JPEE 실천공학교육논문지
JOURNAL OF PRACTICAL ENGINEERING EDUCATION

실천공학교육 특별호
J. Pract. Eng. Educ. 16(1), 21-31, 2024

ChatGPT를 활용한 AI 글쓰기 의사소통 역량 평가도구 개발 과정에 대한 연구: 기술 전문가와의 상호소통을 중심으로

An Exploratory Study on Developing the AI Essay Test Tool based on ChatGPT: Focusing on the Interaction with the Engineer

박소영[1], 이병윤[2]*, 홍유정[3]

[1]숙명여자대학교 교육학부, [2]숙명여자대학교 교육연구소, [3]서울대학교 윤리교육과

So-Young Park[1], ByungYoon Lee[2]*, Yujung Hong[3]

[1]Division of Education, Sookmyung Women's University, Seoul 04310, Korea
[2]Education Research Institute, Sookmyung Women's University, Sookmyung Women's University, Seoul 04310, Korea
[3]Department of Ethics Education, Seoul National University, Seoul 08826, Korea

[요 약]

본 연구는 ChatGPT를 활용한 글쓰기 의사소통 역량을 평가하는 AI 평가도구 개발에 초점을 두었다. 개발 과정에 있어 내용 전문가와 기술 전문가 간의 상호작용을 통해 IT와 인문사회 영역의 융합과정을 탐구하였다. 내용 전문가와 기술 전문가는 긴밀한 소통 및 상호작용을 통해, 글쓰기 의사소통 역량의 채점기준을 탑재하였고 각 영역별(내용의 적절성, 조직의 효과성, 어법의 정확성) 점수와 피드백을 제공하는 AI 평가도구를 개발하였다. 이 과정에서 본 연구는 내용과 기술이 어떻게 결합하는지에 대한 과정을 드러내고, 이후 생성형 AI를 활용한 평가도구 기술 전문가를 포함한 융합연구자들이 유의해야 할 사항 등에 대해 제시하였다.

[Abstract]

This study focused on the development of an AI essay tool for assessing writing-communication competence using ChatGPT. During the development process, the interaction between content expert and technical expert was emphasized to explore the fusion of IT and humanities and social sciences. Through close communication and interaction between the content and technical experts, they incorporated scoring criteria for writing-communication competence and developed an AI essay test tool that provides scores and feedback in the appropriateness of content, effectiveness of organization, and accuracy of grammar. This process revealed how content and technology combine and presented considerations for future fusion researchers, including technical experts in generative AI assessment tools.

Key Words: ChatGPT, AI essay test tool, interaction, writing-communication competence, content experts, technical experts

http://dx.doi.org/10.14702/JPEE.2024.021

Received 15 December 2023; **Revised** 10 January 2024
Accepted 29 January 2024

***Corresponding Author**
E-mail: lee.byungyoon12@gmail.com

I. 서 론

ChatGPT(Generative Pre-trained Transformer)의 등장으로 인공지능(AI)을 활용한 플랫폼 개발은 기하급수적으로 증가하고 있으며, 이는 교육 분야에서도 크게 다르지 않다[1]. 정부는 예비교원과 현직교원을 대상으로 하는 AI 역량 강화 사업을 실시하고 있으며, 이 중 ChatGPT를 활용한 교수학습자료 생성하고 평가하는 등의 내용을 포함하고 있음은 물론이다[2].

ChatGPT의 교육적 활용에 대한 연구도 빠르게 증가하고 있다. 최근 한 선행연구에 따르면, 2022년 12월 이후 2023년 5월까지 6개월의 기간동안 국내와 국외에서 약 72건의 논문이 발표되었다고 보고하였다[3]. 분석 대상 논문에서 교과 내 활동과 과제 등에서 ChatGPT를 활용하는 논문의 비중이 높았지만 본 연구가 평가도구에 관심을 두고 있다는 점을 고려하여 살펴보았을 때, 평가와 관련하여 ChatGPT의 성능을 분석하는 논문도 있었다. 그러나 이들 논문은 모두 이미 만들어진 ChatGPT를 활용한 결과와 관련된 논문으로, ChatGPT를 활용하여 도구를 만드는 과정에 대해서는 다루지 않고 있다. 이는 ChatGPT를 활용한 평가를 일회적으로 사용하거나 아직 이를 연구의 주제로 다루지 않고 있기 때문이다.

한편, 에듀테크 분야에서는 폐쇄형 답안뿐만 아니라 서술형 답안도 채점하는 서비스를 제공하는 기업이 빠르게 증가하고 있다[4]. 서술형 답안에 대한 평가는 역량 평가가 도입되면서 그 중요성이 커지고 있지만, 실제 교육 현장에서는 교사 1인이 채점하는 경우가 많아 학생에 대한 사전정보 및 피로도 등으로 채점의 신뢰도 문제가 발생하기도 한다[5]. 이를 극복하기 위한 하나의 방법으로 AI 기술을 활용한 평가도구 개발에 관심이 높아지고 있는데[6], 이미 여러 에듀테크 기업들에서는 글쓰기 AI 자동평가뿐만 아니라 피드백 기능도 함께 제공하고 있다[7]. 그러나, 수요자들은 이런 채점이 얼마나 정확한 것인지 어떤 개념 하에서 채점되고 있는지에 대해서는 확인하지 못하고 있으며 기업들은 이에 대한 내용을 제공하지 않고 있다.

본 연구에서는 학습자가 습득한 지식을 바탕으로 다양한 주제에 대해 자신의 견해를 글로 명확하게 전달하는 능력인 글쓰기 의사소통 역량을 평가하는 AI 평가도구를 개발하였다. 의사소통은 자신의 생각을 표현하고 이해시키며, 상호교환하는 여러 활동을 포함한다[8]. 본 연구에서 대상으로 하는 평가도구는 글을 매개체로 사용하여 저자의 의견을 효과적으로 전달하는 능력을 평가한다. 이 연구에서는 글쓰기 의사소통 역량을 평가하기 위해, ChatGPT와 같은 AI 기술을 활용하여 내용 전문가의 글쓰기 의사소통 역량과 평가에 관한 전문지식을 통합한 평가도구 개발 과정에 관심을 두고 다음의 사항에 주목하였다.

첫째, 내용과 기술의 결합에 주목하였다. 물론 교육 영역에서 오래전부터 사용되고 있는 LMS(Learning Management System) 역시 교육적 관점과 요구가 적극적으로 반영되었지만[9], 이는 학습 내용의 제시 방법과 보다 밀접한 관련을 맺고 있어, 내용과의 관련성은 상대적으로 낮았다. 그러나 ChatGPT를 활용한 평가도구 개발은 도구가 목적으로 하는 내용적 속성을 기술적으로 구현하는 과정이 필요하기 때문에, 도구에 탑재된 내용(채점기준) 역시 전문적이어야 한다. 즉, 단순한 도구의 개발이 아니라, 기술적 요소와 내용적 요소가 보다 밀접하게 융합된 도구의 개발이 필요한 것이다. 이처럼 다학문 간 융합이 보다 빈번하게 요구되는 시점에서, 본 연구에서는 IT 기술(ChatGPT를 활용)과 인문사회의 전문성(타당하고 신뢰로운 채점기준 마련)의 중요한 융합 사례가 될 수 있는 평가도구 개발 과정에 집중하였다.

둘째, 본 연구에서는 개발 과정에서 나타나는 의사소통 및 상호작용의 방식에도 초점을 두었다. 2010년대 들어, 융합인재교육(STEAM) 및 IT 융합형 인재 육성에 대한 관심이 증가함에 따라, 연구자 간 소통 능력은 핵심 역량으로 인식되었다[10]. 서로 다른 분야의 전문가들이 협업을 할 때, 효과적인 의사소통과 상호작용은 연구의 성과를 크게 좌우한다. 그럼에도 불구하고, 다학제 간 협업과 공동연구를 다룬 선행연구는 의사소통 상의 여러 문제점을 지적해왔다. 첫째, 인문학과 공학과 같이 언어 사용이 상이한 분야의 협업에서는 전문용어에 대한 이해 부족과 이를 이해하기 위한 소통 전략의 부재로 인해 공동연구의 진행에 방해가 될 수 있다[11]. 두번째 문제점은 협력 없는 의사소통이다. 각자 과업에 대해 두 전문가 사이의 충분한 정보 공유가 되지 않는 상황, 과업 수행 과정에서 문제가 생겼을 때 함께 해결하려는 노력의 부족, 그리고 서로 간의 의견을 바탕으로 변화를 받아드리려는 노력의 부재 등 적극적인 상호협력이 빠진 의사소통은 여러 문제를 초래한다[12]. 셋째, 기술적 문제해결에만 초점을 맞춘 의사소통은 오히려 의사소통 빈도를 줄일 수 있다는 점에서 문제가 된다[13]. 그러나, 기존 연구들은 두 전문가 간의 구체적인 의사소통 내용, 어려움 등에 대해서는 깊이 있게 다루지 않았다. 따라서, 본 연구에서는 ChatGPT를 활용한 평가도구 개발 과정에서 드러나는 내용 전문가와 기술 전문가 간의 의사소통 및 상호작용의 내용에 심층적으로 집중하고자 하였다. 이를 통해, 향후 IT 영역과 인문사회 영역의 융합과정에 대한 실천적 시사점을 제시할 수 있을 것이라고 기대하였다.

본 연구는 ChatGPT를 활용한 글쓰기 의사소통 역량 평가도구 개발과, 개발 과정에서 두 전문가 간의 상호작용을 면

밀하게 조명하는 것에 초점을 맞추고 있다. 구체적으로, 두 전문가가 어떻게 의사소통하며, 그 과정에서 발생하는 의견 차이와 이를 극복하는 과정을 깊이 있게 분석하였다. 이를 위해 본 연구에서는 내용분석을 활용하여, 평가도구 개발 과정의 각 단계와 상호작용을 분석하였다. 두 전문가 간 이메일 교환 내용, 화상 회의 기록, 각 회의별로 작성된 회의록 등 다양한 원천 자료를 활용하였다.

본 연구에서 연구자들이 설정한 연구 문제는 다음과 같다.

첫째, ChatGPT를 활용한 평가도구 개발 과정에서 채점기준은 어떻게 설정되었는가?

둘째, ChatGPT를 활용한 평가도구 개발 과정은 어떠한가?

셋째, ChatGPT를 활용한 평가도구 개발 과정에서 내용 전문가와 기술 전문가의 상호소통은 어떠했는가?

II. ChatGPT를 활용한 글쓰기 의사소통 역량 평가도구 개발 과정

A. 채점기준 개발

ChatGPT가 평가할 수 있는 글쓰기 의사소통 역량 채점

표 1. 글쓰기 의사소통 역량 채점기준

Table 1. Essay grading criteria

원문항			→	수정문항	
범주	채점 문항	점수		범주	채점 문항
내용	글의 중심 내용이 명확하고 구체적이며, 세부 내용이 중심 내용에 부합한다. 내용이 독창적이고 신선하며 독자에게 흥미롭고 가치 있는 정보를 제공하고 있다.	5점		내용의 적절성	1. 글의 중심 내용이 명확하다 2. 글의 중심 내용이 구체적이다 3. 글의 내용이 독창적이다 4. 글의 내용을 뒷받침하는 근거나 예시가 다양하게 제시되어 있다 5. 저자의 의도를 분명하게 제시한다 6. 글의 전체 내용에 일관성이 있다
	글의 중심 내용이 드러나지만 효과적으로 초점화되지 않는다. 제시한 세부 내용이 중심 내용에 부합하지만, 독창적이지 않고 독자에게 새로운 정보를 주지 않는다.	3점			
	글의 중심 내용이 명확하지 않고, 세부 내용이 거의 없거나 중심 내용에 부합하지 않는다. 글을 읽고 독자가 새로운 정보를 얻기 어렵다.	1점			
조직	글의 구조가 글의 중심 내용을 더욱 잘 드러내는 역할을 한다. 필자가 제공하는 정보의 순서나 구조가 설명하는 대상의 특성을 고려하여 유기적으로 배열되어 있어서 독자가 자연스럽고 흥미롭게 글을 읽을 수 있다.	5점			
	글의 중심 내용을 효과적으로 드러낼 수 있는 도입이나 결론이 명확하게 드러나지는 않지만, 독자가 큰 혼란 없이 내용을 이해할 수 있도록 구성되어 있다.	3점			
	글의 내용을 유기적으로 이어주는 연결 구조가 없어 독자가 글의 내용을 이해하는 데에 혼란을 가진다.	1점			
표현	독창적이고 흥미로우면서도 독자가 쉽고 정확하게 이해할 수 있도록 표현되었다. 글에서 제공하는 정보의 중요성을 독자가 이해하고 공감할 수 있도록 필자의 주체적인 목소리가 잘 드러나 있어서, 글의 내용에 깊이 몰입할 수 있다.	5점		조직의 효과성	1. 글의 구조가 글의 목적에 부합한다 2. 문단과 문단 간 역할이 분명하다 3. 문단 내 중심 문장과 뒷받침 문장이 명료하게 드러나 있다. 4. 필요한 정보가 유기적으로 배열되어 있다
	독자가 쉽게 이해할 수 있도록 표현되었으나 독창성이나 흥미는 다소 떨어진다. 필자의 주체적인 목소리가 잘 드러나지 않아서, 독자는 글의 정보에 대하여 크게 호감을 가지거나 몰입하기는 어렵다.	3점			
	전반적으로 독자의 이해가 어렵게 표현되어 있고, 어조가 글의 목적 및 유형에 부합하지 않는다.	1점			
단어 선택	내용을 정확하고 자연스럽게 전달할 수 있는 단어가 선택되었다.	5점			
	대체적으로 단어 선택이 내용 전달에 무리가 없으나 부적절한 단어들이 포함되어 있다.	3점			
	내용을 전달하는 단어가 매우 제한적이며 단어의 선택이 풍부하지 못하다.	1점			
형식 및 어법	글쓴이는 표준적인 쓰기 관습(어법, 구두점, 철자, 단락 구분 등)을 잘 이해하고 있으며 독자의 가독성을 고려하여 이러한 관습을 효과적으로 사용하고 있다.	5점		어법의 정확성	1. 맞춤법과 띄어쓰기가 지켜지고 있다 2. 어법을 잘 지키고 있다 3. 이해하기 쉽게 문장을 작성하였다
	제한된 범위에서만 표준적인 쓰기 관습을 지키고 있다.	3점			
	어법,구두점,철자 등에서 잘못된 것이 많아 내용 파악을 방해한다.	1점			

주1. 수정된 채점기준은 문항당 1–5점으로 평가됨.
주2. 채점문항 수정 시 참고문헌 [17], [18], [19], [20]를 참고함.

기준을 개발하는 과정은 다음과 같다. [14]가 제안하고 [15]가 한글로 번안한 설명문 글쓰기 평가항목을 바탕으로 채점기준을 수정하였으며, ChatGPT 평가 결과의 신뢰도를 높이기 위하여 다음의 내용을 고려하였다. a) 한 문장에는 하나의 평가 내용을 포함하며, b) 가능한 한 여러가지 의미로 해석될 수 있는 단어를 지양하고, c) 세부 평가문항마다 평가 결과가 제시되도록 하였다. 이와 함께 ChatGPT는 매 평가마다 새로운 판단을 하게 되므로, 연구자가 동일한 에세이에 대한 ChatGPT 평가를 반복 실시하여 평가 결과가 일관성 있게 나타나는지 확인하였다. 수정된 채점기준은 교육학 전공 교수 3인의 자문 및 검토를 통해 범주의 타당성과 범주 내 문항 구성의 적절성을 확인하였다. 원문항과 위 기준에 따라 개발된 ChatGPT 채점기준은 표 1과 같다.

원문항에서는 채점 범주별 5점, 3점, 1점에 해당하는 채점기준이 제시되어 있었으나, ChatGPT 채점기준 개발에서는 각 범주의 5점에 해당하는 평가문항을 중심으로 세부문항을 수정하였다. 이는 5, 3, 1점에 해당하는 평가문항이 연속된 척도로 보기 어려운 부분이 있었기 때문이다. 예를 들어, 단어선택의 경우 5점은 정확하고 자연스러운 단어를 사용하는 것에 초점을 두고 있으나 1점에서는 풍부한 단어 사용이 새로운 기준으로 등장하는 것을 확인할 수 있었다. 인간 채점자의 경우 5, 3, 1점 평가문항을 총체적으로 파악하여 1점과 5점 사이의 평가를 하는 데 반해, ChatGPT는 주어진 각 평가항목에 근거하여 평가 결과를 제시한다. 따라서 평가점수와 점수 산출의 근거가 분명하게 매칭되고, 에세이에 따른 평가 결과 차이를 줄이기 위하여 5점 평가 문항의 평가 항목에 준해 ChatGPT 채점기준을 개발하였다.

또한, 한 문항에 다양한 채점기준이 포함되어 있는 경우, 동일한 에세이에 대한 평가 결과가 매 평가마다 다르게 나타나는 경향이 나타났다. 예를 들어, 내용 범주 5점 평가문항 안에는 내용의 정확성, 내용 사이의 연관성, 내용의 참신성, 주제의 적절성에 대한 내용이 모두 포함되어 있다. 이는 ChatGPT 평가 시 일부 항목이 평가되지 않거나, 동일한 에세이라도 평가 결과가 상이하게 나타나는 문제가 발생하였다. 따라서 한 문장에 하나의 채점기준이 포함되도록 세부 문항을 수정하였다.

이 과정에서 평가내용이 분명하게 구분되도록 하기 위하여 범주를 재구성하였다. 원문항에서는 내용, 조직, 표현(어조 및 태도), 단어 선택, 형식 및 어법 5개 범주로 채점 문항을 구분하고 있으나, 채점 문항 진술은 여러 범주에 공통적으로 나타나는 경우를 확인할 수 있었다. 원문항의 채점 문항을 살펴보면, 내용 범주의 '흥미롭고 가치있는 정보 제공'과 표현(어조 및 태도) 범주의 '독창적이고 흥미로운 표현'이 매우 유사한 것을 볼 수 있다. 이에 따라 내용 전문가는 ChatGPT 평가 시 범주 간 구분이 분명하지 않은 평가 결과가 산출될 가능성이 높다고 판단하고, 각 범주의 의미와 범주 간 채점 문항을 재구성하였다. 최종적으로 내용의 적절성, 조직의 효과성, 어법의 정확성 3개 범주, 총 13개 채점문항으로 확정하였다.

이 외에도 ChatGPT 평가 결과의 해석을 용이하게 하기 위하여 각 평가문항별 1~5점의 점수를 부여하고 이를 합하여 총점을 계산하는 분석적 채점방법을 적용하였다. 이는 범주별 총점을 한번에 요구하는 경우, 평가점수가 동일한 경우에도 점수 산출 근거를 다르게 제시하는 모습이 나타났기 때문이다.

B. 평가도구 개발

본 연구에서는 개발된 글쓰기 의사소통 역량의 채점기준을 ChatGPT-4를 활용하여, 실제 교수자가 글을 업로드하고 채점 결과를 받아볼 수 있는 평가도구(평가플랫폼)를 개발하였다. 이에 앞서, 내용 전문가는 교육학 배경지식을 갖춘 기술 전문가와의 협업이 중요하다고 판단하였다. 이는 본 평가플랫폼 개발이 교수자가 타당한 채점기준으로 유의미하고 효율적인 채점 경험을 제공하도록 구현하는 데 목적이 있기 때문이다. 따라서, 본 내용 전문가는 컴퓨터공학을 전공하고, 교육학으로 석사학위를 취득한 프로그램 기술 전문가를 섭외하였다. 총 평가도구 개발 기간은 3개월 정도 소요되었다. 주요 과정은 표 2와 같다.

1) Step 1: 평가도구 구현 범위 논의

글쓰기 의사소통 역량 평가플랫폼 개발은 두 단계(phase)로 계획되었다. Phase 1은 교수자가 학생들의 파일을 업로드한 후, ChatGPT-4를 활용하여 글쓰기 의사소통 역량 평가가 이루어지고, 채점결과를 제시하는 일련의 과정을 구현하는 것이다. Phase 2는 Phase 1에서 완성된 UI(user interface)를 토대로, 다른 역량을 추가하고 추가된 역량의 채점기준 등 채점 틀 변경에 따른 평가 결과 조정과 관련된 부분이다.

Phase 1에서는 여러 개의 파일이 한번에 업로드 되고, 다양한 파일 형식(hwp, docx, pdf)이 지원되도록 하였고, OpenAI 기반(ChatGPT-4) 답변 자동 생성 모듈을 구현하여 자동채점이 가능하도록 하였으며, 채점 결과를 점수와 서술형 피드백 형태로 화면에 보여주는 방식을 구현하였다.

Phase 2에서는 평가하고자 하는 역량을 추가하고 연구자가 역량별 채점기준을 수정할 수 있도록 프로그래밍 하는 것을 포함한다. 이는 본 평가플랫폼 개발이 단순히 평가도구

표 2. 글쓰기 의사소통 역량 평가도구 개발 순서

Table 2. Development process of the AI essay test tool

Step 1		Step 2		Step 3
평가도구 구현 범위 논의	→	평가도구 기본 틀 구성	→	평가도구 활용성 검토 (1차버전)
Step 4		**Step 5**		**Step 6**
→ 1차버전 피드백 반영 및 2차버전 완성	→	평가도구 활용성 검토 (2차버전)	→	Phase 1 최종버전 완성

개발에 목적을 둔 것이 아니라, 타당한 채점기준에 따른 신뢰도 높은 평가도구를 개발하는 데 목적을 두고 있기 때문이다. 따라서 역량별 채점기준이 주기적으로 업데이트 될 수 있다는 점을 반영할 예정이다. 단, Phase 2는 Phase 1에서 개발된 기본 평가도구 틀에 추후 역량과 역량별 채점기준을 추가하는 것이므로, 본 연구에서는 Phase 1의 개발 단계에 초점을 두었다.

2) Step 2: 평가도구 기본 틀 구성

우선, 평가도구의 기본 틀을 다음과 같이 구체화하였다. 첫째, 채점기준 틀에 따라 각 세부항목을 ChatGPT-4가 평가할 수 있도록 자동화하는 것이다. 둘째, 다수의 파일이 업로드 되고, 각 파일의 글 내용이 채점되는 것이다. 본 평가플랫폼은 교수자용을 개발하는 것이므로, 채점의뢰자(예: 교사, 교수)가 여러 명의 글을 한꺼번에 업로드하여 자동 채점한 결과를 용이하게 활용할 수 있도록 하기 위함이다. 셋째, 학생들이 글쓰기를 하는 환경 및 저장방식이 상이할 수 있으므로, 채점 파일은 여러 파일 형식(예: hwp, docx, pdf 등)이 모두 인식되도록 하는 것이다. 넷째, 평가 결과는 평가 점수와 함께 서술형의 피드백도 함께 제공할 수 있도록 하는 것이다. 이를 통해 채점기준을 토대로 계산되는 양적인 점수 뿐만 아니라, 채점의뢰자에게 적절한 피드백을 제공하여 더 교육적인 평가도구로써 기능할 수 있기를 기대하였다.

3) Step 3: 평가도구 활용성 검토 (1차버전)

완성된 평가도구의 1차버전을 토대로, 활용성 및 평가의 타당성을 검토하였다. 내용 전문가와 3명의 전문가(교육학 전공 교수)가 초기 버전을 활용하여 실제 글을 채점하였다. 검토 후 1차버전에 나타난 문제와 제안 방향은 다음과 같다.

첫째, 파일 형태 외에 압축파일(zip 파일)도 업로드하는 방향이 제안되었다. 교수자가 학생들의 글을 한꺼번에 다운로드 받는 경우, 그 파일이 종종 압축파일로 저장되는 경우가 있어, 본 연구팀은 여러 개의 개별파일 업로드하는 기능 외에, 압축파일을 업로드 하는 것이 필요하다고 의견을 모았다. 둘째, ChatGPT-4가 채점할 수 있는 글자 수의 제한이 있었다. 제한 용량은 약 8,000자 (8,192 토큰) 정도로 긴 글의 평가를 진행하기에는 한계가 있었다. 따라서 글자 수가 초과되는 경우, 에러메시지를 띄우도록 설정한 상태였으나, 이를 근본적으로 해결할 필요성이 있었다. 셋째, 채점 결과는 화면에 제시되는 것이 아니라, 결과 내용이 담긴 엑셀파일을 다운로드하는 방식으로 변경하는 것이 필요하였다. 이는 교수자가 여러 학생의 결과를 한눈에 확인하고 정리하기에는 화면에 나타내는 방법보다는 엑셀파일로 제시하는 방법이 용이할 수 있기 때문이다. 넷째, 총점 뿐만 아니라 채점 범주별 합산 점수, 각 문항별 점수도 함께 제시할 필요가 있었다. 이와 함께 세부 채점 범주와 문항 번호 등을 함께 제시하는 것이 필요하다고 판단하였다. 다섯째, 본 평가플랫폼은 ChatGPT-4가 채점한 결과이므로, 해석이나 사용 시 주의가 필요하다는 문구 제시가 필요할 수 있다는 의견이 있었다.

4) Step 4: 1차버전 피드백 반영 및 2차버전 완성

내용 전문가와 기술 전문가는 1차버전 검토 사항을 논의하는 과정을 통해, 구현된 1차버전 수정안에는 다음의 사항들이 반영되었다.

첫째, 여러 파일을 한번에 업로드하는 것 뿐만 아니라 압축파일도 업로드할 수 있도록 평가도구를 수정하였다. 둘째, 채점가능한 글자 수와 관련하여, 기술 전문가는 내용 전문가가 글의 길이와 관련하여 다양한 테스트를 해볼 수 있도록 플레이그라운드[1]를 공유해주었다. 내용 전문가는 GPT-3.5 모델과 GPT-4.0 모델[2]을 활용하여 다양한 길이의 글을 실제 평가하는 테스트를 진행하였다. 이를 통해 GPT-3.5 모델이 GPT-4.0 모델에 비해 더 긴 글 채점할 수 있는 것으로 확인되었다. 이에 따라, 본 평가도구의 전체 포맷은 GPT-4.0 모델로 세팅하되, 긴 글은 알아서 GPT-3.5 모델로 전환되는 방식을 적용하였다. 그리고 결과 파일의 비고란에 두 모델 중

[1]사용자가 자유롭게 텍스트를 입력하고 GPT 모델의 응답을 확인하며, 다양한 작동 모드를 활용하고 다양한 GPT 모델 버전을 실험할 수 있는 상호작용형 플랫폼을 의미함. 사용자가 GPT 모델과의 상호작용을 개인화하여 모델의 기능을 관찰할 수 있다는 특징이 있음(출처: https://www.digitalfocus.news/bbs/board.php?bo_table=news&wr_id=1815)

[2]모델을 지칭할 때에만 GPT-4.0이라고 기술함

어떤 것으로 채점되었는지 표기되도록 하였다.

셋째, 채점 결과는 두 가지 방식으로 제시되도록 하였다. 첫번째는, 엑셀파일에 저장되고, 이를 사용자가 다운로드를 받을 수 있도록 하였다. 두번째는, 구글드라이브의 스프레드시트를 연동시켜, 사용자가 [결과 확인하기]를 클릭하면 결과가 저장된 스프레드시트를 볼 수 있도록 구현하였다. 넷째, 결과 파일에는 구체적으로 한 행(row)마다 각 글의 채점 결과가 저장되도록 하였다. 또한 학생을 구분 짓는 학생ID를 부여하고(예: 채점한 날짜와 연번의 조합), 원문파일명과 원문 내용도 결과 파일에 저장되도록 하여, 저장된 채점 결과를 활용하거나 추후 연구에 용이하도록 하였다. 따라서, 결과 파일에 학생ID, 각 채점 범주를 표시하는 이니셜[3], 채점 문항의 번호, 각 채점 범주별 합산 점수, 전체 총점, 서술형 피드백, 원문파일명, 원문 내용이 순서대로 저장되도록 하여, 사용자가 채점한 여러 파일을 쉽게 구별할 수 있도록 하였다. 다섯째, 내용 전문가는 "이 점수는 내용 전문가가 개발한 채점기준을 활용하여 ChatGPT-4가 채점을 시행한 결과로, 향후 채점의 신뢰도와 타당도를 평가하고 개선하기 위한 연구자료로 활용됩니다. 현재 단계에서 ChatGPT-4의 채점 결과는 실제 해당 역량의 특성을 충분히 반영하고 있지 않을 수 있으므로, 해석과 사용 시 주의가 필요합니다."라는 문구를 평가플랫폼 접속 첫 화면에 삽입하였다.

5) Step 5: 평가도구 활용성 검토 (2차버전)

2차버전은 대학에서 실제 강의를 하는 교수진 3인과 교육학 전공 대학원생들에게 활용하게 함으로써, 이 도구를 상용화시켰을 때의 문제점을 최소화할 수 있도록 하는 것을 목표로 하였다. 2차버전을 검토한 피드백과 수정한 사항은 다음과 같다.

첫째, 결과 파일에 함께 기입되는 원문파일명과 본문 내용에 일부 글자 깨짐 현상이 발생했다. 이는 특히 한글파일(hwp)을 사용할 경우 발생되었는데, 실제 본문이나 파일명에는 없는 한자나 특수기호 등에 대한 처리가 필요하였다. 이를 위하여 한글파일의 경우, 파일 업로드 후 후처리 과정을 추가하여 한자, 특수기호를 일괄 제거하도록 하였다. 이때, 실제로 유효한 한자까지 모두 삭제될 수 있다는 문제점이 있으나, 내용 전문가는 실제 평가할 학생의 글에서 한자를 사용하는 경우가 많지 않다는 점, 반면 글자 깨짐 현상으로 인해 원문파일명이나 원문 내용 구별이 어렵다는 점에서, 한자와 특수기호를 후처리로 모두 삭제하였다.

[3] C = Contents (내용의 적절성), S = Structure (조직의 효과성), G = Grammar (어법의 정확성) (예: C_1, S_1, G_1)

둘째, 글쓰기 의사소통 역량에 대한 사전 지식이 없는 사용자에게는 결과파일에 C, S, G 등으로 표현된 각 범주의 이니셜만을 가지고는 정확한 역량에 대한 정보를 파악할 수가 없다는 피드백이 있었다. 따라서 본 평가플랫폼 첫 화면에 글쓰기 의사소통 역량과 각 채점 범주에 대해 소개하는 웹페이지로 접속할 수 있는 링크를 생성하고, 그 웹페이지를 구현하는 방안을 모색하였다.

셋째, 본 평가도구의 GPT 모델 세팅(전체 포맷은 GPT-4.0, 긴 글은 GPT-3.5로 전환)과 관련하여서는 OpenAI의 대규모 업데이트로 인해, GPT-4.0의 최신 모델이 더 긴 글을 채점할 수 있게 되었다(8,192 토큰에서 12,800 토큰으로 증가). 따라서, 본 평가도구의 GPT-4.0도 최신 버전으로 교체되면서, 전체 포맷을 GPT-4.0 모델로만 운영할 수 있게 되었다. 따라서, GPT-3.5 모델과 GPT-4.0 모델 중 선택하는 과정을 삭제하였다.

넷째, 업로드 가능한 글자 수를 제한하지 않는 경우 비용이 크게 발생할 가능성이 있었다. 기술 전문가는 분량에 대한 부분에 대해 채점가능한 최대 글자 수를 제한하는 것이 가능하다고 하였고, 내용 전문가는 내부 협의 끝에 40,000자(A4 30쪽 분량)로 제한을 두기로 하였다.

6) Step 6: Phase 1 최종버전 완성

내용 전문가와 기술 전문가는 1차버전과 2차버전을 바탕으로 각 버전에 대한 다양한 피드백을 수렴하였고, 이에 대한 협의 과정을 통해 글쓰기 의사소통 역량 평가도구(교수자용)를 최종 제작하였다(그림 1 참고). 또한, 본 평가도구의 최종버전을 통해 채점된 결과가 저장된 파일 예시는 그림 2에 제시하였다.

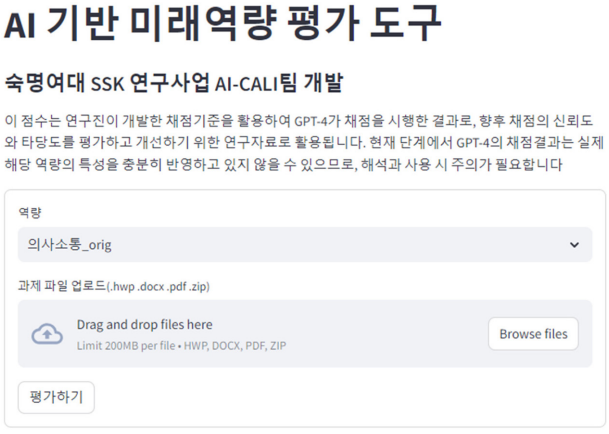

그림 1. 글쓰기 의사소통 역량 평가플랫폼 최종버전 UI 화면

Fig. 1. Final version UI screen of the AI essay test tool.

STU ID	C_1	C_2	C_3	C_4	C_5	C_6	C_total	S_1	S_2	S_3	S_4	S_total	G_1	G_2	G_3	G_total	Total	descriptic	descriptic	descriptic	원문파일명	원문 내용	사용 모델명	비고
231122_020917_1	5	5	4	5	5	5	29	5	5	5	5	20	5	5	5	15	64	중심 내용	글의 구조	맞춤법과	test1	[redacted]	gpt-4-1106-preview	
231122_020917_2	4	4	3	4	4	4	23	4	4	4	4	16	5	5	4	14	53	학생은 능	리포트는	맞춤법과	test2	[redacted]	gpt-4-1106-preview	
231122_020917_3	5	5	4	5	5	5	29	5	5	5	5	20	5	5	5	15	64	학생은 능	리포트는	맞춤법과	test3	[redacted]	gpt-4-1106-preview	
231122_020917_4	4	4	3	4	4	4	23	4	4	4	4	16	5	5	4	14	53	학생은 수	리포트는	맞춤법과	test4	[redacted]	gpt-4-1106-preview	
231122_020917_5	5	5	4	5	5	5	29	5	5	5	5	20	5	5	5	15	64	리포트는	서론, 본론	맞춤법과	test5	[redacted]	gpt-4-1106-preview	
231122_020917_6	4	3	2	3	4	4	20	4	4	4	4	16	5	4	3	12	48	중심 내용	글의 구조	맞춤법과	test6	[redacted]	gpt-4-1106-preview	
231122_020917_7	4	4	3	4	4	4	23	4	4	4	4	16	5	5	4	14	53	리포트는	글의 구조	맞춤법과	test7	[redacted]	gpt-4-1106-preview	
231122_020917_8	5	5	4	5	5	5	29	5	5	5	5	20	5	5	5	15	64	리포트는	글의 구조	맞춤법과	test8	[redacted]	gpt-4-1106-preview	
231122_020917_9	5	5	4	5	5	5	29	5	5	5	5	20	5	5	5	15	64	리포트는	글의 구조	맞춤법과	test9	[redacted]	gpt-4-1106-preview	

그림 2. 글쓰기 의사소통 역량 채점결과 화면

Fig. 2. Results screen of the AI essay test tool.

C. 평가도구 개발을 위한 내용 전문가와 기술 전문가의 상호작용

본 평가도구의 제작 과정에서는 내용 전문가와 기술 전문가 간의 상호작용이 필수적이었다. 특히, 개발 과정에서 의견차가 발생했을 때 이 상호작용은 중요한 역할을 했다. 본 절에서는 내용 전문가와 기술 전문가 간의 주요 상호작용을 기술하였다(표 3 참고). 가장 큰 의견차는 평가플랫폼에 평가할 역량을 어떻게 프로그램화 할 것인지에 대한 논의에서 나타났다. 이 논의는 본 연구의 핵심이며, 본 평가플랫폼의 궁극적인 목적이 채점의뢰자의 요구에 맞춰 여러 역량(예: 교육부의 6대 미래핵심 역량)을 평가할 수 있도록 하는 것이므로, 내용 전문가는 다양한 역량의 채점기준을 탑재하는 기능 구현의 필요성을 강조하였다.

이러한 요청에 대해, 기술 전문가는 평가하고자 하는 역량의 구성(예: 영역의 개수, 세부 채점 문항 개수)을 먼저 파악하는 것이 중요하다고 주장했다. 이는 역량의 채점기준을 프로그래밍하기 위해 각 역량이 몇 개의 층(layer)으로 구분되는지 파악하는 것이 우선이며, 이후 단계에서 추가될 역량도 이 층에 맞춰 정렬되어야 한다는 것이다. 예컨대, 글쓰기 의사소통 역량이라는 대영역을 첫번째 층이라고 하면, 그 아래 중영역(두번째 층, 혹은 채점 범주)과 소영역(세번째 층, 혹은 세부 채점 문항)이 있는지, 그리고 각 층에 몇 개의 요소들이 포함되는지 등을 파악해야 했다.

반면, 내용 전문가는 전문적인 채점기준을 마련한 후 글쓰기 의사소통 역량의 전체 구조를 파악할 수 있다고 주장했다. 구체적인 채점 문항을 구성하면서, 주요 채점요소가 결정되고, 이를 토대로 층의 수가 결정된다고 보았기 때문이다. 결론적으로, 기술 전문가는 평가할 역량을 프로그래밍 하기 위해 top-down식으로 접근하였고, 내용 전문가는 bottom-up식으로 접근하는 주요한 차이를 발견했다. 이를 해결하기 위해, 기술 전문가가 평가도구를 개발하는 동시에, 내용 전문가는 구체적인 채점기준 개발을 하는 병렬적 연구 방향을 선택했다.

내용 전문가과 기술 전문가 사이의 의견 차이는 GPT-4.0 모델로 글자 수가 초과된 글을 처리할 때도 나타났다. 기술

표 3. 평가도구 개발을 위한 내용 전문가와 기술 전문가 간 상호작용

Table 3. Interaction between content experts and technical experts in developing the AI essay test tool

순서	주요 과정	상호작용		
		내용 전문가		기술 전문가
Step 1	평가도구 구현 범위 논의	• 기술적 이해도가 있는 내용 전문가 • 평가도구 기본틀 제시	⟷	• 교육학적 이해도가 높은 기술 전문가 • 기술적 구현 가능한 범위 제안
Step 2	평가도구 기본 틀 구성	• 채점기준 타당화 • 채점기준 업데이트 가능한 시스템 요구	⟷	• 채점기준을 반영한 layer 구현 • 채점기준 변경 및 역량 추가가 인터페이스 구현 제안
Step 3	평가도구 활용성 검토 (1차 버전)	• 압축파일 업로드 필요성 제시 • 평가 상황에서 요구되는 글자 수 확인 • 결과 활용에 필요한 정보 목록 제시	⟷	• 업로드 가능한 파일 형태 확인 • 채점 가능한 글자 수 확인 • 결과 파일 추출가능 범위 확인
Step 4	1차 버전 피드백 반영 및 2차 버전 완성	• 평가 결과 신뢰도를 위하여 하나의 모델 사용 제안	⟷	• 글자 수에 따라GPT-4.0과 GPT-3.5모델 유연하게 활용 제안
Step 5	평가도구 활용성 검토	• 평가 상황을 고려하여 글자 깨짐 현장 처리 방법의 범위 제안	⟷	• 글자 깨짐 현상 등의 처리 방법 제안
Step 6	최종버전 완성			

전문가는 글자 수가 초과된 글을 오류로 처리하거나, GPT-4.0이 처리할 수 있는 만큼만 채점하거나, 긴 글을 선택적으로 GPT-3.5로 채점하는 방법을 제안했다. 하지만 내용 전문가는 부분적인 채점이 평가의 타당성을 해칠 수 있다고 지적했다. 또한, 글자 수에 따라 다른 모델을 사용하는 것은 평가의 신뢰성에 문제를 일으킬 수 있다고 판단했다.

이에 대한 해결책으로 기술 전문가는 플레이그라운드라는 애플리케이션 프로그래밍 인터페이스(API)를 통해 GPT-3.5와 4.0 모델을 비교할 수 있는 방법을 제안했다. 내용 전문가는 이를 통해 두 모델의 채점 능력과 결과의 차이를 여러 차례 테스트해 볼 수 있었다. 그 결과, 두 모델 간에 성능 차이가 크지 않음을 확인하고 긴 글은 GPT-3.5로 채점하기로 결정했다.

결국, OpenAI의 업그레이드로 GPT-4.0의 채점 가능 글자 수가 증가하여 모든 글을 GPT-4.0으로 채점하게 되었다. 이 과정에서 내용 전문가는 채점 결과의 타당성과 내용의 납득 정도에 중점을 둔 반면, 기술 전문가는 기술적 오류 없이 입력에 따른 채점 결과를 효과적으로 제공하는 데 초점을 맞췄다는 것을 알 수 있다.

이 두 사례를 통해, 기술 전문가는 기술의 빠른 변화와 새로운 기술의 적용이 중요하다고 보는 반면, 내용 전문가는 속도보다 정교한 결과를 중시하는 접근 방식의 차이를 보였다. 이러한 차이 인식과 상호작용이 협업의 근본적인 힘이었다. 특히, 교육학 지식을 갖춘 기술 전문가는 내용 전문가의 타당한 결과를 얻기 위해 여러 차례 시험하는 과정의 중요성을 이해하고 기술적 지원을 제공함으로써 협업을 강화했다. 이는 기술과 인문학의 융합에서 단순히 두 분야가 결합된 도구 개발뿐만 아니라, 두 분야의 전문성을 모두 반영하는 도구 개발이 중요함을 보여준다.

III. 결 론

본 연구에서는 ChatGPT-4를 기반으로 하는 교수자용 AI 글쓰기 의사소통 역량 평가도구(평가플랫폼)를 개발하고, 이에 발생하는 내용 전문가와 기술 전문가 간의 구체적인 상호소통 과정을 확인하고자 하였다. 내용 전문가와 기술 전문가는 두 차례에 걸쳐 평가도구를 완성하고, 이를 지속적으로 검토하고 수정하는 작업을 거쳤다. 최종적으로 완성된 평가도구는 40,000자 분량의 글을 다양한 유형의 파일과 압축 파일을 업로드할 수 있는 기능을 포함한다. 이 도구를 통해 교수자는 글의 내용의 적절성, 조직의 효과성, 어법의 정확성 등을 평가받을 수 있으며, 점수와 함께 상세한 피드백을 제공받을 수 있다.

본 평가도구의 개발 과정에서는 연구팀 내부의 협의도 중요했지만, 특히 내용 전문가와 기술 전문가 간의 긴밀한 상호소통과 협력이 필수적이었다. 이는 특히 도구 개발 과정에서 발생한 두 전문가 간 의사소통의 장벽을 극복하고 양측의 요구사항을 반영한 최종 도구를 완성하는 데 핵심적인 역할을 했다. 본 연구에서 발견한 두 전문가 간 의사소통의 가장 큰 어려움은 본 평가도구 개발을 바라보는 두 전문가의 관점 차이에서 비롯되었다. 기술 전문가는 개발된 평가도구가 오류없이 평가 결과가 산출되는 데 집중한 반면, 내용 전문가는 평가 결과를 활용하여 추가 분석을 수행하고 이를 통해 여러 연구로 발전시킬 가능성에 중점을 두었다. 따라서, 내용 전문가는 추후 연구의 깊이를 더하기 위해 필요한 다양한 세부 옵션들, 예를 들어, 교육학 연구에 주로 활용되는 파일 형식, 평가 내용의 저장 형식, 평가기준의 구조, 평가할 글의 길이, 글의 내용, 그리고 학생 ID와 글 내용의 연결 등에 대한 요구사항을 제시하였다. 반면, 기술 전문가는 이러한 요구사항에 대한 이해도가 낮아, 구체적인 옵션 구현에 있어 어려움을 겪었다. 한편, 내용 전문가는 기술적 한계와 구현 가능성을 정확히 이해하는 데 어려움을 겪었으며, 평가도구가 장기적으로 사용될 수 있도록 설계하는 과정에서 평가기준의 설정, ChatGPT를 사용한 프롬프트 입력, 종합된 결과 도출 및 저장 방법 등에 대한 이해가 필요하였다.

그리고 이 어려움을 해결하기 위한 내용 전문가와 기술 전문가 간 소통의 노력은 단순히 요청과 수행의 관계를 넘어섰다. 매 단계에서 인문사회학 분야의 내용 전문가는 전문성을 바탕으로 이 도구를 더 효율적이고 실용적으로 만들기 위해 노력했으며, 기술 전문가도 이러한 전문성을 이해하고 적용하는 과정을 거쳤다. 이는 단순한 평가플랫폼의 개발을 넘어서, 실제 교육 현장에서 교수자들이 효과적으로 사용할 수 있는 도구를 만들기 위한 내용 전문가와 기술 전문가 간의 상호소통이 중요했음을 의미한다. 이는 특히, 학제간 혹은 다학문적 연구의 본질에 대해 두 전문가 간 상호소통의 중요성을 다시 한번 강조하는 계기가 되었다. 산학연 전문가 200여 명을 대상으로 융합기술 분야에 있어 필요한 정책 방향에 대해 탐색한 한 선행연구에 따르면[11], 과거 공동연구 진행 시에는 각 분야의 전문 지식을 단순히 이전하고 결합하는 과정으로 인식되었으나, 각 분야 전문가 간의 상호작용과 상호흡수(co-absorption)를 통한 진정한 융합을 이루는 공동연구가 될 수 있도록 의사소통 능력이 중요함이 밝혀졌다.

본 연구에서는 한발 더 나아가, 기술 전문가와 내용 전문가 간 협업 및 상호소통의 세부 내용과 어려움, 그리고 이를 극복하기 위한 구체적 방안을 탐색하고, 이에 따른 시사점을

도출하였다. 첫 번째로, 전문 분야가 상이한 두 전문가의 협업에서는 간단하고 명확한 언어 사용이 필수적이다. 본 연구에서 기술 전문가는 내용 전문가의 채점기준 설정 과정을 명확히 이해해야 했고, 내용 전문가는 이러한 기준이 기술적으로 어떻게 구현되는지 파악해야 했다. 이 과정에서 전문 용어(jargon)에 대한 충분한 설명이나 다른 용어로 대체함으로써 상호작용을 더 유연하게 하는 데 크게 기여하였다. 이는 특히, 두 전문가 사이에 "교환되는 정보의 정확성"이 의사소통의 질에 영향을 미친다는 점에서[16], 정확한 언어 사용을 통한 정보 교환은 의사소통의 명확성을 높이고, 협업의 효율성을 향상시킴을 밝혔다.

두번째 시사점은 상호지원의 중요성에 관한 것이다. 연구결과, 내용 전문가는 기술 전문가의 지원을 받아 기술적인 면에서 더 깊이 있는 테스트를 수행할 수 있었다. 기술 전문가는 평가도구 제작 뿐만 아니라, 내용 전문가가 프로그래밍을 이해하고 테스트할 수 있도록 적극적으로 지원하였고 이를 통해 프로그래밍과 관련한 의사소통을 원활하게 할 수 있었다. [12]에 따르면, 효과적인 의사소통을 위해서는 타당한 메시지 및 지식 공유 외에도 조직 간의 적절한 지원이 필요하다. 본 연구를 통해 두 전문가 간 부족한 면을 서로 채워주는 형태의 소통 형태를 확인하였다.

셋째, 상호작용 시 서로의 과업에 대한 상세한 기록이 중요하다. 이메일, 화상 회의 내용, 회의록 등을 통해 두 전문가 간의 구체적인 상호작용을 분석할 수 있었다. 이처럼 정밀한 기록과 공유는 수정과 피드백 전달 과정을 효과적으로 만들었다. 공동연구에서 원활한 상호소통을 위한 요건으로 프로젝트 진행 단계에 대한 검토(regular progress review)의 중요성이 강조된다[13]. 즉, 공동연구 진행 과정에서 두 전문가가 자신의 과업과 서로 주고 받은 피드백, 다음 단계 과업, 그 밖에 회의에서 오고 간 대화들을 꼼꼼히 기록하고 이를 서로 교차검토함으로써, 복잡한 협업 과정에서 원활한 의사소통 및 상호작용을 가능하게 하였음을 시사한다.

이처럼 본 연구를 통해, ChatGPT와 같은 생성형 AI를 활용한 평가도구 및 교육적 도구를 개발하려는 미래 기술 전문가와 내용 전문가의 융합연구에 필요한 중요한 지침을 제공했다. 그럼에도 불구하고, 본 연구는 다음의 한계점을 가진다. 첫째, 본 연구는 내용 전문가와 기술 전문가 간 상호작용에 있어 각각의 집단이 어떤 점에서 어려움을 느끼는지에 대해 심층면담을 실시하지는 못하였다. 이는 본 연구가 서로 다른 학제 간 융합과정뿐만 아니라 AI 평가도구를 개발하는 과정을 드러내보이고 싶었기 때문이다. 과정적 측면에서 개인의 인식보다는 그 산출물을 드러내는 데 좀더 초점을 두었다. 후속 연구에서는 각 집단이 느꼈던 인식에 대해 조사함으로써 향후 서로 다른 학문적 집단의 협업에서 필요한 요소를 도출할 수 있을 것이라 기대한다. 둘째, 본 연구에서는 개발된 평가도구의 평가 성능에 대해서는 다루지 않았다. 이는 AI를 활용한 글쓰기 의사소통 역량 평가도구 개발과 해당 과정에서 나타난 두 전문가 간의 상호작용의 중요성을 강조하는 데 더 집중하기 위함이었다. 그러나 본 연구팀은 향후 학생들의 글을 실제로 채점하고 인간 평가자의 평가 성능과 비교함으로써 개발된 평가도구의 성능을 검증하는 추후 연구를 계획하고 있다. 셋째, 현재 이 평가도구는 글쓰기 의사소통 역량만을 평가하도록 설계되어 있다. 교육과정이 역량 중심으로 전환됨에 따라, 다양한 역량 평가에 대한 관심이 증가하고 있다. 이러한 배경을 바탕으로, 향후 연구에서는 다양한 역량을 평가할 수 있는 채점기준을 개발하고, 이를 평가도구에 탑재하는 것을 목표로 하여 평가도구의 확장 가능성을 탐구할 예정이다.

본 연구는 융합연구가 강조되고 앞으로 더 활발해질 것이라고 기대되는 시점에서 융합 연구의 과정을 산출물을 중심으로 드러내는 데 초점을 두었다. 앞으로의 연구를 통해 융합의 과정의 촉진 요소와 저해 요소를 밝힘으로써 융합 연구가 활성화될 수 있는 조건을 탐색할 수 있기를 바란다.

감사의 글

이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020S1A3A2A02095447).

이 연구를 수행하기 위해 많은 전문가들의 협업이 필요하였음. 이를 위해 애써준 내용 전문가 여러분과 평가도구를 빠른 시일 내에 개발하기 위해 애써준 김한길 엔지니어께도 감사를 표함.

참고문헌

[1] S. H. Kim, K. S. Kim, S. Kim, Y. Kim, J. H. Kim, J. B. Kim, H. I. Kim, H. B. Kim, H. C. Kim, H. R. Kim, D. Park, S. J. Park, C. S. Park, W. S. Sohn, S. Song, J. K. Shim, J. H. Lee, J. S. Lee, S. Jun, W. Jun, Y. Jeong, U. Jung, and J. Seo, "Analysis of functions and services for the development of AI education platform," *The Journal of Korean Association of Computer Education*, vol. 24, no. 2, pp. 25-37, March, 2021.

[2] AIEDAP. Monthly Report [Internet]. Available: https://

aiedap.or.kr/

[3] H. Jang and H. J. So, "The analysis of research trends and topics about the educational use of ChatGPT," *Journal of Research in Curriculum & Instruction*, vol. 27, no. 4, pp. 387-401, August, 2023.

[4] AI Times, "AI trends identified at the Edutech Korea Fair," Sep. 22, 2023 [Online]. Available: https://www.aitimes.com/news/articleView.html?idxno=153859

[5] H. Park, S. S. Kim, K. H. Kim, M. Lee, K. K. Kim, and J. Kim, "Substanitializing methods of restricted and extended response essay assessment through enforcing the instruction-assessment alignment," Jincheon: Chungbuk, KICE Report, 2019.

[6] K. Ercikan and D. F. McCaffrey, "Optimizing implementation of artificial intelligence-based automated scoring: an evidence centered design approach for designing assessments for AI-based scoring," *Journal of Educational Measurement,* vol. 59, no. 3, pp. 272-287, September, 2022.

[7] Korea Education Culture News, "Siwon School Lab-Data Bank, providing AI mock test platform 'The Test Glider'," Oct. 18, 2023 [Online]. Available: https://www.kecn.co.kr/news/articleView.html?idxno=1723

[8] S. W. Han and W. J. Kim, "A study on the effects of communicative competence on information literacy of undergraduates," *Journal of the Korean Library and Information Science,* vol. 50, no. 1, pp. 377-394, February, 2016.

[9] C. Lim, "The development and effects of design and implementation strategies for supporting web-based self-regulated learning," *Journal of Educational Technology*, vol. 18, no. 4, pp. 3-23, December, 2002.

[10] B. Kim, Y. Jeon, J. Kim, C. Hong, and T. Kim, "A study of IT convergence desired talent based on computation thinking," in Proceeding of the 2014 Winter Conference of the Korean Association of Computer Education, Seoul, pp. 27-33, 2014.

[11] J. M. Lee, T. Hur, J. B. Lee, G. Hwang, and K. Om, "Analysis of human resources practices and career path movement in the field of convergence technologies," *The Journal of the Korea Contents Association*, vol. 10, no. 5, pp. 446-459, April, 2010.

[12] J. Kim, "BSC (Balanced Scorecard) implementation factors and firm performance: The role of communication, strategic alignment, and cooperation," *Korean Accounting Journal*, vol. 20, no. 1, pp. 233-264, March, 2011.

[13] J. Bae and G. Jeon, "A study on the influencing factors and performance of communication patterns among partners in collaborative R&D projects," in Proceeding of the 1998 Spring Conference of KIEE and KORMS, Pusan, pp. 1-9, 1998.

[14] V. Spandel and R. Culham, "Writing assessment," in *A Handbook for Student Performance Assessment in an Era of Restructuring*. Alexandria, VA: ASCD.

[15] J. Park and Y. Park, "A study on the differences of the anchor papers according the rater consistence," *Korean Writing Association*, vol. 14, pp. 301-338, March, 2012.

[16] I. Park and B. Kim, "Determinants of successful R&D cooperations between SMEs and public research institutes in Korea," *Journal of Korea Technology Innovation Society,* vol. 15, no. 4, pp. 783-814, December, 2012.

[17] B. Min, Y. Oh, S. Lee, S. Ahn, K. Kim, M. Kim, J. Son, J. Lee, and S. Chang, "Development of a writing ability diagnosis system for university first-year students － The Seoul National University essay examination," *The Korean Journal of Literacy Research*, vol. 13, no. 6, pp. 45-76, December, 2022.

[18] Y. M. Park, "The method and procedure of evaluating writing ability," *The Journal of Korean Language and Literature Education*, pp. 1-29.

[19] S. Seo, "Study on setting the components of writing assignment," *Korean Language Education Research*, vol. 33, pp. 449-472, December, 2008.

[20] H. Song, "An analysis of basic grammar evaluations for college students' writing," *The Journal of Korean Language and Literature Education*, vol. 35, pp. 31-57, June, 2015.

박 소 영 (So-Young Park)_정회원

1998년 2월 : 서울대학교 교육학과 졸업
2003년 5월 : University of Wisconsin, Madison 교육행정 박사
2009년 9월 ~ 현재 : 숙명여자대학교 교육학부 교수
〈관심분야〉 교육정책, AI 기반 교육 평가, 교사교육

이 병 윤 (ByungYoon Lee)_정회원

2012년 8월 : University of Minnesota, Twin Cities 심리학과 졸업
2016년 8월 : 서울대학교 교육학과 석사
2023년 2월 : 서울대학교 교육학과 박사
2023년 3월 ~ 현재 : 숙명여자대학교 교육연구소 전임연구원
〈관심분야〉 교육심리, AI 기반 교육 평가

홍 유 정 (Yujung Hong)_정회원

2013년 2월 : 경인교육대학교 사회교육과 졸업
2018년 2월 : 서울대학교 교육학과 석사
2022년 8월 : 서울대학교 교육학과 박사
〈관심분야〉 교육평가, 교육빅데이터 분석, 교육통계