

예술·디자인 학과 졸업작품 심사를 위한 생성형 AI 협업 기반 시스템 구축에 관한 실행연구

비개발자 교육자 사례를 중심으로

An Action Research on Building a Generative AI-Assisted System for Graduation Project Evaluation in Art and Design Departments

A Case Study of a Non-developer Educator

주 저 자 : 나상귀 (Na, Sang Kwee)

계원예술대학교 공간연출과 조교수
nask@kaywon.ac.kr

<https://doi.org/10.46248/kidrs.2026.2.746>

접수일 2026. 05. 20. / 심사완료일 2026. 05. 31. / 게재확정일 2026. 06. 08. / 게재일 2026. 06. 30.

Abstract

This action research analyzes a case in which a non-developer educator designed and operated a graduation project evaluation system for an art and design department through natural language collaboration with generative AI. Without programming expertise, the researcher constructed the system across four stages using Gemini, ChatGPT, and Claude, and examined its operation through a cycle of first-round evaluation, intervention, and second-round application. The findings show that non-developer educators can build systems reflecting their operational contexts, and that the same tool can produce different outcomes depending on how it is used. Between the two rounds, the average score changed from 83.10 to 84.64; however, this change is interpreted as a result of external factors such as the development of students' work. In contrast, the exception case caused by unrecorded scores observed in the first round did not occur in the second round, which more clearly demonstrates the effect of system refinement.

Keyword

Generative AI(생성형 AI), Graduation Project Evaluation(졸업작품 심사), Non-developer Educator(비개발자 교육자), Action Research(실행연구), Digital Evaluation System(디지털 평가 시스템)

요약

본 연구는 비개발자 교육자가 생성형 AI와의 자연어 기반 협업을 통해 예술 디자인 학과의 졸업작품 심사 시스템을 직접 설계하고 운영한 사례를 실행연구의 관점에서 분석한 것이다. 연구자는 프로그래밍 전문성을 갖추지 않은 상태에서 Gemini, ChatGPT, Claude와의 협업을 통해 4단계에 걸쳐 시스템을 구축하였으며, 1차 심사 운영, 1차 결과를 바탕으로 한 개입과 보완, 2차 심사 적용의 사이클을 통해 그 작동을 확인하였다. 분석 결과 시스템 구축의 층위에서는 비개발자도 운영 맥락이 반영된 시스템을 구축할 수 있음이 확인되었으며, 동일한 도구가 다른 방식으로 활용될 때 다른 결과를 낳는다는 점이 드러났다. 심사 운영의 층위에서는 1차에서 2차로 이행하는 동안 평균 점수가 83.10점에서 84.64점으로 변화하였으나, 이러한 점수 변화는 학생 작품의 발전 등 외적 요인이 함께 작용한 결과로 해석된다. 반면 1차에서 확인된 미입력 처리 관련 예외 사례가 2차에서는 나타나지 않은 점은 시스템 정비의 효과를 보다 분명하게 보여 준다. 본 연구는 비개발자 교육자가 생성형 AI 협업을 통해 자신의 운영 맥락에 맞는 시스템을 구축할 수 있는 가능성을 구체적 사례로 제시하였다는 점에서 의의를 지닌다.

목차

1. 서론

- 1-1. 연구 배경 및 목적
- 1-2. 연구 방법 및 구성

2. 이론적 배경

- 2-1. 생성형 AI와 교육기술 환경의 변화
- 2-2. 예술·디자인 교육의 디지털 평가

2-3. 비개발자 교육자의 시스템 구축 실행연구

3. 시스템 설계 및 개발 과정

- 3-1. 시스템 설계의 배경과 출발점
- 3-2. 1단계: Gemini 기반 초기 프로토타입
- 3-3. 2단계: ChatGPT 기반 구조 재설계
- 3-4. 3단계: Claude 협업 시도와 한계
- 3-5. 4단계: 다도구 교차 검증 방식 도입

4. 시스템 적용 및 운영 결과

- 4-1. 시스템 적용 개요
- 4-2. 1차 심사 운영 결과
- 4-3. 2차 심사 운영 결과

1. 서론

1-1. 연구 배경 및 목적

실기 중심 예술디자인 교육에서 졸업작품 심사는 학생의 전공 역량과 학습 성과를 종합적으로 확인하는 핵심 절차이다. 졸업작품은 학생이 습득한 전공 지식과 기술을 바탕으로 기획·설계·제작한 교육 성과이며, 심사는 이를 교육과정의 마지막 시점에서 점검하는 단계로 기능한다.

현재 심사 운영에서 작품의 제출과 열람은 디지털 환경에서 이루어지지만, 평가 결과의 기록과 집계는 여전히 종이 심사서 작성과 전산 재입력에 의존하는 경우가 많다. 본 연구에서는 이처럼 검토는 디지털화되었으나 기록과 집계가 연계되지 않은 상태를 '반(半) 디지털화된 심사 방식'으로 정의한다. 이러한 방식은 심사 집중도와 운영 효율을 동시에 저해한다. 특히 학과별 평가 항목과 심사 구조가 다른 실기 전공에서는 학과 특화형 디지털 심사 시스템의 필요성이 더욱 두드러진다.

한편, 최근 생성형 AI의 발전은 프로그래밍 전문성을 갖추지 않은 교육자, 즉 비개발자 교육자도 자연어 기반 협업을 통해 학습 지원이나 창작 보조를 넘어 교육 운영 시스템의 설계와 실행까지 직접 시도할 가능성을 확대하고 있다. 실제로 선행 연구에서는 AI 기반 맞춤형 피드백 시스템의 효과성을 검증하거나¹⁾ 이미지 생성 AI의 미술교육 활용 가능성을 탐색하는 등²⁾ AI의 교육적 활용을 다각도로 논의해 왔다. 그러나 이러한 연구는 주로 학습 지원이나 창작 보조에 초점을 두고 있어 비개발자 교육자가 생성형 AI와 협업하여 교육

4-4. 1차와 2차 결과의 비교 분석

5. 결론

참고문헌

운영 시스템을 직접 설계·구현·운영한 사례를 분석한 연구는 상대적으로 부족하다.

이에 본 연구는 비개발자 예술 교육자가 생성형 AI와 협업하여 학과 특화형 졸업작품 심사 시스템을 직접 개발·운영한 사례를 분석한다. 이를 통해 반 디지털화된 심사 방식의 한계를 검토하고, 생성형 AI 기반 협업이 맞춤형 심사 시스템의 설계·구현·운영 과정에서 지니는 가능성과 한계를 단일 사례 차원에서 살펴보고자 한다.

1-2. 연구 방법 및 구성

본 연구는 연구자가 시스템 개발의 주체이자 심사 운영자로 직접 참여하여 현장의 문제를 인식하고, 그 해결을 위한 개입을 설계·실행한 뒤, 그 과정과 결과를 성찰적으로 분석하는 실행연구(Action Research) 방법을 채택한다. 실행연구는 연구자가 현장 내부자의 위치에서 문제 해결과 변화 과정을 순환적으로 탐색한다는 점에 특징이 있다.³⁾ 본 연구는 졸업작품 심사 운영 과정에서 발생하는 반 디지털화 문제를 연구자가 직접 경험하고, 이를 해결하기 위한 기술적 개입으로 생성형 AI 협업 기반 심사 시스템을 개발·적용하였다는 점에서 실행연구의 성격에 부합한다.

연구의 범위는 2026년 A대학교 B학과의 졸업작품 심사 운영 사례로 한정하며, 연구 대상 학과는 익명화를 위해 예술디자인 계열 실기 중심 학과로 일반화하여 제시한다. 내용적 범위는 시스템의 기획 및 개발 착수 단계부터 1차 심사 운영, 개입 및 보완, 2차 심사 적용과 데이터 확정 단계까지를 포함한다. 연구 자료는 학생 28명과 심사 교수 4명의 심사 운영 자료를 바탕으로 구성되며, 크게 세 가지로 정리된다. 첫째, Gemini, ChatGPT, Claude⁴⁾와 주고받은 AI 협업 로

1) 김동삼·류다현·박규동·강혜진, 'AI기반 실시간 온라인 교육 맞춤형 피드백 시스템 효과성 평가', 한국컴퓨터교육학회, 컴퓨터교육학회 논문지, Vol.27, No.7, 2024. 10. pp.11-23.

2) 장윤경·윤인화·박지원·제갈유진·이지연, '이미지 생성 AI 도구의 이해와 미술교육의 활용 가능성 탐구', 한국조형교육학회, 조형교육, Vol.88, 2023. 11. pp.277-298.

3) McNiff, J., 『Action Research: All You Need to Know』, SAGE Publications, 2017, pp.9-12.

4) 본 연구에서 활용한 생성형 AI 도구는 Gemini 3 Flash, ChatGPT-4o, Claude Sonnet 4.6이며, 시스템

그는 비개발자의 요구가 시스템 기능과 구현 논리로 구체화되는 과정을 보여 주는 질적 자료이다. 둘째, 실제 심사 운영 과정에서 생성된 시스템 산출 데이터는 점수 입력 결과, 집계 결과, 누락 확인 정보, 피드백 데이터 등을 포함하며 시스템의 운영 성과와 데이터 관리의 개선 양상을 보여 준다. 셋째, 개발 과정의 시행착오와 주요 의사결정의 근거, 운영 중에 발생한 문제와 대응 과정을 기록한 연구자 성찰 기록은 시스템 개발과 적용의 맥락을 해석하는 데 활용된다.

자료의 분석은 질적·정량적 접근을 병행하여 수행한다. AI 협업 로그와 연구자 성찰 기록은 문제 정의, 요구사항 조정, 도구별 협업 방식의 변화, 운영 중 수정·보완 과정을 질적으로 분석한다. 시스템 산출 데이터는 점수 집계, 신속성, 실시간 통계 반영 여부, 누락 확인 가능성, 데이터 확정 및 관리의 용이성 등을 검토하여 1차 심사, 개입 및 보완, 2차 심사 적용 단계에서 나타난 변화 양상을 비교·분석한다.

본 논문은 총 5장으로 구성된다. 제1장에서는 연구의 배경과 목적, 연구 방법 및 논문 구성을 다룬다. 제2장에서는 생성형 AI와 교육기술 도구, 실기 중심 예술교육의 디지털 평가, 비개발자 교육자의 시스템 개발에 관한 선행 연구를 검토한다. 제3장에서는 시스템 설계의 배경과 초기 요구사항을 검토하고, 생성형 AI와의 협업을 통한 단계별 개발 과정을 분석한다. 제4장에서는 시스템의 1차 심사 적용 결과와 그 과정에서 도출된 문제를 살펴본 후, 개입과 보완을 거쳐 2차 심사에 재적용한 결과를 분석한다. 제5장에서는 연구 결과를 종합하고, 비개발자 예술 교육자의 생성형 AI 협업 기반 시스템 구축이 갖는 의의와 향후 과제를 제시한다.

2. 이론적 배경

2-1. 생성형 AI와 교육기술 환경의 변화

대학 교육 현장의 디지털 전환은 주로 범용 학습관리시스템(LMS)이나 상용 소프트웨어를 도입하는 방식으로 이루어져 왔다. 이러한 시스템은 출결, 과제 제출, 성적 입력과 같은 정형화된 행정 절차를 지원하는 데에는 효과적이지만, 시각적 결과물의 검토와 다면적 평가가 동시에 이루어지는 실기 중심 예술 디자인 교육의

심사 맥락을 충분히 반영하기에는 한계가 있었다. 그 결과 교육 현장에서는 기성 시스템의 기능에 맞추어 운영 방식을 조정하거나, 시스템이 수용하지 못하는 절차를 수기 업무로 보완하는 방식이 지속되어 왔다. 이러한 구조에서 교육자는 대체로 시스템의 사용자 위치에 머물렀으며, 현장의 운영 맥락을 시스템 설계에 반영할 여지는 제한적이었다.

생성형 AI의 확산은 이러한 구도에 중요한 변화를 불러오고 있다. 특히 대규모 언어모델(Large Language Model) 기반 도구는 전문적인 프로그래밍 지식이 없는 사용자도 자연어 기반 상호작용을 통해 기능을 구상하고 구현 방향을 구체화할 수 있도록 한다. Sarkar는 이를 ‘생성적 전환 가설(generative shift hypothesis)’로 설명하면서, 생성형 AI가 자연어 프롬프트를 통해 비전문가의 프로그래밍 범위와 규모를 질적·양적으로 확장할 수 있다고 보았다.⁵⁾

이러한 변화는 교육 영역에서 특히 중요한 함의를 지닌다. Mollick과 Mollick은 시뮬레이션, 멘토링, 코칭, 공동창작 등 프롬프트 기반 사례를 통해, 교원의 내용 전문성과 교수법 전문성이 새로운 교육기술 설계의 기반이 될 수 있음을 논의하였다. 이들은 생성형 AI가 교육자를 단순한 도구 사용자에서 학습 경험의 설계 주체로 이동시키고 있다고 보았다.⁶⁾ 다만 이들의 논의는 학습 활동과 상호작용 설계에 초점을 두고 있어, 실제 운영 데이터와 사용자 인증, 관리 화면을 포함하는 교육 운영 시스템의 구축까지를 본격적으로 다루지는 않는다. 이러한 점에서 본 연구는 교육자 주도의 설계 가능성을 교육 운영 시스템의 개발과 적용 차원으로 확장한 사례로 볼 수 있다.

이처럼 생성형 AI는 교육자가 시스템의 수용자에서 설계자로 이동할 수 있는 환경을 열어 가고 있으며, 실기 중심 예술 디자인 교육의 평가 환경에서는 이러한 변화의 의미가 더욱 크게 나타난다.

5) Sarkar, A., ‘Will Code Remain a Relevant User Interface for End-User Programming with Generative AI Models?’, ACM, Proceedings of the 2023 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (Onward! 2023), 2023, 10. pp.153-167.

6) Mollick, E. & Mollick, L., ‘Instructors as Innovators: A Future-focused Approach to New AI Learning Opportunities, With Prompts’, The Wharton School, The Wharton School Research Paper (SSRN 4802463), 2024. 04. pp.1-3.

개발과 1차 심사 후 정비 작업에 사용된 시기는 2026년 2월부터 3월까지이다.

2-2. 예술디자인 교육의 디지털 평가

실기 중심 예술디자인 전공의 졸업작품 평가는 일반적인 이론 교과와 정량 평가와는 다른 특성을 보인다. 작품의 조형성, 기획의 논리성, 구현의 완성도, 발표의 설득력 등을 종합적으로 판단해야 하며, 다양한 형식의 자료가 동시에 검토되고, 복수의 심사위원이 참여하는 다면 평가가 이루어진다는 점에서 일반적인 디지털 평가 체계와 구별된다.

심사 자료의 형식이 다변적이다. 심사위원은 학생이 사전에 제출한 기획서, 포트폴리오, 도면, 이미지 자료, 영상 자료뿐 아니라, 심사 당일 현장에서 제시되는 프레젠테이션 자료와 설치 결과까지 함께 검토해야 하는 경우가 많다. 평가 구조 역시 복잡적이다. 여러 명의 전공 교수가 동시에 심사에 참여하며, 각 심사위원의 평가 결과는 운영 과정에서 신속하게 확인되고 집계될 필요가 있다. 이 과정에서는 점수 입력, 누락 여부 확인, 수정 반영, 평균값과 순위 산출 등이 유기적으로 연결되어야 한다.

미술교육 분야에서도 디지털 환경의 확산에 대응하기 위한 다양한 연구가 축적되어 왔다. 정옥희는 디지털 전환이 미술교육에 약과 독을 동시에 안기는 파르마콘적 실천 딜레마를 지적하면서, 디지털 소양과 심미적 감수성의 균형이 새로운 과제로 부상하였음을 논의하였다.⁷⁾ 또한 이미지 생성 AI의 미술교육 활용 가능성을 탐색한 연구나 AI 기반 맞춤형 학습 피드백 시스템의 효과성을 검증한 연구처럼, 학습 지원과 창작 보조 측면에서 AI를 다룬 논의도 꾸준히 이루어져 왔다. 그러나 이러한 연구는 주로 수업 내용과 학습 과정을 중심으로 전개됐으며, 평가 운영 자체의 디지털 구조를 다룬 논의는 상대적으로 제한적이다.

이러한 맥락에서 평가 체계의 디지털 전환은 단순한 페이퍼리스 환경 구축을 넘어, 심사 행위가 이루어지는 운영 환경 자체를 재구성하는 문제로 이해될 필요가 있다. 자료의 다양성, 심사 흐름의 실시간성, 결과 관리의 정확성을 함께 수용하려면 전공의 특수한 심사 맥락을 반영한 맞춤형 시스템 설계가 요구되며, 이는 학습 지원이나 창작 보조 중심의 기존 AI 활용 논의가 충분히 다루지 못한 영역이자 본 연구가 위치하는 지점이기도 하다.

7) 정옥희, '디지털 전환 미술교육의 파르마콘적 실천 딜레마', 한국예술교육학회, 예술교육연구, Vol.21, No.2, 2023. 06. pp.287-304.

2-3. 비개발자 교육자의 시스템 구축 실행연구

비개발자 교육자가 생성형 AI와의 협업을 통해 직접 시스템을 설계하고 구축하는 행위는 단순한 기술 구현을 넘어, 자신의 교육 철학과 현장의 문제를 구조화하고 성찰하는 과정으로 이해될 수 있다. 이들은 프로그래밍 전문성에는 한계를 가질 수 있으나, 교육 현장의 운영 맥락과 평가 요구에 대한 도메인 전문성을 밀접하게 보유한 주체이기도 하다.

이러한 연구자 주도의 개입은 실행연구의 관점에서 해석될 수 있다. 실행연구는 연구자가 실천 현장에서 문제를 인식하고, 개입을 직접 설계·실행하며, 그 과정을 성찰적으로 분석함으로써 실천과 이론의 간극을 좁히는 연구 접근이다.⁸⁾ 본 연구에서 연구자가 생성형 AI와의 협업을 통해 졸업작품 심사 시스템을 개발한 것은 반 디지털화된 심사 운영의 문제를 해결하기 위한 기술적 개입인 동시에, 그 과정에서의 판단과 시행착오를 분석 가능한 연구 자료로 전환하는 실천적 행위라 할 수 있다.

미술교육 분야에서도 실행연구는 새로운 교육적 상황에 대응하기 위한 방법론으로 활용되어 왔다. 안인기와 이미영은 생성형 AI를 활용한 초등 미술수업을 개발하는 과정에서 예비 운영과 반성, 본 과정의 재설계와 실행으로 이어지는 순환적 구조를 따랐으며, 이를 통해 새로운 기술이 도입되는 교실 환경에서 교사 연구자의 성찰이 갖는 의미를 보여 주었다.⁹⁾ 본 연구는 이와 같은 실행연구의 흐름과 맥을 같이하면서도, 그 적용 대상을 수업 설계에서 평가 운영 시스템의 개발과 적용으로 확장한다는 점에서 차별성을 지닌다.

본 연구는 이러한 세 가지 이론적 토대가 만나는 지점에 있다. 생성형 AI가 비개발자에게 열어 놓은 시스템 설계의 가능성, 실기 중심 예술디자인 평가가 요구하는 맞춤형 디지털 환경, 비개발자 교육자가 자신의 실천을 성찰적으로 분석하는 실행연구의 관점이 결합할 때, 교육자 주도의 시스템 구축이라는 실천 가능성은 보다 구체화된다. 다음 장에서는 이러한 이론적 토대를 바탕으로 실제 시스템이 어떻게 설계되고 단계적으로 발전하였는지를 분석한다.

8) McNiff, Op. cit., 2017, pp.9-12.

9) 안인기·이미영, '생성형 인공지능 활용 초등 미술수업 개발 실행연구', 한국초등미술교육학회, 미술과 교육, Vol.25, No.1, 2024. 01. pp.1-31.

3. 시스템 설계 및 개발 과정

3-1. 시스템 설계의 배경과 출발점

본 연구의 출발점은 두 가지였다. 하나는 학과의 졸업작품 심사 운영에서 반복되던 구조적 한계였고, 다른 하나는 연구자 자신이 프로그래밍 전문성을 갖추지 않은 교육자였다는 점이었다.

학과의 졸업작품 심사는 오랫동안 종이 심사서와 디지털 자료가 분리된 반 디지털화 방식으로 운영되어 왔다. 학생이 제출한 작품 이미지와 기획 텍스트는 디지털 형태로 존재하였으나, 심사위원의 평가는 종이 심사서에 수기로 기록된 뒤 별도의 전산 입력 과정을 거쳐야 했다. 이 과정에서는 매체 간 불일치, 집계 오류, 실시간 진행 상황 확인의 어려움과 같은 문제가 반복되었다. 연구자는 이러한 문제를 자체 시스템 구축을 통해 해결하고자 하였다.

그러나 연구자는 교육과 실무 모두 디자인 영역에 속해 있었으며, 프로그래밍 코드에 대한 기초 지식을 갖추고 있지 않았다. 다만 시스템 개발을 외부에 의뢰하는 방식은 비용 부담이 크고 학과 운영의 세부 맥락을 충분히 반영하기 어렵다고 판단되었다. 이러한 상황에서 생성형 AI의 활용 가능성이 점차 확장되면서, 연구자는 자연어 기반 협업을 통해 직접 시스템 구축을 시도해 볼 수 있겠다는 가능성을 검토하게 되었다.

본 장은 이와 같은 문제 인식과 시도가 실제 시스템 설계 및 개발 과정으로 이어진 경로를 네 단계로 나누어 기술한다. 분석 자료는 각 도구와의 채팅 기록, GitHub 커밋 이력, 그리고 1차와 2차 심사 운영 기록으로 구성된다.

[표 1] 시스템 개발 4단계 요약

단계	시기/도구	핵심 작업	결과
1 단계	2026.2 / Gemini	자연어 기반 웹 환경에서 초기 프로토타입 구축	데이터 휘발 및 협업 이해 단절 확인
2 단계	2026.2~3 / ChatGPT	자료 저장·인증·운영 환경이 분리된 구조로 재설계, 운영 기능 5종 추가	운영 가능 구조 확보 (이미지 휘발 잔존)
3 단계	2026.3 / Claude	1차 심사 직전 단편적 코드 투입을 통한 검증 시도	누적 맥락 부재로 진전 제한
4 단계	2026.3~ / Gemini · ChatGPT · Claude	이미지 저장 휘발 문제 다도구 진단·해결, 1차 심사 후 시스템 4개 영보 역 정비	2차 심사 안정 운영의 기반 확보

3-2. 1단계: Gemini 기반 초기 프로토타입

개발의 첫 단계는 Google AI Studio의 Gemini를 활용한 초기 프로토타입 구축으로 이루어졌다. 별도의 개발 환경을 설정하지 않은 상태에서 연구자가 웹 화면 안에 구현하고자 하는 기능을 자연어로 기술하면 Gemini가 이에 대응하는 코드를 생성하고 미리보기 화면을 통해 결과를 즉시 확인할 수 있는 구조였다.

이 단계에서는 학생 기획안 제출 화면, 심사위원용 평가 입력 화면, 관리자용 통계 대시보드를 포함한 기본 구조가 자연어 요청만으로 구성되었다. 초기 버전은 단일 파일 구조의 코드로 생성되어 Google Cloud Run을 통해 외부에서 접속 가능한 형태로 배포되었으며, 이후에도 명단 일괄 등록과 같은 운영 편의 기능이 채팅 기반 요청을 통해 순차적으로 추가되었다.

그러나 이 단계에서는 협업의 한계도 함께 드러났다. Gemini가 제시하는 답변의 상당 부분은 개발 중심의 전문 용어와 설명 방식으로 이루어져 있었으며, 연구자가 그 의미를 즉시 이해하지 못하는 경우가 반복되었다. 특히 입력된 데이터가 일정 시간 이후 휘발되는 문제가 발생하였을 때, Gemini는 이에 대한 보완 방안을 제시하였다. 그러나 제한된 용어와 작동 방식이 연구자에게 충분히 이해되지 않은 상태에서 수정이 이루어지면서 문제는 해소되지 못하였다.

이 단계는 두 가지 점을 동시에 보여 주었다. 하나는 일상적 언어를 기반으로 한 생성형 AI와의 협업이 비개발자에게도 시스템 구축의 가능성을 열 수 있다는 점이었고, 다른 하나는 AI의 권고가 연구자의 이해 수준과 충분히 연결되지 못할 경우 협업이 더 이상 진전되기 어렵다는 점이었다. 이러한 경험은 이후 협업 도구와 개발 환경 전반을 재검토하고 전환해야 한다는 판단으로 이어졌다.

3-3. 2단계: ChatGPT 기반 구조 재설계

1단계에서 이루어진 보완 시도가 충분히 작동하지 않았던 경험은 시스템 구조 전반에 대한 진단이 필요하다는 판단으로 이어졌다. 이에 연구자는 이미 다른 업무에서 활용하며 비교적 익숙해져 있던 ChatGPT를 주 협업 도구로 선택하였다. 연구자는 1단계에서 생성된 코드 전체를 압축 파일 형태로 정리하여 ChatGPT에 제공하고, 학과 심사 운영의 맥락을 함께 설명한 뒤 구조적 취약점과 개선 방향을 검토하였다. 코드를 전체 단위로 제시하고 진단을 받는 방식은 1단계의 자연어 요청 중심 협업과는 성격이 달랐으며, 이때부터 협업의

단위는 개별 화면이 아니라 시스템 전체로 확장되었다.

ChatGPT의 진단을 통해 세 가지 핵심 문제가 정리되었다. 첫째는 이미지 파일이 운영 서버의 임시 영역에 저장되어 시간이 지나면 휘발될 수 있다는 점이었다. 둘째는 심사위원의 단순 열람이 평가 완료로 처리되는 등 평가 상태가 명확히 구분되지 않는 구조였다. 셋째는 미입력 항목이 0점으로 처리되어 평균 점수가 왜곡되는 산출 방식의 문제였다.

이러한 진단과 함께 학과 심사 운영 경험에서 도출된 요구사항도 구체적인 기능으로 반영되었다. 평가 상태의 단계 분리, 심사 라운드의 단계화, 항목별 점수 입력 기능의 보완, 관리자 화면의 인쇄 기능, 평균 산출 방식의 수정이 그것이다. 이와 같은 기능 추가는 시스템을 단순한 프로토타입 수준에서 실제 운영 가능한 구조로 전환하는 데 중요한 역할을 하였다.

다만 이 단계에서도 완전히 해결되지 않은 문제가 남아 있었다. 텍스트 데이터는 비교적 안정적으로 보존되었으나, 이미지 파일은 시간이 지나면 휘발되는 현상이 계속 관찰되었다. 이 문제는 이후 단계에서 다시 다루어지게 되었다.

3-4. 3단계: Claude 협업 시도와 한계

1차 심사를 앞둔 시점에서 ChatGPT와 Gemini를 통해서도 이미지 휘발 문제가 완전히 해결되지 않아, 연구자는 새로운 도구로 Claude를 활용해 보고자 하였다. 2단계에서 정리된 코드를 Claude에 제공하고, 개선이 필요한 부분을 함께 검토하는 방식으로 작업이 진행되었다.

그러나 이 단계는 앞선 두 단계와는 다른 양상을 보였다. Claude가 제안한 수정안을 반영한 이후 시스템이 정상적으로 작동하지 않는 상황이 반복되었으며, 연구자는 수정된 코드를 다시 원래 상태로 복원하는 작업을 수행해야 했다. 1단계와 2단계의 협업은 각 도구와의 대화가 누적되면서 의사결정의 맥락이 함께 축적되는 과정이었으나, Claude는 개발의 중간 시점에 새롭게 투입된 형태였기 때문에 이러한 누적 맥락이 부재한 상태였다. 코드는 전달되었지만, 왜 그러한 구조가 선택되었는지, 어떤 시행착오를 거쳐 현재의 형태에 이르렀는지에 대한 배경은 충분히 공유되지 못하였다.

이 단계는 추가적인 시스템 발전을 이루기보다, 도구 전환 자체가 지니는 한계를 확인하는 국면이 되었다. 생성형 AI와의 협업에서 중요한 것은 코드 자체가 아니라 그 코드에 이르는 과정의 맥락이며, 이러한 맥

락이 도구 간 연속성을 형성한다는 점이 확인되었다.

3-5. 4단계: 다도구 교차 검증 방식 도입

1차 심사는 이미지 휘발 문제가 충분히 해결되지 못한 상태에서 운영되었다. 심사 당일 시스템에서 작품 이미지가 안정적으로 표시되지 않아, 발표 자료를 통해 이미지를 확인하는 보완적 방식으로 심사가 진행되었다. 이로 인해 시스템 자체의 운영은 가능하였으나, 본래 설계 의도에 부합하는 방식으로 심사가 이루어지지지는 못하였다.

1차 심사 직후 연구자는 AI와의 협업 방식을 재구성하였다. 앞선 단계가 개별 도구를 순차적으로 활용하는 방식이었다면, 이 단계에서는 동일한 질문을 Gemini, ChatGPT, Claude에 각각 제시하고 그 응답을 상호 비교하는 방식을 시도하였다. 이러한 접근은 결과적으로 한 도구의 답변을 다른 도구의 응답과 대조하여 검토하는 교차 검증의 형태로 정착되었다. 비개발자 교육자가 코드 자체의 적절성을 직접 판단하기 어려운 상황에서 복수 도구의 응답을 비교하는 방식은 협업 결과의 신뢰도를 높이는 실질적인 검증 장치로 작동하였다. 이미지 휘발 문제 역시 이러한 교차 검증 과정을 통해 데이터 저장 방식을 외부 저장 구조로 전환하는 방향으로 최종 해결되었다.

이와 같은 협업 방식은 시스템 전반의 정비 과정에도 적용되었다. 학생 자료 제출, 심사위원 평가 입력, 관리자 운영, 서버 처리의 네 영역에 걸쳐 누적되어 있던 문제들이 다도구 교차 검증을 통해 순차적으로 정비되었으며, 이는 2차 심사의 안정적 운영을 가능하게 하는 기반이 되었다. 이 단계는 비개발자 교육자가 단일 도구에 의존하는 협업에서 벗어나, 각 도구의 응답을 비교하고 조정하면서 활용하는 방식으로 협업을 발전시켜 간 국면으로 이해할 수 있다.

4. 시스템 적용 및 운영 결과

4-1. 시스템 적용 개요

본 장은 3장에서 기술한 시스템이 실제 졸업작품 심사 운영에 적용된 결과를 분석한다. 1차 심사는 시스템이 2단계의 ChatGPT 기반 구조 재설계까지 완료된 상태에서 운영되었으며, 이후 1차 심사 직후 진행된 4단계의 다도구 교차 검증 기반 정비를 거쳐 2차 심사에 다시 적용되었다. 이에 본 장에서는 1차와 2차

심사의 운영 결과를 비교함으로써 시스템 적용의 성과와 한계를 함께 검토하고자 한다.

시스템은 2026년 3월의 1차 심사와 2026년 5월의 2차 심사에 적용되었다. 1차 심사에는 대상자 28명 중 미참여 1명을 제외한 27명의 학생과 심사위원 4명이 참여하였고, 2차 심사에는 25명의 학생과 동일한 심사위원 4명이 참여하였다. 학생들은 시스템에 접속하여 학과에서 지정한 제출 항목에 따라 기획안을 업로드하였으며, 심사위원은 동일한 시스템 안에서 학생별 자료를 열람하고 평가 항목에 따라 점수를 입력하였다. 입력된 점수는 실시간으로 집계되어 관리자 화면에 반영되었고, 최종 결과는 엑셀 파일 형태로 출력되어 행정 처리에 활용되었다.

본 장의 분석 자료는 1차와 2차 심사 운영 과정에서 시스템이 산출한 점수 데이터와 평가 항목별 집계 결과로 구성된다. 점수 데이터는 평균 총점, 표준편차, 점수 구간별 분포를 중심으로 분석하였으며, 운영 과정에서 확인된 예외적 데이터 양상과 그 처리 방식도 함께 검토하였다.

4-2. 1차 심사 운영 결과

1차 심사에서 시스템은 기존 수기 방식에서 나타났던 운영상의 한계를 상당 부분 완화하였다. 기존 방식에서는 종이 심사서를 수령한 뒤 이를 별도로 전산 입력하는 과정이 필요하였으나, 본 시스템에서는 심사위원이 입력한 점수가 실시간으로 관리자 화면에 집계되었다. 이를 통해 심사 진행 중 누락 여부를 즉시 확인할 수 있었고, 심사 종료 시점에는 별도의 재입력 과정 없이 최종 결과를 확보할 수 있었다. 집계 결과는 엑셀 파일 형태로 출력되어 이후 행정 처리에 활용되었다.

응시자 27명을 기준으로 점수 분포를 분석한 결과는 [표 2]와 같다. 평균 총점은 83.10점, 표준편차는 4.11점으로 나타났으며, 최고점은 86.83점, 최저점은 76.33점으로 확인되었다. 이는 전체적으로 약 10점 내외의 범위 안에서 학생 간 점수 차이가 형성되었음을 보여 준다.

[표 2] 1차 심사 점수 구간별 분포 (n=27)

점수 구간	학생 수	비율
85점 이상	8명	30%
80~84점	14명	52%
75~79점	5명	19%

시스템 운영 과정에서는 예외 사례도 확인되었다. 000 학생의 평균 총점이 42.5점으로 산출되어 다른 학생들과 현저한 차이를 보였는데, 분석 결과 이는 학생의 실제 수행 결과라기보다 시스템의 평균 산출 방식에서 비롯된 문제로 확인되었다. 한 심사위원이 해당 학생의 평가에 참여하였으나 점수 입력이 누락된 채 저장되었고, 시스템은 입력되지 않은 값을 0점으로 인식하여 평균값에 반영하였다. 그 결과 다른 심사위원의 정상적인 평가 점수와 합산되는 과정에서 평균 총점이 비정상적으로 낮게 산출되었다. 이후 누락된 점수는 수기로 보정하여 처리하였으며, 이 사례는 시스템이 의도적 최하점 부여와 단순 미입력 상태를 구조적으로 구분하지 못하고 있음을 보여 준다.

이러한 발견은 시스템 정비의 필요성을 분명하게 보여 주었으며, 1차 심사 직후 진행된 4단계의 다도교차 검증 기반 정비 과정에서 평균 산출 로직과 평가 상태 구분 방식이 함께 보완되었다.

4-3. 2차 심사 운영 결과

4단계 정비를 거친 시스템은 2026년 5월, 2차 심사에 다시 적용되었다. 1차 심사 이후 보완된 평균 산출 로직과 평가 상태 구분 방식은 실제 운영에서 안정적으로 작동하였으며, 심사위원의 점수 입력, 실시간 집계, 결과 출력의 전 과정도 큰 문제 없이 진행되었다.

응시자 25명을 기준으로 점수 분포를 분석한 결과는 [표 3]과 같다. 평균 총점은 84.64점, 표준편차는 3.97점으로 나타났으며, 최고점은 91.50점, 최저점은 76.50점으로 확인되었다. 1차와 비교할 때 평균은 1.54점 상승하고 최고점은 4.67점 상승한 반면, 표준편차의 변화는 0.14점에 그쳤다. 이러한 점수 변화는 1차 이후 학생들의 작품 발전 등 외적 요인이 함께 반영된 결과로 보아야 하며, 점수 자체를 시스템 정비의 효과로 직접 해석하기는 어렵다.

[표 3] 2차 심사 점수 구간별 분포 (n=25)

점수 구간	학생 수	비율
90점 이상	3명	12%
85~89점	8명	32%
80~84점	10명	40%
75~79점	4명	16%

1차 심사에서 확인된 예외 사례, 즉 미입력 항목이

0점으로 처리되어 평균이 비정상적으로 산출되는 현상은 2차 심사에서는 나타나지 않았다. 이는 4단계 정비를 통해 미입력 항목을 평균 산출의 분모에서 제외하는 방식이 적용되었기 때문으로 볼 수 있으며, 그 결과 시스템은 의도적인 최하점 부여와 단순 미입력 상태를 구조적으로 구분할 수 있게 되었다.

2차 심사의 운영 결과는 1차 심사에서 드러난 시스템상의 한계가 4단계 정비를 통해 일정 부분 해소되었음을 보여 준다. 다음 절에서는 1차와 2차의 운영 결과를 종합적으로 비교하고, 시스템 개입이 가져온 효과를 정리한다.

4-4. 1차와 2차 결과의 비교 분석

1차와 2차 심사의 운영 결과를 종합하면 [표 4]와 같이 정리된다. 평균 총점은 83.10점에서 84.64점으로 1.54점 상승하였고, 표준편차는 4.11점에서 3.97점으로 변화하였다. 최고점은 86.83점에서 91.50점으로 4.67점 상승했지만, 최저점은 76.33점에서 76.50점으로 큰 변화 없이 유지되어 점수 분포의 상단이 확장된 양상이 확인되었다. 또한, 1차에서 발생하였던 미입력 처리로 인한 예외 사례는 2차에서는 확인되지 않았다.

[표 4] 1차2차 심사 운영 결과 비교

구분	1차	2차	변화
응시자 수	27명	25명	—
평균 총점	83.10	84.64	+1.54
표준편차	4.11	3.97	-0.14
최고점	86.83	91.50	+4.67
최저점	76.33	76.50	+0.17
예외 사례	1건	0건	—

이러한 변화는 1차 심사에서 발견된 시스템상의 한계가 4단계의 다도구 교차 검증 기반 정비를 통해 보완되었음을 보여 준다. 특히 미입력 항목과 의도적 최하점 부여를 구조적으로 구분하게 된 점은 시스템 차원의 직접적인 개선에 해당하며, 평가 결과 산출의 정확성을 높이는 데 기여하였다.

다만 1차와 2차의 점수 변화 전부를 시스템 개입의 효과로 단정하기는 어렵다. 평균 상승과 점수 분포의 상단 확장은 1차 심사 이후 학생들의 작품 발전이 함께 반영된 결과로 보아야 하며, 시스템 정비의 효과는 평가 결과 산출의 정확성과 운영의 안정성 측면에서 보다 분명하게 확인된다고 할 수 있다.

5. 결론

본 연구는 실기 중심 예술·디자인 교육의 졸업작품 심사에서 나타나는 반 디지털화된 한계를 배경으로 학과 특화형 디지털 심사 시스템의 설계·적용 과정을 실험연구의 관점에서 분석한 것이다. 연구자는 프로그래밍 전문성을 갖추지 않은 비개발자 교육자임에도 생성형 AI와의 자연어 기반 협업을 통해 시스템을 직접 구축하고 실제 심사 현장에 적용하였다. 그 출발점은 종이 심사서와 디지털 자료가 분리된 운영 방식에서 반복되던 집계의 비효율과 결과 산출의 불안정성을 학과 맥락에 맞게 개선할 필요성에 있었다.

시스템 개발은 네 단계로 전개되었다. 1단계에서는 Gemini를 활용한 초기 프로토타입을 통해 자연어 기반 협업의 가능성과 함께, AI의 설명이 연구자의 이해 범위를 벗어날 때 협업이 정제될 수 있다는 한계를 확인하였다. 2단계에서는 ChatGPT를 중심으로 구조 전반을 재설계하여 운영 가능한 안정성을 확보하였고, 3단계에서는 Claude 협업을 시도하였으나 누락된 개발 맥락이 공유되지 않은 상태의 도구 전환이 작업의 연속성을 약화시킬 수 있음을 확인하였다. 4단계에서는 세 도구의 응답을 비교·대조하는 다도구 교차 검증 방식을 도입하여 1차 심사에서 드러난 한계를 보완하였다.

시스템은 1차 심사에 적용된 뒤 4단계 정비를 거쳐 2차 심사에 다시 적용되었다. 평균 총점은 83.10점에서 84.64점으로 변화하였으나, 이러한 점수 변화는 학생 작품의 발전 등 외적 요인이 함께 작용한 결과로 보아야 하며 시스템 개입만의 효과로 단정하기는 어렵다. 반면 1차에서 미입력 항목이 0점으로 처리되어 평균이 비정상적으로 산출되던 예외 사례는 평균 산출로직과 평가 상태 구분 방식이 보완되면서 2차에서는 나타나지 않았다. 이는 학생 집단의 변화와 무관하게 시스템 정비로 설명되는 결과로, 평가 결과 산출의 정확성과 운영 안정성이 개선되었음을 보여 준다.

본 연구의 의의는 세 가지로 정리할 수 있다. 첫째, 본 연구는 실기 중심 예술·디자인 교육의 졸업작품 심사가 범용 평가 도구만으로는 수용되기 어려운 복합적 운영 구조를 지니며, 전공의 심사 맥락을 반영한 학과 특화형 디지털 평가 환경의 재구성이 필요함을 사례를 통해 보여 주었다. 이는 디지털 전환을 도구 도입이 아니라 평가 환경의 재설계 문제로 이해할 필요가 있음을 시사한다.

둘째, 본 연구는 비개발자 교육자의 생성형 AI 협업

이 단일 도구 의존을 넘어 다도구 교차 검증 방식으로 확장될 수 있음을 확인하였다. 동일한 질문을 복수의 도구에 제시하고 그 응답을 비교·대조하는 방식은 코드의 적절성을 직접 판단하기 어려운 비개발자에게 협업 결과의 신뢰도를 높이는 실질적 검증 장치로 작동하였다.

셋째, 본 연구는 교육 현장의 도메인 전문성이 시스템 설계에 직접 개입할 수 있음을 보여 주었다. 범용 학습관리시스템으로는 반영되기 어려운 심사 흐름, 자료 제시 방식, 평가 항목 구성, 결과 처리 방식을 전공 교육자가 직접 정의하고 반영한 과정은 교육자의 실천적 판단이 디지털 평가 체계의 구성 원리로 전환될 수 있음을 보여 준다.

본 연구는 다음과 같은 한계를 갖는다. 연구 범위가 단일 학과의 심사 운영 사례에 한정되어 있어 확인된 설계 원칙과 운영 경험의 일반화에는 추가 검토가 필요하며, 심사위원과 학생을 대상으로 한 사용자 평가가 이루어지지 않아 사용자 경험 측면의 인식을 정량적으로 확인하지 못하였다. 또한, 연구자가 개발 주체이자 운영자, 연구자를 동시에 담당하였다는 점에서 관찰과 해석의 객관성에 일정한 한계가 있다. 다만 본 연구는 이를 보완하기 위해 분석의 근거를 시스템이 산출한 점수, 집계 결과, 협업 로그 등 검증 가능한 자료에 두었으며, 점수 변화를 전적으로 시스템 효과로 단정하지 않고 학생 작품의 발전 등 외적 요인이 함께 작용한 결과로 해석하였다.

이러한 한계를 바탕으로 향후 과제를 제안하면 다음과 같다. 첫째, 미평가 학생의 자동 감지와 이상 데이터 경고 표시 등 운영 예외 상황에 대응하는 기능의 보완이 필요하다. 둘째, 다도구 교차 검증 방식이 다른 유형의 교육 운영 시스템 개발에서도 유효하게 작동하는지에 대한 후속 검토가 필요하다. 셋째, 본 연구는 비개발자 예술 교육자의 생성형 AI 협업 기반 시스템 구축이라는 동일한 방향 아래 진행되는 시리즈 연구의 1편으로, 후속 연구에서는 동일한 협업 방식으로 개발·운영 중인 디자인 과제 AI 자가진단 시스템의 운영 결과와 그에 대한 AI·학생·전문가 응답의 비교·분석을 통해 본 연구의 발견을 확장하고자 한다.

참고문헌

1. McNiff, J., 『Action Research: All You Need to Know』, SAGE Publications, 2017
2. 김동삼·류다현·박규동·강혜신, ‘AI기반 실시간 온라인 교육 맞춤형 피드백 시스템 효과성 평가’, 한국컴퓨터교육학회, 컴퓨터교육학회 논문지, Vol.27, No.7, 2024
3. 안인가·이미영, ‘생성형 인공지능 활용 초등 미술수업 개발 실행연구’, 한국초등미술교육학회, 미술과 교육, Vol.25, No.1, 2024
4. 장윤경·윤인화·박지원·제갈유진·이지연, ‘이미지 생성 AI 도구의 이해와 미술교육의 활용 가능성 탐구’, 한국조형교육학회, 조형교육, Vol.88, 2023
5. 정옥희, ‘디지털 전환 미술교육의 파르마콘적 실천 딜레마’, 한국예술교육학회, 예술교육연구, Vol.21, No.2, 2023
6. Mollick, E. & Mollick, L., ‘Instructors as Innovators: A Future-focused Approach to New AI Learning Opportunities, With Prompts’, The Wharton School, The Wharton School Research Paper (SSRN 4802463), 2024
7. Sarkar, A., ‘Will Code Remain a Relevant User Interface for End-User Programming with Generative AI Models?’, ACM, Proceedings of the 2023 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (Onward! 2023), 2023