버리지 곡선은 한 집단의 통합 코퍼스에서 상위 r개 n-gram이 전체 토큰에서 차지하는 비중을 r=1…50까지 누적해 그린 것이다. 곡선이 가파르고 위쪽일수록 소수의 상위 표현에 사용이 집중되어 동형화 경향이 크다는 뜻이다. 본 연구에서는 유니그램(unigram) 기준으로, 모든 학생 글을 전처리한 뒤 빈도를 통합 집계하여 상위 r개 항목의 누적 점유율을 계산하고 이를 활용해서 누적 커버리지 곡선을 도출하였다. 다음 〈표 6〉은 상위 표현 개수에 따른 누적 커버리지이며 〈그림 7〉은 누적 커버리지 곡선이다.

〈표 6〉 상위 표현 개수에 따른 커버리지 누적량

학과	세션(조건)	r=10	r=20	r=30	r=40	r=50
국어과	세션 1(필자 숙고)	0.033	0.052	0.072	0.088	0.098
	세션 2(LLM 보조)	0.108	0.138	0.172	0.191	0.212
연합과	세션 1(필자 숙고)	0.037	0.055	0.075	0.091	0.102
	세션 2(LLM 보조)	0.109	0.144	0.181	0.198	0.219



국어과 연합과

〈그림 7〉 누적 커버리지 곡선

〈표 6〉, 〈그림 7〉를 살펴보면, '필자 숙고' 조건보다 'LLM 보조' 조건에서 커버리지 누적량이 더 크다는 것을 알 수 있다. 또한 누적 커버리지 곡선

이 더 가파르고 위쪽에 위치하는 것을 확인할 수 있다. 이는 'LLM 보조' 글쓰기 조건에서의 텍스트가 소수 상위 표현이 집중되어 어휘 분산이 축소되었다는 것을 나타낸다. 정리하면, 두 지표(학생 간 Top-50 중복률, 누적 커버리지 곡선) 모두에서 'LLM 보조' 조건이 '필자 숙고' 대비 상위 어휘의 집중도 증가와 분산 축소를 일관되게 보였다. 이는 LLM 활용이 텍스트의 어휘 선택을 동형화하는 경향을 실증하며, 쓰기 과제 수행 시 다양성 개별성 유지를 위한 교육 방법이 필요함을 시사한다.

2. 의미적 수렴 비교

동일 주제에서 '필자 숙고'와 'LLM 보조'가 생성한 텍스트의 의미적 수렴을 비교하기 위해 과제 배치를 교차 설계하였다. 과제 1은 국어교육과가 '필자 숙고' 조건에서, 연합과가 'LLM 보조' 조건에서 작성하였고 과제 2는 반대로 국어교육과가 'LLM 보조', 연합과가 '필자 숙고' 조건에서 작성하였다. 다음 〈표 7〉은 임베딩 기반 집단 차이 검증 결과이다.

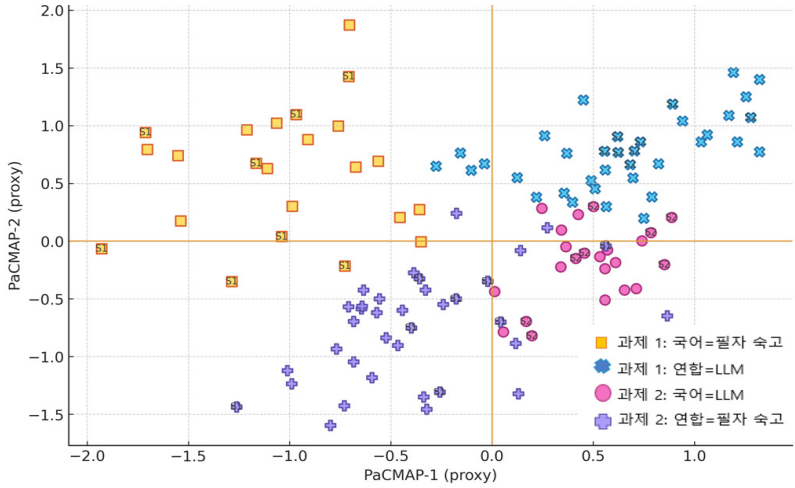
〈표 7〉 집단 차이 검증 결과

분석 대상	비교집단 (수준)	Silhouette (cosine)	PER MANOVA F	R^2	p (perm=1,999)	퍼뮤테이션 층화 단위
전체 주효과	조건×학과 (4수준)	0.580	2978.818	0.987	0.000***	연구 대상자
조건 효과	숙고 vs LLM (2수준)	-	263.101	0.694	0.000***	연구 대상자
학과 효과	국어 vs 연합 (2수준)	-	0.17	0.002	0.999	연구 대상자

〈표 7〉에서 코사인 거리를 기준으로 계산한 Silhouette[9] 계수=0.580

9) 실루엣 계수는 -1~1 범위이며, 일반적 해석으로 0.70 초과=강한 분리, 0.51-0.70=양호로

으로 집단 내 응집도와 집단 간 분리도가 양호하게 나타났다. PERMANOVA(퍼뮤테이션 1,999회, 층화 단위=연구 대상자) 결과, 전체 주효과가 유의했으며($F=2978.818$, $R^2=0.987$, $p=0.000$), 이는 임베딩 공간에서 네 집단(숙고·국어, LLM·국어, LLM·연합, 숙고·연합)의 구조적 차이가 통계적으로 뚜렷함을 의미한다. 대비 분석에서 조건 효과(숙고 vs LLM)가 유의하였고($F=263.101$, $R^2=0.694$, $p=0.000$), 학과 효과(국어 vs 연합)는 유의하지 않음이 확인되었다($F=0.17$, $R^2=0.002$, $p=0.999$). 〈그림 8〉은 조건×학과별 문서 임베딩의 PaCMAP[10] 2차원 투영 시각화이다.



〈그림 8〉 2차원 투영 시각화

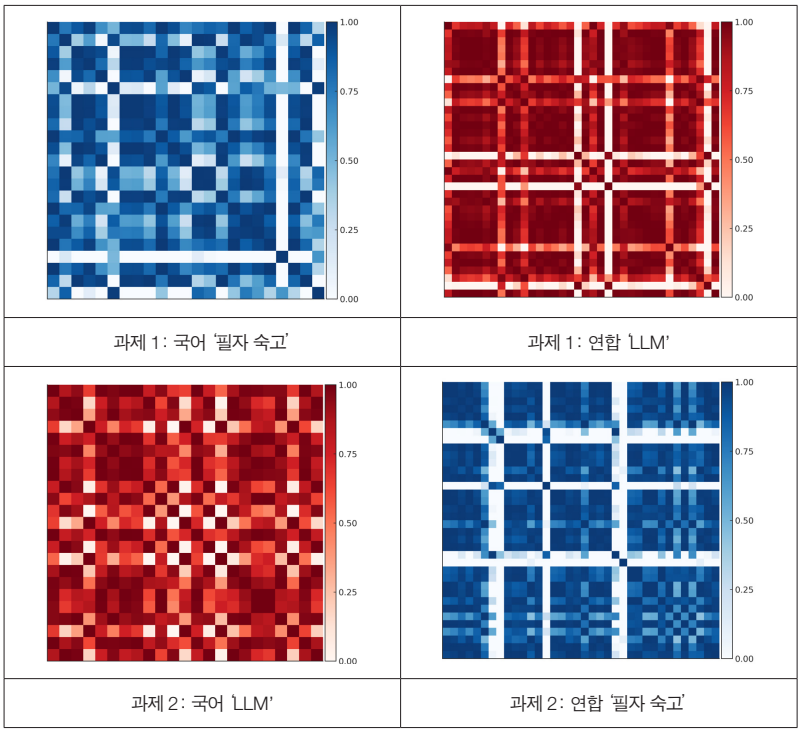
〈그림 8〉을 살펴보면, 같은 과제를 '필자 숙고'와 'LLM 보조' 조건에서 작성했을 때 두 조건 간 임베딩 분리가 명확히 관찰되었다. 특히 필자 숙고 집단은 집단 내 분산이 상대적으로 커, 부분적 이질성이 드러난 반면, LLM

판단한다(Yulisasih, Herman, & Sunardi, 2024).

10) 고차원 문장 의미를 2차원에 보존해 배치하는 투영법으로 점이 가까울수록 의미가 비슷함을 나타낸다.

보조 집단은 서로 가까이 응집하는 경향을 보였다. 이러한 패턴은 LLM 보조 글쓰기에서 언어적 동형화(의미 구조의 수렴)가 발생함을 시사한다.

히트맵은 앞서 산출한 문서 임베딩에 대해 코사인 유사도를 계산하여 구성하였다. 구체적으로 문서 i, j에 대해 $S_{ij}=v_i^T v_j$로 정의하고 이를 NxN 유사도 행렬로 배열하였다. 행·열 순서는 조건(숙고·LLM)×학과(국어·연합) 기준으로 고정하였다. 통계적으로는 동일한 층화 단위(연구 대상자)를 유지한 퍼뮤테이션 검정으로 '집단 내 vs 집단 간' 평균 유사도 차이를 평가하였고, 다중비교는 BH 보정으로 조정하였다. 다음 〈그림 9〉는 동일 주제 글쓰기에서 과제 1(국어·필자 숙고, 연합·LLM)과 과제 2(국어·LLM, 연합·필자 숙고)를 각각 분리하여 산출한 문서 임베딩 코사인 유사도 히트맵이다.



〈그림 9〉 코사인 유사도 히트맵

과제 1에서는 연합과 'LLM 보조' 조건이 평균 유사도 ≈0.72로 비교적 응집된 반면, 국어과 '필자 숙고' 조건은 ≈0.60으로 이질성이 크게 나타났다. 과제 2에서도 국어 'LLM 보조' 조건이 ≈0.75로 가장 높은 응집을 보였고, 연합과 '필자 숙고' 조건은 조정 결과 ≈0.61로 낮은 응집을 보였다. 전반적으로 LLM 보조 조건의 히트맵은 진한 색이 넓게 분포하는 반면, 필자 숙고 조건은 옅은 색의 비중이 커 의미적 수렴(동형화)이 상대적으로 약함을 시사한다. 이는 앞서 제시한 임베딩 산점도의 분리·응집 패턴을 행렬 수준의 정량 근거로 보완해 주며, 'LLM 보조'를 통한 글의 수렴 현상과 '필자 숙고'에서의 글의 이질성에 대한 해석을 강화한다.

3. 쓰기 막힘

세션 2 녹화 영상 분석을 통해 LLM 활용 강도(고·저)에 따라 필자 6명을 다음 〈표 8〉과 같이 분류하였다.

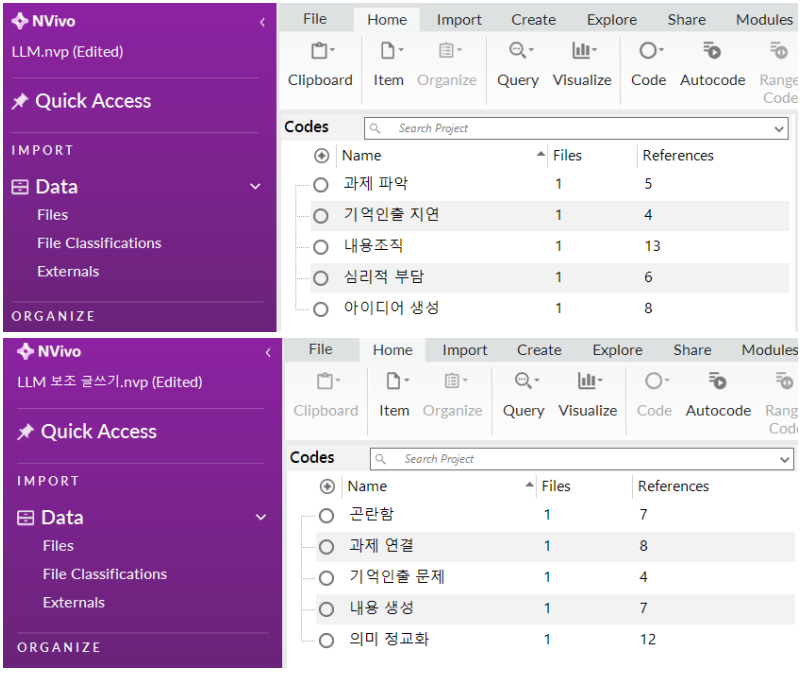
〈표 8〉 쓰기 막힘 연구 대상자

학년	학과	성별	대상자	ChatGPT 활용 양상
3	국어	남	P1	제한적 활용: 근거 조사, 작성 글 평가, 예시 생성
	사회	여	P2	
	수학	남	P3	
	국어	여	P4	전면적 활용: 개요 작성, 내용 생성, 전체 글 재작성
	과학	여	P5	
	국어	남	P6	

이를 토대로 'LLM 보조' 글쓰기 이후 '필자 숙고' 조건으로의 전환 시, 인지의 부채에 의한 쓰기 막힘 발생 여부를 살펴보았다.

1) 쓰기 막힘 맥락 코딩

쓰기 막힘 발생 맥락의 분석은 두 연구자가 Nvivo 15를 활용하여 독립적으로 수행하였으며 1차 코딩 후, 코딩 관련 맥락을 지속 협의하였으며 분석 과정에서 반복 코딩 후 신규 범주가 더 이상 출현하지 않아 주제 포화에 도달하였음을 확인하였다. 다음 〈그림 10〉은 두 코더(연구자)의 쓰기 막힘 발생 맥락의 최종 코딩 결과이다.



〈그림 10〉 최종 코딩 결과

최종 코딩 결과를 바탕으로 두 연구자는 서로 일치하는 코딩 범주와 내용을 확인하고 범주명을 통일하였다. 범주명은 ①내용 조직, ②논리성, ③과제 이해, ④기억인출지연, ⑤심리적 부담으로 재구성하였다. 이는 논리 전개의 문제, 아이디어 생성 곤란, 과제에 대한 이해 문제로 인해 막힘이 발생한다는 고리사(2025)의 연구와 상당히 일치하는 결과였다.

두 코더 간 쓰기 막힘 맥락의 코딩이 완전 일치하는 부분은 28개였으며 불일치하지만, 쓰기 맥락으로 코딩된 사례는 2개로 총 30개였다. 그리고 두 코더 중 한명이 쓰기 막힘으로 코딩하지 않은 것이 A 코더 6개와 B 코더 3개로 총 9개였다. 이 9개 사례는 코드 경계 및 맥락 전환을 탐색하는 별도 분석에 활용하였다. 다음〈표9〉는 두 코더의 범주별 코딩 일치 분할표이다.

〈표 9〉 범주[11]별 코딩 일치

	코더B:①	코더B:②	코더B:③	코더B:④	코더B:⑤	미코딩	합계
코더A:①	13	0	0	0	0	3	16
코더A:②	0	7	1	0	0	2	10
코더A:③	0	1	4	0	0	0	5
코더A:④	0	0	0	3	0	1	4
코더A:⑤	0	0	0	0	2	0	1
미코딩	1	1	1	0	0	0	3
합계	14	9	6	2	2	6	39

코더 간 명목 코딩의 일치 수준은 Cohen's $\kappa = 0.66$[12]로 나타나 '양호한 일치 경계'에 해당하였다. 관찰 일치도 $Po = 0.74$, 우연 일치도 $Pe = 0.24$이었으며 보조지표인 $AC1$[13] $= 0.73$ 역시 유사한 수준을 보여 범주 불균형의 영향을 감안하더라도 코딩의 일관성이 충분히 확보된 것으로 나타났다. 이상의 신뢰도 절차와 결과는 본 연구에서 보고하는 막힘 관련 핵심 지표의 추정이 코더 간 변동성에 과도하게 민감하지 않다는 것을 보장한다.

---

11) 범주명은 ①내용 조직, ②논리성, ③과제이해, ④기억인출지연, ⑤심리적 부담이다.

12) 두 코더의 분류 일치도를 우연 일치 보정까지 반영해 0~1로 나타낸 지표로 0.81 −1.00은 매우 양호, κ=0.61 −0.80은 양호, 0.41 −0.60은 보통 수준의 일치로 본다(Altman, 1991).

13) 희귀 범주에 덜 민감한 대안 일치도 지표로 일반 해석은 κ와 유사하게 0.61 −0.80=양호, 0.81 −1.00=매우 양호로 본다.

2) 쓰기 막힘 맥락과 원인

쓰기 막힘 맥락을 확정하기 위해 코더 A와 B가 막힘 맥락으로 코딩을 했지만, 서로 다른 범주로 코딩한 경우(예, 코더 A는 ③내용 생성으로 코딩하고 코더 B는 ②과제 이해로 코딩)에는 두 연구자의 합의로 하나의 범주에 포함하도록 하였다. 이를 통해 'LLM 보조' 글쓰기에서 '필자 숙고' 조건으로 전환 시, 발생하는 쓰기 막힘의 맥락을 모두 도출하였다. 다음 〈표 10〉은 쓰기 막힘 맥락과 원인이다.

〈표 10〉 쓰기 막힘 맥락과 원인

분류	발생 맥락	원인
인지적	내용 조직	미이행 구조화 작업이 누적
	논리성	필자 내부의 논리 사슬 점검 및 빈틈 메우기 작업이 뒤로 밀림
	과제 이해	미결정된 과제 요소의 부채
	기억인출지연	필자 내부의 미인출 상태의 개념들이 누적
정의적	심리적부담	글의 통제감 저하

'LLM 보조' 조건의 글쓰기에서는 문단 구조와 전개 순서의 내면화가 충분히 이루어지지 않는다. 이로써 담화 스키마가 완결되지 않은 채 남고, '필자 숙고' 조건 전환 시 문장·문단 접속에서 병목이 발생한다. 이에 '내용 조직' 시, 쓰기 막힘이 발생하게 된다. 다음은 연구 대상자 P4와 P1의 인터뷰 내용이다.

> "GPT가 문장을 거의 완성해 주면, 제 사고에서 나온 구조가 아니라서 나중에 이어 쓸 때 전개와 표현을 다시 조직하는 것이 막혀요."(P4)
> "GPT의 의견만을 받아 놓은 글에 제가 쓴 것 같지 않은 글에다가 이제 제 의견을 첨언을 해야 되는 상황이 와서 굉장히 뭔가 막히는..."(P1)

‘논리성’ 측면에서도 주장-근거 정합성 판단과 반증 처리 등 비판적 추론이 뒤로 밀려 검증 과업이 부채화되며 이는 ‘필자 숙고’ 전환 시, 인과적 연결 지점에서 막힘이 반복되는 형태로 나타난다. 다음은 연구 대상자 P2과 P4의 인터뷰 내용이다.

> “GPT가 제시한 주장이나 근거가 제 이야기가 아니다 보니 이어서 글의 논리성을 확보하기가 어려웠다고 생각이 들었어요.”(P2)
> “뒷받침 문장의 관계에 대해 주의 깊게 살펴보지 않아서인지 다시 문단을 읽어볼 때 눈에 걸리는 부분이 더러 있었습니다.”(P4)

‘LLM 보조’ 글쓰기에서는 ‘과제 이해’ 측면에서 주제·범위와 같은 과제 표상 요소가 미결정 상태로 남는다. 그 결과 과제 요구와 산출물 간 정합성 판단이 전환 시점에 집중적으로 요구되며, 새로운 내용 전개 진입 지연과 개요 재편으로 표면화된다. 다음은 연구 대상자 P5와 P6의 인터뷰 내용이다.

> “세 번째 질문을 글의 어디에 위치시켜야 할지 고민하였습니다.”(P5)
> “범위를 제가 직접 정한 게 아니라서 어디까지 써야 할지 기준을 못 잡고...”(P6)

‘기억 인출 측면’에서는 ‘LLM 보조’ 단계에서 표현·어휘·근거 생성이 외주화되면서 장기기억에서 작업기억으로의 인출 경로가 약화된다. 이에 개념들이 미인출 상태로 누적되고, 필요한 표현을 즉각 호출하지 못해 커서 정지와 대체 표현 탐색이 증가한다. 다음은 연구 대상자 P5의 인터뷰 내용이다.

> “여기 ‘누적성’이라고 적었는데 정확한 단어가 바로 떠오르지 않아 계속 망설였습니다.”(P5)

마지막으로 ‘심리적 부담’의 측면에서 ‘LLM 보조’와 ‘필자 숙고’ 텍스트 간 질적 간극 인식이 자기 글에 대한 저자성, 통제감을 저하시켜 심리적 부

담으로 막힘이 발생한다. 다음은 연구 대상자 P6의 인터뷰 내용이다.

"앞의 여러 부분을 수정하며 '이걸 왜 못봤지'라는 생각을 했습니다."(P6)

지금까지 전체 연구 대상자의 쓰기 막힘 맥락과 원인을 살펴보았다. 이어서 'LLM 보조' 저활용 필자와 고활용 필자의 쓰기 막힘 빈도와 지속시간을 비교하였다.

3) 'LLM 보조' 저활용·고활용 필자 비교

다음 〈표 11〉은 'LLM 보조' 저활용·고활용 필자 6인의 쓰기 막힘의 인지적 원인 및 지속 시간이다.

〈표 11〉 쓰기 막힘의 인지적 원인 및 지속 시간

대상자	'LLM 보조'	쓰기 막힘 맥락	빈도	시간(평균)
P1	저활용	내용조직	3	44초
		논리성	1	
		과제이해	2	
P2		-	0	0초
P3		내용조직	3	31초
		논리성	1	
P4	고활용	내용조직	4	27.5초
		논리성	3	
		기억인출지연	1	
P5		내용조직	3	55초
		논리성	3	
		과제이해	2	
		심리적 부담	1	
P6		과제이해	1	49초
		기억인출지연	2	

〈표 11〉을 보면, ‘LLM 보조’ 저활용군(P1-P3)은 전반적으로 인지의 부채가 표면화되는 막힘의 빈도와 범주 폭이 좁게 나타난다. P1은 내용 조직(3회)과 과제 이해(2회), 논리성(1회)에서 국지적 병목을 보였으나, P2는 관찰 기간 동안 막힘이 기록되지 않았다. P3 역시 내용 조직(3회)과 논리성(1회)에서만 제한된 막힘이 나타나, 저활용군의 막힘은 구조·논증 결정을 미루다 전환 시점에 국소적으로 상환되는 양상에 가까운 것으로 해석된다. 이는 LLM 보조 단계에서 내부 구조화와 비판적 점검이 미비하였지만 외주화 비중이 낮아, 미이행 처리분 자체가 크게 누적되지 않았음을 시사한다.

반면, 고활용군(P4-P6)은 막힘 맥락이 다층적으로 확대되고 빈도와 지속 시간도 증가하였다. P4는 내용 조직(4회)과 논리성(3회)에 더해 기억 인출 지연(1회)까지 동반하여, 구조화·논증·표현 경로 전반에서 동시적 인지의 부채 상환 요구가 발생하였음이 나타났다. P5는 내용 조직(3회), 논리성(3회), 과제 이해(2회)에 더해 심리적 부담(1회)까지 나타나 인지적 부채와 정서적 부채의 결합으로 평균 지속 시간이 가장 길게 관찰되었다(55초). P6은 과제 이해(1회)와 기억 인출 지연(2회)이 중심으로, ‘할 일을 머릿속에서 아직 또렷하게 정리하지 못한 상태(과제 표상 미완성)’와 ‘필요한 지식·예시·표현이 곧바로 떠오르지 않는 상태(기억 인출이 더딤)’가 겹쳐서, LLM 보조에서 스스로 쓰기로 넘어가는 순간에 막히는 병목을 만든다는 점을 보여준다.

요컨대 고활용군은 ‘LLM 보조’ 단계에서 구조 설계·논증 검증·표현 생성이 외주화되며 뇌신경 연결성 약화가 미이행 처리 누적(인지의 부채)으로 크게 축적된다. 이어 ‘필자 숙고’ 단계로의 전환에서 그 상환이 빈도 증가와 범주 확장으로 표면화되었다.

V. 결론 및 시사점

본 연구는 생성형 인공지능 시대의 글쓰기 교육에서 필자의 근본적 글쓰기 능력 저하를 완화하고, 개별 필자의 목소리(voice)를 보존·강화할 수 있는 교수·평가 설계의 근거를 마련하고자 했다. 이를 위해 'LLM 보조' 조건의 글쓰기에서 언어적 동형화를 표면 어휘와 의미 구조 차원에서 검증하고 'LLM 보조' 조건에서 '필자 숙고' 조건으로 전환 시, '인지의 부채'에 의해 발생하는 쓰기 막힘의 맥락과 원인을 실증적으로 분석하였다.

실험·분석에서는 먼저 상위 n-gram 분포로 표면 어휘 차원의 동형화를 점검하고 문장 임베딩·군집 구조와 히트맵 분석으로 의미 구조 차원의 수렴과 확산을 확인했다. 다음으로 화면 녹화기록 분석과 면담의 질적 코딩을 통해, 'LLM 보조' 조건에서 '필자 숙고' 조건 글쓰기 전환 시, 쓰기 막힘의 맥락과 원인을 도출하였다. 실험 결과, 'LLM 보조' 조건에서 상위 표현 사용이 중심부로 쏠려 표면 어휘 동형화가 유의하게 증가함을 확인하였다. 또한 같은 주제의 글쓰기에서 'LLM 보조' 조건의 글은 의미 공간의 군집 구조가 '필자 숙고' 조건보다 상대적으로 더 수렴하는 것으로 나타났다.

마지막으로 조건 전환 시점에 누적된 '인지의 부채'가 상환되며 내용 조직, 논리성, 과제 이해, 기억 인출 지연, 심리적 부담의 다섯 맥락에서 쓰기 막힘이 표면화됨을 참여자 자료로 확인하였다. 종합하면, 필자가 LLM 보조를 활용하여 작성 초기의 유창성을 확보하더라도 그 과정에서 AI의 표현을 무비판적으로 수용한다면 표면 어휘의 동형화와 의미 구조의 수렴을 피하기 어렵다. 더 나아가 이러한 필자의 의존적 태도는 핵심 처리(의미 구성·검증·인출) 과정을 생략하게 만들어, 스스로 글을 써야 하는 전환 구간에서 심각한 쓰기 막힘을 야기할 수 있음을 시사한다.

본 연구 결과를 바탕으로 한 제언은 다음과 같다.

첫째, 쓰기 교육에서는 학생들에게 '주체적인 필자로서 숙고를 감내하

고 목소리를 드러내는' 연습과 기회를 충분히 제공할 필요가 있다. 이를 통해 작문 교육은 기계적 산출물에 압도되지 않도록 전통적인 '내면화'와 '자기 수정' 훈련을 통해 필자의 인지적 내구성을 단단히 하고 그 바탕 위에서 LLM을 주체적으로 통제하는 '인지적 결합'을 도모하는 방향이어야 한다.

이는 단순히 LLM 사용을 제한하는 차원을 넘어, LLM이 일상화된 현실에서 '인지의 부채'를 상환하고 '글쓰기 능력'을 신장하기 위한 새로운 교육적 모델이 필요함을 시사한다. 먼저, '인지적 워밍업' 단계의 도입이다. LLM을 활용하기 전, 필자가 핵심 키워드와 문단 구조를 스스로 설계하게 하여 뇌신경망의 초기 점화를 유도해야 한다. 이는 본 연구에서 확인된 예비교사들의 내용 조직 막힘 현상을 예방하는 완충 기제가 될 것이다.

다음으로 '비판적 재구조화' 활동의 강화이다. LLM이 생성한 텍스트의 동형화된 논리를 그대로 수용하는 것이 아니라, 학습자가 자신의 목소리로 문체를 변환하거나 반박하는 과정을 필수 과업으로 포함해야 한다. 즉, LLM의 산출물을 '최종 결과물'이 아닌 '숙고를 위한 초기 자극'으로 재정의하여, 도구 활용이 오히려 필자의 고유성을 강화하는 촉매제가 되도록 교수 설계를 전환해야 한다.

둘째, LLM 활용 글쓰기의 확산이 Bereiter & Scardamalia(1987)의 발달 단계나 Flower & Hayes(1981)의 인지 모형을 새로운 쓰기 환경에 맞게 이론을 정교하게 재구성해야 함을 시사한다. Bereiter & Scardamalia(1987)의 발달 모델 관점에서 볼 때, LLM은 '숙달의 착시' 현상을 설명할 수 있도록 갱신되어야 한다. 본 연구에서 관찰된 언어적 동형화와 표면적 유창성은 LLM 보조 글쓰기가 미숙련 필자의 '지식 나열' 전략을 외관상 고차원적인 '지식 변형'인 것처럼 모사하게 만듦을 보여준다. 필자 내부에서는 실질적인 문제 재정의와 의미 재구성이 부재함에도 산출물은 전문적인 논증 구조를 갖추는 '과정과 결과의 괴리'가 발생하기 때문이다.

따라서 향후 작문 이론은 발달 단계를 단순히 일직선상에 두기보다, '내부 인지에 기반한 지식 변형'과 '외부 도구를 매개로 한 지식 조율'로 이원화

하여 설명할 필요가 있다. 나아가 필자가 LLM의 제안을 단순 수용하는지, 아니면 비판적으로 재구성·혼합하는지에 따라 '도구-조정형 지식 변형'과 같은 새로운 하위 발달 단계를 상정하는 이론적 확장이 요구된다.

또한 Flower & Hayes(1981)의 과정 모형은 '외부 생성기'와의 상호작용을 포함하는 조건부 모형으로 확장되어야 할 필요가 있다. 기존의 '계획-전사-검토' 순환 모형은 필자의 내적 인지자원만을 변인으로 다루었으나, LLM 활용 시대의 글쓰기는 외부 지능이 이 순환 고리에 상시 접속된 상태로 진행된다. 본 연구 결과, 계획 단계에서부터 LLM에 과도하게 의존하거나 전사·검토 단계에서 외부 산출물을 반복적으로 호출할수록, 필자의 핵심 기제인 '상위 모니터'의 조율 기능과 전략적 멈춤이 약화됨이 확인되었다.

이에 따라, 기존 모형에 LLM과의 상호작용을 별도의 하위 모듈로 도식화하고, '어느 단계에서 어떤 강도로 외부 도구를 개입시킬 때 필자의 내생적 조율 고리가 유지되거나 강화되는가'를 설명하는 '조건부 확장 모형'으로의 갱신이 필요하다. 결국, 이러한 이론적 재구성은 전통적 쓰기 이론의 설명력을 포기하는 것이 아니라, LLM이라는 강력한 타자가 개입된 복합적 쓰기 행위를 포괄할 수 있도록 그 설명의 범위를 확장하는 필연적인 이론적 진화라 할 것이다.

LLM을 활용한 글쓰기가 제공하는 효율성은 매력적이지만, 그것이 인간 고유의 '성찰'과 '내면화'라는 인지적 본질을 잠식하도록 방치해서는 안 된다. 진정한 쓰기 역량의 확장은 필자가 도구에 인지를 위탁하는 것이 아니라, 단단한 사고의 근육을 바탕으로 도구를 주체적으로 통제할 때 비로소 가능하기 때문이다. 따라서 작문 교육은 학습자가 '인지의 부채'라는 편의성의 유혹을 넘어, AI라는 거인의 어깨 위에서 자신만의 고유한 목소리를 낼 수 있는 '깨어있는 필자'로 성장하도록 돕는 데 역량을 집중해야 할 것이다.

* 본 논문은 2025.10.31. 투고되었으며, 2025.11.10. 심사가 시작되어 2025.12.03. 심사가 종료되었음.

참고문헌

강동훈(2016), 「쓰기 멈춤의 요인과 발생 양상 분석」, 한국교원대학교 박사학위논문.

강동훈(2023), 「챗지피티(ChatGPT)의 등장과 국어교육의 대응」, 『국어문학』 82, 469-496.

고리사(2025), 「중국인 고급 한국어 학습자의 쓰기 막힘 및 멈춤 양상 연구」, 서울대학교 석사학위논문.

곽수범(2024), 「생성형 AI, 글쓰기의 동반자일까 방해물일까: ChatGPT를 활용한 교육 현장 실험」, 『독서연구』 73, 45-80.

김유민·이경화(2025), 「AI 코스웨어를 활용한 고쳐쓰기 양상」, 『초등 교과교육』 43, 15-28.

나은미(2024), 「챗GPT의 시대, 대학 글쓰기 교육에 대한 성찰 및 교육 방향에 대한 고찰」, 『한성어문학』 51, 79-106.

서종훈(2019), 「쓰기 막힘과 멈춤에 대한 필자의 인식 양상 고찰—갈래와 수준을 고려한 글쓰기를 중심으로—」, 『국어교육』 164, 59-94.

손달임(2023), 「교양 글쓰기 수업에서 ChatGPT의 활용 가능성과 한계」, 『사고와표현』 16(2), 33-65.

오선경(2023), 「대학 교양 글쓰기에서의 챗GPT 활용 사례와 학습자 인식 연구」, 『교양교육연구』 17(3), 11-23.

이미나(2024), 「ChatGPT를 활용한 학술적 글쓰기 사례와 교육 방안 고찰 -H대학교 〈논리적 사고와 글쓰기〉 강좌를 중심으로-」, 『한국언어문학』 128, 265-302.

이승주(2019), 「대화가 고등학생의 쓰기 막힘에 미치는 영향—'가전' 쓰기 중에 나눈 소집단의 대화를 중심으로—」, 『국어교육』 165, 99-140.

홍진옥·전기화(2023), 「글쓰기 수업 내 챗GPT의 활용 가능성—서평 쓰기 수업을 중심으로」, 『작문연구』 59, 7-35.

Altman, D. G. (1991), *Practical Statistics for Medical Research*, London: Chapman & Hall.

Anderson, M. J. (2001). "A new method for non-parametric multivariate analysis of variance", *Austral Ecology 26*(1), 32-46.

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). "Homogenization effects of large language models on human creative ideation", *Creativity and Cognition (C&C '24)*, Chicago, IL: ACM.

Agarwal, D., Vashistha, A., & Naaman, M. (2025). "AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances", In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Yokohama, Japan: ACM.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). "On the dangers of stochastic parrots: Can language models be too big?", *Proceedings of the 2021*

ACM Conference on Fairness, Accountability, and Transparency (FAccT 2021), 610-623.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Bhat, A., Shrivastava, D., & Guo, J. L. C. (2023). "Approach Intelligent Writing Assistants Usability with Seven Stages of Action", In *The Second Workshop on Intelligent and Interactive Writing Assistants (co-located with CHI 2023)*. Hamburg, Germany.

Boice, R. (1990). *Professors as Writers: A Self-Help Guide to Productive Writing*. Stillwater, OK: New Forums Press.

Budiyono, H., Marzuki, Pudjaningsih, W., Prastio, B., & Maulidina, A. (2025). "Exploring long-term impact of AI writing tools on independent writing skills: A case study of Indonesian language education students", *International Journal of Information and Education Technology 15*(5), 1003-1013.

Durak, H., Eğin, F., & Onan, A. (2025). "A comparison of human-written versus AI-generated text in discussions at educational settings: Investigating features for ChatGPT, Gemini and Bing AI", *European Journal of Education 60*(1), e70014.

Flower, L. & Hayes, J. R. (1981). "A cognitive process theory of writing", *College Composition and Communication 32*(4), 365-387.

Fu, L., Newman, B., Jakesch, M., & Kreps, S. (2023). "Comparing Sentence-Level Suggestions to Message-Level Suggestions in AI-Mediated Communication", In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Hamburg, Germany: ACM.

Gries, S. T. (2024). *Frequency, dispersion, association, and keyness: Revising and tupleizing corpus-linguistic measures*. Amsterdam & Philadelphia: John Benjamins.

Hjortshoj, K. (2001). *Understanding Writing Blocks*. New York, NY: Oxford University Press.

Houichi, A. & Sarnou, D. (2020). "Cognitive Load Theory and its Relation to Instructional Design: Perspectives of Some Algerian University Teachers of English", *Arab World English Journal 11*(4), 110-127.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). "Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task", *arXiv preprint arXiv:2506.08872, 4*.

Matsuhashi, A. (1981). "Pausing and planning: The tempo of written discourse production", *Research in the Teaching of English 15*(2), 113-134.

Olive, T. (2014). "Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination", *Journal of Writing Research 6*(2), 173-194.

Padmakumar, V. & He, H. (2024). "Does writing with language models reduce content diversity?", *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*.

Rivera-Novoa, A. & Duarte Arias, D. A. (2025). "Generative Artificial Intelligence and Extended Cognition in Science Learning Contexts", *Science & Education 4*, 1-22.

Rose, M. (1984). Writer's Block: *The Cognitive Dimension*. Carbondale, IL: Southern Illinois University Press

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Amsterdam/Atlanta, GA: Rodopi.

Shaffer, E. L. (2022). *Cognitive Load Theory and Instructional Message Design*. Instructional Message Design: Theory, Research, and Practice (Vol. 2). Old Dominion University Digital Commons.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). "Cognitive architecture and instructional design: 20 years later", *Educational Psychology Review 31*, 261-292.

Vygotsky, L. (2009). Interaction between Learning and Development. *Readings on the Development of Children* (5th ed., pp. 42-48). New York: Worth.

Yulisasih, B. N., Herman, H., & Sunardi, S. (2024). "K-Means Clustering Method for Customer Segmentation Based on Potential Purchases", *Jurnal Teknik Elektro, Teknologi Informasi dan Komputer 8*(1), 83-90.

초록

LLM 활용 글쓰기에서 언어적 동형화 양상
— 뇌신경 연결성 축소와 인지의 부채

장동민 · 박종호

이 연구는 LLM 보조 글쓰기가 필자 개별성에 미치는 영향을 혼합방법으로 검증하였다. 국어교육과 23명, 연합과 36명(총 59명)이 필자 숙고(세션 1·3)와 LLM 보조(세션 2) 조건에서 동일 주제로 작성했다. 표면 수준은 단어 n-gram의 Top-50 중복률과 누적 커버리지 곡선으로 상위 표현 집중도를, 의미 수준은 SBERT 임베딩을 PaCMAP으로 투영하고 코사인 유사도 히트맵으로 집단 응집을 평가했다. 6명의 화면녹화-회상면담을 NVivo 코딩하고 Cohen's κ로 신뢰도를 점검했다. 결과적으로 LLM 보조에서 Top-50 중복률 상승과 커버리지 곡선 급경사화가 나타났고, 임베딩 공간에서도 응집 증가(Silhouette=0.58)와 조건 효과 유의(PERMANOVA $p<.001$)가 확인되었다. 전환 국면(LLM → 숙고)에서는 LLM 고활용 필자의 '내용 조직 · 논리성 · 과제 이해 · 기억 인출 지연 · 심리적 부담' 범주의 막힘이 더 빈번 · 지속되었다.

핵심어 LLM 보조 글쓰기, 뇌신경 연결성, 언어의 동형화, 인지의 부채, 쓰기 막힘

ABSTRACT

LLM-Assisted Writing and Linguistic Homogenization

— Reduced Neural Connectivity and Cognitive Debt

Jang Dongmin · Park Jongho

This mixed-methods study tested how LLM-assisted writing affects writers' individuality. Fifty-nine students wrote on the same topic under deliberation (Sessions 1,3) and LLM assistance (Session 2). Surface convergence used Top-50 n-gram overlap and cumulative coverage curves; semantic cohesion used SBERT → PaCMAP and cosine-similarity heatmaps. Six retrospective interviews were coded in NVivo with Cohen's κ. LLM assistance increased Top-50 overlap and steepened coverage curves; in embedding space, cohesion rose (Silhouette = 0.58) and the condition effect was significant (PERMANOVA $p < .001$). During the transition (LLM → deliberation), high-LLM users showed more frequent, longer blocks.

KEYWORDS LLM-assisted writing, neural connectivity, linguistic homogenization, cognitive debt, writing block