컴퓨터교육학회 논문지 2026년 제29권 제6호
https://doi.org/10.32431/kace.2026.29.6.004

생성형 AI 활용이 대학생의 창의적 아이디어 생성에 미치는 영향: 동일 과제 반복 수행에서의 사용 전환 효과

The Effects of Generative AI Use on University Students' Creative Idea Generation: Switching Effects in Repeated Task Performance

강태곤† · 김경순††
Taegon Kang† · Kyungsoon Kim††

요약

본 연구는 동일 세션에서 동일 과제를 두 차례 수행하는 상황에서 생성형 AI의 효과를 검증하였다. 특히 2차 업그레이드 단계에서의 AI 사용 전환(사용 유지, 사용 중단, 도입, 미사용 유지)에 주목하였다. 대학생 164명은 "2035년, 미래를 바꿀 아이디어 상상하기" 과제를 수행한 뒤 2차에서 산출물을 수정·보완·정교화하였다. 창의성은 4차원 루브릭 총점으로 평가하였다. 결과는 AI의 효과가 단순한 사용 여부뿐 아니라, 과제 수행의 어느 단계에서 어떤 역할로 활용되는지에 따라 달라질 수 있음을 시사한다. 교육적으로는 AI를 일률적으로 허용하거나 금지하기보다, 특히 업그레이드 단계에서 정교화와 보완을 지원하는 도구로 어떻게 설계하고 활용할 것인지에 주목할 필요가 있다. 다만 단일 과제, 짧은 시간, 사용 행태 미측정의 한계가 있으므로, 후속 연구에서는 다양한 과제와 기간, 과정 데이터를 포함할 필요가 있다.

주제어 생성형 인공지능, 창의적 아이디어 생성, 동일 과제 재수행, 사용 전환 효과, 교육 현장 실험연구

ABSTRACT

This study examined the effects of generative AI in a repeated-task setting where the same task was completed twice within a single session. It focused on transitions in AI use during the second-stage upgrade: continued use, discontinuation, adoption, and continued non-use. A total of 164 university students completed the task, "Imagining an Idea That Will Change the Future in 2035," and then revised and refined their outputs in the second stage. Creativity was assessed using the total score of a four-dimensional rubric. The findings suggest that the effects of generative AI depend not only on whether AI is used, but also on when and for what role it is used in the task process. Educationally, rather than treating AI as uniformly allowed or prohibited, it may be more useful to design and guide its use with particular attention to refinement and improvement during the upgrade stage. Because this study was limited to a single task, a short time frame, and no direct measure of usage behavior, future research should include diverse tasks, longer periods, and process data.

Keywords Generative AI, Creative Idea Generation, Repeated Task Performance, Usage Switching Effects, Classroom-Based Experimental Study

†정회원 숭실대학교 일반대학원 IT정책경영학과 박사수료
††정회원 숭실대학교 베어드교양대학 조교수 (교신저자)

논문투고 2026년 02월 15일
심사완료 2026년 04월 17일
게재확정 2026년 04월 22일
발행일자 2026년 06월 22일

1. 서론

생성형 인공지능(Generative AI)과 대규모 언어모델(large language model, LLM)은 개방형 글쓰기, 기획, 아이디어 발상처럼 정답이 정해지지 않은 비정형 과업에서 빠르게 확산되고 있으며, 교육 현장에서도 학생들의 과제 수행 과정과 산출물의 형태를 바꾸는 핵심 도구로 자리 잡았다[1, 2]. 그럼에도 실증 연구와 교육적 논의는 "AI를 쓰면 성과가 좋아지는가"라는 이분법적 비교에 집중하는 경우가 많다[3]. 그러나 실제 수업 과제는 단발 제출로 끝나기보다 초안 작성 이후 수정·보완·재작성(업그레이드)으로 이어지고[4], 이 과정에서 학습자는 AI 사용을 유지하거나 중단하거나 후반에 도입하는 등 전략을 전환한다[5, 6]. 따라서 동일 과제를 반복 수행하며 산출물을 개선하는 맥락에서, AI의 영향은 사용 여부 뿐 아니라 사용 전환(언제/어떤 단계에서 쓰는가) 관점에서 검증될 필요가 있다[5].

본 연구는 창의성을 새로움(originality/novelty)과 유용성·실현가능성(usefulness/feasibility)을 동시에 충족하는 산출로 정의하는 표준 정의를 전제로 하되[7], 반복 수행의 교육 맥락에서는 아이디어가 얼마나 구체화·구조화되는지(정교화/구체성)와 과제의 미래 맥락(본 연구는 2035년 변화 전제)과의 연결성이 얼마나 명확한지(미래 변화 반영성) 또한 결과물의 질을 설명하는 중요한 차원이 될 수 있다는 논의를 반영한다[8, 9]. 이에 따라 본 연구는 새로움, 유용성/실현가능성, 정교화/구체성, 미래 변화 반영성의 4차원 루브릭을 기반으로 창의성 총점을 0–100점으로 산정하여, 반복 개선 과정에서 나타나는 질적 향상을 정량적으로 평가하도록 설계하였다[7, 8].

이론적으로 창의적 수행은 개인의 역량·과정·동기뿐 아니라 사회·환경적 요인(도구, 피드백 등)의 영향을 받는다는 관점이 제시되어 왔으며[10], 생성형 AI는 아이디어의 대안 제시, 관점 전환, 구조화, 표현 정돈, 세부 보완을 지원함으로써 특히 업그레이드 단계에서 정교화 부담을 낮추고 완성도를 높일 가능성이 있다[2, 8, 11]. 동시에 AI의 효과는 과제 유형과 활용 방식에 따라 달라질 수 있고[12], 창의 과제에서는 개인 성과 향상과 집단 수준 산출물의 수렴(유사성 증가) 사이의 긴장이 보고되기도 한다[13, 14]. 이는 생성형 AI의 영향이 단순히 접근 여부가 아니라 어느 단계에서, 어떤 역할로 사용되는지에 의해 달라질 수 있음을 시사한다[13, 14].

최근 생성형 AI와 학습 성과의 관계를 다룬 연구는 빠르게 축적되고 있다. 메타분석 연구는 교육 맥락에서 생성형 AI가 학습 수행에 전반적으로 긍정적 영향을 줄 수 있음을 보고하였으나, 많은 연구가 여전히 AI 사용 여부의 평균 차이를 비교하는 데 집중하고 있다[3]. 반면 실제 학습 장면에서는 학생들이 초안 작성, 수정, 정교화의 각 단계에서 AI를 다르게 활용하며[5, 6], 고등교육 맥락의 교차 설계 연구 역시 AI 보조의 효과가 일률적이지 않고, AI 지원이 제공되는 시점과 활용 방식에 따라 창의적 산출의 양상이 달라질 수 있음을 시사한다[14]. 또한 생성형 AI의 효과는 과업 전반에 균등하게 나타나기보다, AI의 역량 범위 안에 있는 특정 과업 요소에서 더 크게 나타날 수 있다는 점도 보고되었다[12].

이러한 선행연구를 종합하면, 교육 현장에서 중요한 질문은 단순히 "AI를 썼는가"가 아니라 "언제, 어떤 역할로, 어떤 단계에서 사용했는가"라고 볼 수 있다[5, 6, 14]. 특히 실제 수업 과제는 단발 제출보다 초안 작성 이후 수정·보완·재작성의 과정을 포함하는 경우가 많으며[4], 이 과정에서 AI 지원의 투입 시점과 활용 방식은 서로 다른 학습 효과를 만들 수 있음을 시사한다[5, 6, 14].

이에 본 연구는 대학생 164명을 대상으로 "2035년, 미래를 바꿀 아이디어 상상하기" 과제를 실시하되, 1차 수행(10분) 후 약 30분의 간격을 둔 뒤 동일 세션 내에서 2차 업그레이드 수행(10분)을 하도록 설계하였다. 이 30분 간격은 단순한 시간 지연이 아니라, 1차 산출물에 대한 즉각적 반복을 완화하고 잠시 과제에서 주의를 이탈시킨 뒤 다시 돌아오도록 함으로써, 창의적 과제에서 보고되어 온 잠복효과(incubation effect)의 가능성을 부분적으로 반영하기 위한 장치이다[15-18]. 2차 수행에서는 1차 조건 내부에서 무작위로 전환 조건을 분기하여 AI 사용 유지(A1), AI 사용 중단(A2), AI 도입(B1), 미사용 유지(B2)의 4조건을 구성하고, 변화량과 반복측정 혼합효과모형을 통해 도입 효과, 중단 효과, 그리고 두 효과의 비대칭성을 검정하였다.

본 연구는 동일 과제 반복 수행 맥락에서 생성형 AI의 효과를 단순한 사용 여부를 넘어 사용 전환(유지·중단·도입·미사용 유지) 관점에서 비교·검정함으로써, 교육 현장에서 AI 활용을 전면 허용/금지의 이분법만이 아니라 과제 단계별 활용 설계의 문제로도 다룰 필요가 있음을 보여준다. 또한 창의성 평가를 4차원 루브릭으로 구성해 반복 개선 과정에서의 질적 향상을 포착하고자 하였다. 다만 과제 유형과 수행 시간, 표본 범위 및 실제 사용 행태 측정의 제약을 고려할 때, 후속 연구에서는 다양한 과제·기간·학습자 집단과 과정 데이터(사용 목적/행태 등)를 결합한 검증이 필요하다.

2. 이론적 배경

2.1 창의적 아이디어 생성과 평가 틀

창의성은 일반적으로 새로움(originality/novelty)과 유용성·실현가능성(usefulness/feasibility)을 동시에 충족하는 산출로 정의되며, 이는 표준 정의(standard definition)로 널리 수용된다[7]. 그러나 교육 맥락의 개방형 미래 아이디어 과제에서는 최종 산출물의 질이 단순히 참신한 발상 여부만으로 설명되기 어렵다[7, 10]. 반복 수정 과정에서는 아이디어가 얼마나 구체화되고 실현가능성으로 연결되는지, 그리고 과제에서 요구한 미래 맥락을 얼마나 구체적이고 정합적으로 구성하는지가 함께 중요해질 수 있다[9, 19, 20].

특히 본 연구의 과제는 "2035년, 미래를 바꿀 아이디

어 상상하기"로, 현재의 일반적 아이디어가 아니라 미래 사회 변화의 조건을 전제로 한 아이디어 산출을 요구한다. 따라서 '미래 변화 반영성'은 새로움이나 유용성과 별개의 보편적 창의성 핵심 개념이라고 주장하기보다, 본 연구 과제의 맥락 적합성(task-specific relevance) 을 평가하기 위해 추가된 차원으로 이해하는 것이 타당하다[7, 19]. 즉, 동일한 수준의 참신성과 유용성을 가진 아이디어라 하더라도, 2035년의 변화 조건과의 연결이 명확할수록 과제 요구에 더 충실한 산출물로 볼 수 있다.

본 연구의 4차원 루브릭은 창의성의 표준 정의가 제시하는 핵심 축인 새로움과 유용성·실현가능성[7]을 중심으로 하되, 반복 개선 과제에서 중요하게 나타나는 정교화·구체성[8, 9]과 미래지향 과제의 맥락 적합성인 미래 변화 반영성[19]을 보완 차원으로 결합한 평가 틀이다. 다시 말해, 본 루브릭은 일반 창의성의 핵심 요소와 본 연구 과제의 특수성을 함께 반영하도록 구성한 내용타당도 중심의 평가 체계라고 볼 수 있다. 이에 본 연구는 반복 수행의 업그레이드 과정에서 나타나는 질적 향상을 보다 충실히 포착하기 위해 4차원 루브릭을 채택하였다.

2.2 창의적 사고 과정과 반복 수행

창의적 아이디어 생성은 대체로 발산적 사고를 통해 다양한 가능성을 산출한 뒤, 선택·조합·재구성 및 정교화 과정을 거치며 산출물의 질을 높이는 과정으로 이해된다[19, 21]. Geneplore 관점은 창의적 수행을 생성(generation)과 탐색/해석(exploration)의 반복으로 설명하며, 단발 산출보다 반복적 변형과 평가가 결과물의 개선을 촉진할 수 있음을 시사한다[9].

이러한 관점에서 초안 작성 후 업그레이드(수정·보완·정교화)를 요구하는 과제 구조는 실제 수업에서 흔히 발생하는 반복 개선 활동을 실험적으로 모사하면서, 창의적 수행의 핵심 메커니즘(반복 탐색과 정교화)을 반영한 설계로 정당화될 수 있다[8, 9].

2.3 Incubation(잠복) 효과

창의적 문제 해결 연구에서 incubation(잠복) 효과는 일정 시간 동안 문제(과제)에서 주의를 떼고 다른 활동을 하거나 휴식한 뒤 다시 과제에 돌아왔을 때 수행이 향상되는 현상을 의미한다. Sio와 Ormerod의 메타분석(117개 독립 연구)은 전반적으로 잠복 효과가 유의하게 관찰되며, 특히 발산적 사고(divergent thinking) 과제에서 그 효과가 더 두드러질 수 있음을 보고하였으며, 잠복 구간 동안 인지부하가 큰 과제를 수행하면 잠복 효과가 감소하는 경향이 있어, 잠복 구간은 과도한 인지적 요구를 피하는 것이 중요하다고 제안된다[15].

잠복 구간의 작동 기제에 대해서는 고착(fixation) 완화, 연합적 활성(spreading activation), 무의식적 처리(unconscious work), 주의 전환 등이 논의되어 왔으며, 잠복 시간 자체도 수 분의 짧은 간격부터 며칠·수주까지 다양한 범위로 설정될 수 있다. Ritter와 Dijksterhuis는 잠복 연구에서 잠복 구간이 매우 다양하게 조작되며, 인지적으로 덜 요구되는 상태에서의 마음 방황(mind-wandering)이 잠복 과정과 연결될 수 있음을 정리한다[16]. 실제로 Baird 등은 잠복 패러다임에서 과도하게 어렵지 않은(interpolated, undemanding) 과제가 마음 방황을 촉진하는 조건에서, 휴식(rest)이나 고부하 과제 조건보다 이전에 다룬 창의 과제(예: Unusual Uses Task) 수행이 더 향상될 수 있음을 보고하였다. 또한 잠복 효과는 발산적 과제에서 비교적 일관되게 관찰되며, 잠복 구간에 고부하 활동을 배치하지 않는 설계가 잠복 효과를 뒷받침한다는 논의도 제시된다[17].

따라서 본 연구에서 1차 수행 후 약 30분의 간격을 두고 2차 업그레이드 수행을 배치한 설계는, 창의적 아이디어 산출과 같은 발산적 과제에서 잠복 효과가 보고되어 왔다는 점[15], 잠복 구간이 수 분에서 수십 분 이상의 다양한 범위로 운용될 수 있다는 점[16], 그리고 잠복 구간에 과도한 인지부하를 유발하지 않는 활동이나 휴식을 배치할 때 그 효과가 유지될 수 있다는 점[17, 18]과 연결된다. 본 연구의 30분 간격은 장기간의 독립적 재수행을 의미한다기보다, 동일 세션 내에서 1차 초안 작성 이후 잠시 과제에서 벗어났다가 다시 돌아와 수정·보완·정교화를 수행하도록 한 업그레이드 맥락의 설계 장치로 이해할 수 있다. 따라서 본 연구는 반복 수행의 단순 연습 효과뿐 아니라, 짧은 간격 이후 재접근이 아이디어 재구성과 정교화에 어떤 차이를 만드는지도 함께 살펴보는 설계로 해석될 수 있다[4, 15, 18, 20, 22].

2.4 생성형 AI의 창의 수행 지원 메커니즘

구성요소 이론 관점에서는 창의적 수행이 개인의 역량·과정·동기뿐 아니라 환경적 지원(도구, 피드백 등)에 의해 영향을 받을 수 있다[10]. 생성형 AI(LLM 기반 도구)는 아이디어 산출과 정교화 단계에서 대안 제시, 관점 전환, 구조화, 표현 개선, 근거·세부 요소 보완을 제공할 수 있으며, 이는 본 연구의 평가 차원 중 유용성/실현가능성 및 정교화/구체성의 향상과 직접 연결될 여지가 있다[2, 8, 11].

동시에 생성형 AI의 효과는 모든 과제와 모든 수행 단계에서 동일하게 나타나지 않는다. 선행연구는 AI가 특정 과업에서는 생산성과 품질을 높일 수 있지만, 그 효과가 과제 유형과 수행 맥락에 따라 달라질 수 있음을 보여주고[11, 12], 창의 과제에서는 개인 수준의 성과 향상과 집단 수준의 산출물 다양성 감소 사이의 긴장이 발생할 수 있음을 보고하였다[13, 14]. 이러한 결과는 생성형 AI의 교육적 효과를 단순한 사용/미사용 차이로 환원하기보다, 초안 생성과 수정·정교화의 어느 단계에서 어떤 방식으로 사용되었는지를 함께 고려해야 함을 시사한다[5, 6, 13, 14]. 따라서 동일 과제 반복 수행 상황에서의 '사용 전환'은 생성형 AI 효과를 더 정교하게 설명하기 위한 적절한 분석 단위가 된다.

2.5 동일 과제 반복 수행에서의 사용 전환

현실의 학습 상황에서 AI 활용은 일회성보다 반복적이며, 초안 작성 단계와 수정·보완 단계에서 활용 전략이 달라질 수 있다[5, 6]. 동일 세션 내 1차 초안 작성 후 2차 업그레이드 수행을 하는 맥락에서, AI를 AI 사용 유지·AI 사용 중단·AI 도입·미사용 유지로 전환시키는 설계는, AI가 발산(아이디어 확장)과 수렴(구조화·정교화·표현 정리) 중 어느 국면에서 더 큰 기여를 하는지, 그리고 도구 도입과 철회가 성과 변화에 대칭적으로 작동하는지(또는 비대칭인지)를 검정할 수 있게 한다[8, 14].

특히 본 연구의 전환 조건은 다음과 같은 이론적 예측을 가능하게 한다.

·미사용→도입(B1): 1차에서 학습자 고유의 문제 정의·관점이 확보된 상태에서 2차에 AI를 결합하므로, AI의 구조화·정교화·표현 개선 기능이 업그레이드 단계의 병목을 완화할 수 있다[2, 8, 11].

· 사용→중단(A2): 1차에서 AI 보조를 경험한 뒤 2차에서 도구가 제거되면, 동일 시간·인지자원 제약하에서 정교화 부담이 상대적으로 커질 수 있고, 그 결과 개선 폭이 둔화될 수 있다[8].

·AI 사용 유지(A1)·미사용 유지(B2): 1차 초안 작성 후 2차 업그레이드 수행 자체에 따른 학습 효과는 존재할 수 있으나[9], 도구의 지속적 활용 여부는 개선의 양상(정교화, 적용 가능성 보강, 미래 맥락 명료화 등)을 달리 만들 수 있다[2, 8].

2.6 연구가설

이상의 논의를 종합하면, 생성형 AI의 효과는 단순한 사용 여부보다 과제 단계와 사용 전환 방식에 따라 달라질 가능성이 있다. 먼저, 생성형 AI는 개방형 글쓰기와 아이디어 생성 과제에서 수행 향상을 가져올 수 있다는 보고가 축적되고 있으며[2, 3, 11], 본 연구의 1차 수행과 같은 초기 아이디어 산출 단계에서도 AI 사용 집단이 더 높은 성과를 보일 가능성이 있다. 이에 다음과 같은 가설을 설정하였다.

H1. 1차 수행에서 AI 사용 집단(A)은 미사용 집단(B)보다 창의성 총점이 높을 것이다.

다음으로, 초안 작성 이후 수정·보완·재구성을 거치는 반복 수행은 창의적 산출물의 질을 향상시킬 수 있으며[4, 8, 9], 짧은 간격 이후 재접근은 잠복효과와 결합하여 추가적인 향상을 가져올 가능성이 있다[15-18]. 이에 전체 표본 차원에서는 반복 수행 자체의 업그레이드 효과를 예상할 수 있다.

H2. 전체 참여자의 2차 수행 점수는 1차 수행 점수보다 높을 것이다.

또한 실제 학습 장면에서 AI 활용은 초안 작성과 수정 단계 사이에서 다르게 나타나며[5, 6], 생성형 AI는 특히 구조화, 정교화, 표현 개선과 같은 업그레이드 단계 지원에 강점을 가질 수 있다[2, 8, 11]. 따라서 1차에서 스스로 문제 정의와 기본 관점을 확보한 학습자가 2차에서 AI를 도입하는 경우, 미사용을 유지하는 경우보다 더 큰 향상폭을 보일 가능성이 있다[14].

H3. 2차에서 AI를 도입한 집단(B1)은 미사용을 유지한 집단(B2)보다 변화량 Δ가 클 것이다.

반대로 1차에서 AI 지원을 활용하던 학습자가 2차에서 그 도구를 제거당할 경우, 제한된 시간 안에 정교화와 구조화를 수행해야 하는 부담이 커져 AI 사용을 유지한 집단보다 향상폭이 작아질 수 있다[8, 14].

H4. 2차에서 AI를 중단한 집단(A2)은 AI 사용을 유지한 집단(A1)보다 변화량 Δ가 작을 것이다.

마지막으로, 1차에서 이미 학습자 고유의 초안과 문제 정의가 형성된 상태에서 2차에 AI를 결합하는 경우에는, AI가 아이디어 확장, 구조화, 표현 정교화, 수정 지원의 자원으로 작동하면서 기존 초안을 보완할 가능성이 있다[4, 5, 6, 13]. 반면 사용 중단의 경우에는 1차에서 형성된 초안과 아이디어 기반이 이후 수정의 출발점으로 남을 수 있으므로, 향상폭이 둔화되더라도 그 감소 양상이 도입 효과와 동일하게 나타나지는 않을 가능성이 있다.

H5. 도입 효과(B1−B2)는 중단 효과(A1−A2)보다 더 크게 나타날 것이다.

3. 연구가설 및 연구방법

3.1 연구 설계 및 자료 수집

본 연구는 동일한 창의적 아이디어 생성 과제에서 1차 초안 작성 후 약 30분의 간격을 두고 동일 세션 내 2차 업그레이드 수행을 하게 한 뒤, 2차 수행에서 생성형 AI 사용을 AI 사용 유지/AI 사용 중단/AI 도입/미사용 유지로 전환시키는 혼합설계(mixed design) 실험이다. 시점(1차, 2차)은 개인 내 반복측정 요인(within-subject)이며, 2차 전환 조건(A1/A2/B1/B2)은 개인 간 비교 요인(between-subject)이다. 실험 절차는 Figure 1과 같다.

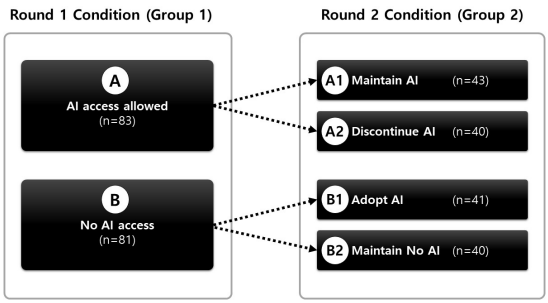

Figure 1. Experimental procedure: round 1 and 2 conditions

(1) 1차 수행(10분): 참여자는 제시된 과제에 대해 아이디어를 작성한다. 1차에서 AI 사용 집단(A)은 생성형 AI를 활용할 수 있었고, 미사용 집단(B)은 생성형 AI 없이 작성하였

다. 본 연구의 AI 사용 조건에서는 ChatGPT 사용을 기본으로 안내하였으며, 사용 당시 기본 접근 가능한 버전은 Free plan이었다. 다만 실제 수업 맥락을 반영하여, 참여자가 원할 경우 다른 생성형 AI 도구의 사용도 허용하였다. 따라서 본 연구의 AI 조건은 특정 단일 도구의 순수 효과라기보다 생성형 AI 활용 조건의 효과로 해석할 필요가 있다.

(2) 휴식 및 과제 간격(약 30분): 단순 휴식 대신 유튜브 동영상을 제시하여 참여자가 과제와 직접 관련 없는 내용에 주의를 분산하도록 하였으며, 이를 통해 즉시 반복 수행에서 발생할 수 있는 단순 기억 및 복사 효과를 완화하고자 하였다.

(3) 2차 수행(10분): 동일 과제를 업그레이드하도록 안내한다. 즉, 1차 산출물을 기반으로 아이디어를 수정·보완·정교화하거나, 새로운 관점을 추가하도록 지시한다. 2차에서는 1차 집단(A/B) 내부에서 무작위로 전환 조건을 분기하여 A는 AI 유지(A1) 또는 AI 중단(A2), B는 AI 도입(B1) 또는 미사용 유지(B2)로 배정한다.

과제는 “2035년, 미래를 바꿀 아이디어 상상하기”이며, 참여자에게는 미래 사회·일상의 변화(예: 기후 변화, 고령화, 도시 과밀, 정신건강 문제, 관계 변화 등)를 전제로 삶을 개선할 수 있는 아이디어를 제안하도록 안내하였다. 아이디어의 형태는 소프트웨어, 제품, 서비스, 정책 등으로 제한하지 않았으며, 각 차수마다 최대 10개까지 작성할 수 있도록 하였다.

1차 수행에서는 제한 시간 10분 동안 가능한 한 다양한 아이디어를 산출하도록 하였고, 2차 수행에서는 1차 산출물을 바탕으로 기존 아이디어를 수정·보완·정교화하거나 새로운 관점을 추가하도록 지시하였다. 즉, 2차는 완전히 새로운 과제를 다시 수행하는 방식이 아니라, 1차 초안을 기반으로 한 업그레이드 과제로 운영되었다.

분석에 포함된 참여자는 총 164명이며, 1차 조건은 A 83명, B 81명으로 구성되었다. 2차 조건은 A1 43명, A2 40명, B1 41명, B2 40명이다. 참여자는 국내 2개 대학 재학생으로 구성되었으며, 성별 분포는 남 121명, 여 43명이다. 다만 본 연구에서는 참여자의 학년 및 전공 정보를 체계적으로 수집하지 못하였으므로, 이러한 배경 변인이 창의적 아이디어 생성에 미칠 수 있는 영향은 본 연구에서 직접 통제하지 못하였다. 이 점은 결과 해석과 일반화에서 고려할 필요가 있다. 분석 단위는 개인이며, 1차 및 2차 산출물을 모두 작성한 참여자만 포함하였다.

본 연구는 실제 수업 맥락에서 수행한 교육 현장 실험에서 진행되었다. 참여자는 IT관련 과목을 학습하는 학생들로, 동일한 과제 안내와 시간 조건 아래 1차 수행과 2차 업그레이드 수행에 참여하였다.

연구참여에 앞서 참여자에게 연구 목적, 수행 절차, 자료 활용 범위, 익명성 보장 여부를 안내하였으며, 서면 동의를 통해 자발적 참여 의사를 확인하였다. 수집된 자료는 분석 과정에서 개인 식별 정보가 드러나지 않도록 처리하였다.

3.2 측정 및 채점(창의성 점수 산정)

3.2.1 채점 단위와 점수 구조

채점은 아이디어 1개를 단위로 수행하였다. 각 아이디어는 4개 차원에 대해 평가한 뒤 이를 합산하여 0–10점 범위의 아이디어 점수로 변환하였다. 4개 차원은 다음과 같다.

·새로움(originality/novelty): 기존에 흔한 해법의 단순 반복이 아니라 새로운 결합·관점·접근이 제시되는 정도이다. 단순한 아이디어 나열이나 익숙한 해결책의 반복은 낮게 평가하고, 기존 요소를 새롭게 결합하거나 문제를 다른 관점에서 재구성한 경우는 높게 평가하였다.

·유용성/실현가능성(usefulness/feasibility): 제안된 아이디어가 실제 문제 해결에 기여할 가능성과 적용 가능성이 얼마나 구체적으로 드러나는지를 평가하였다. 기대 효과가 모호하거나 실행 가능성이 불분명한 경우는 낮게, 적용 대상·작동 방식·기대 효과가 비교적 명확한 경우는 높게 평가하였다.

·정교화/구체성(elaboration/specificity): 아이디어가 얼마나 구체적인 실행 논리와 설명을 갖추었는지를 평가하였다. 단순한 개념 수준의 제시는 낮게, 작동 방식, 대상, 절차, 운영 방식, 제약 조건 등이 보다 구체적으로 제시될수록 높게 평가하였다.

·미래 변화 반영성(future-change fit): 아이디어가 2035년의 사회·기술·일상 변화 맥락과 얼마나 분명하게 연결되는지를 평가하였다. 미래 맥락과의 관련성이 약하거나 단순 언급에 그친 경우는 낮게, 미래 변화 조건이 아이디어의 필요성·작동 방식·효과와 구체적으로 연결된 경우는 높게 평가하였다. 이러한 기준은 각 차원의 정의를 단순 명칭 수준이 아니라 실제 채점 판단이 가능하도록 조작적으로 구체화하기 위한 것이다.

3.2.2 참여자 점수 산정

참여자별 총점은 해당 차수에서 제출한 최대 10개 아이디어의 아이디어 점수를 합산하여 산정하였다(범위 0–100점). 아이디어 개수가 10개 미만인 경우는 제출된 아이디어만 합산하였고, 미제출 아이디어는 점수에 포함하지 않았다.

본 자료에서 학생들의 평균 아이디어 개수는 1차 9.18개, 2차 9.52개였다. 이는 대부분의 참여자가 최대 제출 가능 수인 10개에 가까운 수의 아이디어를 제시했음을 의미한다. 따라서 본 연구의 총점 차이는 아이디어 수의 차이만으로 설명되기보다, 각 아이디어의 새로움·유용성·정교화·미래 변화 반영성과 같은 질적 차이를 상당 부분 반영하는 것으로 해석할 수 있다. 다만 총점이 아이디어의 양과 질을 함께 반영하는 지표라는 점에서, 아이디어 개수의 영향을 완전히 배제할 수는 없으므로 결과 해석에서는 이 점을 함께 고려할 필요가 있다.

3.2.3 중복/유사 아이디어 감점(차수 내)

동일 차수에서 한 참여자가 의미적으로 유사한 아이디어

를 반복 제시한 경우, 새로움 차원에 감점을 적용하고, 감점 기준은 사전에 규정하였다. 구체적으로, 경미한 반복은 핵심 문제의식은 유사하나 적용 대상, 기능, 사용 맥락, 해결 방식 중 일부가 구분되는 경우로 보아 1점을 감점하였다. 상당히 유사는 핵심 아이디어와 해결 방식이 대부분 동일하고 표현이나 세부 예시만 일부 다른 경우로 보아 2점을 감점하였다. 사실상 중복은 적용 대상과 해결 논리가 거의 동일하여 별개의 아이디어로 보기 어려운 경우로 보아 3점을 감점하였다. 감점 이후에도 아이디어 점수는 0-10점 범위로 제한하였다.

3.2.4 평가자 및 신뢰도

평가는 2명의 독립 평가자가 수행하였으며, 평가자는 참여자의 조건 정보를 알 수 없도록 한 상태에서 동일한 루브릭에 따라 채점하였다. 본 채점에 앞서 루브릭 적용 기준과 차원별 판단 원칙, 유사 아이디어 감점 규칙에 대해 사전 협의를 거쳤으며, 본 채점에서는 각 평가자가 독립적으로 점수를 부여하였다. 평가자 간 일치도는 급내상관계수(ICC)로 확인하였다. 개별 평가점수 기준 ICC와 두 평가자 평균 점수 기준 ICC를 함께 산출하였으며, 최종 분석에 두 평가자의 평균점수를 사용하였기 때문에, 평균점수 기준 ICC도 함께 제시하였다. 총점 기준으로 1차 평가는 개별 점수 기준 ICC=.918, 평균 점수 기준 ICC=.957이었고, 2차 평가는 개별 점수 기준 ICC=.920, 평균 점수 기준 ICC=.958이었다. 또한 1·2차 산출물 전체 328건에 대해 차원별 평균 점수 기준 ICC를 산출한 결과, 새로움=.921, 유용성/실현가능성=.929, 정교화/구체성=.912, 미래 변화 반영성=.848로 나타났다. 이는 본 연구의 4차원 루브릭이 두 평가자 간에 전반적으로 높은 수준의 일치도를 보였음을 의미한다.

3.3 통계 분석 방법

모든 가설 검정은 양측 유의수준 α=.05 기준으로 수행하며, 추정치에 대한 95% 신뢰구간을 함께 제시한다.

3.3.1 전환 조건 점검

전환 조건 간 비교가 1차 성과 차이에 의해 편향되지 않았는지 확인하기 위해, 1차 집단 내부에서 다음 비교를 수행한다.

·A 내부: A1 vs A2의 점수 비교

·B 내부: B1 vs B2의 점수 비교

등분산을 가정하지 않는 Welch t-test를 적용하여 1차 점수 차이가 유의하지 않음을 확인한다.

3.3.2 H1: 1차 AI 사용 효과

그룹 1(A vs B)의 1차 점수 평균 차이를 Welch t-test로 검정한다. 평균차와 95% 신뢰구간을 제시하고, 효과크기는 Cohen's d를 함께 보고한다.

3.3.3 H2: 반복 수행(업그레이드) 효과

전체 표본에서 점수_2와 점수_1의 차이를 대응표본 t-test로 검정한다. 평균 증가량(2차 점수−1차 점수), 95% 신뢰구간, 대응표본 효과크기(dz)를 보고한다. 또한 해석 보조를 위해 전환 조건별로(그룹 2) 1차→2차 평균 증가량과 조건 내 대응표본 t-test 결과를 함께 제시한다.

3.3.4 H3-H5: 사용 전환 효과(도입/중단/비대칭성)

반복측정 구조(개인별 2회 점수)를 직접 반영하기 위해 선형 혼합효과모형(무선절편)을 적용한다. 모형은 다음과 같다.

·점수 ~ time × 그룹 2 + (1 | 개인)

여기서 time은 1차(0)와 2차(1)로 코딩하며, 그룹 2의 기준집단은 A1로 설정한다. 관심 효과는 time×그룹 2 상호작용이며, 계획 대비(planned contrasts)로 아래 비교를 핵심 결과로 보고한다.

·도입 효과: (B1의 상승폭) − (B2의 상승폭)

·중단 효과: (A1의 상승폭) − (A2의 상승폭)

·전환 비대칭성: [(B1−B2) − (A1−A2)]

4. 결과

4.1 표본 구성 및 전환 조건 점검

분석대상자는 총 164명이며, 1차 조건은 A 83명, B 81명으로 구성되었다. 2차 전환 조건은 A1 43명, A2 40명, B1 41명, B2 40명으로 집단 간 표본 수가 비교적 균형적이다.

2차 전환 조건 간 비교가 1차 점수 차이에 의해 편향되지 않았는지 확인하기 위해, 1차 점수를 기준으로 A 내부(A1 vs A2)와 B 내부(B1 vs B2)의 동질성을 점검하였으며, 그 결과는 Table 1과 같다.

Table 1. Homogeneity Test Results (Welch's t-test)

Category	df	t	p	ΔM	95% CI (ΔM)	d
A1 vs A2	81.00	0.53	.598	1.78	[−4.91, 8.47]	0.12
B1 vs B2	77.19	0.38	.704	1.44	[−6.08, 8.96]	0.08

즉, A 내부 및 B 내부에서 2차 전환 조건은 1차 성과 수준에서 유의한 차이가 없었으며, 전환 조건 비교는 1차 점수 편향 없이 수행 가능한 상태로 판단된다.

4.2 1차 AI 사용 효과(H1)

1차 수행에서 AI 사용 집단(A)의 평균 점수는 63.07(SD=15.27, n=83), 미사용 집단(B)의 평균 점수는 29.65(SD=16.88, n=81)였다. 두 집단의 평균 차이는 통계적으로 유의하였다(Welch t(159.54)=13.29, p<.001). 평균차는 33.42점이며, 95% 신뢰구간은 [28.45, 38.38]이었다. 효과크기 Cohen's d=2.08로 매우 큰 수준의 차이가 관찰되었다. 이처럼 효과크기가 크게 나타난 것은 두 집단의

평균차가 33.42점으로 매우 큰 반면 표준편차는 각각 15.27과 16.88로 비교적 유사하여, 표준화 평균차가 크게 산출되었기 때문이다. 또한 1차 수행은 10분의 제한된 시간 안에서 초기 아이디어를 발산적으로 산출하는 단계였으므로, 생성형 AI의 아이디어 확장, 구조화, 표현 정리 지원이 집단 간 차이에 일부 기여했을 가능성이 있다. 다만 이러한 해석은 본 연구의 결과에 대한 설명적 추론이며, 실제 사용 행태를 직접 측정한 것은 아니므로 후속 연구에서 과정 자료를 통해 추가 검증할 필요가 있다. 따라서 1차 수행에서 AI 사용 집단이 미사용 집단보다 창의성 총점이 높다는 결과가 확인되어 H1은 지지되었다.

4.3 반복 수행(업그레이드) 효과(H2)

전체 표본에서 1차와 2차 점수의 기술통계는 Table 2와 같으며, 평균 점수의 변화는 Figure 2에 제시하였다.

Table 2. Descriptive statistics for scores

Category	Mean	SD
Round 1	46.57	23.19
Round 2	60.93	19.66

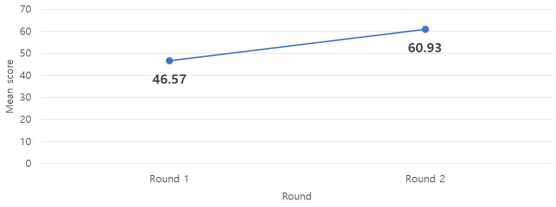

Figure 2. Round 1 and Round 2 mean scores

점수_2와 점수_1을 대응표본으로 비교한 결과, 2차 점수는 1차 점수보다 유의하게 높았다(paired $t(163)=10.34$, $p<.001$). 평균 증가량(Δ의 전체 평균)은 14.36점이며, 95% 신뢰구간은 [11.62, 17.10]이었다. 대응표본 효과크기 $dz=0.81$로 중간 이상 수준의 향상이 관찰되었다. 따라서 반복 수행에 따른 업그레이드 효과(H2)는 지지되었다.

전환 조건별(그룹_2) 조건 내 변화(1차→2차)는 Table 3과 같으며, 전환 조건별 평균 변화량의 차이는 Figure 3에 시각화하였다.

Table 3. Paired t-test results for score changes by switching condition

Cat.	n	⊿M	t	df	p	95% CI (⊿M)	dz
A1	43	8.00	6.65	42	<.001	[5.57, 10.43]	1.01
A2	40	2.50	3.76	39	<.001	[1.15, 3.85]	0.59
B1	41	37.73	12.66	40	<.001	[31.71, 43.75]	1.98
B2	40	9.10	6.81	39	<.001	[6.40, 11.80]	1.08

이 결과는 모든 조건에서 평균적으로는 상승이 존재하되, 상승폭의 크기가 전환 조건에 따라 크게 달라질 수 있음을 시사한다. 전환 조건 간 상승폭 차이에 대한 가설 검정은 다음 절의 혼합효과모형(주분석)에서 제시한다.

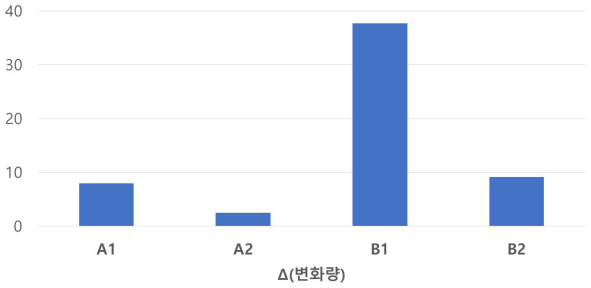

Figure 3. Mean change scores by switching condition

4.4 사용 전환 조건에 따른 변화량 차이(H3~H5)

4.4.1 전환 조건별 기술통계

전환 조건별 1차 점수, 2차 점수, 변화량(Δ)의 기술통계는 Table 4와 같다.

Table 4. Switching-condition means (SD)

Category	n	Mean(SD)		
		Round 1	Round 2	⊿
A1	43	63.93(15.83)	71.93(16.30)	8.00(7.89)
A2	40	62.15(14.79)	64.65(15.08)	2.50(4.21)
B1	41	30.37(15.86)	68.10(11.37)	37.73(19.08)
B2	40	28.93(18.04)	38.03(15.18)	9.10(8.45)

기술통계 수준에서 B1(도입)은 2차 점수가 크게 상승하여 Δ가 가장 컸고, A2(중단)는 Δ가 가장 작았다. 다만 가설 검정은 반복측정 구조를 반영한 혼합효과모형의 계획 대비로 수행하였다.

4.4.2 반복측정 혼합효과모형(time×그룹_2)

기준집단(A1)의 1차 평균 수준(절편)은 63.93점이었고, 기준집단(A1)의 1차→2차 상승폭(time 효과)은 8.00점으로 추정되었다(계수=8.00, SE=1.71, $z=4.68$, $p<.001$, 95% CI [4.65, 11.35]).

관심 효과인 time×그룹_2 상호작용 및 계획 대비 결과는 다음과 같다.

1) 도입 효과(H3): B1(도입) vs B2(미사용 유지)

도입 효과는 "B1의 상승폭 − B2의 상승폭"으로 정의하였다. 계획 대비 결과, 도입 효과는 28.63점으로 유의하였다(추정치=28.63, SE=2.49, $z=11.48$, $p<.001$, 95% CI [23.74, 33.52]). 즉, 2차에서 AI를 도입한 집단(B1)은 미사용을 유지한 집단(B2)보다 1차→2차 향상폭이 유의하게 컸

다. 따라서 H3은 지지되었다.

2) 중단 효과(H4): A1(유지) vs A2(중단)

중단 효과는 “A1의 상승폭 – A2의 상승폭”으로 정의하였다. 계획 대비 결과, 중단 효과는 5.50점으로 유의하였다(추정치=5.50, SE=2.47, z=2.23, p=.026, 95% CI [0.67, 10.33]). 즉, 2차에서 AI 사용을 중단한 집단(A2)은 AI 사용을 유지한 집단(A1)보다 상승폭이 유의하게 작았다. 따라서 H4는 지지되었다.

3) 전환 비대칭성(H5): (B1–B2)와 (A1–A2)의 차이

전환 비대칭성은 “도입 효과(B1–B2) – 중단 효과(A1–A2)”로 정의하였다. 계획 대비 결과, 비대칭성은 23.13점으로 유의하였다(추정치=23.13, SE=3.51, z=6.60, p<.001, 95% CI [16.26, 30.00]). 즉, 본 자료에서는 도입 효과의 크기가 중단 효과의 크기보다 유의하게 크게 나타났다. 따라서 H5는 지지되었다.

4.5 결과 요약

(1) 1차 수행에서 AI 사용 집단(A)은 미사용 집단(B)보다 창의성 총점이 유의하게 높았다(H1 지지).

(2) 전체적으로 2차 점수는 1차 점수보다 유의하게 상승하여 반복 수행(업그레이드) 효과가 확인되었다(H2 지지).

(3) 2차에서 AI를 도입한 집단(B1)은 미사용 유지 집단(B2)보다 향상폭이 유의하게 컸다(H3 지지).

(4) 2차에서 AI를 중단한 집단(A2)은 AI 유지 집단(A1)보다 향상폭이 유의하게 작았다(H4 지지).

(5) 도입 효과(B1–B2)가 중단 효과(A1–A2)보다 유의하게 크게 나타나 전환 효과의 비대칭성이 확인되었다(H5 지지).

5. 결론 및 한계점

5.1 결론 및 시사점

본 연구는 동일한 창의적 아이디어 생성 과제에서 1차 초안 작성 후 동일 세션 내 2차 업그레이드 수행을 하도록 설계하고, 생성형 AI의 효과를 단순한 사용/미사용 비교(H1)를 넘어 사용 전환(AI 사용 유지/AI 사용 중단/AI 도입/미사용 유지) 관점에서 검증하였다. 참여자(총 164명)는 1차에서 AI 사용(A)과 미사용(B)으로 과제를 수행한 뒤, 2차에서 A는 AI 사용 유지(A1)/AI 사용 중단(A2), B는 AI 도입(B1)/미사용 유지(B2)로 무작위 분기되어 1차 산출물을 기반으로 업그레이드 과제를 수행하였다. 창의성은 새로움, 유용성/실현가능성, 정교화/구체성, 미래 변화 반영성의 4차원 루브릭을 바탕으로 총점(0–100점)으로 산정하였다. 주요 결과는 다음과 같다. 첫째, 본 연구에서는 1차 초안 작성 후 2차 업그레이드 수행 자체에 따른 전반적 향상 효과가 확인되었으며(H2), 이와는 별도로 2차에서 AI를 어떤 방식으로 전환했는가에 따라 향상폭이 달라지는 추가 효과도 확인되었다(H3~H5). 즉, 본 연구 결과는 반복 개선 자체에 따른 일반적 향상 효과와, 그 위에서 추가로 나타난 AI 사용 전환 효과를 구분하여 해석할 필요가 있음을 보여준다. 따라서 교육 현장에서 AI 활용은 단순 허용·금지의 정책 문제만이 아니라, 과제 수행 단계에 맞춘 교수설계의 문제로도 다룰 필요가 있다.

둘째, 1차 초안 작성 후 2차 업그레이드 수행 자체가 학습자의 아이디어를 더 구체화하고 완성도를 높이는 데 기여할 수 있으므로, 수업 설계에서 초안 제출로 끝내기보다 수정·보완 단계가 포함되도록 과제를 구성하는 것이 유익하다. 이는 전체 표본에서 2차 점수가 1차보다 유의하게 높았고(H2), 전환 조건과 무관하게 모든 조건에서 평균적으로 상승이 관찰되었다는 결과에 근거한다(Table 3). 특히 학생이 1차에서 스스로 아이디어의 방향성과 문제 정의를 충분히 확보한 뒤, 2차에서 개선·보완을 수행하도록 하는 구조는 학습 과정의 질을 높이는 데 도움이 된다. 다만 이러한 향상은 AI 사용 여부와 무관하게 일정 부분 나타날 수 있는 반복 개선 효과이며, 본 연구의 AI 전환 효과는 이러한 일반적 향상폭 위에서 조건별 추가 차이로 해석하는 것이 타당하다.

셋째, 업그레이드 단계에서의 AI 도입은 반복 개선 자체의 평균적 향상 효과를 넘어, 학생이 이미 만든 초안을 바탕으로 아이디어를 구조화하고 보완하는 데 추가적으로 유리하게 작동할 가능성을 보여준다. 이는 B1의 향상폭이 B2보다 유의하게 컸고(H3), 반복측정 혼합효과모형에서도 도입 효과가 유의하게 확인되었다는 결과에 근거한다. 즉, 본 자료에서는 AI가 초안이 어느 정도 형성된 상태에서 논리 전개, 세부 설계, 적용 가능성 보강, 표현 정돈 등 정교화 역할에 활용될 때 특히 유용할 가능성이 시사되었다.

넷째, 반대로 AI를 사용하던 학습자가 업그레이드 단계에서 AI 사용을 중단하게 되면, 개선 과정이 둔화되거나 완성도 향상이 제한될 가능성이 있다. 이는 학생이 AI에 의존해서라기보다, 제한된 시간 안에 정교화·구체화를 수행해야 하는 상황에서 도구가 제공하던 일부 지원이 제거되었기 때문일 수 있다. 따라서 수업 운영에서 사용 중단이 필요하다면, 그 효과를 단순 금지로만 다루기보다 체크리스트, 템플릿, 피어 피드백과 같은 대체 전략을 함께 제공하는 방안도 고려할 수 있다.

다섯째, 공정성과 학습 효과를 함께 고려하기 위해서는 “AI를 어디에 쓰면 좋은지/쓰면 안 되는지”를 보다 구체적으로 안내하고, 결과물 평가 역시 그에 맞춰 설계하는 방안을 검토할 수 있다. 예컨대 아이디어의 핵심 개념과 문제 정의는 학습자 기여를 중심으로 평가하고, 정교화·구체성(실행 계획, 작동 방식, 이해관계자, 제약조건 등)과 표현의 명료성은 AI 활용을 포함한 개선 과정과 함께 해석하는 방식도 하나의 대안으로 고려할 수 있다. 또한 AI 사용 여부뿐 아니라 사용 목적(예: 구조화, 표현 개선, 대안 탐색)을 함께 기록하면, 지도와 피드백에 도움이 될 가능성이 있다.

여섯째, 본 연구는 반복 수행 맥락에서 전환 전략의 차이가 성과 변화와 관련될 수 있음을 보여주므로, 후속 연구에서는 단순 성과 비교를 넘어 어떤 사용 방식이 어떤 차원의 개선과 연결되는지를 보다 정교하게 추적할 필요가 있다.

이를 위해서는 AI 사용 로그(프롬프트/응답 요약), 사용 목적 체크, 수정 전·후 비교, 자기 성찰 보고 등을 결합하여 학습 과정 데이터를 함께 수집·분석하는 접근을 후속 연구에서 검토할 필요가 있다.

5.2 연구의 한계점 및 후속 연구 과제

첫째, AI 사용의 강도/방식 측정 부재와 관련하여, 본 연구는 조건 수준에서 사용/미사용 및 전환을 통제했으나 사용 조건 내부에서 실제 사용 시간, 프롬프트 수, 사용 목적(아이디어 확장/구조화/문장화 등), 반복 상호작용 횟수 같은 사용 행태를 정량화하지는 않았다. 후속 연구는 로그(프롬프트·응답 요약), 사용 목적 체크리스트, 사용 시간·횟수 기록 등을 결합하여 어떤 사용 방식이 어떤 차원의 향상과 연결되는지 메커니즘 수준에서 분석할 필요가 있다.

둘째, 평정 기반 창의성 점수의 주관성 및 채점 규칙 민감도와 관련하여, 2명의 독립 평가자 블라인드 채점과 사전 합의 절차를 적용했으나 루브릭 기반 평정은 본질적으로 평가자 판단의 영향을 완전히 제거하기 어렵다. 또한 차수 내 유사 아이디어 감점 규칙은 중복 산출의 과대평가를 줄이기 위한 장치지만, 유사성 판단 기준과 감점 폭에 따라 결과가 달라질 수 있다. 후속 연구는 감점 기준의 예시 확장, 규칙 변화에 따른 민감도 점검 등을 통해 재현성과 설득력을 강화할 필요가 있다. 아울러 본 연구의 총점은 최대 10개 아이디어의 합산 점수로 산정되었으므로, 후속 연구에서는 아이디어 개수와 아이디어당 평균 질 점수를 분리하여 분석함으로써 양과 질의 효과를 보다 명확히 구분할 필요가 있다.

셋째, 일반화의 한계(표본·과제·시간 제약)와 관련하여, 표본은 2개 대학 재학생(총 164명)이며 과제는 “2035년 미래 아이디어”라는 단일 주제, 제한된 수행 시간(각 10분)에서 수행되었다. 과제 유형(논증형/설계형/데이터 기반 문제해결형), 수행 기간(장기 프로젝트), 학습자 특성(전공/비전공, 학년 등)이 달라지면 전환 효과의 크기와 양상이 달라질 수 있으므로, 다양한 과제와 집단으로 재현 연구가 필요하다.

넷째, 학습 과정 변수 미포함과 관련하여, 본 연구는 산출물 점수(총점 및 변화량)를 중심으로 효과를 검정했으며 동기·몰입·자기효능감·성찰 수준 같은 과정 변수는 충분히 포함하지 못했다. 또한 본 연구는 1차 초안 작성 후 2차 업그레이드 수행이라는 구조를 취했으므로, 관찰된 점수 향상에는 반복 개선 자체에 따른 향상과 AI 사용 전환 효과가 함께 반영되어 있을 가능성이 있다. 후속 연구에서 성과 지표와 과정 지표를 결합하면, 왜 도입이 유리했는지/왜 중단이 불리했는지를 교육적으로 더 구체적으로 설명할 수 있다.

참고문헌

[1] Liang, W., Zhang, Y., Codreanu, M., Wang, J., Cao, H., & Zou, J. (2025). The widespread adoption of large language model-assisted writing across society. *Patterns, 6*(12), 101366. https://doi.org/10.1016/j.patter.2025.101366

[2] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

[3] Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

[4] Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.58680/ccc198115885

[5] Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2025). How do students use ChatGPT as a writing support? *Journal of Adolescent & Adult Literacy, 68*(5), 445–457. https://doi.org/10.1002/jaal.1373

[6] Jang, H., Jin, L., & So, H.-J. (2026). Examining University Students' Engagement with ChatGPT in English Essay Writing: Interaction Patterns and Perceptions. *The Asia-Pacific Education Researcher*, 1–14. https://doi.org/10.1007/s40299-025-01076-9

[7] Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

[8] Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM, 50*(12), 20–32. https://doi.org/10.1145/1323688.1323689

[9] Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. MIT Press.

[10] Amabile, T. M. (2012). *Componential theory of creativity* (Working Paper No. 12-096). Harvard Business School. http://www.hbs.edu/faculty/Publication%20Files/12-096.pdf

[11] Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942. *https://doi.org/10.1093/qje/qjae044*

[12] Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality* (Working Paper No. 24-013). Harvard Business School Technology & Operations Mgt. Unit Working Paper. https://ssrn.com/abstract=4573321

[13] Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

[14] Baltà-Salvador, R., Brasó-Vives, E., & Peña, M. (2026).

Evaluating AI-assisted creative ideation: A crossover study in higher education. *Thinking Skills and Creativity, 59*, 101958. https://doi.org/10.1016/j.tsc.2025.101958

[15] Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135*(1), 94-120. https://doi.org/10.1037/a0014212

[16] Ritter, S. M., & Dijksterhuis, A. (2014). Creativity—the unconscious foundations of the incubation period. *Frontiers in Human Neuroscience, 8*, 215. https://doi.org/10.3389/fnhum.2014.00215

[17] Gilhooly, K. J. (2016). Incubation and intuition in creative problem solving. *Frontiers in Psychology, 7*, 1076. https://doi.org/10.3389/fpsyg.2016.01076

[18] Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science, 23*(10), 1117-1122. https://doi.org/10.1177/0956797612446024

[19] Thakral, P. P., Yang, A. C., Addis, D. R., & Schacter, D. L. (2021). Divergent thinking and constructing future events: Dissociating old from new ideas. *Memory, 29*(6), 729–743. https://doi.org/10.1080/09658211.2021.1940205

[20] Baruah, J., Paulus, P. B., & Kohn, N. W. (2021). The effect of the sequence of creative processes on the quality of the ideas: The benefit of a simultaneous focus on originality and feasibility. *The Journal of Creative Behavior, 55*(4), 946-961.

[21] Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

[22] Gilhooly, K. J., Georgiou, G. J., Garrison, J., Reston, J. D., & Sirota, M. (2012). Don't wait to incubate: Immediate versus delayed incubation in divergent thinking. *Memory & Cognition, 40*(6), 966-975. https://doi.org/10.3758/s13421-012-0199-z

강태곤

· 1997년 동의대학교 수학과(이학사)
· 2023년 성균관대학교 글로벌경영학전공(경영학석사)
· 2026년 숭실대학교 컴퓨터공학전공(공학박사 수료)
· 2001년~현재 ㈜우리은행, ㈜우리금융지주
· 2025년 신한대학교 소프트웨어융합학과 겸임교수

관심분야 : 인공지능, 금융공학, 컴퓨터교육
comjaby@naver.com

김경순

· 1997년 성신여자대학교 전산학과(이학사)
· 1999년 성신여자대학교 전산교육전공(교육학석사)
· 2025년 숭실대학교 컴퓨터공학전공(공학박사)
· 2026년~현재 숭실대학교 베어드교양대학 조교수
· 2023년~2025년 숭실대학교 베어드교양대학 겸임교수
· 2018년~2025년 신한대학교 소프트웨어 융합학과 겸임교수
· 1999년~2010년 성신여자대학교 컴퓨터정보학부 강사

관심분야 : 인공지능, 블록체인, 컴퓨터교육
anamorne@ssu.ac.kr