

국립국어원 2023-01-37

발간등록번호

11-1371028-000970-01

2023년 국민의 글쓰기 능력 진단 체계 개발

연구책임자
민 병 곤



국립국어원

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구 용역 계약에 따라 '2023년 국민의 글쓰기 능력 진단 체계 개발'에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2023년 3월 15일 ~ 2023년 12월 15일

2023년 12월 15일

연구책임자: 민병곤(서울대학교)

연구 기관: 서울대학교 국어교육연구소

연구책임자: 민병곤

공동연구원: 남가영, 김선희, 장성민, 이성준, 권은선

보조연구원: 오예림, 이수진, 정은선, 김동섭, 송규민, 민소연, 홍태경

<국문 요약>

‘2023년 국민의 글쓰기 능력 진단체계 개발 연구’ 과제는 국민의 글쓰기 능력 향상을 도모할 수 있는 효과적이고 신뢰할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있다. 세부 과제로 글쓰기 능력 진단 지표 고도화, 신뢰도 높은 채점 전문 인력 양성 프로그램 개발, 고품질의 논증적 글쓰기 원자료 수집과 채점 자료 구축, 자동 채점을 지원하기 위한 인공지능 모델 개발을 실행하였다. 세부 연구 결과는 다음과 같다.

□ 글쓰기 능력 진단 지표 고도화

「인공지능 활용 국어능력 진단체계 개발 기초 연구」(국립국어원·서울대학교 국어교육연구소, 2022)의 결과로서 개발된 논증적 글쓰기 문항의 프로토타입을 수정·보완하여 동형의 논증적 글쓰기 문항 15개를 추가 개발하였다. 개발한 문항은 작문 교육 전문가 및 대학 글쓰기 강좌 교강사의 검토, 대학생 글쓰기 참여자의 예비 평가 등의 과정을 거쳐 성인을 대상으로 한 글쓰기 능력 진단 도구로서의 타당성이 검증되었다.

논증적 글쓰기 채점 기준표(rubric)의 고도화를 위해 채점 기준표 상세화, 표현 범주 기준 정량화를 시행하였다. 상세화를 위해 채점 기준표에서 유의 사항 등 기술 내용을 명료화하고, 예시 글쓰기 자료를 선별하여 제시하였다. 표현 범주 정량화를 위해 2022년 기초 연구의 채점 결과를 분석하여 ‘문장과 어휘’, ‘어문 규범과 관습’ 준수를 양적으로 평가할 수 있는 기술문(description)을 마련하고, 질적 기준을 추가하는 보완 작업을 통해 기술문을 최종 정련하였다. 채점 기준표의 타당성을 검증하기 위해 채점 기준표를 적용한 분석적 채점 결과와 외부 전문가 3인의 총체적 채점 결과를 비교 분석하여 통계적으로 유의한 차이가 없는 것을 확인하였다. 나아가 진단 목적이나 진단 대상의 특성에 따라 내용, 조직, 표현 범주별 또는 하위 준거별로 가중치를 부여하기 위한 방안을 연구하였다. 내용 준거에 가중치를 부여하는 내용 중심 모형, 조직과 표현 준거에 가중치를 두는 형식 중심 모형과 이들을 비교적 유사한 수준으로 고려하는 균형적 모형을 제안하였다.

설명적 글쓰기 문항 프로토타입을 개발하고, 채점 준거와 기준표를 개발하는 기초 연구를 수행하였다. 설명적 글쓰기의 한 유형으로서 비교-대조의 구체적인 대상, 주제, 목적, 분량 등의 기준이 포함된 프로토타입을 개발하였다. 채점 준거는 논증적 글쓰기와 공통 지표로서 내용, 조직, 표현 범주를 설정하고, 내용 범주는 설명문의 특성에 부합하도록 ‘맥락에 대한 이해’, ‘중심 내용’, ‘뒷받침 내용의 적절성’, ‘뒷받침 내용의 충분성’을 하위 준거로 설정하였다.

□ 글쓰기 원자료 수집

글쓰기 참여자를 효율적으로 수집하고 안정적으로 관리하기 위해 대학 협의회를 운영하여 맞춤형 디지털 홍보 자료를 개발하여 지원하고, 글쓰기 자료 수집 플랫폼을 개발·활용하였다. 플랫폼 작성자 페이지는 글 작성 외에 글쓰기 참여자 정보 입력 및 개인정보 활용 동의, 저작권 이용 허락 계약서 작성, 작성 글 다운로드 기능을 갖추었다. 또 관리자 페이지에는 글쓰기 자료 수집 현황 모니터링, 글쓰기 자료 다운로드 기능을 구축하고, 나아가 글쓰기 자료 간 유사도 검사, 글쓰기 자료 자동 정제 등 인공지능 학습 데이터 구축을 위해 필요한 일련의 데이터를 수집·관리하는 기능을 갖추도록 하였다.

1학기과 2학기에 걸쳐 총 9,439건의 글쓰기 원자료를 수집하였다. 글쓰기 능력 진단을 위한 글 내용 및 형식 요건을 충족하지 않았거나, 외국인 학생이 작성한 글쓰기 원자료를 제외하는 검수를 진행하였다. 최종적으로 글쓰기 자료 요건을 충족하는 9,371건의 유효 글쓰기 자료를 구축하여, 목표로 한 9천여 건을 충족하였다.

□ 글쓰기 채점 전문 인력 양성

글쓰기 채점 전문 인력을 양성하기 위해 실무 중심 논증적 글쓰기 채점 교육 프로그램을 개발·운영하였다. 먼저 쓰기 채점에 필요한 기초적인 지식과 원리를 체득하는 데 목표를 둔 온라인 교육과정(12차시)과 채점 워크숍 형태로 진행한 오프라인 교육과정(8차시)으로 구성된 블렌디드 교육과정을 설계하였다. 프로그램 개발 단계에서는 교육 플랫폼 구성, 교수학습 자료 개발, 학습 진단을 위한 과제 개발이 이루어졌다. 온라인 교육을 운영할 플랫폼으로서 네이버 웨일클래스를 채택하였다. 강의 자료, 퀴즈 문항, 온-오프라인 교육 간 연계를 위한 사전 채점 과제 9편을 포함한 온라인과 오프라인 교육용 자료는 자료집으로 제작하여 배부하였다.

프로그램 운영을 위해 대학의 글쓰기 교강사를 대상으로 전국의 약 22개 대학에서 총 109명의 교육 참여자를 모집하였다. 온라인 교육과정은 6월-7월에 실행하였으며, 오프라인 교육은 7월 중 4개 권역에서 각각 진행되었다. 이수 사정을 실시하여 최종적으로 96명이 '2023년 채점 전문 인력 양성 프로그램 1기'를 공식 수료한 것으로 확정하였다. 만족도 설문 결과 만족도가 90%를 상회하는 것으로 나타났으며 채점자 교육에 대한 필요성이 사전 설문 대비 높아진 것을 확인할 수 있었다. 프로그램 운영 결과를 바탕으로 채점자 매뉴얼을 개발하였다.

□ 글쓰기 자료 채점 실행

체계적이고 일관된 채점 실행을 위한 채점 모형과 자동 채점 인공지능 학습 데이터로 활용하기 위한 채점 자료 선정 모형을 제시하였다. 채점 모형은 글쓰기 자료 목록 확정, 채점 방법 결정, 채점 할당 설계, 채점 자료의 관리로 구성되었다.

다음으로 채점자들에게 일관된 채점 환경을 제공할 수 있는 채점 플랫폼을 개발하였다. 채점자 페이지에는 기본 정보 입력 화면, 채점 자료 목록 확인 화면, 채점 실행 화면으로 구성하였다. 채점 관리자 페이지는 채점자별 정보 확인 화면, 채점 대상 글쓰기 자료 할당을 위한 화면, 채점 진행 상황 모니터링 화면으로 구성하였다.

본 채점은 채점의 효율적인 관리를 위해 1-5차로 나누어 진행이 되었으며, 채점자 간 편차가 큰 사례에 대한 재채점까지 총 6회에 걸쳐 시행하였다. 1-4차 채점에서는 상반기에 수집한 글쓰기 자료 4,002편을 대상으로 가교 채점과 짝 채점을 하는 글쓰기 자료를 편성하고, 개인별로 총 100-108편을 채점하도록 하였다. 하반기에 진행한 5차 채점에서는 개인별로 42-56편씩, 총 998편을 채점하였다. 짝 채점 사례 중 두 채점자 간 평균 표준편차가 0.9 이상인 92편의 글에 대해서는 재채점이 이루어졌다.

채점자들은 채점 수행 직전 온라인 채점자 워크숍에 참석하여 사전 채점 연습, 실시간 채점 연습, 소집단별 토의 활동에 참여하여 채점 기준표 사용과 문항 특성, 응답 특성에 관한 인식을 공유하였다. 이후 채점 일관성 유지를 위하여 가교 채점 결과에 대한 개인별 채점 피드백을 제공하였고, 채점자들은 제공받은 피드백 내용을 숙지한 후 2인이 하나의 글쓰기 자료를 채점하는 짝 채점에 참여하였다.

□ 글쓰기 자료 채점 결과 분석 및 선정

글쓰기 능력 평가 문항 및 채점 도구, 채점의 특징을 파악하고, 나아가 채점 자료에 대한 대표 채점 결과를 선정하기 위해 5,000편의 글쓰기 자료에 대한 채점 결과를 분석하였다. 먼저 고전검사이론에 따라 기술통계, 신뢰도, 상관 분석, 요인 분석을 진행하였다. 기술통계 분석 결과 문항 5(조력 존엄사 허용)의 곤란도가 가장 높고, 문항 2(혐오 표현 규제)가 가장 낮았으며, 준거 중에는 내용 5(다른 입장 고려)가 가장 높고, 표현 2(어문 규범과 관습)의 곤란도가 가장 낮았다. 가교 채점에 대한 채점자 신뢰도 분석 결과 전체 그룹에서 급내 상관이 0.9 이상으로 매우 높은 수준이었다. 총점과의 상관에서는 내용 3(이유/근거의 적절성)의 상관이 0.825로 가장 높았으며, 내용 5(다른 입장 고려)의 상관이 0.459로 가장 낮았다. 각 문항별로 총점-준거의 상관을 비교하였을 때는 전반적으로 평균 0.7 정도의 양호한 상관을 갖고 있는 것으로 나타났다. 다음으로 9개의 채점 준거

를 대상으로 확인적 요인분석을 실시한 결과 9개 준거 모두 모형을 구성하고 있는 것으로 확인이 되었다.

문항반응이론에 따라 다국면 라쉬 모형을 기반으로 채점자의 채점 특성을 분석하였다. 채점자들은 채점 척도를 안정적으로 사용하고 있었으나, 일부는 높은 수준의 관대성을 보였고, 채점 척도를 부적합하게 사용한 사례로 발견하였다. 다음으로 문항별 준거 특성에 대한 분석 결과에서는 전체 문항에서 내용 5(다른 입장 고려)의 곤란도가 가장 높은 것으로 나타났으며, 변별도가 가장 낮은 준거는 공통적으로 내용 5(다른 입장 고려)였다.

자동 채점 인공지능 학습용 데이터를 구축하기 위해 복수의 채점 결과 중 균형적이고 안정적인 결과를 나타낸 채점자의 채점 결과를 대푯값으로 선정하였다. 다만 짝 채점 사례 중 두 채점자 간 의견 불일치 정도가 큰 92편의 경우 우수한 채점자인 제3자의 재채점 결과를 대푯값으로 채택하였다.

□ 글쓰기 자료의 정제 및 주석화

자동 채점 인공지능 개발용 학습 데이터 구축을 위해 비유효 데이터를 걸러내고, 개인 정보를 비식별화하는 등 글쓰기 자료를 정제하는 단계를 진행하였다. 데이터 정제는 플랫폼 내 데이터베이스화 과정에서 단락 간격, 글자 간 간격 등 ‘형식 요소를 자동 편집’하는 1단계, 심각한 논제 이탈, 심각한 표절, 과제 장르 이탈, 글자 수 기준 미충족 등의 ‘부적합 데이터를 선별’하는 2단계, 제목이나 주제문, 목차와 같은 구성 표지, 인사말과 자기소개와 같은 글쓰기 상황 표현, 글 전개와 무관한 내용과 기호 등 ‘불필요 항목을 삭제’하는 3단계, 개인 식별 정보, 혐오와 차별 조장 내용 등 ‘개인정보나 비윤리적 내용을 삭제’하는 4단계로 진행되었다.

그다음으로 정제된 글쓰기 자료에 파일 메타 정보, 글쓰기 자료 정보, 채점 결과 정보 등 주요 정보를 주석화하고 최종 검수하는 단계를 진행하였다. 파일 메타 정보에는 생성자, 배포자, 말뭉치 구축 연도, 글 유형, 길이, 분석 층위, 샘플링 방식이 포함되었다. 글쓰기 자료 정보에는 직업, 성별 등 글쓰기 참여자 정보, 수집 방식, 글 유형, 수집 일시, 글쓰기 과제 발문, 문서 통계가 포함되었다. 채점 정보는 채점자 정보와 채점자별/채점 영역별 세부 점수가 포함되었다.

□ 인공지능 기반 채점 지원 도구 개발

채점 지표별 점수 예측 인공지능 프로그램을 개발하기 위해 베이스라인 모델, 분류 및 점수 맞추기 멀티태스크 모델, 채점 기준 기반 분류 및 맞추기 멀티태스크 모델을 설계

하고 모델 간 점수 예측 성능을 비교하였다. 그 결과 채점 기준까지 고려한 멀티태스크 모델이 48.1%의 정확도로 가장 높은 성능을 보였다. 나아가 딥러닝 모델의 크기로 생기는 성능 향상 정도를 비교하였는데, 큰 규모 클루-로버타(KLUE-RoBERTa Large) 사전학습 언어모델을 활용한 분류 및 점수 맞히기 멀티태스크 모델에서 가장 높은 정확도(43.8%)를 보였다. 다음으로 채점 점수의 신뢰도 예측 AI 프로그램을 개발하고, 온도(temperature) 보정 기법을 활용하여 신뢰도 구간별 정확도를 전반적으로 개선할 수 있음을 확인하였다.

채점 지표별 점수 예측 AI를 채점자 지원 도구로 활용하기 위하여, 상위 2개의 점수를 범위(boundary) 형태로 제공하는 방안을 제안하였다. 상위 1개의 점수를 예측하는 정확도는 모든 채점 지표에 대해서 50% 미만이었지만, 상위 2개의 점수를 제시할 경우, 개발된 AI의 채점 지표별 점수 범위 예측 성능은 30% 이상 향상되었다. 최종적으로 자동 점수 예측 AI 및 신뢰도 예측 AI 결과를 외부 채점 시스템과 연동하기 위하여, API 서비스를 개발하여 데이터 전달 예제를 제시하였다.

□ 후속과제 제언

첫째, 글쓰기 능력 진단 도구 고도화 차원에서 설명적 글쓰기 진단 도구 타당화, 정교한 문항 분석을 위한 글쓰기 능력 평가 시행 고도화, 논증적 글쓰기의 다양한 평가 상황과 목적에서 채점 기준표를 활용할 수 있는 방안 마련을 제안하였다. 둘째, 통제된 조건에서 다양한 수준에서 글쓰기 자료를 수집하고, 문항별로 균형 있는 글쓰기 자료를 수집하기 위해 관리 체계를 강화하기 위해 방안을 제시하였다. 셋째, 채점자 교육과정 및 운영 개선 방안으로 온라인 교육과 오프라인 교육 간 교육 내용을 긴밀히 연계하고, 온-오프라인 학습 참여도를 제고하기 위한 방안을 제시하고, 온라인 교육 플랫폼의 지속성과 안정화를 위한 정책 마련의 필요성을 논하였다. 넷째, 채점 플랫폼 기능 개선을 통해 채점 효율성과 안정성을 제고하고, 채점 결과 선정 방식 연구와 자동 채점 인공지능 개발 연구를 연계하여 인공지능의 성능을 제고할 수 있는 채점 결과 선정에 대한 연구가 필요함을 논하였다. 마지막으로 최근 인공지능 연구가 거대 언어모델을 중심으로 이뤄지는 동향을 고려하여 오픈소스 거대 언어모델 기반의 자동 채점 인공지능 연구를 제안하였다.

주요어: 국민의 글쓰기 능력 진단 체계, 글쓰기 원자료 수집, 쓰기 채점 인력 양성, 쓰기 평가 데이터 세트, 인공지능 활용 자동 평가

<Abstract>

Development of National Evaluation System for Writing aims to develop an effective and reliable writing proficiency evaluation system that can help improve one's writing skills. Specific tasks for the research include advancing the writing proficiency evaluation system, developing writing rater training program to train reliable evaluation experts, collecting high-quality argumentative raw writing data and scoring materials and developing artificial intelligence based models to support automatic evaluation. The results of the study are as follows:

□ Advancing Writing Proficiency Evaluation System

This study developed additional 15 questions of argumentative writing by revising and supplementing the prototype of argumentative writing as the result of basic research in 2022. The developed questions were reviewed by the experts of writing education professors and lecturers of college writing courses and preliminary evaluation of college students were conducted to ensure the validity as a tool for diagnosing writing skills of adults.

In order to advance the rubric of argumentative writing the study refined the scoring criteria and quantified the categories of expression. The guidelines were clarified and example of writing data were presented to develop the detailed scoring criteria. For the quantification of expression category, the description was developed to help evaluate the compliance of sentences and vocabularies and language norms and custom quantitatively. Additionally the description was refined by the additional qualitative criteria.

To verify the validity of the rubric, the analytic evaluation done by three external experts and overall grading result using the rubric were compared and could find that there were no statistically significant differences between them.

Furthermore, the ways of weighting of the content, organization, and expression, categories or sub-categories during evaluation was studied depending on the purpose of evaluation or the characteristics of the evaluation target.

Three types of models were derived including a content-centric model that weighs the content criteria a format-centric model that weighs the organization and expression criteria and a balanced model that considers both criteria

relatively similar.

The basic research developing the prototype of expository writing scoring criteria, and rubric were conducted. A prototype including the criteria for the specific object of compare and contrast, topic, purpose, and length was developed as a type of expository writing. The criteria for expository writing was set similarly with that of argumentative writing which are content, organization, and expression. The content category is subcategorized into 'context understanding', 'main idea', 'appropriateness of supporting content,' and 'sufficiency of supporting content' to match the characteristics of expository writing.

□ Collecting Raw Writing Data

In order to recruit the participants effectively and manage them, a council of university was organized to operate the customized digital promotional materials and develop and utilize a writing data collection platform.

The platform page for the writer includes the functions for collecting information of the writer's basic information, consent form for agreeing to utilize personal information, creating a copyright transfer, and function of downloading the writing in addition to the writing function. The administrator's page includes the function to collect and manage the data necessary for building AI learning data including monitoring function of collection of data, data downloading function, function to check the similarity between data and function of refining data automatically.

Total of 9,439 raw data was collected through the first and second semester of the year. Data that did not meet the content or the formatting requirements of the evaluation were excluded and were not analyzed. The goal of collecting about 9,000 data was achieved since 9,371 valid writing data met the requirements and were analyzed.

□ Developing and Training Experts for Writing Evaluation

A practical writing rater training program was developed and operated to train and develop writing evaluation experts.

First, a blended curriculum consisting of an online course (12 sessions) to teach basic knowledge and principles for writing evaluation and an off-line course (8 sessions) in the form of a grading workshop was designed. On the stage of

developing the program, the development of online educational platform, teaching and learning materials and tasks for evaluation were developed. Naver Wales was selected for the online platform for the online education. Materials for online and off-line including course materials, questions for the quiz, nine pre-assignments for linking online and offline training were distributed in a form of a textbook.

To operate the program, 109 professors and lecturers for writing courses were recruited from about 22 universities around the country. The online course was held from June to July and the off-line course was held in July in four different regions. After the completion of the program, 96 people officially completed the first phase of the 2023 Grading Professional Training Program. The results of the satisfaction survey showed that the satisfaction rate was over 90%, and the need for rating training had increased compared to the pre-survey. The manual for raters was developed based on the results of the program.

□ Grading the Writing Data

A grading model for systematic and consistent grading and another model for material selection model to be used as training data for AI-based automatic evaluation were presented. The grading model consists of determining the list of writing data, grading method designing the grading assignment and grading managing the grading materials.

Furthermore a new grading platform that can offer a consistent grading environment was developed. The rater's page includes a page for entering basic information, a list of grading materials, and a grading execution page. The administrative page includes a page for checking information of each rater, a page for assigning writing materials, and a page for monitoring grading progress.

In order to manage the grading process efficiently, the grading was divided into 5 rounds and in total of 6 rounds if re-grading was needed for large deviations between raters. In the first to 4th rounds of grading, 4,002 writing materials collected in the first half of the year were organized into writing data for cross-checking and paired grading, and each individual was asked to grade a total of 108 to 100 writing data. In the fifth round of grading that happened in the second half of the year, each individual graded 42 to 56 data, which were 998 data in total. The data was re-graded if the average standard deviation between the two raters was greater than 0.9.

The rater participated in the online workshop and practiced the preliminary evaluation and live scoring program, and participated in small group discussions and shared their opinions regarding the use of the scoring criteria, characteristics of questions and responses prior to the evaluation.

To ensure consistency of evaluation, individual raters received feedback for the cross-checking result and participated in pair scoring in which two raters score the single data after reviewing the feedback.

□ Analyzing and Selecting Grading Results for Writing Data

5,000 grading results for writing data were analyzed in order to identify the characteristics of the writing skill assessment and scoring tools, and to select representative grading results.

First of all, descriptive statistics, reliability, correlation, and factor analysis were conducted according to classical test theory. The descriptive statistics analysis showed that content area 5 (Allowing assistant for death) had the highest difficulty and content area 2 (Regulating hate speech) had the lowest difficulty, and among the standards, content area 5 (level of consideration and diversity of other positions) had the highest difficulty while content area 2 (Language norms and custom) had the lowest difficulty.

The rater reliability analysis for the cross-checking grading showed very high intraclass correlations of over 0.9 across all groups. In terms of correlation with the total score, content area 3 (relevance of connection of reason/evidence) had the highest correlation at 0.825, while content area 5 (level of consideration and diversity of other positions) had the lowest correlation at 0.459. When compared the correlation between the total score and the criteria for each question, the correlation was relatively good, with an average of 0.7. Afterwards, a confirmatory factor analysis on the nine scoring criteria was run and was found that all nine criteria formed a certain model.

Based on item response theory, the scoring characteristics of the raters based on a multifaceted Rasch model was analyzed.

The raters were mostly using the scale reliably but some showed a high degree of generosity and used the scale inappropriately. Then, the scoring characteristics of each question were analyzed and were found that content area 5 (level of consideration and diversity of other positions) was the most

difficult across all questions, and content area 5 (level of consideration and diversity of other positions) was also the least discriminating factor.

To build the data for training the automatic scoring AI, two raters who showed balanced and stable results among multiple scores were selected as the representative value. However, if the result for the paired scoring cases where the degree of disagreement between the two raters was too excessive, the re-scoring results of a third rater who was also an excellent scorer was selected as a representative value for 92 writing data.

□ Refining and Annotating Writing Data

In order to build the training data for the development of automatic grading AI, filtering out invalid data and de-identifying personal information steps were taken. The filtering includes the following four stages. The first stage was editing the formatting elements automatically. The second stage was to select the invalid data which do not meet the requirements such as writing data that are off topic, or data that includes content with serious plagiarism, data that did not meet the assignment genre or number of words requirements. The third stage was deleting the unnecessary information or symbols that are not related to the writing assignment such as title, topic sentences and table of contents or expressions of writing situations such as greetings and personal introduction. The fourth stage was deleting the personal or unethical information such as personal identification information hate and discriminate words.

The next step was to annotating and finalizing the refined writing data with key information including file meta-information, writing(corpus) information, and grading results. File meta-information included the creator, distributor, the year of corpus construction, article type, length, stratum of analysis, and sampling method. Article information included writing participant information such as occupation and gender, collection method, article type, date of collection, writing assignment, and document statistics. Grading information included rater's information and specific scores per rater and per scoring area.

□ Development of AI-based Evaluation System

To develop an AI-based Evaluation program for score prediction by grading metrics, the following models had been designed and the score prediction performance among the following models were compared; a baseline model, a multi-task model with classification and score matching, and a multi-task model with classification and score matching considering scoring criteria.

As a result, the matching multi-task model considering the scoring criteria performed the best with an accuracy of 48.1%. Furthermore, the performance improvement as the size of the deep learning model increases was compared and the highest accuracy (43.8%) was achieved by the classification and scoring multi-task model using the KLUE-RoBERTa Large pre-trained language model. Furthermore, an AI program for predicting the confidence of grading scores was developed and was found that the overall accuracy of each confidence interval can be improved by utilizing temperature correction techniques.

In order to utilize the score prediction AI as a grading support tool, top two scores were provided the form of boundaries. The accuracy of predicting the top one score was less than 50% for all scoring metrics, but when the top two scores were provided the performance of the developed AI in predicting the score range for each scoring metric improved by more than 30%. Finally, to link the results of the automatic score prediction AI and reliability prediction AI with an external scoring system, we developed an API service and presented an example of data delivery.

□ Implications

First, in terms of improving writing ability diagnostic tools, a validate descriptive writing diagnostic tools was proposed and in order to advance the administration of writing ability assessments for sophisticated questions analysis was developed, and scoring rubrics to be used in various evaluation situations and purposes of argumentative writing were presented. Second, methods were proposed to strengthen the management system to collect writing data at various levels under controlled conditions and to collect balanced writing data per question.

Third, as a way to improve the training and program for the raters, the

contents between online and offline training program were closely related and the methods of improving online and offline learning participation were proposed. Also, the necessity of rules for the keeping the continuity and stabilization of the online training platform was discussed. Fourth, the necessity of efficient and stable evaluation system by improving the function of the grading platform was proposed and the need for research on the selection of grading results that can improve the performance of AI by linking research on grading result selection methods with research on the development of AI-based automatic evaluation. Finally, considering the recent trend of AI research centered on large-language models, a study on AI-based automatic evaluation on open source with large language models was proposed.

Key-words: National Evaluation System for Writing, Collecting Raw Writing Data, Writing Rater Training, Writing Evaluation Dataset, Artificial Intelligence -Based Automatic Evaluation

연구 사업 자문위원

가은아(공주대)	배영환(제주대)
간호윤(인하대)	백민정(충남대)
강미영(전남대)	백수진(경상국립대)
강수경(부산대)	손대익(충북대)
강지연(서원대)	송미지(서울대)
김기영(충남대)	송원용(인천대)
김다연(성포고)	신용권(인천대)
김소산(당곡고)	안서현(서울대)
김승현(홍익대)	양수연(서울대)
김영건(수원대)	양영희(전남대)
김정은(서울대)	윤석민(전북대)
김정인(충남대)	윤준섭(충남대)
김정자(경인교대)	이상재(서울대)
박고운(부산대)	이선영(경인교대)
박광규(경상국립대)	이윤빈(덕성여대)
박민호(한국교육과정평가원)	장시광(경상국립대)
박보연(인천대)	장정민(제주대)
박성석(춘천교대)	장지혜(춘천교대)
박소영(이화여대)	장충덕(충북대)
박영민(한국교원대)	정민주(한남대)
박재희(충남대)	조유영(제주대)
박종임(한국교육과정평가원)	차성현(전남대)
박종훈(부산교대)	차윤정(부산대)
박지은(전북대)	최숙기(한국교원대)
박현이(배재대)	최영인(서울교대)
박혜영(한국교육과정평가원)	편지윤(청주교대)

차 례

I. 서론

1. 연구의 필요성 및 목적 3
2. 연구의 내용 및 범위 8
3. 연구의 방법 12

II. 글쓰기 진단 지표 고도화

1. 논증적 글쓰기 동형 문항 개발 17
2. 채점 도구의 정교화 33
3. 준거별 가중치 연구 62
4. 설명적 글쓰기 문항 및 준거 관련 연구 64

III. 글쓰기 자료 수집

1. 글쓰기 자료 수집을 위한 대학 협의체 운영 87
2. 글쓰기 자료 수집 계획 수립 94
3. 글쓰기 자료 수집 플랫폼 개발 및 운영 101
4. 글쓰기 자료 수집을 위한 홍보 자료 개발 112
5. 글쓰기 자료 수집의 결과 118

차 례

IV. 글쓰기 채점 전문 인력 양성

1. 채점자 교육 프로그램 설계 126
2. 채점자 교육 프로그램 개발 140
3. 채점자 교육 프로그램 운영 149
4. 채점자 교육 프로그램 평가 158

V. 글쓰기 자료 채점 실행

1. 채점 모형 설계 181
2. 채점 실행 플랫폼 개발 및 운영 189
3. 본 채점 진행 및 관리 201

VI. 채점 결과 분석 및 선정

1. 채점 결과 분석 221
2. 채점 결과 선정 239

VII. 인공지능 기반 채점 지원 도구 개발

1. 인공지능 학습 데이터 구축 247
2. 자동 채점 프로토타입 개발 279

차 례

VIII. 결론

- 1. 요약 293
- 2. 후속 과제 제언 300

참고 문헌 307

부록

[부록 1] 채점 매뉴얼 개요 311

[부록 2] 저작권 이용 허락 동의서 양식 313

[별책 1] 국민의 논증적 글쓰기 채점 매뉴얼

[별책 2] 채점 전문 인력 양성 프로그램 운영 개요서

표 차례

<표 I -1> 연구의 범위	10
<표 I -2> 연구의 방법	12
<표 II -1> 논증적 글쓰기 문항의 프로토타입 수정·보완	17
<표 II -2> 논증적 글쓰기 동형 문항 개발 절차	19
<표 II -3> 문항 개발 지침	20
<표 II -4> 2023년도 개발 논증적 글쓰기 문항 프로토타입	20
<표 II -5> 문항 프로토타입에 대한 전문가 협의회 점검 항목	25
<표 II -6> 2023년도 개발 논증적 글쓰기 수정 문항	27
<표 II -7> 2023학년도 글쓰기 원자료 수집용 논증적 글쓰기 문항 ..	31
<표 II -8> 채점 기준표 상세화 절차	33
<표 II -9> 논증적 글쓰기 채점 기준표(내용, 조직) 유의 사항 수정 전후 대조표	35
<표 II -10> 2022년 개발 표현 범주 채점 준거의 점수별 채점 기준 ..	37
<표 II -11> 2023년 개발 표현 범주 채점 준거 1차 수정안의 점수별 채점 기준	37
<표 II -12> 표현 범주 채점 준거 유의 사항의 2022년 개발안과 2023년 1차 수정안의 대조표	38
<표 II -13> 정량화 이후 보완한 표현 범주 채점 기준표	41
<표 II -14> 문항 관련 전문가 검토 의견 및 연구진 의견	43
<표 II -15> 채점 기준표 전반 관련 전문가 검토 의견 및 연구진 의견 ..	43
<표 II -16> 채점 준거의 위계 관련 전문가 검토 의견 및 연구진 의견	45

표 차례

<표 II-17> 하위 준거 관련 전문가 검토 의견 및 연구진 의견	45
<표 II-18> 예시문 관련 전문가 검토 의견 및 연구진 의견	47
<표 II-19> 채점자 관련 전문가 검토 의견 및 연구진 의견	47
<표 II-20> 정책 관련 전문가 검토 의견 및 연구진 의견	47
<표 II-21> 인공지능 학습 관련 전문가 검토 의견 및 연구진 의견 ..	48
<표 II-22> 논증적 글쓰기 채점 기준표(최종)	49
<표 II-23> 채점 기준표 타당성 확보를 위한 외부 전문가 재채점 절차	51
<표 II-24> 외부 전문가 3인의 총체적 채점 및 하위 범주별 채점 평균과 표준편차	52
<표 II-25> 구간별 외부 전문가 3인의 총체적 채점 및 채점 위원 2인 채점 평균과 표준편차	53
<표 II-26> 내용, 조직, 표현의 외부 전문가 3인 채점 평균값(ABC)과 채점 위원 2인 채점 평균값(D)	54
<표 II-27> 내용, 조직, 표현의 외부 전문가 3인 평균값(ABC)과 채점 위원 2인 평균값(D)의 차이 검정	54
<표 II-28> 구간별 총체적 채점(A, B, C) 및 9개 준거 평균(D)의 일치율 및 상관 비교	55
<표 II-29> 2023년 가교 채점 문항 유형에 따른 준거별 총점과의 상관 계수 예시	63
<표 II-30> 설명적 글쓰기 문항 및 채점 준거 개발 절차	64
<표 II-31> 설명적 글쓰기 관련 국내외 시험 및 평가 특성	67

표 차례

<표 II-32> 설명적 글쓰기 문항 소재 탐색의 예	71
<표 II-33> 설명적 글쓰기 문항 프로토타입의 형식	71
<표 II-34> 설명적 글쓰기 채점 준거	75
<표 II-35> 설명적 글쓰기 프로토타입 문항 및 채점 준거에 대한 전문가 검토 결과	77
<표 III-1> 실무자 회의 운영 개요	89
<표 III-2> 하계 대학 협의회 운영 개요	92
<표 III-3> 글쓰기 자료 수집 이원화 모델	95
<표 III-4> 글쓰기 자료 수집 모델별 참여 대학(1학기)	95
<표 III-5> 글쓰기 자료 수집 모델별 참여 대학(2학기)	97
<표 III-6> 2학기 글쓰기 자료 수집 계획	100
<표 III-7> 글쓰기 자료 수집 플랫폼의 기능	102
<표 III-8> 대학별 글쓰기 원자료 수집 결과	118
<표 III-9> 문항별 글쓰기 원자료 수집 결과	118
<표 III-10> 2학기 글쓰기 자료 문항별 수집 현황	120
<표 IV-1> 글쓰기 채점자 교육과정 개요	129
<표 IV-2> 채점자 교육용 사전·사후 설문 도구	130
<표 IV-3> 온라인 채점자 교육과정 개요	134
<표 IV-4> 쓰기 평가 이론 강의 요목	135
<표 IV-5> 쓰기 채점 이론 강의 요목	136
<표 IV-6> 쓰기 평가 준거별 채점 방법 강의 요목	136
<표 IV-7> 오프라인 채점자 교육과정 개요	138

표 차례

<표 IV-8> 채점 도구 이해 실습 강의 요목	138
<표 IV-9> 채점 신뢰도 이해 실습 강의 요목	139
<표 IV-10> 채점 실습 교육의 환류 강의 요목	139
<표 IV-11> 온라인 강의 개발에 필요한 자료	141
<표 IV-12> 강의 영상 촬영 일정	143
<표 IV-13> 차시별 퀴즈 예시	145
<표 IV-14> 온라인 교육용 자료집 목차	146
<표 IV-15> 채점자 교육과정 이수 조건	149
<표 IV-16> 권역별 오프라인 교육 일정 및 장소	150
<표 IV-17> 채점 전문 인력 양성 프로그램 최종 모집 현황	152
<표 IV-18> 권역별 오프라인 교육 개요	154
<표 IV-19> 오프라인 교육 세부 일정	154
<표 IV-20> 대학별 프로그램 이수 인원	158
<표 V-1> 채점 실행 플랫폼의 기능	190
<표 V-2> 1-4차 채점 그룹별 배정 인원(명)	201
<표 V-3> 본 채점 대상 글쓰기 자료 수	202
<표 V-4> 1-4그룹 글쓰기 자료 할당 현황	203
<표 V-5> 가교 설계 예시	203
<표 V-6> 1-4차 채점 그룹별 일정	204
<표 V-7> 1-4그룹 채점자 워크숍 세부 일정	205
<표 V-8> 기술 통계 채점 피드백	208
<표 V-9> 채점 특성에 따른 채점 피드백	209

표 차례

<표 V-10> 짝 채점 관리 일정	212
<표 V-11> 1-5그룹 글쓰기 자료 할당 현황	214
<표 V-12> 5그룹 채점자 워크숍 세부 일정	215
<표 V-13> 짝 채점 글쓰기 자료별 평균 표준편차	215
<표 V-14> 재채점 대상 글쓰기 자료 수	216
<표 V-15> 채점자별 재채점 편수	217
<표 VI-1> 문항별 채점 대상 글쓰기 자료	221
<표 VI-2> 기술 통계 분석 결과	222
<표 VI-3> 총점 기준 기술 통계 결과	223
<표 VI-4> 채점 그룹별 급내 상관 평균	227
<표 VI-5> 문항별 채점자 간 급내 상관 평균	227
<표 VI-6> 전체 채점 결과의 준거 간 상관 계수	228
<표 VI-7> 문항별 총점-준거 상관 계수	229
<표 VI-8> 문항 1의 준거 간 상관 계수	230
<표 VI-9> 문항 2의 준거 간 상관 계수	230
<표 VI-10> 문항 3의 준거 간 상관 계수	231
<표 VI-11> 문항 4의 준거 간 상관 계수	231
<표 VI-12> 문항 5의 준거 간 상관 계수	232
<표 VI-13> 문항 6의 준거 간 상관 계수	232
<표 VI-14> 채점 준거 모델 분석 결과	234
<표 VI-15> 채점 요인 간 상관 분석 결과	234
<표 VI-16> 채점 요인 모형 적합도 분석 결과	235

표 차례

<표 VI-17> 그룹별 채점 특성에 대한 다국면 라쉬 모형 분석 결과	236
<표 VI-18> 문항별 채점자 효과 분석 결과	236
<표 VI-19> 문항별 곤란도 분석 결과	237
<표 VI-20> 문항별 변별도 분석 결과	238
<표 VI-21> 채점 자료 선정을 위한 채점 특성 활용 예시	240
<표 VI-22> 재채점 자료-기존 채점 자료 특성	242
<표 VI-23> 문항별 재채점 자료 특성	243
<표 VII-1> 글쓰기 자료 데이터 세트 제이슨 파일 구조	262
<표 VII-2> 글쓰기 채점 자료 데이터 세트 제이슨 파일 구조	263
<표 VII-3> 최종 데이터 세트 문항별 건수	278
<표 VII-4> 자동 채점 인공지능 프로그램 유형별 입력 및 출력	282
<표 VII-5> 자동 채점 인공지능 프로그램 입력 예시	282
<표 VII-6> 인공지능 점수 예측 방법 간 비교 실험 결과	284
<표 VII-7> 사전학습 언어모델 간 비교 실험 결과	285
<표 VII-8> 온도 보정 기반 신뢰도 예측 개선 결과	286
<표 VII-9> 온도 보정 기반 신뢰도 개선 예시	286
<표 VII-10> 채점 준거별 점수 범위 예측 성능	287
<표 VII-11> 자동 채점 프로토타입 에이피아이(API) 서비스 예시 ..	289

그림 차례

[그림 I -1] 연구의 하위 목표	7
[그림 II -1] 2022년 논증적 글쓰기 자료 중 내용 및 조직 범주 예시 답안 선별	34
[그림 II -2] 5개 구간별 총체적 채점(A-B-C) 및 9개 준거 평균(D) ..	53
[그림 II -3] 준거 수(비율) 조정을 통한 가중치 적용 예시	62
[그림 III-1] 글쓰기 자료 수집 협약 대학과의 협력 체계	88
[그림 III-2] 글쓰기 자료 수집을 위한 협조 사항 안내	93
[그림 III-3] 글쓰기 자료 수집 플랫폼 기능 구조도3	102
[그림 III-4] 글쓰기 자료 수집 플랫폼의 개발 과정	103
[그림 III-5] 글쓰기 자료 수집 플랫폼-프로필 정보 입력 화면	104
[그림 III-6] 글쓰기 자료 수집 플랫폼-동의서 작성 화면	105
[그림 III-7] 글쓰기 자료 수집 플랫폼-글쓰기 수행 화면	106
[그림 III-8] 글쓰기 자료 수집 플랫폼-글쓰기 완료 화면	107
[그림 III-9] 글쓰기 자료 수집 관리 플랫폼-전체 현황 요약 화면 ..	108
[그림 III-10] 글쓰기 자료 수집 관리 플랫폼-개별 제출 현황 화면 ..	109
[그림 III-11] 글쓰기 자료 수집 관리 플랫폼-제출 자료 다운로드 화면	110
[그림 III-12] 글쓰기 능력 진단 체계 개발 홍보 동영상	113
[그림 III-13] 홍보 포스터	114
[그림 III-14] 1학기 홍보 리플릿	115
[그림 III-15] 2학기 홍보 리플릿	116
[그림 III-16] 플랫폼 이용 문서 매뉴얼	117

그림 차례

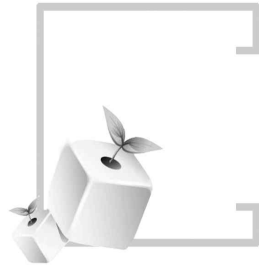
[그림 III-17] 플랫폼 이용 영상 매뉴얼 화면	117
[그림 IV-1] 웨일클래스 진입 화면 예시	140
[그림 IV-2] 강의 자료 템플릿(PPT) 예시	142
[그림 IV-3] 강의 자료 최종본 예시	142
[그림 IV-4] 강의 영상 최종본 예시 화면	144
[그림 IV-5] 강의 자료 탑재 화면	146
[그림 IV-6] 오프라인 강의 자료 예시	147
[그림 IV-7] 오프라인 교육용 자료집 표지 및 목차	148
[그림 IV-8] 채점 전문 인력 양성 프로그램 참여자 모집 문건	151
[그림 IV-9] 온라인 학습 현황 관리 화면	153
[그림 IV-10] 패들렛을 통한 소그룹 활동 결과 공유 화면 예시	155
[그림 IV-11] 채점 활동 1, 2의 신뢰도 분석 결과 제시 내용	156
[그림 IV-12] 1-4권역 오프라인 교육 현장 사진	156
[그림 IV-13] 2023년 채점 전문 인력 양성 프로그램 수료증 양식	159
[그림 V-1] 채점 실행 플랫폼의 기능 구조도	190
[그림 V-2] 채점 실행 플랫폼 개발 과정	191
[그림 V-3] 채점 수행 플랫폼 프로필 입력 화면	192
[그림 V-4] 채점 수행 플랫폼 동의서 화면	193
[그림 V-5] 채점 수행 플랫폼 글쓰기 자료 목록 화면	193
[그림 V-6] 채점 플랫폼 채점 수행 화면1	194
[그림 V-7] 채점 플랫폼 채점 수행 화면2	195
[그림 V-8] 채점 플랫폼 글쓰기 자료 다운로드 화면	195

그림 차례

[그림 V-9] 채점 관리 플랫폼 채점자 목록 화면	196
[그림 V-10] 채점 관리 플랫폼 글쓰기 자료 할당 화면1	197
[그림 V-11] 채점 관리 플랫폼 글쓰기 자료 할당 화면2	197
[그림 V-12] 채점 관리 플랫폼 글쓰기 자료 할당 화면3	197
[그림 V-13] 채점 관리 플랫폼 글쓰기 자료 할당 화면4	198
[그림 V-14] 채점 관리 플랫폼 메시지 송수신 화면	199
[그림 V-15] 채점자 서약서 양식	206
[그림 V-16] 사전 채점 결과 공유 사례	207
[그림 V-17] 가교 채점 결과 입력 사례	208
[그림 V-18] 채점 피드백 사례	209
[그림 V-19] 플랫폼상 채점 완료 화면	211
[그림 V-20] 채점 진행률 점검 문자 발송 사례	212
[그림 V-21] 짝 채점 글별 평균 표준편차 구간 비율	216
[그림 VI-1] 문항 1 총점 분포	224
[그림 VI-2] 문항 2 총점 분포	224
[그림 VI-3] 문항 3 총점 분포	225
[그림 VI-4] 문항 4 총점 분포	225
[그림 VI-5] 문항 5 총점 분포	226
[그림 VI-6] 문항 6 총점 분포	226
[그림 VI-7] 확인적 요인 분석의 경로 모형 다이어그램	233
[그림 VII-1] 인공지능 학습용 글쓰기 평가 데이터 세트 구축 과정 ..	247
[그림 VII-2] 플랫폼 데이터 관리 화면	248

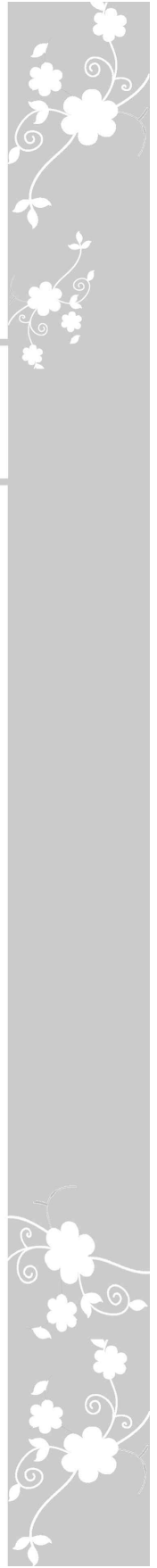
그림 차례

[그림 VII-3] 데이터 정제 방안	251
[그림 VII-4] 데이터 정제 단계별 수행 주체와 정제 항목	252
[그림 VII-5] 데이터 자동 편집 결과 글자 수 변경 사례	253
[그림 VII-6] 자동 정제 후 글자 수 변경 사례	254
[그림 VII-7] 논제 이탈 글쓰기 자료 사례	255
[그림 VII-8] 데이터 유사도 검사 결과 예시	257
[그림 VII-9] 지피티(chatGPT) 표절 사례	258
[그림 VII-10] 데이터 주석 절차	261
[그림 VII-11] 글쓰기 자료 데이터 주석 예시	272
[그림 VII-12] 글쓰기 채점 자료 데이터 주석 예시	277
[그림 VII-13] 자동 채점 프로토타입 기술 활용 프로세스	279
[그림 VII-14] 자동 채점 프로토타입 개발 내용	280



제 I 장

서론



1. 연구의 필요성 및 목적

1) 연구의 필요성

(1) 성인의 글쓰기 능력을 타당하게 평가하는 진단 체계의 필요

학문 및 산업 분야와 관련하여 업무 역량의 전문화와 고도화가 요구되는 현 사회에서 글쓰기 능력은 성인이 주제 탐구, 학술 연구, 과제 제안, 업무 기획, 수행 결과의 보고 등 다양한 과업을 성공적으로 수행함에 있어 필수적으로 요구되는 핵심 역량이다. 이에 대학 및 기업에서 인재를 선발·관리하고, 기관 내 글쓰기 교육의 질을 제고할 수 있도록 성인 대상 글쓰기 능력 진단 체계가 상용화될 필요가 있다. 국립국어원의 ‘인공지능 활용 국어능력 진단 체계 개발 기초 연구(국립국어원·서울대학교 국어교육연구소, 2022)’에서는 대학 교양 교육 담당자 및 기업 인사 담당자를 대상으로 면담 조사를 수행하여 대학 글쓰기 강의에서 학생들의 쓰기 능력 진단, 기업에서 쓰기 과제 수행 능력 고양 및 채용 시 쓰기 능력 인증을 위해 글쓰기 능력 진단 체계가 필요함을 강조하고 있다.

그러나 기존의 성인 대상 언어 능력 진단이나 검정을 위한 체계는 글쓰기 능력을 직접 평가하는 구성형 평가 도구로서의 기능을 하지 못하고 있다. 국어능력인증시험(TOKL)이나 KBS의 ‘한국어능력시험’, ‘한국실용글쓰기검정(KLATA)’ 등에서는 작문 영역에서 선택형 문항 또는 짧은 글 구성을 활용한 평가만이 시행되고 있으며 그 결과가 교육적으로 활용되지 못하고 있어 한계가 명확하다(국립국어원·서울대학교 국어교육연구소, 2022: 25-36).

따라서 성인의 글쓰기 능력 진단의 본질에 부합하는 글쓰기 유형별 평가 문항을 설계하고 채점 준거를 개발하며, 신뢰성과 타당성을 갖춘 고도화된 진단 체계를 갖추기 위한 연구가 필요하다. 또 실제 글쓰기 진단 평가를 반복적으로 수행하고 채점 자료를 대규모로 확보하여 피험자 집단의 글쓰기 능력에 대한 분석, 신뢰성과 타당성에 대한 지속적인 모니터링, 문항 유형별 채점 준거의 정교화, 평가 점수 의미 규명을 통한 진단 지표 가중치 산정 등 진단 체계의 고도화를 위한 단계적 목표를 설정하여 체계적으로 연구를 수행할 필요가 있다.

(2) 한국어 글쓰기 자동 채점 시스템 개발에 대한 요구

성인의 글쓰기 능력 제고를 위해 대규모 글쓰기 능력 진단을 정례화하려면 진단의 효율성과 경제성을 강화하기 위해 자동 채점(Automated Essay Scoring) 인공지능 시스템을 개발하여 채점 과정에 활용할 필요가 있다. 기존의 공공 및 민간에서 진행하는 대규모 글쓰기 능력 진단은 시간 소요 및 비용 문제와 맞물려 객관식을 통한 간접 평가나 한 단락 수준의 짧은 글을 쓰는 정도로 평가하는 데 머물고 있다. 이에 기존에 영어권, 프랑스어권 국가에서 주로 연구되었던 쓰기 자동 채점 기술을 한국어 쓰기 평가에 확대 적용할 필요성이 제기되어 왔다. 그간 한국어로 작성된 서답형 문항에 대해 자동 채점 기술을 개발·적용하는 일련의 연구가 수행되었으나(노은희 외, 2012, 2013, 2014, 2015, 2016) 문장 단위의 자동 채점이라는 점에서 활용도가 제한적이었다.

호주 교육 연구 위원회(ACER)에서는 성인을 대상으로 학업 및 업무 수행을 위해 일정 수준 이상의 영어 글쓰기 능력을 갖추고 있는지를 진단하기 위해 자동 채점 기술을 활용한 성인 대상 온라인 글쓰기 평가(OWA) 체계를 갖추어 상용화하고 있다. 또한 미국 교육부에서는 영어 자동 채점 시스템인 크레이즈(CRASE)를 성취도 향상 평가를 위한 보조금(EAG: Enhanced Assessment Grant) 제도의 운영에 활용한 바(Mughal & Samreen, 2018) 있다. 이와 같이 국민의 글쓰기 능력 증진을 위한 진단 체계를 효율적으로 도입, 운영하기 위해서는 자동 채점 시스템의 개발 및 활용이 필요하다.

현재 우리나라에서 성인을 대상으로 실시하는 연구 중 국립국어원의 ‘국민의 국어능력 실태 조사’는 그 대상이 5천여 명에 한정되어 있고, 교육적 차원에서 개인별로 채점 결과를 즉각적으로 제공하는 직접적인 환류 체계를 운영하지 못하는 한계가 있다. 호주와 같이 자동 채점에 기반한 성인 대상 글쓰기 능력 진단 체계를 갖추게 되면, 쓰기의 중요성에 대한 국민들의 인식을 제고하고 글쓰기 능력의 실제적인 증진에 기여할 수 있다.

(3) 진단 체계의 고도화를 위한 대규모 학습 데이터 구축

국민의 글쓰기 능력을 타당하고 신뢰성 높게 진단하는 평가 도구를 개발하며, 인공지능을 활용하여 글쓰기 자료를 자동 채점하는 고도화된 진단 체계를 개발하려면 글쓰기 자료를 대량으로 획득하고, 신뢰도가 담보된 채점 결과 데이터를 확보해야 한다. 현재 국내에서는 직장인 및 대학생과 같은 성인 대상 논증적 글쓰기 원자료가 대규모로 축적되

어 있지 않고, 하나의 통일된 진단 지표에 기반하여 채점된 쓰기 평가 자료 또한 공개된 데이터로 구축되어 있지 않다. 현재 사용되고 있는 글쓰기 자동 채점 시스템과 학습용 데이터 세트는 대부분 영어를 대상으로 하고 있으므로, 한국어를 포함한 영어 외 언어의 글쓰기 자동 채점 인공지능 개발용 학습 데이터 세트의 구축과 자동 평가 시스템의 개발은 실용적 가치뿐만 아니라 학문적 연구 가치도 크다(Wang et al., 2022).

획득할 데이터의 규모는 영어 자동 채점 모델 개발에 활용되고 있는 영어권의 청소년 글쓰기 평가 자료인 아삽(ASAP, Automated Student Assessment Prize) 데이터가 1만 7천여 건인 것을 기준으로 삼을 수 있다. 성인의 글쓰기 자료가 내용이 더 복잡하고 글의 길이가 더 길 수밖에 없는 특성을 고려하여 진단 체계를 개발하려면 중장기적으로 최소 3만 건 이상의 데이터 획득을 목표로 글쓰기 원자료를 축적할 필요가 있다.

대규모 학습 데이터 세트를 구축하는 초기 단계에서는 안정적인 데이터 수집 경로를 구축하는 것이 중요하다. 성인 대상 글쓰기 자료를 대규모로 수집하기 위한 현실적인 방안으로 주요 대학과의 협약, 성인 대상 대규모 평가 등을 모색할 필요가 있다. 또 초기 단계에서 비교적 적은 데이터로 높은 성능의 자동 채점 인공지능 프로토타입을 개발하려면 동일한 지표를 바탕으로 개발된 문항을 활용하여 글쓰기 자료와 채점 자료를 구축할 필요가 있다. 그러나 하나의 프롬프트 즉, 하나의 주제나 하나의 읽기 자료와 쓰기 지침으로 구성된 하나의 과제에 대한 글을 데이터 세트로 구축하여 자동 채점 시스템을 개발하면 다른 프롬프트에 적용하기 어렵게 된다. 중장기적으로는 다양한 글쓰기 유형을 고려한 추가 문항을 개발하여 교차-프롬프트 채점(cross-prompt scoring) 연구가 가능하도록 글쓰기 자료를 수집할 필요가 있다.

(4) 글쓰기 채점 전문 인력 양성

인공지능을 활용하여 글쓰기 자료를 자동 채점하는 고도화된 진단 체계를 개발하려면 신뢰성이 높은 채점 자료를 구축하는 것이 중요하다. 쓰기 자동 채점 시스템은 전문 채점자의 점수와의 일치도를 통해 그 성능을 평가하게 되므로, 애초 전문 채점자 그룹의 채점이 내적 일관성이나 채점자 간 신뢰도를 확보하지 못한다면 높은 성능을 가진 시스템을 개발할 수 없게 된다. 숙련된 채점자가 갖추고 있을 것으로 기대되는 채점의 전문성은 채점을 하는 목적과 목표 수준에 대한 이해, 그리고 이를 달성하기 위한 기술과 원리 등을 포함하며, 채점자는 채점 상황 가운데 이를 가용하여 최선의 채점 결과를 산출

할 수 있다. 반대로 채점 전문성이 결여된 경우에 채점자는 자신의 채점 행위가 무엇을 목표로 어떤 절차와 원리로 진행해야 하고, 어떻게 문제 상황에 대처해야 하는지 등에 대한 자원 부족 속에 부적합한 채점 결과를 산출하게 되며, 이는 채점의 신뢰도를 떨어뜨리고, 결과적으로 연구의 타당성을 위협하는 원인이 된다.

따라서 이러한 채점의 전문성을 갖춘 채점자를 단락 수준 이상의 긴 글을 읽고 신뢰할 수 있는 수준으로 일관된 채점을 하는 숙련된 인력으로 보고, 이를 양성하는 것이 주요 과제이다. 타당한 평가 문항과 객관적인 채점 기준과 방법이 마련되더라도 채점자의 전문성에 따라 타당한 채점 전략을 갖추고 있지 않을 수 있고, 피험자의 특성이나 글쓰기 과제 유형에 따라 채점 기준을 적절하게 적용하기 위해서는 체계적이며 집중적인 교육 및 훈련 과정을 통한 전문성 개발이 중요하기 때문이다. 또한 대규모 글쓰기 자료 채점 데이터를 확보하기 위해서는 일정 기간에 집중적으로 활용할 수 있는 전문 인력의 확보가 전제되어야 한다.

전문적인 쓰기 채점 인력 양성을 위한 프로그램의 사례로 언어교육 전문가 그룹인 미국외국어교육위원회(ACTFL)에서 제공하는 쓰기 능력 평가 채점자 인증(Writing Proficiency Test Rater Certification) 워크숍을 들 수 있다. 이 기관에서 쓰기 능력 평가 채점자로서 인증을 받기 위해서는 자체 제공하는 워크숍에 참여해야 할 뿐만 아니라, 평가 대상이 되는 언어에 대한 숙련도를 입증해야 하고, 자체 제공하는 쓰기 채점 가이드에 따라 쓰기 능력 평가 채점을 수행하고 채점 신뢰성을 입증해야 한다. 주로 외국어의 글쓰기 능력 평가의 채점 인력을 확보하기 위함이지만, 채점 전문성을 확보하기 위해서 전문적인 교육이 중요함을 시사한다.

2) 연구의 목적

‘2023년 국민의 글쓰기 능력 진단 체계 개발 연구’ 과제는 국민의 글쓰기 능력 향상을 도모할 수 있도록 효과적이고 신뢰할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있다.

이러한 연구 목적을 실현하기 위해서 다음과 같은 세부 연구 목표를 단계적으로 달성하고자 한다.

- 세부 목표 ① 국민의 글쓰기 능력 진단 지표 고도화
- 세부 목표 ② 고품질의 논증적 글쓰기 원자료 수집 및 채점 자료 구축
- 세부 목표 ③ 신뢰도 높은 채점 전문 인력 양성 프로그램 개발과 운영



[그림 I -1] 연구의 하위 목표

2. 연구의 내용 및 범위

1) 연구의 내용

(1) 국민의 글쓰기 능력 진단 지표 고도화

본 과업에서는 국민의 글쓰기 능력 진단 체계의 이용자의 관점에서 활용성과 타당성, 채점자의 관점에서 객관성과 유용성을 확보할 수 있는 글쓰기 능력 진단 지표와 채점 도구를 개발하고자 한다. 이에 아래와 같이 연구를 진행하였다.

2022년 기초 연구의 결과를 토대로 논증적 글쓰기 문항의 프로토타입을 보완하고, 프롬프트 구성과 난이도의 유사성을 유지한 문항 15개를 추가 개발하였다. 채점자 특성과 채점자 교육과의 관계를 고려하여, 2022년에 수집한 글쓰기 자료와 채점 자료를 활용하여 데이터 기반으로 채점 준거를 정교화하고 채점 기준표를 상세화하였다. 또한 2023년 상반기에 수집한 글쓰기 자료에 대한 채점 결과를 바탕으로 채점 준거에 대한 해설과 함께 채점 준거의 적용 사례를 제시한 채점 매뉴얼을 개발하였다.

다양한 장르의 글쓰기 능력을 종합적으로 진단할 수 있도록 범용 진단 지표의 가능성을 모색하며 설명적 글쓰기 문항과 채점 기준을 초안 수준에서 개발하였고, 타당성 확보를 위해 전문가 검토가 이루어졌다.

또 외부 전문가의 총체적 채점 결과와 본 연구에서의 채점 결과를 비교하여, 진단 지표의 타당성을 확보하였다. 마지막으로 채점 척도 구성의 이론적 타당성을 기반으로 실제 글쓰기 능력 추정의 안정성과 기대하는 글쓰기 능력 추정 효과를 강화하기 위해 점수 및 글쓰기 능력 수준의 의미에 대해 탐색하며 준거별 가중치 부여 방안을 제안하였다.

글쓰기 능력 진단 도구 고도화의 차원에서 이상 채점 데이터의 탐지 및 채점 인력 교육 도구로 활용할 수 있도록, 채점 준거별 점수 예측 인공지능 프로그램을 개발하였다. 더불어 향후 점수 예측 인공지능 프로그램을 신뢰할 수 있는 채점 지원 도구로 활용하기 위해 자동 채점 점수의 신뢰도 예측 프로그램, 평가 시스템과 연계를 위한 API 서비스를 개발하였다.

(2) 고품질의 글쓰기 원자료 수집 및 채점 자료 구축

본 과업에서 글쓰기 자료는 다양한 수준의 글쓰기 능력을 타당하게 진단할 수 있도록 진단 지표와 도구를 고도화하는 데 자료로서 활용되고, 자동 채점 기술을 개발하기 위한 데이터로 활용된다. 따라서 이러한 목적에 부합하는 고품질의 데이터가 구축될 수 있도록 ‘진단 목적에 부합하는 타당성’, ‘글쓰기 수준의 다양성’, ‘수준별 양적 충분성’을 고려하여 글쓰기 원자료 수집 모델과 글쓰기 자료 채점 모형을 설계하여야 한다(과학기술정보통신부, 2022).

글쓰기 원자료 수집과 정제, 채점이 유기적으로 이어질 수 있는 데이터 수집 종합 플랫폼을 개발하였다. 또한 채점의 신뢰성 확보를 위한 글쓰기 자료 채점 모형을 설계하였다. 이를 토대로 본 과업의 채점자 교육 프로그램 이수자를 대상으로 채점자를 선정, 문항에 특성화된 추가적인 워크숍 진행, 가교 채점과 채점 특성 피드백 제공, 복수 채점의 다양한 전략을 활용하여 글쓰기 자료 5,000여 건을 채점하였다. 이후 채점 정보를 포함하여 메타 정보를 주석화하는 제이슨(json) 파일 변환을 통해 학습 데이터 세트를 구축하였다.

(3) 신뢰도 높은 채점 전문 인력 양성 프로그램 개발과 운영

공신력을 갖춘 글쓰기 능력 진단 체계를 개발하고, 운영하기 위해서는 채점 전문성을 갖춘 인력을 충분히 확보하는 것이 중요하다. 본 과업은 대규모 글쓰기 원자료를 수집하고 채점하는 1차 연도 과업으로서, 협약 대학과의 협력을 통해 채점 인력으로 각 대학에서 글쓰기 과목을 가르치고 있는 교강사 100여 명을 모집하여 온라인 및 오프라인 채점 교육을 실시하였다.

본 과업의 세부 추진 과제는 채점자 교육 프로그램의 설계, 개발, 실행, 분석의 네 단계로 구분된다. 프로그램의 설계 단계에서는 채점자 교육에 대한 문헌 연구를 토대로 하되, 채점자 집단의 성격, 본 과업의 채점 맥락 등을 고려해 교육과정의 개요와 강의안을 구성하였다. 개발 단계에서는 채점자 교육 프로그램 운영 플랫폼을 확정하고, 강의 영상, 강의 자료 및 스크립트, 평가 문항 등 교육에 필요한 자료를 개발·제작하였다. 다음은 프로그램의 실제 운영 단계로서 참여자를 모집하고 온라인과 오프라인 교육을 실시하였다. 마지막 분석 단계에서는 채점자들의 자기 점검 및 평가를 위한 사전·사후 설문 도구를

활용하여 교육 실행의 결과를 분석하고 개선안을 도출하였다. 이러한 연구 결과를 종합하여 프로그램 운영 개요서를 개발하였다.

2) 연구의 범위

본 과업의 내용은 다음의 16개 하위 과제로 단계적으로 수행되었다.

<표 I -1> 연구의 범위

목표	하위 과제	세부 과제 내용
목표① 글쓰기 진단 지표 고도화	논증적 글쓰기 동형 문항 개발	<ul style="list-style-type: none"> 2022년 개발 논증적 글쓰기 문항 프로토타입 수정·보완 논증적 글쓰기 15개 문항 추가 개발
	채점 도구의 정교화	<ul style="list-style-type: none"> 논증적 글쓰기 채점 준거 타당화와 정교화 채점자 교육을 위한 채점 준거 수준별 글쓰기 자료 예시 구성 논증적 글쓰기 채점 기준표(rubric) 해설 및 채점 매뉴얼 개발
	설명적 글쓰기 문항 및 채점 준거 개발	<ul style="list-style-type: none"> 설명적 글쓰기 문항 개발 공통 지표 기반 설명적 글쓰기 채점 준거 개발
	진단 지표 가중치 연구	<ul style="list-style-type: none"> 진단 지표 가중치 연구의 목표 정립 가중치 설정 목표에 부합하는 가중치 방안 탐색
	인공지능 기반 채점 지원 도구 개발	<ul style="list-style-type: none"> 채점 준거별 점수 예측 인공지능(AI) 프로그램 개발 자동 채점 점수의 신뢰도 예측 프로그램 개발
목표② 글쓰기 자료 수집 및 채점 자료 구축	데이터 종합 관리 온라인 플랫폼 개발	<ul style="list-style-type: none"> 대규모 글쓰기 평가·채점 관리에 최적화된 플랫폼 개발 데이터 정제 및 가공 시스템 개발과 적용
	글쓰기 수집을 위한 진단 체계 홍보	<ul style="list-style-type: none"> 다양한 매체의 홍보 자료 개발 온라인 및 오프라인 활용 글쓰기 수집 참여 독려
	대학 협력 기반 글쓰기 자료 수집	<ul style="list-style-type: none"> 협약 대학 협의체와의 협력 체계 구축 및 상시 소통 온라인 플랫폼을 활용한 글쓰기 자료의 체계적 수집
	채점 자료 구축 모형 설계	<ul style="list-style-type: none"> 글쓰기 자료 목록, 채점 방법, 채점 설계를 포함한 채점 자료 구축 모형 개발과 절차 수립 채점 자료 선정(단일 점수 확정) 모형 설계 및 적용
	채점 실행	<ul style="list-style-type: none"> 채점자 선정 및 채점 일정 수립 워크숍 기반의 채점자 교육 실시 채점 특성 피드백의 제공 및 초점 대상 개별 면담 채점 설계에 따른 채점 대상 자료 할당

	채점 결과 분석 및 채점 자료 선정	<ul style="list-style-type: none"> ▪ 고전검사이론에 따른 채점 결과 분석 ▪ 문항반응이론에 따른 채점 결과 분석 ▪ 최종 채점 결과 자료의 선정
	데이터 정제 및 가공	<ul style="list-style-type: none"> ▪ 데이터 정제 절차 및 기준 마련 ▪ 데이터 자동/수동 정제 도구 개발과 적용 ▪ 메타 정보 주석화 및 데이터 검수
목표③ 글쓰기 채점 전문 인력 양성 프로그램 개발과 운영	프로그램 설계	<ul style="list-style-type: none"> ▪ 채점자 교육과정 설계 ▪ 차시별 강의안 구성
	프로그램 개발	<ul style="list-style-type: none"> ▪ 채점자 교육 프로그램 운영 플랫폼 확정 ▪ 강의 동영상 및 교수·학습 자료 제작 ▪ 자기 점검 및 평가를 위한 사전·사후 설문 도구 제작
	프로그램 운영	<ul style="list-style-type: none"> ▪ 참여자 모집 및 선정 ▪ 사전 설문 조사 시행 ▪ 온라인 및 오프라인 교육과정 실행 ▪ 현장 채점 실습 결과 분석
	프로그램 분석	<ul style="list-style-type: none"> ▪ 사후 설문 조사 시행 ▪ 이수 사정 및 이수증 제작·송부 ▪ 프로그램 개선안 마련

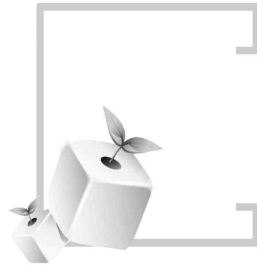
3. 연구의 방법

본 과업을 추진하기 위한 연구 방법을 16개의 하위 과제별로 <표 I -2>에서와 같이 구체화하였다.

<표 I -2> 연구의 방법

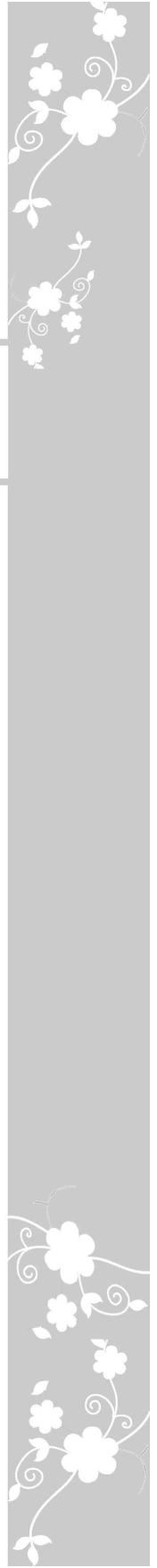
목표	하위 과제	연구의 방법
목표① 글쓰기 진단 지표 고도화	논증적 글쓰기 동형 문항 개발	<ul style="list-style-type: none"> 전문가 협의회를 통한 문항 개발 및 수정 협약 대학 교강사의 현장 적합성 검토 협약 대학 대학생 참여자 예비 평가 시행
	채점 도구의 정교화	<ul style="list-style-type: none"> 2022년 채점 자료 활용 채점 준거별 예시 글쓰기 자료 추출 2022년 채점 자료에서 채점자 간 채점 일치도가 높은 글쓰기 자료에서 어휘 및 문법 오류 정량화 외부 전문가의 총체적 채점 결과와의 통계적 비교 분석 전문가 검토와 내부 논의를 통한 채점 준거 수정 보완
	설명적 글쓰기 문항 및 채점 준거 개발	<ul style="list-style-type: none"> 설명적 글쓰기 평가 관련 국내외 연구 및 사례 고찰 대상, 주제, 목적 등을 고려한 문항 프로토타입 개발 및 채점 기준표(rubric) 개발 전문가 검토와 내부 논의를 통한 문항 및 준거 수정 보완
	진단 지표 가중치 연구	<ul style="list-style-type: none"> 이론적 탐색 및 내부 논의를 통한 가중치 모형 고안
	인공지능 기반 채점 지원 도구 개발	<ul style="list-style-type: none"> 한국어 사전학습 언어모델(BERT 계열)을 활용하여 사후학습(fine-tuning)을 통한 점수 예측 예측 점수의 확률 분포 교정(calibration)을 통한 신뢰도 예측
목표② 글쓰기 자료 수집 및 채점 자료 구축	데이터 종합 관리 온라인 플랫폼 개발	<ul style="list-style-type: none"> 협약 대학별 글쓰기 자료 수집 계획을 고려한 플랫폼 개발 대규모 글쓰기 평가 및 채점, 데이터 관리에 최적화된 온라인 플랫폼 기능 정의 연구진-플랫폼 개발 업체 간 상시 소통 체계 구축
	글쓰기 수집을 위한 진단 체계 홍보	<ul style="list-style-type: none"> 학교별 수요 조사에 기반한 맞춤형 홍보 방안 수립 다양한 매체 환경 및 이용 문화를 고려한 홍보 자료의 다양화
	대학 협력 기반 글쓰기 자료 수집	<ul style="list-style-type: none"> 대학 협의회 및 실무자 회의 운영 등 상시 소통 체계 구축 협약 대학별 계획 기반 자료 수집 모델 다양화
	채점 자료 구축 모형 설계	<ul style="list-style-type: none"> 국내외 대규모 수행 평가 채점 자료 구축 선행 연구 검토 문항 유형 및 채점자 수, 일정 등을 고려한 최적화된 채점 설계 개발 채점 자료 선정 모형 수립을 위한 예비 분석(2022년 자료)

	채점 실행	<ul style="list-style-type: none"> ▪ 워크숍 기반의 채점자 교육 실시 ▪ 행렬 표집 가교 채점 설계 ▪ 채점자 간 편차를 고려한 재채점 자료 선정
	채점 결과 분석 및 채점 자료 확정	<ul style="list-style-type: none"> ▪ 고전검사이론 및 문항반응이론에 따른 채점 결과 분석 ▪ 채점자 신뢰도(개인-평균), 기술통계(문항 및 준거별 평균, 표준편차), 준거 간 상관 분석 및 확인적 요인 분석, 채점 특성(채점 엄격성, 척도 사용의 일관성) 제공 ▪ 채점 특성 비교를 통한 대표 채점자 및 대표 채점 결과 선정
	데이터 정제 및 가공	<ul style="list-style-type: none"> ▪ 정제 기준 설정 및 보완 ▪ 데이터 수집 플랫폼 내 자동 정제 및 전문 인력 활용 수동 정제 ▪ 데이터 수집 플랫폼 내 프로그램 및 데이터베이스 활용 자동 주석화
목표③ 글쓰기 채점 전문 인력 양성	프로그램 설계	<ul style="list-style-type: none"> ▪ 채점자 훈련 프로그램 관련 국내외 문헌 및 사례 검토 ▪ 고경력 쓰기 평가 전문가 활용 강의안 구성
	프로그램 개발	<ul style="list-style-type: none"> ▪ 전문성 개발 프로그램 평가 설문 도구 검토 ▪ 쓰기 평가 전문가의 자료 개발-연구진의 피드백의 협력 체계
	프로그램 운영	<ul style="list-style-type: none"> ▪ 연구진-프로그램 참여자 간 상시 소통 체계 구축 및 온라인 교육 이수 현황 점검과 관리 ▪ 오프라인 교육 사전 점검을 위한 연구진 시뮬레이션 ▪ 채점자 일관성 및 채점자 간 신뢰도 분석을 포함하는 양적 분석 및 채점 과정 질적 자료 분석 실시
	프로그램 분석	<ul style="list-style-type: none"> ▪ 사전·사후 설문을 통한 프로그램 효과 검증 ▪ 설문 조사 분석을 통한 교육 프로그램 점검 및 개선안 도출



제 Ⅱ 장

글쓰기 진단 지표 고도화



1. 논증적 글쓰기 동형 문항 개발

1) 기개발된 논증적 글쓰기 문항의 프로토타입 수정·보완

2022년 개발된 논증적 글쓰기 문항(3개)은 발문(prompt)을 통해 쓰기 주제와 과제의 수사적 상황을 중립적으로 제시하고, 이에 대한 필자의 관점을 논리적으로 전개하도록 하는 ‘쟁점 제시형’의 방식으로 개발되었다. 이를 수정, 보완하여 논증적 글쓰기 문항의 프로토타입(prototype)을 확정하였다.

2023년 초부터 대중적으로 사용되는 대화형 생성 인공지능 기술(예: ChatGPT) 활용으로 인한 데이터 오염의 문제가 집중적으로 논의되었다. 그러나 과제 발문의 맥락을 구체적으로 제시할수록 대화형 생성 인공지능 기술이 더 우수한 글을 산출해 낼 수 있다는 우려로 인해 최종적으로 이를 문항 차원의 수정으로 반영하는 대신, 글쓰기 자료 수집 환경을 통제하는 쪽으로 고려하였다. 가령 응시자에게 통제된 환경에서 과제를 수행하도록 하거나, 카메라 등을 활용한 감독관 기능을 구현함으로써 이러한 문제가 일정 부분 해소될 수 있을 것으로 기대하였다.

과제 발문을 통해 소재에 대한 배경지식, 특히 쟁점에 대한 찬반 양측의 입장을 모두 제공하는 것이 자칫 응시자의 아이디어 생성에 방해가 될 수 있음을 고려하여, 기개발된 3개 문항 가운데 일부의 과제 발문을 250자 이내의 통일된 형태로 보다 간명하게 수정하였다. 이는 <표 II-1>과 같다.

<표 II-1> 논증적 글쓰기 문항의 프로토타입 수정·보완

문항 번호	수정 전	수정 후
1	로봇의 발달로 일자리가 줄어들 것이라는 사람들의 불안이 커지면서 최근 로봇세 도입에 대한 논의가 활발하다. 로봇세는 로봇의 노동으로 생산하는 경제적 가치에 부과하는 세금이다. 로봇 기술의 발달로 인해 우리 생활이 편리해진 부분도 있지만 그로 인해 일자리를 잃는 사람들이 갈수록 많아질 수 있기 때문에, 로봇세를 부과함으로써 로봇으로 인	로봇의 발달로 일자리가 줄어들 것이라는 사람들의 불안이 커지면서 최근 로봇세 도입에 대한 논의가 활발하다. 로봇세는 로봇의 노동으로 생산하는 경제적 가치에 부과하는 세금이다. 로봇 기술의 발달로 인해 <u>일 자리를 잃는 사람들이 갈수록 많아질 수 있기 때문에, 그런 사람들을 지원하거나 사회 안전망을 구축하기 위해 예산을 마련하자</u> 는 것이 로

문항 번호	수정 전	수정 후
	<p>해 일자리를 잃은 사람들을 지원하거나 사회 안전망을 구축하기 위해 예산을 마련하자는 것이 로봇세 도입의 목적이다.</p> <p>로봇세 도입에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. (288자)</p>	<p>봇세 도입의 목적이다.</p> <p>로봇세 도입에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. (236자)</p>
2	<p>혐오 표현이란 정당화될 수 없는 혐오 감정에 근거하여 특정 집단이나 개인에 대해 공개적으로 모욕, 비하, 멸시, 위협을 하거나 차별과 폭력을 선동하는 언어적, 비언어적 행위를 말한다. 혐오 표현은 혐오의 대상이 되는 개인 또는 집단의 존엄성을 부정하고 사회 전체에 미치는 해악이 크기 때문에 혐오 표현을 법적으로 규제해야 한다는 주장이 제기되고 있다.</p> <p>혐오 표현의 법적 규제에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. (239자)</p>	<p>좌동(수정 없음)</p>
3	<p>우리나라에서는 만 18세 이상이면 누구나 일정한 자격 구비를 통해 자동차 운전 면허를 취득할 수 있도록 하고 있다. 현행법상 자동차 운전 면허에 대한 연령 제한의 상한선은 없다. 그러나 최근 고령자의 대처 미흡으로 일어난 자동차 사고가 사회적 관심을 받으면서 일정 연령에 이르면 운전 면허를 반납하도록 해야 한다는 주장이 제기되고 있다. 우리 사회가 고령화되고 있는 상황에서 고령자에 대한 운전 면허 문제를 더 이상 외면할 수는 없다는 것이다.</p> <p>고령자 운전 면허 연령 제한에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. (294자)</p>	<p>우리나라에서는 <u>만 18세 이상이면 누구나 자동차 운전 면허를 취득할 수 있도록 하고 있다.</u> 그러나 최근 고령자의 대처 미흡으로 일어난 자동차 사고가 사회적 관심을 받으면서 일정 연령에 이르면 운전 면허를 반납하도록 해야 한다는 주장이 제기되고 있다. 우리 사회가 고령화되고 있는 상황에서 고령자에 대한 운전 면허 문제를 더 이상 외면할 수는 없다는 것이다.</p> <p><u>고령 운전자 면허 제한에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</u> (244자)</p>

2) 논증적 글쓰기 문항 추가 개발

수정, 보완한 논증적 글쓰기 문항의 프로토타입을 바탕으로 채점 준거와 채점 기준표를 공유하는 논증적 글쓰기 문항을 추가로 15개 개발하였다. 이는 글쓰기 채점 준거의 객관성과 타당성을 확보하기 위한 데이터 세트를 효과적으로 구축하는 데 기여할 것으로 기대된다.

논증적 글쓰기 동형 문항 개발은 <표 II-2>와 같은 절차로 진행되었다.

<표 II-2> 논증적 글쓰기 동형 문항 개발 절차

수행 단계	세부 내용
문항 개발 관련 전문가 섭외 및 개발 요청	· 문항 개발 전문가 섭외(5인), 기존 문항과의 구성과 양식 통일. 다양한 주제 분야를 반영한 15문항 개발
연구진 검토 및 전문가 협의회	· 이 연구에서 의도하는 논증적 글 과제에 맞게 개발되었는지 내부 검토 및 전문가 협의회 - 문항의 난이도, 간결성, 명확성 - 쟁점의 적절성, 시의성, 편향성, 다양성 - 예상 답안 수준의 다양성 - 채점 기준과의 관련성
전문가 및 연구진의 개발 문항 수정·보완	· 논의를 통해 수렴한 바를 반영하여 문항 수정 및 보완(1차)
교강사의 문항 검토	· 1차 수정 및 보완된 문항 대상으로 교강사의 현장 적합성 검토
문항 예비 평가	· 예비 평가 과정에 협약 대학의 협조 진행 · 문항당 대학생 글쓰기 참여자 2인의 글쓰기 수행을 통한 예비 평가 진행 - 일정 기간 내에 온라인으로 글쓰기 결과물 제출, 글쓰기 완료 후 문항 피드백 요청 - 문항 피드백은 문항 이해의 어려움, 문항 주제의 적절성, 답안 작성 과정에서의 어려움 등을 중심으로 질문
예비 평가 결과 검토, 문항 보완 및 확정	· 예비 평가 결과물(글쓰기 자료, 문항 피드백 자료)을 바탕으로 연구진 및 문항 개발 전문가의 문항 수정 및 보완(2차) 후 최종 문항 확정

(1) 문항 개발 관련 전문가 섭외 및 개발 요청

기개발한 논증적 글쓰기 문항과 발문 구성과 난이도의 유사성을 유지하며, 인문, 사회, 자연, 공학 등 다양한 분야의 주제를 포함하되, 융합적인 성격을 갖는 주제로 개발하고자 하였다. 문항 개발 전문가 5인에게 제공한 문항 개발 지침은 <표 II-3>과 같다.

<표 II-3> 문항 개발 지침

- 쓰기 주제 및 과제의 수사적 상황을 발문을 통해 중립적으로 제시한 후, 응시자가 자신의 관점(의견)을 논증하는 의견 제시형 에세이 과제를 개발할 것.
- 문항별 발문 길이는 250자 이내로 작성하며, 발문을 간결하게 작성하고 가급적 겹문장 사용을 지양할 것.
- 유의 사항은 기존 프로토타입 문항과 동일하게 제시할 것.

전문가 5인이 개발하고 연구진 검토를 거쳐 수정, 보완한 15개 문항 초안은 <표 II-4>와 같다.

<표 II-4> 2023년도 개발 논증적 글쓰기 문항 프로토타입

융합 A	<p>디지털 미디어에서 잊힐 권리란, 인터넷에서 사생활 침해 또는 명예훼손이 발생한 개인정보에 대해 검색 엔진에 정보 삭제를 요청할 수 있는 권리를 의미한다. 잊힐 권리가 대두된 배경에는 온라인에서 부정확하거나 부적절한 정보가 확산되어 피해를 입은 사람들의 권리 회복에 대한 관심이 자리하고 있다. 그러나 개인정보라 할지라도 공인(公人)에 대한 정보는 국민의 알 권리나 언론 기사의 저작권 문제와 충돌을 피하기 어려운 면도 있다.</p> <p>디지털 미디어에서의 잊힐 권리에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
융합 B	<p>인공지능(AI) 기술에 기반한 다양한 프로그램이나 생성형 로봇이 상용화됨에 따라 그 활용 분야도 넓어졌다. 문화예술 창작 분야에서도 AI가 만들어낸 창작물이 생산·유통되기 시작했다. 저작권이란 사람이 자신이 만든 창작물에 대해 갖는 법적 권리를 의미한다. 데이터를 학습해서 창작물을 만들어낸 AI에게 저작권을 인정해야 할지에 대한 사회적 논의가 필요한 시점이다.</p> <p>AI 저작권 인정에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p>

	<p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p>
<p>융합 C</p>	<p>메타버스는 3차원의 가상 세계를 의미한다. 이곳에서 이용자들은 특정 환경을 설정하고, 정교하게 구현된 아바타를 통해 현실처럼 상호 교류한다. 최근 메타버스 이용자가 늘면서 아바타 간에 발생하는 범죄를 처벌해야 한다는 주장이 제기되고 있다. 가상 세계에서 아바타가 겪는 폭력, 성범죄 등의 경험을 현실의 이용자가 겪는 실제적 경험과 동일하게 보아야 한다는 것이다. 반면에 가상 세계는 현실 세계와는 별개의 공간이므로 이를 범죄로 보기 어렵다는 상반된 입장도 있다.</p> <p>메타버스 내 범죄 처벌에 관한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p>
<p>융합 D</p>	<p>기술 도핑이란 스포츠에서 첨단 기술이 적용된 장비나 도구가 선수의 기록 향상에 영향을 주는 것을 말한다. 수영이나 달리기에서 기록 단축에 결정적인 영향을 주었던 전신 수영복이나 탄소섬유 운동화는 스포츠의 본질에 어긋나는 이유로 착용이 금지되기도 했다. 그러나 골프에서 기술 도핑 논란으로 금지되었던 채나 공이 시간이 흘러 보편적으로 사용되고 금지 또한 풀리게 된 사례도 있다. 이런 이유로 기술 도핑에 대한 규제를 첨단 기술에 대한 일시적 반발 정도로 보는 시각도 있다.</p> <p>기술 도핑에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p>
<p>인문 A</p>	<p>동물원은 동물의 보호와 연구를 진행하고, 관람을 통해 동물에 대한 지식을 넓히며 관람객들에게 오락 및 휴식을 제공하기 위한 목적으로 동물을 모아 기르는 곳이다. 그러나 최근에 동물 복지와 윤리에 대한 관심이 높아지면서, 동물원의 열악한 환경과 방치되거나 학대당하는 동물들의 상황을 우려하는 목소리들이 커지고 있다. 동물원이 멸종 위기종 보호라는 명분만 내세울 뿐 실제 동물의 삶의 질 향상에 도움이 되지 않으며, 희귀 동물에 대한 인간이 소유욕이 밀려 부추겨 생태계 파괴를 유발하고 있다는 것이다.</p> <p>동물원 폐지에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p>
<p>인문 B</p>	<p>존엄사란 인간으로서 최소한의 품위와 존엄성을 유지하며 죽음을 맞이하는 것을 말한다. 이와 관련해 우리나라는 2018년부터 임종 과정에 있는 환자에</p>

	<p>한해 무의미한 연명의료를 보류하거나 중단할 수 있도록 하고 있다. 그러나 불치병으로 인해 극심한 고통에 시달리더라도, 임종 과정에 있지 않은 환자는 여전히 연명치료를 중단할 수 없는 실정이다. 이에 가망 없는 환자가 의사의 도움을 받아 삶을 마감할 수 있도록 하는 조력존엄사를 허용하자는 주장이 제기되고 있다.</p> <p>조력존엄사에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 A</p>	<p>노키즈존(No Kids Zone)이란 영유아나 어린이의 출입을 금지하는 구역을 말한다. 최근에 식당이나 카페에서 어린이들이 가게의 다른 손님들에게 피해를 준다는 이유로 노키즈존을 설정하는 사례들이 늘고 있다. 이러한 행위는 손님들의 편익은 물론 어린이의 안전, 업소의 이익을 위해 필요한 조치로 보기도 한다. 그러나 하나의 인격체인 어린이를 차별하고 나아가 모든 어린이에 대해 사회적으로 부정적 인식을 조장한다는 측면에서 논란이 되고 있다.</p> <p>노키즈존 설정에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 B</p>	<p>성적이 우수한 학생들에게 주어지는 장학금을 없애고 저소득층 학생들을 위한 장학금을 늘리겠다는 대학이 많아지고 있다. 공부를 잘하는 학생이 아니라 형편이 어려운 학생에게 장학금을 주어서, 저소득층 학생들이 학비 때문에 학업을 소홀히 하는 일이 없도록 하겠다는 것이다. 즉 소득분위를 기준으로 장학금을 지급함으로써 교육받을 기회를 공평하게 제공해야 한다는 주장이다. 그러나 성적에 따른 장학금 제공이 학생들의 학업 성취 욕구나 보상을 위해 필요하다는 의견도 있어, 성적 우수 장학금 폐지를 둘러싼 논란이 커지고 있다.</p> <p>성적 우수 장학금을 폐지하는 것에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 C</p>	<p>블라인드 채용은 지원자의 출신 지역이나 학력, 성별, 연령 등 직무 능력과 무관한 요소를 배제하는 채용 방법이다. 블라인드 채용은 균등한 고용 기회를 보장하는 방법으로 자격증 및 관련 경험만으로 지원자를 평가할 수 있어 채용 과정의 투명성과 공정성을 높이는 데 기여하였다. 그러나 지원자의 전문성 판단에 핵심 요소인 학문적 성과나 실적 등을 파악할 수 없어 업무에 적합한 인재를 뽑기 어려운 ‘깜깜이 채용’이라는 비판이 제기되고 있으며, 연구 기관을 중심으로 블라인드 채용을 폐지하려는 움직임이 나타나고 있다.</p> <p>블라인드 채용을 폐지하는 것에 대한 자신의 의견을 논리적으로 제시하는 글</p>

	<p>을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 D</p>	<p>우리나라는 인구 고령화 현상과 청년 인구 감소가 동시에 급격하게 진행되고 있다. 산업 현장에서의 생산 가능 인구가 급격하게 줄어드는 위기를 대비하기 위해 정부는 정년 연장을 고려하고 있다. 일정한 나이가 되면 퇴직을 하던 근로자들이 노동을 계속할 수 있는 제도를 마련하여 더 많은 인구가 생산 활동에 참여하도록 하자는 것이다. 이런 정책이 오히려 청년 일자리를 감소시키거나 청년 실업을 심화시킬 수 있다는 반론도 제기되고 있다.</p> <p>정년 연장에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 E</p>	<p>촉법소년은 만 10세 이상에서 만 14세 미만의 미성년자 중 범법 행위자를 말한다. 우리나라에서 촉법소년은 형법에 따라 범죄를 저질러도 형사 처벌을 받지 않고 보호 처분의 대상이 된다. 그러나 최근 촉법소년의 범죄가 증가하면서 촉법소년의 연령을 낮추어야 한다는 주장이 제기되고 있다. 늘어나는 청소년의 범죄에 대해 방관하고 있을 수는 없다는 것이다. 반면에 어린 나이의 소년범이 사회에 복귀하여 건설한 사회인으로 성장할 기회를 주어야 한다는 의견도 있다.</p> <p>촉법소년 연령 하향에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 F</p>	<p>최근 청년 및 노인 동거인, 성 소수자 동거인 등 가족의 형태가 다양해졌다. 우리나라에서는 혼인·혈연·입양만을 가족의 조건으로 본다. 따라서 이에 해당하지 않는 구성원은 주거, 의료 등 법과 행정 제도로부터 보호받지 못한다. 그러나 생활동반자* 역시 가족으로 인정해야 한다는 주장이 제기되고 있다. 법적 가족 개념과 시민적 권리를 확대하자는 것이 생활동반자법의 취지이다. 반면에 기존 가족 제도의 변화로 인해 사회의 혼란을 야기할 수 있다는 입장도 존재한다.</p> <p>생활동반자법 도입에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. *생활동반자: 1명과 동거하며 부양하고 협조하는 관계를 맺고 있는 성인</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>

<p>사회 G</p>	<p>헌법재판소는 낙태죄를 규정한 형법 조항에 대해 2019년 헌법 불합치 결정을 내렸다. 이후 여러 개정안이 충돌하며 법이 개정되지 못해 임신 중지에 관한 형법 조항은 존재하지 않게 되었다. 이에 산부인과 등 관련 현장에서는 혼란이 계속되고 있다. 법무부에서는 태아의 생명권과 여성의 자기 결정권에 관한 의견들을 절충하여, 임신 14주까지는 임신 중지를 허용하고 15주-24주까지는 법에서 정하는 사유에 해당될 경우에만 임신 중지를 허용하자는 개정안을 내놓았다.</p> <p>임신 중지 허용에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>과학·기술 A</p>	<p>온실가스 감축을 위한 세계적인 움직임 속에서 원자력 에너지가 다시 주목받고 있다. 원자력 에너지는 온실가스의 주범인 화석 연료를 대체하는 에너지이자 산업 분야의 안정적인 전력 공급원이기 때문이다. 그러나 원전을 둘러싼 안전 문제에 대한 우려는 원자력 에너지 사용의 걸림돌이기도 하다. 원전 폭발 사고가 발생하면 국가적 피해는 물론 주변국의 피해까지 막대하다는 점에서 원자력 발전소의 가동을 중지하거나 폐기를 추진하자는 의견도 존재한다.</p> <p>원자력 에너지 사용에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>과학·기술 B</p>	<p>동물 실험이란 교육, 시험, 연구 및 생물학적 제제의 생산 등 과학적 목적을 위해 동물을 대상으로 하는 실시하는 실험 또는 그 과학적 절차를 가리킨다. 지금까지 동물 실험이 의학을 비롯한 다양한 산업 및 학문의 발전에 기여한 부분은 분명하다. 그러나 동물 실험이 이루어지는 분야가 넓어지면서 실험에 동원되는 동물의 개체 수가 증가하였고, 지나치게 잔인한 방식으로 진행되는 실험 사례 등이 알려지면서 동물 실험의 정당성에 대한 의문이 제기되기도 하였다.</p> <p>동물 실험에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>

(2) 전문가 협의회

개발된 15개의 문항에 대하여 국어교육 전공의 전문가 2인과 전문가 협의회를 실시하였다. 전문가 협의회는 <표 II-5>와 같은 항목에 대한 의견을 공유하는 형태로 이루어졌다.

<표 II-5> 문항 프로토타입에 대한 전문가 협의회 점검 항목

대주제	점검 항목 및 세부 요소
평가 문항 주제에 관한 의견	(1) 평가 문항의 논제는 각각 대학생의 논증적 글쓰기를 위한 주제로 적절한가? <ul style="list-style-type: none"> ○ 각 논제는 대학생의 논증적 글쓰기 주제로 적절한가? ○ 만약 적절하지 않다면 대안적 주제로 제시될 만한 논제로는 어떤 것이 가능한가?
	(2) 평가 문항의 주제는 대학생의 다양한 반응을 이끌어 내기에 적절한가? <ul style="list-style-type: none"> ○ 대학생으로부터 찬성과 반대의 의견을 균형 있게 수집할 수 있는가? ○ 만약 적절하지 않다면 어떻게 보완할 수 있는가?
평가 문항의 발문에 관한 의견	(1) 평가 문항의 발문 방식은 적절한가? <ul style="list-style-type: none"> ○ 각 문항의 발문 방식이 문고자 하는 바를 명료하게 드러내는가? ○ 특정 답안(찬성 혹은 반대)을 유도할 수 있는 여지는 없는가?
	(2) 평가 문항의 발문이 대학생의 글쓰기 능력을 평가하는데 적절한가? <ul style="list-style-type: none"> ○ 평가 문항의 발문이 참여자의 선행 지식에 따라 다르게 영향을 줄 여지는 없는가? ○ 글쓰기 능력 외의 다른 요소가 평가에 영향을 주지는 않는가?
평가 문항 난이도에 관한 의견	(1) 평가 문항의 분량 및 난이도는 적절한가?
	(2) 평가 문항 간 난이도를 비교하였을 때 유사한 수준이라 볼 수 있는가?
기타 의견	

(3) 교강사 문항 검토

개발된 15개의 문항에 대해서는 협약 대학의 교강사(16명)로부터 현장 적합성 검토를 받았다. 이들은 각 문항에 대해 아래와 같은 질문에 답하였다.

- ① 이 문항에 대하여 대학생들의 글쓰기 결과, 찬반의 균형은 어떠할 것으로 예상하시나요?
- ② 대학생들에게 이 문항의 화제는 어느 정도로 친숙할 것으로 예상하시나요?

- ③ 대학생들에게 이 문항의 난이도는 어느 정도일 것으로 예상하시나요?
- ④ 이 문항이 대학생들의 논증적 글쓰기 능력을 평가하기에 어느 정도로 타당하다고 생각하시나요?
- ⑤ 이 문항과 관련하여 의견이 있으신 경우 자유롭게 말씀해 주십시오.

(4) 문항 예비 평가

15개 문항에 대해서는 문항당 협약 대학의 대학생 글쓰기 참여자 2인(총 30명)으로부터 예비 평가도 받았다. 이들은 실제 검사 환경과 동일하게 글을 작성하고, 다음 질문에 답하였다.

- ① 이해하기 어려웠거나, 의미가 모호한 부분이 있었다면 구체적으로 언급하여 의견을 기술해 주시기 바랍니다.
- ② 답안을 작성하는 데 소요된 시간은 몇 분이었으며, 적절하지 않았다면 주된 이유를 기술해 주세요.
- ③ 답안을 작성하는 데 있어 가장 큰 어려움은 무엇이었으며, 그 어려움을 어떻게 해결하셨나요?
- ④ 대학생들이 이 문제를 해결할 때, 찬성과 반대의 입장이 균형 있게 나올 것으로 생각하시나요? 특정 입장에 쏠릴 것 같다면, 그렇게 판단한 이유를 기술해 주세요.
- ⑤ 대학생들에게 문제의 난이도는 어떠하다고 생각하시나요? 그렇게 판단한 이유를 기술해 주세요.
- ⑥ 문항과 관련된 의견이 있으시거나 수정이 필요하다고 생각하시는 부분이 있으면, 자유롭게 적어 주세요.

(5) 문항 재수정

전문가 의견, 교강사 검토, 대학생 글쓰기 참여자들의 예비 평가 결과(글쓰기 자료, 문항 피드백 자료)를 바탕으로 연구진과 문항 개발 전문가가 협의하여 문항을 수정하였다. 수정 작업을 거친 문항은 <표 II-6>과 같다.

<표 II-6> 2023년도 개발 논증적 글쓰기 수정 문항

<p>융합 A</p>	<p>디지털 세계 또는 온라인에서 잊힐 권리란, 인터넷에서 사생활 침해 또는 명예훼손이 발생한 개인정보에 대해 검색 엔진에 정보 삭제를 요청할 수 있는 권리를 말한다. 이는 온라인에서 부정확하거나 부적절한 정보가 확산되어 피해를 입은 사람들의 권리 회복에 대한 관심에서 대두되었다. 그러나 개인정보라 할지라도 공인(公人)에 대한 정보에 잊힐 권리를 주장하는 것은 국민의 알 권리나 언론 기사의 저작권을 보장하는 것과 충돌할 수 있다.</p> <p>디지털 미디어에서의 잊힐 권리 보장에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>융합 B</p>	<p>인공지능(AI) 기술에 기반한 다양한 프로그램과 로봇이 상용화되면서 활용 분야가 넓어졌다. 문화예술 창작 분야에서도 AI가 만들어낸 작품이 생산·유통되기 시작했다. 그러나 저작권은 창작물을 만든 사람의 법적 권리이다. AI가 데이터를 학습해 창작물을 만들어낸 경우, 저작권을 누구에게 인정해야 할지(예: 원본 창작자, AI, AI 이용자 등)에 대한 사회적 논의가 필요하다.</p> <p>AI가 만들어 낸 창작물에 대해 저작권을 인정하는 것에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>융합 C</p>	<p>메타버스는 3차원의 가상 세계를 의미한다. 이곳에서 이용자들은 특정 환경을 설정하고, 정교하게 구현된 아바타를 통해 상호 교류한다. 최근 메타버스 이용자가 늘면서 아바타 간에 발생하는 범죄를 처벌해야 한다는 주장이 제기되고 있다. 가상 세계에서 아바타가 겪는 폭력, 성범죄 등을 현실의 이용자에게 가해지는 범죄와 동등하게 다루어야 한다는 것이다. 반면 가상 세계는 현실 세계와 별개의 공간이기 때문에 범죄로 간주하기 어렵다는 반대 의견도 있다.</p> <p>메타버스 내 범죄에 대해 현실 세계에서 처벌하는 것에 관한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>융합 D</p>	<p>기술 도핑이란 스포츠에서 선수의 기록 향상에 기여하지만 사용이 금지되어 있는 첨단 기술이나 장비를 사용하는 것을 말한다. 전신 수영복이나 탄소섬유 운동화 등 일부 장비는 스포츠의 본질에 어긋나는 것으로 여겨져 사용이 금지되었었다. 그러나 기술 도핑 논란이 있었던 골프용 채나 공과 같은 일부 장비가 보편적으로 사용됨에 따라 그 금지가 풀리는 경우도 있다. 이에 기술 도핑 규제를 첨단 기술에 대한 일시적 반발 정도로 보는 시각도 있다.</p>

	<p>기술 도핑 인정에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
인문 A	<p>동물원은 동물을 모아 보호, 연구, 교육, 오락 및 휴식 등의 목적으로 운영되는 곳이다. 그러나 최근에 동물 복지와 윤리에 대한 관심이 높아지면서, 동물원의 열악한 환경과 동물 학대 문제를 우려하는 목소리가 커지고 있다. 동물원이 멸종 위기종 보호라는 명분만을 내세울 뿐 실제 동물의 삶의 질 향상에 도움이 되지 않으며, 희귀 동물에 대한 인간의 소유욕이 밀렵을 부추겨 생태계 파괴를 유발한다는 것이다.</p> <p>동물원 폐지에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
인문 B	<p>존엄사란 인간으로서 최소한의 품위와 존엄성을 유지하며 죽음을 맞이하는 것을 말한다. 우리나라에서는 2018년부터 임종 과정에 있는 환자에 한해 무의미한 연명치료를 보류하거나 중단할 수 있도록 규정하고 있다. 그러나 불치병으로 인해 극심한 고통에 시달리더라도, 임종 과정에 있지 않은 환자는 여전히 연명치료를 중단할 수 없는 상황이다. 이에 가망 없는 환자가 의사의 도움을 받아 삶을 마감할 수 있도록 하는 조력존엄사를 허용하자는 주장이 제기되고 있다.</p> <p>조력존엄사 허용에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
사회 A	<p>노키즈존(No Kids Zone)은 영유아나 어린이 출입이 제한된 구역을 말한다. 최근에는 어린이들이 다른 손님에게 피해를 준다는 이유로 사업자가 자신의 식당이나 카페를 노키즈존으로 설정하는 사례가 늘고 있다. 이러한 조치는 손님들의 편의와 어린이의 안전, 그리고 업소의 이익을 위한 필요한 조치로 보기도 한다. 그러나 하나의 인격체인 어린이를 차별하고 나아가 모든 어린이에 대해 사회적으로 부정적인 인식을 조장한다는 측면에서 논란이 되기도 한다.</p> <p>노키즈존 설정에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항] △ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
사회 B	<p>많은 대학들이 성적이 우수한 학생들을 위한 장학금을 줄이고, 대신 저소득층 학생들을 위한 장학금을 늘리는 방향으로 나아가고 있다. 이는 학생들이 경제적 어려움으로 학업을 소홀히 하는 일이 없도록 하려는 조치이다. 소득분위</p>

	<p>를 기준으로 장학금을 지급함으로써 교육받을 기회를 공평하게 제공해야 한다는 것이다. 그러나 학생들의 학업 성취 욕구나 보상을 위해 성적 우수 장학금 유지가 필요하다는 의견도 있다.</p> <p>성적 우수 장학금의 비율을 축소하는 것에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 C</p>	<p>블라인드 채용은 출신 지역, 학력, 성별, 연령, 외모 등과 같이 직무 능력과 무관한 요소를 배제하는 채용 방법이다. 자격증 및 관련 경험만으로 지원자를 평가하여 채용 과정의 공정성과 투명성을 높이고 균등한 고용 기회를 보장하려는 것이다. 그러나 이로 인해 전문성 판단의 핵심 요소들을 파악하기 어려워 '깜깜이 채용'이라는 비판이 제기되고 있다. 연구기관을 중심으로 블라인드 채용을 폐지하려는 움직임도 나타나고 있다.</p> <p>블라인드 채용을 폐지하는 것에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 D</p>	<p>우리나라는 인구 고령화와 청년 인구 감소가 동시에 급격히 진행되고 있다. 산업 현장에서의 생산 가능 인구가 급격하게 줄어드는 위기를 대비하기 위해 정부는 정년 연장을 고려하고 있다. 일정한 나이가 되면 퇴직을 하던 근로자들이 노동을 계속할 수 있는 제도를 마련하여 더 많은 인구가 생산 활동에 참여하도록 하자는 것이다. 그러나 이러한 정책이 오히려 청년 일자리를 감소시키거나 청년 실업을 심화시킬 수 있다는 반론도 제기되고 있다.</p> <p>정년 연장에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 E</p>	<p>촉법소년은 만 10세 이상에서 만 14세 미만의 미성년자 중 범법 행위자를 말한다. 우리나라에서 촉법소년은 형법에 따라 범죄를 저질러도 형사 처벌을 받지 않고 보호 처분의 대상이 된다. 그러나 최근 촉법소년의 범죄가 증가하면서 촉법소년의 연령을 낮추어야 한다는 주장이 제기되고 있다. 늘어나는 청소년의 범죄에 대해 방치할 수 없다는 것이다. 반면 어린 나이의 소년범이 사회에 복귀하여 건실한 사회인으로 성장할 기회를 주어야 한다는 의견도 있다.</p> <p>촉법소년 연령 하향에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>

<p>사회 F</p>	<p>가족의 형태가 청년 및 노인 동거인, 성 소수자 동거인 등으로 다양해지고 있다. 우리나라에서는 가족의 조건을 혼인, 혈연, 입양으로 제한적으로 규정하고 있으며, 이에 해당하지 않는 가족 구성원은 법적으로 보호받지 못한다. 최근 생활동반자* 역시 가족으로 인정해야 한다는 주장이 제기되고 있다. 생활동반자를 가족으로 인정하고, 시민적 권리를 보호하자는 것이다. 반면에 기존 가족 제도의 변화로 인해 사회의 혼란을 야기할 수 있다는 입장도 존재한다.</p> <p>생활동반자법 도입에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. *생활동반자: 1명과 동거하며 부양하고 협조하는 관계를 맺고 있는 성인</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>사회 G</p>	<p>헌법재판소는 2019년에 낙태죄를 규정한 형법 조항에 대해 헌법 불합치 결정을 내렸다. 그러나 이후 여러 개정안이 충돌하며 법을 개정하지 못해 산부인과 등 관련 현장에서는 혼란이 이어지고 있다. 법무부에서는 태아의 생명권과 여성의 자기 결정권에 관한 의견들을 절충하여, 임신 14주까지는 임신 중지를 허용하고 15-24주에는 법에서 정하는 사유에 해당될 경우에만 임신 중지를 허용하자는 개정안을 내놓았다.</p> <p>임신 중지 허용에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>과학·기술 A</p>	<p>온실가스 감축을 위한 세계적인 움직임 속에서 원자력 에너지가 다시 주목받고 있다. 원자력은 화석 연료를 대체하는 대체 에너지이자 산업 분야의 안정적인 전력 공급원이기 때문이다. 그러나 원전 안전 문제에 대한 우려는 원자력 에너지 사용의 걸림돌이기도 하다. 원전 폭발 사고 발생 시 국가적인 피해는 물론 주변국의 피해까지 커질 수 있으므로 원자력 발전소의 가동을 중지하거나 폐기를 적극적으로 추진해야 한다는 의견도 존재한다.</p> <p>원자력 에너지 존폐에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>
<p>과학·기술 B</p>	<p>동물 실험은 과학적 목적을 위해 동물을 대상으로 하는 실험 또는 그 과학적 절차를 말한다. 지금까지 동물 실험이 의학을 비롯한 다양한 산업 및 학문의 발전에 기여한 부분은 분명하다. 하지만 동물 실험이 이루어지는 분야가 넓어지면서 실험에 동원되는 동물의 개체 수가 증가하고, 지나치게 잔인한 방식으로 진행되는 실험 사례 등이 알려지면서 동물 실험의 정당성에 대한 의문이 제기되고 있다.</p> <p>동물 실험 폐지에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.</p> <div style="border: 1px solid black; padding: 5px;"> <p>[유의 사항]</p> <p>△ 서론, 본론, 결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것).</p> <p>△ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분</p> </div>

(6) 플랫폼 문항 입력

2023학년도 1학기에는 6개의 문항을 활용하고, 2학기에는 3문항을 추가로 활용하는 것으로 결정하였다. 이에 따라 2023학년도에 글쓰기 자료를 수집하는 문항은 총 9개의 문항이다. 응시자들의 플랫폼 문항 선택을 위한 발문과 용어 정의는 <표 II-7>과 같이 제시하였다.

<표 II-7> 2023학년도 글쓰기 원자료 수집용 논증적 글쓰기 문항

문항 번호 Q1) 로봇세를 도입해야 하는가?

* 로봇세: 로봇의 노동으로 생산하는 경제적 가치에 부과하는 세금

문항 번호 Q2) 혐오 표현을 법적으로 규제해야 하는가?

* 혐오 표현: 특정 집단이나 개인에 대해 공개적으로 행하는 차별적/폭력적인 언어적, 비언어적 행위

문항 번호 Q3) 고령 운전자의 면허를 제한해야 하는가?

* 운전 면허 제한: 일정 연령 이상인 운전자의 운전 면허를 반납하도록 하는 제도

문항 번호 Q4) 디지털 세계에서 잊힐 권리를 보장해야 하는가?

* 잊힐 권리: 디지털상의 개인정보 보호와 관련된 개념으로, 개인이 자신의 정보를 삭제하거나 숨길 수 있는 권리

문항 번호 Q5) 조력존엄사를 허용해야 하는가?

* 조력존엄사: 회생 가능성이 없는 환자가 의사의 도움을 받아 삶을 마감하는 행위

문항 번호 Q6) 블라인드 채용을 폐지해야 하는가?

* 블라인드 채용: 출신 지역, 학력, 성별, 연령, 외모 등과 같이 직무 능력과 무관한 요소를 배제하는 채용 방법

문항 번호 Q7) AI를 활용하여 창작물을 만들어낸 사람에게 저작권을 부여해야 하는가?

* 저작권: 인간의 생각이나 감정을 표현한 글, 그림, 영상 등의 창작물을 만든 이가 자기 저작물에 대해 가지는 배타적인 법적 권리

문항 번호 Q8) 동물 실험을 폐지해야 하는가?

* 동물 실험: 과학적 목적을 위해 동물을 대상으로 하는 실험 또는 그 과학적 절차

문항 번호 Q9) 촉법소년의 연령을 하향해야 하는가?

* 촉법소년: 만 10세 이상에서 만 14세 미만의 미성년자 중 범법 행위자

<표 II-7>의 문항에서 Q1부터 Q6의 문항은 1학기에 글쓰기 자료를 수집하고, Q4부터 Q9의 문항은 2학기에 글쓰기 자료를 수집하였다. 약 9,000건의 글쓰기 자료를 수집한 후 그중 5,000건을 채점하여 자동 채점 모델 개발에 활용하였다.

2. 채점 도구의 정교화

1) 기개발된 논증적 글쓰기 채점 기준표의 상세화

일반적으로 채점자는 채점 기준표를 내부적으로 재구성하여 표상한 기준을 적용하는 경향이 있기 때문에(Bejar, 2006), 연구진의 의도를 담아 만든 채점 기준표와, 이를 바탕으로 채점에 적용하는 채점자 간의 간극을 줄이기 위해서는 채점자 교육 프로그램을 체계적으로 설계할 뿐만 아니라, 먼저 채점 기준표를 정제하고 상세화하여 채점자의 이해와 편의를 제고할 필요가 있다.

논증적 글쓰기 채점 기준표를 상세화하는 작업 절차는 <표 II-8>과 같다. 먼저 예시 답안 선정과 채점 기준표 정제를 위한 외부 전문가를 섭외하였다. 채점자 교육용 자료를 제작한다는 목표를 견지하며 채점 기준표를 구체화 및 상세화하고 예시 답안 자료를 선별하였다. 이러한 일련의 과정을 연구진과 함께 협력하고 의논하며 지속적으로 검토하고 논의를 수렴해 갔다.

<표 II-8> 채점 기준표 상세화 절차

수행 절차		
· 예시 답안 선별 및 채점 기준표의 정제 작업을 위한 전문가 섭외 및 의뢰	→ ←	연구진 및 전문가의 협의회를 통한 지속적인 검토 및 논의 수렴
· 채점 기준표의 하위 준거별, 수준별 기술 내용을 뒷받침하는 예시 자료 선별 및 채점자 교육 자료로의 구성 - 하위 수준별, 기술별로 전형성을 보여주는 예시 자료 선별 - 예외적 속성의 특이한 답안까지 선별 범위에 포함		
· 채점자 교육 자료를 고려하여 채점 기준표의 정제 및 상세화 - 용어의 일관적이고 명확한 사용 지향, 추상의 정도가 높은 표현은 보다 구체적인 표현으로 정련.		

(1) 채점자 교육용 자료 제작을 위한 예시 답안 선정

먼저, 채점자 교육용 자료 제작을 위하여 채점 기준표 정제 작업과 예시 답안 자료 선별 작업을 함께 진행할 외부 전문가 2인을 섭외하였다. 이들 전문가는 각각 현직 국어 교사와 국어교육 전공 박사 학위 소지자로서 글쓰기 및 국어 규범 관련 전문성이 있다.

외부 전문가 2인과 협력하여, 2022년에 수집한 다양한 수준의 논증적 글쓰기 자료와 채점 자료 300여 건 중 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.

이들 전문가와 협력한 예시 답안 선정 과정은 ‘내용’과 ‘조직’ 범주의 예시 답안을 우선적으로 선별하는 것으로 진행되었다. 이는 올해 과업 중 하나로 후술할 ‘표현 범주 채점 준거 정량화 작업’이 있으므로, ‘표현’ 범주의 예시 답안은 해당 작업에 참여하는 전문가와 함께 진행하는 것이 보다 정교한 작업이 가능하고 효율적으로 진행할 수 있을 것이라는 판단을 하였기 때문이다. 일차적으로 선별된 예시 자료는 내부 연구진이 검토하였고, 이후 전문가 협의회를 거쳐 최종 확정하였다. 이러한 예시 자료는 채점자 교육과정의 자료로서 투입하여 신뢰도 높은 채점 결과를 산출하는 데 활용하고자 하였다. 선별된 예시 답안 자료의 개략적인 모습은 [그림 II-1]과 같다.

항목	내용 1: 문제 상황 이해	내용 2: 주장	내용 3: 이유/근거의 특성	내용 4: 이유/근거의 충분성	내용 5: 다른 입장에 대한 고려	조직 1: 문단 구성	조직 2: 문단 구성 요건
사	<p>문제 상황 이해</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>	<p>내용 2: 주장</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>	<p>내용 3: 이유/근거의 특성</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>	<p>내용 4: 이유/근거의 충분성</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>	<p>내용 5: 다른 입장에 대한 고려</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>	<p>조직 1: 문단 구성</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>	<p>조직 2: 문단 구성 요건</p> <p>2022년 1월 1일부터 2022년 12월 31일까지 1년간 실시된 '2022년 국민 글쓰기 능력 진단'의 결과, '표현' 범주 채점 기준표의 하위 준거와 수준별 기술 내용을 뒷받침하는 예시 자료를 선별하는 작업을 진행하였다. 이 답안 예시 자료는 수준별로 전형적인 답안뿐만 아니라 채점자들이 손쉽게 판단하기 어려운 예외적 속성의 특이한 답안까지를 포함하여 구성한다. 다양한 속성의 답안을 예시 자료로 선별함으로써 채점자 교육의 효과성을 높이고자 하였다. 또한 채점자 교육과정을 제작할 때 참고할 수 있도록 각 예시 답안을 하위 준거 및 수준별 기술에 따라 선별한 이유를 함께 밝힘으로써 선별한 의도가 잘 전달될 수 있도록 하였다.</p>

[그림 II-1] 2022년 논증적 글쓰기 자료 중 내용 및 조직 범주 예시 답안 선별

(2) 논증적 글쓰기 채점 기준표의 정제 및 상세화

이후 채점자 교육과의 관계를 고려하면서, 논증적 글쓰기 채점 기준표를 정제하고 상세화하는 작업을 실시하였다. 예시 답안 선정 작업을 함께한 외부 전문가 2인과 협력하여 채점 기준표의 표현을 채점자가 쉽게 이해 가능한 용어로 정제하고 일부 모호했던 표현들을 명확히 다듬었다.

채점 기준표는 주로 ‘유의 사항’ 수준에서 보완 작업이 많이 진행되었다. 채점 기준표 내 ‘유의 사항’은 주로 하위 준거의 기술만으로는 적용하기 어려운 답안들, 즉 채점자들이 채점 과정에서 어려움에 봉착하는 지점들에 대하여 일종의 지침과 안내를 제공하는 성격을 지닌다. 이러한 부분은 채점자 교육과정에서도 주요한 내용으로 다루어질 수 있는 부분이므로 교육의 효과성을 높이고자 내용과 조직 범주에서는 유의 사항을 중심으로 정제 및 상세화가 진행되었다. 유의 사항 중심의 대략적인 수정 전후 대조표는 <표 II-9>와 같다.

<표 II-9> 논증적 글쓰기 채점 기준표(내용, 조직) 유의 사항 수정 전후 대조표

	수정 전	수정 후
내용 1: 문제 상황 이해	※ 유의 사항 서론을 중심으로 보되, 글의 중간중간에 드러날 수 있음. (1) 불필요한 정보가 포함되었거나, (2) 불필요한 정보로 인해 문제 상황에 대한 설명의 분량이 길어진 경우 높은 점수를 부여하기 어려움.	※ 유의 사항 - 서론을 중심으로 보되, 글의 중간중간에 드러날 수 있음. - 문제 상황 이해에 불필요한 정보가 포함되었거나, 불필요한 정보로 인해 문제 상황에 대한 설명의 분량이 길어진 경우 높은 점수를 부여하기 어려움.
내용 2: 주장	※ 유의 사항 주장은 논제에 대한 입장(글 전체의 주장)으로 한정하여 판단.	※ 유의 사항 주장은 논제에 대한 입장(글 전체의 주장)으로 한정하여 판단. 하위 주장의 경우 ‘이유/근거의 적절성’에서 평가함.
조직 2: 문단 구성 요건	(없음)	※ 유의 사항 형식 문단을 기준으로 채점하되, 형식 문단은 구별되지 않지만 내용 문단이 분명하게 드러나는 경우에는 내용 문단을 기준으로 각 구성 요건을 판단할 수 있음.

2) 표현 범주 채점 준거의 정량화 작업 및 보완

표현 범주 채점 준거를 정량적으로 적용하는 방안을 검토하였다. 표현 범주는 내용이 나 조직 범주보다는 정확성과 규범성 등과 같이 보다 양적으로 측정될 수 있는 속성이 더 많으므로 판단 근거를 비교적 객관적으로 확인할 수 있는 지점에 대하여 정량적인 적용 방안을 탐색하고자 하였다. 이는 채점자 간 일치도를 높이고, 정성적 적용에 투입되는 시간과 노력을 효율적으로 관리하려는 목적으로 점수별 급간에 관한 명확한 기준을 설정하고자 한 것이다. 이러한 방식은 이용상 외(2022)에서 한국어 쓰기 자동 채점을 위해 탐색한 자질 추출 방식이기도 하다.

2022년에 개발한 표현 범주 채점 준거에 대한 정량화 작업 역시 외부 전문가 2인과 협력하여 진행하였다. 이들 전문가도 각각 현직 국어 교사와 국어교육 전공 박사 학위 소지자로서 글쓰기 및 국어 규범 관련 전문성이 있다.

(1) 표현 범주 채점 준거 정량화 작업의 과정

표현 범주 채점 준거 정량화 작업은 표현 범주의 예시 답안 선정 작업과 채점 기준표의 정제 및 상세화 작업과 함께 진행되었다. 먼저, 이전 연도의 과업에서 수집한 글쓰기 참여자 답안 301건 중 채점자 2인 간 점수 평정이 같았던 답안을 상(5점)/중(3점)/하(1-2점) 수준으로 구분하여, 각각의 답안에서 ‘표현 1’과 ‘표현 2’에 해당하는 감점 요소가 얼마나 포함되어 있는지를 분석하였다.

이후, 점수 급간별 답안에 포함되어 있는 평균적인 감점 요소 개수를 중심으로 ‘표현 1’과 ‘표현 2’의 정량화 기준을 일차적으로 설정하였다. 그리고 이렇게 설정한 일차적 기준을 연구진과 함께 오류 개수에 문제가 없는지를 검토하여 최종 확정하였다. 2022년 개발 표현 범주 채점 준거의 점수별 채점 기준은 <표 II-10>과 같고, 위와 같은 과정으로 설정된 표현 범주 채점 준거 1차 수정안 점수별 채점 기준은 <표 II-11>과 같다.

<표 II-10> 2022년 개발 표현 범주 채점 준거의 점수별 채점 기준

점수	표현 1: 문장과 어휘	표현 2: 어문 규범과 관습
5	문장 표현이 자연스럽게 효과적이며, 어휘 사용이 적절함.	어문 규범과 글쓰기 관습을 매우 정확히 준수함.
4	문장 표현이 대체로 자연스럽게 효과적이며, 어휘 사용이 대부분 적절함.	어문 규범과 글쓰기 관습을 대체로 준수함.
3	자연스럽지 않거나 효과적이지 않은 문장 표현과 부적절한 어휘가 일부 사용되었으나, 글을 읽는 데 크게 방해가 되지 않음.	어문 규범과 글쓰기 관습에 일부 문제가 있으나, 중대한 오류가 발견되지 않음.
2	자연스럽지 않거나 효과적이지 않은 문장과 부적절한 어휘가 다수 사용됨.	어문 규범과 글쓰기 관습에 오류가 많아 자연스러운 읽기에 방해가 됨.
1	대부분의 문장이 그 효과를 따지거나 글의 의미를 파악하기 어려울 정도로 부자연스럽고, 어휘 대부분이 글의 맥락에 어울리지 않음.	어문 규범과 글쓰기 관습의 오류가 심각하여 글의 의미를 파악하기가 어려움.

<표 II-11> 2023년 개발 표현 범주 채점 준거 1차 수정안의 점수별 채점 기준

점수	표현 1: 문장과 어휘	표현 2: 어문 규범과 관습
5	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 1개 이하임.	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 1개 이하임.
4	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 2-3개임.	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 2-4개임.
3	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 4-6개임.	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 5-7개임.
2	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 7-9개임.	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 8-10개임.

점수	표현 1: 문장과 어휘	표현 2: 어문 규범과 관습
1	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 10개 이상임.	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 11개 이상임.

또한, 내용 및 조직 범주와 마찬가지로 표현 범주의 채점 기준표의 정제 및 상세화 작업과 함께, 채점자 교육용 자료 제작을 위한 예시 답안을 선정하는 작업도 진행하였다. 이들 작업 역시 표현 범주 채점 준거 정량화 작업을 진행한 외부 전문가 2인과 협력하여 진행하였다.

채점 기준표 정제 및 상세화 작업은 정량화된 점수별 기준에 따라 보완할 점을 유의 사항을 중심으로 내용을 추가하는 것으로 진행하였고, 예시 답안 선정은 점수별로 전형적인 답안과 비전형적인 답안을 추가하여 교육 자료로 활용될 때의 교육적 효용을 높이고자 하였다. 표현 범주 채점 준거의 채점 기준표 정제 및 상세화 결과는 <표 II-12>와 같다.

<표 II-12> 표현 범주 채점 준거 유의 사항의 2022년 개발안과 2023년 1차 수정안의 대조표

	2022년 개발 채점 준거	2023년 1차 수정 채점 준거
표현 1: 문장과 어휘	※ 유의 사항 - 문장의 길이의 적절성에 대한 판단은 길이 자체보다는 의미 전달이 잘 되는지를 기준으로 함.	※ 유의 사항 - 문장의 길이의 적절성에 대한 판단은 길이 자체보다는 의미 전달이 잘 되는지를 기준으로 함. - 한 문장 내에서 ‘의미 전달이 잘 되지 않을 정도로 문장이 긴 경우’와 ‘문맥에 적절하지 않은 어휘 사용’이 나타나는 경우 각각을 오류 수로 판정함.

	2022년 개발 채점 준거	2023년 1차 수정 채점 준거
<p>표현 2: 어문 규범과 관습 - 어문 규범 준수</p>	(없음)	<p>※ 유의 사항</p> <ul style="list-style-type: none"> - ‘명사+명사’의 지나친 붙여쓰기, ‘관형사형+명사(의존 명사 포함)의 붙여쓰기’, 접사 ‘-되-’, ‘-하-’류, 어간+어미의 띄어쓰기 미준수는 감점함. - 본용언+보조 용언의 띄어쓰기는 감점하지 않음. - 인용의 부사격 조사 ‘고/라고’의 혼동은 감점하지 않음.
<p>표현 2: 어문 규범과 관습 - 글쓰기 관습 준수</p>	<p>※ 유의 사항</p> <ul style="list-style-type: none"> - 격식체(-습니다) 사용은 감점하지 않음. - 구어체(-요)는 감점함. - 기타 단정적인 어조 등 관습과 무관한 문체상의 특징에 대해서는 감점하지 않음. 	<p>※ 유의 사항</p> <ul style="list-style-type: none"> - 격식체(-습니다) 사용은 감점하지 않음. - 종결 어미의 불일치나 구어체(-요)는 감점함. - 기타 단정적인 어조 등 관습과 무관한 문체상의 특징에 대해서는 감점하지 않음.

(2) 정량화 이후 질적 기준에 따른 보완 작업 진행

표현 범주 채점 준거의 정량화 작업 이후, 이전 연도 과업에서 수집한 글쓰기 참여자 답안 301건을 3명의 새로운 채점자에게 의뢰하여 변경한 채점 기준을 적용하여 재채점하였다. 이후 이 수집된 글쓰기 자료를 대상으로 코빅버드(KoBigBird) 언어모델을 사후 학습(Fine-tuning)하여 정답 점수를 예측하는 작업을 실시하였다. 측정 방법은 피어슨 상관계수를 활용하였다.

그러나 점수 예측 AI 모델 개발 중, 표현 영역에서 점수 예측에 일반적이지 않은 결과가 나타났다. 특히 ‘표현 2’의 경우 예측 작업에 활용한 글쓰기 자료의 수가 부족한 상황임을 고려하면 추후 AI 모델이 양호하게 예측될 수 있을 것으로 예상되는 수준이지만, ‘표현 1’은 정답 점수와 AI 모델의 예측 점수가 상관도가 없는 것으로 평가되었다. 비록 예측 데이터가 301건으로 적은 수임을 감안하더라도 이는 정량화 기준을 세워 새롭게 마련한 표현 범주 채점 준거가 다른 범주의 채점 준거들과 다소 동떨어져 있음을 방증하는 결과이다.

7월 14일부터 22일 중 4일 동안 진행된 오프라인 채점자 교육에서도 일부 참여자들로

부터 정량화된 표현 범주의 채점 준거 적용이 어렵다는 의견도 제기된 바 있다. 오프라인 채점자 교육에서 표현 범주 채점 준거의 채점자 간 신뢰도가 높지 않았으며, 글의 특성에 따라서 정량화된 기준을 실제 글에 적용하기가 어렵다는 의견이 많았다. 이러한 특성 때문에 채점자 내 신뢰도와 채점자 간 신뢰도가 모두 높지 않은 수준으로 나타나, 당초 의도했던 채점이 개선된 양상은 확인할 수 없었다.

정량화된 기준은 글쓰기 자료의 분량이 제각기 다른 경우가 많아 문제가 된다. 과제에서 글의 분량으로 1,000자를 요구하였으나 수집된 글쓰기 자료의 분량은 750자부터 1,200자까지 다양하게 존재한다. 최소 분량과 최대 분량의 차이가 1/3 수준인데, 이러한 글의 전체 분량 차이나 문장 개수 등을 고려하지 않고 산술적으로 정해진 오류 수만으로 점수를 달리 평정하는 것은 채점 기준 적용의 일관성을 저해하는 요인이 되었다.

정량화된 기준은 동일 오류가 반복되는 사례를 처리하는 경우에도 문제가 된다. 하나의 오류를 개별 오류로 판정할 것인지, 어떤 표기를 오류로 판정할 것인지 등 오류 수는 정량적으로 기준이 설정되었어도 그것을 적용하는 채점자마다 판정 기준이 다른 경우가 많다. 이러한 경우는 오류 수의 양적 기준만으로는 포괄하기 어려운 문제이다.

이와 같은 결과는 채점 기준표의 채점 준거는 정량화 작업이 효용이 크지 않으며 오히려 문제가 발생하는 경우가 많음을 시사한다. 이에 표현 범주는 질적인 측면을 강조한 2022년 채점 준거를 준용하되, 1차 수정안에서 정량화 작업을 통해 마련한 양적인 기준을 보완적으로 사용하기로 결정하였다. 이는 채점자들의 채점 기준 적용에 요구되는 피로도를 줄이고 표현 범주 준거가 전체 글과의 상관성이 보다 높아질 수 있도록 한 것이다.

구체적으로, ‘표현 1’에 해당하는 문장과 어휘 준거는 정량화 작업 이전의 기준을 준용하되 개수 기준을 보조적으로 활용하는 것으로 수정하였다. 특히 ‘상’ 수준에 해당하는 점수를 4점으로, ‘하’ 수준에 해당하는 점수를 2점으로 설정하고, ‘상’ 수준에서 문장과 어휘 사용이 논증에 효과적인 경우 5점으로, ‘하’ 수준에서 문장과 어휘 사용이 의미를 파악하는 데 크게 방해가 되면 1점으로 판정할 수 있도록 기술하였다.

‘표현 2’에 해당하는 어문 규범과 관습 또한 정량화 작업 이전의 기준을 준용하되 내부적으로는 ‘어문 규범 준수’와 ‘글쓰기 관습 준수’를 구분하여 각각의 경우에서 최대 2점씩 감점할 수 있는 여지를 두는 식으로 수정하였다. ‘어문 규범’과 ‘글쓰기 관습’을 모두 정확하게 지키는 경우 5점을 부여하되, 어문 규범의 위반 정도에 따라 최대 2점을 감점할 수 있고, 글쓰기 관습 미준수 유형에 따라서도 최대 2점을 감점할 수 있도록 기술하였다. 이를 표로 정리하면 <표 II-13>과 같다.

<표 II-13> 정량화 이후 보완한 표현 범주 채점 기준표

		2023년 1차 수정안	2023년 2차 수정안
표현 1: 문장과 어휘	5	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 1개 이하임.	문장과 어휘 사용이 자연스럽게 논증에 효과적임.
	4	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 2-3개임.	문장이 대체로 자연스럽게 어휘 사용이 대부분 적절하나, 논증에 효과적으로 보이지 않음. *글의 길이에 따라 2-3개 어색한 문장과 어휘가 있을 수 있음.
	3	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 4-6개임.	자연스럽지 않은 문장과 부적절한 어휘가 사용되었으나, 문장의 의미를 파악하는데 크게 방해가 되지 않음. *한 단락에 1-2개의 오류가 있을 수 있음.
	2	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 7-9개임.	자연스럽지 않은 문장과 부적절한 어휘가 사용되어 문장의 의미를 파악하는데 방해가 됨. *한 단락에 2개 이상의 심각한 오류가 있을 수 있음.
	1	‘문장 표현의 자연스러움’과 ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’ 기준에 따른 오류 수가 10개 이상임.	문장 대부분이 부자연스럽고, 어휘 대부분이 글의 맥락에 어울리지 않아 문장의 의미를 파악하는데 크게 방해가 됨.
표현 2: 어문 규범과 관습	5	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 1개 이하임.	어문 규범과 글쓰기 관습을 매우 정확히 준수함.
	4	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 2-4개임.	어문 규범 위반이 가끔 나타나거나 한 가지 유형의 글쓰기 관습 위반이 나타남.

		2023년 1차 수정안	2023년 2차 수정안
			* 어문 규범과 글쓰기 관습, 둘 중 하나만 문제가 있음.
	3	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 5-7개임.	어문 규범 위반이 가끔 나타나며, 한 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범은 정확히 준수하였으나 두 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범 위반이 빈번하게 나타나지만 글쓰기 관습은 정확히 준수하였음.
	2	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 8-10개임.	어문 규범 위반이 빈번하게 나타나며, 한 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범 위반이 가끔 나타나며, 두 가지 유형의 글쓰기 관습 위반이 나타남.
	1	‘어문 규범 준수’와 ‘글쓰기 관습 준수’ 기준에 따른 오류 수가 11개 이상임.	어문 규범 위반이 빈번하게 나타나며, 두 가지 유형의 글쓰기 관습 위반이 나타남.

이렇게 보완한 표현 범주 채점 준거의 2차 수정안을 적용하여 이전 연도의 과업에서 수집한 글쓰기 참여자 답안 301건을 2023년 11월에 재채점하였다. 2차 수정안을 적용하는 재채점자로는 올해 과업의 채점에 참여한 채점자 중에서 채점 특성이 안정적인 채점자를 선정하였다. 글쓰기 자료당 한 명의 채점자가 단독 채점을 진행하였으며, 문항 유형 별로는 2명의 채점자를 선정하여 총 6명의 채점자가 재채점에 참여하였다.

3) 전문가 의견 반영

중간 보고회(2023년 8월 말) 이후 국립국어원에서 진단 지표 고도화와 관련하여 외부 전문가 3인으로부터 의견을 수렴하고, 그 결과를 연구진에 전달하였다. <표 II-14>-<표 II-21>은 여기에서 제시된 전문가 의견을 일정한 범주에 따라 나누어 살피고, 그에 대한 연구진의 의견을 정리한 것이다. 전문가 의견을 검토하면서 채점 준거를 보다 정교하게 보완할 수 있었다.

<표 II-14> 문항 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
‘찬반형 정책 논제, 발문(논제 및 도입 배경 설명), 과제 요건(서, 본, 결을 갖춘 1,000자 내외의 글), 과제 환경(자료 미참조, 온라인 글쓰기, 90분 이내 작성)’이라는 단일 유형의 평가 문항이 연령별, 국적별, 학력별 국어 글쓰기 능력을 변별하는 데 두루 쓰일 수 있는가에 대한 재고가 필요함.	본 과제는 글쓰기 자동 채점을 위한 대규모 학습 데이터 구축을 목표로 함. 본 과제에서는 단일 유형의 평가 문항을 주제만 달리하여 개발함으로써 자동 채점 학습 자료의 효율성을 높이고자 하였음. 주로, 대학생의 논증적 글쓰기를 대상으로 하였으며, 추후 전 국민을 대상으로 한 다양한 종류의 글쓰기 유형을 포함함으로써 문제를 해결할 수 있음.
유사 문항으로써 작문 평가를 지속하여 수험자가 문항 유형에 대해 선지식을 갖게 될 경우, 처음 응시한 수험자와 2회, 3회 응시한 수험자의 작문 능력이 타당하게 변별될 것인지 궁금함. 수험자 유형별 쓰기 문항을 개발할 필요가 있지 않은가?	고부담 시험 체제를 전제로 하지 않고 자신의 글쓰기 능력 점검을 목적으로 하기 때문에 문제 되지 않을 것으로 봄.

<표 II-15> 채점 기준표 전반 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
채점 기준표가 비전공자가 적용하는 것까지를 염두에 두는 것이라면, 전반적으로 단순화할 필요가 있어 보임.	채점 기준표에 대한 요구가 단순화와 상세화라는 두 가지 모순되는 측면이 있음. 채점 기준표의 사용 방식을 고려하여 적정 수준의 명료화가 필요할 것으로 보임.
대학, 기업, 방송사 등의 수요 조사 결과 바탕으로 논증적 글쓰기 능력 채점 준거를 개발했는데, 이를 ‘국민’의 글쓰기 능력 진단	본 과제의 논증적 글쓰기 채점 기준표는 논증적 글쓰기 채점을 목적으로 개발되었음. 다양한 글의 유형과 참여자를 대상으로 한

전문가 검토 의견	연구진 의견
<p>지표로 일반화하는 것은 문제임. 차라리 선행 연구를 통해 조사된 여러 능력 중 ‘비판적 사고 및 논증’의 고등 사고 역량 측정만을 목표로 하고서 고등 교육과정 입문을 위한 ‘학술적 글쓰기 능력 진단 지표’로 평가 문항 및 준거 내용을 초점화할 필요가 있음.</p>	<p>일반화된 글쓰기 능력 진단 지표를 의미하는 것은 아님.</p>
<p>준거 간 난도(점수를 얻기 쉬운 준거와 어려운 준거)가 달라 보임. 이에 대한 정보를 유의 사항이나 참고 사항 정도로 제공해 주어야 할 필요는 없는가?</p>	<p>본 과제의 채점 방식은 분석적 채점이므로 별도로 채점할 때 난이도는 문제가 되지 않음. 준거에 따른 개별적 판단은 분석적 채점의 기본 전제이므로 별도의 판단이 필요하지 않음.</p>
<p>‘내용 5 - 다른 입장에 대한 고려’의 유의 사항처럼 어떠한 경우에 몇 점 부여와 같은 전문가적 판단과 결정이 채점자 간 주관적 편향을 줄이는 데 기여할 것 같음. 다른 채점 기준에도 모두 이러한 방식이 필요하지 않은가?</p>	<p>어떤 것에 점수를 부여하고, 어떤 것에 점수를 부여하지 않았느냐에 대한 선택의 근거를 제공할 필요는 있겠음. 다른 채점 기준에 대해서는 일관성 있게 점수를 부여하는 데 문제가 없었음.</p>
<p>채점 기준표에 사용된 ‘적절하다, 충분하다, 자연스럽다’ 같은 형용사 표현들도 어느 정도 정량적인 수치로서 보완적인 설명이 필요하지 않은가?</p>	<p>정량적인 정보 제공은 이미 일부 이루어지고 있음. ‘충분하다’는 정량적으로 제공하고 있음. ‘자연스러움’도 어느 정도 정량화되어 있음. ‘적절하다’는 정성적 표현이므로 정량적으로 제공하기에 어려움.</p>
<p>숙달된 채점자의 정성적 채점 준거를 정량적 데이터로 환산할 채점 등급 설명이 필요하지 않은가?</p>	<p>표현 준거의 경우, 정량적 데이터로 수준을 기술한 바 있으나, 채점자들이 채점을 하는 과정에 일관성과 효율성 측면에서 문제점이 드러남. 정성적 채점 준거를 정량적 채점 준거로 일부 보완하는 방식으로 기술하였음. 정량적 준거가 인간 채점자의 채점에서 가장 효율적이지는 않고, 비효율을 초래하는 경우도 많음.</p>

<표 II-16> 채점 준거의 위계 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
<p>‘(상위) 준거’가 두 층으로 구성되어 있는데, 이를 구별해서 지칭하는 것이 좋겠다. (예: ‘내용’과 그 아래 ‘문제 상황 이해’가 모두 ‘준거’로 묶여 있음.)</p>	<p>세 개 ‘영역’과 아홉 개 ‘준거’로 구분하여 명명. 수용함.</p>
<p>이 준거가 ‘절대적 준거’인지 ‘상대적 준거’인지 결정되어야 함. 절대적 준거로 작동하기 위해서는 채점자의 엄격성이 일관성이 있어야 하며, 따라서 채점의 기준점(anchor point)이 동일하게 나타나야 함. 이를 위한 여러 가지 장치가 필요해 보임.</p>	<p>척도는 절대적인 것으로 구안한 것이고, 그것의 일관된 적용을 위하여 채점 정확성을 높이기 위하여 약 30시간 이상의 채점자 교육을 실시하여 신뢰도를 높이는 체계를 활용하였음.</p>
<p>‘and’로 연결된 하위 준거 항목의 위계가 균등하지 않아 보임. 정량적 제한이 없으면 주관적 해석이 많이 개입되지 않는가? (예: ‘주장의 적절성과 명료성’에서 주장은 적절한데 덜 명료한 경우 등 각 요소별로 경우의 수가 많아 보임.)</p>	<p>주장의 적절성은 논제와 무관한 것을 주장하면 안 되는 것이 핵심임. 주장의 명료성에 대한 용어 기술은 조금 더 명확히 해야 할 것 같음. 일관되지 않으면 명료하지 않은 것은 맞는데, 명료성이 적합한 용어인지는 모르겠음. 그리고 적절성과 명료성을 조합해서 경우를 나눠주는 것이 필요. ‘적절성’과 ‘명료성’으로 다섯 가지의 레벨을 만들어야 할 것.</p>

<표 II-17> 하위 준거 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
<p>하위 준거의 개수 또한 (상위) 준거마다 서로 달라 체계성이 떨어져 보임.</p>	<p>쓰기 평가는 일반적으로 ‘내용, 조직, 표현’의 세 범주로 나뉘며, 모어 사용 필자 대상의 쓰기 평가에서는 내용이, 제2언어 사용 필자 대상의 쓰기 평가에서는 표현 범주를 상세화하여 이에 가중치를 두고 평가하는 것이 보편적임.</p>
<p><내용-문제 상황 이해> : ‘문제 상황 이해’라는 준거는 ‘이해’로 표현되어 있어 ‘표현’에 해당하는 글쓰기와는 다소 거리가 있어 보임.</p>	<p>‘문제 상황 제시’로 수정.</p>

전문가 검토 의견	연구진 의견
<p>‘내용 1 - 문제 상황 이해’에서 ‘문제 상황에 대한 설명의 정보성’이 보편적인 글쓰기 과정에서 필수적인 요소인지 점검할 필요가 있음. 가령, 글쓰기 참여자가 해당 쓰기 평가를 평가자와의 ‘대화’로 인식하여 서로 문제 상황이 충실히 공유된 것으로 간주한다면, 문제 상황을 재진술할 필요를 느끼지 않고 본론으로 들어갈 가능성도 있지 않은가?</p>	<p>현재의 과제에서는 대화가 아닌 ‘한 편의 글’을 작성하도록 안내하였으며, 과제를 ‘대화’로 인식하여 완결된 글을 작성하지 않는 경우 글의 일반적인 관습을 충실히 따른 것으로 보기 어려움. ‘대화’와 같이 상호작용적인 상황을 가정하더라도 면대면 상호작용이 아닌 상황에서 충분한 맥락을 공유하지 않은 채 함축에 기대어 의사소통을 하는 것은 정당화될 수 없음.</p>
<p>‘내용 1 - 문제 상황 이해’의 표준적 척도 사용을 위해 설명의 적정 분량을 제시하는 방안 검토. (예: 혹은 ‘문제 상황 이해에 대한 적정 설명 분량은 전체 글의 30% 미만임’.)</p>	<p>글의 분량이 최소 750자에서 최대 1,200자로 다양하므로 분량의 비율을 제시하기에는 난점이 있음. 또한 전체 글에서 문제 상황과 관련된 부분을 적절하게 드러내는 것도 글쓰기 능력의 일부로 보았음.</p>
<p><조직>의 하위 준거에서는 ‘구성’이라는 용어(예: 글 전체 구성, 문단 구성 요건)를 사용하고 있는데, 상위-하위의 개념인지 의문.</p>	<p>‘글 전체 조직, 문단 내 조직’으로 수정.</p>
<p><조직 - 글 전체의 구성>에서 ‘문단의 유기적 배열’은 내용과의 관계가 잘 드러나지 않음.</p>	<p>일반적으로 ‘내용’은 ‘생성된 아이디어(의 질)’를, ‘조직’은 ‘생성된 아이디어의 배열(의 질)’을 의미함. 개념적으로는 ‘내용, 조직, 표현’이 모두 글의 질을 평가하는 하위 요소들로 수렴되는 것이 맞으며, 채점 과정에서 이를 측정학적으로 타당하게 분리하여 수량화하는 것임.</p>
<p>‘표현 1’의 어휘 사용과 관련하여, 문맥 적합도 외에 ‘어휘의 수준’에 대해서도 평가가 가능하지 않는가? 인공지능 기반의 평가를 도입한다면 어휘 수준이야말로 자동 채점이 가능한 요소 중 하나이지 않은가?</p>	<p>어휘의 절대적인 수준 자체가 중요한 것이 아니라, 자신이 쓴 글 내에서 맥락에 비추어 어휘 사용이 적절하고 효과적인지에 초점을 맞추는 것이 타당함.</p>
<p>‘표현 2 - 어문 규범 준수’에서 ‘거의 나타나지 않는 경우’, ‘가끔 나타나는 경우’, ‘빈번하게 나타나는 경우’는 다소 막연한 구분으로, 인간 채점이든 기계 채점이든 전체 글 분량 대비(예: 어절 단위) 몇 회 또는 몇 %와 같은 형태로 수치화하는 것이 채점자 간 신뢰도를 높이는 방안이 아닌가?</p>	<p>글의 분량이 최소 750자에서 최대 1,200자로 다양하므로 일률적으로 횟수나 비율을 산정하기에는 난점이 있음. 필요시 협의를 통해 유의 사항에 추가하는 방식으로 차후 해결.</p>

<표 II-18> 예시문 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
<p>채점자의 일관된 채점을 위하여 준거별로 1-5점 수준을 전형적으로 보여주는 예시문을 제시할 필요가 있어 보임.</p>	<p>국민에게 수준에 대해 안내하는 목적으로 개발하는 것이라면 전형적 자료 선별이 타당함. 2023년 수집 글쓰기 자료 채점을 마친 이후 예시문을 준비할 예정. 다만 전형적인 예시문이 채점 단계에서 활용될 시 예외적인 글쓰기 자료 채점에 편향된 판단을 줄 가능성을 검토할 필요가 있음.</p>
<p>준거별 점수의 예시문을 제시함으로써 각 준거가 의도하는 글의 모습을 국민에게 안내할 필요가 있음.</p>	

<표 II-19> 채점자 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
<p>채점자 입장에서 채점하기 어려운 준거와 쉬운 준거가 있을 텐데, 이를 밝혀 유의 사항에 포함하거나 채점자 교육 단계에서 활용할 필요가 있을 듯함.</p>	<p>2022년 채점 및 2023년 상반기 채점 단계에서 채점자 교육 및 채점 실행 과정에서 문의가 들어온 내용을 중심으로 유의 사항을 단계적으로 추가해 왔음.</p>
<p>채점자 집단의 구성이 이질적인 경우 채점자 간 신뢰도가 높지 않을 수 있음. 교수, 한국어 교사, 중등학교 교사는 글쓰기 참여자에 대한 각기 다른 표상이 있기 때문. 채점자 간 신뢰도를 높이기 위해서는 연령대별 평가 문항과 채점 준거를 차별적으로 개발할 필요가 있지 않은가?</p>	<p>본 채점 준거는 일상적 생활에서 글쓰기 능력을 필요로 하는 성인을 대상으로 개발하였음. 다른 연령대를 고려한 문항은 별도의 개발이 필요함. 이에 대학생 이상 성인 대상 글쓰기 교육 및 평가 경험이 있는 대학의 글쓰기 교강사를 중심으로 채점자 교육을 수행함.</p>

<표 II-20> 정책 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
<p>이 준거를 절대적 준거로 사용한다면 지역적 편차, 가령 지역 격차 및 도농 격차, 학력에 따른 격차, 성별에 따른 격차, 소득 수준에 따른 격차 등이 발견될 것이 자명하고, 이에 따른 사회적 갈등을 어떻게 관리하고 대처할 것인지, 그러한 격차는 어떠한 방법으로 해소할 것인지에 대한 방안도 마련해 두어야 할 것으로 보임.</p>	<p>본 채점 준거는 대학생이나 직장인과 같이 학업이나 업무를 위해 글쓰기 능력을 필요로 하는 성인을 대상으로 활용될 것으로 예상됨. 향후 활용 대상이 확대될 경우 실제 격차의 내용과 정도를 확인하여, 교육적 지원 등 대응 방안을 수립할 필요가 있음.</p>

<표 II-21> 인공지능 학습 관련 전문가 검토 의견 및 연구진 의견

전문가 검토 의견	연구진 의견
<p>자동 평가를 위한 채점 데이터를 모으기 위해서는 국내외 자동 채점 채점 준거를 수집하고 이를 한국어 작문 평가 기준으로 정량화할 필요가 있는데 이 연구의 선행 연구 분석은 1981년부터 2000년대까지 쓰기 빅데이터 분석 이전의 인간 채점자를 위한 채점 준거 정리에 그치고 있음. 이러한 인간 중심의 정성적 평가 기준을 인공지능이 학습하기 위해서는 유사 문항으로 작성된 거대 규모의 쓰기 결과물이 필요하므로 채점자 간 신뢰도를 높일 정량적 채점 준거로의 수정이 필요함.</p>	<p>인간 채점자를 활용한 채점 자료 구축에서 정량적 채점의 방식이 채점의 신뢰도를 높일 수 있는 실제적인 방안이 될 수 있는지에 대한 학술적 근거가 부족함. 정량화는 채점 준거를 구성하는 모든 세부 요소에 대한 구체적인 기준을 수립하고 각 세부 기준에 대한 채점자 간 지식이 동일하고, 글의 분량이나 특성이 유사하다는 전제가 필요함. 2023년 상반기에 실제 표현 영역을 정량화하여 채점을 시행하였을 때, 신뢰도가 높아진 증거를 확인할 수 없었음.</p>

이상의 의견을 반영하여 최종적으로 확정된 논증적 글쓰기 채점 기준표는 <표 II-22>와 같다. 이 연구에서 채점 매뉴얼을 개발할 때에는 이 기준표를 따랐다.

<표 II-22> 논증적 글쓰기 채점 기준표(최종)

영역 / 준거	내 용					조 직		표 현	
	내용 1 문제 상황 제시	내용 2 주장	내용 3 이유/근거의 적절성	내용 4 이유/근거의 충분성	내용 5 다른 입장에 대한 고려	조직 1 글 전체 조직	조직 2 문단 내 조직	표현 1 문장과 어휘	표현 2 어문 규범과 관습
하위 준거	<p>▶ 문제 상황에 대한 제시의 적절성 : 문제 상황에 대한 제시가 <u>논제와 밀접하게</u> 관련되는가?</p> <p>▶ 문제 상황에 대한 설명의 정보성 : 문제 상황을 설명하는 데 <u>필요한 만큼의</u> 정보를 제공하는가?</p> <p>※ 유의 사항 - 서론을 중심으로 보되, 글의 중간중간에 드러날 수 있음. - 문제 상황 이해에 불필요한 정보가 포함되었거나, 불필요한 정보로 인해 문제 상황에 대한 설명의 분량이 길어진 경우 높은 점수를 부여하기 어려움.</p>	<p>▶ 주장의 적절성 : 주장이 논제에 부합하는가?</p> <p>▶ 주장의 명료성 : 주장이 글 전체에 일관되고 뚜렷한가?</p> <p>※ 유의 사항 주장은 논제에 대한 입장(글 전체의 주장)으로 한정하여 판단. 하위 주장의 경우 '이유/근거의 적절성'에서 평가함.</p>	<p>▶ 주장-이유/근거 연결(추론)의 적절성 : 주장 또는 하위 주장 (=주장의 이유)을 뒷받침하는 이유나 근거가 적절한가?</p>	<p>▶ 이유나 근거의 충분성 : 주장이나 하위 주장(=주장의 이유)을 뒷받침하는 적절한 이유나 근거가 충분한가?</p> <p>▶ 이유나 근거의 전개 수준 : 이유나 근거의 아이디어를 깊이 있게 전개하는가?</p> <p>※ 유의 사항 한 단락으로 구성된 경우에도 전체 주장을 뒷받침하는 이유가 2개 이상일 수 있음.</p>	<p>▶ 접근의 다양성 : 다른 입장에 대한 접근이 이루어졌는가?</p> <p>▶ 입장 고려의 수준 : 다른 입장에 해당하는 아이디어를 깊이 있게 전개하는가? (예: 비교, 논박, 종합 등)</p> <p>※ 유의 사항 - 다른 입장이 존재할 수 있음을 구체적인 정보 없이 형식적으로만(예: 다르게 생각하는 사람도 존재할 수 있으나) 드러낸 경우 2점 부여함. - 다른 입장을 뒷받침하는 구체적인 정보가 포함된 경우 3점 이상 부여함(하위 단락 안에 반론의 형식으로 포함될 수도 있음).</p>	<p>▶ 문단 구분의 적절성 : 서-본-결의 구조를 갖추고 있는가? (형식 문단과 내용 문단 모두 포함)</p> <p>▶ 문단 배열의 유기성 : 문단 배열이 유기적이고 자연스러운가?</p> <p>※ 유의 사항 글 전체의 흐름이 대체로 자연스러우나 형식 문단이 적절히 나뉘지 않은 경우 3점 부여함.</p>	<p>▶ 완결성 : 뒷받침 문장을 사용하여 문단의 중심 문장을 뒷받침하였는가?</p> <p>▶ 통일성 : 문단별로 하나의 중심 생각(소주제)이 전개되는가?</p> <p>▶ 일관성 : 문단을 이루는 여러 문장들이 서로 긴밀한 관계로 연결되는가? (예: 시간적, 논리적 순서)</p> <p>※ 유의 사항 형식 문단을 기준으로 채점하되, 형식 문단은 구별되지 않지만 내용 문단이 분명하게 드러나는 경우에는 내용 문단을 기준으로 각 구성 요건을 판단할 수 있음.</p>	<p>▶ 문장 표현의 자연스러움 : 문장 길이가 적절하고 그 표현이 자연스러운가? (예: 주술호응)</p> <p>▶ 문장 표현의 효과성 : 문장의 의미가 명료하고 그 표현이 효과적인가?</p> <p>▶ 어휘 사용의 적절성 : 어휘 사용이 문맥에 맞는가?</p> <p>※ 유의 사항 문장의 길이의 적절성에 대한 판단은 길이 자체보다는 의미 전달이 잘 되는지를 기준으로 함.</p>	<p>▶ 어문 규범 준수 : 맞춤법, 띄어쓰기 등을 준수하였는가? (오타자 여부 포함)</p> <p>※ 유의 사항 - 어문 규범 위반이 거의 나타나지 않는 경우 감점 없음. - 어문 규범 위반이 가끔 나타나는 경우 -1점. - 어문 규범 위반이 빈번하게 나타나는 경우 -2점.</p> <p>▶ 글쓰기 관습 준수 : 장르 관습에 맞는 문어적 표현을 사용하는가?</p> <p>※ 유의 사항 - 격식체(-습니다) 사용은 용인함. - 구어체(-요)를 사용하거나 종결어미를 일관되게 사용하지 않은 경우 -1점. - 자기 소개, 인사말 등 장르 관습에 맞지 않는 표현을 사용한 경우 -1점. - 기타 단정적인 어조 등 관습과 무관한 문체상의 특징에 대해서는 감점하지 않음.</p>

	내용 1 문제 상황 제시	내용 2 주장	내용 3 이유/근거의 적절성	내용 4 이유/근거의 충분성	내용 5 다른 입장에 대한 고려	조직 1 글 전체 구성	조직 2 문단 내 구성	표현 1 문장과 어휘	표현 2 어문 규범과 관습
5	문제 상황에 대한 이해가 논제에 밀접하게 매우 효과적으로 제시되며, 반드시 필요한 정보로 구성됨.	주장이 논제에 부합하며, 글 전체에 걸쳐 매우 일관되고 뚜렷함.	이유나 근거가 매우 적절하여 설득력이 높음.	주장에 대한 적절한 하위 주장(=주장의 이유)을 2가지 이상 제시하고, 각각의 하위 주장에 대한 적절한 이유나 근거를 깊이 있게 제시함.	다른 입장을 고려하고, 그 내용을 매우 깊이 있게 전개함.	형식 문단과 내용 문단 모두에서 서-본-결의 구조가 매우 적절하게 구분되고, 문단 배열이 매우 유기적임.	모든 문단에서 완결성, 통일성, 일관성을 갖춘.	문장과 어휘 사용이 자연스럽고 논증에 효과적임.	어문 규범과 글쓰기 관습을 매우 정확히 준수함.
4	문제 상황에 대한 이해가 대체로 논제에 밀접하게 효과적으로 제시되며, 대부분 필요한 정보들로 구성됨.	주장이 논제에 부합하며, 글 전체에 걸쳐 대체로 일관되고 뚜렷함.	이유나 근거가 대체로 적절하여 다소 설득력이 있음.	주장에 대한 적절한 하위 주장(=주장의 이유)을 2가지 이상 제시하였고, 각각의 하위 주장에 대한 적절한 이유나 근거를 제시함.	다른 입장을 고려하고, 그 내용을 대체로 깊이 있게 전개함.	형식 문단과 내용 문단 모두에서 서-본-결의 구조가 대체로 적절하게 구분되고, 문단 배열이 대체로 유기적임.	대부분의 문단에서 완결성, 통일성, 일관성을 갖춘.	문장이 대체로 자연스럽고 어휘 사용이 대부분 적절하나, 논증에 효과적으로 보이지 않음. * 글의 길이에 따라 2-3개 어색한 문장과 어휘가 있을 수 있음	어문 규범 위반이 가끔 나타나거나 한 가지 유형의 글쓰기 관습 위반이 나타남. * 어문 규범과 글쓰기 관습, 둘 중 하나만 문제가 있음.
3	문제 상황에 대한 이해가 논제와 관련되어 있으나, 정보성이 다소 떨어짐.	주장이 논제에 대체로 부합하나, 일관성이 다소 부족하거나 뚜렷하지 않음.	이유나 근거 가운데 일부가 적절하지 않으나, 대체로 수용할 만함.	주장에 대한 적절한 하위 주장(=주장의 이유)을 1가지 제시하고, 그에 대한 적절한 이유나 근거를 비교적 깊이 있게 제시함.	다른 입장을 고려하였으나, 그 내용의 깊이가 다소 떨어짐.	형식 문단 또는 내용 문단의 구분과 배열이 일부 자연스럽지 않으나, 글을 읽는 데 크게 방해가 되지 않음.	일부 문단에서 완결성, 통일성, 일관성을 갖춘.	자연스럽지 않은 문장과 부적절한 어휘가 사용되었으나, 문장의 의미를 파악하는 데 크게 방해가 되지 않음. * 한 단락에 1-2개의 오류가 있을 수 있음.	어문 규범 위반이 가끔 나타나며, 한 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범은 정확히 준수하였으나 두 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범 위반이 빈번하게 나타나지만 글쓰기 관습은 정확히 준수하였음.
2	문제 상황에 대한 이해가 논제와 관련성이 떨어지는 정보들로 구성됨.	주장이 논제에서 다소 어긋남. 또는 주장이 논제와 관련은 있지만 일관성이나 명료성이 매우 부족함.	이유나 근거가 적절하지 않은 경우가 많아 설득력이 떨어짐.	주장에 대한 적절한 하위 주장(=주장의 이유)을 1가지 제시하였으나, 그에 대한 적절한 이유나 근거를 깊이 있게 제시하지 못함.	다른 입장이 존재할 수 있음에 대한 인식이 간접적으로 드러남. (예: 주장의 강도를 낮추거나 단정적이지 않은 표현)	글의 체계가 다소 부족하여 완성도가 낮고, 문단 배열 가운데 부자연스러운 부분이 많음.	문단의 완결성, 통일성, 일관성이 전반적으로 부족하여 완성도가 낮음.	자연스럽지 않은 문장과 부적절한 어휘가 사용되어 문장의 의미를 파악하는 데 방해가 됨. * 한 단락에 2개 이상의 심각한 오류가 있을 수 있음.	어문 규범 위반이 빈번하게 나타나며, 한 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범 위반이 가끔 나타나며, 두 가지 유형의 글쓰기 관습 위반이 나타남.
1	문제 상황에 대해 설명하지 않았거나 이해했지만 매우 빈약함.	주장이 논제에 부합하지 않음. 또는 주장이 논제와 관련되지 않는 것이거나 일관성과 명료성이 매우 부족함.	이유나 근거가 대부분 적절하지 않아 설득력이 매우 낮음.	주장에 대한 하위 주장(=주장의 이유)이 모두 적절하지 않음.	단일한 입장에 대해서만 단정적으로 언급됨.	내용 문단이 거의 구분되지 않아, 글의 체계가 없음.	대부분의 문단에서 구성요건을 갖추지 못함.	문장 대부분이 부자연스럽고, 어휘 대부분이 글의 맥락에 어울리지 않아 문장의 의미를 파악하는 데 크게 방해가 됨.	어문 규범 위반이 빈번하게 나타나며, 두 가지 유형의 글쓰기 관습 위반이 나타남.

4) 채점 기준표의 타당성 확보를 위한 외부 전문가 재채점

국립국어원은 중간 보고회(2023.8.29)에서 논의된 채점 준거 설정에 대한 근거를 재검토할 것과 타당성을 확보하기 위하여 외부 전문가를 통해 수집한 글쓰기 자료 답안의 일부를 재채점하도록 결정하였다. 이를 위하여 본 과업에 참여하지 않은 외부 글쓰기 및 작문 교육 관련 분야의 전문가 3인을 선정하였으며, 이들 전문가에게 채점용 글쓰기 자료 50편을 선정하여 전달하였다. 선정한 글쓰기 자료 답안은 기재점된 글쓰기 자료 중에서 점수대(9점-45점)를 다섯 구간으로 나누어 구간별로 10편씩을 선정한 것이다. 이후 전문가에게 인상 비평에 따른 총체적 채점을 1-5점 수준으로 평정한 이후, 내용, 조직, 표현 준거별로 1-5점 수준에서 평정할 것을 요청하였다. 준거별 채점 시에는 총체적 채점 점수를 고려하지 않고 독립적인 과제로 수행할 것을 당부하였다. 전문가의 준거별 채점은 본 과업에서 개발한 논증적 글쓰기 채점 준거 기준표를 적용하지 않고 전문가의 판단에 따른 평정 과정을 거쳤다. 끝으로 전문가의 채점 결과와 본 과업의 기재점 결과를 통계적으로 비교하여 유의미한 차이가 있는지를 검정하는 과정을 통해 논증적 글쓰기의 채점 기준표의 타당화 작업을 진행하고자 하였다. 이상의 절차를 표로 나타내면 <표 II-23>과 같다.

<표 II-23> 채점 기준표 타당성 확보를 위한 외부 전문가 재채점 절차

총체적·분석적 채점을 위한 글쓰기 자료 선정	외부 전문가 채점 (총체적 채점)	외부 전문가 채점 (분석적 채점)	전문가 채점 결과와 본 과업 채점 비교
전문가 채점을 위한 점수대별(다섯 구간) 글쓰기 자료 50편 선정	⇒ 전문가의 판단에 따른 평가 요소 및 수준 설정을 적용한 총체적 채점	⇒ 전문가의 판단에 따른 평가 요소 및 수준 설정을 적용한 내용, 조직, 표현 채점	⇒ 통계적 검정 과정을 통한 채점 결과 비교 및 해석

외부 전문가 3인(A, B, C)의 총체적 채점 및 준거별 채점(내용, 조직, 표현) 결과를 본 연구에 참여한 채점자 2인의 9개 준거에 대한 기재점 결과(D)와 비교하였다. 이때, D의 채점 결과를 A, B, C의 채점 결과와 비교하기 위하여, A, B, C의 총체적 채점에 대한 대응값은 D의 9개 준거의 평균값을 사용하였고, A, B, C의 내용, 조직, 표현 준거별 채점 결과의 경우 ‘내용’은 5개 준거의 평균, ‘조직’과 ‘표현’은 각각 2개 준거의 평균을 사

용하였다.

먼저, 기술통계 값은 <표 II-24>와 같다.

<표 II-24> 외부 전문가 3인의 총체적 채점 및 하위 범주별 채점 평균과 표준편차

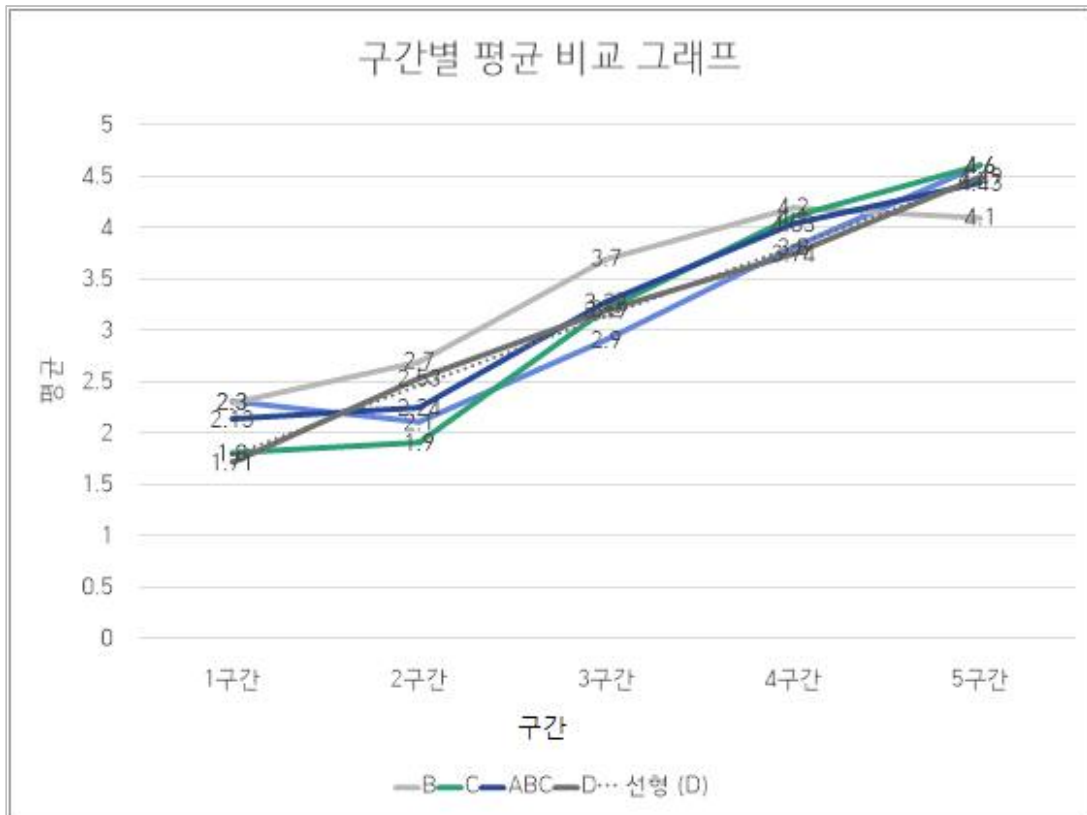
구분	총체적 채점		하위 범주별 채점					
			내용		조직		표현	
	평균	표준 편차	평균	표준 편차	평균	표준 편차	평균	표준 편차
A	3.14	1.43	3.04	1.35	3.32	1.28	3.42	1.26
B	3.40	1.12	3.38	1.14	3.58	0.95	3.60	0.93
C	3.12	1.38	3.20	1.40	3.32	1.28	3.34	1.15
합계	3.22	1.19	3.21	1.18	3.41	1.05	3.45	1.00

외부 전문가 3인의 채점 결과를 분석한 결과, 총체적 채점의 평균은 3.22였으며, 영역별 채점의 평균은 표현(3.45) > 조직(3.41) > 내용(3.21)의 순이었다. 채점자별로는 전문가 B가 전문가 A, C보다 총체적 채점과 영역별 채점에서 모두 점수가 높았다. A, B, C 세 채점자 간 신뢰도를 분석하기 위해 급내상관분석(ICC)을 실시하였다. 급내상관분석을 통해 도출하는 급내상관계수는 두 명 이상의 평가자(채점자) 간 일치도(신뢰도)를 확인하기 위해 구하는 값으로, 평균 측도 급내상관분석(ICC)을 해석한다. 합일된 기준은 없으나 일반적으로 그 값이 0.6 이상이면 양호하다고 볼 수 있으며 0.75 이상이면 매우 좋은 신뢰도로 볼 수 있다. 분석 결과, 평균 측도 급내상관분석(ICC)은 0.884로 매우 양호한 수치로 나타났다.

보다 정밀한 분석을 위하여 5개 점수 구간별로 A, B, C의 총체적 채점의 결과와 이들 셋 간의 평균값을 D의 9개 준거 평균값과 비교하였다. 이를 표로 정리하면 <표 II-25>와 같고, 그래프로 나타내면 [그림 II-2]와 같다.

<표 II-25> 구간별 외부 전문가 3인의 총체적 채점 및 채점 위원 2인 채점 평균과 표준편차

구간 (총점)		1 (9-17.99)	2 (18-24.99)	3 (25-31.99)	4 (32-38.99)	5 (39-45)
A	평균	2.30	2.10	2.90	3.80	4.60
	표준편차	1.00	0.83	1.37	1.33	0.49
B	평균	2.30	2.70	3.70	4.20	4.10
	표준편차	0.90	1.00	0.78	0.75	0.54
C	평균	1.80	1.90	3.20	4.10	4.60
	표준편차	0.60	0.70	0.87	0.94	0.66
ABC	평균	2.13	2.24	3.27	4.03	4.43
	표준편차	0.76	0.60	0.88	0.92	0.34
D	평균	1.71	2.53	3.19	3.74	4.49
	표준편차	0.27	0.20	0.20	0.15	0.14



[그림 II-2] 5개 구간별 총체적 채점(A-B-C) 및 9개 준거 평균(D)

위와 같이 글쓰기 자료 50편의 기재점 결과에 따른 5개 점수 구간별 평균과 표준편차

를 확인하였다. 각 점수 구간별로 A-B-C-ABC의 총체적 채점과 D의 9개 준거 평균값의 경향은 대체로 유사하게 나타났으며, A의 경우 2구간에서 다소 역전 현상이 나타났으나 이상치는 아니었다.

이상의 기술통계치에 더하여, 실제로 외부 전문가의 채점 결과의 평균(ABC)과 본 과업의 채점자 2인의 평균(D)이 통계적으로 유의한 차이가 있는지 차이 검정을 진행하였다. 전체 글쓰기 자료 50건에 대하여 내용, 조직, 표현에서 ABC와 D의 평균 및 표준편차, 평균의 표준오차는 <표 II-26>과 같고, t-검정 결과는 <표 II-27>과 같다.

<표 II-26> 외부 전문가 3인의 채점 평균값(ABC)과 채점 위원 2인 채점 평균값(D)

	집단	N	평균	표준편차	평균의 표준오차
내용	ABC	50	3.21	1.18	.167
	D	50	3.02	.99	.140
조직	ABC	50	3.41	1.06	.149
	D	50	3.13	1.20	.169
표현	ABC	50	3.45	1.00	.142
	D	50	3.42	1.03	.145

<표 II-27> 내용, 조직, 표현의 외부 전문가 3인 평균값(ABC)과 채점 위원 2인 채점 평균값(D)의 차이 검정

		Levene의 등분산 검정		평균의 동일성에 대한 t-검정						
		F	유의 확률	t	자유도	유의 확률 (양쪽)	평균차	차이의 표준오차	차이의 95% 신뢰구간	
									하한	상한
내용	등분산이 가정됨	3.788	.054	.865	98	.389	.189	.218	-.244	.622
조직	등분산이 가정됨	.574	.450	1.227	98	.223	.277	.226	-.171	.725
표현	등분산이 가정됨	.039	.843	.189	98	.850	.038	.203	-.364	.441

t-검정 결과, 내용, 조직, 표현 모든 범주에서 등분산 가정이 성립하였으며, t 통계량의 유의 확률이 0.05보다 높게 나타났다. 이는 외부 전문가의 점수인 ABC와 채점 위원의 점수인 D 간의 내용, 조직, 표현의 점수 차이가 통계적으로 유의하지 않음을 나타낸다.

마지막으로, 구간별 총체적 채점(A, B, C)과 9개 준거 평균(D)의 일치율과 상관을 비교하였다. 결과는 <표 II-28>과 같다.

<표 II-28> 구간별 총체적 채점(A, B, C) 및 9개 준거 평균(D)의 일치율 및 상관 비교

구간 (총점)		1 (9-17)	2 (18-24)	3 (25-31)	4 (32-38)	5 (39-45)
A/D	일치율	0.30	0.44	0.30	0.30	0.60
	상관	0.22	0.40	0.63	0.43	0.34
B/D	일치율	0.50	0.36	0.50	0.40	0.40
	상관	0.49	0.40	0.54	0.12	-0.37
C/D	일치율	0.70	0.41	0.50	0.40	0.40
	상관	0.33	0.20	0.20	0.33	0.19
ABC/D	일치율	0.30	0.57	0.20	0.50	0.50
	상관	0.38	0.60	0.56	0.35	0.09

분석 결과는 구간별 사례 수가 적어 관계의 의미를 일반화하여 해석할 수는 없으나, 5개 점수 구간별 'A-B-C-ABC'와 D 사이의 일치율 및 상관을 분석한 결과, 전반적으로 약한 상관이 있는 것으로 나타났다. B와 D 사이의 5점 구간을 제외하면 어느 정도 약한 상관이 나타나는 것은, 비교 분석을 위한 채점 사례 수가 더 늘어난다면 그 상관의 정도가 더 커질 것으로 예상할 수 있다.

채점 기준표의 타당성 확보를 위한 외부 재채점 결과를 정리하면 다음과 같다. 먼저, 외부 전문가 3인 간의 급내상관계수는 매우 양호한 수치로 나타났다. 이는 전문가 사이에서 전문가의 관점에 따른 평가 요소와 수준이 대체로 비슷하게 설정되어 공유되고 있음을 보여준다. 그리고 <표 II-25>의 기술통계치나 <표 II-27>의 차이 검정 결과는 그러한 외부 전문가의 채점 결과가 본 과업의 채점 위원의 채점 결과와 크게 다르지 않음을 나타낸다. 이는 본 과업에서 설계한 채점 기준표에 대하여 채점자 교육을 받은 채점자의 채점 결과가 글쓰기 전문가의 입장이나 관점과 유사하다는 점을 시사한다. 채점자

교육을 체계적으로 수행한 채점자가 본 과업의 채점 준거 기준표를 적용하여 논증적 글쓰기 자료를 채점한 결과가 글쓰기 관련 전문가의 채점 결과와 유사하다는 점은 본 과업의 논증적 글쓰기 채점 기준표가 신뢰성이 높고 타당하게 설계되었음을 드러낸다.

5) 논증적 글쓰기 채점 기준표에 대한 설명

(1) 내용 1: 문제 상황 제시

‘문제 상황 제시’는 해당 논제를 사회적으로 공론화하여 해결해야 할 필요성으로서의 논의 배경에 대한 설명을 의미한다. 즉, 논제가 제시된 사회문화적 배경에 대해 필자가 일정 수준으로 이해하고 그 이해를 바탕으로 글을 작성하고 있는가에 대하여 판단하는 준거이다.

‘문제 상황 제시’를 적용할 때의 핵심은 논제와 연결된 문제 상황에 대한 정보, 즉 논제와 관련된 배경 정보가 1) 적절하게 그리고 2) 충분히 글에 제시되었는지를 판단하는 것이다. 이들 각각은 ‘문제 상황에 대한 제시의 적절성’, ‘문제 상황에 대한 설명의 정보성’으로 채점 기준표에 제시되었다. 이러한 문제 상황은 주로 글의 도입 부분에 제시되는 경우가 많으나 다른 부분에서도 제시될 수 있다.

이러한 문제 상황에 대한 제시는 필자에 따라 매우 다양한 양상으로 나타날 수 있다. 가령, 로봇세 도입에 관한 문항의 경우 로봇세가 논의가 되기 시작한 배경으로서의 로봇 기술이 발달하고 그것이 상용화된 최근의 상황 등을 제시하면서 글을 시작할 수도 있을 것이고, 로봇세라는 것이 무엇인지에 대한 질문을 제기하면서 그 개념을 간단히 밝히는 것도 가능할 것이다. 핵심은, 이러한 문제를 왜 공론화해야 하는지에 대한 필요성을 얼마만큼 어떻게 드러내는지를 판단하는 것이라고 볼 수 있다.

(2) 내용 2: 주장

‘주장’은 논증을 구성하는 핵심 요소 중 하나로 논증적 글쓰기에서 필자가 강조하는 중핵적인 내용을 의미한다. 이러한 주장에 대한 채점 준거의 하위 요소로서 ‘주장이 논제에 부합하는가’와 ‘주장이 글 전체에 일관되고 뚜렷한가’가 포함된다. 이들 각각은 ‘주장의

적절성'과 '주장의 명료성'으로 채점 기준표에 제시되었다.

'주장의 적절성'은 제시된 문항에서 요구하는 과제 상황에서 벗어난 주장을 하는 경우 배치된다고 판단할 수 있다. 가령, '혐오 표현의 법적 규제'에 대하여 자신의 입장을 논증해야 하는 문항에서 '인종 차별의 문제'를 중심에 두고 논증하거나, '혐오 표현의 판정 기준'에 대하여 논증한다면, 이는 그러한 초점에 대하여 일관된 논지를 전개하였어도 주장이 논제에 부합한다고 보기 어려우므로 '주장의 적절성' 측면에서 높은 점수를 받을 수 없다.

'주장의 명료성'은 하나의 글에서 하나의 주장이 일관되게 제시되어야 명료하다고 판단할 수 있다. 그러나 가령, '로봇세 도입'에 대한 자신의 입장을 논증해야 하는 문항에서 로봇세 도입에 반대하는 로봇 업계의 입장을 제시하다가, 중반부부터 로봇세 도입을 지지하는 과학자들의 입장을 제시하면서 글을 끝맺는 경우, 로봇세 도입에 대하여 필자가 찬성하는지 반대하는지 입장을 명확히 파악하기 어렵다. 이러한 경우에는 '주장의 명료성' 측면에서 높은 점수를 받을 수 없다.

(3) 내용 3: 이유/근거의 적절성

'이유/근거의 적절성'은 글에 나타난 주장이나 하위 주장을 뒷받침하기 위해 제시된 이유나 근거가 질적으로 타당하고 적절한가에 대한 판단을 요구하는 준거이다. 이유는 주로 '~이기 때문이다.'의 형식으로 제시되는 명제로 볼 수 있는데, '~는 공평하지 않기 때문', '~에게 부담이 되기 때문', '~에게만 특혜를 주기 때문', '~기술 발전을 저해하기 때문'과 같은 형식으로 논증적 글에 나타난다. 그리고 이러한 이유를 다시 입증하기 위해 동원하는 외부 자료들을 '근거'라고 한다.

글에서 필자가 주장하는 하나의 대명제를 '주장'으로 본다면, 그 주장은 대체로 각 문단들을 통해 뒷받침되는 구조로 구성된다. '주장-이유'의 관계에서 각 문단들은 여러 개의 '이유'의 성격을 지닌다. 그러한 개별 문단 차원에서 본다면 그러한 이유는 다시 하나의 중심 생각을 지닌 '주장'으로 작용할 수도 있다. 이러한 관점으로 인해 글 전체의 '주장'과 구별하기 위하여 문단의 중심 생각으로서의 '이유'는 '하위 주장'이라는 용어로 설명하기도 한다.

(4) 내용 4: 이유/근거의 충분성

이 채점 준거는 ‘내용 3’의 ‘적절성’과 밀접하게 연관된다. 합리적인 논증에서 적절하고 타당한 ‘이유와 근거’의 세트를 기준으로 할 때, 그 세트가 양적으로도 충분한지를 판단하는 준거이다. 이유와 근거의 질적인 측면을 ‘내용 3’에서 평가하였다면, 그것의 양적인 측면은 ‘내용 4’에서 평가한다. 구체적으로, 주장에 대한 하위 주장을 충분하게 제시하였는지, 그리고 그것의 이유와 근거를 깊이 있게 제시하였는지를 중심으로 평가한다.

본 과업에서는 1,000자 내외 글쓰기라는 글의 분량상의 조건을 고려하여 이유나 근거가 ‘충분하게’ 제시되었다는 기준을 2개로 설정하였다. 2가지 이상의 하위 주장을 제시하고, 각각의 하위 주장에 대한 이유나 근거를 구체적으로 뒷받침하여 깊이 있게 제시하였다면, 그 논증은 일정 수준의 충실성, 풍부성을 갖춘 것으로 볼 수 있다고 판단하였다. 이때, 하나의 이유에 대해 두세 가지 근거가 제시되는 경우나 단일 근거로 뒷받침하는 것 모두 가능하다.

(5) 내용 5: 다른 입장에 대한 고려

‘다른 입장에 대한 고려’는 해당 논제에 대해 필자와 다른 생각을 가진 독자들이 자신의 글을 읽으면서 어떤 질문을 떠올릴 것인가를 예상하고 그 예상을 글에 반영하고 있는지를 평가하는 준거이다. 즉, 상대가 제기할 가능성이 있는 반론들에 대해 다루는 부분과 밀접한 연관이 있다.

이러한 다른 입장에 대한 고려 역시 다양한 방식으로 논증적 글에 제시될 수 있다. 필자가 자신의 입장에 대한 반론을 미리 요약적으로 제시하면서 반영할 수도 있고, 자신의 입장과 반대 입장을 동일선상에서 비교 및 대조하면서 자신의 주장이 더 우위에 있음을 입증하는 방식으로 글을 전개할 수도 있다. 이를 종합적으로 고려하여 반영한 이 준거의 하위 평가 요소로는 ‘접근의 다양성’과 ‘입장 고려의 수준’이 있다.

‘접근의 다양성’은 자신과 다른 입장에 대해서 논의하였는지를 평가하는 요소이다. ‘입장 고려의 수준’은 그러한 다른 입장을 비교하기나 논박하기 등의 다양한 전개 방식을 활용함으로써 깊이 있게 전개하였는지를 평가하는 요소이다. 가령, ‘고령자 운전 면허 제한’에 대하여 자신의 입장이 제한해서는 안 된다는 입장이라고 한다면, 자신과 반대되는 입장인 ‘제한하여야 한다’라는 입장에 대하여 필자 자신에게 제기할 수 있는 질문들을 예

상하여 답하는 식으로 1,000자라는 분량 내에 적절하게 반영하여 진술하였다면, ‘다른 입장에 대한 고려’는 충분히 반영된 것이라고 판단할 수 있다.

이러한 ‘다른 입장에 대한 고려’는 본 과업의 글쓰기 분량상 간소화되어 글에 반영될 가능성이 있다. 글쓰기 참여자가 본인의 입장을 제시하고 그 입장을 뒷받침하기 위한 이유나 근거를 제시하는 것으로도 글의 분량이 초과할 수 있기 때문이다. 그러나 논증은 서로 다른 입장 간의 견해 차이를 전제로 한다는 점에서 대화적인 성격을 띠며(장지혜, 2016), 따라서 다양한 관점에 대한 고려를 토대로 전개되는 논증이 필자의 입장을 강화하며 설득력을 높일 수 있다는 점에서 논증적 글쓰기에서 중요하게 요구되는 지점이다. 한 정적인 분량 내에서도 글쓰기 참여자가 다른 입장을 고려하는 것이 논증에서 중요한 부분임을 알고, 이를 자신이 세운 입장과 연결 지어 제시하는 것 또한 논증적 글쓰기 능력을 평가하는 데 있어 핵심적인 부분 중 하나이므로 ‘내용’의 채점 준거 중 하나로 설정하였다.

(6) 조직 1: 글 전체 조직

‘조직’은 글을 이루는 내용들 간의 연결이 자연스럽고 유기적이며 응집성을 갖추고 있는가를 의미한다. 글의 조직은 거시 조직과 미시 조직으로 대별할 수 있다. 거시 조직은 글 전체를 구성하고 있는 하위 부분들이 유기적으로 잘 연결되어 있는지를 의미하고, 미시 조직은 문단 내의 응집성이나 유기성이 확보되어 있는지를 의미한다.

‘조직 1’에 해당하는 ‘글 전체 조직’의 경우 ‘서론-본론-결론’의 구조가 적절하게 갖추어져 있는지, 그리고 그 문단 간 연결 관계가 적절한지가 핵심적인 판단 기준이 된다. 이때 ‘서론’은 논증을 전개해야 할 배경이나 필요성을 설명하는 역할을 하는 문단이다. ‘본론’은 필자가 특정 입장을 지지하는 이유와 근거를 잘 제시하는 역할을 하는 문단이다. ‘결론’은 앞서 논증한 내용을 정리하며 필자의 주장을 다시 강조하거나 마무리하는 역할을 하는 문단이다.

줄 바꿈을 통해 문단 구분이 되는 ‘형식 문단’과 각 문단 내의 내용적인 완결성과 통일성이 충족된 ‘내용 문단’이 글 전체에 갖추어져야 ‘글 전체 조직’에서 높은 점수를 부여할 수 있다. 다만 그러한 문단이 형식적으로 잘 갖추어지지 않은 경우의 글도 존재할 수 있다. 전반적으로 글의 흐름이 자연스럽더라도 형식 문단의 구분이나 제시가 부적절하다면 평균 수준의 점수를 부여할 수 있다.

(7) 조직 2: 문단 내 조직

‘조직 2’에 해당하는 ‘문단 내 조직’의 경우 하나의 문단을 기준으로 할 때, 그 문단 내의 구조 요소 세 가지를 중심으로 평가하게 되는 준거이다. 이는 ‘완결성’, ‘통일성’, ‘일관성’으로 구성된다.

‘완결성’은 뒷받침 문장을 사용하여 문단의 중심 생각을 뒷받침하였는지를 판단하는 요소이다. 가령, ‘고령 운전자의 면허를 제한해야 한다’라는 입장의 글에서 한 문단의 중심 생각이 ‘고령 운전자의 사고를 줄이기 위해 운전 면허 반납 등 관련 법안을 마련해야 한다’와 같다고 할 때, 그러한 법안을 마련해야 하는 이유나 효과성에 대하여 뒷받침하면 해당 문단은 완결성을 갖춘 문단이라고 할 수 있다.

‘통일성’은 문단별로 하나의 중심 생각이 전개되는지를 기준으로 판단하는 요소이다. 가령, ‘고령 운전자의 면허를 제한해야 한다’라는 입장의 글에서 한 문단의 전반부에서는 ‘개인의 인지 능력 저하 속도가 사람마다 다르다’라고 기술하였으나, 후반부에서 ‘특정 연령이 되면 면허증을 반납하게 해야 한다’라고 기술한다면 두 문장의 내용이 서로 배치되어 통일성을 갖추었다고 보기 어렵다.

‘일관성’은 문단을 이루는 여러 문장들이 서로 긴밀한 관계로 연결되는지를 기준으로 판단하는 요소이다. 가령, ‘고령 운전자의 면허를 제한해야 한다’라는 입장의 글에서 고령 운전자의 사고를 사례로 제시한다고 하였을 때, 해당 사고가 발생한 시간적 흐름이나 인과 관계와 상관없이 흐름을 뒤섞어 제시한다면 그것은 일관성을 갖추었다고 보기 어렵다.

(8) 표현 1: 문장과 어휘

‘표현 1’에 해당하는 ‘문장과 어휘’는 어휘와 문장 수준에서 그것이 적절하고 효과적으로 표현되었는지를 판단하는 준거이다. 이는 크게 세 가지 하위 요소인 ‘문장 표현의 자연스러움’, ‘문장 표현의 효과성’, ‘어휘 사용의 적절성’으로 나뉜다.

‘문장 표현의 자연스러움’은 문장 성분들이 서로 잘 호응하는지를 중점적인 기준으로 삼아 판단하는 요소이다. 가령 주술 호응 관계가 맞지 않아 문장이 어색하게 기술된 경우 효과적인 문장 사용이라고 보기 어렵다.

‘문장 표현의 효과성’은 문장 표현이 명료한지를 중점적인 기준으로 삼아 판단하는 요

소이다. 문장 내에서 접속어가 빈번하게 사용되어서 그 길이가 길어지는 경우, 지나치게 내용이 병렬적으로 열거될 경우 문장의 초점이 불명료하게 되고, 이는 문장의 효과성을 저해하는 지점으로 판단하여 해당 요소에서 높은 점수를 받을 수 없다.

‘어휘 사용의 적절성’은 문맥에 걸맞은 어휘를 사용하였는지를 중점적인 기준으로 삼아 판단하는 요소이다. 전문 분야의 어휘를 필자가 잘못 이해하여 문맥에 어울리지 않게 사용하는 경우나, 한자어들을 병렬적으로 연결하여 만연하게 사용할 때 어색한 어휘 사용으로 볼 수 있으며, 이러한 경우 해당 요소에서 높은 점수를 받을 수 없다.

(9) 표현 2: 어문 규범과 관습

‘표현 2’에 해당하는 ‘어문 규범과 관습’은 사회적으로 통용되는 기본적인 국어 규범을 준수하여 글을 작성하였는지에 대해 판단하는 준거이다. 이 준거의 하위 요소는 ‘어문 규범 준수’와 ‘글쓰기 관습 준수’로 대별된다.

‘어문 규범 준수’의 경우 기본적으로는 한글 맞춤법 규정을 따르는지를 기준으로 판단하며, 이 경우 오타자 여부도 오류로 판정한다. 채점 기준표상에서는 이러한 규범 미준수 오류가 글 전체에 거의 나타나지 않으면 감점이 없으며, 가끔 나타나는 경우에는 1점을, 빈번하게 나타나는 경우에는 2점을 감점한다. ‘가끔’과 ‘빈번하게’를 구분하는 기준점은 글의 분량이 제각기 다르기 때문에 명시적으로 설정되지는 않았으나 대체로 한 문단 내에 한두 개의 규범 미준수가 나타나는 경우를 ‘가끔’, 이를 초과하는 경우를 ‘빈번하게’로 판단할 수 있다.

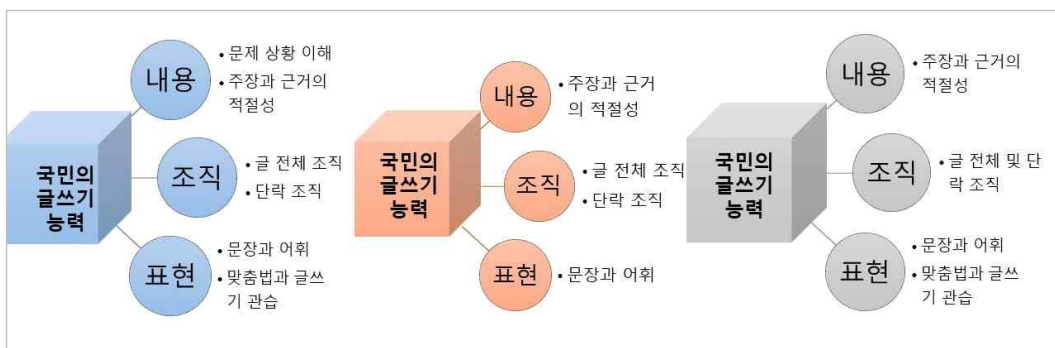
‘글쓰기 관습 준수’는 논증적 글쓰기라는 장르 관습에 적합한 문어적 표현을 일관되게 사용하였는지를 기준으로 판단한다. 가령, 일관된 종결 어미를 글 전체에 걸쳐 사용하지 않는다면, 구어체를 사용하는 경우는 논증적 글쓰기에 부적합한 사례이므로 감점 요소에 해당한다. 글쓰기 관습 역시 크게 종결 어미 부적합 유형과 장르 관습에 맞지 않는 표현 사용 유형으로 나누어, 각각의 유형이 나타날 때마다 1점씩 감점하여, 최대 2점이 감점될 수 있도록 설정하였다.

3. 준거별 가중치 연구

채점 준거의 가중치 연구는 채점 척도 구성의 이론적 타당성을 기반으로 실제 글쓰기 능력 추정의 안정성과 기대하는 글쓰기 능력 추정 효과를 강화하기 위한 목적으로 이루어졌다. 기본적으로 준거별 가중치를 결정하기 위해서는 먼저 가중치 설정의 목표를 수립해야 하며, 이는 수집한 글쓰기 자료가 어떤 성격을 갖고 있으며, 또한 어떤 성격이기를 기대하는지를 확립하는 것에 기반하여 지향점을 모색해야 한다.

먼저, 본 연구에서 지향하는 논증적 글쓰기의 상(像)이 무엇인지를 정립할 필요가 있었다. 한 편의 논증적인 글을 구성하는 데 필수적인 요소가 무엇이며, 그것을 어떤 준거로 측정하여야 하는지가 합의되어야 하며 그것들의 총합이 본 과업에서 지향하는 논증적 글쓰기임을 확인하는 과정이 선행되어야 했다. 이를 위해서는 본 연구에서 수집한 글쓰기 자료의 특성을 충분히 반영할 수 있는 글쓰기란 무엇인지 정의하기 위해 선행 연구를 검토하고, 논증적 글쓰기의 상에 대한 전문가들의 협의를 통한 숙의의 과정을 거쳤다.

현재의 채점 준거는 내용 준거가 5개, 조직 준거가 2개, 표현 준거가 2개로서, 총 9개의 준거가 논증적 글쓰기를 구성하는 것으로 설정된 것이다. 이는 준거 개수의 비율을 기준으로 삼는다면 상대적으로 내용 요소가 조직이나 표현 요소보다 강조된 것을 의미하며, 준거 개수 자체를 기준으로 삼는다면 논증적 글의 가능한 준거를 망라한 총괄형 모형에 가깝다고 할 수 있다. 만약 사전 검토와 협의를 통해 준거 간 비중을 조정할 필요가 있다면 비슷한 성격의 준거를 통합하는 식으로 조정하여 어떠한 준거에 더 큰 비중을 둘 것인지를 결정하는 방안도 가능할 것이다.



[그림 II-3] 준거 수(비율) 조정을 통한 가중치 적용 예시

위와 같은 작업에는 통계 분석 등을 활용한 양적 접근이 가능하다. 가령, 본 과업에서 수집한 채점 자료 5천 건에 대하여 준거별 점수와 총점 간의 상관을 분석하는 등 가중치 적용 방안을 탐색하기 위한 통계적 처리를 진행할 수 있다. <표 II-29>와 같은 이러한 상관 분석은 모든 준거를 대상으로 시행할 수도 있고, 특정 준거만을 대상으로 시행할 수도 있으며, 상관 분석의 통계적 결과는 이 진단 도구를 활용했을 때 이익이 커지는 방향으로 해석하는 것이 좋다.

<표 II-29> 2023년 가교 채점 문항 유형에 따른 준거별 총점과의 상관 계수 예시

문항 유형	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2
1	0.69	0.76	0.81	0.82	0.50	0.78	0.82	0.72	0.65
2	0.70	0.76	0.85	0.80	0.59	0.86	0.85	0.77	0.63
3	0.75	0.80	0.88	0.83	0.22	0.90	0.87	0.84	0.80
4	0.73	0.70	0.79	0.73	0.63	0.85	0.84	0.64	0.55
5	0.70	0.85	0.89	0.86	0.60	0.90	0.88	0.74	0.59
6	0.79	0.84	0.88	0.85	0.47	0.90	0.91	0.84	0.80
평균	0.73	0.79	0.85	0.82	0.50	0.86	0.86	0.76	0.67

준거 간 비중을 조정하는 것이 바람직하다고 판단될 시 통계 분석 자료를 그대로 활용하는 방법 이외에도 보조적으로 활용함으로써 어떠한 방향으로 가중치를 둘지 결정하는데 도움을 얻을 수 있다.

이러한 채점 준거의 가중치 연구는 채점 자료 수집이 모두 마무리된 시기에 채점 결과에 대한 통계적 분석의 결과를 활용하는 것을 병행함으로써 정교하게 수행할 수 있다. 잠정적으로 내용 준거(1-5)에 가중치를 부여하는 내용 중심 모형, 조직(1-2), 표현(1-2) 준거에 가중치를 두는 형식 중심 모형과 이들을 비교적 유사한 수준으로 고려하는 균형적 모형으로 나누어 볼 수 있을 것이다.

4. 설명적 글쓰기 문항 및 준거 관련 연구

1) 논증적 글쓰기와 설명적 글쓰기의 연결 지점 탐색

이 연구에서는 중장기적인 목표로서의 글쓰기 공통 채점 준거 개발을 목적으로 한다. 현재 단계에서는 안정화된 논증적 글쓰기의 채점 준거를 바탕으로 글쓰기 장르의 특성을 고려하여 설명적 글쓰기와의 연결 지점을 이론적으로 탐색하고, 설명적 글쓰기 문항의 프로토타입과 관련 채점 준거를 개발하는 기초 연구를 수행하였다. 이론적인 기초 작업으로, 논증적 글쓰기에 관한 채점 준거가 안정화된 이후 논증적 글쓰기와 설명적 글쓰기가 공유될 수 있는 준거와 현저한 차이가 있는 준거를 명확하게 구별하는 작업이 진행될 필요가 있다.

설명적 글쓰기는 설명 대상과 설명 목적에 따라 글의 구성 양상이 사뭇 다르다는 점에서 다소 외연이 포괄적인 장르이다. 이에, 이 연구에서는 국민의 글쓰기 능력 진단을 목적으로 하는 설명적 글쓰기 유형의 개발을 위한 기초 작업으로서 이론적인 탐색과 프로토타입 수준의 문항 및 채점 준거를 개발하는 기초 연구를 수행하였다. 이러한 작업은 이후 연구의 토대가 되어, 논증적 글쓰기 및 설명적 글쓰기의 채점 준거를 기반으로 다양한 장르에 적용 가능한 공통 채점 준거의 가능성을 탐색할 수 있다.

설명적 글쓰기 문항 및 채점 준거 개발은 <표 II-30>과 같은 절차로 진행되었다.

<표 II-30> 설명적 글쓰기 문항 및 채점 준거 개발 절차

수행 단계	세부 내용
논증적 글쓰기와 설명적 글쓰기의 연결 지점 이론적 탐색	<ul style="list-style-type: none"> · 기개발한 논증적 글쓰기 채점 준거 검토 · 설명적 글쓰기 및 설명적 글쓰기 평가 관련 국내외 연구 검토
설명적 글쓰기 문항 프로토타입 개발	<ul style="list-style-type: none"> · 국민의 글쓰기 능력 진단을 목적으로 하는 설명적 글쓰기 문항의 프로토타입 개발 - 구체적인 대상, 주제, 목적, 분량 등을 고려하여 문항 개발의 다양한 방식 검토 - 인공지능을 활용한 자동 채점 상황에서의 활용 가능성 고려 - 글쓰기 참여자의 수준 고려

수행 단계	세부 내용
관련 채점 준거 개발	<ul style="list-style-type: none"> · 이론적 탐색의 결과와 문항 프로토타입의 형식을 고려한 채점 준거 개발 - 개발한 문항 프로토타입에 대한 적용 가능성 검토 - 설명적 글쓰기 능력의 타당하고 신뢰로운 측정 가능성 검토
설명적 글쓰기 프로토타입 문항 및 관련 채점 준거 검토를 위한 전문가 섭외 및 요청	<ul style="list-style-type: none"> · 이 연구에서 의도하는 설명적 글 과제에 맞게 문항과 채점 준거가 개발되었는지 내부 검토 및 전문가 검토 진행, 지속적인 수정 및 보완

(1) 논증적 글쓰기와 설명적 글쓰기의 연결 지점 이론적 탐색

논증적 글쓰기와 설명적 글쓰기는 중심 내용(일반적 진술, 주제, 주장)과 뒷받침 내용(특수 진술, 근거)이 결합된 ‘핵심-상술’의 위계적 관계를 지닌다는 공통점이 있다. 이는 완결성의 원리로 설명되는 문단과 문단, 문장과 문장 사이의 일반적인 조직 원리이기도 하다(김봉균, 2002). 가령, 현재의 글쓰기 채점 준거에서 ‘주장(주장의 적절성, 주장의 명료성)’을 ‘주제(주제의 적절성, 주제의 명료성)’와 같이 바꾸어 구현해 볼 수 있다.

그러나 설명적 글쓰기는 설명하고자 하는 대상의 정보를 명확하게 구성하고 핵심 개념을 드러내는 어휘의 정확한 사용이 강조된다는 점에서 논증적 글쓰기와는 사뭇 다른 측면이 있다. 설명적인 글에서는 설명하고자 하는 대상을 일정한 기준에 따라 얼마나 구체적이고 정확하게 묘사하느냐가 중요한 요소로 작용하며, 설명의 목적에 따라 내용 구성의 방법도 달라질 수 있다(이정찬, 2013). 따라서 설명적 글쓰기의 장르적 특성을 고려하여 채점 준거의 구성이나 비중은 달라질 여지가 존재한다.

따라서 구체적으로 설명적 글쓰기의 문항을 개발하고 관련 채점 준거를 개발하기 위해서는 기존 유관 평가에서 설명적 글쓰기 문항을 어떻게 제시하고 있는지, 어떤 요소에 주목하여 문항의 평가를 수행해 오고 있는지를 전반적으로 검토할 필요가 있다. 이들 문항을 검토함으로써 설명적 글쓰기 문항과 채점 준거 개발에 관한 시사점을 도출할 수 있으며, 이로부터 현재 논증적 글쓰기 채점 준거의 내용, 조직, 표현의 측면과 설명적 글쓰기의 채점 준거가 어떤 측면에서 서로 비슷하고 달라질 수 있는지를 확인할 수 있을 것이다.

(2) 국내외 유관 평가에서의 설명적 글쓰기 문항 검토

설명적 글쓰기 문항 프로토타입 개발을 위해 국내외 유관 평가들에서 설명적 글쓰기 문항이 어떻게 구성되는지 검토하였다. 각 평가에서 확인할 수 있는 공통 요소들을 중심으로 한 분석 결과는 <표 II-31>과 같이 요약할 수 있다.

<표 II-31> 설명적 글쓰기 관련 국내외 시험 및 평가 특성

시험/평가 이름	예시 문항 소재	분량 및 시간	내용 조직 방식	시험/평가 대상	평가 요소	비고	자료 제공 여부
2018년 국민의 국어능력 실태 조사(민병곤 외, 2018)	전통 탈 모양 설명	300자 / 10분	묘사(나열 구조, 순서 구조)	대한민국 국민 (20세-69세)	내용 조직 표현	채점 요소에 '분량' 요건이 '내용' 범주에 포함되어 있음	△
평가를 활용한 초·중등학생 글쓰기 능력 신장 방안 연구(II)(박혜영 외, 2021)	미세먼지 감축 대책	800자 / 40분	나열 구조, 비교-대조 구조, 원인-결과 구조, 문제-해결 구조	고등학생	과제 내용 조직 표현	하나의 문항에 총체적 채점과 분석적 채점 방식 모두 적용	○
한국실용글쓰기검정 ((사)한국국어능력평가협회 주관)	부동산 시장 전망 보고서 작성	800자 / 90분(서술형 4문항 포함)	비교-대조 구조, 원인-결과 구조	일반인(고등학교 수준 포함)	구체적인 평가를 공개되지 않음.	직무 관련 글쓰기 중심(기안서, 품의서, 사내제안서, 보고서, 기획서, 사외제안서, 프레젠테이션, 홍보·광고문안, 거래·계약서, 기술문 등의 쓰기 능력)	○
토픽(TOPIK, 한국어능력시험)	온라인 쇼핑 시장의 변화	2-300자 / 50분(6-700자의 논술문 작성 포함)	비교-대조 구조	한국어가 제1언어가 아닌 재외동포 및 외국인	내용(내용 및 과제 수행) 조직(글의 전개 구조) 표현(언어 사용)		○

시험/평가 이름	예시 문항 소재	분량 및 시간	내용 조직 방식	시험/평가 대상	평가 요소	비고	자료 제공 여부
미국 NAEP (NCES 시행)	의사소통/정보 테크놀로지 사용에 관하여 대학입학위원회에 보내는 에세이	미확인 / 30분	묘사하기(나열 구조, 순서 구조), 비교-대조 구조	미국 학생 (4/8/12학년)	내용 생성 내용 조직 언어 사용과 규범	‘예상 독자’가 중요한 요소로 작용하여 실제적이고 구체적인 독자를 상정하도록 과제에서 요구함.	○
TOEFL iBT	침팬지와 보노보의 특성 비교	150-225 단어 / 20분	비교-대조 구조, 나열 구조	영어를 모국어로 하지 않는 외국인	내용 조직 표현	읽기 자료와 듣기 자료를 토대로 설명적 글쓰기 문항 제시	○
호주 OWA	시끄러운 이웃집 개가 있는 상황에 대해 의회에 서신 보내기	미확인 / 20-25분	묘사하기(나열 구조, 순서 구조), 문제-해결 구조	직업 학습자, ESL 학습자, 고등 교육 프로그램 등록 학습자, 대학원생 및 직장 내 전문직 종사자 등 성인 대상	내용(대상 및 목적, 아이디어 질적 수준) 조직(텍스트 응집성) 표현(언어 선택, 문장 구조, 문장에서의 부호, 문장 내 구두점, 철자법)	보고서 과제는 제시된 사실이나 수치, 그림 등을 바탕으로 요약된 보고서의 내용을 선택, 정리, 제시하는 것이 요구됨(제공되는 예시 문항은 보고서 과제와는 다소 거리가 많.)	○

설명적 글쓰기 문항 프로토타입 개발을 위해 국내외 유관 평가들에서 설명적 글쓰기 문항이 어떻게 구성되는지 검토하였다. 분석 결과는 다음과 같이 요약해 볼 수 있다. 먼저, 글의 분량을 기준으로 하면 200-300자 내외의 짧은 글을 요구하는 문항과 800자 정도의 상대적으로 긴 글을 요구하는 문항으로 양상이 대별된다. 예시 문항이나 분량을 공개하지 않는 평가도 있으나, 토픽(TOPIK)이나 토플 아비티(TOEFL iBT)에서의 짧은 글쓰기의 경우 그 대상이 주로 외국인이며, 박혜영 외(2021)의 연구나 한국실용글쓰기검정과 같은 상대적으로 긴 글쓰기 평가의 경우 800자 정도의 분량을 요구하고 있었다. 이러한 글의 분량은 설명적 글쓰기 문항이 전체 평가 문항에서 차지하는 위치나 평가 목적에 따라 조금씩 다른 것으로 보인다. 본 과업의 설명적 글쓰기 문항은 논증적 글쓰기 문항과 평가 기준표를 어느 정도 공유하고자 하며 응시자에게 설명적 글쓰기 단일 문항을 제시하는 형식으로 진행되므로 분량은 800자 이상의 글쓰기를 요구하는 문항이 적합할 것으로 보인다.

설명적 글쓰기는 설명 대상에 따라 다양한 설명 방식, 즉 글의 조직 방식이 가능하다. 이에 따라 여러 시험 및 평가에서도 문항의 소재에 따라 다양한 내용 조직 방식을 명시적으로 혹은 문항 스크립트에서 요구하고 있음을 확인할 수 있다. 가령, 국립국어원·서울대학교 국어교육연구소(2018)에서는 전통 탈 모양을 주로 묘사의 방식으로 설명할 것을 요구하고 있고, 토픽(TOPIK)에서는 온라인 쇼핑 시장의 ‘변화’에 따라 변화의 전후가 잘 드러날 수 있도록 비교-대조 방식으로 설명할 것을 요구하고 있다. 박혜영 외(2021)에서는 ‘미세먼지 감축 대책’이라는 국면이 다양한 종합적인 소재를 채택하여 나열뿐만 아니라 비교-대조나 원인-결과와 같은 다양한 내용 조직 방식을 활용할 수 있도록 의도하였다. 이처럼 설명적 글쓰기는 설명 대상에 따라 다양한 조직 방식이 가능하므로 대상의 어떤 측면을 설명할 때 어떤 조직 방식을 활용할 것인지도 설명적 글쓰기 능력의 한 측면이라 할 수 있다.

평가 요소는 세부 요소의 경우 시험 및 평가가 지향하는 목표에 따라 조금씩 다르지만, 거시적인 틀은 ‘내용’, ‘조직’, ‘표현’으로 구분될 수 있다. 대체로 ‘내용’의 경우 설명하는 내용의 질적인 수준, 설명 대상이나 목적에 따른 내용 기술의 충실성이 주를 이루며, ‘조직’의 경우 텍스트 내외의 응집성이, ‘표현’의 경우 언어 사용 방식과 규범 준수 등이 주된 세부 요소로 포함되어 있다.

자료 제공 여부에 관해서는, 대부분의 시험 및 평가에서 짧은 글이나 도표 등을 함께 제시하는 형식으로 문항을 구성하였다. 이는 설명적 글쓰기의 특성상 내용 생성에 배경

지식이 크게 관여하는바, 응시자 간 배경지식 정도의 차이를 어느 정도 최소화하기 위한 장치로 보인다. 그러나 자료 제시형의 경우 그러한 글과 도표에 내용이 지나치게 한정되어 내용 평가 요소에 대해서는 응시자 간의 수준 변별이 어려울 수도 있다는 난점이 있다.

2) 설명적 글쓰기 문항 및 채점 준거 개발

(1) 설명적 글쓰기 문항 프로토타입 개발

본 과업에서는 논증적 글쓰기와의 연결 지점을 고려하면서, 국내외 유관 평가에서의 설명적 글쓰기 특성을 검토하여 설명적 글쓰기 문항 프로토타입을 개발하고자 하였다.

글쓰기 분량으로는 90분 내외의 시간 동안 1,000자 내외의 글쓰기를 수행하도록 하였다. 이는 본 과업의 문항이 설명적 글쓰기 단일 문항을 응시자에게 제시하는 환경임을 고려한 것이며, 논증적 글쓰기와의 문형 유사성을 염두에 둔 것이다.

문항에서 요구하는 내용 조직 방식은 초기에 ‘비교-대조’ 방식으로 한정함으로써 체계적인 설명 방식을 안정적으로 사용할 수 있도록 의도하였다. 그러나 ‘비교-대조’ 방식이 반드시 설명적 글쓰기의 전형적이고 중핵적인 설명 방식이라고 볼 근거가 부족하며, 동일한 설명 대상에 대해서도 어떤 측면을 설명하고자 하는지에 따라 설명 방식이 다양하게 사용될 수 있다. 이에 따라 응시자가 내적인 기준을 세워 설명 대상을 그 의도와 목적에 맞게 다양하게 설명할 수 있도록 문항 차원에서 내용 조직 방식을 한정하지 않고, 이를 채점 기준표에서 고려할 수 있도록 변경하였다.

문항에서 별도의 자료는 제시하지 않는 것으로 결정하였다. 이 또한 논증적 글쓰기 문항과의 유사성을 고려한 것이며, 자료를 제시할 경우 글의 내용이 대부분 자료를 충실하게 기술하는 것으로 한정될 가능성이 커 내용 요소 수준 간 변별이 어려울 수 있으리라는 우려가 있었다. 대신 자료를 제시하지 않는 경우 응시자 간 배경지식 유무에 차이가 크게 발생할 수 있으므로 문항 소재는 응시자 대부분이 경험적으로 알 수 있을 만한 친숙한 것으로 선정하고자 하였다.

문항의 소재는 배경지식 여부에 크게 구애받지 않으며, 두 대상의 다양한 측면에 대하여 필자가 나름의 분석 기준을 설정하고, 이에 따라 두 대상의 특성을 깊이 있게 분석한

다음, 분석 결과를 바탕으로 논리적으로 설명할 수 있을 만한 화제를 선정하고자 하였다. 이를 설명의 대상과 분야에 따른 <표 II-32>에 따라 소재를 탐색하였다. <표 II-32>의 가로축은 다양한 분야의 소재를 선정하기 위하여 설정한 축으로서, 작년 과업에서 논증적 글쓰기 문항의 프로토타입을 개발할 때 고려했던 방향과 유사하게, ‘인문·예술’, ‘사회·문화’, ‘과학·기술(또는 융합)’로 설정하였다. 세로축은 소재를 설명하기에 적절한 중심 속성을 설정한 축으로서, 각각의 소재가 어떠한 성격을 띠고 있는지, 그래서 해당 소재를 체계적으로 설명하기에 적합한 설명 방식은 어떠한 수 있는지를 고려한 축이다.

<표 II-32> 설명적 글쓰기 문항 소재 탐색의 예

	인문·예술	사회·문화	과학·기술(또는 융합)
개념	예: ‘오마주’와 ‘표절’		
사실			예: 인간의 노동과 로봇의 노동
절차	예: 윗놀이와 체스의 규칙		
사건		예: 코로나 이전과 이후의 삶의 모습	

설명적 글쓰기 문항 개발을 위한 프로토타입에는 글쓰기의 상황 맥락(배경 설명과 예상 독자, 설명의 목적, 행동 및 과제)이 포함될 수 있도록 다음의 형식을 고려하여 문항을 구성하였다. 문항을 개발하기 위한 형식을 정리한 내용은 <표 II-33>과 같다.

<표 II-33> 설명적 글쓰기 문항 프로토타입의 형식

문항 형식	
배경 설명	영화 감상은 사람들이 오랫동안 즐겨온 문화 생활이다. 영화를 보기 위하여 지금까지는 영화관에 가서 영화를 감상하였지만, 최근에는 온라인 스트리밍 서비스(OTT)를 통해 장소에 구애받지 않고 감상하는 경우가 늘어나고 있다.
예상 독자	영화 감상을 영화관에 가서 할지, OTT 플랫폼을 통해 할지 고민하는 독자가
설명 목적	다양한 영화 감상 방법을 이해할 수 있도록
행동	이 두 가지 방식의 특성을 설명하고자 한다.
과제	유의 사항을 고려하여 글을 쓰시오.
유의 사항	△ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분

최종적으로 문항 간 발문의 글자 수를 서로 비슷하게 다듬고, 문항의 의도가 명확하게 전달될 수 있도록 수정하였다. 이와 같은 과정을 통해 개발한 설명적 글쓰기 문항 프로토타입 총 3문항은 다음과 같다.

문항 1. (인문·예술) 영화관에서의 영화 감상과 OTT 플랫폼에서의 영화 감상

영화 감상은 사람들이 오랫동안 즐겨온 문화 생활이다. 영화를 보기 위하여 지금까지는 영화관에 가서 영화를 감상하였지만, 최근에는 온라인 스트리밍 서비스(OTT)를 통해 장소에 구애받지 않고 감상하는 경우가 늘어나고 있다. 영화 감상을 영화관에 가서 할지, OTT 플랫폼을 통해 할지 고민하는 독자가 다양한 영화 감상 방법을 이해할 수 있도록 이 두 가지 방식을 설명하고자 한다. 유의 사항을 고려하여 글을 쓰시오. (231자)

[유의 사항]

- △ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것).
- △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분

문항 2. (사회·문화) 코로나19 이전과 이후의 삶의 모습

2019년 말 출현한 코로나 바이러스는 세계적인 팬데믹을 초래하였다. 인류는 코로나19를 극복하기 위해 다양한 노력을 기울였으며, “포스트 코로나”라는 용어가 생겨날 만큼 코로나19 이전과 이후의 삶의 모습에는 적지 않은 변화가 나타났다. 미래의 독자가 그 변화를 이해할 수 있도록 코로나19 전후 삶의 모습을 설명하고자 한다. 유의 사항을 고려하여 글을 쓰시오. (208자)

[유의 사항]

- △ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것).
- △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분

문항 3. (과학·기술) 인간의 노동과 로봇의 노동

기술의 발달은 산업 분야의 여러 측면에 변화를 가져 왔다. 특히 로봇의 등장은 노동의 자동화에 큰 기여를 했으며, 로봇의 영향력은 산업 현장에서 나아가 노동 시장으로까지 확대되고 있다. 자신의 사업(공장)에 어떤 노동 주체(로봇 또는 인간)를 선택할지 고민하는 독자가 각각의 특징을 이해할 수 있도록 두 주체 노동의 특징을 설명하고자 한다. 유의 사항을 고려하여 글을 쓰시오. (210자)

[유의 사항]

- △ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것).
- △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분

(2) 설명적 글쓰기 채점을 위한 채점 준거 개발

설명적 글쓰기 채점을 위한 채점 준거는 기개발된 논증적 글쓰기 채점 기준표와 대응될 수 있도록 개발하였다. 앞서 설명적 글쓰기를 다룬 유관 시험 및 평가에서의 평가 요소를 검토한 결과 대체로 내용, 조직, 표현의 큰 틀을 공유하고 있음을 확인할 수 있었다. 이때, ‘조직’과 ‘표현’은 설명적 글쓰기에서 요구되는 평가 요소가 논증적 글쓰기의 그것과 거의 동일하므로 논증적 글쓰기의 채점 기준표와 요소를 공유하며 대응될 수 있는 지점이라고 보았다.

반면, ‘내용’은 설명과 논증의 차이로 인하여 구체적인 평가 요소는 다소 상이할 수밖에 없는데, 이 또한 논증적 글쓰기의 평가 요소와의 거시적인 틀을 유지하고자 하였다. 먼저, ‘내용 1’에 해당하는 논증적 글쓰기의 ‘문제 상황 이해’는 필자가 논리를 풀어 나감에 있어 문제에서 요구하는 상황을 명확하게 파악하였는지, 파악한 상황을 설명하는데 필요한 정보를 제공하는지를 요구하는 준거이다. 이를 고려하여 설명적 글쓰기의 ‘내용 1’은 ‘맥락에 대한 이해’로 대응하고자 하였다. 글의 내용에서 예상 독자와 설명 목적을 고려하였는지, 필자가 설정한 설명 목적을 글의 전체 맥락 속에 편입하였는지를 중심으로 요구하는 준거로 설정하였다.

‘내용 2’에 해당하는 논증적 글쓰기의 ‘주장’은 글 전체를 관통하는 핵심적인 주장이다. 이는 설명적 글쓰기의 ‘중심 내용’에 해당하는 성격으로서 중심 내용의 적절성과 명료성을 중심으로 ‘내용 2’를 설정하였다.

‘내용 3’과 ‘내용 4’에 해당하는 논증적 글쓰기의 채점 준거는 각각 이유나 근거의 적절성과 충분성으로서, 적절성은 하위 주장을 질적으로 평정하는 기준을, 충분성은 하위 주장을 다소 양적으로 평정하는 기준을 의미한다. 이들 하위 주장은 설명적 글쓰기의 ‘뒷받침 내용’에 해당하는 성격으로서 대응된다. 이에 따라 ‘내용 3’은 ‘뒷받침 내용(정보)의 적절성’으로, ‘내용 4’는 ‘뒷받침 내용(정보)의 충분성’으로 설정하였으며, 적절성 측면에서는 뒷받침 내용이 중심 내용과 관련성이 깊으며 출처나 근거를 통해 뒷받침이 되는지를 평가할 수 있도록 하였고, 충분성 측면에서는 중심 내용을 뒷받침하는 내용이 충분하며 설명 대상의 특성이 적절하고 다양한 방식으로 제시되었는지를 평가할 수 있도록 하였다.

한편, 논증적 글쓰기 ‘내용 5’의 ‘다른 입장에 대한 고려’는 논증 과정에서의 필자의

사고의 깊이를 측정하는 준거로서의 성격을 띤다. 자신의 주장과 반대되는 입장을 고려하는 것이 필자의 논증에 설득력을 더하고 논의 전개가 깊이 있고 풍성해지기 때문이다. 그러나 이러한 요소는 설명 목적의 글쓰기에서는 대응하여 다루기 어려운 요소이므로 설명적 글쓰기 채점 기준표에서는 배제하였다. 이상의 논의를 종합하여 최종적으로 설정한 설명적 글쓰기 채점 준거는 <표 II-34>와 같다.

<표 II-34> 설명적 글쓰기 채점 준거

준거	내 용				조 직		표 현	
	내용 1 맥락에 대한 이해	내용 2 중심 내용	내용 3 뒷받침 내용(정보)의 적절성	내용 4 뒷받침 내용(정보)의 충분성	조직 1 글 전체 구성	조직 2 문단 구성 요건	표현 1 문장과 어휘	표현 2 어문 규범과 관습
하위 준거	<p>▶ 예상 독자와 설명 목적에 대한 고려 : 구성한 내용이 예상 독자와 설명 목적에 부합한가?</p> <p>▶ 맥락 설정의 적절성 : 설명 대상을 일정한 맥락 속에 편입시키는가?</p> <p>※ 유의 사항 도입을 중심으로 보되, 글의 중간중간에 드러날 수 있음. (1) 불필요한 정보가 포함되었거나, (2) 불필요한 정보로 인해 맥락에 대한 설명의 분량이 길어진 경우 높은 점수를 부여하기 어려움.</p>	<p>▶ 중심 내용의 적절성 : 중심 내용이 논제에 부합하는가?</p> <p>▶ 중심 내용의 명료성 : 중심 내용이 글 전체에 일관되고 뚜렷한가?</p> <p>※ 유의 사항 용어 설명: '정보'는 필자 외부에 객관적으로 존재하는 사실이나 데이터를, '중심 내용'은 이러한 정보를 바탕으로 필자가 궁극적으로 전달하려는 주요 생각(주제)을 의미함.</p>	<p>▶ 정보의 관련성 : 뒷받침 내용(정보)이 중심 내용과 관련되는가?</p> <p>▶ 정보의 신뢰성 : 설명한 정보가 출처나 근거를 통해 뒷받침되는가?</p> <p>※ 유의 사항 - '정보의 관련성' 판단은 중심 내용(주제)에 대한 뒷받침 내용으로 식별함. - '정보의 신뢰성' 판단은 중심 내용을 전달하기 위해 필자가 사용한 외부의 사실이나 데이터로 식별함.</p>	<p>▶ 뒷받침 내용(정보)의 충분성 : 중심 내용을 뒷받침하는 적절한 내용(정보)이 충분한가?</p> <p>▶ 뒷받침 내용(정보)의 다양성 : 중심 내용을 뒷받침하는 내용(정보)이 다양한가?</p> <p>※ 유의 사항 - 정보는 중심 내용(주제)에 대한 뒷받침 내용으로 식별함. - '다양성'은 뒷받침 내용이 예시나 상세화, 인용 등의 여러 방식으로 제시되는 것으로 식별함.</p>	<p>▶ 문단 구분의 적절성 : 도입-전개-정리의 구조를 갖추고 있는가? (형식 문단과 내용 문단 모두 포함)</p> <p>▶ 문단 배열의 유기성 : 문단 배열이 유기적이고 자연스러운가?</p> <p>※ 유의 사항 글 전체의 흐름이 대체로 자연스러우나 형식 문단이 적절히 나뉘지 않은 경우 3점 부여함.</p>	<p>▶ 완결성 : 뒷받침 문장을 사용하여 문단의 중심 문장을 뒷받침하였는가?</p> <p>▶ 통일성 : 문단별로 하나의 중심 생각(소주제)이 전개되는가?</p> <p>▶ 일관성 : 문단을 이루는 여러 문장들이 서로 긴밀한 관계로 연결되는가?</p> <p>※ 유의 사항 형식 문단을 기준으로 채점하되, 형식 문단은 구별되지 않지만 내용 문단이 분명하게 드러나는 경우에는 내용 문단을 기준으로 각 구성 요건을 판단할 수 있음.</p>	<p>▶ 문장 표현의 자연스러움 : 문장 길이가 적절하고 그 표현이 자연스러운가? (예: 주술호용)</p> <p>▶ 문장 표현의 효과성 : 문장의 의미가 명료하고 그 표현이 효과적인가?</p> <p>▶ 어휘 사용의 적절성 : 어휘 사용이 문맥에 맞는가?</p> <p>※ 유의 사항 문장의 길이의 적절성에 대한 판단은 길이 자체보다는 의미 전달이 잘 되는지를 기준으로 함.</p>	<p>▶ 어문 규범 준수 : 맞춤법, 띄어쓰기 등을 준수하였는가? (오탈자 여부 포함)</p> <p>※ 유의 사항 - 어문 규범 위반이 거의 나타나지 않는 경우 감점 없음. - 어문 규범 위반이 가끔 나타나는 경우 -1점. - 어문 규범 위반이 빈번하게 나타나는 경우 -2점.</p> <p>▶ 글쓰기 관습 준수 : 장르 관습에 맞는 문어적 표현을 사용하는가?</p> <p>※ 유의 사항 - 격식체(-습니다) 사용은 용인함. - 구어체(-요)를 사용하거나 종결어미를 일관되게 사용하지 않은 경우 -1점. - 자기 소개, 인사말 등 장르 관습에 맞지 않는 표현을 사용한 경우 -1점. - 기타 단정적인 어조 등 관습과 무관한 문체상의 특징에 대해서는 감점하지 않음.</p>

5	예상 독자와 설명의 목적이 글에 효과적으로 제시되었으며, 설명 대상에 관한 맥락 설정이 반드시 필요한 정보로 구성됨.	중심 내용이 논제에 부합하며, 글 전체에 걸쳐 매우 일관되고 뚜렷함.	뒷받침 내용이 중심 내용과 관련성이 매우 높으며 신뢰할 만한 출처나 근거와 함께 제시되어 있음.	중심 내용을 뒷받침하는 내용을 2가지 이상 충분히 제시하고, 각각의 뒷받침 내용을 그에 걸맞은 적절한 방식으로 다양하게 제시함.	형식 문단과 내용 문단 모두에서 도입-전개-정리의 구조가 매우 적절하게 구분되고, 문단 배열이 매우 유기적임.	모든 문단에서 완결성, 통일성, 일관성을 갖춘.	문장과 어휘 사용이 자연스럽고 설명에 효과적임.	어문 규범과 글쓰기 관습을 매우 정확히 준수함.
4	예상 독자와 설명의 목적이 글에 제시되었으며, 설명 대상에 관한 맥락 설정이 반드시 대체로 필요한 정보로 구성됨.	중심 내용이 논제에 부합하며, 글 전체에 걸쳐 대체로 일관되고 뚜렷함.	뒷받침 내용이 중심 내용과 관련성이 높으며 출처나 근거와 함께 제시되어 있음.	중심 내용을 뒷받침하는 내용을 2가지 이상 제시하고, 각각의 뒷받침 내용을 그에 걸맞은 적절한 방식으로 다양하게 제시함.	형식 문단과 내용 문단 모두에서 도입-전개-정리의 구조가 대체로 적절하게 구분되고, 문단 배열이 대체로 유기적임.	대부분의 문단에서 완결성, 통일성, 일관성을 갖춘.	문장이 대체로 자연스럽고 어휘 사용이 대부분 적절하나, 설명에 효과적으로 보이지 않음. * 글의 길이에 따라 2-3개 어색한 문장과 어휘가 있을 수 있음.	어문 규범 위반이 가끔 나타나거나 한 가지 유형의 글쓰기 관습 위반이 나타남. * 어문 규범과 글쓰기 관습, 둘 중 하나만 문제가 있음.
3	예상 독자와 설명의 목적이 글에 제시되었으나, 다소 불필요한 정보로 인해 맥락 설정이 흐려짐.	중심 내용이 논제에 대체로 부합하나, 일관성이 다소 부족하거나 뚜렷하지 않음.	뒷받침 내용이 중심 내용과 관련성이 있으나 출처나 근거와 함께 제시되지 않음.	중심 내용을 뒷받침하는 내용을 1가지 제시하고, 뒷받침 내용을 그에 걸맞은 다양한 방식으로 제시함.	형식 문단 또는 내용 문단의 구분과 배열이 일부 자연스럽지 않으나, 글을 읽는 데 크게 방해가 되지 않음.	일부 문단에서 완결성, 통일성, 일관성을 갖춘.	자연스럽지 않은 문장과 부적절한 어휘가 사용되었으나, 문장의 의미를 파악하는 데 크게 방해가 되지 않음. * 한 단락에 1-2개의 오류가 있을 수 있음.	어문 규범 위반이 가끔 나타나며, 한 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범은 정확히 준수하였으나 두 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범 위반이 빈번하게 나타나지만 글쓰기 관습은 정확히 준수하였음.
2	예상 독자나 설명의 목적에 대한 고려가 거의 드러나지 않으며, 맥락 설정에 대한 인식이 없음.	중심 내용이 논제에서 다소 어긋남. 또는 중심 내용이 논제와 관련은 있지만 일관성이나 명료성이 매우 부족함.	뒷받침 내용이 중심 내용과 관련성이 없는 경우가 많아 설명력이 떨어지며, 출처나 근거가 제시되어 있더라도 신뢰하기 어려움.	중심 내용을 뒷받침하는 내용을 1가지 제시하였으나, 뒷받침 내용을 그에 걸맞은 방식으로 제시하지는 못함.	글의 체계가 다소 부족하여 완성도가 낮고, 문단 배열 가운데 부자연스러운 부분이 많음.	문단의 완결성, 통일성, 일관성이 전반적으로 부족하여 완성도가 낮음.	자연스럽지 않은 문장과 부적절한 어휘가 사용되어 문장의 의미를 파악하는 데 방해가 됨. * 한 단락에 2개 이상의 심각한 오류가 있을 수 있음.	어문 규범 위반이 빈번하게 나타나며, 한 가지 유형의 글쓰기 관습 위반이 나타남. 혹은 어문 규범 위반이 가끔 나타나며, 두 가지 유형의 글쓰기 관습 위반이 나타남.
1	예상 독자 및 설명의 목적이 글에 제시되지 않음.	중심 내용이 논제에 부합하지 않음. 또는 중심 내용이 논제와 관련되지 않는 것은 아니나 일관성과 명료성이 매우 부족함.	뒷받침 내용이 대부분 중심 내용과 관련성이 없어서 설명력이 매우 낮음.	중심 내용을 뒷받침하는 내용이 모두 충분하지 않음.	내용 문단이 거의 구분되지 않아, 글의 체계가 없음.	대부분의 문단에서 구성요건을 갖추지 못함.	문장 대부분이 부자연스럽고, 어휘 대부분이 글의 맥락에 어울리지 않아 문장의 의미를 파악하는 데 크게 방해가 됨.	어문 규범 위반이 빈번하게 나타나며, 두 가지 유형의 글쓰기 관습 위반이 나타남.

이후에는 위와 같은 과정을 통해 개발한 설명적 글쓰기 프로토타입 문항과 관련 채점 준거에 대하여 전문가를 섭외하여 타당성과 적합성 측면에서 큰 문제는 없는지, 보완하거나 수정해야 할 지점은 없는지를 중심으로 협의회를 통해 검토하며 수정하고자 하였다.

3) 설명적 글쓰기 프로토타입 문항 및 채점 준거에 대한 전문가 검토

개발된 설명적 글쓰기 프로토타입 3문항과 관련 채점 준거에 대해서는 전문가 2인을 섭외하여 검토를 요청하였다. 전문가 검토를 요청할 때에는 예상되는 필자의 반응을 가늠해 볼 수 있도록 생성형 인공지능 기술을 활용하여 예시 답안을 작성하고, 이를 참고로 함께 송부하였다.

<표 II-35>는 전문가 2인으로부터 받은 주요 지적 사항을 세부 범주별로 나누어 정리한 것이다.

<표 II-35> 설명적 글쓰기 프로토타입 문항 및 채점 준거에 대한 전문가 검토 결과

범주	전문가 검토 결과
문항 공통	<p>문항의 배경 설명을 응시자가 거의 그대로 활용하여 갖다 쓰는 경우(몇 단어만 조금 바꿔서 인용하는 경우)는 어떻게 채점해야 할 것인지에 대한 고려가 필요함. 유의 사항에 가령, 문항의 설명을 그대로 글에 쓰지 않아야 한다는 점 등이 반영될 필요는 없는지?</p> <p>설명 대상에 관한 맥락 설정이 중요한 채점 준거라면 문항의 배경 진술 부분이 평가 결과에 영향을 줄 수 있음. 설명 대상만을 제시할지 적절한 맥락과 함께 제시할지는 이 과제의 목적과 그리고 두 가지 유형의 이점 중 어느 것이 더 중요한지에 대한 연구진의 관점에 따라 결정될 사항임.</p>
문항 1: 영화관에서의 영화 감상과 OTT에서의 영화 감상	최근의 추세가 OTT 방식의 비중이 높아지고 있어 설명의 비중이 애초에 OTT에 놓여 있을 수 있다는 문제는 감안해야 할 듯함.

범주	전문가 검토 결과
<p>문항 2: 코로나19 이전과 이후의 삶의 모습</p>	<p>과제에서 말하는 ‘코로나 이후’가 현재와 같은 일상 회복의 때를 말하는 것인지, 코로나를 한창 겪고 있을 때를 말하는 것인지 다소 모호함. 예시 답안의 변화 양상도 ‘코로나를 겪고 있는 때’를 서술한 것으로 보임. 코로나 종식 직후에는 변화의 양상이 감지되었지만, 현재 상황에서는 변화의 모습을 알아채기가 어려워 보이기 때문임.</p> <p>문항 2의 독자는 문항 1과 3의 독자와는 다른 성격을 지님. 문항 1과 3은 두 가지 대상 중에 무언가를 선택해야 하는 목적성을 지닌 독자임. 그러나 문항 2의 독자는 단지 삶의 변화를 이해하기만 하면 되는 독자임. 이에 따라 문항 1과 3은 각 대상의 장단점을 비교하며 서술될 가능성이 높으나, 문항 2는 코로나19 이후의 삶의 변화를 중심으로 서술될 가능성이 높음. 연구 목적상 세 문항이 동등해야 한다면 문항 2의 성격이 다소 다르다는 점은 고려해야 함.</p>
<p>문항 3: 인간의 노동과 로봇의 노동</p>	<p>노동의 문제를 다루고 있어 다른 쓰기 과제보다는 전문적인 분야에 속한다고 판단됨. 따라서 다른 쓰기 과제보다 다소 난도가 높아 보임. 로봇 노동을 경험해보지 않은 글쓰기 참여자는 좀 더 어려움을 느낄 것으로 예상됨.</p> <p>문항 3의 “자신의 사업(공장)에 어떤 노동 주체(로봇 또는 인간)를 선택할지 고민하는 독자가 각각의 특징을 이해할 수 있도록 두 주체 노동의 특성을 설명하고자 한다.”는 진술에서 ‘주체’란 말의 표현이 조금 어색함. 물론 실제 현실에서는 당연히 인간의 노동과 로봇의 노동 중 이익에 따라 선택하겠지만, 인간과 로봇을 두 가지 노동 주체로 보고 선택하는 것에서 인간이 로봇과 같이 수단으로 여기거나 아니면 로봇을 의식과 생각이 있는 주체로 보는 것처럼 여겨져서 어색하게 느껴짐.</p>
<p>채점 기준표 ‘내용 1’ - 맥락에 대한 이해</p>	<p>내용 1 준거명인 “맥락에 대한 이해”에서의 ‘맥락’과 하위 준거인 “맥락 설정의 적절성”에서의 ‘맥락’은 서로 다른 성격을 지닌 용어로 보임. 전자는 쓰기의 맥락(상황 맥락 또는 수사적 맥락)인 독자, 목적, 주제, 매체 등을 의미하지만, 후자는 설명 대상의 중요성이나 설명의 필요성 등에 대해 진술하는 것과 관련 있어 보임.</p>

범주	전문가 검토 결과
	<p>5점 진술인 “예상 독자와 설명의 목적이 글에 효과적으로 제시되었으며, 설명 대상에 관한 맥락 설정이 반드시 필요한 정보로 구성됨.”에서 예상 독자가 글에 효과적으로 제시되었다는 것보다는 예상 독자를 고려하여 글을 썼다는 것이 더 적합할 것 같음.</p> <p>‘설명 대상을 일정한 맥락 속에 편입’한다는 것이 어떤 의미인지 명료하지 않음.</p> <p>‘예상 독자 조건’이 채점 과정에서는 실질적인 요소로 기능하지 못할 것 같음. 결국은 쓰기 과제에서 요구하는 대상별 특징을 얼마나 잘 설명하고 있는가가 관건이 되지 않을까 싶음.</p>
<p>채점 기준표 ‘내용 2’ - 중심 내용</p>	<p>‘중심 내용의 적절성’에서 ‘논제’라는 표현을 사용하고 있는데, 논제는 주로 논증적 글쓰기나 토론 등에서 사용하는 용어이고 글쓰기 참여자가 작성하는 글이 ‘설명문’이므로 다른 용어로 대체하는 것이 낫겠음.</p> <p>‘중심 내용의 명료성’에서 중심 내용이 일관되고 뚜렷한가라는 요구 사항도 논설문에 좀 더 적합한 듯함. 설명문이라면, 얼마나 정보성이 있는가, 즉 얼마나 구체적인가, 얼마나 유용한가와 같은 요구 사항이 좀 더 적절할 듯함.</p> <p>‘중심 내용의 명료성’의 준거 진술이 “중심 내용이 글 전체에 일관되고 뚜렷한가?”인데, 이 진술은 ‘조직 2: 문단 구성 요건’의 ‘일관성’과 용어가 중복되나 다른 지점을 가리키는 것 같음. 너무 세밀하게 따지는 것일 수도 있지만 굳이 따져 보자면 조직 2의 일관성에서 “문장들이 서로 긴밀하게 연결되는가”가 의미하는 바를 명확히 해야 할 것 같음.</p>
<p>채점 기준표 ‘내용 3’ - 뒷받침 내용(정보)의 적절성</p>	<p>‘정보의 신뢰성’에서 정보의 출처나 근거를 설명문에서 확인하기는 어려워 보임. 학술 논문이라면 주석이나 출처를 인용하겠지만, 주어진 글쓰기 과제에서는 이러한 방식을 채택하지 않는 것 같음.</p> <p>‘내용 3’은 ‘정보의 관련성’과 ‘정보의 신뢰성’을 하위 채점 준거로 설정하였는데, 이 두 요건을 조합하여 5점 척도로 진술문을 구성하였음. 그러다 보니 5개의 평가 진술문만으로 모든 상황을 포괄하기 어려운 문제가 있을 듯함. 가령, 5점인 경우는 “뒷받침 내용이</p>

범주	전문가 검토 결과
	<p>중심 내용과 관련성이 매우 높으며 신뢰할 만한 출처나 근거와 함께 제시되어 있음.”인데, “뒷받침 내용이 중심 내용과 관련성이 (3점의 ‘있으나’와 달리) ‘매우 높으나’ 신뢰할 만한 출처나 근거와 함께 제시되어 있지 않은 경우”도 있을 것임.</p> <p>현재 설명문 쓰기 과제가 ‘비교-대조’의 구조를 적용해야 하는데, 이와 관련된 사항이 ‘적절성’에 포함되어야 하지 않을까 싶음. ‘영화’를 예로 든다면, 비교-대조할 때 ‘영화관 방식’과 ‘OTT 방식’의 항목(내용)이 부합하는지, 내용 항목들은 서로 배타적으로 이루어져 있는지 등이 포함되어야 할 듯함.</p>
<p>채점 기준표 ‘내용 4’ - 뒷받침 내용(정보)의 충분성</p>	<p>‘내용 4’는 ‘내용의 충분성’을 다루고 있는데, 첫째 번 하위 준거는 직접적으로 이를 반영하고 있으나, 둘째 번 하위 준거는 ‘다양성’으로서 여기에 부합하지 않는 듯함. 1번 하위 준거의 ‘충분성’을 다른 요소로 수정하면, 내용 4의 ‘충분성’은 수정된 그 요소와 ‘다양성’으로 구성되는 요인으로 본다면 준거의 안정성이 더 높아지지 않을까 싶음.</p> <p>‘다양성’의 유의 사항은, “뒷받침 내용이 예시나 상세화, 인용 등의 여러 방식으로 제시되는 것으로 식별한다”라고 제시되었는데, 이 유의 사항은 뒷받침 내용 즉 정보의 다양성이 아니라 설명 방법의 다양성을 의미하는 것이 아닌가?</p> <p>설명문에서는 설명 내용의 ‘균형’도 중요한 듯합니다. 두 개의 대상으로 비교-대조하는 구조로 글을 작성해야 하므로, 두 대상에 대한 설명이 균형을 이루어야 하는데, 한쪽에 편중되어 있다면, 잘 작성한 설명문이 될 수 없을 듯함(정보의 균형 또는 내용의 균형).</p> <p>설명 방법에 대한 것은 채점 준거로 설정되어 있지 않은 것 같음.</p>
<p>채점 기준표 ‘조직 1’ - 글 전체 구성</p>	<p>5점: “형식 문단과 내용 문단 모두에서 서-본-결의 구조가 매우 적절하게 구분되고~”라는 진술문은 형식 문단과 내용 문단 안에 서-본-결의 구조가 있는 것처럼 읽힐 가능성이 있음.</p> <p>설명문 글쓰기 과제의 경우, 모두 비교-대조를 활용해야 하는데, 채점 준거, 즉 문단 배열이나 문단 구성</p>

범주	전문가 검토 결과
	<p>준거에 이에 관한 사항이 포함되어야 할 듯함. ‘영화 감상’을 예로 들면, 영화관 방식의 특징을 전체적으로 설명한 후에, OTT 방식의 특징을 비교-대조하여 서술할 수도 있지만, 내용 항목별로 영화관 방식과 OTT 방식을 비교-대조하여 서술할 수도 있습니다. 이 둘은 문단 배열 또는 문단 구성이 매우 달라 보임.</p>
<p>채점 기준표 ‘조직 2’ - 문단 구성 요건</p>	<p>하위 준거 중 “완결성: 뒷받침 문장을 사용하여 문단의 중심 문장을 뒷받침하였는가?”는 내용 준거인 “뒷받침 내용(정보)의 충분성”과 중복될 가능성이 있지 않은가? 내용 문단을 고려할 수 있다는 유의 사항에 따라, 문단이 구분되어 있지 않아도 뒷받침 문장을 사용하여 문단의 중심 문장을 뒷받침하였는지 판단하는 것은 결국 뒷받침 내용이 충분한지를 보는 것과 겹치는 것은 아닌지 염려됨.</p> <p>유의 사항에서 ‘형식 문단’과 ‘내용 문단’이라는 용어에 대해 재고가 필요함. 유의 사항에서 “형식 문단은 구별되지 않지만 내용 문단이 분명하게 드러나는 경우”라는 의미가 문단 구분은 되어 있지 않은 글이지만 글의 내용으로 보아 몇 개의 문단으로 나눌 수 있는 글을 의미한다면, 다른 표현이 필요하지 않을까 싶음.</p> <p>‘완결성’, ‘통일성’에서 언급하고 있는 문단이 형식 문단을 뜻한다면, 비교-대조의 서술 시 고려해야 할 사항이 더 많은 듯함. 예를 들어 영화관 방식과 OTT 방식을 비교-대조할 때, ‘접근 방식의 차이’라는 문단에 영화관 방식의 접근 방식과 OTT 방식의 접근 방식이 같이 담길 텐데, 글쓰기 참여자에 따라 이를 한 문단으로 묶을 수도 있고 이를 별개의 문단으로 묶을 수도 있을 듯함. 이러한 구성을 어떻게 처리할지를 안내가 필요하지 않을까 싶음.</p>
<p>채점 기준표 ‘표현 1’ - 문장과 어휘</p>	<p>문장과 관련된 채점 준거로 ‘어법에 맞게 문장을 쓰는 것과 관련된 내용’이 필요하지는 않은가? 물론 “문장 길이가 적절하고 그 표현이 자연스러운가? (예: 주술 호응)”이 있기는 하지만 ‘자연스러운 문장’과 ‘문법에 맞는 문장’은 그 의미에 있어 차이가 있으므로.</p> <p>어휘의 경우, 설명문에 사용한 어휘의 ‘세련도’를 반영할 필요가 있지 않은가?</p> <p>쓰기 과제에 따라 외국어가 쓰일 수 있는 상황이라면, 이에 대한 준거는 별도로 필요하지 않은가?</p>

범주	전문가 검토 결과
채점 기준표 '표현 2' - 어문 규범과 관습	어문 규범의 준수와 글쓰기 관습(장르 관습)의 준수 여부를 따져 가며 여러 경우를 수를 따져 나열한 채점 준거 진술문이 평가자에게 평가의 어려움을 주지 않을까 하는 염려가 듭.
기타 의견	쓰기 과제별로 난도의 차이가 있어 보이므로, 쓰기 과제를 수정하거나 활용할 때 이를 고려하면 좋을 듯함. 생활 밀착형의 쓰기 과제는 배경지식(경험)의 차이의 영향을 줄일 수 있으므로, 이러한 부담을 줄일 수 있다고 판단됨.
	설명문의 구조(또는 설명 방식) 중에서 '비교-대조'를 활용하도록 하는 과제를 구성하였는데, 이 구조만을 다루도록 할 예정이라면 채점 기준표의 준거를 이에 맞도록 수정할 필요가 있어 보임. 현재는 범용인 것처럼 보입니다. 범용도 범용으로서 장점이 있지만 쓰기 과제와 거리가 있어 보이는 문제가 있으므로 이 거리를 좁힐 필요가 있어 보임.
	생성형 AI를 사용해 글의 기초를 구성하고 난 후 이를 활용하거나 수정하여 예시 답안을 만들었다고 한다면, 수정에 좀 더 주의를 기울여 예시 답안에서 문법적으로 부적절하거나 표현이 부적절한 것이 없도록 해야 할 듯함.

전문가 2인이 지적한 사항 가운데에는 즉각적으로 수용 가능한 부분도 있고, 중장기적인 시각에서 시간을 두고 검토가 필요한 부분도 있다. 설명적 글쓰기 문항의 경우 올해 과업에 실제 투입하지 않을 예정이며, 문항 2(코로나19 이전과 이후의 삶의 모습)에 대한 전문가 반응을 통해 가늠할 수 있듯 과제의 실제성에 대한 인식과 감각이 차후 실제 평가 적용 시 달라질 수 있음을 감안하여, 그 가능성을 다각도로 열어 놓고 수정 폭을 결정하는 것이 필요하다는 판단에 이르게 되었다. 설명적 글쓰기 채점 준거는 논증적 글쓰기의 채점 준거와 대응될 수 있도록 기초를 유지하되 여러 선행 연구와 사례를 검토하여 고안한 것이지만 설명적 글쓰기와 관련한 실제 과업 수행 시 이러한 전문가 검토 의견을 참조하여 세밀하게 조율할 필요가 있음을 확인하였다.

다만 설명적 글쓰기 프로토타입 문항의 경우 차후 연구에 바로 적용될 수 있음을 고려하여 문항 2를 중심으로 수정한 후, 다음과 같이 확정하였다.

문항 1. (인문·예술) 영화관에서의 영화 감상과 OTT 플랫폼에서의 영화 감상

영화 감상은 사람들이 오랫동안 즐겨온 문화 생활이다. 영화를 보기 위하여 지금까지는 영화관에 가서 영화를 감상하였지만, 최근에는 온라인 스트리밍 서비스(OTT)를 통해 장소에 구애받지 않고 감상하는 경우가 늘어나고 있다. 영화 감상을 영화관에 가서 할지, OTT 플랫폼을 통해 할지 고민하는 독자가 다양한 영화 감상 방법을 이해할 수 있도록 이 두 가지 방식을 설명하고자 한다. 유의 사항을 고려하여 글을 쓰시오. (231자)

[유의 사항]

- △ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것).
- △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분

문항 2. (사회·문화) 김밥과 햄버거

김밥은 한국인에게 친숙한 음식으로, 밥 위에 다양한 속 재료를 넣고 김으로 말아서 만든다. 김밥은 조리의 절차가 복잡하지 않고 간편하게 먹을 수 있으며 다양한 재료가 사용된다는 점에서 햄버거와 비슷한 측면이 많다. 그러나 김밥을 먹어 보지 못한 사람에게는 다소 낯선 음식일 수 있다. 햄버거는 알지만 김밥은 잘 모르는 외국인이 김밥의 다양한 측면을 이해할 수 있도록 김밥과 햄버거를 함께 설명하고자 한다. 유의 사항을 고려하여 글을 쓰시오. (245자)

[유의 사항]

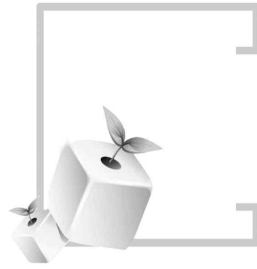
- △ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것).
- △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분

문항 3. (과학·기술) 인간의 노동과 로봇의 노동

기술의 발달은 산업 분야의 여러 측면에 변화를 가져왔다. 특히 로봇의 등장은 노동의 자동화에 큰 기여를 했으며, 로봇의 영향력은 산업 현장에서 나아가 노동 시장으로까지 확대되고 있다. 자신의 사업(공장)에 어떤 노동 주체(로봇 또는 인간)를 선택할지 고민하는 독자가 각각의 특징을 이해할 수 있도록 두 주체 노동의 특성을 설명하고자 한다. 유의 사항을 고려하여 글을 쓰시오. (210자)

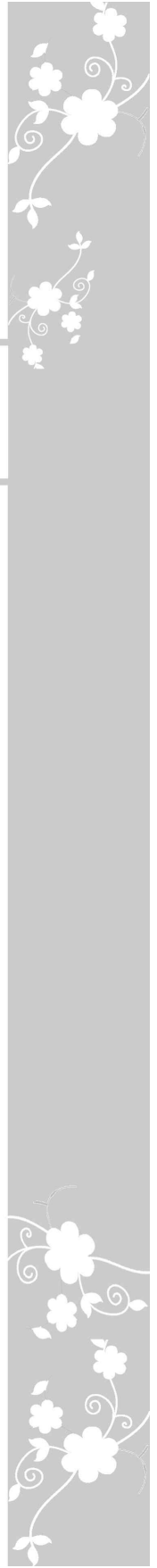
[유의 사항]

- △ 도입, 전개, 정리의 구조로 완결된 글을 쓸 것(제목을 쓰지 말 것).
- △ 분량: 1,000자 내외(±200자, 공백 포함), 시간: 90분



제 Ⅲ 장

글쓰기 자료 수집



1. 글쓰기 자료 수집을 위한 대학 협의체 운영

본 연구의 목적인 신뢰성과 타당성을 갖춘 고도화된 글쓰기 능력 진단 체계를 개발하기 위해서는 연구의 목적에 따라 설계된 글쓰기 문항에 대한 글쓰기 자료를 대규모로 확보하여 채점 결과를 분석하는 것이 필수적이다. 글쓰기 원자료 구축은 인공지능 학습 데이터의 기반이 되는 원시 데이터를 확보하기 위한 단계로서도 중요하다. 인공지능 개발에서 그 품질은 80-90%가 데이터 구축 과정에서 결정되기 때문에(과학기술정보통신부, 2023) 원시 데이터 확보에는 신중하고 체계적인 계획을 필요로 한다. 이에 본 연구에서는 대학 협의체 운영을 통해 글쓰기 참여자를 안정적으로 확보하고, 글쓰기 자료 수집 플랫폼을 개발하여 대규모 글쓰기 자료를 체계적으로 수집하고 관리하며, 데이터 검수 및 정제 팀을 상시 운영하여 원시 데이터의 질을 담보하였다.

성인의 글쓰기 능력 진단 체계 개발을 위해 요구되는 글쓰기 자료는 진단의 특성에 부합하는 특정한 글쓰기 조건하에서 성인이 작성한 글쓰기 자료여야 한다. 또한 인공지능 학습용 데이터로 구축되기 위해서는 충분한 수량과 데이터의 다양성 확보가 중요하다.

이에 본 연구에서는 발주처인 국립국어원이 글쓰기 자료 수집을 위해 맺은 9개의 국립 대학과의 업무 협약에 따라 협약 대학의 대학생들을 대상으로 글쓰기 자료를 수집하였다. 원활한 글쓰기 자료 수집을 위해 협약 대학 협의체를 구축하고, 정기적인 협의회 및 실무자 회의를 운영하고, 상시 소통을 위한 통로를 개설하여 체계적인 협력이 이루어지도록 하였다.

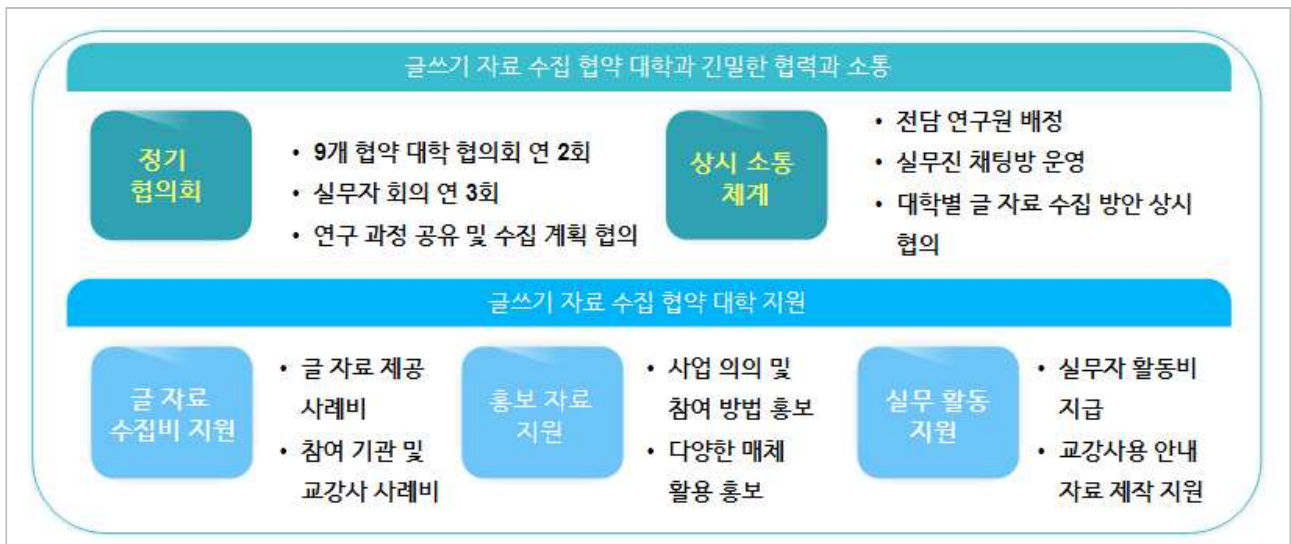
1) 대학 협력 체계 운영의 개요

연구진은 협약 대학과의 긴밀한 협업을 위해 대학 협의체를 구성하고, 대학 및 국어원과 상시 소통하기 위한 전담 연구원을 배정하였다. 대학 협의체는 각 대학의 의사결정 책임자인 대표자와 연구진 및 교강사와 직접 소통하며 실무를 담당하는 실무자로 구성되었다.

연구진은 대학 협력 체계를 통해 협약 대학과 긴밀히 소통하고 협력하였다. 우선 정기 협의회를 통해 연구 과정 공유 및 수집 계획을 협의하였다. 각 대학의 대표자와 실무자로 구성된 대학 협의회를 연 2회, 실무자로만 구성된 실무자 회의를 연 3회 운영하였다.

또한 상시 소통 체계를 구축하기 위해 소통을 위한 전담 연구원을 배치하여 이메일 및 단체 채팅방(SNS)을 활용하여 대학별 글쓰기 자료 수집 방안을 상시 협의하였다.

이 외에도 글쓰기 자료 수집의 실무를 담당하는 협약 대학에 글쓰기 자료 제공 사례비 및 교강사 사례비를 지급하고, 다양한 매체의 홍보 자료를 지원하였다. 또한 실무 활동을 지원하기 위해 실무자 활동비를 지급하고 교강사용 안내 자료를 제작하여 배포하였다.



[그림 Ⅲ-1] 글쓰기 자료 수집 협약 대학과의 협력 체계

2) 대학 실무자 회의

담당 연구진 및 협약 대학 실무자가 참여하는 ‘실무자 회의’는 3월, 7월, 11월에 각 1회씩, 총 3회 운영되었다. <표 Ⅲ-1>에 실무자 회의 운영 개요를 제시하였다.

<표 Ⅲ-1> 실무자 회의 운영 개요

차수	항목	세부 내용
제1차 실무자 회의	일시 및 장소	<ul style="list-style-type: none"> 일시: 3월 16일(목) 14:00-15:00 장소: 실시간 온라인 비대면 회의
	참석자	<ul style="list-style-type: none"> 담당 연구진, 국립국어원 담당 연구관 및 연구사, 협약 대학 실무자(A대학, B대학, C대학, E대학, G대학, H대학, K대학)
	주요 회의 내용	<ul style="list-style-type: none"> 연구 과제 소개: 연구 목적과 세부 과제, 대학 협의체 운영 방안 소개 글쓰기 자료 수집 안내: 태블릿 대여, 플랫폼 사용, 동의서 양식, 글쓰기 자료 수집 모델, 예산 지급 안내 1학기 계획 조사: 홍보 자료 수요, 학교별 1학기 수집 계획, 플랫폼 개발 요구 사항 조사
제2차 실무자 회의	일시 및 장소	<ul style="list-style-type: none"> 일시: 7월 3일(월) 15:00-16:00 장소: 실시간 온라인 비대면 회의
	참석자	<ul style="list-style-type: none"> 담당 연구진, 국립국어원 담당 연구관 및 연구사, 협약 대학 실무자(A대학, B대학, C대학, D대학, F대학, G대학, H대학, K대학)
	주요 회의 내용	<ul style="list-style-type: none"> 글쓰기 자료 수집 종료 현황 보고: 대학별 수집 글쓰기 자료 수, 문항 선택 현황, 사례비 책정 글쓰기 자료 수집상의 문제 상황 보고: 글쓰기 자료 사전 탑재 방식의 문제, 글쓰기 수행 및 과제 요건의 미통제, 문항 풀림 현상 2학기 운영 개선을 위한 협의: 문항 배정 방식 논의
제3차 실무자 회의	일시 및 장소	<ul style="list-style-type: none"> 일시: 11월 1일(수) 18:00-19:00 장소: 실시간 온라인 비대면 회의
	참석자	<ul style="list-style-type: none"> 담당 연구진, 협약 대학 실무자(A대학, B대학, C대학, D대학, E대학, F대학, G대학, H대학, K대학)
	주요 회의 내용	<ul style="list-style-type: none"> 글쓰기 자료 수집 현황 보고: 문항별 수집 현황, 수집 부족분 글쓰기 자료 수집상의 문제 상황 보고: 플랫폼 오류, 문항 풀림 현상 글쓰기 자료 수집 부족분에 대한 해결 방안 논의: 수집 기한 연장, 글쓰기 참여자 추가 모집

(1) 제1차 실무자 회의

제1차 실무자 회의는 3월 16일(목) 14:00-15:00에 실시간 온라인 비대면 회의를 통해 진행되었다. 주요 안건은 대학별 수집 모델 확인 및 플랫폼 기능에 대한 간략 소개 및 의견 청취이며, 연구진 외 국립국어원과 각 대학 실무자가 참석하였다. 부득이한 사정으

로 참석하지 못한 경우 회의 자료와 회의록을 전달하고 서면으로 의견을 수렴하였다. (이하 2, 3차 실무자 회의도 동일)

제1차 실무자 회의에서는 연구 과제를 소개하고 글쓰기 자료 수집의 방법을 안내한 후 1학기 글쓰기 자료 수집의 계획을 조사하였다. ‘연구 과제 소개’의 내용으로는 연구 목적과 세부 과제, 대학 협의체 운영 방안이 있었으며, ‘글쓰기 자료 수집 안내’의 내용으로는 태블릿 대여, 플랫폼 사용, 동의서 양식, 글쓰기 자료 수집 모델, 예산 지급이 있었으며, ‘1학기 계획 조사’의 내용으로는 홍보 자료 수요, 학교별 1학기 수집 계획, 플랫폼 개발 요구 사항 등이 있었다.

본 연구에서는 데이터 획득 및 정제 도구로서 플랫폼을 사용하였다. 온라인 플랫폼은 내장된 온라인 서명 플랫폼을 통해 각종 동의서를 손쉽게 체계적으로 관리할 수 있으며, 채점 도구와의 확장 가능성이 있다는 장점이 있다. 이러한 플랫폼에 접근하기 위해서는 개인 전자 기기가 있어야 하는바, 대학이 플랫폼 업체인 엔에스데블(NSDevil)로부터 일정 수량의 태블릿을 대여하여 필요한 글쓰기 참여자에게 제공할 수 있도록 하였다. 플랫폼 개발 및 운영에 대한 자세한 사항은 뒤에서 기술한다.

제1차 실무자 회의를 통해 본 과업에 참여하는 교강사의 수 및 글쓰기 자료 수집 시기 등 각 대학의 상황을 보다 정확하게 파악할 수 있었다. 또한 제1차 실무자 회의의 시기가 플랫폼이 개시되기 이전이기 때문에, 대학의 요구 사항을 청취하여 자료 수집을 위한 플랫폼 개발에도 반영하였다.

(2) 제2차 실무자 회의

제2차 실무자 회의는 7월 3일(월) 15:00-16:00에 실시간 온라인 비대면 회의로 진행되었으며, 주요 안건은 1학기 글쓰기 자료 수집 현황 보고 및 2학기 운영 협의이다. 제2차 실무자 회의에서는 글쓰기 자료 수집이 종료된 이후의 시기이므로 한 학기의 글쓰기 자료 수집을 정리하고 운영의 어려움을 공유하며 해결을 논의하였다. 1학기 ‘글쓰기 자료 수집 종료 현황 보고’ 내용으로는 대학별 수집 글쓰기 자료 수, 문항 선택 현황, 사례비 책정 등이 있었으며, ‘글쓰기 자료 수집상의 문제 상황 보고’ 내용으로는 글쓰기 자료 사전 탑재 후 동의서 작성 방식의 문제, 글쓰기 수행 및 과제 요건의 미통제, 문항 쓸림 현상 등이 있었다. ‘2학기 운영 개선을 위한 협의 사항’으로는 주로 문항 배정 방식에 대해 활발한 논의가 이루어졌다. 연구진은 연구 초기 글쓰기 참여자들이 문항을 선택할 수 없

도록 임의 배정을 통해 문항의 불균형을 방지하고자 하였으나, 1학기에 대학 측의 의견을 반영하여 교강사 또는 글쓰기 참여자들이 문항을 플랫폼에서 선택할 수 있도록 운영하였다. 2학기에도 문항 랜덤 배정 방식이 어렵다는 대학 측의 의견을 반영하여 글쓰기 자료 수집 계획 조사 후 문항의 불균형을 조정하는 중재안을 협의하였다.

(3) 제3차 실무자 회의

2학기에는 1학기와 동일 또는 유사한 방식으로 글쓰기 자료 수집이 진행되었으므로, 수집 전후로 각 1회씩 회의를 진행한 1학기와 달리 글쓰기 자료 수집이 이루어지는 중간 과정을 점검하기 위해 실무자 회의가 진행되었다. 제3차 실무자 회의는 11월 1일(수) 18:00-19:00에 실시간 온라인 비대면 회의를 통해 진행되었으며, 주요 안건은 글쓰기 자료 수집 현황 및 문제 사항 보고, 수집 부족분 대책 마련, 최종 보고회 일정 조사이다.

제3차 실무자 회의에서는 글쓰기 자료 수집 상황을 점검하고 글쓰기 자료 수집의 부족분이 발생한 것과 관련하여 문제의 해결을 논의하였다. 2학기 ‘글쓰기 자료 수집 현황 보고’ 내용으로는 문항별 수집 현황 및 수집 부족분 발생이 있었으며, ‘글쓰기 자료 수집 상의 문제 상황 보고’ 내용으로는 플랫폼의 오류 및 문항 쏠림 현상 등이 있었다. 글쓰기 자료 수집 부족분에 대한 대책 마련을 위해 수집 기한 연장 및 글쓰기 참여자의 추가 모집이 논의되었다.

3) 하계 대학 협의회

하계 대학 협의회는 2023년 8월 29일(화) 11:30-15:30에 대면 회의로 진행되었다.

<표 III-2> 하계 대학 협의회 운영 개요

항목	세부 내용		
일시	• 일시: 2023. 8. 29.(화), 11:30-15:30		
참석자	서울대학교 국어교육연구소	민병곤 책임 교수 외 연구진 12명	
	국립국어원	<ul style="list-style-type: none"> • 국립국어원 장소원 원장 • 기획연수부 박위진 부장 • 교육연수과 이은영 과장 • 교육연수과 황용주 학예연구관 • 교육연수과 박미영 학예연구사 	
	협약 대학	A대학	백수진(실무자)
		B대학	박고운(실무자)
		D대학	신용권(대표자), 박보연(실무자)
		F대학	윤석민(대표자), 박지은(실무자)
		G대학	조유영(대표자), 장정민(실무자)
		H대학	박재희(대표자), 김정인(실무자)
	K대학	장충덕(대표자), 손대익(실무자)	
주요 협의 내용	<ul style="list-style-type: none"> • 협력 방안 안내: 사업 추진 계획, 협약 대학 지원 사항, 2학기 글쓰기 문항 공개 등 • 협조 요청 사항 안내: 외국인 글쓰기 자료 수집 제외, 글쓰기 요건 준수 및 환경 통제 • 논의 사항 토의: 목표 수량 및 문항별 불균형 해소 방안 • 2학기 글쓰기 자료 수집 준비: 홍보 자료 및 태블릿 수요 조사 		

하계 대학 협의회에서는 1학기에 이루어진 글쓰기 자료 수집 현황을 정리하고, 2학기의 추가적인 글쓰기 자료 수집 협력 방안이 논의되었다. 연구진은 대학 협의회에서 2023년 사업 추진 계획 및 자료 수집, 협약 대학 지원 사항에 대해 안내하였다. 2학기의 글쓰기 자료 수집에 사용될 글쓰기 문항 역시 공개하였다. 또한 대학 협의회에서는 대학의 실무자뿐만 아니라 대표자가 참석하며, 연구진과 국어원이 모두 참석하였기에 대학의 협조 사항과 논의 사항에 대한 심도 있는 토의가 이루어졌다.

4. 협조 사항	자료 수집 시 유의사항
<p>➤ 학생 글쓰기 안내 사항</p> <ul style="list-style-type: none"> ▪ 외국인 학생 글자료는 최종 데이터 세트에 미포함: 강의별 안내 필요 ▪ 글쓰기 요건 준수: 글 분량 및 논제 준수, 제목 및 목차 미포함, 인사말 미포함 ▪ 글쓰기 환경 통제: 진단 평가로 진행(인터넷 검색 및 표절 금지) <p>※ 강의 운영 상 논제에 대한 학습이 진행되더라도 글 작성 단계에서 자료 참고 미허용</p>	
<p>➤ 온라인 플랫폼 이용 시 유의사항</p> <ul style="list-style-type: none"> ▪ 정확한 프로필 기입(전화번호 잘못될 시 2학기에는 지급하지 않음) ▪ 문항 번호 확인(1학기 문항 번호 불일치 사례 50건 이상) 	

[그림 Ⅲ-2] 글쓰기 자료 수집을 위한 협조 사항 안내

1학기의 경우, 외국인 글쓰기 자료를 예상하지 못하여 우리 국민 외에도 외국인까지 포함하여 글쓰기 자료가 수집되었다. 그러나 외국인 글쓰기 자료의 수집은 국민의 글쓰기 능력이라는 본 연구의 취지에 벗어나기 때문에 최종 데이터 세트에 포함하지 않았다. 따라서 이 부분에 대한 안내 필요를 논의하였으며, 수업의 과정에서 외국인 참여자들이 부당하게 제외되지 않으면서 수업이 운영될 수 있도록 각 대학의 상황을 확인하였다. 또한 여전히 글쓰기 요건을 준수하지 않거나 환경이 통제되지 않은 글쓰기 자료가 발견된다는 점에서 원데이터의 품질 관리를 위해 2학기 글쓰기 자료 수집이 시작되기 전에 연구진이 교강사 및 글쓰기 참여자가 참고할 수 있는 매뉴얼을 작성하여 배포하기로 결정하였다.

2학기의 목표 수집 수량은 2023년의 목표 수집 수량인 9,000건의 글쓰기 자료 중 1학기의 수집 수량을 제외한 4,820건으로 책정되었다. 문항별로 글쓰기 자료 수집 쏠림을 방지하기 위해 각 대학이 대략적으로 문항별 목표 수집 수량을 정하고, 연구진이 조정하는 방안이 논의되었다. 대학 협의회에서는 이 수량을 확인 및 조정하고, 대학의 의견을 연구진이 청취하여 반영하였다. 또한 2학기 대학별 홍보 자료 배부 및 태블릿 대여에 대한 안내가 이루어졌다.

2. 글쓰기 자료 수집 계획 수립

1) 글쓰기 자료 수집 모델 설계

(1) 1학기 운영: 이원화된 글쓰기 자료 수집 경로

본 연구에서는 글쓰기 참여자가 온라인 플랫폼상에서 직접 글을 작성할 수 있는 글쓰기 자료 플랫폼을 중심으로 글쓰기 자료 수집 체계를 확립하였다. 글쓰기 자료 수집 결과를 체계적으로 데이터베이스화하고 관리하기 위해서는 온라인상에서 디지털화된 자료로 글을 수집할 수 있는 온라인 플랫폼을 구축하는 것이 필수적이다. 글쓰기 자료 수집 온라인 플랫폼은 참여자들이 일관된 환경 가운데 글쓰기를 수행할 수 있도록 지원하는 장점이 있다. 또한 자료 수집 과정에서 글쓰기 참여자들이 연구 윤리 관련 사항을 확인하고, 저작권에 동의하는 절차를 비대면으로 진행함으로써 관리 전반을 용이하게 한다. 나아가 채점 플랫폼과 연동되는 데이터베이스화를 통해 종합적인 자료 수집 및 관리 체계를 구축할 수 있다.

그러나 1학기에는 연구진이 글쓰기 자료 수집 계획을 수립하기 전에 대학의 교강사별로 강의 기획 단계에서 자료 수집을 위한 학생들의 글쓰기 수행 방안을 개별적으로 수립한 이후 플랫폼 개발 완료 일정이 4월로 잡히면서, 온라인 플랫폼을 통해 직접 글을 쓰는 모델 외에 오프라인으로 글쓰기를 수행하는 이원화된 모델을 적용하게 되었다. 오프라인에서 글 작성을 먼저 수행한 협약 대학이 있었으며, 해당 협약 대학의 사정에 따라 글쓰기 참여자는 우선 수기 또는 디지털 문서로 글을 작성 후 추후 플랫폼에서 동의서를 작성하였다. 글쓰기 참여자가 오프라인으로 작성한 글은 실무자가 파일 형태로 수집한 후 연구진을 통해 업체에 전달하여 바로 플랫폼에 업로드되었다.

A대학과 D대학이 이러한 예외 사례에 해당하였다. B대학의 경우 오프라인 글 작성을 하였으나 학생들이 자신이 수기로 작성한 글쓰기 자료를 추후 스스로 타이핑하여 재제출하는 방식으로 운영하였으므로 실제적으로는 온라인 과제 제출 방식으로 운영된 것과 큰 차이가 없었다. 아울러 오프라인 방식으로 글쓰기 자료를 수집한 대학의 참여 교강사 중 플랫폼 업체를 통하지 않고 직접 플랫폼을 통해 글쓰기 자료를 올리기를 희망하는 경우를 위해 플랫폼 입력 필드의 보안 기능을 해제하고, 복사 및 붙여 넣기 기능을 통해 과

제처럼 글쓰기 자료를 제출할 수 있도록 하였다.

<표 Ⅲ-3> 글쓰기 자료 수집 이원화 모델

글쓰기 수행 환경	참여 방식	데이터베이스화 방안	동의서 제출 방안	글쓰기 자료 플랫폼 제출 확인
온라인	글쓰기 강의 기반 집합 참여	강의 시간 중 플랫폼에 접속하여 글쓰기 수행 후 제출	플랫폼에서 저작권 이양 동의서 전자 서명 후 제출	대학 코디네이터 확인 후 명단 전달 → 연구진 확인 후 답례품 발송
	글쓰기 강의 과제 수행 또는 개인별 참여	사전에 안내된 시간에 개인별로 자신의 디지털 도구를 활용하여 플랫폼에 접속, 글쓰기 수행 후 제출		
오프라인 (예외 사례)	글쓰기 강의 기반 집합 참여	수기 또는 디지털 문서로 작성 후 담당 교강사에게 디지털 문서 파일로 제출 → 플랫폼에 우선 탑재	자신의 글 확인 → 플랫폼에서 IRB 및 저작권 이양 동의서 전자 서명 제출	플랫폼을 통해 최종 제출자 확인 후 답례품 발송

<표 Ⅲ-4>는 1학기 글쓰기 자료 수집 모델별 대학을 정리한 표이다. 하나의 대학에서 다수의 교강사가 서로 다른 수집 모델을 선택할 수 있다는 점에서 대학은 복수로 표기되었다.

<표 Ⅲ-4> 글쓰기 자료 수집 모델별 참여 대학(1학기)

글쓰기 수행 환경	강의 기반 여부	집합 여부	참여 대학
온라인 플랫폼상에서 글쓰기	강의 기반	집합	A대학, D대학, F대학, G대학, K대학
		개별	A대학, B대학, D대학, E대학, F대학, G대학, H대학
	강의 무관	개별	C대학
오프라인 글쓰기	강의 기반	집합	A대학, D대학

1학기에는 글쓰기 강의 외 시간에 집합 참여 방식을 운영한 대학은 없었으며, 참여자가 게시 글 등의 홍보 자료를 통해 참여 의사를 밝히고 참여하는 글쓰기 강의 외 시간 개별 참여의 경우 통제의 어려움으로 인해 권장되지 않았다. 결과적으로 참여 교강사가

없어 관리 및 통제가 어려운 C대학만 강의 기반이 아닌 강의 외 시간 개별 참여 방식으로 글쓰기 자료 수집을 운영하였다.

(2) 2학기 운영: 수집 경로의 단순화

1학기에 참여 대학교의 다양한 상황을 고려하여 플랫폼 수집 경로 외의 예외적인 방식들을 수용한 결과 플랫폼에 글쓰기 자료를 재작성하거나 추후 동의서를 수집하는 경우 참여율이 떨어진다는 한계가 있었다. 특히 파일을 플랫폼에 사전 탑재한 경우, 글쓰기를 수행한 글쓰기 참여자 중 동의서를 작성하지 않는 비율이 높게 나타나 추가적인 관리 및 독려가 필요하였다. 또한 처음 파일을 수합할 때 글자 수 제한 등의 글쓰기 요건을 지키지 않은 파일들이 많아 이러한 파일들을 수동으로 정제하는 데 추가적인 시간 및 인력이 필요하였다.

반면 학생들이 강의실에 집합하여 온라인 플랫폼상에서 직접 글을 쓰게 하는 경우, 위와 같은 문제 상황이 발견되지 않았고, 글쓰기 상황 통제 및 자료 제출 확인이 용이하다는 점을 확인할 수 있었다. 따라서 온라인 글쓰기 수집 플랫폼 구축이 완료된 2학기에는 온라인 플랫폼을 통한 글쓰기 자료를 수집하는 일원화된 모델만을 운영하였다.

다만 온라인 참여 방식 중 과제 수행 등의 방식으로 개별적으로 참여할 경우 글쓰기 시간을 엄수한다거나 다른 자료를 참고하지 않는 등 글쓰기 환경을 통제하는 데 한계가 있다. 이에 2학기에는 글쓰기 자료 생성 환경의 일관성 확보를 위해서 강의 기반 집합 참여를 권장하였지만, 참여 교강사가 계획한 커리큘럼이 다양하다는 현실적인 한계로 인해 대부분의 협약 대학에서 집합 참여와 개별 참여 방식 모두 활용되었다.

B대학의 경우 1학기과 마찬가지로 강의 시간에 수기로 글쓰기 자료를 작성하고 학생이 개별적으로 추후 플랫폼에 재작성하는 방식을 유지하였다. 이 경우 참여자가 직접 온라인 플랫폼에서 글을 작성하는 기본 모델을 준수한다고 할 수 있으나, 수기 작성 참여자수에 비해 실제 플랫폼을 통해 수집된 글쓰기 자료의 수량이 현저히 적다는 문제점이 있었다. C대학의 경우 2학기에도 강의 내 홍보를 통해 자율 과제 참여 방식을 선택했으며 특정 강의에서만 홍보를 진행하였으므로 강의 기반 참여와 다름이 없이 운영되었다.

<표 III-5>는 2학기 글쓰기 자료 수집 모델별 대학을 정리한 표이다. 하나의 대학에서 다수의 교강사가 서로 다른 수집 모델을 선택할 수 있다는 점에서 대학은 복수로 표기되었다.

<표 Ⅲ-5> 글쓰기 자료 수집 모델별 참여 대학(2학기)

글쓰기 수행 환경	강의 기반 여부	집합 참여	참여 대학
온라인 플랫폼상에서 글쓰기	강의 기반	집합	A대학, D대학, E대학, F대학, G대학, K대학
		개별	A대학, B대학, C대학, D대학, E대학, F대학, H대학

2) 강의별 글쓰기 자료 수집 계획 협의

본격적인 글쓰기 자료 수집 전 자료 수집을 체계적으로 관리하기 위해 협력 대학 실무자 및 참여 교강사를 대상으로 강좌별 글쓰기 자료 수집 계획에 대해 조사하고, 일정 및 방안 등에 대해 협의하였다.

1학기에는 웹 기반 설문 조사를 실시하여 대학별 자료 수집 일정, 참여 방식, 교강사명, 강의명, 태블릿 대여 여부, 플랫폼 입력 방식, 참여 학생 수 등의 구체적인 사항을 확인하였다. 수업을 기반으로 한 자료 수집 계획은 각 교강사가 직접 입력하며, 전체적인 운영 상황 및 수업을 기반으로 하지 않은 자료 수집 계획은 각 대학의 실무자가 입력하였다. 또 플랫폼 개발에 대학 측 의견을 반영하기 위해 관련 내용을 ‘기타 요청 사항’에 기입하기를 요청하였다. 2학기에는 실무자 회의를 통해 글쓰기 자료 수집 플랫폼을 활용한 모델만을 운영하는 것으로 협의되었기에, 교강사명, 강의명, 강의 정원, 실시 일정, 참여 방식으로 조사 항목을 줄여서 조사를 시행하였다. 조사 결과 및 협의 내용은 다음과 같다.

(1) 글쓰기 자료 수집 일정

1학기의 경우 플랫폼 개발 이전에 자료 수집이 완료되었거나 시작되는 사례를 고려하여, 정확한 “자료 수집 일정”을 파악하여 대응할 필요가 있었다. 플랫폼 개발 전 수기로 글쓰기를 완료한 B대학의 경우 플랫폼이 개발된 후 글쓰기 참여자로 하여금 자율적으로 수기로 작성한 자신의 글을 플랫폼에 입력하여 제출하도록 협의하였다. D대학의 모든 강의와 A대학의 일부 강의에서는 컴퓨터 문서 파일의 형태로 글쓰기 자료를 이미 수집해

두었기에, 플랫폼 업체를 통해 글쓰기 자료를 플랫폼에 우선 업로드 후, 추후 글쓰기 참여자들이 플랫폼에서 자신의 글을 확인하고 저작권 이용 허락 계약서를 작성하는 방식으로 진행되도록 협의하였다.

2학기에는 예산 집행 기한으로 인해 글쓰기 자료 수집을 학기 말까지 연장할 수 없는 어려움이 존재하였다. 따라서 글쓰기 자료 수집의 일정을 조사하여 협약 대학과의 협의를 거친 후 예산 집행 기간을 고려하되 협약 대학의 사정을 최대한 반영하여 수집 기간을 설정하였다.

(2) 글쓰기 자료 수집 참여 방식

대학별 글쓰기 수집 환경의 다양성과 요구를 파악하기 위하여 대학별 글쓰기 자료 수집의 “참여 방식”을 조사하였다. 1학기의 조사 결과 참여 방식은 ①강의 기반 집합 참여, ②강의 외 시간 집합 참여, ③강의 기반 개별 참여(과제), ④강의 외 시간 개별 참여 중 ①의 강의 기반 참여가 가장 많았으며, 강의를 맡은 교강사 중 과반이 개별 과제로 진행할 계획이라고 답하였다. 또한 C대학의 경우 참여 교강사가 없기 때문에 참여의 의지가 있는 글쓰기 참여자들의 신청을 통해 강의 외 시간에 개별 참여하는 방식으로 계획하였다.

2학기의 참여 방식 모두 강의 기반으로 이루어졌으며 강의 내 집합 참여와 강의 외 과제 참여로 나눌 수 있었다. 환경의 통제를 위해 강의 내 집합 참여를 독려하였으나, 협약 대학 또는 교강사의 사정으로 인해 과제 참여 방식이 여전히 많이 선택되었다.

(3) 글쓰기 자료 수집 도구 지원

“태블릿 대여” 항목을 통해 태블릿 대여의 수량 및 시기를 미리 파악하고자 하였다. 강의 기반 글쓰기 자료 수집을 진행하는 교강사들의 태블릿 대여 수요는 약 30%였으며, 실제로 플랫폼을 통한 수집이 시작된 후에는 태블릿을 신청하였으나 사용하지 않는 교강사도 있었다. 태블릿의 수요가 적은 것은 강의를 기반으로 글쓰기 자료를 수집하기로 한 교강사들이 다수 수업 후 과제로 글쓰기를 수행하도록 하였으며 강의 내 집합 참여의 경우에도 개인 노트북이나 태블릿을 지참하는 글쓰기 참여자들이 많았기 때문이다. 1학기 와 2학기 모두 동일하게 A대학, G대학, K대학이 태블릿 대여를 신청하여 활용하였으며,

대여 태블릿의 총 수량은 2학기에 감소하는 양상을 보였다. 또한 “플랫폼 입력 방식” 조사 결과, 1학기에는 대학 중 자료 수집 시기나 환경으로 인해 다른 양식으로 글쓰기 자료를 수집하고 차후 글쓰기 참여자들이 자율적으로 플랫폼에 업로드하는 방식을 희망하는 교강사들도 있었다.

(4) 글쓰기 자료 수집 참여자 수

“참여 학생 수” 항목은 글쓰기 자료 수집 수량을 예측하기 위해 설정되었다. 물론 강의 외 자율 모집 참여 인원을 예측하기 어려우며, 강의 기반의 수집 모델을 선택하였더라도 강의 안에서 글쓰기 참여자의 자율 참여를 원칙으로 하는 만큼 정확한 참여 인원을 파악할 수 없다는 한계가 있었다. 이처럼 조사 시점에 참여 학생 수 예측이 어렵다는 실무자의 의견을 반영하여 2학기에는 “참여 학생 수” 대신 “강의 정원”을 조사하였다.

(5) 홍보 자료

“홍보 자료 요청” 항목을 두어 홍보물의 종류에 따른 수요를 파악하고자 하였다. 1학기의 수요 조사 결과 실물 포스터와 이미지 파일 요청이 많았으며, 추가적인 별도의 영상물에 대한 수요는 없는 것으로 파악되었다. 강의 시간에 글쓰기를 수행하는 경우, 진도 문제 등으로 홍보에 시간을 많이 할애할 수 없다는 의견이 있었다. 2학기에도 협약 대학의 요청에 따라 홍보 자료를 제공하였다.

(6) 기타

“기타 요청 사항” 중에서는 수업과 연계하여 운영하기를 희망하는 대학의 경우 특정 글쓰기 문항으로만 자료 수집을 하게 해달라는 요청이 다수 있었다. 1학기에는 이를 반영하여 글쓰기 참여자의 문항 선택을 교강사의 자율로 두었다. 그러나 이는 결과적으로 수집 자료의 문항 불균형 문제의 원인이 되기도 하였다. 이에 따라 2학기에는 수집 자료의 문항 불균형 문제를 방지하기 위해 글쓰기 자료 예상 수집 규모 파악 후, 문항별 권장 수집 규모를 제공하고자 계획하였다.

2학기 글쓰기 자료 수집에서 중요한 것은 문항별 수집 글쓰기 자료 수의 불균형을 해

결하는 것이기 때문에 2학기의 자료 수집 계획 조사는 문항별 권장 수집량을 책정하기 위한 조사의 의미를 가진다. 문항별 권장 수집량을 책정하기 위한 대학별 수집 규모는 각 대학의 강의 정원을 참고하였다. 플랫폼 입력 이외의 입력 방식을 운영하지 않는 관계로 플랫폼 입력 방식은 조사하지 않고 참여 방식만을 조사하였다. 다음은 2학기 글쓰기 자료 수집 계획을 바탕으로 수립한 문항별 목표 수집량이다. C대학의 경우 의무 참여 교강사가 없기 때문에 목표 수량을 책정하지 않고 진행하였다.

<표 III-6> 2학기 글쓰기 자료 수집 계획

	A대학	B대학	C대학	D대학	E대학	F대학	G대학	H대학	K대학	목표 수집량 총합
Q4		250		260	150					660
Q5		250		170	200					620
Q6		300		210	150					660
Q7	127			300	110	100		183	200	1,020
Q8	227					210	150	233	200	1,020
Q9	246					290	150	234	100	1,020
총	600	800	-	940	610	600	300	650	500	5,000

3. 글쓰기 자료 수집 플랫폼 개발 및 운영

1) 글쓰기 자료 수집 플랫폼 개발의 목적과 기능

(1) 글쓰기 자료 수집 플랫폼 개발의 목적

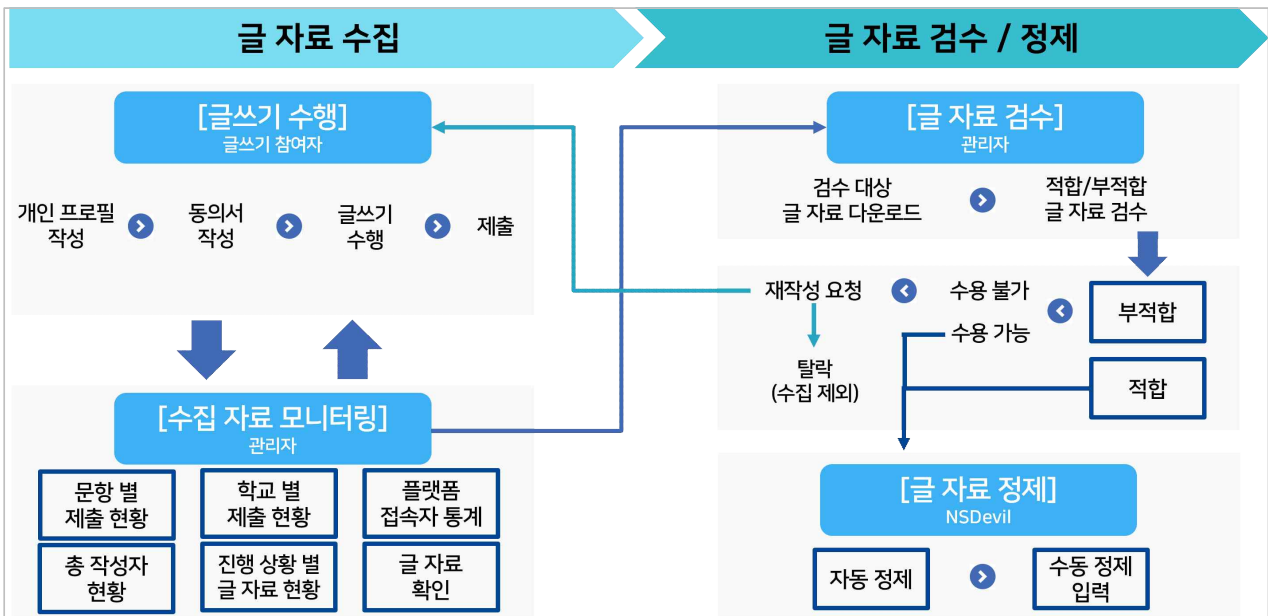
본 연구의 글쓰기 자료 수집 플랫폼 개발 목적은 두 가지로 기술할 수 있다. 첫째, 대규모 데이터의 효율적인 수집 및 관리이다. 글쓰기 자료 데이터는 9,000건의 수집을 목표로 하였으며, 체계적인 수집 및 관리를 위해 글쓰기 자료 외의 기타 정보를 함께 수집하고자 했다. 글쓰기 참여 인원과 수집되는 정보의 부피를 고려할 때 매우 큰 규모의 데이터가 수집될 것으로 예상하였으며, 이러한 데이터를 웹에서 접근할 수 있는 플랫폼을 구축해 관리함으로써 그 데이터 사용의 효율성 및 안정성을 확보하고자 하였다. 둘째, 글쓰기 자료 수집 플랫폼은 일정 수준의 글쓰기 수행 환경 통제에 기여한다. 글쓰기가 이루어지는 환경에 대한 제약을 고려하지 않은 채 글쓰기 자료를 수집한다면 균일하지 못한 데이터 수집의 위험이 존재한다. 다양한 협약 대학 소속의 글쓰기 참여자로부터 자료를 수집하기에 글쓰기 수행 환경의 엄격하고 완벽한 통제는 어려우나, 플랫폼 내 글쓰기 수행 제한 시간 및 글자 수 제한 등을 설정하여 수집되는 글쓰기 자료가 일정 정도의 공통성을 유지할 수 있도록 하였다.

(2) 글쓰기 자료 수집 플랫폼의 기능

양질의 글쓰기 자료를 구축하기 위해선 원활한 글쓰기 자료 수집과 동시에 수집된 글쓰기 자료에 대한 체계적인 관리가 필요하다. 이를 위해 본 연구에서는 플랫폼 개발 업체와의 지속적인 논의와 연구진 내부 검토를 거쳐 글쓰기 자료 수집 기능과 글쓰기 자료 수집 관리 기능을 모두 포함한 글쓰기 자료 수집 플랫폼을 개발하였다. 글쓰기 자료 수집 기능에서는 글쓰기 참여자가 글을 작성하는 전 과정에 필요한 제반 기능을 제공하며, 글쓰기 자료 수집 관리 기능에서는 관리자가 자료 수집 현황을 파악하고, 수집된 글쓰기 자료를 다운로드할 수 있도록 기능을 구성하였으며, 중복된 글쓰기 자료 선별을 위한 유사도 검사 기능 및 자동 정제 기능을 탑재하였다.

<표 III-7> 글쓰기 자료 수집 플랫폼의 기능

	글쓰기 자료 생성		글쓰기 자료 수집 및 관리
주체	글쓰기 참여자		연구진 플랫폼 개발 업체(NSDevil) 협약 대학 실무자 및 교강사
기능	글쓰기 전 단계	개인 프로필 작성 동의서 작성	학교별·문항별 글쓰기 자료 수집 현황 글쓰기 자료 수집 단계별 분류 기준 항목별 정렬 글쓰기 자료 다운로드 유사도 검사 자동 정제
	글쓰기 단계	글쓰기 문항 선택 제한 시간 표시 글쓰기 중 임시 저장	
	글쓰기 완료 단계	작성 글 다운로드	

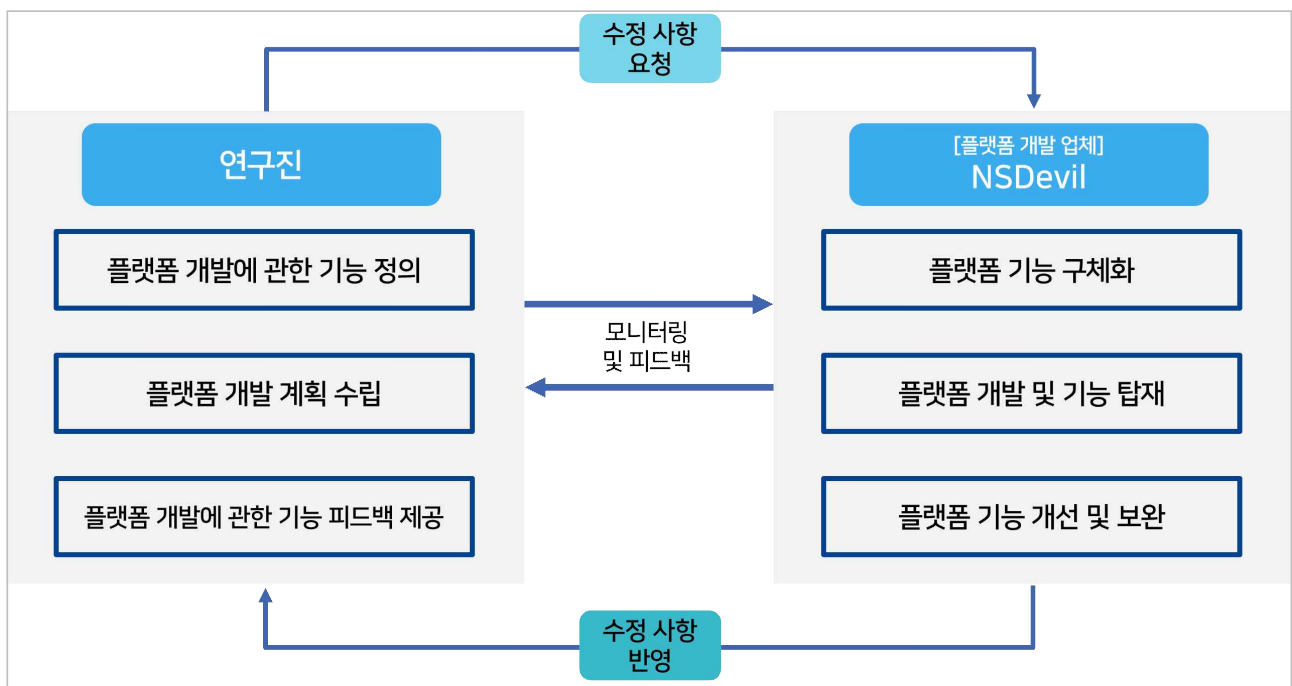


[그림 III-3] 글쓰기 자료 수집 플랫폼 기능 구조도

2) 글쓰기 자료 수집 플랫폼 개발 과정

본 연구에서는 글쓰기 자료에 대한 체계적인 수집과 관리, 채점 및 데이터 정제 등의 기능을 연동할 수 있는 웹 플랫폼을 개발하였다. 연구진은 플랫폼 개발 계획을 수립하고, 업무 협약 대학으로부터 원활하게 글쓰기 자료가 수집될 수 있도록 웹 기반 교육 및 평

가 도구 전문 업체와 계약을 체결하여 구체적인 플랫폼 기능 정의 및 실제 기능 구축에 관한 제반 사항을 전달하고 관리하는 업무를 담당하도록 하였다. 글쓰기 자료 수집 플랫폼은 다음과 같은 세 단계에 따라 개발되었다. 첫째, 연구에 필요한 플랫폼 개발에 관한 기능의 정의를 확정한다. 둘째, 플랫폼 개발 업체에 해당 기능 정의를 전달하여 구현 여부 및 활용 양상에 대한 대략적인 범위를 확정한다. 셋째, 플랫폼 개발 업체가 연구진이 전달한 기능을 구체화하여 플랫폼을 개발하고, 연구진은 이를 검토하여 실제 글쓰기 자료 수집 데이터 구축에 유효한지를 점검한다. 이 모든 과정에는 연구진과 플랫폼 개발 업체의 지속적인 소통과 모니터링 및 피드백이 전제되어 있으며, 플랫폼 개발 업체는 연구진이 요청한 수정 사항을 반영하여 플랫폼 기능을 보완하는 작업이 이루어졌다.



[그림 Ⅲ-4] 글쓰기 자료 수집 플랫폼의 개발 과정

3) 글쓰기 자료 수집 플랫폼의 실제

본 절에서는 글쓰기 자료 수집 플랫폼을 이용하는 실제 모습을 보여주고자 한다. 글쓰기 자료 수집 플랫폼은 ①글쓰기 자료 생성 플랫폼과 ②글쓰기 자료 수집 관리 플랫폼으로 나누어지는데, ①글쓰기 자료 생성 플랫폼은 글쓰기 참여자들이 계정을 부여받아 할

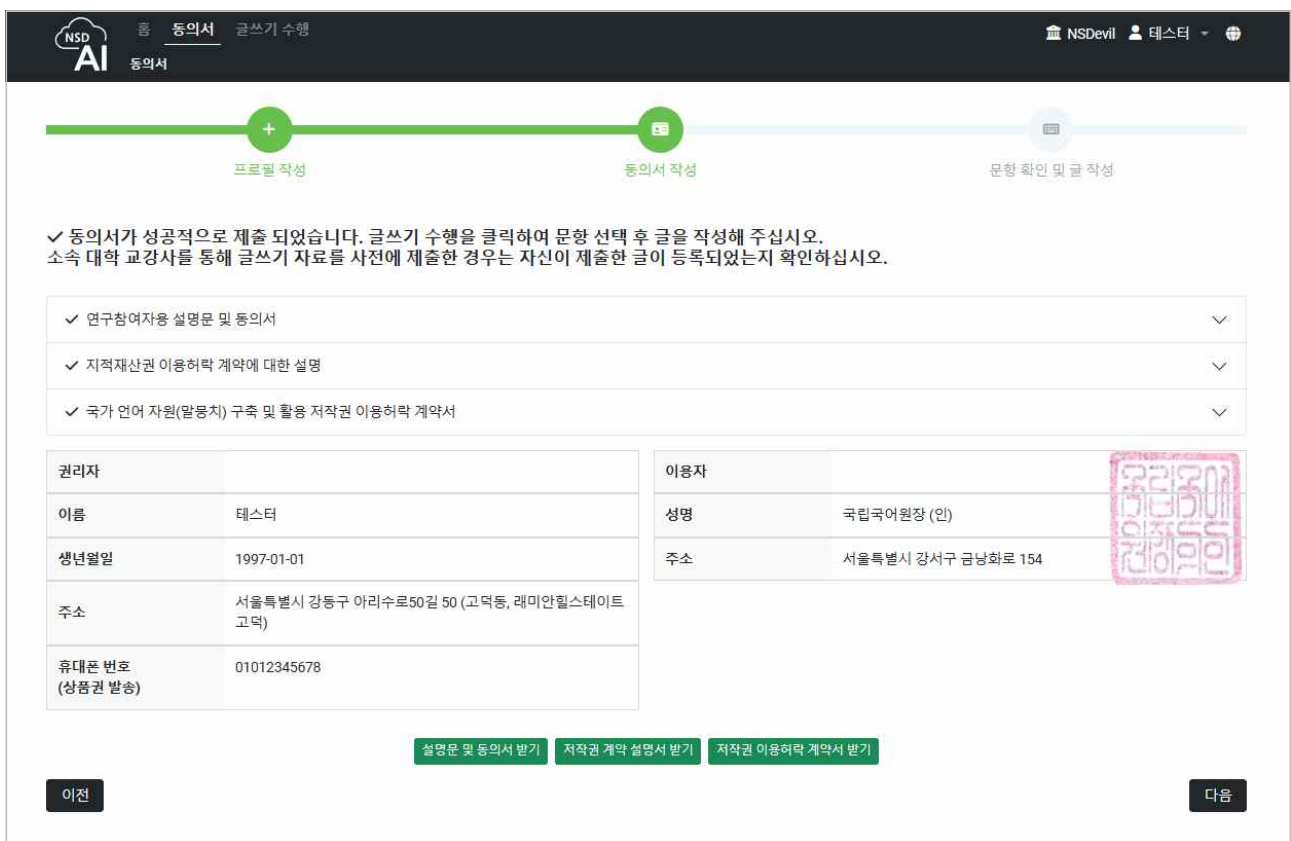
당된 문항 번호에 맞게 글쓰기를 수행함으로써 글쓰기 자료를 수집하는 플랫폼이다. ② 글쓰기 자료 수집 관리 플랫폼은 연구진 및 플랫폼 개발 업체, 협약 대학 실무자 및 교강사가 접근할 수 있는 플랫폼으로 글쓰기 자료 수집 현황을 파악하고 관리하는 기능을 담당한다. 이상 두 가지 글쓰기 자료 수집 플랫폼의 실제적 운용 양상 및 특징에 대해서는 아래에서 자세히 서술하도록 한다.

(1) 글쓰기 자료 생성

글쓰기 자료 작성에 앞서 체계적이고 윤리적인 글쓰기 자료 수집을 위해 개인 프로필 생성 및 동의서 작성 단계를 구성하여 글쓰기 자료 수집 현황을 파악하고 문제 상황 발생 시 원활한 대처가 가능하도록 하였다. 글쓰기 참여자가 플랫폼에 접속하면 글을 작성하기 전, 개인 프로필을 생성할 수 있도록 하였다. 기본 정보(이름, 생년월일, 성별), 학교 정보(학교명, 단과대학, 학년), 연락처 정보(이메일, 휴대폰 번호)를 입력하고 내용 확인 후 [저장 후 다음] 버튼을 클릭하면 다음 단계로 이동할 수 있다.

[그림 Ⅲ-5] 글쓰기 자료 수집 플랫폼-프로필 정보 입력 화면

개인 프로필 생성이 완료된 이후엔 본격적인 글쓰기 시작에 앞서, 글쓰기 참여자가 자신이 작성한 글을 연구에 활용하는 것에 대한 동의서를 작성할 수 있도록 하였다. [연구 참여자용 설명문 및 동의서]와 [지적재산권 이용허락 계약에 대한 설명], [국가 언어 자원(말뭉치) 구축 활용 및 저작권 이용허락 계약서]의 내용을 확인한 후 주소 및 상세 주소를 입력하고 서명하면 동의서 작성이 완료되며, 제출된 동의 서류를 내려받을 수 있도록 하였다.



[그림 Ⅲ-6] 글쓰기 자료 수집 플랫폼-동의서 작성 화면

글쓰기 참여자는 여러 개의 문항 목록 중에서 자신이 작성하고자 하는 문항 번호를 선택할 수 있으며, 한 번 선택한 문항은 변경할 수 없도록 하였다. 문항 번호를 선택한 후 명에 서약서 내용을 확인하면 글쓰기를 수행할 수 있다.

글쓰기를 시작하면 팝업을 통해 글 작성이 시작되었다는 안내 문구를 확인할 수 있으며, 90분의 타이머가 작동하도록 설정하여 글쓰기 참여자가 남아 있는 시간을 쉽게 확인

할 수 있도록 하였다. 글쓰기를 수행하며 글의 주제와 배경에 대한 설명을 수시로 확인할 수 있도록 답안 작성란 위에 해당 내용을 삽입하였다. 글쓰기 참여자는 글을 작성하는 중간에 [임시 저장] 기능을 활용할 수 있는데, 불가피한 이유로 글 작성을 중단해야 할 경우 [임시 저장] 버튼을 클릭하면 작성 중 상태로 글이 저장되며, 글 작성을 다시 시작할 때 저장된 상태의 글을 불러와 이어서 작성할 수 있도록 하였다. 제한 시간 역시 이에 맞추어 조정될 수 있도록 기능을 구성하였다.



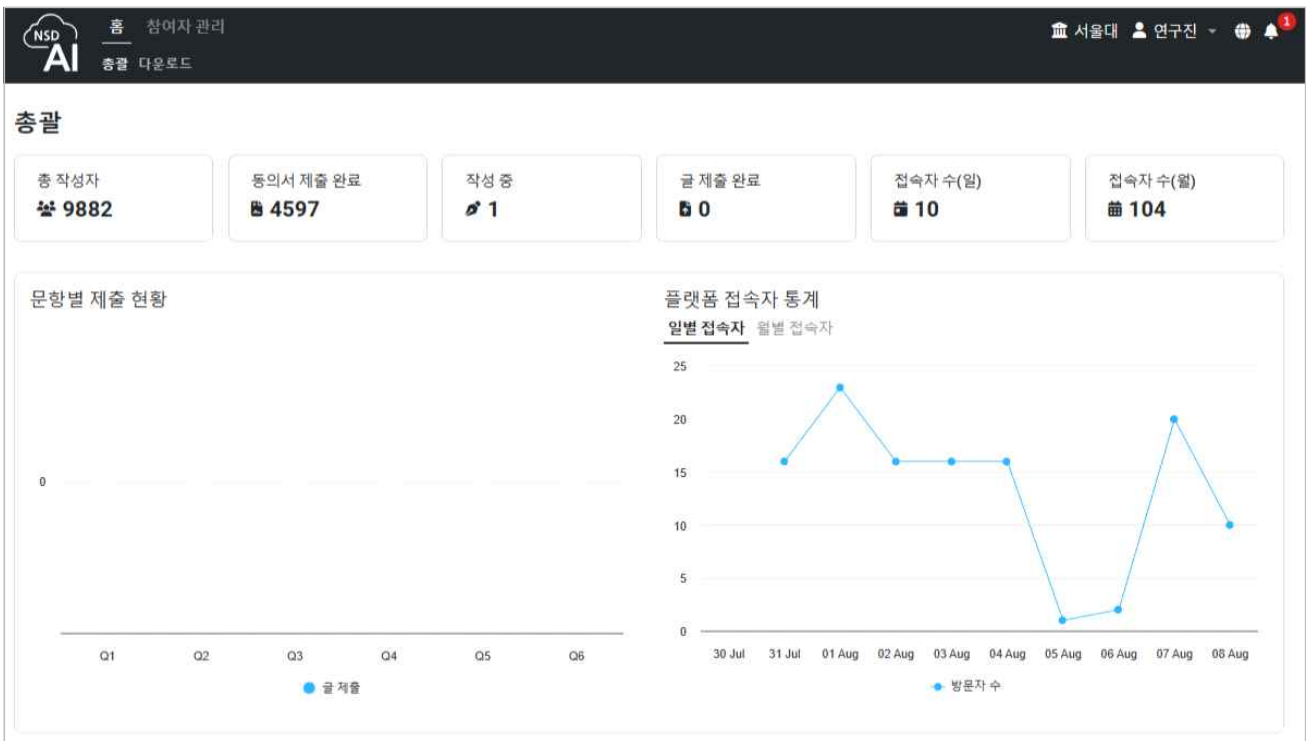
[그림 Ⅲ-7] 글쓰기 자료 수집 플랫폼-글쓰기 수행 화면

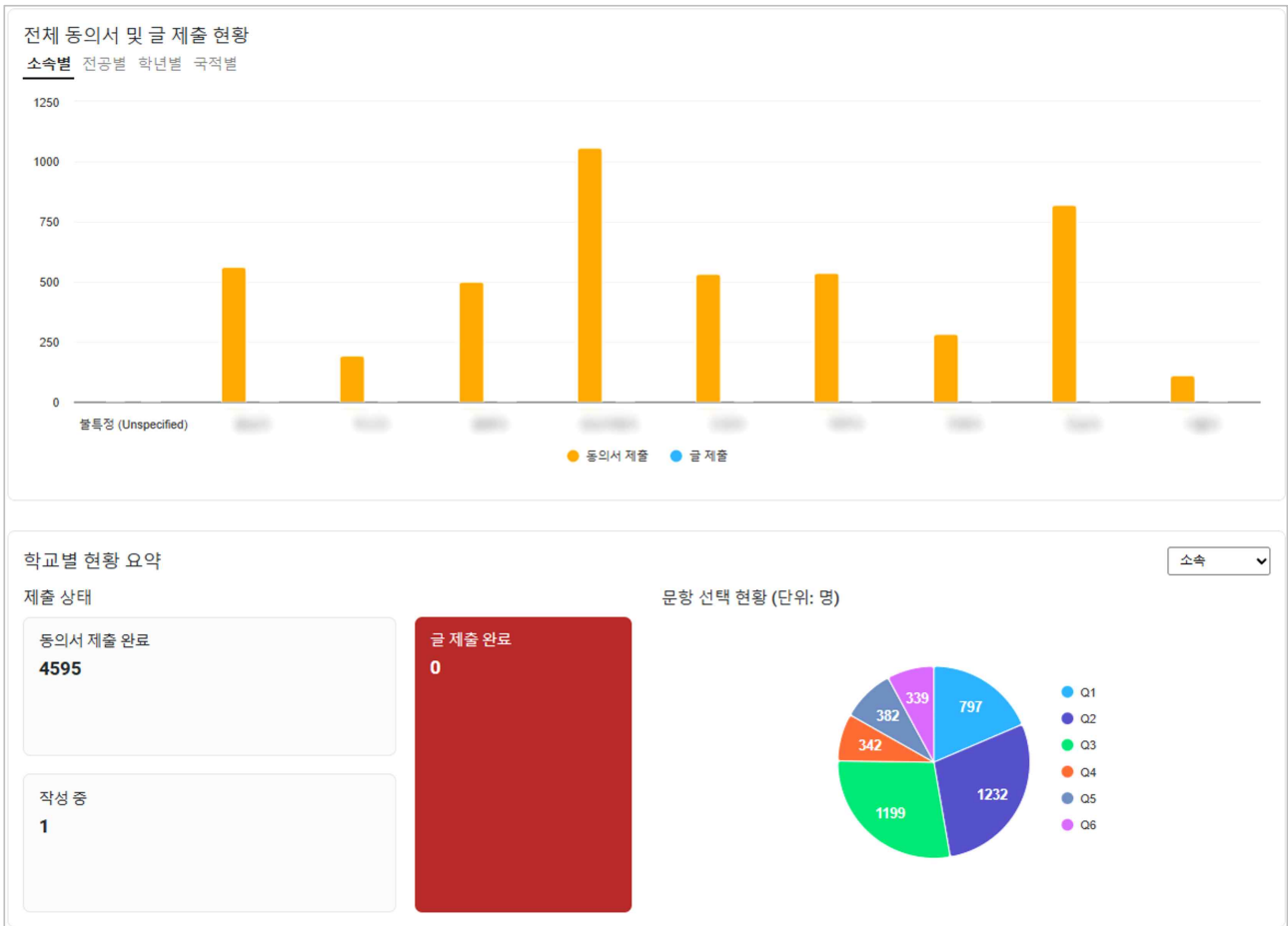
글 작성이 완료되면 [제출하기] 버튼을 클릭해 작성된 글을 제출할 수 있으며, 글 제출 완료 안내 메시지를 확인 후 [나가기] 버튼을 클릭하면 글쓰기 자료 수집이 최종 완료된다. 글쓰기 참여자는 제출된 글을 다시 확인할 수 있으며 필요시 작성 글 유출 금지 메시지 팝업을 확인 후 자신이 작성한 글을 내려받을 수 있도록 글쓰기 자료 다운로드 기능을 탑재하였다.



[그림 Ⅲ-8] 글쓰기 자료 수집 플랫폼-글쓰기 완료 화면

(2) 글쓰기 자료 수집 관리





[그림 III-9] 글쓰기 자료 수집 관리 플랫폼-전체 현황 요약 화면

관리자 계정으로 접속할 시 글쓰기 자료 수집 현황을 한눈에 파악할 수 있도록 메인 화면을 구성하여 글쓰기 자료 수집 상황에 따라 빠르고 정확한 대처가 가능하도록 하였다. 학교별 제출 현황 및 문항별 제출 현황을 파악하여 추가 수집이 필요한 대학에 글쓰기 참여 독려 요청을 하거나, 자료 수집량이 부족한 문항에 대하여 해당 문항 선택을 독려하는 용도로 활용하였다. 플랫폼 접속자 통계를 확인하여 글쓰기 자료 수집 진행률 및 속도를 파악하였고, 문항 선택 현황을 원형 그래프로 나타내어 문항 쏠림 현상을 진단하였다.

번호	ID	소속	동의서	문항	글자 수	최초 시작 시간	마지막 저장 시간	상태	정제본 수정 여부	선택
1	A0001		제출	Q3	907	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
2	A0002		제출	Q2	854	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
3	A0003		제출	Q3	861	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
4	A0004		제출	Q3	1139	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
5	A0005		제출	Q2	973	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
6	A0006		제출	Q2	974	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
7	A0007		제출	Q3	850	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
8	A0008		제출	Q1	1003	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
9	A0009		제출	Q2	810	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
10	A0010		제출	Q3	993	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
11	A0011		제출	Q3	948	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
12	A0012		제출	Q3	997	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기
13	A0013	미입력	미제출		0				X	
14	A0014		제출	Q2	1190	2023-04-27 15:18:32	2023-04-27 15:18:32	검수 완료	0	보기

[그림 Ⅲ-10] 글쓰기 자료 수집 관리 플랫폼-개별 제출 현황 화면

관리자 플랫폼에서는 개별 참여자의 글 작성 진행 상황을 즉각적으로 확인할 수 있도록 분류 및 정렬 기능을 탑재하였다. 글 작성 여부와 상관없이 플랫폼에 접속이 가능하도록 생성된 전체 ID를 [총 작성자] 탭에서 확인할 수 있으며, 글 작성이 이루어진 ID의 경우 해당 ID를 사용한 글쓰기 참여자의 소속, 동의서 제출 여부, 선택한 문항 번호, 작성된 글쓰기 자료의 글자 수, 글 작성 시작 시간, 글 작성 종료 시간, 글 작성 진행 상태, 정제본 수정 여부를 확인할 수 있도록 하였다.

[작성 중]은 글쓰기를 시작하였으나 글 작성이 완료되지 않아 제출하지 않은 상태를 의미하며, [글 제출 완료]는 글 작성이 완료되어 제출까지 이루어진 상태를 의미한다. [검수 중]은 제출이 완료된 글쓰기 자료 중 연구진이 검수를 진행하고 있는 상태를 의미하고, [검수 완료]는 제출이 완료되었으며 연구진의 검수 또한 진행되어 최종적으로 수집된 글쓰기 자료를 의미한다. [최종 보류]는 글 제출은 완료되었으나 검수 결과 글에 이상이 있

어 연구에 활용할 글쓰기 자료로서 수집 여부가 보류된 상태를 의미한다.

번호	ID	소속	문항	글자 수	최초 시작 시간	마지막 저장 시간	상태
1	D0546		Q2	851/1200	2023-05-09 14:00:41	2023-05-09 14:04:50	검수 완료
2	G0021		Q5	998/1200	2023-05-08 23:03:05	2023-05-08 23:37:15	검수 완료
3	G0777		Q6	1147/1200	2023-05-04 22:56:12	2023-05-04 22:58:53	검수 완료
4	E0510		Q2	1170/1200	2023-05-11 09:03:06	2023-05-11 09:03:37	검수 완료
5	G0510		Q5	1001/1200	2023-05-09 10:24:12	2023-05-09 10:39:08	검수 완료
6	G0200		Q6	947/1200	2023-05-10 14:21:11	2023-05-10 15:08:17	검수 완료
7	A0603		Q6	1063/1200	2023-05-04 23:36:31	2023-05-05 00:02:46	검수 완료
8	K0207		Q2	814/1200	2023-05-03 15:12:32	2023-05-03 15:38:48	검수 완료
9	K0209		Q2	1205/1200	2023-05-04 15:05:30	2023-05-04 15:53:13	검수 완료
10	A0152		Q6	810/1200	2023-05-08 16:14:40	2023-05-08 16:28:12	검수 완료

[그림 Ⅲ-11] 글쓰기 자료 수집 관리 플랫폼-제출 자료 다운로드 화면

필요시 관리자가 선택한 글을 내려받아 일괄적으로 확인할 수 있는 기능을 탑재하였다. [검색 항목 전체 다운로드] 버튼을 클릭하여 플랫폼 내에 업로드된 글쓰기 자료 전체를 다운받을 수 있으며, 특정 학교 및 특정 문항만을 선택하여 다운받을 수 있는 기능까지 포함하였다. 화면에 나타난 글쓰기 자료 목록 중 특정 글쓰기 자료만 선택한 뒤 [선택 항목 다운로드] 버튼을 클릭하면 해당 글쓰기 자료만 내려받을 수 있다. 다운로드 기능을 통해 글쓰기 참여자의 동의서, 글쓰기 참여자 인적 사항, 글쓰기 참여자가 작성한 글쓰기 자료를 내려받을 수 있다. 또한, 다운로드 시 작성된 글쓰기 자료뿐만 아니라, 글쓰기 자료를 작성한 ID, 대학명, 문항 번호, 글자 수, 글 작성 시작 시간, 글 작성 종료 시간, 글 작성 진행 상태, 동의서 작성 여부, 동의서 작성 시간 등이 일괄 받아지도록 기능을 구성하였다.

또한 중복 혹은 표절 글쓰기 자료를 선별할 수 있도록 유사도 검사 기능을 탑재하였다. 들

여쓰기 및 문단별 줄 바꿈 등 형식상 문제가 있는 글의 경우 이를 자동으로 포착하여 수정해 주는 자동 정제 기능을 탑재하여 안정된 형태로 글쓰기 자료를 수집할 수 있도록 하였다.

4) 플랫폼 운영상의 제약 개선

글 작성 완료 시기와 동의서 작성 완료 시기가 다를 경우, 이를 각각 명확히 확인하기 어려운 상황이 발생하였다. 동의서 작성이 완료되어야 이후 채점 글쓰기 자료로 활용이 가능하므로, 동의서 작성이 되지 않은 것으로 확인되어 채점 글쓰기 자료에서 배제하였으나 추후에 동의서 작성이 이루어져 뒤늦게 채점 글쓰기 자료로 포함하는 과정이 있었다. 문제 해결을 위해 업체에 요청하여 ‘글 작성 시작 시간’, ‘글 작성 종료 시간’ 외에 ‘동의서 작성 시간’을 확인할 수 있는 기능을 추가로 탑재하였다.

또한 문항 번호 선택 화면의 오류로 글쓰기 참여자들이 잘못된 문항 번호를 선택하여 글을 제출한 경우가 있었다. 강의에서 배정된 문항이 있기 때문에 글쓰기 참여자가 부주의하게 다른 문항 번호를 선택하고 수업에서 배정된 문항에 대한 답을 제출한 경우도 있겠으나, 문항 번호 선택 화면의 오류도 빈번하게 발견되었다. 문항 번호 선택 화면에서 특정 문항이 뜨지 않는 오류가 발견되었으며, 강의 기반 집합 수집 또는 과제 수집의 경우 시간이 제한되어 있거나 도움을 요청하여 바로 기능을 수정하지 못한다는 점에서 다른 문항을 우선 선택하고 글을 작성한 것으로 파악된다. 문항과 글쓰기 자료가 불일치하는 사례는 연구진에서 검수 과정을 거치며 글 내용에 맞는 문항 번호로 수정하여 입력하였고, 문제 상황을 업체에 알려 문항 번호 선택에서 오류가 나타나지 않도록 해당 기능을 보완하였다.

마지막으로, 글자 수 미달 및 초과 글쓰기 자료가 발생하였다. 연구에 활용할 수 있는 글쓰기 자료의 조건으로 글자 수를 1,000자 내외로 설정하고 그 범위를 800자 이상 1,200자 미만으로 한정하였다. 플랫폼 내에서 글쓰기를 수행할 시 800자에 미달되거나, 1,200자를 초과하는 경우 글쓰기 자료 제출이 안 되도록 설정하였지만, 띄어쓰기 중복 및 자동 정제 기능으로 인해 800자 미만 글쓰기 자료와 1,200자 초과 글쓰기 자료가 발생하였다. 또한 모바일 환경에서 ‘복사-붙여 넣기’ 기능을 사용할 시 제출을 위한 글자 수 제한 기능이 활성화되지 않는 것이 발견되었다. 이미 발생한 오류는 연구진이 검수 및 정제 과정을 거치며 글자 수를 조정하였으며 업체에 요청하여 해당 기능을 보완하였다.

4. 글쓰기 자료 수집을 위한 홍보 자료 개발

본 연구에서 글쓰기 자료 수집은 다양한 수준의 글쓰기 자료를 대규모로 구축하는 것을 목표로 한다. 이를 위해서는 체계적인 홍보를 통해 다수의 글쓰기 참여자를 모집할 필요가 있다.

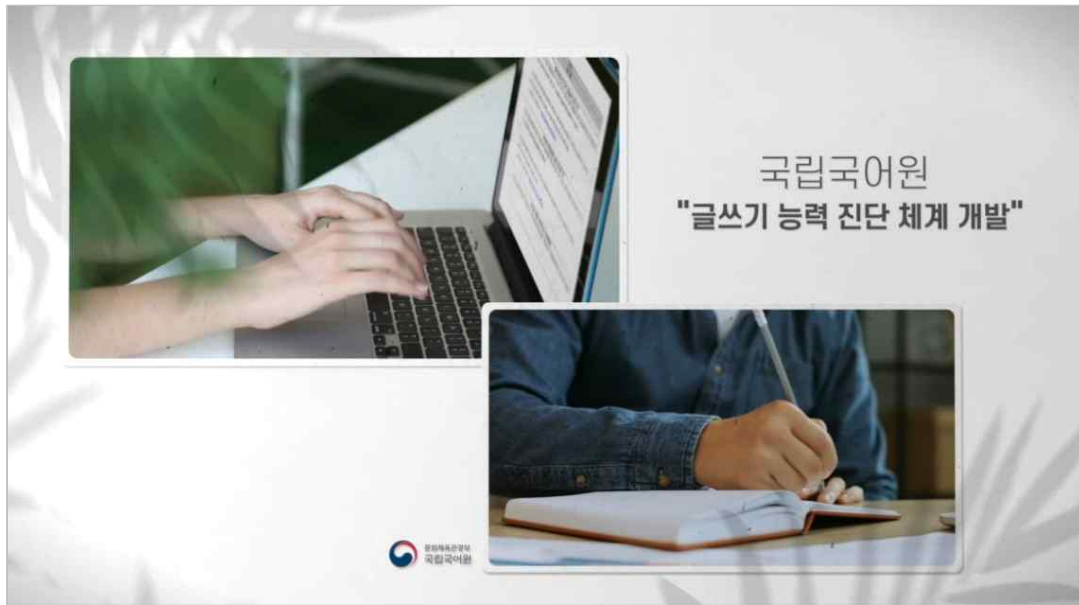
대다수의 협약 대학은 참여 교강사가 강의 시간에 다양한 홍보 자료를 활용하여 글쓰기 자료 수집 참여를 독려했다. 이에 따라 1학기에는 홍보에 대한 학교별 수요 조사를 통해 맞춤형 홍보 방안을 수립하였다. 디지털 홍보 자료에 대한 수요가 컸으며, 다양한 매체 환경 및 이용 문화를 고려하여 디지털 홍보 자료를 개발하였다. 홍보 자료는 각각의 특성과 목적을 고려하여 다양한 형태로 개발하여 협약 대학의 다양한 수요를 충족하고자 하였다.

그러나 C대학의 경우 의무 참여 교강사가 따로 존재하지 않기 때문에 이러한 홍보가 불가능하다는 어려움이 있었다. 따라서 대학 실무자가 직접 다수의 강의에서 글쓰기 자료 수집에 대한 홍보와 안내를 직접 진행하기도 하였다. 이때 학생의 아이디 부여 및 플랫폼 접속을 원활하게 하기 위해 실무자가 특별히 제작한 카드 명함이 사용되기도 하였다. 이 카드 명함에는 아이디 및 플랫폼 링크가 적혀 있어, 참여를 원하는 학습자들이 쉽게 글쓰기 자료 수집 플랫폼에 접근할 수 있다.

그 외에는 모두 참여 교강사가 자신의 수업에서 국립국어원 또는 연구진이 제작하여 배포한 홍보 자료를 공통적으로 사용하여 홍보하였다. 다음은 각각의 홍보 자료이다.

1) 글쓰기 능력 진단 체계 홍보 동영상

글쓰기 자료의 수집이 진행되기 전에 참여 대학의 학생들에게 글쓰기 자료 수집의 취지와 의의를 충분히 알릴 필요가 있다. 이를 위해서는 국립국어원이 제작하고 배포한 글쓰기 진단 체계 개발 연구의 주요 내용과 의의를 홍보하는 동영상을 활용하였다.



[그림 Ⅲ-12] 글쓰기 능력 진단 체계 개발 홍보 동영상

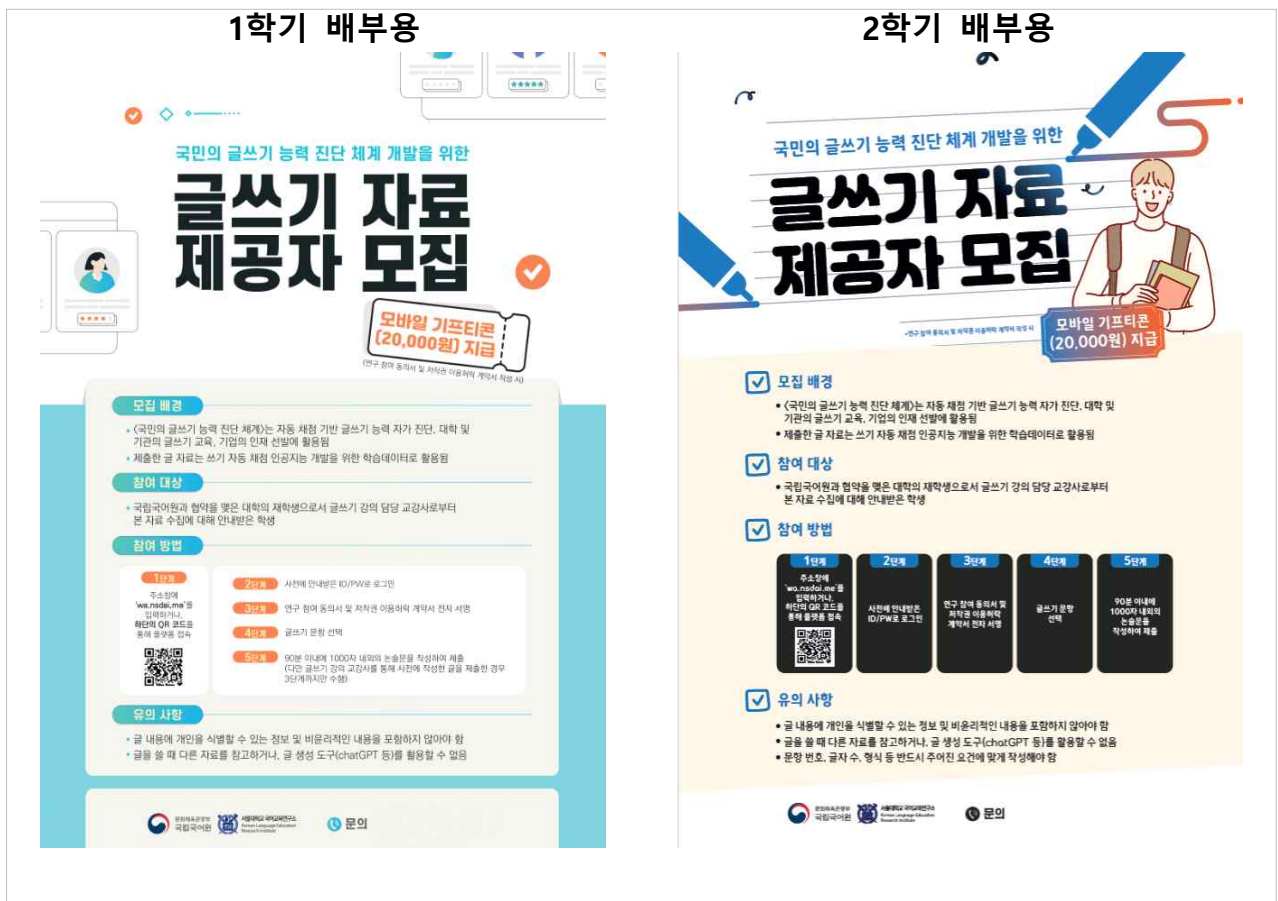
홍보 동영상은 짧은 시간 내 많은 정보를 전달할 수 있으며, 시청자의 흥미를 끄는 점에서 연구를 알리는 초기 단계의 홍보 수단으로 가장 적절하다. 홍보 동영상은 학기 초 글쓰기 참여자들에게 연구의 취지를 소개하는 자료로서 강의 중 상영할 수 있도록 실무자를 통해 대학 교강사들에게 전달되어 수업 중 활용되었다.

2) 글쓰기 참여 안내 포스터 및 리플릿

글쓰기 참여를 독려하면서 참여 방법을 안내하기 위해 포스터 및 리플릿을 제작하였다. 글쓰기 참여자들은 글쓰기 자료 수집 과정에서 개인정보 활용 동의서 및 저작권 이용 허락 계약서를 작성해야 하며, 이에 따라 필수적으로 플랫폼을 이용해 동의서를 작성 및 제출해야 한다. 홍보 동영상은 연구의 취지와 의의를 홍보하지만, 구체적인 참여 방법을 포함하지 않고 있다. 이에 글쓰기 참여자들에게 플랫폼을 통한 참여 방법과 절차를 안내하는 홍보 자료로서 포스터와 리플릿 제작이 요구되었다.

협약 대학은 강의실에 게시 가능한 실물 포스터와 인터넷 공간에 게시 가능한 이미지 파일의 리플릿을 모두 요청하였기에 각각의 홍보 자료를 개발하여 배포하였다. 실물 포스터는 강의실에 부착되어 다량의 정보를 전달하는 것보다는 시선과 흥미를 끄는 것이

가장 중요한 기능이므로 포스터의 내용 정보는 요약적으로 제시하였으며 보상으로 글쓰기 사례비를 시각적으로 강조하도록 구성되었다. 글쓰기를 수행할 뿐만 아니라 연구 참여 동의서 및 저작권 이용 허락 계약서를 작성해야만 사례비를 지급받을 수 있다는 점을 함께 표시하였다. 또한 이미 글쓰기에 대해 안내를 받은 참여자들이 부착된 포스터를 통해 쉽게 플랫폼에 접근할 수 있도록 큐알(QR) 코드를 포스터의 ‘참여 방법’ 설명 부분에 포함하였다. 1학기에 문항 번호, 글자 수, 형식 등 주어진 요건을 지키지 않은 글쓰기 자료 사례가 다수 있었으므로 이와 관련된 내용을 2학기 포스터 및 리플릿에 글 작성 시 ‘유의 사항’으로 포함하였다.



[그림 III-13] 홍보 포스터

하단의 이미지는 1학기에 배포한 양면 리플릿으로, 1면에는 사업의 취지 및 글쓰기 자료 수집의 의의와 참여 방법을, 2면에는 그림과 흐름도로 참여 방법 및 과정을 담았다. 참여를 더욱 쉽고 간편하게 하기 위해 상단에 큐알(QR) 코드를 제공하였다. 이를 통해

글쓰기 참여자들이 개인 전자 기기의 큐알 코드 스캔 기능을 통해 바로 글쓰기 자료 수집에 참여할 수 있도록 하였다.



[그림 Ⅲ-14] 1학기 홍보 리플릿

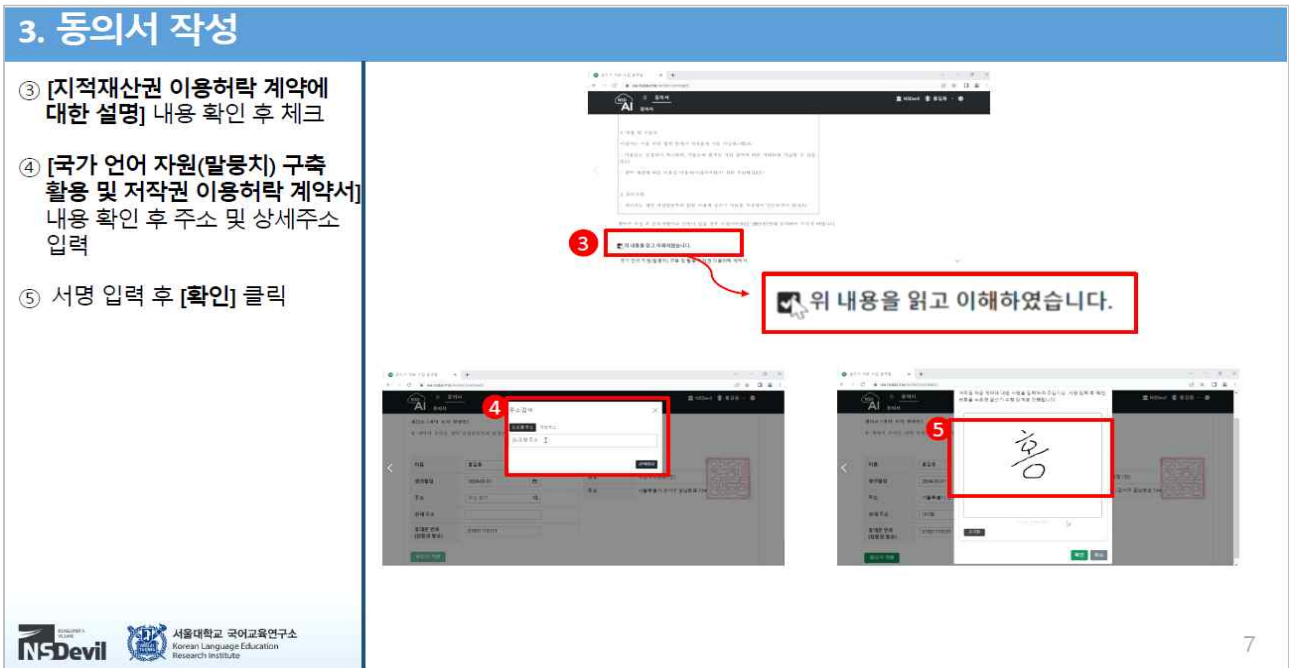
1학기의 경우 플랫폼상에서 글쓰기를 수행하고 바로 제출하는 방식과 사전에 작성한 글을 파일로 제출하는 방식을 혼용하였다. 리플릿 등의 홍보 자료 역시 이러한 두 경로를 담게 되며 한눈에 글쓰기 자료 제출의 과정이 파악되기 어렵다는 아쉬움이 있었다. 2학기의 경우 수집 경로가 단순화되기 때문에 이러한 점이 보완되었다.



[그림 III-15] 2학기 홍보 리플릿

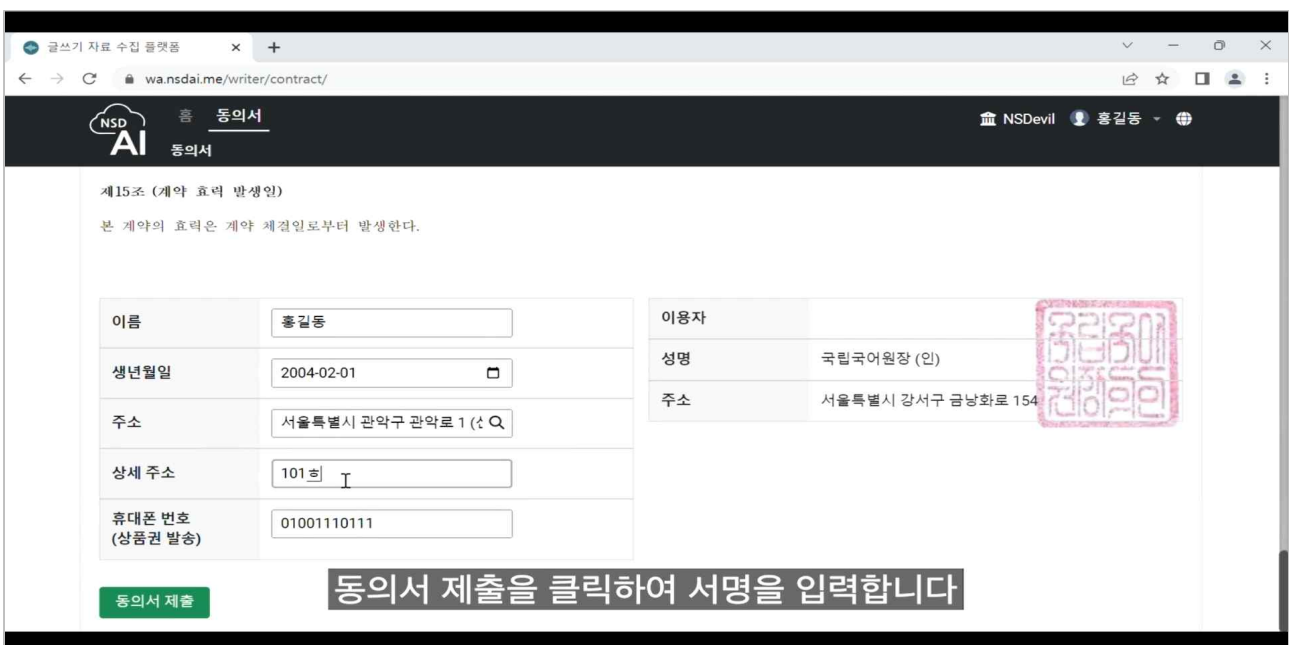
3) 플랫폼 이용 안내 동영상 및 문서 매뉴얼

플랫폼상에서 글을 직접 작성하는 경우는 물론, 사전에 작성한 파일을 업로드하는 방식에서도 연구 참여 및 저작권 이용 허락 계약서 작성을 위해 참여자가 필수적으로 플랫폼을 사용해야 하므로 플랫폼을 이용하는 방법을 숙지하여 오류를 방지하는 것은 중요하다. 글쓰기 자료 수집 플랫폼을 이용해 본 적이 없는 참여자들의 입장을 고려하면, 이용 중 문의 사항이 발생할 것을 예상할 수 있다. 이에 연구진의 이메일 주소를 통해 이용 중 궁금한 사항을 참여자가 문의할 수 있도록 하였으나 실시간 응대가 불가능하기에 상세한 이용 방법을 담은 안내 자료와 영상을 개발하여 배포하여 글쓰기 참여자들의 이해를 돕고자 하였다. 협약 대학의 교강사 역시 플랫폼을 통한 수집 관리 및 참여자의 문의 응대를 위해 동영상 및 문서 매뉴얼을 요청하였으며 동일한 자료가 참여 대학교의 교강사들에게도 공유되었다.



[그림 Ⅲ-16] 플랫폼 이용 문서 매뉴얼

영상 매뉴얼 제작을 위해 글쓰기 참여의 전 과정을 영상 녹화하였으며 각 단계마다 연구 참여자가 수행해야 하는 사항을 자막으로 달아 글쓰기 참여자들이 혼동할 수 있는 부분을 최소화하였다. 다음 그림은 영상 매뉴얼 중 동의서 작성 화면이다.



[그림 Ⅲ-17] 플랫폼 이용 영상 매뉴얼 화면

5. 글쓰기 자료 수집의 결과

1) 글쓰기 원자료 수집 현황

글쓰기 자료와 동의서를 모두 제출한 글쓰기 원자료는 1학기에 총 4,322건, 2학기에 5,117건이었다. 그중 글자 수 등 글 작성 요건을 충족하지 못한 글을 확인하여 44건을 제외한 9,395건의 유효 글을 선별하였다. 아래는 글쓰기 자료의 대학별 수집 결과이다.

<표 III-8> 대학별 글쓰기 원자료 수집 결과

	A대학	B대학	C대학	D대학	E대학	F대학	G대학	H대학	K대학	총합
1학기	965	164	92	524	766	259	508	538	487	4,303
2학기	631	772	49	925	1,027	337	233	621	497	5,092

국민의 글쓰기 능력 진단 체계를 개발하기 위한 본 연구의 목적상 유효 글쓰기 자료 중 외국인 글쓰기 참여자를 선별하여 배제하였다. 1학기에는 외국 국적을 가진 글쓰기 참여자 글쓰기 자료가 15건이었다. 2학기에도 외국인 글쓰기 참여자는 본 연구의 취지에 적합하지 않으므로 포함될 수 없음을 대학마다 사전에 안내하였음에도 외국인이 참여한 5건의 사례가 확인되었다. 1학기에는 자동 정제 후 띄어쓰기 등이 수정되면서 분량 미달 글쓰기 자료 또한 4건 발생하였다.

글쓰기 자료 요건을 모두 충족하는 글쓰기 원자료는 9,371건으로, 문항별 수집 결과는 다음과 같다.

<표 III-9> 문항별 글쓰기 원자료 수집 결과

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	총합
1학기	794	1,231	1,198	341	381	339	0	0	0	4,284
2학기	0	0	0	693	648	870	932	1,110	834	5,087
총합	794	1,231	1,198	1,034	1,029	1,209	932	1,110	834	9,371

2) 글쓰기 자료 수집 개선을 위한 제언

(1) 통제된 조건에서의 글쓰기 자료 수집

본 과업에서 수집하는 글쓰기 자료는 채점 과정을 거쳐 인공지능 학습용 데이터로 활용될 수 있는 요건을 갖추어야 한다. 글쓰기 과제 요건으로서 인터넷 검색 및 표절을 제한하는 평가 환경에서 작성하여야 하며, 글 분량, 작성 시간, 논제를 준수하고, 제목 및 목차 등 글 본문 외 형식 요소를 포함하지 않는 것을 상정하였다.

그런데 평가 환경의 측면과 관련하여, 협약 대학의 글쓰기 강의를 기반으로 자료를 수집하는 상황에서 강의별 커리큘럼에 따라 온라인 자료 활용을 통제하기 어려운 개별 과제 글쓰기의 비율은 절반 이상이 되었다. 결과적으로 인터넷상의 자료를 참고하거나 표절 수준에서 포함하는 글이 적지 않았다. 이에 채점 일관성 유지를 저해하고, 표절이나 챗지피티(ChatGPT) 등의 생성 인공지능을 활용한 글을 감별해 내야 한다는 점에서 글쓰기 자료 정제에 대한 부담이 높았다.

향후 대학과의 협약을 통해 글쓰기 자료를 수집하는 안정성과 효율성을 담보하되, 평가 환경 통제에 대해 사전에 충분히 협의할 필요가 있다. 협조가 가능한 강의를 중심으로 자료 수집을 한정하였을 때 수량 확보가 어렵다면 협약 대학을 점차 확대하는 방안도 적극 검토해야 할 것이다. 또 글쓰기 자료 요건 미충족으로 최종 데이터에서 제외되는 수량을 줄이기 위해서도, 미충족의 구체적인 사례를 포함한 매뉴얼을 제작하고 홍보하는 등 대학과의 협력을 보다 강화할 필요가 있다.

(2) 문항 풀림 현상의 해소

본 연구에서는 인공지능 학습용 데이터로 활용하기 위해서 문항별로 글쓰기 자료를 고르게 수집하는 것을 목표로 하였다. 그런데 문항별로 주제에 대한 선호도가 달라 문항 선택의 자율성이 높았던 1학기에는 문항별 글쓰기 자료 수집에서 풀림 현상이 나타났다. 이를 보완하기 위해 2학기의 글쓰기 자료 수집은 보다 체계적으로 기획되었다. 사전에 각 대학별 수집 역량을 확인하고, 문항을 배정한 후 확인 및 조정을 거쳐 최종 수집 목표량을 확정하였다. 가령 아래 표에서 G대학은 8, 9번 문항의 글쓰기 자료를 각각 150건

씩 총 300건 수집하는 것이 수집 목표에 해당한다. 수집 현황을 살펴보면 각 문항마다 수집 목표량과 실제 수집량이 차이가 있음을 확인할 수 있다. 8번 문항의 경우 총 1,020건을 수집하는 것이 수집 목표였으나 총 1,110건을 수집하여 목표량을 초과하였다. 9번 문항의 경우 총 1,020건을 수집하는 것이 수집 목표였으나 총 834건을 수집하여 목표량에 미달하였다.

<표 III-10> 2학기 글쓰기 자료 문항별 수집 현황

구분		A대학	B대학	C대학	D대학	E대학	F대학	G대학	H대학	K대학	총합
문항 번호	수집 목표량	600	800	-	940	610	600	300	650	500	5,000
Q4	660	14	219	9	252	181	3	7	3	5	693
Q5	620	0	236	13	162	227	4	0	6	0	648
Q6	660	1	304	2	203	354	3	2	0	1	870
Q7	1,020	143	3	9	263	105	76	0	156	177	932
Q8	1,020	301	4	9	16	102	135	113	223	207	1,110
Q9	1,020	172	6	7	29	57	116	111	232	104	834
총합	5,000	631	772	49	925	1,026	337	233	620	494	5,087

수집 기한을 연장하고 추가 참여자를 모집하여 전체적으로 수집 목표량에 근사한 수치의 글쓰기 자료를 수집하였으나 문항별 목표에 균형 있게 도달하는 것에는 큰 어려움이 있었다. 글쓰기 참여가 원칙적으로 자율이라는 점에서 학습자들의 참여도를 예측하는 것이 어려웠으며, 대학별 수집 문항이 달랐기 때문에 대학별 글쓰기 참여자의 참여율 편차가 문항별 수집 불균형으로 이어지기도 하였다.

또한 글쓰기 참여자들의 문항 선호도가 수집 결과에 영향을 미치기도 하였다. 상대적으로 어렵거나 낮은 주제의 문항을 배정받은 글쓰기 참여자들의 참여율이 떨어지는 것을 관찰할 수가 있었다. 대학 실무자들이 문항의 난이도에 따라 특정 문항의 수집에서 어려움이 있었음을 호소한 사실이 이를 방증한다.

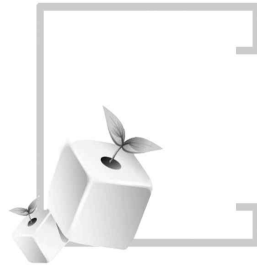
특정 문항의 초과분이 발생하는 경우 플랫폼 시스템상 해당 문항을 숨김 처리하여 해당 문항의 참여를 중단하는 것이 가능하다. 그러나 수집 종료 후 특정 문항의 미달분이 발생하는 경우 해결 방안이 없다. 2학기에는 특정 문항의 미달로 인한 문제를 해결하기

위해 협약 대학과의 소통이 활발하게 이루어졌다. 따라서 연구진 및 실무자가 긴밀히 협력하여 글쓰기 자료 수집 상황을 모니터링하고, 추가 모집 등을 선제적으로 유연하게 사용할 필요가 있다. 또한 협약 대학의 수집 기한을 앞당긴 후, 부족분만 다양한 통로를 통해 추가 모집하는 등 수집 단계의 이원화 역시 고려해 볼 수 있다.

(3) 수집 수량의 예측 가능성 담보

2023년 글쓰기 자료 수집의 최종 목표 수량인 9,000건을 달성하기 위해서 연구진은 지속적인 모니터링 및 수집 계획을 수립하였다. 협약 대학 학생들의 글쓰기 참여는 자율이기 때문에 교강사의 계획 조사 결과와 실제 참여율의 차이가 존재하였다. 따라서 이를 고려하여 아이디어를 넉넉히 발급하되 지속적인 모니터링이 필요하다. 1학기의 사례를 바탕으로 2학기에는 수집 계획 조사에서 수집 수량과 보다 정확한 수집 시기를 조사하였다.

1학기의 경우 학기가 끝날 때까지 수집이 가능했으나 2학기에는 예산 집행 기한 등의 한계로 인해 학기 중 수집을 종료해야 하는 어려움이 있었다. 협약 대학의 교강사 중 다수가 중간고사가 끝난 이후 글쓰기 자료 수집을 계획했기 때문에 2학기 수집 글쓰기 자료의 대부분은 수집 종료 직전에 수집되었다. 수집이 지연되면서, 수집 종료 시기에 가까워질수록 추가 홍보, 추가 아이디어 발급, 수집 기한 연장 등을 통해 목표 수집 총량을 확보하고자 하였다. 그러나 수집 기한 마감 2-3일 전 300명 이상의 참여자들이 집중적으로 글쓰기 자료를 제출하면서 결국은 목표 수량을 다소 상회하는 글쓰기 자료를 최종적으로 수집하게 되었다.



제 IV 장

글쓰기 채점 전문 인력 양성



본 과업에서는 ‘프로그램 개발과 실행의 형성적 순환에 입각한 글쓰기 채점 전문 인력의 양성’을 기본 방향으로 삼아 글쓰기 채점 전문 인력 양성을 추진한다. 다음은 본 과업에서 달성하고자 하는 하위 목표이다.

첫째, 실제 채점 데이터를 활용한 채점 워크숍 기반의 실무 중심 교육 프로그램을 개발한다. 2022년 기초 연구에서 수집한 채점 데이터를 활용하여 채점 기준의 이해 및 적용, 다양한 답안의 해석과 처리 등, 채점 워크숍 기반의 실무 중심 채점자 교육 프로그램을 개발한다. 또한 채점 실습의 과정과 결과를 모니터링하고 그 내용을 바탕으로 채점 교육 참여자에게 맞춤형 피드백을 제공함으로써 채점 문식성에 대한 이해도를 제고한다.

둘째, 실무 중심 교육 프로그램을 통해 전문적 채점 훈련을 시행하고 채점 전문성 신장을 도모한다. 실제 채점 데이터를 활용한 채점 실습과 채점 과정에 대한 맞춤형 피드백을 통해 채점 교육 참여자는 다음과 같은 점에서 채점 문식성을 신장시킬 수 있다.

- 문항 및 진단 지표에 대한 이론적 이해도 제고
- 진단 지표의 해석 및 적용 과정의 공유와 점검
- 채점자 간 협의를 통한 채점 결과의 조정 체험
- 채점 문식성의 신장 및 채점자 인식 변화

셋째, 채점자 교육 프로그램의 실행 과정에 대한 모니터링 및 결과 분석에 기반하여 프로그램 개선안을 도출한다. 채점 실습의 과정과 결과에서 수집된 자료, 채점자 교육용 사전·사후 설문 조사 결과에 대한 양적·질적 분석을 통해 프로그램의 개선 방향을 확인하고 이를 종합하여 채점자 교육 프로그램에 대한 개선안을 도출한다.

본 과업의 세부 추진 과제는 채점자 교육 프로그램의 설계, 개발, 운영, 평가의 네 단계로 구분된다. 프로그램의 설계 단계에서는 교육과정의 개요와 강의안을 구성하고, 개발 단계에서는 교육에 필요한 자료를 개발 및 제작한다. 운영 단계에서는 프로그램 참여자를 모집하고 교육을 실시한다. 마지막 평가 단계에서는 채점자 교육 실행의 결과를 분석하고 개선안을 도출한다.

1. 채점자 교육 프로그램 설계

1) 문헌 검토

채점자 훈련의 방식 및 절차에 대해서는 다양한 논의가 있어 왔고(Herman et al., 1992; Bachman & Palmer, 1996; Carr, 2011), 국내에서도 채점자 훈련 프로그램이나 모형의 개발에 관한 연구가 이루어졌다(이영식, 2000; 박종임, 2013; 최윤곤·허유리, 2013). 따라서 이와 관련된 문헌 연구를 토대로 하여 효과적인 채점자 교육 워크숍 프로그램 개발이 이루어져야 한다. 이에 채점자 교육 프로그램과 관련된 국내외 문헌 및 사례를 검토하여 채점자 교육을 위한 교육과정의 개요를 설계하고자 하였다.

먼저 채점자 훈련의 방식 및 절차와 관련된 연구로는 헤르만 외(Herman et al., 1992), 바흐만과 파머(Bachman & Palmer, 1996), 카(Carr, 2011) 등이 있다. 바흐만과 파머(1996)에서는 채점의 비일관성을 해결하는 효과적인 방법으로 채점자의 적절한 선발과 훈련을 언급하였다. 선발된 채점자들이 서로 다른 배경이 가지고 있더라도, 채점자들에게 적절한 훈련을 제공하는 일이 중요한데, 그것은 (1) 채점 척도를 읽고 함께 토의하기, (2) 전문 채점자들에 의해 사전 채점된 샘플을 검토하고 척도에 대해 토의하기, (3) 연습 채점 실시하기, (4) 추가적인 표본을 채점하고 토의하기, (5) 각각의 예비 채점자들이 동일한 표본을 채점하고 채점 소요 시간 및 일관성을 확인하기, (6) 신뢰할 만하고 효과적인 채점을 수행할 수 있는 채점자를 선별하기 등 여섯 단계 절차로 이루어진다(Bachman & Palmer, 1996: 222).

헤르만 외(1992)에서도 채점자 훈련은 일관된 채점을 위한 선결 요건이라고 보았다. 그 훈련의 매뉴얼은 (1) 충분히 설명된 채점 준거, (2) 각 채점 점수를 대표하는 예시, (3) 실제 채점 중 참고 가능한 축약된 1페이지의 준거, (4) 점수 기록의 표본 형식 등을 포함할 것을 제안하였다(Herman et al., 1992: 83). 아울러, 채점자 훈련 시 수행되어야 할 내용으로서 평가 과제에 대한 오리엔테이션, 채점 기준의 명료화, 실제적 채점 실습, 채점 기준의 수정 가능성, 채점 결과의 기록, 채점자 간 신뢰도 확보, 그리고 채점 시간의 고려 등 7가지가 언급된다(Herman et al., 1992: 83-4).

한편 카(2011)에서는 채점자의 훈련(training)과 표준화(norming)의 필요성을 언급하였다. 채점자 훈련의 목적은 서로 다른 채점자들이 채점 척도를 동일한 방식으로 이해

및 적용하고, 그 기술어(descriptor)를 동일하게 해석할 수 있도록 돕는 것이다(Carr, 2011: 145). 아울러, 숙련된 채점자라고 할지라도 특정 채점 척도를 사용해 본 지 오래 되었을 경우에는 특히 그 척도를 다시 복습하고 재사용하는 절차가 필요한데, 이는 '채점자 규준화(rater norming)'라고 명명된다. 채점자 훈련을 위해 채점자들이 대면하는 것이 때로는 가능하지 않을 수 있는데, 적절한 자기 훈련 자료를 제공함으로써 높은 일관성을 성취하는 것 또한 가능하다. 한편, 웹 기반의 채점자 훈련 역시 보편화되고 있으며, 역시 효과적일 수 있음이 보고되었다(Carr, 2011: 146).

국내 채점자 훈련 프로그램 및 모형 개발 연구에서도 쓰기 평가에서 채점자 훈련의 필요성을 인식하고 그 실제적 방안을 모색한 연구가 이루어졌다(이영식, 2000; 박종임, 2013; 최윤곤·허유리, 2013). 박종임(2013: 420)에서는 쓰기 평가 채점자 훈련 프로그램의 구성 내용으로 '전문 채점자와의 일치도 정보', '채점자 엄격성, 일관성, 편향성 정보', '채점 척도 활용 정보', '채점 분량에 따른 일관성 변화 정보'를 제시하였다. 특히 개별 채점자의 엄격성에 대한 정보 제공을 위해서는 채점자 훈련 프로그램 참여자의 채점 결과 데이터와 전문 채점자의 데이터를 통합적으로 분석할 필요가 있음을 주장하였다. 이 연구는 쓰기 채점자 훈련을 전국의 국어 교사를 대상으로 대규모로, 그리고 지속적으로 시행되어야 한다고 주장하면서 면대면 훈련의 한계점(공간, 시간, 비용 등)을 극복하기 위해 온라인 쓰기 채점자 훈련 프로그램(OWRT, Online Writing Rater Training) 모형을 제시하였다. 이 모형에서는 전문 채점자의 채점 결과와의 비교 및 다국면 라쉬 분석을 통해 채점자 정보를 제공하여 훈련을 하도록 하고, 채점자 간 일치도 향상과 채점자 내 일관성 향상을 바탕으로 평가 전문성 신장을 도모하고자 하였다.

최윤곤·허유리(2013)에서는 한국어 쓰기 평가의 채점 신뢰도 신장을 목표로 하여 채점자 훈련 모형을 논의하였다. 대학 기관에서 한국어를 교육하는 15인의 채점자를 대상으로 채점 수행을 의뢰하였으며, 채점 데이터를 수집함과 동시에 채점자 신뢰도 및 채점자 훈련에 관한 설문 조사를 실시하였다. 이 연구에서는 5단계의 채점자 교육 절차를 제시하였는데, <채점 방법 토론 → 채점 기준 토론 → 채점자 내 신뢰도를 위한 채점 연습 → 채점자 간 신뢰도를 위한 채점 연습 → 채점과 관련된 모든 요소에 대한 정리>로 해당 모형을 제시하였으며, 이러한 모형을 통해 궁극적으로 채점자 간 일치도 향상 및 채점자 내 일관성 향상을 목표로 하였다.

이상의 문헌 검토를 통해 채점자 훈련 및 교육의 필요성과 채점자 간 일치도와 채점자 내 일관성을 확보하기 위한 공통된 절차(평가 도구에 대한 이해, 채점 연습, 채점자 간

토의, 신뢰도 확인 등)를 확인할 수 있었다. 그러나 이는 본격적인 채점을 수행하기에 앞서 실제 채점에 투입되는 채점자들을 대상으로 실시하는 채점자 훈련의 절차이다. 본 과업에서는 본 채점에 투입될 수 있는 전문 채점 인력을 양성하는 것을 목적으로 하므로 채점자 훈련의 절차들이 효과적으로 제 기능을 발휘하기 위해서는 각 절차들이 포함하고 있는 요소에 관한 지식과 원리를 간과할 수 없다. 가령 ‘평가 도구에 대한 이해’라는 채점자 훈련의 첫 번째 절차에서 쓰기 평가에 관한 일반적 지식(쓰기 평가의 목적, 구인, 평가 지표 등)이 선행되어야 해당 절차가 채점자의 이해도를 제고하는 역할을 제대로 수행할 수 있다. 채점 연습과 신뢰도 확인의 절차도 마찬가지이다. 채점 과정에서 영향을 미치는 요인들이 무수히 많으며, 그에 따라 채점의 신뢰도와 적합도가 달라진다는 일종의 채점 이론을 갖추고 있을 때 채점자가 채점 준거 판정에서의 자기 점검과 조절을 더 효율적으로 할 수 있다. 이에 본 교육과정에서는 쓰기 채점을 수행하는 데 필요한 기초적인 지식을 바탕으로 하여 그에 관한 원리를 체득하고, 채점 과정에서 직면할 수 있는 여러 문제 상황들을 점검하고 동료 채점자들과 협의할 수 있도록 하는 채점자 교육 프로그램을 구안하고자 하였다.

2) 교육과정 설계

어떤 채점자가, 어떤 글을 채점하게 되는지가 본 교육과정의 대상과 교육 내용이 된다. 본 교육과정은 앞서 명시한 바와 같이 글쓰기 채점 전문 인력 양성이라는 독자적 목표를 지니는 동시에, 본 사업에서 실시하는 글쓰기 자료 채점에 참여할 인력을 선발하고 교육하는 목적으로도 활용된다. 본 사업의 자료 수집과 채점에 참여하게 되는 인력은 9개 협약 대학의 글쓰기 교강사이며, 자료는 논증적 글쓰기 문항을 통해 수집된 대학생의 글쓰기 자료이다. 이를 반영하여 본 교육과정은 논증적 글쓰기를 중심으로 하되, 대학의 교강사를 교육 대상으로 하여 개발 및 설계하였다. 또한 교육 대상이 전국의 대학 교강사라는 점을 고려하여 시·공간적 제약을 극복하고 교육의 효율성을 높이기 위해 온-오프라인 교육을 병행하는 블렌디드(blended) 교육과정을 채택하였다. 자기 주도적 학습이 필요한 이론 중심의 강의는 온라인 교육과정으로, 채점 실습과 협의가 중심이 되는 강의는 오프라인 교육과정으로 설계하였다. 그리고 교육 참여자의 채점 경험과 자기 인식을 점검하고 교육을 통한 인식 변화와 프로그램의 개선점 등을 확인하기 위한 목적으로 채점자 교

육용 사전·사후 설문 도구도 함께 개발하였다.

(1) 온-오프라인 교육과정

교육 프로그램 전반에서 글쓰기 평가 및 채점에 관한 ①이론 강의와 ②채점 적용, ③채점 실습의 세 차원을 모두 다루는 교육 내용을 포함하도록 하여 쓰기 채점에 필요한 지식과 원리, 적용을 단계별로 경험할 수 있도록 하였다. ①이론 강의와 ②채점 적용은 온라인 교육과정으로 구성하여, <쓰기 평가 이론>과 <쓰기 채점 이론>을 통해 쓰기 채점에 필요한 지식과 원리를 체득하고 이를 <쓰기 평가 준거별 채점 방법>에서 실제로 적용하도록 하였다. 채점자 훈련에 해당하는 ③채점 실습은 오프라인 교육과정으로 구성하여 채점 실습 과정을 교육 참여자가 함께 경험하고 공유할 수 있도록 하였다. 또한 온라인 교육과정과 오프라인 교육과정을 유기적으로 연계하기 위해 ②채점 적용 단계에서 채점 과제를 부여하여 채점의 과정과 결과에 관하여 오프라인 교육과정의 <채점 신뢰도 이해 실습>에서 확인 및 논의할 수 있도록 설계하였다. 온라인 교육 12차시와 오프라인 교육 8차시 총 20차시로 구성된 전체 개요는 <표 IV-1>과 같다.

<표 IV-1> 글쓰기 채점자 교육과정 개요

교육 방식	교육 단계	교육 내용	세부 교육 내용	차시
온라인	채점 이해	쓰기 평가 이론	쓰기 능력과 쓰기 평가의 이해 논증적 글쓰기의 평가 지표	4
		쓰기 채점 이론	쓰기 평가를 위한 채점 절차 채점 신뢰도 및 영향 요인	4
	채점 적용	쓰기 평가 준거별 채점 방법	내용, 조직, 표현 범주 채점 방법 채점 도구의 활용 방법	4
오프라인	채점 실습	채점 도구 이해 실습	쓰기 평가 준거의 일반적 이해 채점 도구의 실제 적용	2
		채점 신뢰도 이해 실습	예시 답안 채점 채점 신뢰도 분석 및 점검	5
		채점 실습 교육의 환류	채점 실습에 관한 종합 토론 채점자 자기 평가 및 성찰	1

(2) 채점자 교육용 설문 조사 도구

채점자 교육과정에는 온-오프라인 교육과정뿐만 아니라 교육 참여자들의 채점 인식 점검 및 평가를 위한 사전·사후 설문 도구가 포함된다. 설문 도구는 다음의 목적을 지닌다: ①채점 및 채점자 훈련 경험 확인, ②교육 참여자의 인식 변화 확인, ③프로그램 효과 검증, ④프로그램 개선 방안 모색. 이때 사전 조사 도구는 채점자의 채점 전문성에 대한 인식 조사를 주된 목적으로 하고, 사후 조사 도구는 채점자 교육 프로그램에 대한 만족도 조사를 주된 목적으로 한다. 조사 대상은 사전 설문의 경우 채점 전문 인력 양성 프로그램의 이수 대상자이며, 사후 설문의 경우 본 프로그램의 이수자이다. 조사 시기는 사전 설문은 온라인 교육 시행 전, 사후 설문은 오프라인 교육 종료 후 실시하며, 조사 방식은 구글 설문을 통한 온라인 방식을 택하였다. 이를 위해 교육 참여자의 채점에 관한 경험과 인식, 프로그램 전반에 관한 평가 등을 확인할 수 있는 문항의 초안을 개발하였으며, 연구진 내부 검토와 관련 전문가 3인의 자문을 구해 최종 설문 도구를 <표 IV-2>와 같이 확정하였다.

<표 IV-2> 채점자 교육용 사전·사후 설문 도구

사전 설문 도구

I. 쓰기 채점 및 채점자 훈련(교육) 경험

1. 쓰기 채점 경험(과제에 대한 복수의 일정한 답안을 채점한 경험)

- 있다
 없다

(있다면 응답한 경우)

1-1. 채점을 경험하신 기관 또는 사업을 선택하여 주십시오(복수 응답 가능).

- 소속 기관 수업 및 강의
 국내 대규모 (한)국어 평가
 연구 사업
 기타()

1-2. 위와 같은 채점 경험의 빈도를 선택하여 주십시오.

- 1-2회 3-5회 6-8회 8-10회 10회 이상

2. 쓰기 채점자 훈련(교육) 경험

- 있다
 없다

사전 설문 도구

(‘있다’에 응답한 경우)

2-1. 쓰기 채점자 훈련(교육)을 경험하신 기관 또는 사업을 선택하여 주십시오(복수 응답 가능).

- 소속 기관
 국내 대규모 (한)국어 평가
 연구 사업
 기타()

2-2. 쓰기 채점자 훈련(교육)에 참여하신 빈도를 선택하여 주십시오.

- 1회 2회 3회 4회 5회 이상

II. 쓰기 평가 전문성 및 채점자 교육 필요성에 대한 인식

1. 선생님의 쓰기 평가 전문성의 현재 수준은 어느 정도입니까?

(1점)매우 낮다 - 낮다 - 보통이다 - 높다 - 매우 높다(5점)

쓰기 평가 전문성 하위 항목	1	2	3	4	5
쓰기 채점에 대한 지식 차원					
쓰기 채점의 실제 수행 차원					
쓰기 채점 결과의 활용 차원					

2. 쓰기 평가 전문성 신장을 위한 채점자 교육이 얼마나 필요하다고 생각하십니까?

(1점)전혀 필요하지 않다 - 그렇지 않다 - 보통이다 - 그렇다 - 매우 필요하다(5점)

본 설문의 결과는 채점 전문 인력 양성 프로그램의 평가 및 개선, 채점자의 인식 변화 및 성찰을 확인하기 위한 목적으로만 활용됩니다. 선생님께서 작성하신 설문 결과의 활용에 동의하십니까?

- 동의함 동의하지 않음

사후 설문 도구

I. 채점 전문 인력 양성 프로그램 평가

(1점)전혀 그렇지 않다 - 그렇지 않다 - 보통이다 - 그렇다 - 매우 그렇다(5점)

1. 온라인 교육

1-1. 플랫폼을 통한 강의 수강이 편리했다.

1-2. 차시별 강의 시간이 적절하였다.

(‘전혀 그렇지 않다, 그렇지 않다’에 응답한 경우)

1-2-1. 강의 시간이 적절하지 않다고 생각하신 이유를 간단히 적어 주십시오.

()

사전 설문 도구

1-3. 다음의 온라인 교육 내용이 쓰기 채점 전문성 향상에 도움이 되었다.

교육 내용	세부 내용	1	2	3	4	5
쓰기 평가 이론	쓰기 능력과 쓰기 평가의 이해/ 논증적 글쓰기의 평가 지표					
쓰기 채점 이론	쓰기 평가를 위한 채점 절차/ 채점 신뢰도 및 영향 요인					
쓰기 평가 준거별 채점 방법	내용, 조직, 표현 범주 채점 방법/ 채점 도구의 활용 방법					

1-4. 온라인 교육 내용의 수준이 적절하였다.

(‘전혀 그렇지 않다, 그렇지 않다’에 응답한 경우)

1-4-1. 온라인 교육 내용 수준이 적절하지 않다고 생각하신 이유를 간단히 적어 주십시오.

()

1-5. 교수진의 설명이 교육 내용을 이해하기 쉽게 이루어졌다.

1-6. 강의 자료가 교육 내용을 이해하는 데 도움이 되었다.

2. 오프라인 교육

2-1. 워크숍 형태의 교육 방식이 적절하였다.

(‘전혀 그렇지 않다, 그렇지 않다’에 응답한 경우)

2-1-1. 교육 방식이 적절하지 않다고 생각하신 이유를 간단히 적어 주십시오.

()

2-2. (오프라인 교육) 교육 시간(8차시)이 적절하였다.

(‘전혀 그렇지 않다, 그렇지 않다’에 응답한 경우)

2-2-1. 교육 시간이 적절하지 않다고 생각하신 이유를 간단히 적어 주십시오.

()

2-3. 다음의 오프라인 교육 내용이 쓰기 채점 전문성 향상에 도움이 되었다.

교육 내용	세부 내용	1	2	3	4	5
채점 도구 이해 실습	쓰기 평가 준거의 일반적 이해 / 채점 도구의 실제 적용					
채점 신뢰도 이해 실습	예시 답안 채점 / 채점 신뢰도 분석 및 점검					
채점 실습 교육의 환류	채점 실습에 관한 종합 토의 / 채점자 자기 평가 및 성찰					

사전 설문 도구

2-4. 오프라인 교육 내용의 수준이 적절하였다.

(‘전혀 그렇지 않다, 그렇지 않다’에 응답한 경우)

2-4-1. 오프라인 교육 내용의 수준이 적절하지 않다고 생각하신 이유를 간단히 적어 주십시오.

()

2-5. 교수진의 설명이 교육 내용을 이해하기 쉽게 이루어졌다.

2-6. 강의 자료가 교육 내용을 이해하는 데 도움이 되었다.

3. 프로그램 전반

3-1. 본 프로그램에 대해 전반적으로 만족한다.

3-2. 본 프로그램은 쓰기 채점 전문성 향상에 도움이 된다.

4. (자기 평가) 나는 본 프로그램에 적극적으로 참여하였다.

5. 본 프로그램에서 향후 다루어졌으면 하는 사항이나 개선점이 있다면 기술하여 주십시오.

II. 채점자 교육 필요성에 대한 인식

1. 쓰기 평가 전문성 신장을 위한 **채점자 교육**이 얼마나 필요하다고 생각하십니까?
(1점)전혀 그렇지(필요하지) 않다 - 그렇지 않다 - 보통이다 - 그렇다 - 매우 그렇다(필요하다)(5점)

본 설문문의 결과는 채점 전문 인력 양성 프로그램의 평가 및 개선, 채점자의 인식 변화 및 성찰을 확인하기 위한 목적으로만 활용됩니다. 선생님께서 작성하신 설문 결과의 활용에 동의하십니까?

동의함

동의하지 않음

(3) 프로그램 운영 개요서(별책 부록)

본 과업에서 개발한 글쓰기 채점자 교육 프로그램을 운영하는 데 필요한 내용을 담은

운영 개요서를 작성하여 운영 주체가 바뀌더라도 안정적으로 프로그램을 운영할 수 있도록 한다. 운영 개요서의 목차는 다음과 같다: 1) 프로그램 운영 절차 및 전략, 2) 프로그램 개요, 3) 참여자 모집 문건, 4) 온라인 교육용 자료집, 5) 오프라인 교육용 자료집, 6) 채점자 교육용 사전·사후 설문 도구. 운영 개요서를 통해 채점자 교육 프로그램 전반의 운영 절차, 각 단계별 핵심 과업, 프로그램 관리 및 운영 전략 등을 공유한다.

3) 온라인 교육 내용 설계

온라인 교육을 위한 강의 교수진으로서 채점 워크숍을 운영해 본 경험이 많고 경력이 높은 쓰기 교육 전문가 3인을 선정하였다. 교육 내용 설계는 문헌 연구를 바탕으로 연구진이 개요 초안을 개발하였으며, 이후 교수진과 연구진이 협력하여 최종 강의안을 완성하였다.

<표 IV-3> 온라인 채점자 교육과정 개요

교육 방식	교육 단계	교육 내용	세부 교육 내용	차시	강의 목차	교수자
온라인	채점 이해	쓰기 평가 이론	쓰기 능력과 쓰기 평가의 이해 / 논증적 글쓰기의 평가 지표	4	쓰기 능력과 구인	장지혜 (춘천교대)
					쓰기 평가의 이해	
					채점 기준표의 기본 구조 논증적 글쓰기의 평가 지표	
	채점 이론	쓰기 평가를 위한 채점 절차 / 채점 신뢰도 및 영향 요인	4	채점 오차와 채점 절차	박성석 (춘천교대)	
				채점 신뢰도 및 영향 요인		
				채점 시 유의 사항과 채점자 훈련 채점 경향과 채점 피드백		
채점 적용	쓰기 평가 준거별 채점 방법	내용, 조직, 표현 범주 채점 방법 / 채점 도구의 활용 방법	4	내용 범주 채점 방법1	최영인 (서울교대)	
				내용 범주 채점 방법2		
				조직 범주 채점 방법		
				표현 범주 채점 방법		

온라인 교육과정은 채점에 대한 이해와 적용의 단계로 구성되며, 채점 이해 단계는 쓰기 평가 이론과 쓰기 채점 이론이 각각 4차시, 채점 적용 단계는 준거별 채점 방법 4차시가 교육 내용으로 다뤄진다. 이때 국립국어원·서울대학교 국어교육연구소(2022)의 기초 연구에서 수집된 글쓰기 자료와 채점 결과를 일부 활용하여 실제성 있는 교육 내용을 설계하고자 하였다. 교육 내용별 세부 강의 요목은 다음과 같다.

(1) 쓰기 평가 이론

쓰기 평가 이론 강의에서는 평가 상황에 존재하는 다양한 요소에 대해 이해하고 채점자의 지식이 쓰기 평가 상황에서 어떻게 활용될 수 있는지에 대한 방향을 제시하는 것을 강의 목표로 삼는다. 강의 내용은 쓰기 평가에 대한 학술적 논의(Purves, 1984; Weigle, 2001; Bachman & Palmer, 1996)를 참고하여 다음 항목들이 포함되도록 강의안을 구성하였다.

<표 IV-4> 쓰기 평가 이론 강의 요목

강의 목차	세부 항목
쓰기 능력과 구인	<ul style="list-style-type: none"> □ 평가 구인(construct)의 개념 및 정의 □ 쓰기 능력의 구인: 지식/수행/태도 차원 □ 쓰기 능력 진단의 일반적 지표: 내용/조직/표현 차원 □ 글의 장르에 따른 쓰기 능력 진단 지표의 차이: 논증적/설명적/친교적 글쓰기 구인
쓰기 평가의 이해	<ul style="list-style-type: none"> □ 쓰기 평가의 과정: 설계, 평가 도구 제작, 실행 □ 신뢰도(채점 신뢰도)와 타당도(구인 타당도, 채점 타당도) □ 쓰기 평가의 유형: 직접식/간접식/대안적 평가(포트폴리오, 자기 평가)
채점 기준표의 기본 구조	<ul style="list-style-type: none"> □ 채점 준거(criteria), 기준(standard), 척도(scale)의 이해 □ 채점 기준표의 구성 요소: 구인(대준거), 하위 준거, 척도(점수), 기준 □ 채점 기준표의 개발 과정
논증적 글쓰기의 평가 지표	<ul style="list-style-type: none"> □ 논증적 글쓰기의 장르적 특징: 논증의 개념, 설득 목적의 글쓰기, 반론 고려의 내용 및 방법 □ 논증적 글쓰기의 주요 평가 지표: 내용/조직/표현 차원

(2) 쓰기 채점 이론

쓰기 채점 이론에서는 쓰기 평가 전체의 맥락에서 채점의 위치와 역할에 대해 이해하되, 이론에 대한 세부적인 지식을 강화한다기보다 신뢰할 만한 채점을 진행하기 위해 알아야 할 핵심 개념을 다룸으로써 채점 과정 전반에 대한 이해도를 높이는 것을 목표로 삼는다. 강의 내용은 채점과 관련된 학술적 논의(Eckes, 2008; 2012; Myford & Wolfe, 2004)를 참고하여 다음의 항목들이 포함되도록 강의안을 구성하였다.

<표 IV-5> 쓰기 채점 이론 강의 요목

강의 목차	세부 항목
채점 오차와 채점 절차	<ul style="list-style-type: none"> □ 쓰기 평가에 개입하는 오차: 발현 오차, 채점 오차, 해석 오차 □ 쓰기 평가의 일반 절차
채점 신뢰도 및 영향 요인	<ul style="list-style-type: none"> □ 채점 신뢰도(객관도): 채점자 내 신뢰도, 채점자 간 신뢰도 □ 채점에의 영향 요인: 채점자 내적 요인, 과제 환경 요인
채점 시 유의 사항과 채점자 훈련	<ul style="list-style-type: none"> □ 채점 시 유의해야 할 사항: 관대성-엄격성, 후광 효과, 중앙 경향성, 무작위 효과, 편향 효과 □ 채점자 훈련: 필요성, 유형, 방법, 절차
채점 경향과 채점 피드백	<ul style="list-style-type: none"> □ 채점 경향: 엄격성, 일관성, 편향성 □ 채점 피드백의 내용: 전문 채점자와의 일치도, 채점 경향, 채점 척도 활용, 채점 분량에 따른 일관성 변화

(3) 쓰기 평가 준거별 채점 방법

쓰기 평가 준거별 채점 방법에서는 채점 준거의 실제적 적용과 준거 적용 시의 쟁점 및 판단에 대한 이해도 제고를 목표로 삼는다. 강의 내용은 2022년 기초 연구에서 기개발한 문항 및 채점 기준표와 척도별 예시 답안들을 활용하여 다음의 항목들이 포함되도록 강의안을 구성하였다.

<표 IV-6> 쓰기 평가 준거별 채점 방법 강의 요목

강의 목차	세부 항목
내용 범주 채점 방법1	<ul style="list-style-type: none"> □ 작문 과제 분석: 작문 과제 분석의 필요성과 과제 분석의 실제 □ 평가 준거 전반 개관: 작문 평가 준거의 구성 □ 내용 범주 평가 준거①: 문제 상황 이해[상황성]의 채점 준거 이해 및 적용
내용 범주 채점 방법2	<ul style="list-style-type: none"> □ 논증 구조에 대한 이해: 논증의 개념과 논증 구조 □ 내용 범주 평가 준거②: 주장[명료성] 채점 준거 이해 및 적용 □ 내용 범주 평가 준거③: 이유/근거의 적절성[적절성] 채점 준거 이해 및 적용 □ 내용 범주 평가 준거④: 이유/근거의 충분성[충분성] 채점 준거 이해 및 적용
조직 범주 채점 방법	<ul style="list-style-type: none"> □ 내용 범주 평가 준거⑤: 다른 입장에 대한 고려[다양성] 채점 준거의 이해 및 적용 □ 조직 평가 준거①: 글 수준 채점 준거 이해 및 적용 □ 조직 평가 준거②: 문단 수준 채점 준거 이해 및 적용
표현 범주 채점 방법	<ul style="list-style-type: none"> □ 표현 평가 준거①: 문장과 어휘[적절성] 채점 준거 이해 및 적용 □ 표현 평가 준거②: 어문 규범과 관습[규범성] 채점 준거 이해 및 적용 □ 답안의 채점 연습: 채점 시 유의 사항

4) 오프라인 교육 내용 설계

오프라인 교육은 채점자 훈련 프로그램이나 모형에 관한 연구(이영식, 2000; 박종임, 2013; 최윤곤·허유리, 2013; 홍은실 외, 2020) 등을 참고하여 채점자 워크숍 형태로 구성하였다. 이영식(2000)은 협의에 기반을 둔 워크숍을 실시하는 것은 채점자들을 커뮤니티로 조직하여 채점자들이 가지고 있는 이질적인 배경과 개념을 중립화함으로써 ‘채점자의 사회화’에 기여한다고 하였다. 이러한 채점자 협의에 대한 강조는 다른 연구에서도 확인된다. 한국어 쓰기 평가의 채점자 훈련 프로그램을 다룬 최윤곤·허유리(2013: 311)에서 제안한 채점자 훈련 모형에서는 채점 방법 및 채점 기준에 대한 토론이 포함되어 있으며, 채점자 내 신뢰도를 위한 채점 연습, 채점자 간 신뢰도를 위한 채점 연습이 모형의 주요한 단계를 구성하고 있다. 박종임(2013)은 온라인으로 이루어지는 채점자 훈련 프로그램의 모형을 논하였으나, 온라인 프로그램에서도 채점자의 정보를 제공하고 피드백을 제공함으로써 채점자 간 일치도 향상과 채점자 내 일관성 향상을 목표로 하고 있다는 점은 최윤곤·허유리(2013)가 제안한 모형과 공통되는 부분이다. 홍은실 외(2020)에서는 한국어 말하기 평가를 대상으로 하여, 채점자 간 협의에 바탕을 둔 워크숍 기반의 훈련을 제시하였다. 이 연구에서 채점자 간 협의 과정에 대해 질적 분석을 시행한 결과, 채점자가 협의를 통해서 서로를 보완하는 실제적인 양상이 확인되었다.

이상의 논의를 참고하여 오프라인 교육에서는 평가 도구에 대한 이해, 채점 실전 연습, 연습 결과 정리 등의 과정을 밟아 나가면서, 실제 채점 과정을 체험해 보고 채점 과정에서 맞닥뜨릴 수 있는 다양한 문제 상황을 해결하는 방법과 태도를 익힐 수 있도록 교육 내용을 구성하였다. 특히 국립국어원·서울대학교 국어교육연구소(2022)에서 유의미한 채점 자료를 추출 및 활용함으로써 채점의 실제성을 높이고 전문적인 채점 훈련이 이루어질 수 있도록 하였다.

<표 IV-7> 오프라인 채점자 교육과정 개요

교육 방식	교육 단계	교육 내용	세부 교육 내용	차시	강의 목차	교수자
오프라인	채점 실습	채점 도구 이해 실습	- 쓰기 평가 준거의 적용 방안 - 채점 도구의 실제 적용	2	채점 도구의 이해	장성민 이성준
					채점 도구의 적용	
		채점 신뢰도 이해 실습	- 채점 활동 1, 2 - 채점자 간 일치도 점검	5	채점 활동 1(사전, 개별)	
					채점 신뢰도의 이해	
					채점 신뢰도 분석	
		채점 실습 교육의 환류	- 종합 토의 - 채점자 자기 평가 및 성찰	1	채점 활동 2(사후, 조별)	
					채점 신뢰도의 향상 방안	
종합 토의						
자기 평가 및 성찰						

오프라인 교육과정은 채점 실습 단계에 해당하며, 채점 도구 이해 실습 2차시, 채점 신뢰도 이해 실습 5차시, 채점 실습 교육의 환류 1차시로 구성하였다.

(1) 채점 도구 이해 실습

채점 도구 이해 실습에서는 채점 실습에 필요한 채점 도구에 대한 이해를 도모하고 실제 평가 도구에 적용해 보는 것을 목표로 삼는다. 이를 위해 일반적인 쓰기 평가 준거의 특성 및 적용 방안을 제시하고 그에 대한 적용으로서 채점자가 채점 과정에서 경험하게 될 쓰기 평가 도구(개발 목적, 문항, 채점 기준표 등)를 소개한다.

<표 IV-8> 채점 도구 이해 실습 강의 요목

강의 목차	세부 항목
채점 도구의 이해	<ul style="list-style-type: none"> □ 쓰기 평가 준거에 대한 일반적 이해(내용, 조직, 표현) □ 쓰기 장르별 특성에 따른 쓰기 평가 준거 적용 방안(정보 전달, 설득)
채점 도구의 적용	<ul style="list-style-type: none"> □ 실제 평가 도구(문항, 척도, 채점 기준표) 소개

(2) 채점 신뢰도 이해 실습

채점 신뢰도 이해 실습에서는 채점 신뢰도 해석에 필요한 기본 개념(평균, 표준편차, 척도 활용의 적합도 등)을 이해하고 실제 채점과 소그룹 활동을 통한 채점자의 인식 조

정을 경험하는 것을 목표로 삼는다. 이를 위해 신뢰도의 분석과 해석, 채점 과정에서의 신뢰도 확보 방법, 채점 실습 및 소그룹 활동 등 다음의 항목들이 교육 내용으로 포함되도록 강의안을 구성하였다. 특히 본 강의에서 온라인 교육 종료 후 실시한 채점 과제의 결과를 제시하고, 그 과정에서 경험한 내용들을 공유하여 온라인 교육과정과의 연계성을 확보하고자 하였다.

<표 IV-9> 채점 신뢰도 이해 실습 강의 요목

강의 목차	세부 항목
채점 활동 1(사전, 개별)	<ul style="list-style-type: none"> □ 온라인 채점 과제 신뢰도 분석 결과 공유(집단-개인) □ 개인별 채점 연습(총체적-분석적 방법, 문항 유형별 각 1개 글)
채점 신뢰도의 이해	<ul style="list-style-type: none"> □ 언어 평가의 신뢰도 유형과 쓰기 평가에서의 확보 방법 □ 채점 신뢰도에 영향을 미치는 변인
채점 신뢰도 분석	<ul style="list-style-type: none"> □ 기술통계를 통한 신뢰도 분석 □ 상관, 라쉬 모형을 통한 채점 신뢰도의 이해
채점 활동 2(사후, 조별)	<ul style="list-style-type: none"> □ 채점 활동 1의 결과에 나타난 기술통계 값 확인 □ 조별 채점 연습(총체적-분석적 방법, 문항 유형별 각 1개 글)
채점 신뢰도의 향상 방안	<ul style="list-style-type: none"> □ 총체적 채점 활동 경험 공유 □ 분석적 채점 활동 경험 공유

(3) 채점 실습 교육의 환류

채점 실습 교육의 환류에서는 본 교육 경험과 실제 평가 및 채점 상황 간의 연결점을 확인하고 활용 방안을 탐색하는 것을 목표로 한다. 이를 위해 종합 토의 및 질의를 통해 전체 오프라인 교육 내용을 정리한다. 그리고 기개발한 사전·사후 설문 도구를 통해 교육 만족도 조사와 자기 평가를 실시하여 채점 교육 참여자의 인식을 스스로 점검 및 확인할 수 있도록 강의안을 구성하였다.

<표 IV-10> 채점 실습 교육의 환류 강의 요목

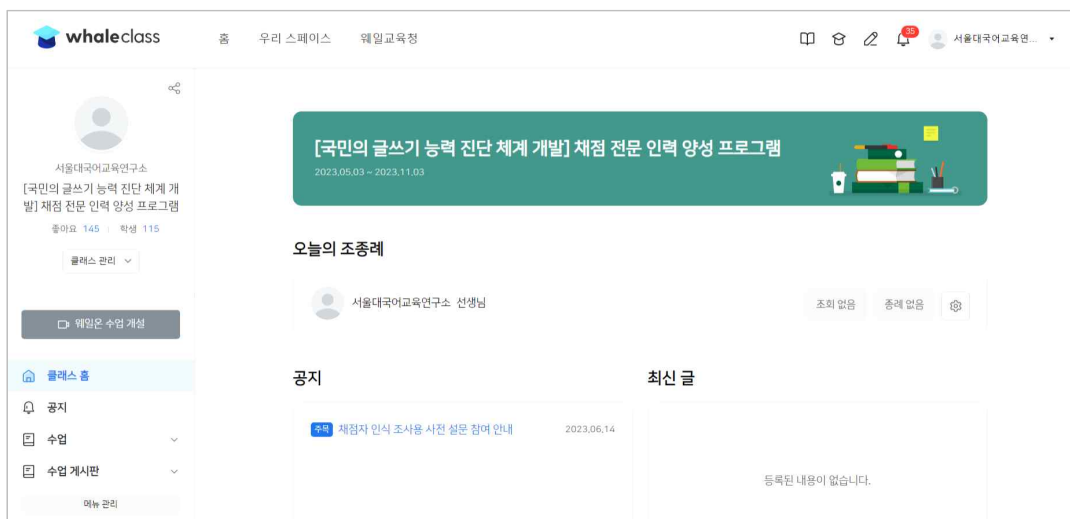
강의 목차	세부 항목
종합 토의 및 자기 평가	<ul style="list-style-type: none"> □ 채점 도구 및 신뢰도 이해 실습에 대한 종합 토의 및 질의 □ 교육 결과의 활용 방안에 대한 안내 □ 자기 평가를 통한 인식 점검 및 조정

2. 채점자 교육 프로그램 개발

1) 온라인 강의 자료 개발

(1) 채점자 교육 프로그램 운영 플랫폼 확정

온-오프라인 블렌디드로 개발된 본 프로그램의 운영을 위해서는 온라인 강의 플랫폼이 필수적이다. 온라인 강의 플랫폼은 12차시 분량의 온라인 강의 동영상 탑재와 과제 부여 및 제출, 이수 사정을 위한 진도율 확인 등의 기능이 필요하다. 이러한 기능을 갖춘 온라인 강의 운영 플랫폼으로서 네이버 웨일클래스(whaleclass)를 채택하였다. 웨일클래스는 네이버 웨일 스페이스에서 제공하는 온라인 수업 관리 서비스(LMS)로 동영상, 문서, 실시간 방송 등을 통한 수업을 구성할 수 있으며, 설문, 퀴즈, 과제 개설이 가능하다. 또한 수강생별로 동영상 강의 수강률을 포함한 진도율을 확인할 수 있고 이를 기반으로 관리자가 수수료 기준을 설정할 수 있다. 이와 같이 웨일클래스가 본 프로그램의 운영 및 관리에 필요한 기능을 두루 갖추고 있다고 판단하여 네이버 웨일스페이스 측과 플랫폼 사용에 관한 협의를 진행하였고, 최종적으로 23년 12월까지, 약 100명의 수강생이 플랫폼을 이용하기로 하였다.



[그림 IV-1] 웨일클래스 진입 화면 예시

(2) 강의 영상 제작 의뢰

온라인 강의 영상 제작은 서울대학교 내의 예비 교원 및 현직 교원의 원격교육 지원 시설인 디스페이스(D-Space, <https://snudspace.modoo.at/>)에 촬영 및 편집을 의뢰하여 크로마키형 영상 촬영 방식으로 진행하였다.

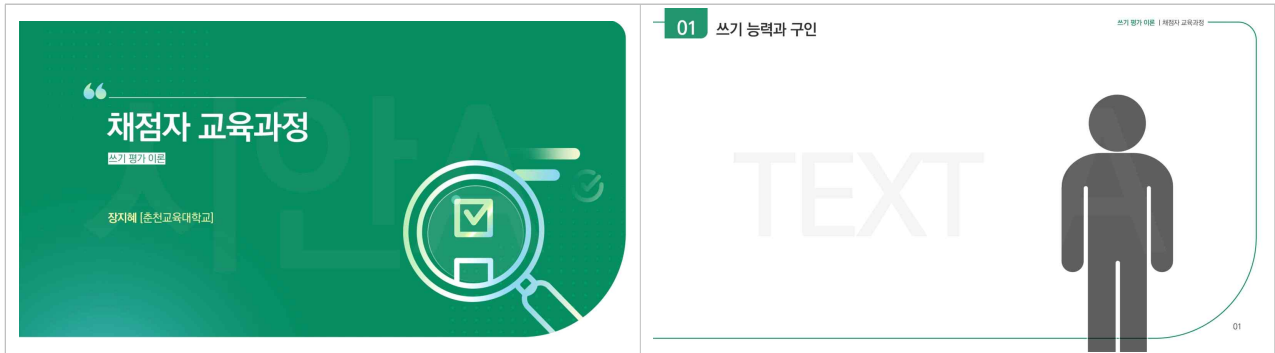
(3) 강의 자료 및 스크립트 개발

본격적인 강의 자료 제작에 앞서 연구진과 교수진이 온라인 강의 자료 개발에 관한 회의를 진행하였다. 연구진 측에서는 강의 개발에 필요한 자료(<표 IV-11>)를 사전에 교수진에 전달하였으며, 연구진과 교수진이 협업하여 개발한 강의 설계안을 바탕으로 교수자가 강의 자료 및 스크립트를 작성하였다.

<표 IV-11> 온라인 강의 개발에 필요한 자료

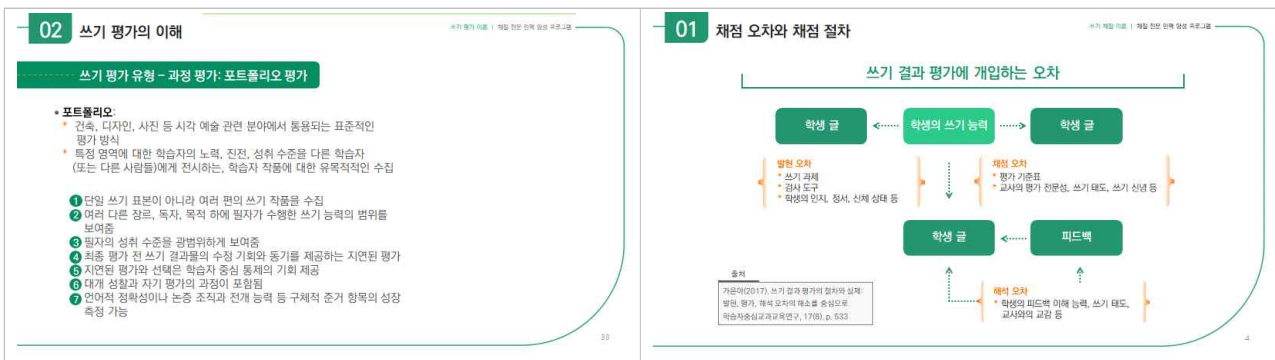
- ㉠ 채점 기준표 개발 과정 등을 포함하여 사업의 전반적인 맥락을 파악할 수 있는 자료(2022년 기초 연구 보고서, 2023년 착수 보고서 등)
- ㉡ 채점자 간 불일치가 높았던(예: 3점 이상 점수 차) 사례
- ㉢ 총점 기준 ‘상-중-하’에 해당하는 대표적 답안
- ㉣ 준거-척도별 대표 샘플
- ㉤ 오프라인 교육과정과의 연계를 위한 사전 채점용 글쓰기 자료(3개 문항당 상-중-하 수준 각 3편, 총 9편)
- ㉥ 정제된 채점 기준표(표현 영역 계량화 결과 포함)

강의 자료 개발 과정에서 글쓰기 자료 및 예시 답안은 2022년 연구에서 산출된 자료를 적극적으로 활용하도록 하였다. 강의 자료는 PPT 형태로 제작하였으며, 이때 연구진이 업체에 의뢰한 강의 자료 형식(PPT 템플릿)을 교수자에게 제공하여 전체적인 온라인 강의 영상 화면이 통일된 형식을 지닐 수 있도록 하였다.



[그림 IV-2] 강의 자료 템플릿(PPT) 예시

강의 자료 제작이 완료된 후에는 교수진과 연구진 간의 교차 검토를 진행하여 최종 강의 자료를 완성하였다. 완성된 강의 자료는 템플릿을 디자인한 업체에 의뢰하여 슬라이드 내 글꼴, 크기, 여백, 정렬 등 전반적인 정렬 작업을 진행하였다.



[그림 IV-3] 강의 자료 최종본 예시

(4) 강의 영상 촬영 및 편집

강의 영상 촬영은 서울대 디스페이스 측과 교수자의 일정을 고려하여 촬영 장소(서울대학교 10-1동 301호)를 대관하고 촬영을 진행하였다.

<표 IV-12> 강의 영상 촬영 일정

교수자	장지혜	박성석	최영인
강의명	쓰기 평가 이론	쓰기 채점 이론	쓰기 평가 준거별 채점 방법
촬영 일정	5월 14일(일) 오후 1시-5시	5월 8일(월) 오후 1시-5시	5월 18일(목) 오후 1시-5시
촬영 장소	서울대학교 10-1동 301호 스튜디오1		

촬영 이후 디스페이스 측에서 크로마키 배경에서 촬영된 강의 영상과 강의 자료(PPT)를 합성하고 인트로, 아웃트로, 음원 및 음성, 자막 등의 편집을 진행하였으며, 강의 영상 편집본은 오타자, 음성-영상 간 일치 등 영상 전반의 완성도 차원에서 연구진과 교수진의 최종 검토를 거쳐 확정되었다.

“ **채점 전문 인력 양성 프로그램** ”

쓰기 평가 이론

장지혜 [춘천교육대학교]



04 채점 경향과 채점 피드백

쓰기 채점 이론 | 채점 전문 인력 양성 프로그램

..... 채점 척도 활용 정보

- 채점자가 채점 척도를 얼마나 고르게 사용하였는가에 관한 정보.
- 일반적으로 쓰기 채점 시 채점 척도 사용에서 중앙 경향성을 보이게 됨.
- 중앙 경향성이 강하다면, 극단적인 점수를 피하려는 우유부단함의 극복, 잘 쓴 글과 잘못 쓴 글에 대한 변별 능력 신장 등의 노력을 기울여야 함.



61

03 '조직_문단 수준' 평가 준거의 이해

쓰기 평가 준거별 채점 방법 | 채점 전문 인력 양성 프로그램

뒷받침 문장을 사용하여 문단의 중심 문장을 뒷받침하였는가?
문단별로 하나의 중심 생각(소주제)이 전개되는가?
문단을 이루는 여러 문장들이 서로 긴밀한 관계로 연결되는가?
(예: 시간적, 논리적 순서)

[문단 수준]

형식 문단을 기준으로 채점하되, 형식 문단은 구별되지 않지만 내용 문단이 분명하게 드러나는 경우에는 내용 문단을 기준으로 각 구성 요건을 판단할 수 있음.

5	4	3	2	1
모든 문단에서 완결성, 통일성, 일관성을 갖춘.	대부분의 문단에서 완결성, 통일성, 일관성을 갖춘.	일부 문단의 완결성, 통일성, 일관성이 부족하나, 글을 읽는 데 크게 방해가 되지 않음.	문단의 완결성, 통일성, 일관성이 전 반적으로 부족하여 완성도가 낮음.	대부분의 문단에서 구성 요건을 갖추지 못함.



50

[그림 IV-4] 강의 영상 최종본 예시 화면

(5) 차시별 퀴즈 및 과제 개발

온라인 학습의 효과를 극대화하기 위해 채점 교육 참여자가 학습 내용을 스스로 점검할 수 있도록 퀴즈 풀이와 같은 교육 장치가 필요하다고 판단하였다. 이를 위해 온라인 강의 자료 및 영상의 최종본을 참고하여 연구진이 차시별 2문항씩 총 24개 문항을 제작하고 교수자의 검토를 거쳐 확정하였다.

<표 IV-13> 차시별 퀴즈 예시

<쓰기 채점 이론>

1차시. 채점 오차와 채점 절차

1) 다음 중 쓰기 결과 평가에 개입하는 오차가 아닌 것은?

- ① 발현 오차 ② 채점 오차 ③ 표준 오차 ④ 해석 오차

정답: ③ (PPT p.4)

2) 다음은 쓰기 평가 및 채점의 일반 절차에 해당하는 항목이다. 항목을 절차에 맞게 바르게 나열한 것은?

- ㄱ. 글쓰기
- ㄴ. 채점, 성취 수준 확정, 피드백
- ㄷ. 평가 결과 해석
- ㄹ. 평가 목적 설정 및 평가 계획 수립
- ㅁ. 쓰기 과제, 평가 기준표 구안
- ㅂ. 채점자 협의회

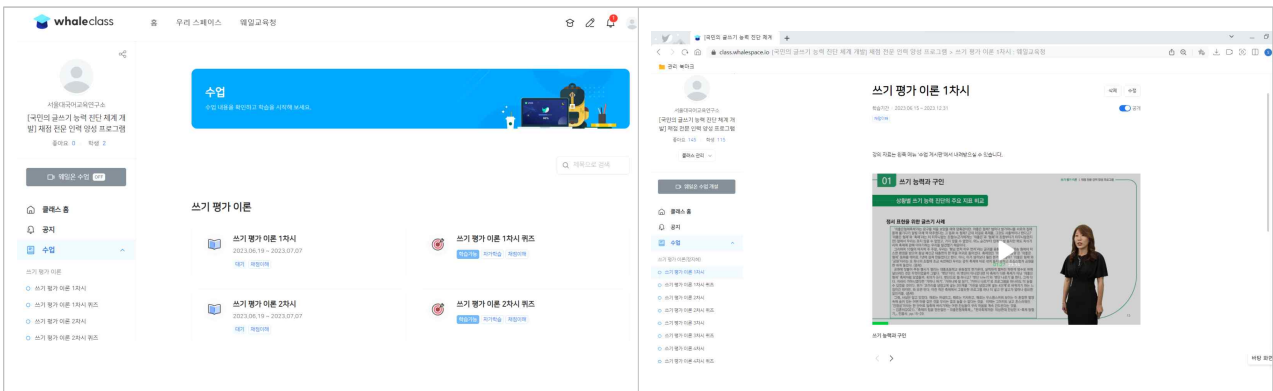
- ① ㄴ-ㄱ-ㅂ-ㄱ-ㄴ-ㄷ ② ㄴ-ㄹ-ㄱ-ㅂ-ㄴ-ㄷ
 ③ ㄴ-ㄱ-ㄱ-ㅂ-ㄴ-ㄷ ④ ㄴ-ㄹ-ㅂ-ㄱ-ㄷ-ㄴ

정답: ③ (PPT p.19)

또한 온라인 교육과정에서 학습한 내용을 적용하고 온라인 교육과 오프라인 교육 간 연계를 위한 과제를 개발하였다. 과제의 내용은 국립국어원·서울대학교 국어교육연구소(2022) 기초 연구에서 수집된 글쓰기 자료 9개를 채점하는 것으로 온라인 교육의 <쓰기 평가 준거별 채점 방법> 마지막 차시 강의를 학습한 후 채점 결과를 온라인으로 제출하도록 하였다. 9개의 글쓰기 자료는 위 기초 연구의 3개 문항을 통해 수집된 글쓰기 자료 중 각각 상-중-하 수준으로 선별하여 제시하였다.

(6) 강의 자료 탑재 및 배부

이상 개발한 강의 영상 및 자료(PPT), 차시별 퀴즈, 사전 채점 과제를 웨일클래스에 탑재하고 강의 영상 재생, 과제 제출, 진도율 확인 등 프로그램 이수 관리에 필요한 기능의 구동 유무를 점검하였다.



[그림 IV-5] 강의 자료 탑재 화면

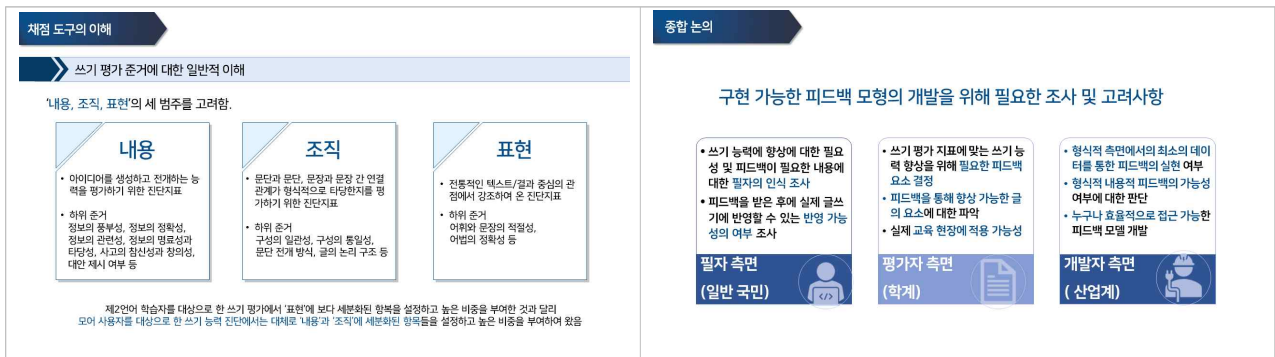
또한 교육 프로그램 소개 및 차시별 개요, 온라인 강의 이수 안내, 플랫폼 이용 방법, 온라인 강의 자료(PPT) 등을 포함하는 온라인 교육용 자료집(PDF)을 제작하여 사전 배부하였다. 교육용 자료집은 안내 자료와 교육 자료로 구분하여 다음 항목들로 구성되었다.

<표 IV-14> 온라인 교육용 자료집 목차

안내 자료	교육 자료
I. 채점 전문 인력 양성 프로그램 안내	I. <쓰기 평가 이론> 강의 자료
II. 세부 교육 내용	II. <쓰기 채점 이론> 강의 자료
III. 온라인 교육 플랫폼 이용 안내	III. <쓰기 평가 준거별 채점 방법> 강의 자료
IV. 과제 안내	IV. 과제① 차시별 퀴즈
V. 사전·사후 설문 조사 안내	V. 과제② 채점 실습

2) 오프라인 강의 자료 개발

오프라인 강의의 세부 내용은 앞서 보고한 바와 같이 채점자 훈련 프로그램이나 모형에 관한 연구(이영식, 2000; 박종임, 2013; 최윤곤·허유리, 2013; 홍은실 외, 2020) 등을 참고하여 채점자 워크숍 형태로 구성되도록 설계하였으며, 진단 지표에 대한 해석과 적용 과정에 대해서는 채점자 개인뿐만 아니라 소그룹 내에서 공유하고 조정을 경험할 수 있도록 하였다. 그리고 채점 과정에 대한 교수진(연구진으로 구성)의 즉각적인 모니터링에 의거해 채점 교육 참여자의 채점 성향별 맞춤형 피드백을 제공해 줌으로써, 채점 과정에 대한 이해도를 제고하고 채점 문식성이 신장될 수 있도록 교육 내용을 구안하였다.



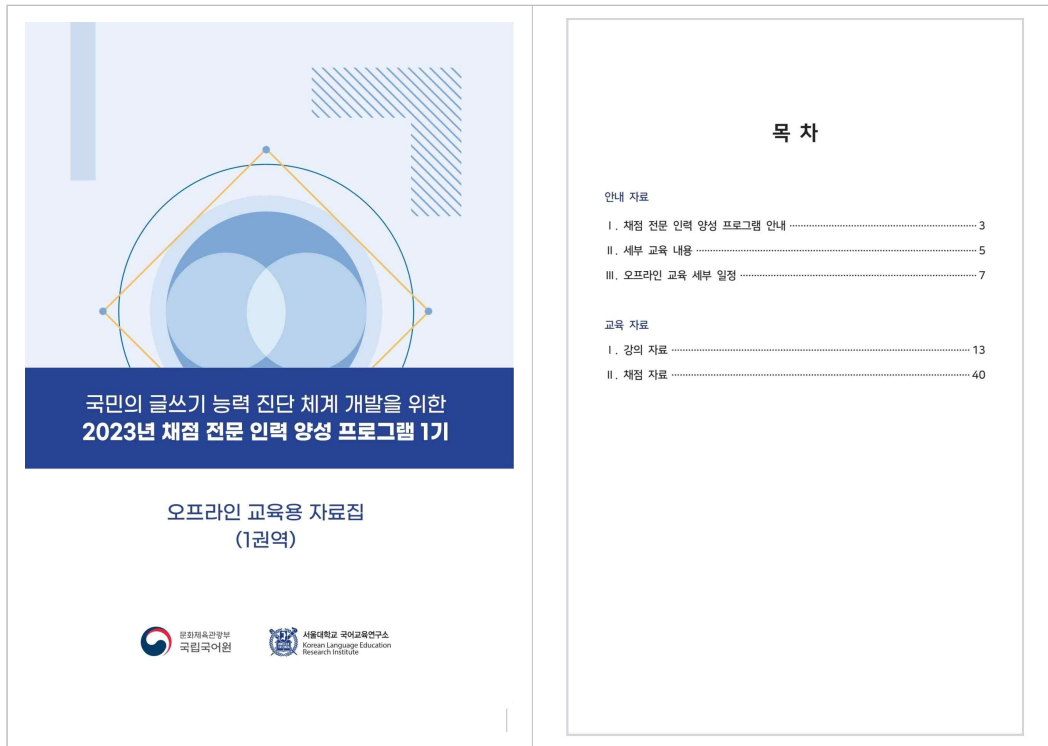
[그림 IV-6] 오프라인 강의 자료 예시

평가 도구의 해석과 적용 과정에 관한 활발한 소그룹 활동과 그 결과를 공유하기 위한 도구로서 '패드렛(<https://padlet.com/>)'을 활용하였다. 이를 통해 그룹별로 토의한 내용을 온라인 게시판에 작성 및 게시하고 전체 집단이 이를 공유함으로써 소그룹 활동의 효과를 극대화하고자 하였다.

오프라인 교육의 핵심이라 할 수 있는 채점 실습 자료로는 2022년 기초 연구에서 수집한 글쓰기 자료 중 1번과 2번 문항의 상-중-하 수준의 글을 선별하였다. 글쓰기 자료 선정 시에는 온라인 교육에서 예제로 다루어진 자료와 과제로 제시된 사전 채점 자료는 제외하였다.

이상의 자료들을 종합하여 채점자 교육용 프로그램 소개 및 차시별 개요, 오프라인 교육 일정 및 장소, 오프라인 강의 자료(PPT), 채점 실습 자료를 포함하는 채점자 교육용 자료집을 제작하였으며, 제작 이후 전체 연구진 검토 과정을 거쳐 확정된 후에 자료집을

인쇄하였다.



[그림 IV-7] 오프라인 교육용 자료집 표지 및 목차

3. 채점자 교육 프로그램 운영

1) 교육과정 운영 방식 확정

본 과업에서 개발한 채점자 교육과정이 채점 전문 인력 양성과 <세부 목표 V. 글쓰기 자료 채점 실행>을 위한 채점자 선발이라는 실제적 기능을 수행하기 위해서는 프로그램 이수자 관리가 필수적이다. 이를 위해 온라인 강의 수강률(12차시, 100%), 오프라인 강의 출석률(8차시, 100%), 강의 참여도(P/F)를 포함하는 채점자 교육 프로그램의 이수 조건을 마련하였다.

<표 IV-15> 채점자 교육과정 이수 조건

교육 방식	온라인(12차시)	오프라인(8차시)	강의 참여도
항목	수강률	출석률	P/F
이수 조건	100%	100%	
비고	교육 대상자에게 특별한 사정이 있는 경우, 이수 사정 회의를 거쳐 수업 결손을 보완하는 방식으로 이수 조건을 충족할 수 있음.		

위의 이수 조건에서 확인할 수 있듯이 프로그램 이수자는 20차시 온-오프라인 교육을 모두 수강해야 하며, 교육과정 설계상 온라인 교육은 오프라인 교육 이전에 수강 완료되어야 한다. 강의 참여도는 ①차시별 퀴즈 24문항 풀이와 ②채점 실습 과제 제출 여부로 평가한다.

본 프로그램의 모집 인원이 100명이라는 점과 전국 9개 협약 대학 글쓰기 교강사를 대상으로 한다는 점에서 오프라인 교육을 4개 권역으로 나누어 진행하였으며, 교육 장소의 경우 권역별로 참여 인원이 많은 협약 대학으로 섭외하였다.¹⁾

1) 단, 1권역의 경우 협약 대학(서울대, 인천대, 제주대) 참여자의 동선을 고려하여 국립국어원에서 교육을 진행하였다.

<표 IV-16> 권역별 오프라인 교육 일정 및 장소

권역	참여 협약 대학	교육 장소	일정	실무진
1권역	서울대, 인천대, 제주대	국립국어원	7월 14일(금)	남가영, 이성준, 오예림, 송규민
2권역	충남대, 충북대	충남대	7월 15일(토)	장성민, 권은선, 김동섭, 홍태경
3권역	전북대, 전남대	전남대	7월 21일(금)	장성민, 권은선, 정은선, 홍태경
4권역	부산대, 경상국립대	경상국립대	7월 22일(토)	민병곤, 이성준, 오예림, 송규민

2) 참여자 모집

본 프로그램의 모집 정원은 100명으로, 총 3차에 걸쳐 모집이 이루어졌다. 1차 모집(23년 5월 1일-12일)은 9개 협약 대학에서 글쓰기 자료 수집을 담당하는 글쓰기 교강사를 대상으로 이루어졌으며, 대학 실무자를 통해 프로그램 안내 사항이 담긴 모집 문건을 전달하였다. 이때 프로그램 지원자에게 본인이 참여 가능한 오프라인 교육 일정에 표시하도록 하였으며, 전체 교육 일정과 이수 조건에 대해 반드시 확인하도록 하였다.

국민의 글쓰기 능력 진단체계 개발을 위한 채점 전문 인력 양성 프로그램 참여자 모집

☑ 참여 대상 및 인원
- 협약 대학의 글쓰기 자료 수집 참여 교강사 100명 내외

☑ 프로그램 내용 및 운영 방식
- 채점 실무-실습 중심의 온라인 교육 12차시, 오프라인 교육 8차시, 총 20차시
- 온라인 교육: 플랫폼을 통한 동영상 강의 시청(차시당 25분 내외)
- 오프라인 교육: 4개 권역별(서울·경기, 충청, 호남, 경남) 현장 집합 교육

교육 방식	교육 단계	교육 내용	세부 내용	차시
온라인	채점 이해	쓰기 평가 이론	쓰기 능력과 쓰기 평가의 이해 - 논중적 글쓰기의 평가 지표	4
	채점 적용	쓰기 채점 이론	쓰기 평가를 위한 채점 절차 - 채점 신뢰도 및 영향 요인	4
	채점 적용	준거별 채점 방법	내용, 조직, 표현 범주 채점 방법	4
오프라인	채점 도구 이해 실습	채점 도구 이해 실습	쓰기 평가 준거의 적용 방안 - 채점 도구의 실제 적용	2
	채점 실습	채점 신뢰도 이해 실습	예시, 답안 채점 - 채점자 간 일치도 점검	5
	채점 실습	교육의 환류	종합 토론 및 교육 결과의 활용 - 채점자 자기 평가 및 성찰	1

☑ 교육 일정
- 온라인 강의 수강 기간: 6월 19일(월) ~ 7월 7일(금), 3주간
- 오프라인 교육 일정 및 장소
· 권역별 진행(09:30~17:30)
· 각 권역 대학 중 참여 인원이 많은 곳으로 장소 확정 후 통보 예정
· 개인 노트북 지참 필수

권역	학교명	일정	비고
1권역	서울대, 인천대, 제주대	7월 14일(금)	
2권역	충남대, 충북대	7월 15일(토)	참가자 개인 사정에 따라 소속 대학 권역이 아닌 다른 권역 교육에 참여할 수 있음.
3권역	전북대, 전남대	7월 21일(금)	
4권역	부산대, 경상국립대	7월 22일(토)	

☑ 이수 조건
① 온라인 강의 수강률 100% ② 오프라인 강의 출석률 100% ③ 강의 참여도(P/F)

☑ 활용
- 본 프로그램의 이수자는 국립국어원의 <2023년 국민의 글쓰기 능력 진단체계 개발> 사업의 글쓰기 채점자로 참여 가능

☑ 지원
- 오프라인 교육 참여 수당 및 소정의 교통비 지급

☑ 모집 기간
- 5월 1일(월) ~ 5월 12일(금), 2주간

☑ 참여자 모집 및 선정 방법
- 아래 링크를 통해 참여자 모집 정보 작성
· <https://forms.gle/rqkULx91rVt6h15j9>
- 소속 기관 권역의 오프라인 교육 일정에 참여 가능한 사람을 1순위로 선발 예정
- 권역별 균형 등을 고려해 최종 참여자 선정 후 통보 예정

문의: 연구보조원 오예림 oy174190@snu.ac.kr, 송규민 songgm26@snu.ac.kr

[그림 IV-8] 채점 전문 인력 양성 프로그램 참여자 모집 문건

1차 모집 결과 49명이 지원하여 2차 모집은 협약 대학의 글쓰기 교강사 전체로 대상을 확대하여 13명이 추가로 모집되었다. 정원 충원을 위해 계속해서 3차 모집을 진행하게 되었다. 3차 모집에서는 협약 대학의 인근 대학의 글쓰기 교강사로 모집 대상을 확대하였으며, 전국의 약 30개 대학에서 총 109명을 모집하였다. 최종 모집 현황은 다음과 같다.

<표 IV-17> 채점 전문 인력 양성 프로그램 최종 모집 현황

권역	학교	인원(명)	
1권역*(22)	서울대	3	32
	인천대	1	
	제주대	4	
	인하대	2	
	수원대	4	
	이화여대	3	
	연세대	2	
	성균관대	4	
	기타**	9	
2권역(23)	충남대	8	29
	충북대	5	
	청주대	1	
	배재대	5	
	공주대	7	
	서원대	2	
	국립한국교통대	1	
3권역(36)	전북대	8	19
	전남대	6	
	순천대	1	
	조선대	3	
	기타	1	
4권역(46)	부산대	9	29
	경상국립대	17	
	대구대	2	
	기타	1	
		109	

*()=권역별 글쓰기 자료 수집 참여 교강사 수

**기타=해당 권역 외 대학 참여자

3) 프로그램 시행

최종 선정된 채점 교육 참여자 109명을 대상에게 교육 일정, 수강 방식, 이수 조건 등 채점자 교육과정 전반과 프로그램 참여 확정에 대해 안내하였으며, 교육과정이 본격적으로 시작되기 전에 온라인 교육용 자료집(PDF)을 메일로 송부하였다.

(1) 온라인 교육과정 실행(6월-7월)

온라인 교육 기간은 6월 19일(월)부터 7월 7일(금)까지 3주간으로, 정해진 기한 내에 교육 참여자가 웨일클래스에 접속하여 온라인 강의를 수강하도록 하였다. 강의 참여도(P/F)

의 판정 기준에 해당하는 차시별 퀴즈와 채점 과제는 7월 12일(수)까지 제출하도록 하였다. 교육 시작 전 안내 과정에서 2명이 포기하여 총 107명이 온라인 교육을 시작하였다.

온라인 교육이 시작되기 전에 참여자별 웨일클래스 접속 계정을 생성하고 최초 접속 시 관리자의 승인을 받은 후에 클래스에 진입이 가능하도록 설정하였다. 온라인 교육 기간에는 담당 연구원이 개인별 학습 현황을 수시로 확인하여 필요한 경우 개별 연락을 통해 학습을 독려하고 수강 문제 등을 지원하였다. 특히 오프라인 교육과정이 시작되기 전에 온라인 교육과정을 모두 이수할 수 있도록 교육 참여자의 온라인 학습 과정을 체계적으로 관리하였다. 온라인 교육 기간에 교육 시작 인원 107명 중 6명이 중도 포기하여 101명이 온라인 교육과정을 이수하였다.

수업	동영상	접속시간						
수업	학습기간	상태	이름	아이디	동영상명	재생회수	총 재생시간	재생률(%)
쓰기 평가 이론(강지현) :: 쓰기 평가 이론 1차시[LECTURE]	2023.06.15 00:00 ~ 2023.12.3 1 23:59	공개	간보은	writing_sco66@whalespace.org	장지현교수님1차시_수정.mp4	4	00:23:35	100
쓰기 평가 이론(강지현) :: 쓰기 평가 이론 1차시[LECTURE]	2023.06.15 00:00 ~ 2023.12.3 1 23:59	공개	김미영	writing_sco40@whalespace.org	장지현교수님1차시_수정.mp4	4	00:39:40	100
쓰기 평가 이론(강지현) :: 쓰기 평가 이론 1차시[LECTURE]	2023.06.15 00:00 ~ 2023.12.3 1 23:59	공개	강민정	writing_sco42@whalespace.org	장지현교수님1차시_수정.mp4	2	00:24:43	100
쓰기 평가 이론(강지현) :: 쓰기 평가 이론 1차시[LECTURE]	2023.06.15 00:00 ~ 2023.12.3 1 23:59	공개	김수영	writing_sco14@whalespace.org	장지현교수님1차시_수정.mp4	1	00:23:14	100
쓰기 평가 이론(강지현) :: 쓰기 평가 이론 1차시[LECTURE]	2023.06.15 00:00 ~ 2023.12.3 1 23:59	공개	김수원	writing_sco72@whalespace.org	장지현교수님1차시_수정.mp4	2	00:26:00	100
쓰기 평가 이론(강지현) :: 쓰기 평가 이론 1차시[LECTURE]	2023.06.15 00:00 ~ 2023.12.3 1 23:59	공개	김지연	writing_sco59@whalespace.org	장지현교수님1차시_수정.mp4	1	00:23:14	100

[그림 IV-9] 온라인 학습 현황 관리 화면

(2) 오프라인 교육과정 실행(7월)

오프라인 교육과정은 7월 중 이루어졌으며 모집 당시 참여자가 표시한 오프라인 교육 일정과 인원수 등을 고려하여 권역별 인원을 배분하였다. 오프라인 교육 참여 대상 인원은 온라인 교육과정을 모두 이수한 101명이며, 참여자들에게 오프라인 교육 권역과 일정에 대해 안내하는 과정에서 일부 참여자의 권역 이동 요청이 있을 경우 원활한 교육을 위한 최대 수용 인원을 넘지 않는 선에서 최대한 허용하였으며 이에 다음과 같이 권역별 인원을 확정하였다.

<표 IV-18> 권역별 오프라인 교육 개요

권역	인원	장소	일자	시간
1권역	32명	국립국어원 4층 강의실	7월 14일(금)	09:30-17:00
2권역	27명	충남대학교 인문대학 254호	7월 15일(토)	
3권역	16명	전남대학교 진리관 507호	7월 21일(금)	
4권역	26명	경상국립대 인문대학 101동 352호	7월 22일(토)	

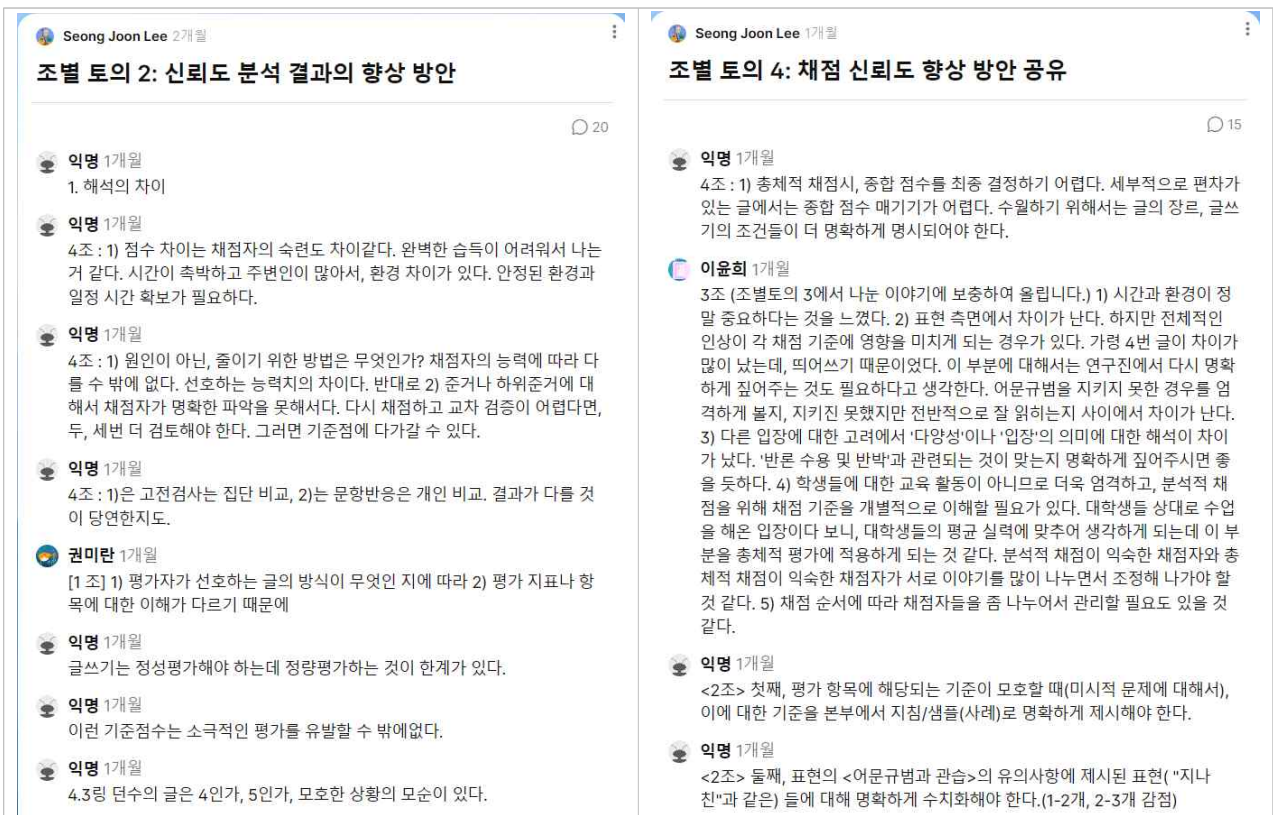
오프라인 교육의 원활한 진행을 위해 담당 연구진이 권역별 교육 장소의 실무자와 소통하여 강의실 대여와 강의실 장비 확인, 점심 식사, 주차 등 사전 준비를 진행하였다. 또한 권역별 교수자가 다르다는 점(1/4권역: 이성준, 2/3권역: 장성민)을 보완하기 위해 연구진 내부에서 오프라인 교육 사전 시뮬레이션을 진행(23년 7월 11일)하여 교육 내용을 표준화하였다. 이 과정에서 채점 활동 1과 2 진행을 위한 현장 채점 신뢰도 분석 도구를 마련하였으며 사전 시뮬레이션의 결과를 바탕으로 다음과 같이 오프라인 교육의 세부 일정을 확정하였다.

<표 IV-19> 오프라인 교육 세부 일정

교육 내용	차시	목차	시간(1차시=45분)	교수자	
채점 도구 이해 실습	2	1. 채점 도구의 이해	9:30 - 11:00	1/4권역 이성준 (성균관대)	
		2. 채점 도구의 적용			
		쉬는 시간	11:00 - 11:15		
채점 신뢰도 이해 실습	5	3. 사전 채점 결과 검토	11:15 - 12:00		2/3권역 장성민 (인하대)
		점심시간	12:00 - 13:00		
		4. 채점 활동 1	13:00 - 13:45		
		5. 채점 신뢰도 분석	13:45 - 14:30		
		6. 채점 활동 2	14:30 - 15:15		
쉬는 시간	15:15 - 15:30				
7. 채점 신뢰도 향상 방안 협의	15:30 - 16:15				
채점 실습 교육의 환류	1	8. 종합 논의 및 자기 평가	16:15 - 17:00		

오프라인 교육 당일 진행을 위해 권역별로 교수자를 포함한 담당 연구진이 교육 시작 전에 현장 점검 및 강의에 필요한 장비 준비를 하였으며 자료집 배부, 출석 확인, 소그룹 활동 보조, 현장 채점 신뢰도 분석 등의 업무를 분담하였다.

오프라인 교육 중에는 4차례의 조별 토의 활동을 통해 소그룹 내에서 참여자 간 의견을 교환하고 그 결과를 온라인 패들렛에 기록함으로써 소그룹에서 토의한 내용을 전체 집단과 공유할 수 있도록 하였다.



[그림 IV-10] 패들렛을 통한 소그룹 활동 결과 공유 화면 예시

또한 두 차례에 걸친 현장 채점을 통해 얻어진 결과는 교육 참여자가 실시간으로 제출한 채점 결과값을 모니터링 및 분석하여 그에 관한 기술통계값(평균, 표준편차)과 신뢰도(전체 채점자-개별 채점자의 상관계수)를 확인하였다. 특히 1차 채점에서 이러한 확인 및 공유를 거치고 2차 채점에서 그 결과값을 다시금 현장에서 공유함으로써 교육 참여자 간 협의가 이들의 인식에 영향을 미쳤는지를 확인할 수 있도록 하였다.

현장 채점 결과 검토								
현장 채점 과제 신뢰도 분석(채점활동1)								
<ul style="list-style-type: none"> 상관 분석 전체 채점자-개별 채점자의 상관 분석 해석: 0.4(약한 상관), 0.5~0.6(보통 상관), 0.7~0.8(좋은 상관), 0.9~(강한 상관) 								
RA1	RA2	RA3	RA4	RA5	RA6	RA7	RA8	
0.85	0.86	0.89	0.65	0.48	0.26	0.75		
RA9	RA10	RA11	RA12	RA13	RA14	RA15	RA16	
0.63	0.47	0.84	0.71	0.77	0.82	0.70	0.80	
RA17	RA18	RA19	RA20	RA21	RA22	RA23	RA24	
0.79	0.64	0.63	0.76	0.81	0.60	0.79	0.67	

현장 채점 결과 검토								
현장 채점 과제 신뢰도 분석(채점활동2)								
<ul style="list-style-type: none"> 상관 분석 전체 채점자-개별 채점자의 상관 분석 해석: 0.4(약한 상관), 0.5~0.6(보통 상관), 0.7~0.8(좋은 상관), 0.9~(강한 상관) 								
RA1	RA2	RA3	RA4	RA5	RA6	RA7	RA8	
0.86	0.86	0.56	0.68	0.71	0.79	0.67	0.61	
RA9	RA10	RA11	RA12	RA13	RA14	RA15	RA16	
0.72	0.80	0.48	0.74	0.86	0.67	0.78	0.81	
RA17	RA18	RA19	RA20	RA21	RA22	RA23	RA24	
0.72	0.79	0.86	0.76	0.56	0.52	0.82	0.79	

[그림 IV-11] 채점 활동 1, 2의 신뢰도 분석 결과 제시 내용

위의 [그림 IV-11]은 채점 활동 1과 2 직후 교육 참여자들의 현장 채점 결과를 분석하여 얻은 참여자 상관 분석 결과이다. 이 상관계수는 전체 참여자와 개별 참여자의 상관을 분석한 것으로 채점자 간 신뢰도를 의미한다. 참여자 간 토의를 거치지 않은 채점 활동 1의 결과에는 0.26, 0.47 등 상관이 낮은 참여자가 발견되었으나 토의 후 진행한 채점 활동 2에서는 대체로 높은 수준의 상관을 확인하였다. 위와 같은 결과를 모든 참여자가 공유함으로써 채점 활동에서의 인식 조정과 그것이 채점 신뢰도에 영향을 미치는 과정을 직접 경험할 수 있었다.



[그림 IV-12] 1-4권역 오프라인 교육 현장 사진

(3) 채점자 교육용 사전·사후 설문 조사 실시

채점자 교육과정을 시행하기 전과 후에 교육 참여자를 대상으로 설문 조사를 실시하여 쓰기 평가 및 채점에 관한 경험과 인식, 교육 프로그램 평가, 채점 인식 변화 등을 확인하고, 교육 프로그램 개선에 필요한 자료를 수집하였다.

사전 설문의 경우 프로그램 안내 메일에 구글 설문 링크를 함께 송부하여 온라인 교육 과정 시작 전에 작성 및 제출할 수 있도록 하였으며, 사후 설문은 오프라인 교육 당일 설문 링크와 큐알(QR) 코드를 전달하여 작성 및 제출하도록 하였다. 사전 설문은 응답 제출 확인을 위해 이메일 주소를 수집하였으나 사후 설문의 경우 만족도 조사의 성격이 강하다는 점을 고려하여 익명으로 진행하였다.

4. 채점자 교육 프로그램 평가

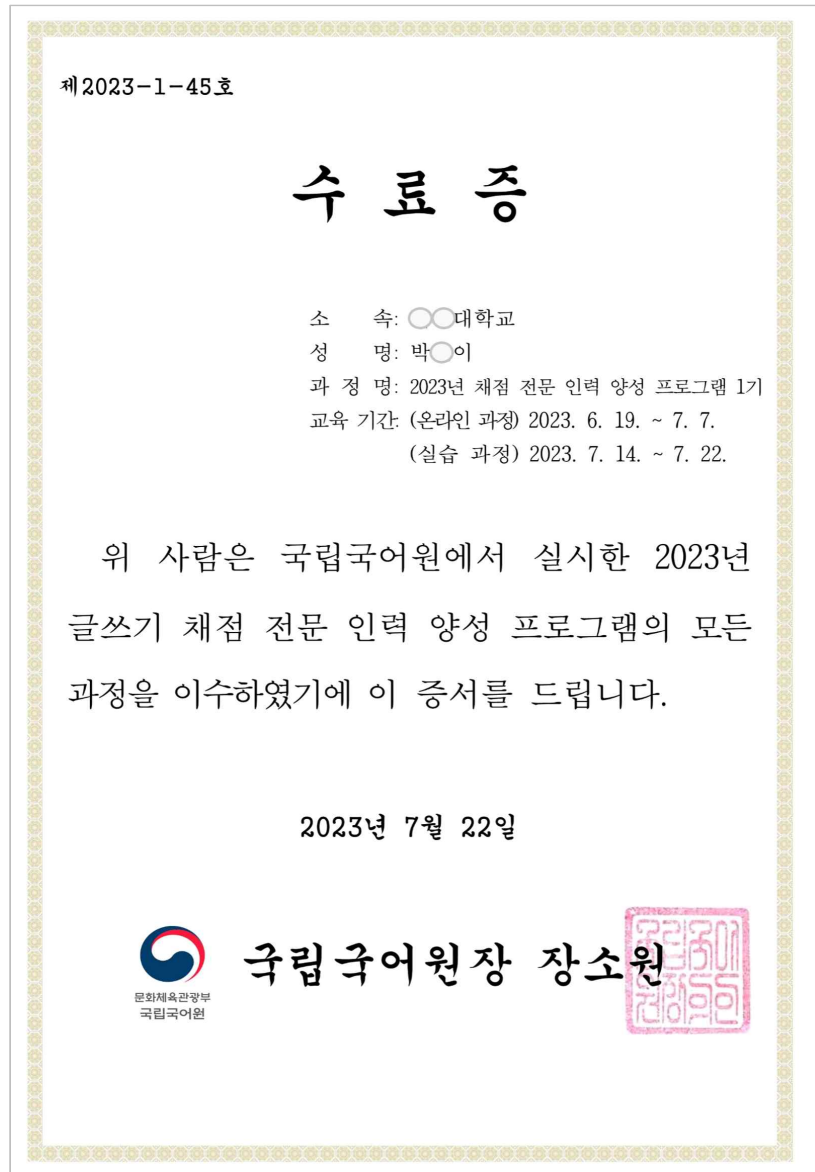
1) 이수 사정 및 수료증 배부

채점자 교육 프로그램 운영 결과에 의거하여 정해진 이수 기준에 따라 채점자 교육과정 이수 사정을 실시하였다. 우선 온라인 강의 12차시 수강률(100%), 차시별 퀴즈 풀이(총 24문항)와 사전 채점 과제 제출 여부를 포함하는 교육 참여도(P/F), 오프라인 강의 8차시 출석률(100%)을 바탕으로 총 96명²⁾의 이수자 목록을 작성하였다. 이수자를 대상으로 <2023년 채점 전문 인력 양성 프로그램 1기 수료증>을 전자 문서로 제작하여 개별 이메일을 통해 이수자에게 발송하였다. 대학별 프로그램 이수 인원은 다음과 같다.

<표 IV-20> 대학별 프로그램 이수 인원

대학	이수 인원(명)	
건국대학교	1	96
경상국립대학교	16	
고려대학교	1	
국립공주대학교	8	
국립한국교통대학교	1	
대구대학교	1	
배재대학교	4	
부산대학교	8	
서울대학교	3	
서원대학교	2	
성균관대학교	4	
수원대학교	4	
순천대학교	2	
연세대학교	1	
이화여자대학교	1	
인천대학교	1	
인하대학교	2	
전남대학교	6	
전북대학교	6	
제주대학교	4	
조선대학교	3	
청주대학교	1	
충남대학교	8	
충북대학교	7	
카이스트	1	

2) 온라인 교육을 이수한 101명 중 5명이 오프라인 교육 당일 불참하여 최종 수료 인원은 96명이다.



[그림 IV-13] 2023년 채점 전문 인력 양성 프로그램
수료증 양식

2) 채점자 교육에 관한 설문 조사 결과 분석

(1) 사전 설문

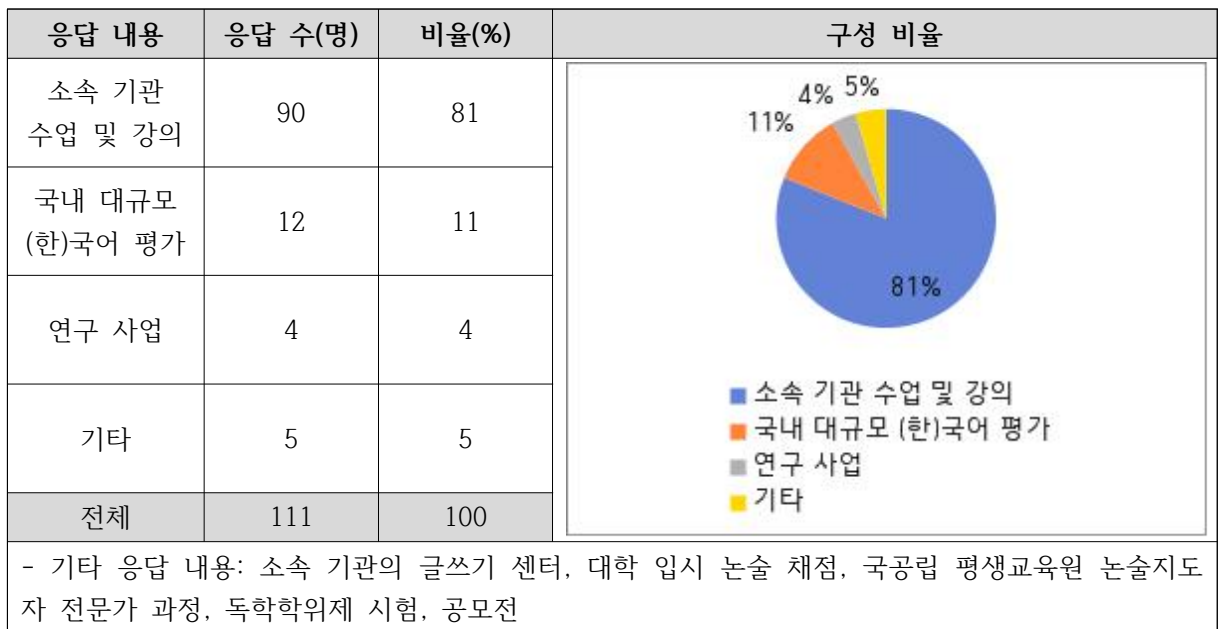
온라인 교육 전 실시한 사전 설문에는 총 107명³⁾이 참여하였으며 채점 및 채점자 훈

련 경험, 쓰기 평가 전문성에 관한 자기 인식, 채점자 교육의 필요성 등을 조사하였다. 응답 결과는 다음과 같다.

① 쓰기 채점 경험 여부

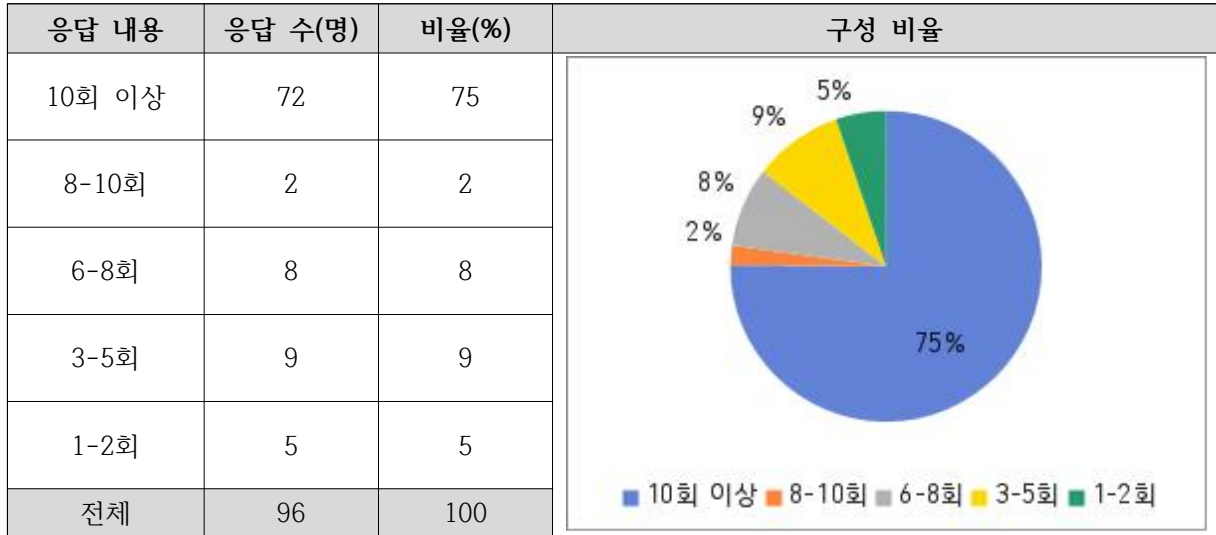


①-1 채점 경험 기관 및 사업(복수 응답)



3) 설문 응답 인원은 109명이나 그중 2명이 설문 결과 활용에 대해 '동의하지 않음'으로 응답하여 분석 및 보고 대상에서 제외하였다.

①-2 채점 경험 횟수



쓰기 채점 경험에 대해 조사한 결과, 응답 인원의 90%가 채점 경험이 있다고 응답하였으며 대부분이 소속 기관에서 수업 및 강의 글쓰기 자료를 채점한 것으로 나타났다(81%). 대규모 (한)국어 평가에서의 쓰기 채점 경험이 있는 경우는 107명 중 12명에 불과했다. 채점 경험 횟수는 채점 경험이 있는 참여자(96명)의 75%가 ‘10회 이상’으로 응답하였다.

② 쓰기 채점자 교육 경험 여부



②-1. 쓰기 채점자 교육 경험 기관 및 사업(복수 응답)

응답 내용	응답 수(명)	비율(%)	구성 비율
소속 기관 수업 및 강의	10	38	
국내 대규모 (한)국어 평가	11	42	
연구 사업	4	15	
기타	1	4	
전체	26	100	

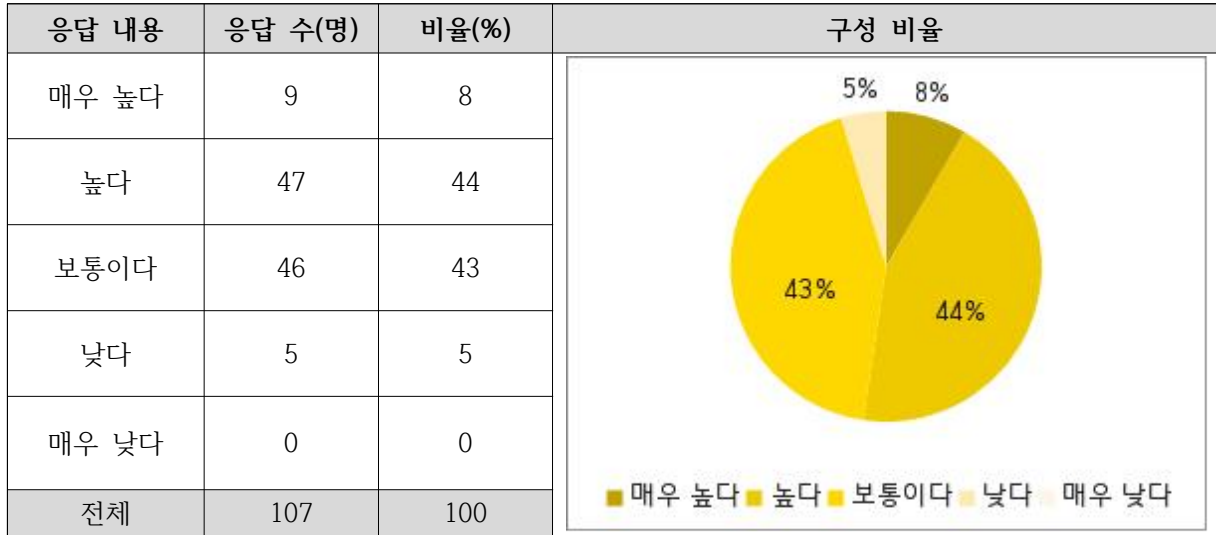
- 기타 응답 내용: 입학사정관 연수 과정

②-2 쓰기 채점자 교육 참여 횟수

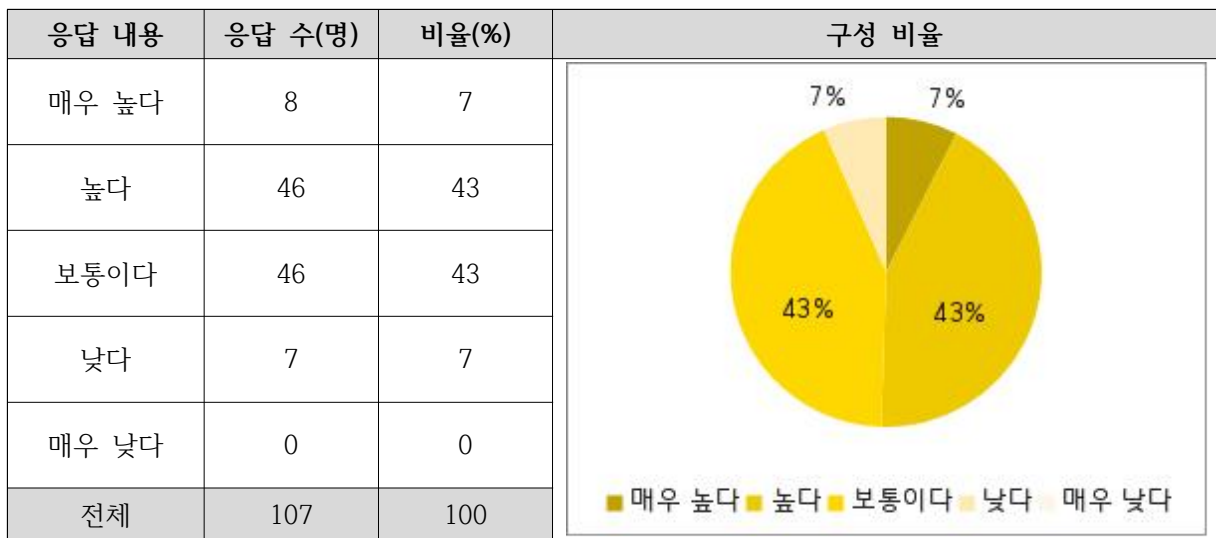
응답 내용	응답 수(명)	비율(%)	구성 비율
5회 이상	7	32	
4회	0	0	
3회	3	14	
2회	9	41	
1회	3	14	
전체	22	100	

쓰기 채점과 관련된 채점자 교육 경험에 대해서는 79%가 경험이 없다고 응답하였다. 쓰기 채점자 교육 경험이 있다고 응답한 22명 중 일부는 국내 대규모 (한)국어 평가에 참여한 경험이 있는 경우였으며(11명, 42%), 소속 기관 수업 및 강의에서도 채점자 교육을 받은 것으로 나타났다(10명, 38%). 쓰기 채점자 교육 참여 횟수에 대해서는 채점자 교육 경험이 있는 참여자의 약 70%가 '3회 이하', 32%가 '5회 이상'으로 응답하였다.

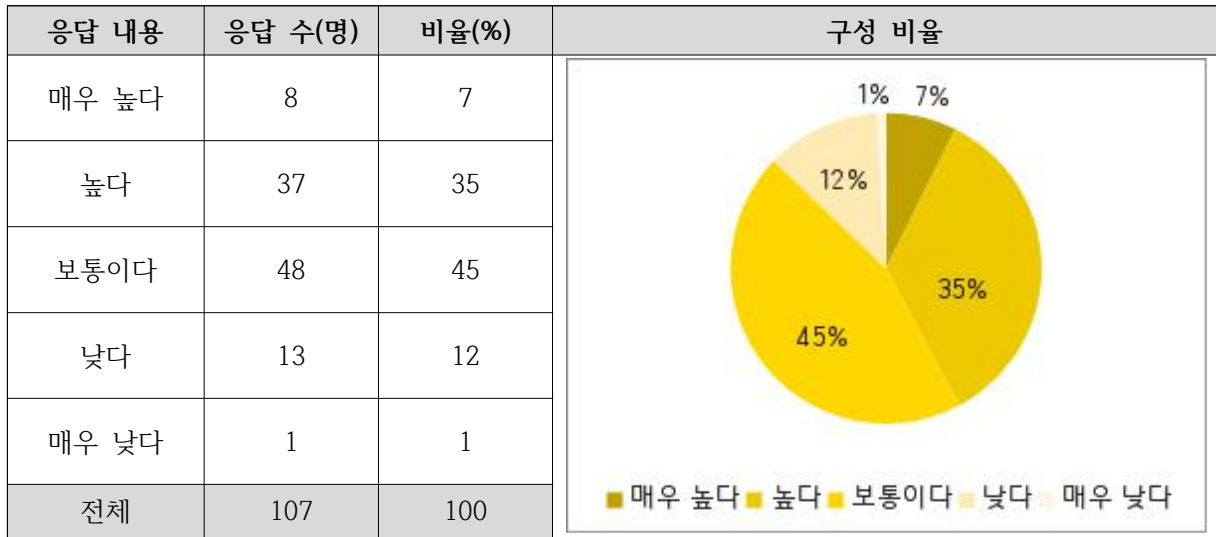
③-1 쓰기 평가 전문성 수준(지식 차원)



③-2 쓰기 평가 전문성 수준(수행 차원)

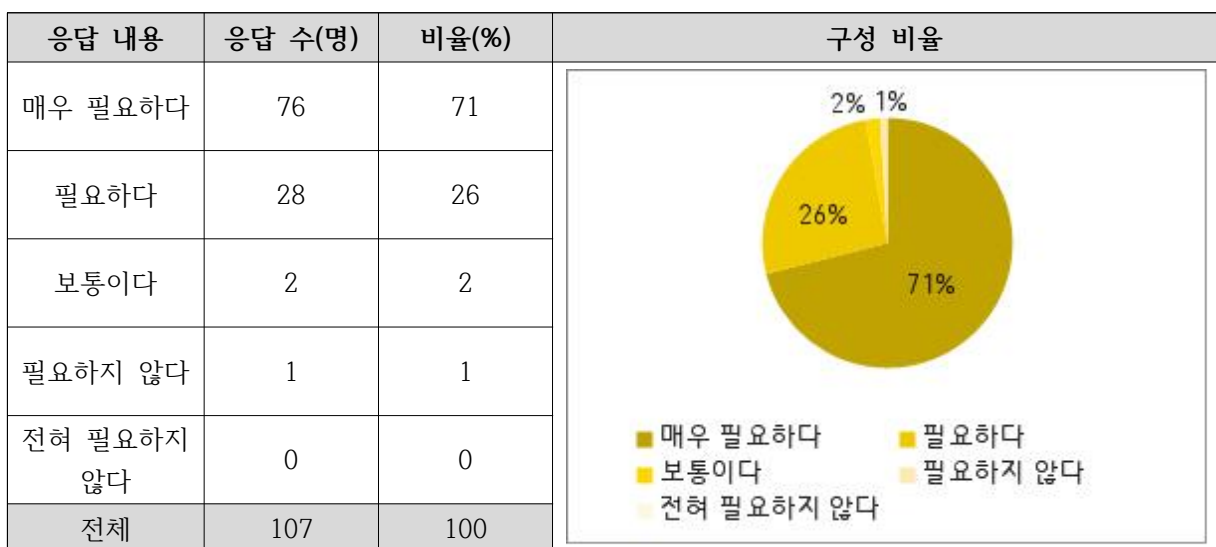


③-3 쓰기 평가 전문성 수준(활용 차원)



쓰기 평가 전문성에 대한 자기 인식 수준에 대한 조사 결과를 살펴보면 쓰기 채점에 대한 ‘지식’ 차원과 실제 ‘수행’ 차원, 결과 ‘활용’ 차원에서 80% 이상의 채점자가 ‘높다’와 ‘보통이다’에 응답한 것으로 나타났다. ‘매우 높다’에 응답한 비율은 7-8%였다. 이를 통해 응답자의 대부분이 자신의 쓰기 평가 전문성에 대해 보통 수준 이상으로 인식하고 있음을 알 수 있었다. 한편, 쓰기 채점 결과의 ‘활용’ 차원에서의 전문성 수준은 ‘낮다’에 응답한 비율이 12%로 참여자가 인식하는 ‘활용’ 차원에서의 전문성 수준이 ‘지식’ 차원(5%)과 ‘수행’ 차원(7%)에 비해 더 낮게 인식되고 있는 것으로 확인되었다.

④ 쓰기 평가 전문성 신장을 위한 채점자 교육의 필요성



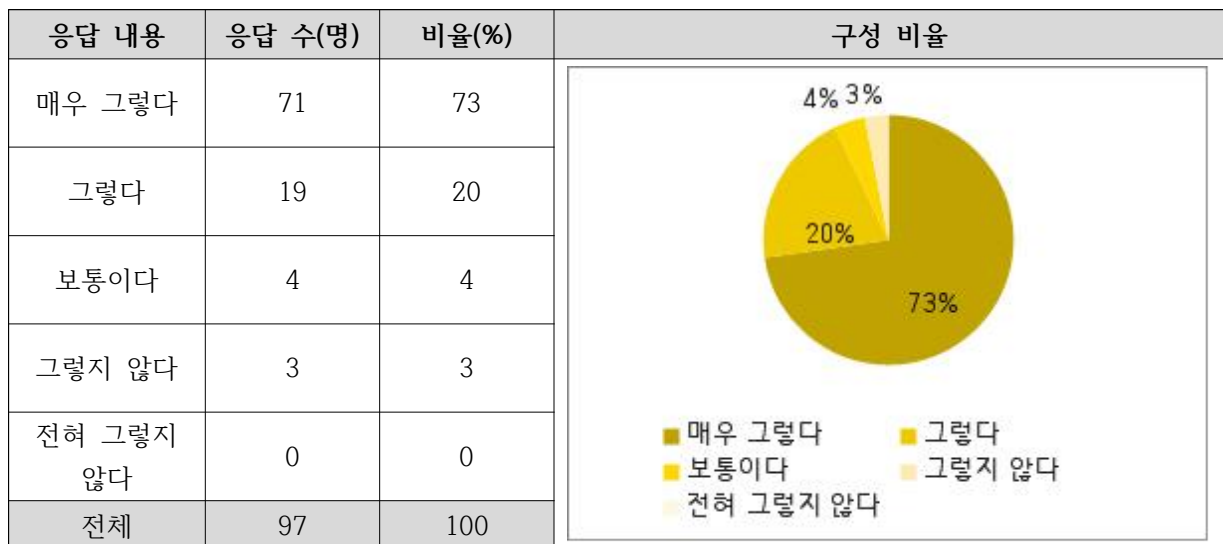
쓰기 평가 전문성 신장을 위한 채점자 교육의 필요성에 대해서는 ‘매우 필요하다’에 71%, ‘필요하다’에 26%가 응답하여 채점자 교육에 대한 필요성 인식 수준이 매우 높은 것으로 확인되었다.

이상의 사전 설문 결과를 요약하자면, 채점 교육 참여자의 약 90%가 쓰기 채점 경험이 있었으나, 쓰기 채점자 교육 경험이 있는 경우는 21%에 불과했다. 또한 참여자가 인식하는 쓰기 평가 전문성의 수준은 쓰기 채점에 대한 지식, 수행, 활용 차원에서 중-상 수준의 전문성 인식을 확인할 수 있었다. 쓰기 평가 전문성 신장을 위한 채점자 교육에 대한 필요성 인식은 매우 높게 나타났다.

(2) 사후 설문

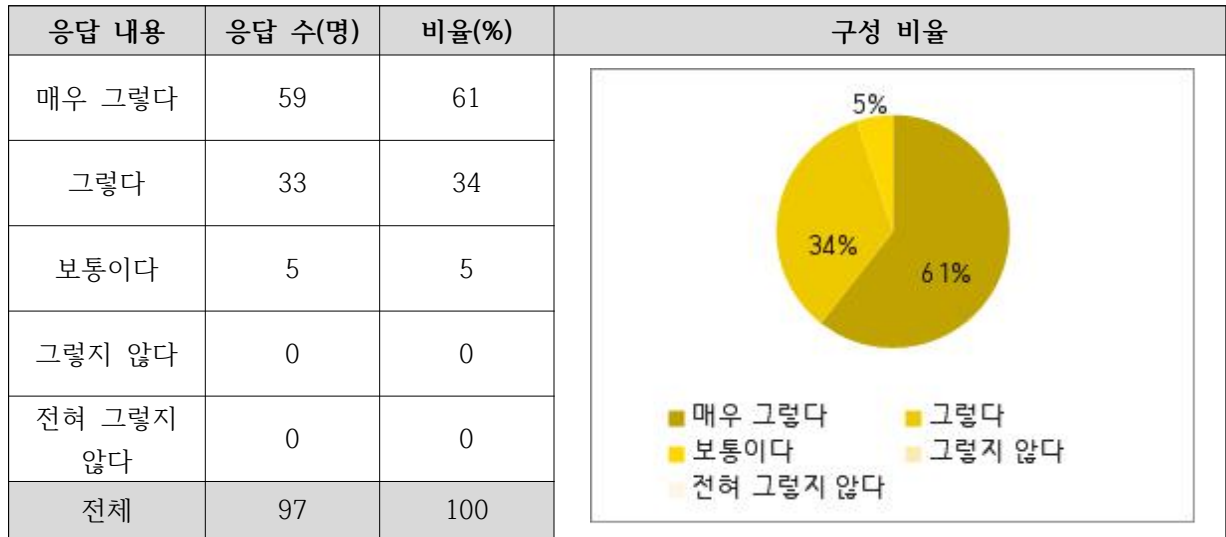
오프라인 교육 종료 후 실시한 사후 설문에는 총 97명⁴⁾이 참여하였다. 사후 설문은 온-오프라인 교육의 내용, 수준, 방식 등에 관한 만족도 조사와 프로그램의 개선점을 묻는 개방형 문항을 포함하고 있으며, 사전 설문과 동일하게 채점자 교육의 필요성을 표시하게 하여 교육 전과 후의 채점 인식 변화를 살펴보고자 하였다.

① (온라인 교육) 플랫폼을 통한 강의 수강의 편리성



4) 설문 응답 인원 99명 중 2명은 설문 결과 활용에 대해 ‘동의하지 않음’으로 응답하여 분석 및 보고 대상에서 제외하였다. 또한 오프라인 교육 참여 인원은 96명이나 만족도 조사의 특성상 익명으로 진행하였기 때문에 중복값을 확인할 수 없었다.

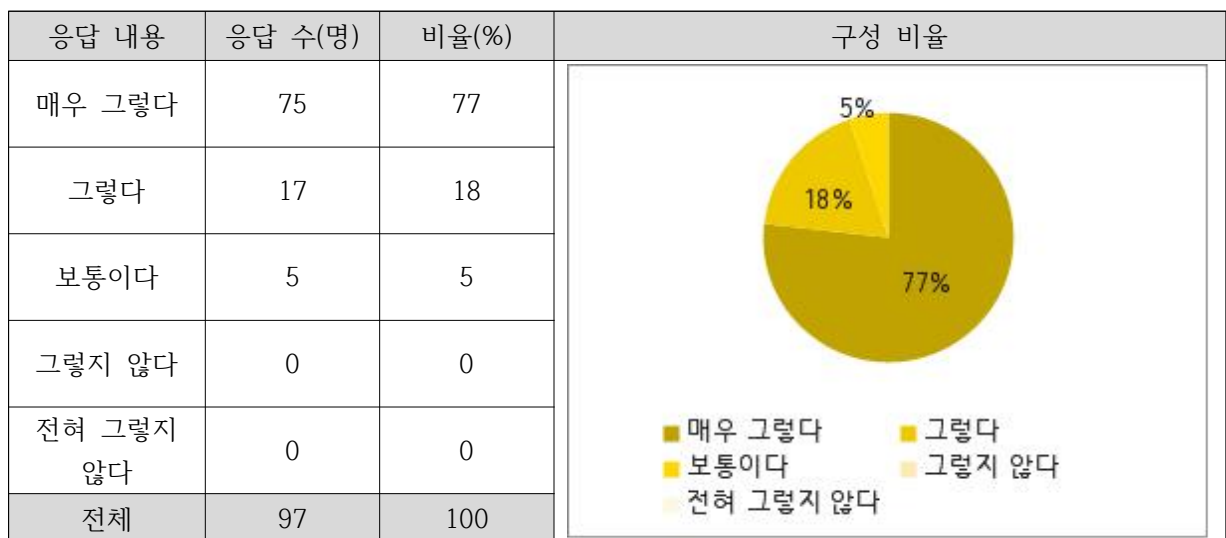
② (온라인 교육) 차시별 강의 시간의 적절성



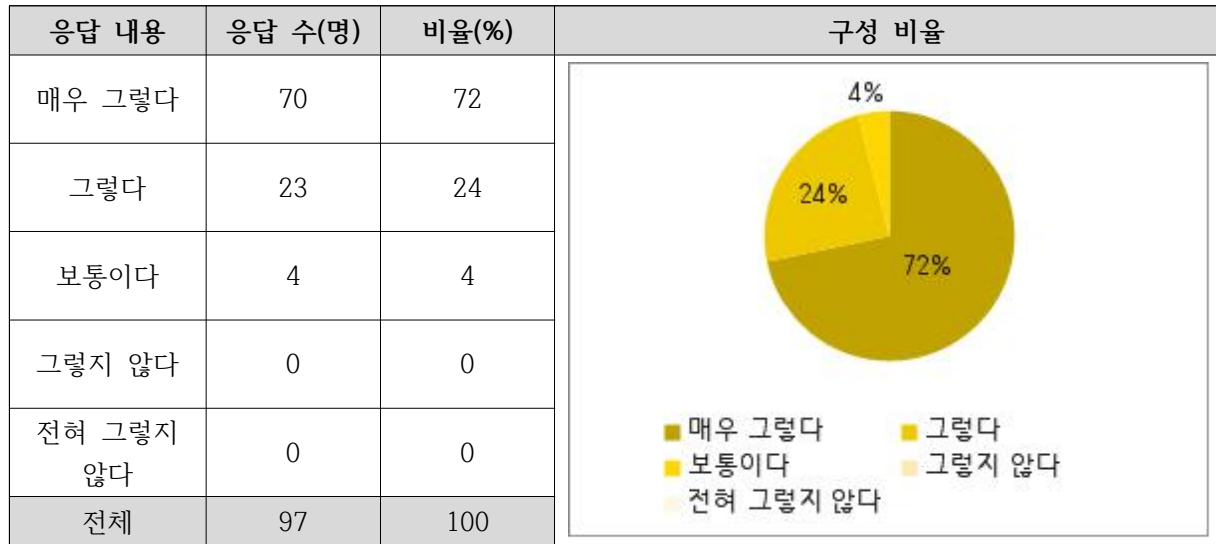
온라인 교육에 관해 웨일클래스 플랫폼을 통한 수강의 편리성에 대해 ‘매우 그렇다’에 응답한 비율이 73%, ‘그렇다’가 20%로 90% 이상의 참여자가 플랫폼을 통한 강의 수강에 만족한 것으로 나타났다. 또한 25분 내외로 구성된 12차시 온라인 강의 시간의 적절성에 대해서도 ‘매우 그렇다’에 응답한 비율이 61%, ‘그렇다’가 34%로 온라인 강의 시간 또한 적절했음을 알 수 있었다.

③ 온라인 교육 내용이 쓰기 채점 전문성 향상에 도움이 되는지 여부

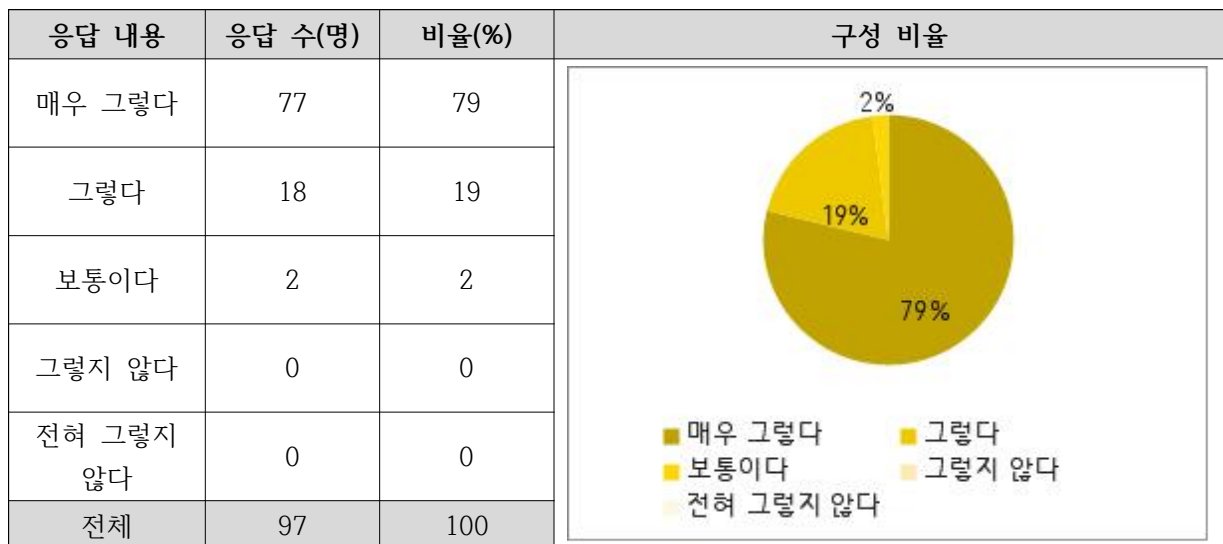
③-1 쓰기 평가 이론: 쓰기 능력과 쓰기 평가의 이해, 논증적 글쓰기의 평가 지표



③-2 쓰기 채점 이론: 쓰기 평가를 위한 채점 절차, 채점 신뢰도 및 영향 요인

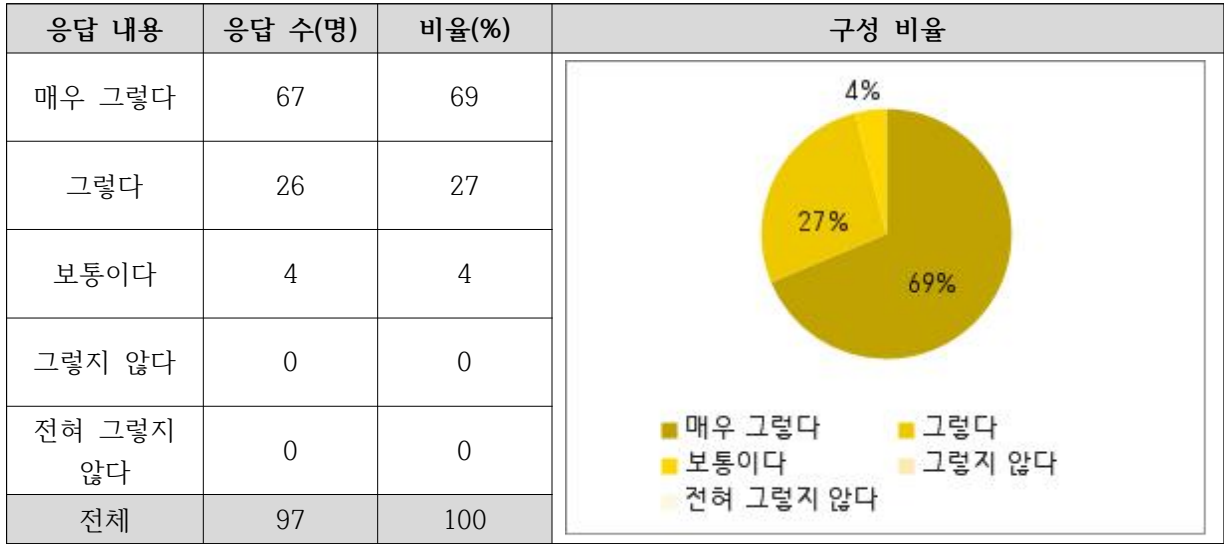


③-3 쓰기 평가 준거별 채점 방법: 내용-조직-표현 범주 채점 방법, 채점 도구의 활용 방법

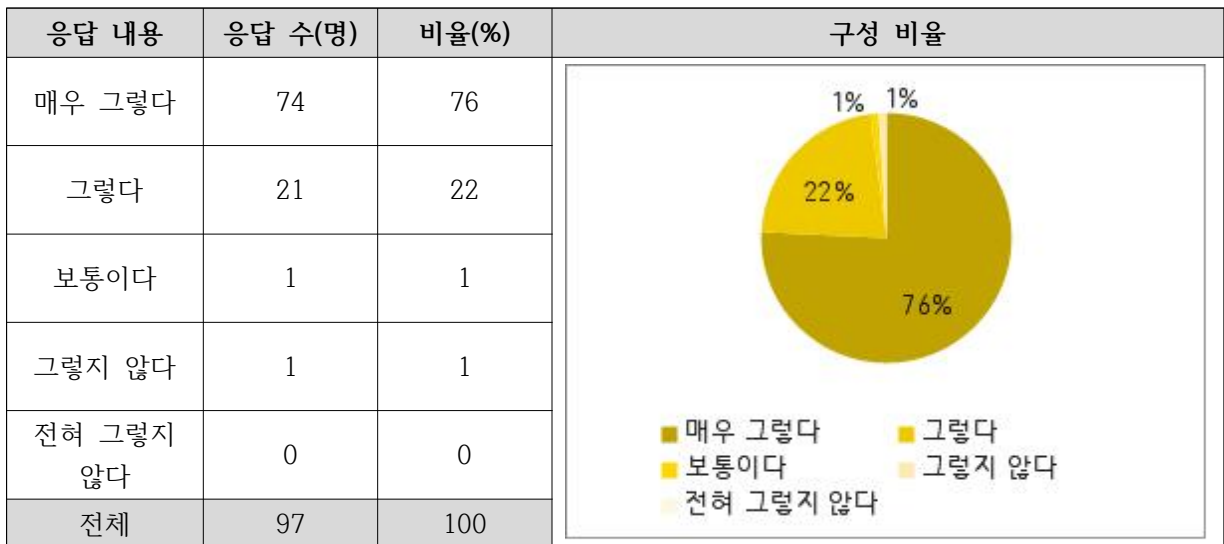


온라인 교육 내용이 채점자의 쓰기 채점 전문성 향상에 도움이 되었는지에 대해서는 세 개 강의 모두에서 ‘매우 그렇다’에 응답한 비율이 72% 이상으로 나타났다. ‘매우 그렇다’와 ‘그렇다’에 응답한 비율을 합하면 모두 95% 이상으로 온라인 교육이 쓰기 채점 전문성 신장에 매우 긍정적인 역할을 했음을 알 수 있다.

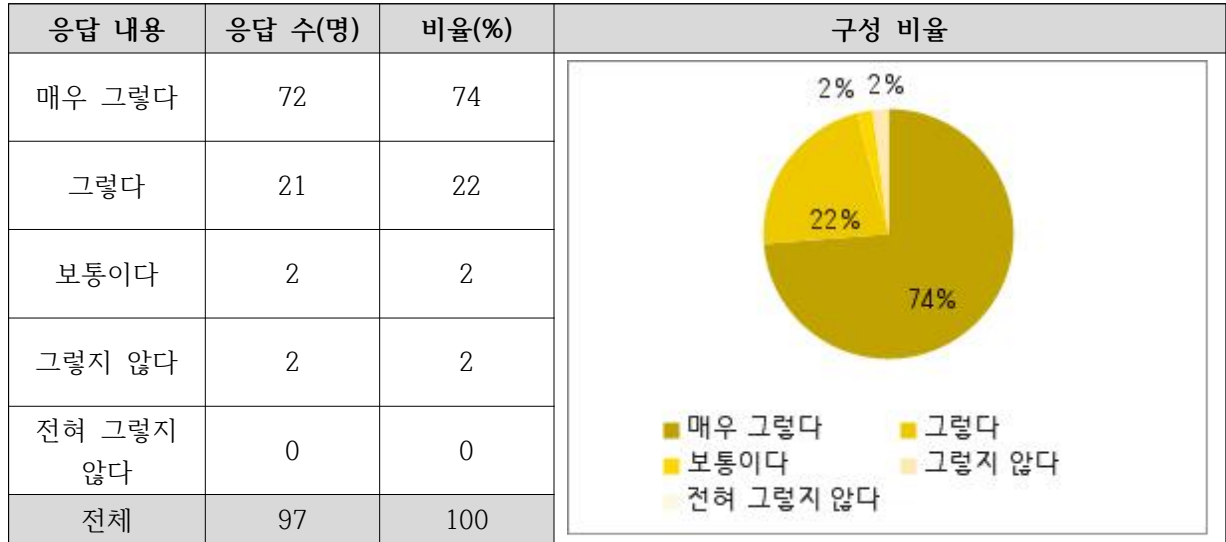
④ 온라인 교육 내용 수준의 적절성



⑤ (온라인) 교수진 설명의 적절성

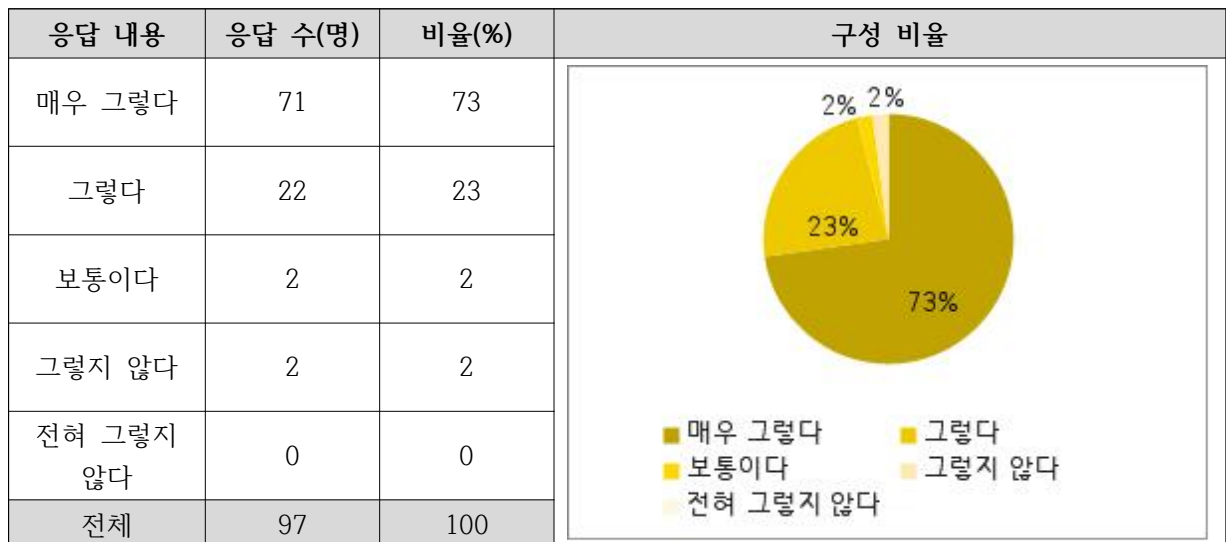


⑥ (온라인) 강의 자료의 적절성



온라인 교육 내용의 수준과 교수진의 설명, 강의 자료의 적절성에 관하여서는 ‘매우 그렇다’에 응답한 비율이 약 70% 이상, ‘매우 그렇다’와 ‘그렇다’에 응답한 비율의 합이 모두 96% 이상으로 교육 자료 구성과 교육 내용 전달이 잘 이루어졌다는 것을 알 수 있다.

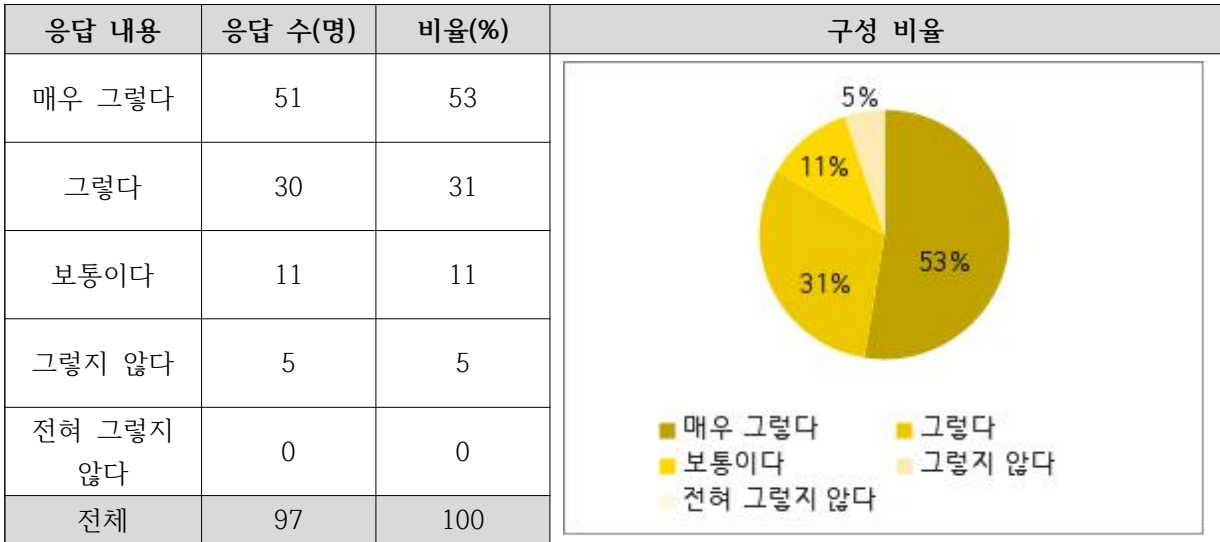
⑦ (오프라인 교육) 워크숍 형태의 교육 방식의 적절성



⑦-1 교육 방식이 적절하지 않다고 생각한 이유

- 신뢰도, 엄격성 등에 대한 개인별 정보를 본인에게는 알려주면 좋겠다.
- 하루에 몰아 강의를 하다보니 피로도가 높아집니다

⑧ (오프라인 교육) 교육 시간(8차시)의 적절성



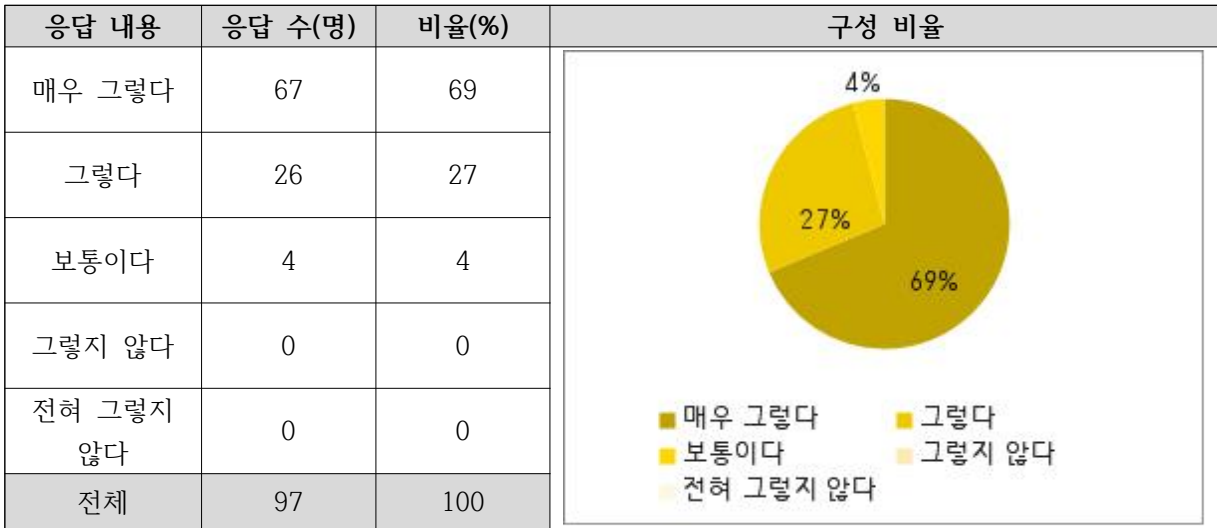
⑧-1 교육 시간이 적절하지 않다고 생각한 이유

- 비가 와서 이동도 일찍부터 했는데 워크숍 시간이 너무 길어 오후 채점은 집중하기 힘들 지경이었다.
- 이론적인 부분에 시간을 할애하기보다, 실전으로 바로 들어가서 진행 시간을 단축하는 것이 효율적일 것 같다.
- 하루에 몰아 강의를 하다 보니 집중이 어렵습니다
- 조금 시간을 줄이면 좋을 것 같습니다.
- 전체적인 교육 시간이 길어서 집중력이 떨어짐

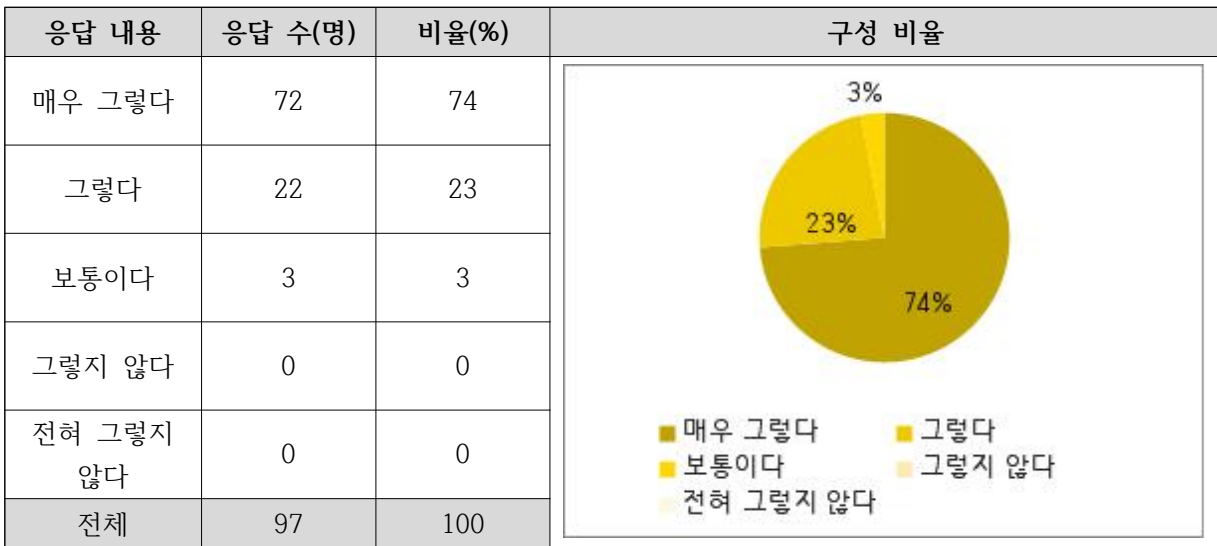
오프라인 교육의 경우 워크숍 형태의 집합 교육에 대해 96%가 적절하다고 응답하였으며(매우 그렇다 73%, 그렇다 23%), 8차시로 이루어진 교육 시간에 대해서는 84%가 적절하다고 응답하였다(매우 그렇다 53%, 그렇다 31%). ‘보통이다’에 응답한 비율은 11%, ‘그렇지 않다’에 응답한 비율은 5%로 나타났는데 그 이유로 ‘교육 시간이 길다’는 의견이 있었다.

⑨ 오프라인 교육 내용이 쓰기 채점 전문성 향상에 도움이 되었는지 여부

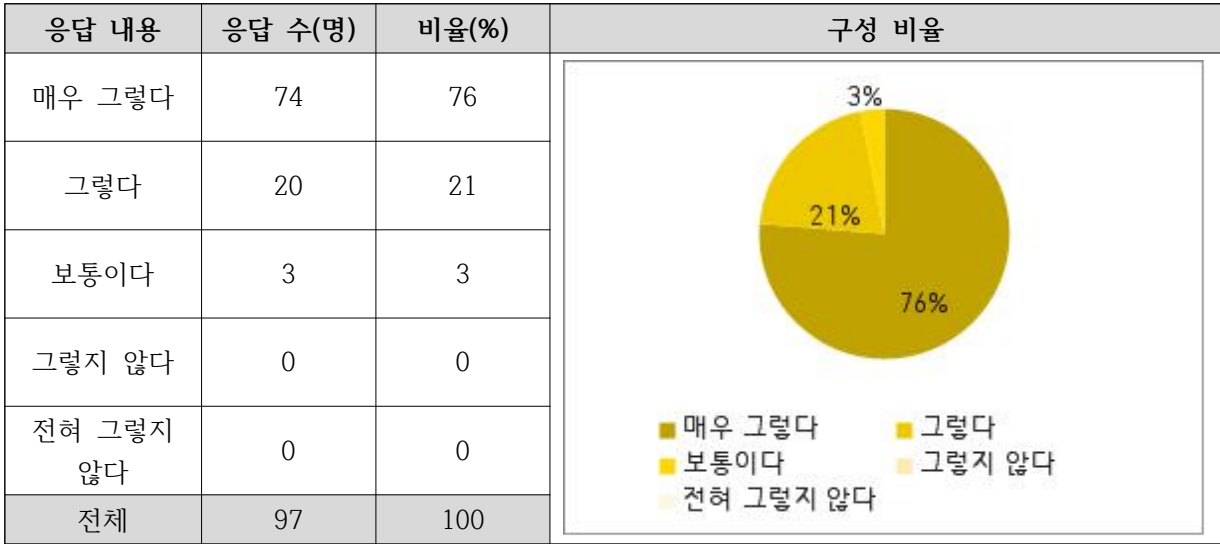
⑨-1 채점 도구 이해 실습: 쓰기 평가 준거의 일반적 이해, 채점 도구의 실제 적용



⑨-2 채점 신뢰도 이해 실습: 예시 답안 채점, 채점 신뢰도 분석 및 점검

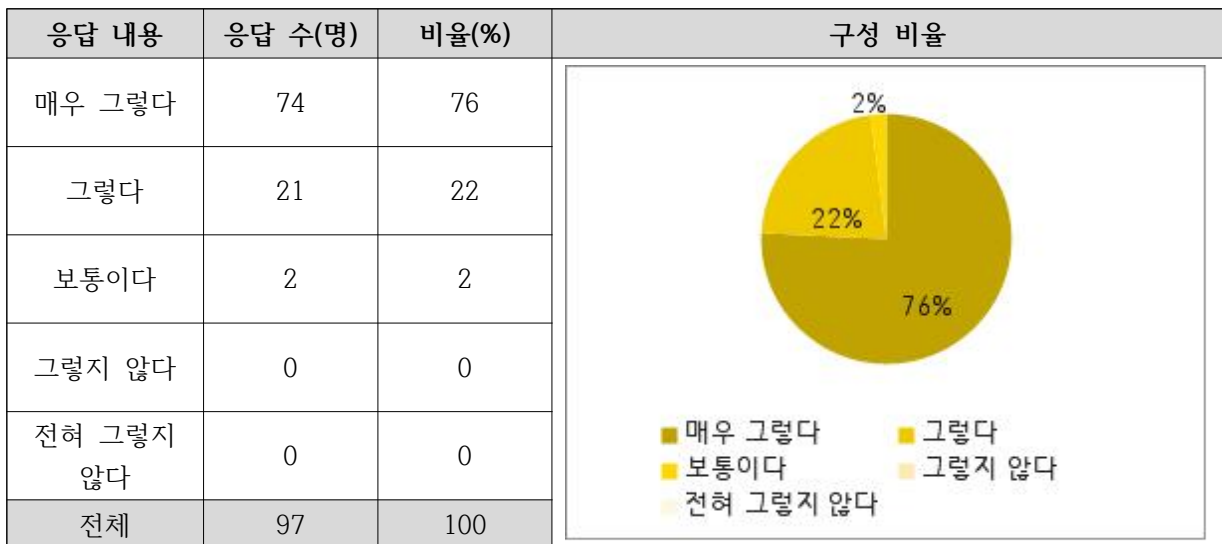


⑨-3 채점 실습 교육의 환류: 채점 실습에 관한 종합 토론, 채점자 자기 평가 및 성찰

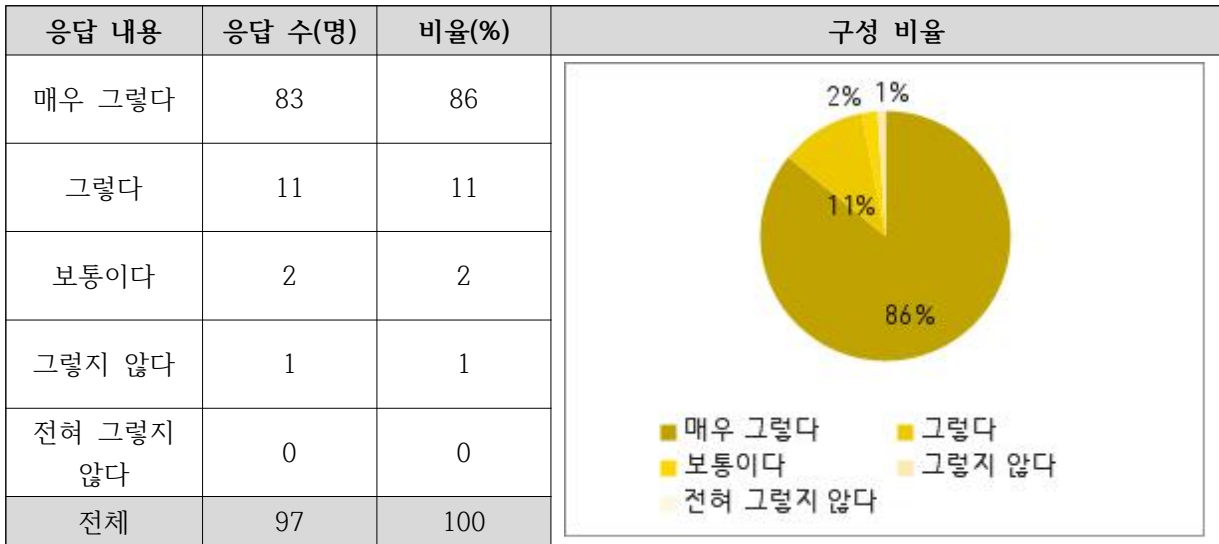


오프라인 교육 내용이 쓰기 채점 전문성 향상에 도움이 되었는지에 대해서는 ‘매우 그렇다’에 응답한 비율이 약 70% 이상으로 나타났으며, ‘매우 그렇다’와 ‘그렇다’에 응답한 비율을 합하면 모두 96% 이상으로 오프라인 교육 또한 쓰기 채점 전문성 신장에 매우 긍정적인 역할을 했음을 알 수 있다.

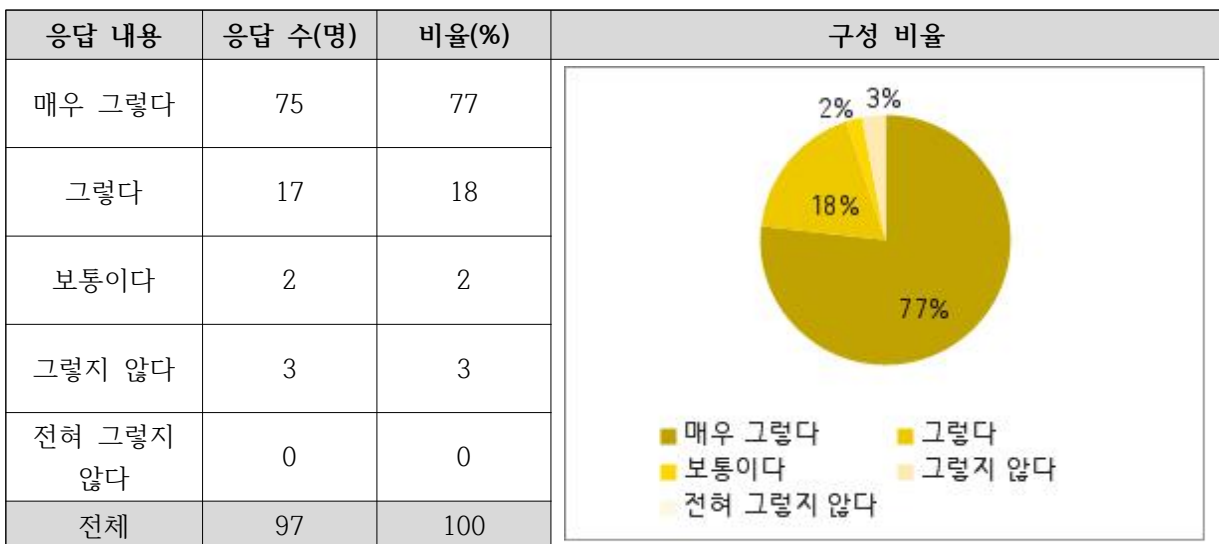
⑩ 오프라인 교육 내용 수준의 적절성



⑪ (오프라인) 교수진의 설명이 교육 내용을 이해하기 쉽게 이루어졌는지 여부

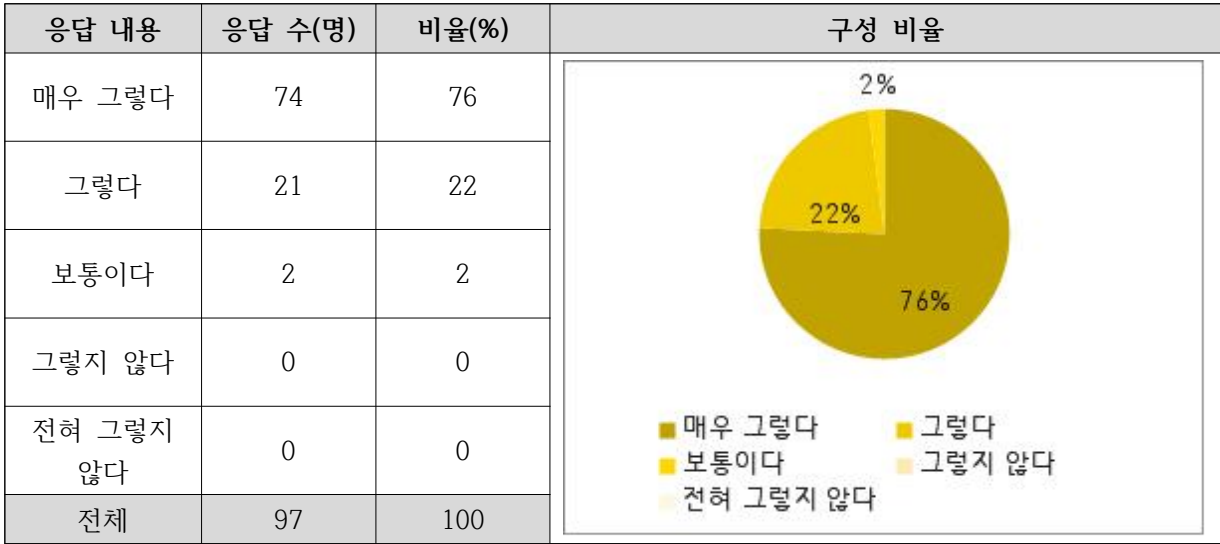


⑫ (오프라인) 강의 자료가 교육 내용을 이해하는 데 도움이 되었는지 여부

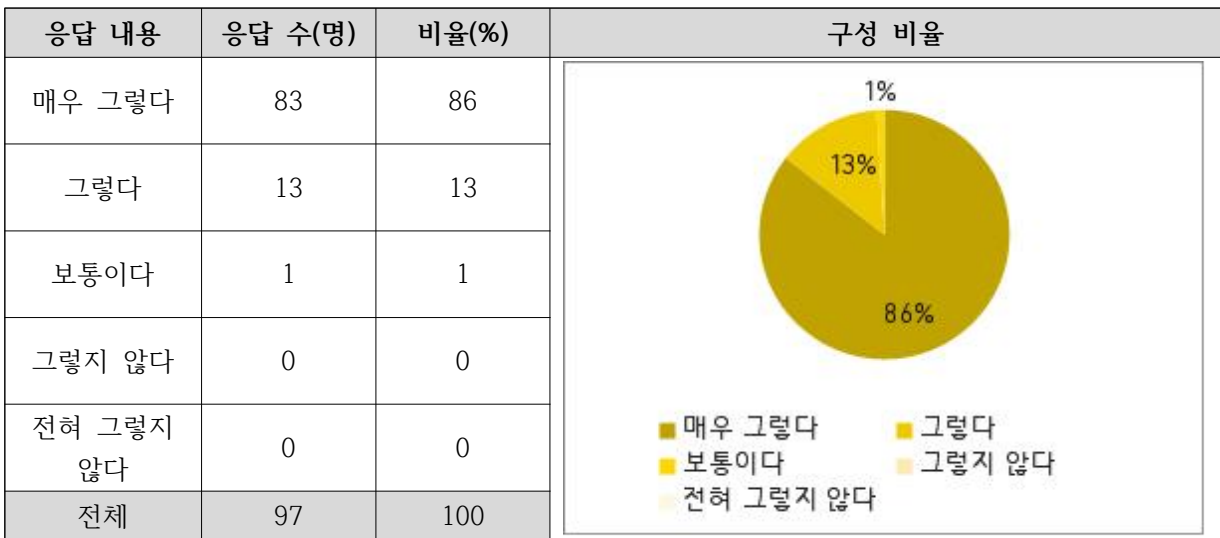


오프라인 교육 내용의 수준과 교수진의 설명, 강의 자료의 적절성에 관하여서는 '매우 그렇다'에 응답한 비율이 76% 이상, '매우 그렇다'와 '그렇다'에 응답한 비율의 합이 모두 95% 이상으로 교육 자료 구성과 교육 내용 전달이 잘 이루어졌다는 것을 알 수 있다.

⑬ 본 프로그램에 대한 전반적인 만족도



⑭ 본 프로그램의 참여도에 관한 자기 인식



프로그램에 대한 전반적인 만족도는 ‘매우 그렇다’에 응답한 비율이 76%, ‘그렇다’에 응답한 비율이 22%로 매우 높은 수준으로 나타났다. 또한 채점 교육 참여자 스스로 본 프로그램에 적극적으로 참여했는지에 대해서는 ‘매우 그렇다’에 응답한 비율이 86%, ‘그렇다’에 응답한 비율이 13%로 거의 모든 참여자가 프로그램에 적극적으로 참여했음을 알 수 있었다.

⑮ 본 프로그램에서 향후 다루어졌으면 하는 사항 및 개선점

- 실제 다양한 채점 예시와 채점 근거에 대한 공유
- 다양한 채점 실습과 충분한 협의
- 채점 준거에 관한 구체적이고 명확한 설명
- 논란이 되는 글(문단)에 대한 평가 의견을 공개적으로 토론할 수 있는 익명 토론방 개설
- 점수 편차가 큰 부분에 관한 집중적인 논의
- 오프라인 교육의 온라인 진행
- 권역별 이동 동선, 냉방 등 오프라인 워크숍 환경 개선

프로그램에 대한 개선점 및 제안에 대해서는 채점 과정에서 필요한 사항(채점 예시, 근거, 준거 관련 설명 등) 제공, 채점에 관한 협의 확대, 온-오프라인 교육 환경 개선 등과 관련한 의견이 제시되었다.

⑯ 쓰기 평가 전문성 신장을 위한 채점자 교육의 필요성

응답 내용	응답 수(명)	비율(%)	구성 비율
매우 필요하다	79	81	
필요하다	17	18	
보통이다	1	1	
필요하지 않다	0	0	
전혀 필요하지 않다	0	0	
전체	97	100	

쓰기 평가 전문성 신장을 위한 채점자 교육의 필요성에 대해서는 99%가 필요하다고 응답하여(‘매우 필요하다’ 81%, ‘필요하다’ 18%) 채점자 교육에 대한 필요성 인식 수준이 매우 높은 것으로 확인되었다. 특히 ‘매우 필요하다’에 응답한 비율이 81%로 사전 설문에서 71%가 ‘매우 필요하다’에 응답했던 것과 대비하면 그 필요성에 대한 인식이 더 높아졌다는 것을 확인할 수 있었다.

이상의 설문 분석 결과를 요약하자면 온-오프라인 교육을 통틀어 프로그램 전반에 관한 만족도가 90%를 상회하는 것을 알 수 있었으며, 채점자 교육의 필요성에 관한 인식

수준이 사전 설문 대비 더 높아진 것을 확인할 수 있었다. 다만 오프라인 교육 시간의 적절성에 대해서는 다른 문항에 비해 만족도가 다소 낮게 나타났는데, 교육 시간이 적절하지 않은 이유에 대해 교육 시간이 길어서 피로도가 높아진다는 의견이 있었다. 이와 관련하여 프로그램의 개선점에 관한 개방형 문항에서는 채점 활동과 소그룹 활동 시간이 더 길었으면 좋겠다는 다수의 의견이 있어 이를 반영한 오프라인 교육 일정 조정 및 개선의 필요성을 확인하였다.

3) 채점자 교육과정 점검 및 개선안 도출

이상의 프로그램 설계, 개발, 운영, 평가에 이르는 과정을 통해 프로그램의 설계 및 운영안을 점검할 수 있었으며 이를 바탕으로 다음과 같은 개선안을 도출하였다.

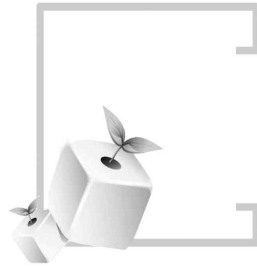
첫째, 온라인 교육과 오프라인 교육 내용의 연계를 긴밀히 하면서 두 교육과정의 목적과 특성이 뚜렷하게 드러날 수 있도록 설계안을 구축할 필요가 있다. 본 프로그램 설계 당시 연구진과 교수진이 협업하여 교육과정을 설계하였으나 연계성을 강화하기 위해서는 설계안과 강의 요목에 관한 보다 적극적인 교차 검토가 필요하겠다. 또한 교육 내용의 중복(특히 평가 도구 및 준거와 관련하여)을 지양하면서 온라인 교육에서는 쓰기 평가 이론과 채점 기준표 및 척도에 관한 충분한 이해에 초점을 두고, 오프라인 교육에서는 채점 실습과 교육 참여자 간 협의에 초점을 두어야 한다. 이와 관련된 요구는 아래와 같이 사후 설문 결과에서도 확인할 수 있었다.

- 실제 채점 실습을 진행하기 전에 1편의 글을 함께 분석할 수 있는 시간이 있으면 좋겠다. 해당 내용이 온라인에 일부 있기는 했지만, 현장에서 글을 함께 분석해 보고 어떻게 평가했는지를 공동으로 점검해 볼 수 있는 시간이 추가적으로 진행된다면 더 알찬 시간이 될 것 같다.
- 팀별 토의 시간이 짧았고, 선생님들에게 발언 기회가 적절하게 돌아가지 않는 것이 아쉬웠음. 개별 의견을 취합하는 시간도 있으면 좋을 것 같음.
- 오프라인 채점 시간을 좀 더 배정해 주시면 좋겠습니다.

둘째, 오프라인 교육 시간의 재분배 및 조정이 필요하다. 사후 설문 결과 오프라인 교육 시간이 길어 피로도가 높다는 의견과 함께 위에서 제시한 바와 같이 채점 협의 시간 확대에 대한 요구를 확인할 수 있었다. 이를 반영하여 오프라인 교육 일정에 대한 피로도는 줄이면서, 채점자 간 협의 시간은 충분히 확보할 수 있도록 교육 시간을 잘 안배할 필요가 있다. 특히 교육 참여자의 채점 문식성에 관하여서는 온라인 교육을 통해 충분히

갖추었다는 전제하에 오프라인 교육에서는 평가 문항 및 진단 지표에 관한 설명은 최대한 간소화하고, 교육 참여자가 채점 사회화를 경험할 수 있을 만큼 충분한 채점 협의가 진행될 수 있도록 오프라인 교육의 전체적인 교육 일정을 조정하는 것이 필요하겠다.

셋째, 교육과정 전반에서 교수·학습의 효과와 학습 참여도를 높이기 위한 노력이 요구된다. 오프라인 교육 전에 이루어지는 온라인 교육에 성실히 임한 채점자는 채점 문식성을 갖추게 되어 오프라인 교육에 대한 동기와 오프라인 교육에서 진행되는 소그룹 활동에서의 참여도가 높을 수밖에 없다. 이에 온라인 교육 참여도와 효과를 극대화하기 위해 차시별 수강 기간을 별도로 정하거나(단기간에 몰아서 수강하는 것 방지), 강의 참여도를 P/F 형태가 아닌 평가제로 변경하여 차시별 퀴즈 풀이에 배점을 부여하는 등 보다 적극적인 수강 장치를 마련할 필요가 있다. 또한 온라인 교육 종료 후에 부여한 사전 채점 과제가 오프라인 교육 내용으로서 제대로 기능하기 위해서는 교육 참여자가 본인이 제출한 결과를 별도로 소지하고, 기억에 남는 글쓰기 자료와 척도 적용 경험 등 간단한 채점 특이 사항이나 채점평을 작성하도록 하여 오프라인 교육에서 참여자 간 채점 경험 공유가 효과적으로 이루어질 수 있도록 해야 할 것이다.



제 V 장

글쓰기 자료
채점 실행



1. 채점 모형 설계

본 과업에서는 수집된 글쓰기 자료에 대한 체계적이고 일관된 채점을 진행하기 위하여 먼저 채점 결과 수집에 관한 모형을 수립하고, 이를 바탕으로 채점 작업이 이루어질 수 있도록 기초를 확립하였다.

1) 채점 모형 설계의 목적과 구성

(1) 목적

채점 모형은 수집한 글쓰기 자료에 대한 타당한 평가 결과를 수집하기 위해 구안되었다. 이를 위하여 본 연구에서는 연구 계획에 따라 수집된 글쓰기 자료 중 5,000건에 대해 타당하고 신뢰로운 채점을 실시할 수 있는 형식과 절차를 마련하였다. 채점 모형을 통해 체계적으로 채점 자료를 수집함으로써 채점 과정에서 발생할 수 있는 오차 요인을 최소화할 수 있다. 이를 구체적으로 제시하면 첫째, 설계된 채점 모형을 통해 글쓰기 자료가 특정 채점자나 채점자 집단에게 불균일하게 할당됨으로 인하여 발생하는 채점 결과의 편향을 방지할 수 있다. 둘째, 채점 정확성이 높은 채점자들이 바람직한 채점 특성을 유지할 수 있도록 관리할 수 있다. 셋째, 수집된 글쓰기 자료 판정의 신뢰도를 높은 수준으로 확보할 수 있다.

(2) 구성

본 연구의 목적에 따라 채점 모형은 채점 자료 수집 모형과 채점 자료 선정 모형으로 구성하였다. 모형 구성을 위하여 교육 평가 및 글쓰기 평가, 인공지능 기반의 자동 채점 연구와 관련된 선행연구를 검토하였으며, 본 과제에서 수집되는 자료의 특성을 고려하여 실행 가능한 모형을 수립하였다.

채점 모형은 채점자들이 글쓰기 자료에 대해 신뢰할 수 있는 채점을 할 수 있도록 지원하기 위하여 수립된다. 관련하여 핵심 개념인 신뢰도의 개념을 구상하기 위하여 기존

의 심리 측정 분야의 신뢰도 관련 문헌을 살펴보았다. 기존의 교육 평가 연구나 글쓰기 평가 연구에서는 채점자 간 신뢰도 또는 채점자 내 신뢰도가 중시되었으며, 이는 평균 중심의 결과적 일관성에 초점을 두는 것이 특징이며, 이를 평가의 정확성을 가리키는 개념으로 보아 왔다(Haertel, 2004). 본 과업에서는 이를 고도화하기 위하여 재현가능성(reproducibility)과 모방(replicability)으로서의 채점을 동시에 지향할 필요가 있다고 판단하였다(Meng, 2020). 이는 채점자가 자신에게 맡겨진 채점을 잘 해낸다는 의미로, 스스로 채점 환경을 조절 및 통제하며, 다른 채점자들과 같은 방식으로 채점 척도를 사용하여 채점할 수 있도록 교육 및 훈련을 유도하고, 결과적으로 본 연구의 채점이 지향하는 채점자의 태도를 갖추는 것과 관련이 있다. 신뢰도의 확보는 조사의 규모와 수집되는 자료의 성격, 양, 그리고 다양성의 정도, 채점 기간과 환경, 채점 척도, 교육과 훈련 등과 관련이 되며, 이러한 채점에 영향을 미치는 다양한 원인이 있음을 채점자들이 인지하고, 이를 관리하며 채점에 임할 수 있도록 채점 과정의 지원이 있어야 한다.

다음으로 본 연구에서 제공하는 채점 모형은 채점 자료를 수집만 하는 것에서 나아가 최적화된 채점 자료를 선정하는 일을 수행한다. 이는 본 과업에서 복수 채점을 실시한 것과 관련이 있다. 일반적으로 교육 평가 상황에서 결과의 공정성을 확보하기 위하여 실시하는 것과 달리, 본 과업에서는 해당 글에 대한 최적의 결과를 산정하기 위하여 복수 채점의 결과에 대한 통계 분석을 실시하고, 복수의 채점 자료 중 최적화된 값을 선정하고자 한다. 이는 기존의 글쓰기 자동 평가 연구들에서 채점 결과를 수집할 때 단일 채점자의 결과만을 사용하여 편향성의 영향 문제를 고려하지 않았고, 원점수를 다른 등급이나 척도로 변환하는 경우 사람이 실제 채점한 결과로 볼 수 없어 실용화에 한계가 있기 때문이다(Zhang, 2013), 따라서 본 과업에서는 자동 채점의 결과로 사람 채점자와 같은 채점 척도상의 값을 산출하는 기술 개발이 가능하도록 각 글쓰기 자료에 대한 최적화된 채점 값을 선정할 수 있는 모형을 설계하고자 한다.

2) 채점 자료 수집 모형

채점 자료 수집은 글쓰기 자료에 대한 전문 채점자의 채점 수행 정확성을 높이기 위한 기반을 제공하는 것으로 구성된다. 이 과정에서 고려되는 문항과 작성된 글쓰기 자료의 다양한 특성을 전문 채점자가 일관되고 정확하게 판단할 수 있도록 하는 체계적인 실천

모형을 구축하고자 하였다.

(1) 채점 대상 글쓰기 자료 목록 확정

채점 자료 수집의 첫 단계는 채점을 실시할 글쓰기 자료 목록의 확정이다. 채점 대상 글쓰기 자료는 채점자가 채점을 할 만한 상태라고 판단한 글로서, 요구한 분량 기준을 충족하며, 문항에 부합하는 내용으로 작성된 것을 가리킨다.

본 사업에서는 채점 대상 글쓰기 자료의 확정을 위해 수집된 글을 컴퓨터 소프트웨어로 형식을 맞추도록 자동 정제한 후, 오류 여부 및 타당성을 검증하는 수동 정제를 거쳐 최종 채점 대상 글 목록을 확정하였다.

(2) 채점 방법 결정

다음은 확정된 채점 대상 글쓰기 자료 목록을 기준으로 실제 채점이 필요한 양과 채점자의 수, 요구되는 신뢰도 확보의 수준 등을 고려하여 채점 방법 및 채점 설계(scoring protocol design)를 마련하는 것이다. 글쓰기 자료에 대한 채점 방법은 채점의 목표와 채점의 여건 등을 토대로 결정할 수 있으며, 대표적으로는 글의 전반적인 인상을 중심으로 접근하는 총체적 채점과 글의 준거별 특성을 나누어 접근하는 분석적 채점, 그리고 이 둘을 병행하는 통합적 채점이 있다. 본 과업에서는 채점의 목표가 글의 수준 구별이라는 종합적인 목표를 달성하는 것과 관련하여 하위 준거별 점수를 신뢰할 수 있게 결정될 수 있도록 표준화된 형식의 채점 척도를 적용하고, 향상된 수준의 교육 및 훈련을 제공해야 한다. 이와 같은 체계적 접근은 향후 인공지능 기반의 채점 모형 개발에도 유의미한 증거들을 제공할 수 있을 것으로 기대된다. 이러한 특성을 고려하여 본 과업에서는 분석적 채점 방법을 적용하여 전체 채점을 진행하였다.

(3) 채점 설계 결정

채점의 설계는 해당 회차별 채점 자료의 수와 채점자의 인원수, 채점 기간, 채점의 환경 등을 고려하여 결정된다. 수행 평가에서 글에 대한 직접식 평가는 해당 글의 수준을 몇 가지로 나누어 부여하는 다분 반응형 채점(polytomous scoring) 결과를 기반으로 하

기 때문에 객관적이고 신뢰할 수 있는 채점 결과를 수집하는 것이 매우 중요하다. 다분 반응형 채점을 효율적으로 진행하기 위해서는 채점자 교육 및 훈련의 효과를 지속시키고, 채점 기간에 자신의 채점 특성을 성찰할 수 있는 피드백을 제공하는 방법, 그리고 주관적인 채점 결과를 보완하기 위하여 단일 자료에 대해 복수의 채점자가 채점을 실시하는 방법을 동시에 고려할 필요가 있다.

본 과제에서는 최종 수집 목표 글쓰기 자료 수인 9,000편 가운데 5,000편을 채점하는 것이 목표이며, 이는 6개 문항별로 약 700-1,000여 편의 글을 채점하는 것이다. 글쓰기 자료에 대한 채점이 정확하고 일관되게 이루어지기 위해서는 채점자가 채점 기준을 고려하여 글을 읽는 전문성을 가질 수 있도록 교육 및 훈련과 관리를 통한 접근이 필요하다. 특히 이번 채점 자료 수집에 참여한 채점자들은 전국의 9개 협약 대학 및 기타 대학에서 글쓰기 교육을 담당하고 있는 교강사이며, 전공 배경과 연령대가 다양하다는 특징이 있었다. 이러한 다양한 배경의 채점자들이 높은 신뢰도 수준을 유지하면서 정확하게 채점하기 위해서는 채점자들이 갖고 있는 글과 글 채점에 대한 인식을 제고 및 유지할 수 있도록 체계적인 교육과 집중적인 훈련, 그리고 채점 상황에 대한 관리가 이루어져야 한다. 또한 채점자는 자신이 수행해야 하는 채점 행위의 범위와 목표, 방법을 숙지하고 있어야 하며, 채점 수행에 필요한 사항을 관리자와 소통하고, 확인할 수 있는 체계를 제공받을 수 있어야 한다.

(4) 채점 자료 수집 모형의 운용

채점 자료 수집 모형을 목적에 맞게 운용하기 위해서는 수집되는 정보의 성격과 이를 산출하는 채점 행위의 특성, 그리고 결과물이 도출되는 환경과 절차를 고려한 관리가 이루어져야 한다.

가. 글쓰기 자료와 채점 척도의 활용 지원

본 연구에서는 연구 참여자가 작성한 논증형 글쓰기 자료를 수집하였다. 채점자는 채점 척도에 의거하여 해당 글쓰기 자료에 대해 9개 준거별로 1-5점을 부여해야 한다. 이와 같은 분석적 채점 방법을 적용하기 위해 채점자들에게 제공한 채점 기준표는 각 준거와 수준별 기술어와 강도나 빈도의 차이를 드러내는 구체적인 기술문의 형태로 작성되었

다. 기술문 형태의 채점 기준표를 활용할 때 잘못된 이해나 혼란을 줄이기 위해서는 채점자들에게 이를 숙지할 수 있는 시간이 요구되며, 본 연구에서는 채점자 온라인 및 오프라인 교육, 온라인 워크숍, 채점 피드백 등을 통해 채점 기준표에 대한 교육과 반복적인 사용 연습을 통해 이에 대한 채점자들의 인식을 향상시키기 위해 노력하였다. 또한 채점자들이 숫자로만 표시되는 점수 값을 채점 척도를 구성하는 각 준거와 점수대별로 구분하여 인식할 수 있도록 오프라인 교육과 온라인 워크숍에서 소집단 토의와 전체적인 의견 공유의 기회를 제공하였다. 또한 본 채점 중에는 사회관계망 서비스(카카오톡 오픈 채팅)를 활용하여 다양한 개별 글쓰기 자료의 사례나 채점하는 과정에서 발생할 수 있는 채점 기준표 적용의 어려움이나 문제를 실시간으로 수렴하고, 이에 대한 유의 사항을 마련하여 채점자들이 겪을 점수 결정 과정의 혼란을 최소화하기 위한 관리 소통 체계를 구축하였다.

나. 채점 수행에 대한 지원

다음으로 채점 모형을 통해 수집하는 채점 자료는 채점자들이 반복하여 할당된 임의의 글쓰기 자료를 읽고, 이를 적용하기 위해 채점 기준표를 확인 및 적용하여 점수를 결정하는 행위를 통해 구축된다. 채점 과정에서 시각적으로 입력되는 정보들은 읽는 순간에는 높은 확률로 기억에 남을 수 있지만, 글의 길이가 길어지고, 구조나 내용이 복잡해지면 기억을 방해할 가능성이 커진다. 뿐만 아니라 선형적으로 갖고 있는 채점 행위에 대한 경험이 채점자가 실제로 수행해야 하는 채점을 위한 읽기 행위를 방해할 수도 있다. 따라서 채점자가 집중적으로 글쓰기 자료를 읽고, 채점 기준표를 일관되게 읽고 적용할 수 있는 채점의 환경을 구축하고, 안정적으로 유지할 수 있도록 돕는 플랫폼이 채점자들에게 제공될 필요가 있다.

본 연구에서는 채점자마다 서로 다른 상황과 환경에서 채점 행위가 진행되는 상황 가운데 균일한 조건을 유지할 수 있도록 채점자 교육 및 훈련에서 이에 대한 관리가 신뢰도 확보에 중요하다는 점을 강조하고, 효과적으로 글쓰기 자료와 채점 기준표를 확인할 수 있는 채점 플랫폼을 구축하였다. 채점 플랫폼에서는 자신이 읽어야 하는 글쓰기 자료의 글자 크기를 조절할 수 있으며, 점수를 입력해야 하는 준거나 점수를 직관적으로 확인할 수 있는 기술문의 팝업 기능 등을 제공하였으며, 또한 화면이 아니라 출력물로 글을 읽기 원하는 채점자들의 요구를 고려하여 인쇄하여 볼 수 있는 기능을 채점 플랫폼에

넣는 등 채점 행위를 돕기 위한 시각적인 차원의 지원을 제공하였다.

다. 채점 자료 수집에 대한 지원

채점 자료의 수집은 채점 설계에 따라 회차별(또는 과정별)로 순차적으로 진행이 이루어지며 또한 동시에 여러 회차가 동시에 진행되기에 각 회차별로 진행 단계에 따라 맞춤형으로 필요한 사항을 소통할 수 있는 관리 수단을 갖출 필요가 있다. 특히 채점자의 수가 많고, 채점자 간 거리가 먼 상황에서 일정 기간 이상 채점 자료 수집이 이루어져야 하는 경우에는 소통이 필요한 정보의 유형에 따라 채점자가 관리자와 직접 연락하고, 관리자도 채점자의 채점 수행을 점검하고 보조할 수 있는 장치들을 사전에 마련해 두어야 한다.

본 과업에서는 각 회차별로 채점 진행 상황을 모니터링하기 위해 채점 플랫폼의 관리자 화면에서 실시간으로 각 채점자의 채점 진행 편수와 진행률에 관한 정보를 확인할 수 있는 기능을 추가하고, 동시에 채점자가 플랫폼에서 채점을 진행하는 중에 발생하는 문제를 플랫폼상에서 바로 관리자에게 문의할 수 있도록 하는 메시지 송수신 기능을 추가하였다. 아울러 회차별로 채점자들이 특정 글에 대한 채점 행위의 일관성을 유지 및 향상시키기 위하여 채점 특성에 대한 피드백을 제공하고, 추가 설명이 필요한 채점자에게는 이에 관한 개별 전화 면담을 실시하여 채점자가 채점 기간 중에 채점 수행에 집중하여 참여할 수 있는 환경을 구축하였다.

3) 채점 자료 선정 모형

본 연구에서 수집되는 채점 자료는 개발 중인 논증적 글 자료에 대한 자동 채점 기술 개발과 관련하여 사용될 수 있다. 기존의 글쓰기에 대한 자동 채점이 갖고 있는 채점 자료 선정 과정의 한계를 극복하기 위하여 본 과제에서는 사람 채점(human scoring)의 결과를 고도화하고, 이를 체계적으로 선정하는 방법을 적용하였다.

(1) 원점수 기반 접근

채점 자료 선정 모형은 수집된 채점 자료를 기반으로 이를 활용할 때 원점수 기반으로 활용할 것인지 아니면 척도점수로 일원화하여 사용할 것인지에 따라 구분할 수 있다. 원점수(raw score) 또는 관찰 점수 기반의 접근의 경우, 채점자들이 부여한 점수를 그대로 사용하는데, 이때 채점을 단독으로 수행하였는지의 여부에 따라서 선정의 절차에 차이가 있다. 먼저 하나의 글에 대해 한 명의 채점자가 단독으로 점수를 부여하는 경우, 해당 점수는 해당 글에 대해 부여된 유일한 전문가 판정의 결과로서 그 타당성 여부에 대한 추가적인 판단 없이 그대로 최종 채점 자료로 선정이 된다. 이와 달리 두 명 이상의 채점자가 하나의 글을 채점하는 경우에는 각 채점 준거마다 둘 이상의 점수가 존재하기 때문에 이를 일반적인 교육 평가의 방식처럼 평균을 낼 것인지, 아니면 보다 정확한 채점을 한 채점자의 채점 결과를 선택할 것인지에 따라라도 차이가 있다. 평균을 활용하는 경우에는 해당 글에 대한 주관적인 편차를 어느 정도 상쇄한 결과로서 채점 자료를 구성할 수 있다는 장점이 있으나, 산출된 수치가 실제 사용된 척도와 달리 소수점 이하로 나타나기 때문에 그 의미를 해석하는 데 한계가 있다. 따라서 평균값을 산출하는 경우에는 이를 실제 사용된 척도와 대응시키기 위하여 정수 범주로 재편성하기 위해 소수점 자리에 대한 버림이나 올림을 하는 방법도 고려되어 왔으나, 이러한 접근의 논리적인 설명이 어렵다는 한계가 있다. 이를 보완하기 위해 글쓰기 자료에 부과된 채점자들의 채점 결과를 평균으로 산출한 후에 해당 문항 내에서 이에 대한 회귀 분석을 실시하여 기존 척도와 다른 새로운 척도를 토대로 수준을 구분하는 방안도 논의된 바가 있다. 그러나 이러한 접근은 종합적인 글의 수준을 구별하는 데에는 유용하나, 실제적으로 요구되는 글에 대한 구체적인 채점 결과로서는 정보가 제한적이라는 한계가 있다.

(2) 척도 점수 기반 접근

척도 점수(scaled score)의 경우, 원점수를 일정한 기준에 따라 다른 척도로 변환한 점수를 가리킨다. 본 연구에서는 문항반응이론(Item Response Theory; IRT)을 적용하여 채점자의 채점 수행에 나타난 채점 특성에 대한 객관화된 수치를 확보한다. 이러한 접근은 채점 자료의 유용화에 대한 신뢰성을 보장할 수 있다는 장점이 있으나, 실제 채점에 사용된 척도가 아니라, 새로운 척도로 변형된 것이기 때문에 해당 결과를 원점수와

직접 연계시킬 수 없다는 어려움이 있다. 척도 점수를 활용한 접근에서는 사람 채점자가 부여한 원점수를 새로운 등급이나 수준으로 재편하고, 그 결과를 다시 해석하는 절차를 제안하기도 한다. 그러나 본 과업의 경우에는 사람 채점자의 판정 결과를 기계가 모방하는 기술 개발과 관련이 있으므로, 척도 점수로 산출된 결과를 활용한 후 이를 다시 원점수와 대응시켜야 한다.

이에 본 연구에서는 척도 점수를 기반으로 해당 글쓰기 자료에 대해 산출된 결과를 활용하는 방식이 아니라, 우수한 채점 결과를 선정하는 방안을 마련하였다. 이는 기존의 자동 채점 연구들이 단독 채점자를 배치하거나 새로운 등급이나 점수로 결과를 활용한 것과 달리 복수의 채점자를 활용하면서 동시에 그에 따른 선택의 문제를 극복하고자 한 것이다. 산출된 해당 글에 대한 복수의 채점 결과 중에 가장 정확한 판단을 받은 것을 선정하기 위해서는 채점 결과에 대한 비교 가능성을 확보하는 것이 필수적이다. 본 연구에서는 한 편의 글에 대해 복수의 채점자들로부터 수집한 채점 결과 중 대푯값을 결정하기 위하여 다국면 라쉬 모형(MFRM) 분석 결과를 바탕으로 채점자들의 채점 특성을 비교한다. 채점 특성에는 글 수준 판정의 정확도와 채점 척도 사용의 적합도, 그리고 각 준거별 척도 사용의 오류 여부가 포함되며, 해당 결괏값을 통해 채점자 간 차이를 비교하여 대표 채점값을 결정한다. 이러한 접근을 통해 선정된 대표 채점 결과는 최적화된 채점 결과로서 사람 채점자와 같은 채점 결과를 제동하는 자동 채점 기술 구축의 기초가 된다.

2. 채점 실행 플랫폼 개발 및 운영

1) 채점 실행 플랫폼 개발의 목적과 기능

(1) 채점 실행 플랫폼 개발의 목적

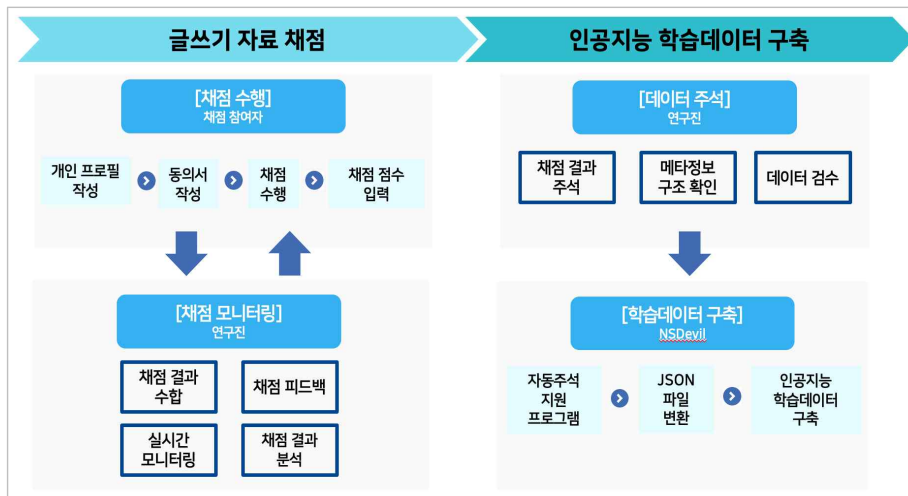
본 연구의 채점 실행 플랫폼 개발 목적은 두 가지로 기술할 수 있다. 첫째, 대규모 채점 자료의 효율적인 수집 및 관리이다. 채점 자료는 개별 글쓰기 자료에 대한 복수 채점자들의 점수 입력값에 대한 대응물로서, 그 글쓰기 자료와 채점자 각각의 인원을 고려할 때 매우 큰 규모의 데이터로 구성된다. 이러한 데이터를 웹에서 접근할 수 있는 플랫폼을 구축해 관리함으로써 사용의 효율성 및 안정성을 확보하고자 하였다. 둘째, 채점 실행 플랫폼은 일정 수준의 평가 환경 통제에 기여한다. 채점자 간 신뢰도를 제고하는 여러 방안 가운데, 채점 환경 및 제반 조건에서 일정 정도의 공통성을 유지하도록 하는 것을 고려할 수 있다. 플랫폼을 통한 채점은 채점자들이 글쓰기 자료에 대해 연구진이 사전에 자동 및 수동 정제한 형태로 채점을 수행할 수 있도록 지원한다.

(2) 채점 실행 플랫폼의 기능

채점 실행 플랫폼은 채점 수행 플랫폼(채점자 페이지)과 채점 관리 플랫폼(관리자 페이지)으로 구분된다. 채점 수행 플랫폼을 통해서 채점자들은 기본 정보, 경력 정보를 입력하고 연구 참여에 대한 설명문 및 동의서를 작성한다. 또한, 채점 점수를 입력하고 저장하여, 플랫폼을 통해 채점 데이터가 구축되도록 한다. 채점 관리 플랫폼을 통해서 연구진은 연구에 참여한 채점자들의 평가 경력 등 정보를 수집하고, 글쓰기 자료를 채점자별로 할당한다. 아울러, 채점이 이루어지는 과정에서 마감 기한에 따른 실시간 진행률을 파악한다. 이를 통해 그룹별로 지정된 기한에 일정 정도의 규칙적인 진행이 이루어지고 있는지를 파악함으로써 마감 기한에 임박하여 과도한 분량을 채점하게 되는 사례가 발생하는 것을 방지한다. 채점자들의 점수 입력이 완료되면 연구진은 관리자를 통하여 채점 결과를 일괄적으로 다운로드받아 채점 데이터를 구축한다.

<표 V-1> 채점 실행 플랫폼의 기능

	채점 수행		채점 수행 관리
주체	채점자		연구진 플랫폼 개발 업체(NSDevil)
기능	채점 전 단계	개인정보 작성 동의서 작성	채점자 그룹 생성 채점자별 글쓰기 자료 할당 채점자 플랫폼 이용 문제 접수 채점 점수 다운로드
	채점 단계	글쓰기 자료 다운로드 채점 마감 시한 확인 채점 점수 입력	
	채점 완료 단계	채점 점수 입력 확인	



[그림 V-1] 채점 실행 플랫폼의 기능 구조도

2) 채점 실행 플랫폼 개발 과정

채점 실행 플랫폼은 다음과 같은 세 단계에 따라 개발되었다. 첫째, 연구진에서 필요로 하는 기능에 대한 최초 정의를 확정한다. 둘째, 연구팀과 협력하는 플랫폼 개발 업체 엔에스데블(NSDevil)에 해당 기능 정의를 전달하여 구현 여부 및 활용 양상에 대한 대략적인 범위를 확정한다. 셋째, 플랫폼 개발 업체의 결과에 대해 연구진이 검토하여 실제 채점 수행 및 데이터 구축에 유효한지를 점검한다.



[그림 V-2] 채점 실행 플랫폼 개발 과정

본 채점 참여자들의 채점 과정에서 수반된 문제 및 개선 방안을 수렴하여 지속적인 플랫폼 개선에 기여한다. 채점 실행 플랫폼은 그 운영 과정에서 초기 개발 단계에서 예측하지 못한 문제가 발생할 가능성이 존재한다. 이에 연구진이 채점 실행 플랫폼 기능을 지속 관리함과 동시에 채점자들로부터의 피드백을 바탕으로 플랫폼 개선을 실시한다. 파악된 문제들을 플랫폼 개발 업체에 전달하고, 업데이트 및 수정을 통해 문제를 개선하여 그 현황을 다시 채점자들에게 제공하는 과정을 거친다.⁵⁾

3) 채점 실행 플랫폼의 실제

채점 실행 플랫폼은 글쓰기 자료 플랫폼과의 연동성을 바탕으로 한다. 글쓰기 자료 플랫폼에서 정제 및 검토를 거친 데이터를 채점 실행 플랫폼에서 채점에 배정할 수 있도록 시스템이 구축되었다. 따라서 채점 실행 플랫폼은 III장에서 제시한 글쓰기 자료 수집 플랫폼과 도메인은 분리되어 있으나 서로 연동된 시스템으로 작동한다. 본 절에서 다루는 채점 실행 플랫폼 자체는 두 가지로 나누어지는데 (1) 채점 수행 플랫폼과 (2) 채점 관리 플랫폼이다. (1) 채점 수행 플랫폼은 본 채점 참여자들이 계정을 부여받아 할당된 글쓰기 자료를 읽고 점수를 부여함으로써 채점 데이터를 입력할 수 있는 플랫폼이다. (2) 채점 관리 플랫폼은 채점자들에게 글쓰기 자료를 할당하며, 채점자들의 채점 현황을 관리하는 기능을 한다.

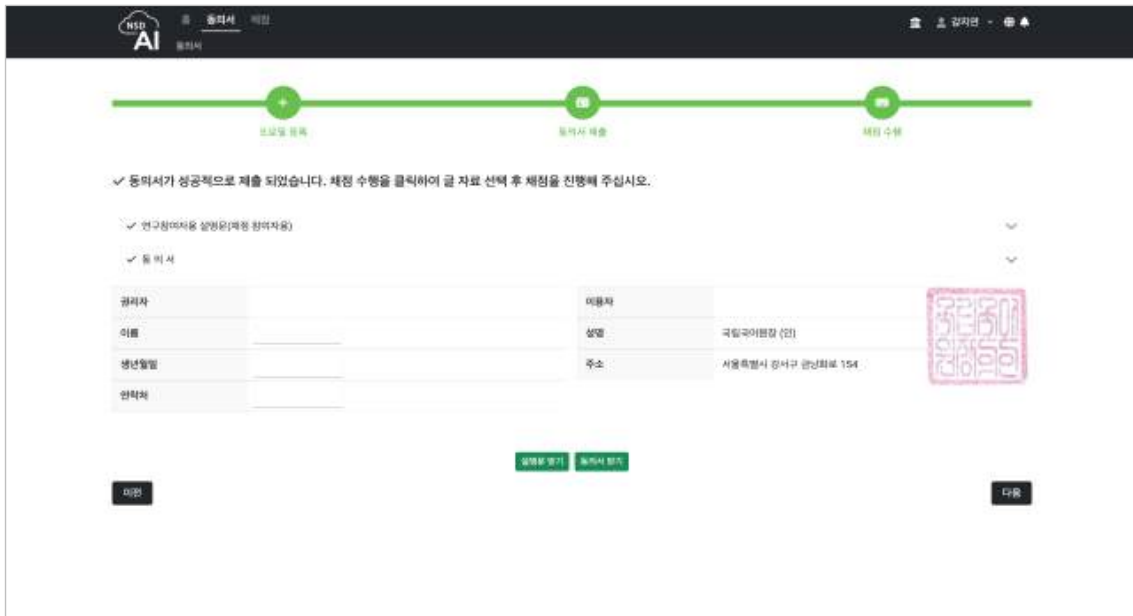
5) 업데이트는 채점 활동이 비교적 드물게 이루어질 것으로 예상되는 지정 시각에 진행되었다. 아울러, 업데이트에 대한 알림을 채점자들이 1회 로그인할 때 게시함으로써 플랫폼 이용 편의를 도모하였다.

(1) 채점 수행 플랫폼

채점 수행 플랫폼은 프로필 등록, 동의서 제출, 채점 수행의 세 하위 항목으로 구성된다. 채점자들은 연구진으로부터 미리 계정(ID, PW)을 부여받아 최초 로그인을 시도한다. 로그인을 시도하면 프로필을 입력하게 되고, 동의서를 작성하게 된다. 프로필 입력을 통해 기본 신상 정보 및 평가 교육 경력 정보를 기입한다. 프로필 작성 후 동의서 작성 단계로 넘어가게 되면 연구 참여 설명문 및 동의서가 제시된다. 본 연구 사업은 서울대학교 연구윤리위원회(IRB) 승인을 획득하여 채점 데이터 수집을 시행하므로, 해당 연구의 목적과 개요, 계획을 포함하는 설명문 및 동의서를 참여자들로부터 확보하도록 하였다. 본 연구에서 재채점 등 연구 진행에 따라 발생하는 추가적 채점을 위해 채점자 아이디 및 그룹 개설 기능을 구현할 필요가 있었다. 이에 따라서, 동일한 신원의 채점자에 대해서는 개별 아이디에 대해서도 개인정보가 연동되어 효율적으로 채점자 정보 파악이 가능하도록 플랫폼 기능을 구현하였다.

[그림 V-3] 채점 수행 플랫폼 프로필 입력 화면

[그림 V-3]의 프로필 입력 페이지를 통해 채점자는 본인의 기본 신상 정보 및 평가 경력 파악에 요구되는 경력 정보를 입력하게 하였다.



[그림 V-4] 채점 수행 플랫폼 동의서 화면

채점 수행 플랫폼을 통해 채점자들은 연구 참여 설명문과 동의서를 읽고 제출할 수 있다. 동의서 작성을 마친 채점자는 채점 항목에서 채점 수행을 시작할 수 있다.



[그림 V-5] 채점 수행 플랫폼 글쓰기 자료 목록 화면

[그림 V-5]에서 볼 수 있는 것처럼 채점자들은 플랫폼을 통해 본인에게 할당된 글쓰

기 자료 목록을 확인할 수 있다. 글쓰기 자료의 문항 및 ID, 글자 수 정보가 제시되며 할당된 글쓰기 자료의 채점 마감 기한 또한 표시된다.



[그림 V-6] 채점 플랫폼 채점 수행 화면1

채점자가 목록에서 개별 글쓰기 자료를 클릭하면 [그림 V-6]에 제시된 채점 수행 페이지로 이동한다. 글쓰기 자료의 문항을 확인할 수 있으며, 학생이 제출한 글쓰기 자료 답안이 정제된 형태로 제공되어 채점을 시작할 수 있다.

연구 진행 과정에서 재채점 등 추가적인 채점이 진행됨에 따라, 사전에 채점한 자료와 신규로 채점하는 자료 간의 적절한 구분이 플랫폼상에 구현될 필요가 있었다. 이에 따라, 재채점에 참여하는 채점자들이 본인이 사전에 채점한 글쓰기 자료에 접근하여 기입력된 채점 점수를 의도하지 않게 훼손하는 일을 방지할 필요가 있었다. 시기를 달리하여 수행된 채점 자료 묶음에 대해서는 그 접근의 허용 및 열람을 구별하는 기능을 구현하였다. 이러한 기능에 의해 재채점에 참여한 채점자들이 사전에 참여한 채점 자료를 제외한 당면 채점 과제에 대해 집중적인 채점 수행을 하도록 하였다.



[그림 V-7] 채점 플랫폼 채점 수행 화면2

[그림 V-7]과 같이 채점자들이 플랫폼에서 직접 점수를 입력할 수 있도록 페이지 기능이 구현되어 있다. 개별 점수 화면에 마우스를 가리키면 채점 기준을 확인할 수 있고, 채점 척도를 PDF 문서로 내려받을 수 있다. 아울러, 특이 사항이 있을 메모를 기재할 수 있으며, 플랫폼의 기능과 관련된 문의가 있을 때 관리자에게 메시지를 전송할 수 있도록 개발하였다.



[그림 V-8] 채점 플랫폼 글쓰기 자료 다운로드 화면

채점자들은 플랫폼상에서 본인에게 할당된 글쓰기 자료를 PDF 문서로 일괄 다운로드 받아 다른 기기나 출력본으로 볼 수 있었다. 다운로드된 글쓰기 자료는 문항 순서, 그리고 글쓰기 자료 ID 순으로 배열되었다.

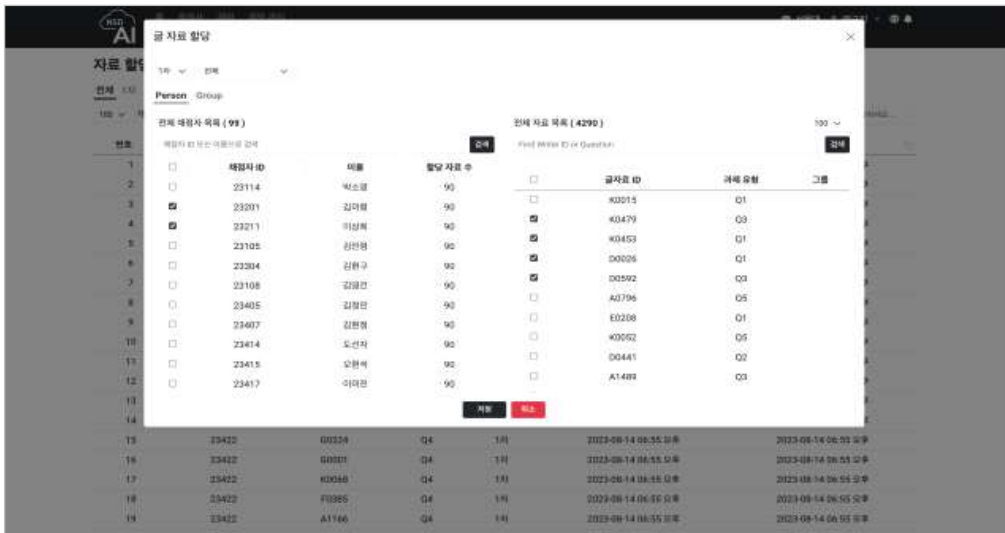
(2) 채점 관리 플랫폼

채점 관리 플랫폼의 주요한 기능은 채점자들에 대한 글쓰기 자료 할당과 실시간 채점 현황에 대한 관리이다. 채점자 워크숍이 지난 후 동일한 간격으로 글쓰기 자료가 할당되고 채점 수행이 배정되었으므로, 워크숍을 통한 채점자 간 협의의 효과가 지속되는 것을 기대하기 위해서는 규칙적인 채점 수행이 요구된다. 이에 채점 진행 속도가 느린 채점자들에게는 연락을 통해 채점 수행을 독려하고 채점 수행에서 난점으로 인식되는 문제들을 확인할 필요가 있다. 이를 위해 플랫폼에서 채점자별 채점 진행률을 실시간으로 확인할 수 있는 기능을 개발하였다.

번호	ID	이름	소속	동역사	회사	채점자 별 진행률 확인		상태	선택
						진행률	마감 기한		
99	23100	박수민	1st Rater	미래솔	1차	0.00% (0/90)	2023-08-26 16:14	진행 중	보기
86	23101	김지연	1st Rater	제출	1차	13.33% (12/90)	2023-08-26 16:14	진행 중	보기
1	23102	고해일	1st Rater	제출	1차	8.89% (8/90)	2023-08-26 16:14	진행 중	보기
23	23103	손기범	1st Rater	제출	1차	0.00% (0/90)	2023-08-26 16:14	진행 중	보기
3	23104	다희정	1st Rater	제출	1차	0.00% (0/90)	2023-08-26 16:14	진행 중	보기
95	23105	김민정	1st Rater	미래솔	1차	0.00% (0/90)	2023-08-26 16:40	진행 중	보기
24	23106	양성희	1st Rater	제출	1차	2.22% (2/90)	2023-08-26 16:40	진행 중	보기
74	23107	박연서	1st Rater	제출	1차	25.56% (23/90)	2023-08-26 16:40	진행 중	보기

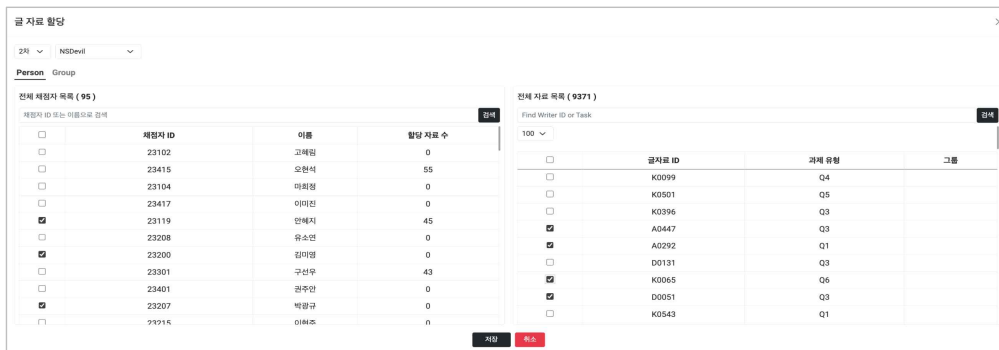
[그림 V-9] 채점 관리 플랫폼 채점자 목록 화면

관리자 계정으로 접속하면 [그림 V-9]와 같이 채점자 목록 및 실시간 진행률, 마감 기한 정보를 확인할 수 있다. 해당 정보를 통해 연구진에서 실시간 진행률이 낮은 채점자들을 모니터링하고, 채점 진행 요청 및 일정 안내를 할 수 있다.

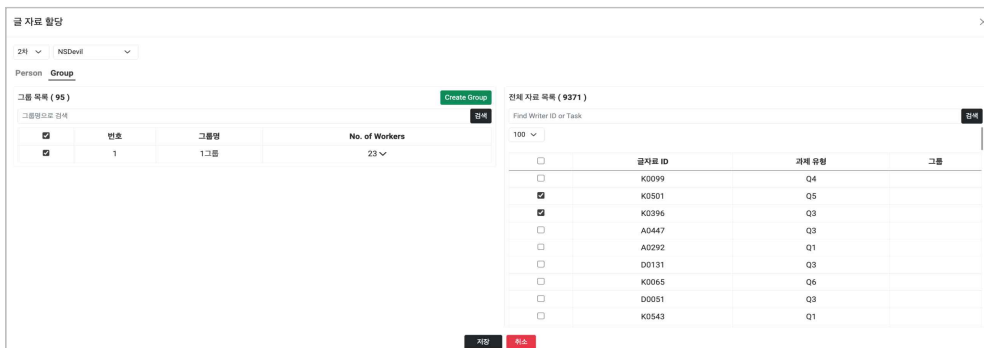


[그림 V-10] 채점 관리 플랫폼 글쓰기 자료 할당 화면1

관리자 계정에서는 글쓰기 자료를 개별 채점자에게 할당할 수 있다. [그림 V-10]과 같이, 채점 설계에 따라 글쓰기 자료를 개별 채점자에게 배정한다.



[그림 V-11] 채점 관리 플랫폼 글쓰기 자료 할당 화면2



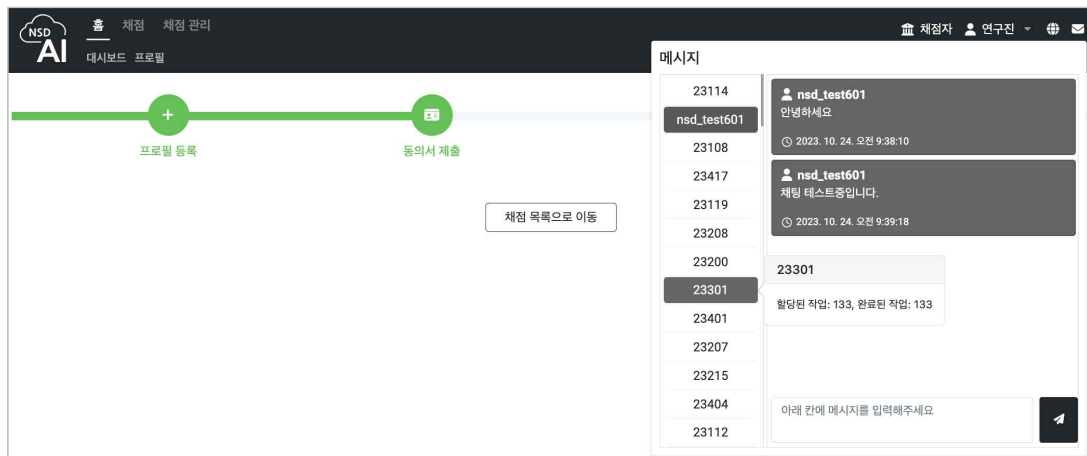
[그림 V-12] 채점 관리 플랫폼 글쓰기 자료 할당 화면3

[그림 V-11]과 같이 플랫폼상 기능을 통하여, 한 명의 채점자에게 복수의 글쓰기 자료를 배정하였으며, 복수의 채점자를 지정하여 동일한 글쓰기 자료를 배정하는 것도 가능하도록 하였다. 또한, 복수의 채점자에 대하여 복수의 글쓰기 자료를 배정하는 기능을 추가하였다. 아울러, 채점 설계 및 실행의 과정에서는 특정한 채점자들을 그 채점 경향 및 수준에 따라 집단으로 분류하고, 집단별로 채점자들을 관리하는 일이 요구된다. 이러한 연구 특성의 필요에 따라 연구진의 채점 계획 설계를 바탕으로 생성된 그룹에 대하여 일괄적인 배정을 수행할 수 있도록 채점 할당 페이지 내에서 그룹을 생성하는 기능을 추가하였다([그림 V-12]). 이를 통해, 특정 그룹에 속한 채점자에게 대량의 글쓰기 자료 할당이 가능하도록 구현하였다.

번호	채점자 ID	글자료 ID	문항	글 자료 할당 세부 코드	회사	최초 시작 시간	마지막 저장 시간	선택
401	23415	D0704	Q4	D0704-0002	2차	2023-10-19 09:52:35	2023-10-29 09:07:57	할당 취소
402	23415	D0687	Q4	D0687-0002	2차	2023-10-19 09:52:35	2023-10-29 09:02:42	
403	23415	D0669	Q4	D0669-0002	2차	2023-10-19 09:52:35	2023-10-29 08:53:38	
404	23415	D0654	Q4	D0654-0002	2차	2023-10-19 09:52:35	2023-10-29 08:45:34	
405	23415	D0637	Q4	D0637-0002	2차	2023-10-19 09:52:35	2023-10-29 08:30:52	
406	23415	D0621	Q4	D0621-0002	2차	2023-10-19 09:52:35	2023-10-29 08:24:36	
407	23415	C0111	Q4	C0111-0002	2차	2023-10-19 09:52:35	2023-10-27 13:15:36	
408	23415	B1581	Q4	B1581-0002	2차	2023-10-19 09:52:35	2023-10-27 13:01:58	
409	23415	B1283	Q4	B1283-0002	2차	2023-10-19 09:52:35	2023-10-27 12:54:15	
410	23415	B0725	Q4	B0725-0002	2차	2023-10-19 09:52:35	2023-10-27 12:30:27	
411	23415	A2014	Q4	A2014-0002	2차	2023-10-19 09:52:35	2023-10-27 12:12:36	
412	23415	H0131	Q2	H0131-0002	2차	2023-10-19 09:52:35	2023-10-26 13:53:58	
413	23415	H0017	Q2	H0017-0002	2차	2023-10-19 09:52:35	2023-10-26 13:41:19	
414	23415	B0258	Q2	B0258-0002	2차	2023-10-19 09:52:35	2023-10-26 13:30:33	
415	23415	A1003	Q2	A1003-0002	2차	2023-10-19 09:52:35	2023-10-26 13:14:05	
416	23415	D1351	Q4	D1351-0009	2차	2023-10-19 09:52:35	2023-10-29 10:31:01	

[그림 V-13] 채점 관리 플랫폼 글쓰기 자료 할당 화면4

[그림 V-13]에서 볼 수 있는 것과 같이 플랫폼 관리자 페이지 내에서 기 할당된 글쓰기 자료에 대한 할당을 취소할 수 있다. 이는 만약 할당 작업에 오류가 생겼을 경우, 오류를 수정할 수 있도록 한다. 또한, 채점 계획 설계가 연구 진행 과정에서 변경되어 할당을 다시 수행할 필요가 있을 경우에도 유용한 기능이 된다.



[그림 V-14] 채점 관리 플랫폼 메시지 송수신 화면

[그림 V-14]에서 볼 수 있는 것처럼, 채점 플랫폼 관리자 페이지에서는 채점자들이 송신한 메시지를 일괄적으로 확인할 수 있다. 채점자들은 플랫폼 이용 과정에서 기술적인 문제가 발생하였을 경우 연구진에게 보고하게 되며, 연구진은 해당 문제에 대하여 자체 해결 또는 업체에의 전달을 통해 채점 과정에 지장이 없도록 지원하였다. 해당 메시지 창에서는 연구진이 채점자들의 메시지를 수신하는 것뿐만 아니라, 완료 여부 또는 해결 방안을 메시지로 송신할 수 있으므로 쌍방향 소통이 가능하도록 구현하였다. 이러한 메시지 기능 창에서는 채점자의 ID 화면에 마우스를 가리킬 경우, 할당된 글쓰기 자료의 총 수량과 완료된 작업의 총 수량도 확인이 가능하도록 하였다.

4) 플랫폼 운영상의 제약과 개선

채점 플랫폼을 운영하는 과정에서, 플랫폼만을 이용해 실시간으로 채점자들의 모든 문의를 접수하는 데에는 한계가 발생하였다. 채점 기한이 임박하고 실시간 대응이 촉박할 경우, 연구진의 비상연락망을 통해 문의 접수 창구를 다원화하였다. 또한, 채점 척도나 글쓰기 자료 해석 등 기술적인 차원을 넘어선 논의를 요구하는 경우에는 연구진 대표 이메일 및 메신저를 병행하여 운영함으로써, 해당 논의를 접수하고, 채점 수행에 대한 지침 및 방향을 제공하는 노력을 하였다.

또한, 채점자들로부터 메시지를 효율적으로 수렴하도록 할 목적에 맞게 메시지 기능을 구현한 결과, 채점자들이 보고한 기술적인 문제가 어떠한 개별 글쓰기 자료에 대한 것인

지를 표기하는 데에 제약이 있었다. 이러한 부분에 대해서는 채점자들에게 문제가 발생한 글쓰기 자료의 번호를 보고할 필요가 있을 경우, 메시지에 정보로 포함하도록 요청함으로써 원활한 소통을 도모하였다.

아울러, 100여 편의 글쓰기 자료를 채점 수행 플랫폼 내에서 읽고 즉각적으로 점수 입력을 동시에 진행할 경우에는 원활한 채점이 어려울 수 있다는 의견이 수렴됨에 따라, 플랫폼 이용을 보조할 수 있는 엑셀 파일 양식을 마련하여 채점을 지원하였다. 엑셀 파일 양식은 플랫폼 화면에 표시되는 채점 준거 화면과 최대한 유사하게 작성하여, 점수를 입력하는 과정에서 혼동이 발생하는 것을 방지하였다. 이러한 엑셀 파일 양식은 유사시 플랫폼 서버의 기술적인 문제 등이 발생하여 채점 입력 데이터가 유실되는 경우가 생기더라도, 채점자 본인이 전산화된 형태의 채점 점수 입력값을 보존하도록 함으로써 데이터 안정성 확보에 기여한다.

3. 본 채점 진행 및 관리

1) 1-4차 채점

(1) 채점 준비

① 채점자 선정 및 채점 그룹 배정

본 채점에 참여하는 채점자의 경우 앞선 채점 전문 인력 양성 프로그램 수료자를 대상으로 본 채점 참여 일정 및 수당에 관하여 안내하고 참여 여부를 조사하였다. 조사 결과, 수료 인원 96명 중 88명이 본 채점 참여 의사를 밝혔다. 이들의 교육 프로그램 참여도, 채점 실습에서의 신뢰도 등을 확인하여 채점자 위촉의 결격 사유가 없음을 확인하였다.

2023년 1학기에 수집된 글쓰기 자료 채점은 일정 조사 결과에 따라 4개 그룹으로 구분하여 진행하였다.⁶⁾ 채점 그룹 배정 과정에서 채점자가 표시한 일정 외에도 신뢰도 수준을 함께 고려하여, 그룹별 채점 신뢰도 수준이 유사할 수 있도록 채점자들을 배치하였다. 이때 채점 신뢰도는 온라인 교육과정의 과제로 제시한 사전 채점 과제 9편과 오프라인 교육 중에 실시한 현장 채점 과제 6편을 종합하여 전체 채점자 집단과 개별 채점자 간의 상관 계수를 통해 산출하였다. 그룹별 배정 인원은 <표 V-2>와 같다.

<표 V-2> 1-4차 채점 그룹별 배정 인원(명)

		채점 그룹			
		1그룹	2그룹	3그룹	4그룹
채점자 교육 채점 신뢰도	상	6	7	4	6
	중	9	8	7	8
	하	9	6	9	9
인원		24	21	20	23
총 인원		88			

6) 2그룹은 '9-10월경에 실시할 5차 채점 가능 여부'를 조사 시에 채점이 불가하다고 응답한 채점자(13명)를 중심으로 배치하였다. 이는 채점 그룹별로 채점하게 될 문항이 다르므로 2학기에 주로 수집하게 될 4-6번 문항을 2그룹을 제외한 나머지 그룹에 할당하기 위함이다.

② 채점 자료 선정

본격적인 채점 설계에 앞서 수집 및 정제된 글쓰기 자료를 수합하고 채점자 워크숍과 공통 채점, 짝 채점에 이르는 각 채점 단계에 필요한 글을 선별하는 과정을 거쳤다. 1학기에 수집된 글쓰기 자료는 4,284건으로 이 중 채점자 워크숍 전에 채점자들이 사전에 진행하는 사전 채점 자료와 채점자 워크숍 당일 현장에서 진행하는 현장 채점 자료를 각각 선별하였다. 이를 위해 연구진 내부에서 문항별로 전형적인 수준에 해당하는 글을 우선 선정하고, 각 글의 특성을 분석하여 가장 전형적이거나 혹은 점수 판정에 관한 논쟁의 여지가 있는 글 등 워크숍 진행에 적절한 글을 각각 사전 채점, 현장 채점 자료로 선별하였다.

③ 채점 설계 구축 및 자료 할당

본 채점 대상 글쓰기 자료 4,284건을 문항별로 구분하면 <표 V-3>과 같다.

<표 V-3> 본 채점 대상 글쓰기 자료 수

문항	1	2	3	4	5	6
자료(건)	794	1,231	1,198	341	381	339
총 자료(건)	4,284					

위의 표에서 살펴볼 수 있듯이 1학기에는 1-3번 문항을 중심으로 글쓰기 자료가 수집되었으며(3,223건), 본 사업의 수집 목표에 해당하는 5,000건의 글 중 약 700건이 4-6번 문항에서 2학기에 수집되었다. 이에 앞서 살펴본 바와 같이 2그룹을 제외한 나머지 그룹에서 4-6번 문항을 채점할 수 있도록 글쓰기 자료를 할당하였다.

<표 V-4> 1-4그룹 글쓰기 자료 할당 현황

문항	전체 자료	1-4그룹 가교	1그룹	2그룹	3그룹	4그룹	합계	나머지 (5그룹 할당)
1	794	6	309	259	220		794	0
2	1,231	6	398	286		295	985	246
3	1,198	6		400	352	440	1,198	0
4	341	8				300	308	33
5	381	8	373				381	0
6	339	8			328		336	3
총	4,284	42	1,080	945	900	1,035	4,002 ⁷⁾	282

위의 표는 그룹별로 채점하게 되는 문항과 자료 편수를 정리한 것이다. 전체 채점 대상 글쓰기 자료 중 가교 채점 대상은 전체 자료 중 1%⁸⁾ 수준에 해당하는 42건이다.

본 채점의 설계는 가교 피험자 설계로 그룹별 채점자가 공통으로 채점하는 글을 문항별로 배분하고, 남은 글들은 2인의 채점자가 하나의 글을 채점하는 짝 채점 방식으로 진행하도록 설계하였다. <표 V-5>는 가교 채점 및 짝 채점 설계의 예시이다.

<표 V-5> 가교 설계 예시

글쓰기 자료 ID	구분	채점자1	채점자2	채점자3	채점자4	채점자5	채점자6	...
A0001	가교	할당	할당	할당	할당	할당	할당	할당
A0002	짝	할당	할당					
A0003	짝		할당	할당				
A0004	짝			할당	할당			
A0005	짝				할당	할당		
A0006	짝					할당	할당	
A0007	짝						할당	할당

7) 당초 계획은 본 사업의 목표 글쓰기 자료 수인 5,000건의 80% 수준에 맞춰 4,000건의 글쓰기 자료에 대한 채점을 진행할 계획이었으나, 회차별 채점자의 수의 변동으로 인해 각 그룹별 채점 수량에 변화하여 최종 4,002건에 대한 채점을 진행하게 되었다.

8) 당초 가교 글쓰기 자료를 2% 수준으로 할당하고자 하였으나, 예산 사정과 채점 완료 시기의 단축 등을 고려하여 1% 수준에서 진행하였다.

위의 설계에 따라 가교 채점에서 개인별 18-20편, 짝 채점에서 개인별 90편을 채점하여 개별 채점자는 전체 채점 기간 동안 총 108-110편의 글쓰기 자료를 채점하였다.

④ 채점 일정 수립

본 채점은 ‘사전 워크숍-공통 채점(가교 채점)-채점 피드백 제공-짝 채점’의 흐름으로 진행되었으며 그룹별 채점 일정은 다음과 같다.

<표 V-6> 1-4차 채점 그룹별 일정

1그룹	2그룹	3그룹	4그룹
사전 워크숍 일정			
8월 7일(월) 13:00-17:00	8월 8일(화) 13:00-17:00	8월 10일(목) 13:00-17:00	8월 11일(금) 13:00-17:00
본 채점 일정			
①공통 채점 ②채점 피드백 ③짝 채점			
①8월 8일-10일 ②8월 13일 ③8월 14일-20일	①8월 9일-11일 ②8월 14일 ③8월 15일-21일	①8월 11일-13일 ②8월 16일 ③8월 17일-23일	①8월 12일-14일 ②8월 17일 ③8월 18일-24일

사전 워크숍은 온라인 화상 회의로 진행하였으며, 워크숍 후 다음날부터 3일간 개인별 18-20편의 ①공통 채점을 진행하고, 그 결과를 바탕으로 개별 채점자의 특성을 분석하여 ②채점 피드백을 제공하였다. 채점자는 채점 피드백을 받은 후 일주일간 개인별 90편씩 ③짝 채점을 진행하였다.

(2) 채점 진행

① 채점자 워크숍

1-4그룹 채점자 워크숍은 1그룹은 8월 7일, 2그룹은 8월 8일, 3그룹은 8월 10일, 4그룹은 8월 11일에 모두 온라인으로 진행되었으며 세부 일정은 <표 V-7>과 같다.

<표 V-7> 1-4그룹 채점자 워크숍 세부 일정

시간	내용
13:00-13:10	워크숍 일정 안내
13:10-13:15	활동 안내
13:15-14:00	소집단 활동 1
14:00-14:20	활동 결과 공유(전체)
14:20-14:30	쉬는 시간
14:30-15:20	채점 연습(실시간)
15:20-15:30	쉬는 시간
15:30-15:35	활동 안내
15:35-16:10	소집단 활동 2
16:10-16:30	활동 결과 공유(전체)
16:30-17:00	일정 및 본 채점 방법 안내

본격적인 워크숍 진행에 앞서 채점자 서약서 작성을 통해 평가 도구와 글쓰기 자료에 관한 보안 유지에 대해 인지할 수 있도록 하였다.

[국민의 글쓰기 능력 진단 체계 개발] 채점자 서약서

본인은 “ 2023년 국민의 글쓰기 능력 진단 체계 개발 ” 사업의 채점자로 참여하면서 다음 사항을 준수할 것을 서약합니다.

1. 채점 그룹별로 정해진 채점 기한을 준수하여 안내받은 양식 및 절차에 따라 평가 결과를 제출한다.
2. 채점 과정에서 소유하게 된 피평가자의 글 자료 및 사업팀의 지적, 물적 재산을 사업팀의 허가 없이 무단으로 사용하지 않는다.
3. 채점 과정에서 소유하게 된 피평가자의 평가 결과를 사업팀의 허가 없이 자신 또는 제3자를 위하여 사용하지 않는다.
4. 본 사업에서 개발한 평가 도구(문항 유형 포함)에 대하여 사업팀의 허가 없이 공개 또는 누설하지 않는다.

서약인 성 명 :
서약 연월일 :

국민의 글쓰기 능력 진단 체계 개발 사업팀 귀중

[그림 V-15] 채점자 서약서 양식

채점자 워크숍에서는 평가 도구(문항 및 채점 기준표)에 대한 공동 인식을 신장시키고 이를 공유하기 위하여 사전 및 실시간 채점 결과에 관한 협의를 진행하였다. 특히 개인 별로 채점 결과의 편차가 큰 순으로 소그룹 내에서 채점 경험을 공유하도록 하였다. 이때 채점 결과를 구글 시트 형태로 공유하여 자신이 입력한 점수와 다른 채점자가 입력한 점수, 최댓값 및 최솟값, 중앙값, 평균, 표준편차 등을 확인할 수 있도록 하였고, 패들렛을 활용하여 소집단 내 협의 내용을 전체 집단이 공유하였다.

사전	실시간	그룹	이름	소속	소집단	글 1	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	평균	표준편차
1-1	평균	4.58	4.38	4.33	4.25	3.67	4.25	4.00	3.96	3.96	4.15						
1-1	표준편차	0.50	0.71	0.70	0.90	1.13	0.79	0.78	0.69	0.75	0.77						
2-1	평균	1.79	1.75	1.88	1.79	1.75	1.50	1.50	1.96	2.04	1.77						
2-1	표준편차	0.93	0.74	0.80	0.88	0.74	0.59	0.66	0.81	0.91	0.78						
3-2	평균	4.58	4.54	4.29	4.33	2.83	4.50	4.42	4.29	4.08	4.21						
3-2	표준편차	0.72	0.78	0.81	0.70	1.05	0.83	0.78	0.62	0.65	0.77						
4-2	평균	3.17	3.33	2.88	3.13	2.04	3.00	3.00	3.67	3.46	3.07						
4-2	표준편차	1.01	0.87	0.80	1.08	1.23	0.66	0.88	0.82	0.93	0.92						
5-5	평균	4.33	4.33	3.83	3.75	3.29	4.29	3.96	4.25	4.58	4.07						
5-5	표준편차	0.64	0.87	0.92	0.99	1.04	1.04	0.86	0.85	0.65	0.87						
6-5	평균	2.92	3.00	2.33	2.29	1.88	2.08	2.25	2.42	2.79	2.44						
6-5	표준편차	0.93	1.02	0.76	0.81	0.85	0.72	0.85	0.97	1.10	0.89						
7-5	평균	2.75	2.42	2.63	2.46	2.54	1.54	1.79	2.42	2.58	2.35						
7-5	표준편차	0.85	0.83	1.01	1.02	0.98	0.66	0.88	0.72	0.97	0.88						

[그림 V-16] 사전 채점 결과 공유 사례

사전 채점에 관한 협의를 통해 채점자 간 인식을 조정한 후 실시간 채점 연습을 진행하였다. 실시간 채점에서는 그룹별로 배당된 문항의 글쓰기 자료 6-7편을 채점하도록 하였다. 이후 실시간 채점 결과를 공유하고 사전 채점과 마찬가지로 소집단 내에서 실시간 채점 결과에 관한 협의를 진행하였다. 모든 협의를 마친 후에는 본 채점 수행 방법과 채점 수행 시 유의 사항에 대해 안내하는 것으로 채점자 워크숍을 마무리하였다.

② 가교 채점⁹⁾

그룹별 공통 글쓰기 자료를 대상으로 진행된 가교 채점을 위해 워크숍 종료 직전 해당 글쓰기 자료와 채점 결과 입력 링크를 채점자에게 메일로 전달하였다. 가교 채점은 채점자별로 생성된 구글 스프레드시트에 글쓰기 자료별 채점 결과를 직접 입력하는 방식으로 진행되었다.

9) 가교 채점은 채점 플랫폼 기능 구현 오류 점검 문제로 채점 플랫폼을 통해 채점을 실시하지 않았다.

번호	문항	글 자료 ID	내용1: 문제 식	내용2: 주장	내용3: 이유/근거	내용4: 이유/근거	내용5: 다른	다른	글	전치	문단	표현1: 문장과 표현2: 어문	어문
1	Q1	E0049	4	4	4	4	3	4	3	4	4	4	도입부가 본론의
2	Q1	E0761	4	5	5	4	2	5	4	4	4	5	마지막 단락이
3	Q1	A1392	4	4	3	4	1	4	3	4	4	4	같은 내용의 문
4	Q1	K0452	3	5	3	4	2	4	3	4	4	5	도입과 결론 부
5	Q1	E0890	3	3	2	2	1	2	2	2	3	3	기제와 로봇의
6	Q1	H0685	3	4	4	3	3	3	3	3	4	4	질문과 대답의
7	Q2	H0755	4	4	4	4	3	4	4	4	4	4	이유나 근거가
8	Q2	H0699	3	4	3	3	2	4	4	3	3	3	주장을 의문시
9	Q2	E0861	4	3	3	3	2	4	3	3	3	4	1단락의 3번째
10	Q2	D0186	3	4	3	3	2	3	3	3	3	4	근거 사례를 제
11	Q2	E0803	4	5	4	4	2	5	4	5	5	5	주장이 명료하
12	Q2	E0999	4	4	4	4	2	3	3	3	3	3	생각한다'의 표
13	Q5	G0393	4	4	3	4	3	4	4	4	4	4	3번째와 4번째
14	Q5	A1368	5	5	4	4	3	5	4	4	4	4	도입의 개념 설
15	Q5	E0963	3	4	3	2	2	3	3	3	3	3	핵심 주장의 분
16	Q5	G0566	4	4	4	2	1	4	3	4	4	4	하위 주장의 근
17	Q5	G0573	4	5	4	4	2	5	4	5	4	4	문장이 간결하
18	Q5	K0295	4	4	4	3	1	3	3	3	3	3	고통이 핵심어
19	Q5	A0940	4	3	3	3	2	3	3	3	4	4	인권의 개념이
20	Q5	C0137	4	4	4	3	2	3	3	3	4	4	5단락은 도입 부

[그림 V-17] 가교 채점 결과 입력 사례

③ 채점 피드백

그룹별로 가교 채점 기간 종료 후 연구진에서 채점자 집단과 개별 채점자의 채점 특성을 분석하여 개인별 채점 특성 피드백을 메일로 송부하였다. 채점 피드백에는 채점자 간 신뢰도 정보와 각 문항별 채점 준거의 평균과 표준편차가 포함되어 있다. 채점자 간 신뢰도를 통해서는 그룹 신뢰도 평균과 채점자 본인의 신뢰도를 확인할 수 있으며, 기술 통계 정보를 통해서는 채점자가 속한 채점 그룹에서 부여한 점수의 평균과 본인이 부여한 점수의 평균을 문항별로 살펴볼 수 있다. 연구진에서는 이 정보를 바탕으로 채점자가 앞으로 수행해야 할 채점에 참고할 수 있도록 구체적인 피드백을 추가로 제공하였다. <표 V-8>은 기술 통계 결과에 대한 채점 피드백 제공 내용이다.

<표 V-8> 기술 통계 채점 피드백

유형	피드백
문항별로 해당 준거의 채점자 평균과 그룹 평균의 편차가 0.5-1점(파란색) 또는 1점 이상(붉은색)일 때	그룹 평균과 편차가 큰 준거(음영)의 채점 척도 및 유의 사항을 다시 한번 꼼꼼히 확인한 후 채점해 주십시오.
문항별로 해당 준거의 채점자 평균과 그룹 평균의 편차가 0.5 이내일 때	기존과 같이 채점 척도 및 유의 사항을 따라 채점해 주십시오.

채점 피드백											해석 방법			
작성일:	2023-08-11										1. 채점자 간 신뢰도			
채점자 ID:	23101										의미	평균 값을 기준으로 채점자들이 부여한 모든 점수의 일치도		
채점 자료:	G1-A(공통 채점)										0.85 이상	매우 우수		
1. 채점자 간 신뢰도											0.75-0.84	우수		
그룹 신뢰도 평균											0.65-0.74	양호		
내 신뢰도											0.64 이하	보통 및 이하 수준		
2. 기술 통계											2. 기술 통계			
분량	항목	내용1	내용2	내용3	내용4	내용5	조각1	조각2	표현1	표현2	평균	의미	각 문항별 준거에 따라 평균 및 표준편차 값으로 집단적인 경향성을 나타내는 정보	
Q1	그룹 평균	3.52	3.19	3.76	3.28	1.96	3.26	3.17	3.42	3.68	3.25	평균 ↑	해당 준거에서 점수를 받기가 수월함.	
	표준편차	1.14	1.07	1.02	1.12	1.02	1.13	1.10	0.88	0.90	1.04	평균 ↓	해당 준거에서 점수를 받기가 어려움.	
	나의 평균	3.83	2.83	4.33	3.33	1.00	3.50	3.17	3.83	3.33	3.24	표준편차 ↑	채점자 간 점수의 차이가 큼.	
Q2	그룹 평균	3.43	3.44	3.77	3.53	2.19	3.49	3.31	3.38	3.51	3.34	표준편차 ↓	채점자 간 점수의 차이가 작음.	
	표준편차	1.13	1.05	1.00	1.03	1.13	1.24	1.19	1.01	1.01	1.09	3. 채점 특성		
	나의 평균	3.50	3.17	4.17	4.17	1.67	3.83	3.00	3.33	3.33	3.35	의미	글 수준 판정의 정확도	글 수준을 엄격/관대하게 판단하였느냐에 대한 정보
Q5	그룹 평균	3.24	3.20	3.65	3.13	1.91	3.23	3.13	3.47	3.56	3.17	+1.0 이상	엄격함	글의 수준보다 점수를 낮게 주는 경향
	표준편차	1.28	1.19	1.14	1.23	0.91	1.25	1.21	1.00	0.97	1.13	-0.9~+0.9	정확함	글 수준에 맞게 점수를 주는 경향
	나의 평균	3.38	3.13	4.13	3.63	1.38	3.25	3.13	3.25	3.25	3.17	-1.0 이하	관대함	글의 수준보다 점수를 높게 주는 경향
수행 결과:											3. 채점 특성			
피드백											의미	채점 척도 활용의 적합도	채점 척도를 충분히 적절하게 사용하였느냐에 대한 정보	
그룹 평균과 편차가 큰 준거(유형)의 채점 척도 및 유의사항을 다시 한번 꼼꼼히 확인한 후 재검토 주십시오.											1.45 이상	부적합	수준별로 일관되지 않은 척도 사용 경향	
3. 채점 특성											0.55-1.44	적합	수준별로 적합한 척도 사용 경향	
글 수준 판정이 정확하며, 준거별 척도를 수준에 맞게 사용하는 것으로 나타났습니다. 지금과 같이 안정적인 환경에서 채점 척도와 글을 꼼꼼히 읽어 채점을 진행해 주시기 바랍니다.											0.54 이하	과적합	수준별로 범위를 한정하여 척도를 사용하는 경향	

국립국어원 국민의 글쓰기 능력 진단 체계 개발

[그림 V-18] 채점 피드백 사례

채점자가 제공받은 채점 피드백의 양식은 [그림 V-18]과 같다. 위의 사례에서 보이는 왼쪽 표에서는 채점자의 신뢰도와 평균을 제시하여 채점자가 속한 그룹 내에서 비교해 볼 수 있도록 하였으며, 오른쪽 표에서는 각각의 수치를 해석하는 방법에 대해 안내하였다. 또한 기술 통계에 따른 채점 피드백뿐만 아니라 채점 특성에 따른 피드백을 함께 제시하여 채점자 스스로 자신의 채점 특성을 이해할 수 있도록 하였다. 채점 특성에 따른 피드백은 아래 <표 V-9>과 같이 유형별로 피드백을 제공하였다.

<표 V-9> 채점 특성에 따른 채점 피드백

유형	피드백
판정이 엄격하며, 척도 활용이 부적합 범주일 때	글 수준 판정이 엄격하고, 준거별 척도 사용의 일관성이 부족한 경향이 나타났습니다. 글과 채점 척도 기술문을 꼼꼼히 확인한 후 안정적인 환경에서 채점을 진행해 주십시오.
판정이 엄격하며, 척도 활용이 적합 범주일 때	글 수준 판정이 엄격하고, 준거별 척도를 수준에 맞게 사용하는 것으로 나타났습니다. 기술 통계에서 엄격한 준거(평균 이하)가 무엇인지 살펴보고, 해당 준거의 채점 척도와 유의 사항을 다시 한번 확인 후 채점을 진행해 주십시오.
판정이 엄격하며, 척도 활용이 과적합 범주일 때	글 수준 판정이 엄격하고, 준거별 척도를 제한적인 범주로 좁혀 판정하는 경향이 나타났습니다. 1점과 5점을 포함한 전체 준거별 채점 척도와 유의 사항을 다시 한번 확인 후 안정적인 환경에서 채점을 진행해 주십시오.

판정이 정확하며, 척도 활용이 부적합 범주일 때	글 수준 판정은 정확한 편이며, 준거별 척도 사용의 일관성이 부족한 경향이 나타났습니다. 채점 척도와 유의 사항을 다시 한번 꼼꼼히 읽은 후 안정적인 환경에서 채점을 진행해 주시기 바랍니다.
판정이 정확하며, 척도 활용이 적합 범주일 때	글 수준 판정이 정확하며, 준거별 척도를 수준에 맞게 사용하는 것으로 나타났습니다. 지금과 같이 안정적인 환경에서 채점 척도와 글을 꼼꼼히 읽으며 채점을 진행해 주시기 바랍니다.
판정이 정확하며, 척도 활용이 과적합 범주일 때	글 수준 판정은 정확한 편이며, 준거별 척도를 제한적인 범주로 좁혀 판정하는 것으로 나타났습니다. 기존과 같이 채점을 진행하시되, 1-5점을 포함한 전체 준거별 채점 척도와 유의 사항을 다시 한번 확인 후 안정적인 환경에서 채점을 진행해 주십시오.
판정이 관대하며, 척도 활용이 부적합 범주일 때	글 수준 판정이 관대하고, 준거별 척도 사용의 일관성이 부족한 것으로 나타났습니다. 글과 채점 척도 기술문을 꼼꼼히 확인한 후 안정적인 환경에서 채점을 진행해 주십시오.
판정이 관대하며, 척도 활용이 적합 범주일 때	글 수준 판정이 관대하고, 준거별 척도를 수준에 맞게 사용하는 것으로 나타났습니다. 기술 통계에서 관대한 준거(평균 이상)가 무엇인지 확인하시고, 해당 준거의 채점 척도와 유의 사항을 꼼꼼히 읽은 후 채점을 진행해 주십시오.
판정이 관대하며, 척도 활용이 과적합 범주일 때	글 수준 판정이 관대하고, 준거별 척도를 제한적인 범주로 좁혀 판정하는 것으로 나타났습니다. 전체 준거별 1-5점의 채점 척도와 유의 사항을 다시 한번 확인 후 안정적인 환경에서 채점을 진행해 주십시오.

피드백을 제공받은 채점자는 채점 신뢰도를 포함한 자신의 채점 특성을 인지하여 이어지는 짝 채점을 수행하게 된다.

④ 짝 채점

짝 채점은 채점 피드백을 받은 후부터 일주일간 채점 플랫폼상에서 진행되었다. 채점자들은 정해진 기한 내에 자신이 부여받은 아이디로 채점 플랫폼에 접속하여 글쓰기 자료에 해당하는 채점 결괏값을 입력하도록 하였다. 다만 채점 플랫폼에 점수를 입력하는 과정에서 플랫폼상의 오류가 발생할 가능성을 배제할 수 없고, 또 채점자가 채점 결과와 다른 점수를 클릭하는 실수도 발생할 수 있기 때문에 이를 대비하기 위해 채점 입력 결과를 별도의 문서(양식 제공)에 저장하거나 메모하는 등 점수를 기록할 것을 권장하였다.

NSD AI 홈 채점 채점 관리 1차 채점 연구진 1

채점 (점수 및 특이사항 입력)

글자 ID 목록

- A0479 Q1
- A0549 Q1
- A0555 Q1
- A0835 Q1
- A0863 Q1
- A1050 Q1
- A1121 Q1
- A1521 Q1
- B0128 Q1
- D0470 Q1
- E0161 Q1
- E0759 Q1
- F0118 Q1
- F0145 Q1**
- F0177 Q1

전체 목록 화면으로 돌아가기

글자 ID: F0145 문항: Q1 글 자료 할당 세부 코드: F0145-0002 진행 완료

2. 점수 입력

구분	준거	1	2	3	4	5
내용	문제 상황 이해	1	2	3	4	5
	주장	1	2	3	4	5
	이유/근거의 적절성	1	2	3	4	5
	이유/근거의 충분성	1	2	3	4	5
	다른 입장에 대한 고려	1	2	3	4	5
조직	글 전체 구성	1	2	3	4	5
	문단 구성 요건	1	2	3	4	5
표현	문장과 어휘	1	2	3	4	5
	어문 규범과 관습	1	2	3	4	5

3. 채점 시 특이사항

주장의 적절성과 근거의 충분성을 고려함

4. 관리자에게 메시지 전송

아직 여기에 전송된 메시지가 없습니다. 전송된 모든 메시지가 여기에 표시됩니다.

[그림 V-19] 플랫폼상 채점 완료 화면

(3) 채점 관리

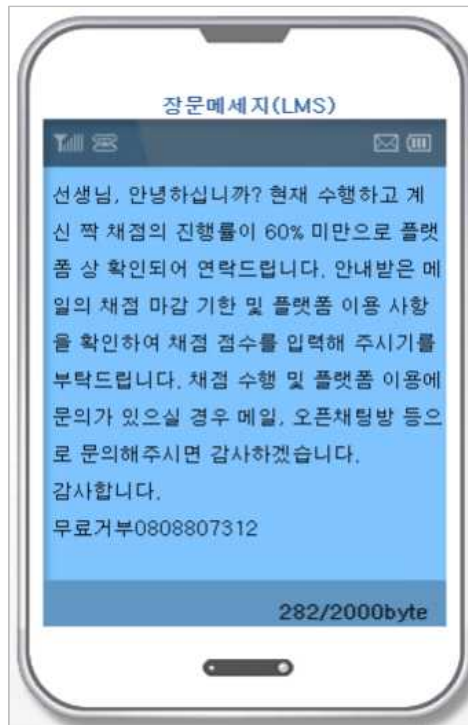
대규모로 진행되는 채점의 경우 한정적인 기한 내에 다수의 채점자가 채점을 진행하게 되므로 추후 원활한 채점 결과 분석 및 연구 진행을 위해서는 채점 과정에서의 관리가 필요하다. 이를 위해 연구진에서는 다음의 네 가지 차원에서 채점 관리가 이루어지도록 하였다.

① 채점 진행률 점검

채점 진행률 점검은 원활한 연구 진행뿐만 아니라 채점자의 신뢰도를 확보하기 위해서도 필수적이다. 채점자가 마감 기한에 임박하여 많은 양을 채점하게 될 경우 피로도가 증가하고, 채점 기준표를 적절하게 적용하지 못하는 등 결과적으로 채점 신뢰도에 부정적인 영향을 미치기 때문이다. 이에 따라 연구진에서는 채점 진행 단계별로 채점 진행률을 점검하고 필요한 경우 개별 채점자에게 마감 기한에 대해 다시 안내하였다. 가교 채점의 경우 3일간 진행되었으므로 가교 채점 종료일 오전 9시에 채점 분량의 50% 미만 수행한 채점자에게 종료일에 대해 재공지하였다. 이후 일주일간 진행된 짝 채점의 경우 <표 V-10> 및 [그림 V-20]과 같이 중간 점검일을 정해 역시 60% 미만 수행한 채점자에게 개별 연락을 취하였다.

<표 V-10> 짝 채점 관리 일정

일정	1그룹	2그룹	3그룹	4그룹
짝 채점 시작	8월 14일	8월 15일	8월 17일	8월 18일
짝 채점 중간 점검	8월 17일	8월 18일	8월 21일	8월 22일
짝 채점 종료	8월 20일	8월 21일	8월 23일	8월 24일



[그림 V-20] 채점 진행률 점검
문자 발송 사례

② 개별 특성 진단 피드백

앞서 살펴본 바와 같이 가교 채점 이후 그 결과를 바탕으로 채점자의 개별 채점 특성을 진단하여 피드백을 제공하였다. 이는 집중적으로 채점을 수행하게 되는 짝 채점에 앞서 채점자가 자신의 채점 특성을 인식하고 필요한 경우 채점 기준표에 관한 인식을 일부 조정할 수 있는 기회를 제공한다. 이 과정에서 자신의 피드백에 관한 추가 설명 및 조정 방향에 대해 문의하는 채점자도 있었으며 채점자 고유의 특성을 저해하지 않는 선에서 필요한 정보를 제공하였다.

③ 부적합/과적합 채점자 개별 면담

채점자에게 제공한 채점 특성에는 글 수준 판정의 정확도와 채점 척도 활용의 적합도가 있다. 이때 지나치게 엄격/관대한 채점자나, 척도 활용에 있어 부적합(1.45 이상) 또는 과적합(0.54 이하)으로 나타난 채점자는 별도로 선별(그룹별 5명 내외)하여 연구진에서 개별 연락을 취해 채점자의 채점 환경을 확인하고 채점 피드백의 해석 방법과 적용 방향에 관하여 안내하였다.

④ 플랫폼 메시지 및 카카오톡 오픈채팅방 활용

채점 과정에서 발생하는 채점자의 다양한 문의에 대해 신속히 대응하고 채점과 관련한 주요 공지 사항을 전달하기 위해 플랫폼상의 메시지와 카카오톡 오픈채팅방(그룹별 개설)을 활용하였다. 플랫폼상에서 관리자에게 직접 메시지를 보낼 수 있는 기능을 통해 채점 플랫폼상에서 발생할 수 있는 문제들에 관하여 플랫폼 관리자에게 문의하고, 카카오톡 오픈채팅방은 채점 기준표 사용에 관한 문제가 발생했을 때 연구진에게 직접 문의할 수 있도록 안내하였다.

2) 5차 채점

(1) 채점자 선정 및 채점 그룹 배정

23년 2학기에 수집된 글쓰기 자료에 대한 5차 채점은 채점자 그룹을 5그룹으로 구분하여 진행하였다. 5그룹 채점자의 경우 채점자 교육 종료 이후 실시한 본 채점 참여 조사에서 5차 채점 참여(9-10월경 진행 예정)에 대해 ‘가능’으로 표시한 채점자 76명을 1차적으로 선별하였다. 이 중 지난 1-4차 채점에서 1-3번 문항만을 채점한 2그룹 16명을 제외한 채점자 60명을 대상으로 하여 참여 가능 여부에 대해 재조사하였다. 재조사 결과 60명 중 55명이 참여 의사를 밝혔으며, 문항별 채점 수량 및 예산, 본 채점에서의 채점자 신뢰도 등을 고려하여 최종적으로 44명을 5차 채점자로 선정하였다.

(2) 채점 자료 할당

2학기에 수집된 글쓰기 자료는 757건으로 4-6번 문항을 중심으로 수집되었다. 지난 1-4그룹 채점 후 남은 자료와 2학기에 수집된 자료를 5그룹 채점 자료로 할당하였으며, 전체 자료 할당 현황은 다음 표와 같다.

<표 V-11> 1-5그룹 글쓰기 자료 할당 현황

문항	전체 자료	1-4그룹 가교	1그룹	2그룹	3그룹	4그룹	5그룹 가교	5그룹	합계	나머지
1	794	6	309	259	220				794	0
2	1,231	6	398	286		295	3	243	1,231	0
3	1,198	6		400	352	440			1,198	0
4	640	8				300	4	287	599	41
5	590	8	373				4	205	590	0
6	588	8			328		7	245	588	0
총	5,041	42	1,080	945	900	1,035	18	980	5,000	41

5그룹 채점 역시 가교 피험자 설계(<표 V-5>)에 따라 하반기 채점 건수의 약 2%에 해당하는 18편의 글을 가교 채점 대상으로 할당하였다. 개별 채점자는 가교 채점에서 개인별 7편, 짝 채점에서 개인별 35-49편을 채점하여 총 42-56편의 글쓰기 자료를 채점하였다.

(3) 채점 진행

5그룹 채점자 워크숍은 10월 21일(토)에 오전과 오후 중 채점자가 선호하는 시간대를 택일하여 참여하도록 하였다. 1-4그룹 채점자 워크숍과 달리 5그룹 채점자 워크숍에서는 실시간 채점 연습을 생략하고 1회의 소집단 활동을 중심으로 진행하였다. 세부 일정은 다음과 같다.

<표 V-12> 5그룹 채점자 워크숍 세부 일정

시간	내용
10:00-10:10	워크숍 및 활동 안내
10:10-10:40	소집단 활동 1
10:40-10:50	활동 결과 공유(전체)
10:50-11:00	일정 및 채점 안내

5그룹은 지난 1-4그룹 채점과 동일한 문항을 채점한다는 점과 채점 특성이 양호한 채점자가 참여한다는 점, 5차 채점 일정 등을 고려하여 별도의 채점 피드백을 제공하지 않았으며, 공통 채점 및 짝 채점의 구분 없이 10월 22일부터 29일까지 일주일간 개인별 약 50편의 글쓰기 자료를 채점하였다.

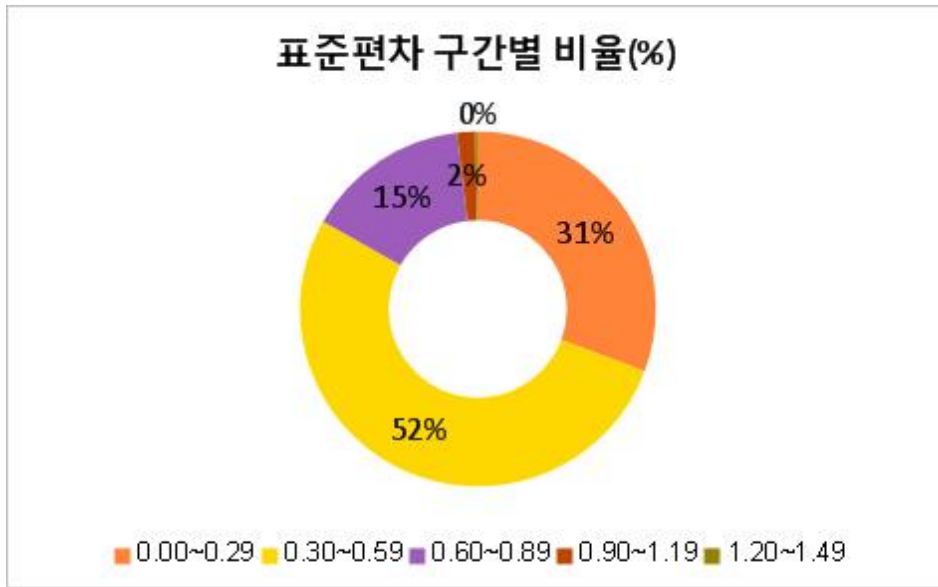
3) 재채점

1-5차 채점 결과는 채점 설계에 따라 한 편의 글에 대한 대표 채점값을 결정하게 된다. 대표값을 최종 결정하기에 앞서 채점자 간 점수 편차를 살펴보고 편차가 큰 경우에는 해당 글쓰기 자료에 대한 재채점을 실시하였다.

이를 위해 먼저 1-5차 채점에서 짝 채점을 실시한 글에 대한 채점자 간 점수 편차를 분석하였다. 분석 결과 각 글에 대한 두 채점자 간의 평균 표준편차는 두 채점자가 모두 일치한 것을 의미하는 0.00에서 1.39까지 나타났다. 짝 채점 글쓰기 자료에 따른 평균 표준편차를 구간별로 살펴보면 다음과 같다.

<표 V-13> 짝 채점 글쓰기 자료별 평균 표준편차

표준편차 구간	0.00-0.29	0.30-0.59	0.60-0.89	0.90-1.19	1.20-1.49	합계
자료 수(편)	1,517	2,594	738	80	12	4,942
비율(%)	30.70	52.50	14.94	1.62	0.24	100.00



[그림 V-21] 짝 채점 글별 평균 표준편차 구간 비율

위의 표와 그림에서 살펴볼 수 있듯이 짝 채점 자료 중 약 83%에서 평균 표준편차가 0.6 미만으로 나타나 대체로 편차가 극히 낮은 수준으로 나타났다. 채점자 간 편차가 1.00을 넘는 자료는 짝 채점자 사이에 전체 9개 준거에서 2점 이상의 차이가 있다는 것을 의미한다. 이에 본 과업에서는 채점자 간 점수의 평균 표준편차가 0.9 이상인 것을 기준으로 하여 92편의 글쓰기 자료를 재채점 대상 자료로 선정하였다. 92편의 자료를 채점 차수별, 문항별로 구분하여 살펴보면 다음 표와 같다.

<표 V-14> 재채점 대상 글쓰기 자료 수

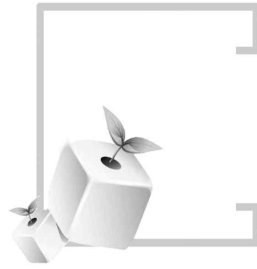
문항 번호	1	2	3	4	5	6	합계
1-4차 채점	22	25	13	-	18	6	84
5차 채점	-	1	-	4	-	3	8
합계	22	26	13	4	18	9	92

재채점에 참여할 채점자 후보군은 5차 채점에 참여하였으며, 채점 특성이 안정적인 채점자를 문항 번호에 따라 구분하여 1순위와 예비 순위 채점자를 선정하였다. 섭외 결과 총 4명의 채점자가 재채점에 참여하였으며, 재채점 기간은 11월 17일부터 20일까지 총 3일간 진행하였다. 재채점 또한 본 채점과 마찬가지로 채점 플랫폼을 통해 이루어졌으며, 채점자별 재채점 편수는 아래 표와 같다.

<표 V-15> 채점자별 재채점 편수

문항 번호	1	2, 4	3, 6	5
채점자	a	b	c	d
채점 편수	22	30	22	18

재채점 대상 자료는 채점자가 본 채점에서 채점한 문항 번호로서 본인이 채점하지 않았던 글을 배정받도록 하였다.



제 VI 장

채점 결과 분석 및 선정



1. 채점 결과 분석

1) 분석 대상 및 방법

(1) 분석 대상

본 장에서는 올해 수집하여 채점 대상으로 선정한 글쓰기 자료 5,000편에 대한 채점 결과를 분석한다. 채점 대상 글쓰기 자료의 수량은 수집 시기와 수집 상황 특성으로 인하여 문항별로 차이가 있었으며, 전체 6개 문항에서 최소 590편(문항 5)에서 최대 1,231편(문항 2)이었다.

<표 VI-1> 문항별 채점 대상 글쓰기 자료

문항	1	2	3	4	5	6	전체
글쓰기 자료(편)	794	1,231	1,198	599	590	588	5,000
비중(%)	15.88	24.62	23.96	11.98	11.80	11.76	100

(2) 분석 방법

채점 결과 분석은 확인하고자 하는 정보의 성격에 따라 원점수 기반의 고전검사이론(Classical Testing Theory) 분석과 척도화된 점수(이하 척도 점수)를 사용하는 문항반응이론(Item Response Theory) 분석으로 진행하였다.

고전검사이론 분석은 전체 데이터의 특성을 나타내는 기술 통계 분석(descriptive statistics)과 채점자 사이의 채점 일관성을 알아보는 신뢰도 분석(reliability), 채점 준거 간 관련성 정도를 가늠하기 위한 상관 분석(correlation analysis), 채점 준거 타당성 검증을 위한 확인적 요인 분석(confirmatory factor analysis)을 실시하였다. 기술 통계 분석에서는 각 문항 및 채점 준거별 평균과 표준편차 정보, 각 문항별 총점 히스토그램을 제시하였다. 신뢰도 분석에서는 가교 채점 결과에 나타난 채점자들의 일치 경향성을 알아보기 위하여 급내 상관분석을 실시하였다. 그리고 채점에 적용한 9개 준거가 단일 모형으로서 어느 정도의 적합도를 가지고 있는지의 타당성을 통계적으로 확인하기 위하여 확인적 요인 분석을 실시하였다.

다음으로 문항반응이론 분석에서는 채점자들의 채점 특성을 분석하기 위하여 다국면 라쉬 모형(Multi Faceted Rasch Model) 기반의 패시트(Facets, 3.86.0)를 사용하여 채점자 국면 정보 중 채점 엄격성(rater severity)과 척도 사용 적합도(Infit MnSq)를 확인하였으며, 문항별 준거 특성은 파스케일(Parscale)을 사용하여 각 문항에서 9개 채점 준거의 곤란도와 변별도를 확인하였다.

2) 분석 결과

(1) 고전검사이론 분석

① 기술 통계

<표 VI-2> 기술 통계 분석 결과

문항	항목	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	평균
1	평균	3.45	3.46	3.16	3.14	2.34	3.20	3.05	3.35	3.46	3.18
	표준 편차	0.99	1.04	0.96	1.00	1.09	1.02	0.97	0.91	0.93	0.99
2	평균	3.53	3.54	3.32	3.19	2.45	3.32	3.21	3.48	3.63	3.30
	표준 편차	0.88	0.96	0.91	0.92	1.06	0.98	0.90	0.91	0.94	0.94
3	평균	3.49	3.48	3.32	3.16	2.42	3.21	3.13	3.56	3.68	3.27
	표준 편차	0.88	0.95	0.91	0.90	1.04	0.98	0.95	0.84	0.91	0.93
4	평균	3.24	3.24	3.08	2.92	2.59	2.94	2.82	3.37	3.50	3.08
	표준 편차	0.86	0.84	0.84	0.85	1.07	0.98	0.87	0.88	0.92	0.90
5	평균	3.13	3.51	3.08	3.07	2.27	2.97	2.89	3.12	3.20	3.03
	표준 편차	0.97	0.94	0.91	0.95	0.97	1.02	0.95	0.89	0.94	0.95

6	평균	3.34	3.57	3.08	3.04	2.45	3.10	2.98	3.35	3.46	3.15
	표준편차	0.88	0.92	0.94	0.95	0.99	1.05	0.98	0.90	0.97	0.95
전체											
평균		3.36	3.47	3.17	3.09	2.42	3.12	3.01	3.37	3.49	3.17
표준편차 평균		0.91	0.94	0.91	0.93	1.04	1.01	0.94	0.89	0.94	0.94

기술 통계 분석 결과, 6개 문항 중 곤란도가 가장 높은 것은 5번 문항(조력존엄사 허용)(M: 3.03)이었으며, 가장 낮은 것은 2번 문항(혐오 표현의 법적 규제)(M: 3.30)이었다. 표준편차 평균은 전체 문항에서 1 미만으로 나타난 가운데 가장 큰 문항은 1번(로봇세)(SD: 0.99)이었으며, 4번(디지털 세계에서의 잊힐 권리)(SD: 0.90)이 가장 작았다.

채점 준거별로는 9개 준거 중 곤란도가 가장 높은 것은 ‘다른 입장 고려’에 관한 내용 5(M: 2.42, SD: 1.04)였으며, 이는 1-6번 문항에서 공통적으로 나타난 특징이었다. 다음으로 ‘어문 규범과 관습’에 관한 표현 2(M: 3.49, SD: 0.94)의 곤란도가 가장 낮았으며, 이는 1-4번 문항에서 나타난 특징과 일치하였다. 표준편차가 가장 큰 채점 준거는 곤란도가 가장 높았던 ‘다른 입장 고려’에 관한 내용 5(SD: 1.04)였으며, 1-4번 문항에서 내용 5의 표준편차가 가장 크게 나타났다. 표준편차가 가장 작은 채점 준거는 표현 1(문장과 어휘)(SD: 0.89)이었으며, 문항 1, 3, 5에서 이와 같은 특징이 나타났다.

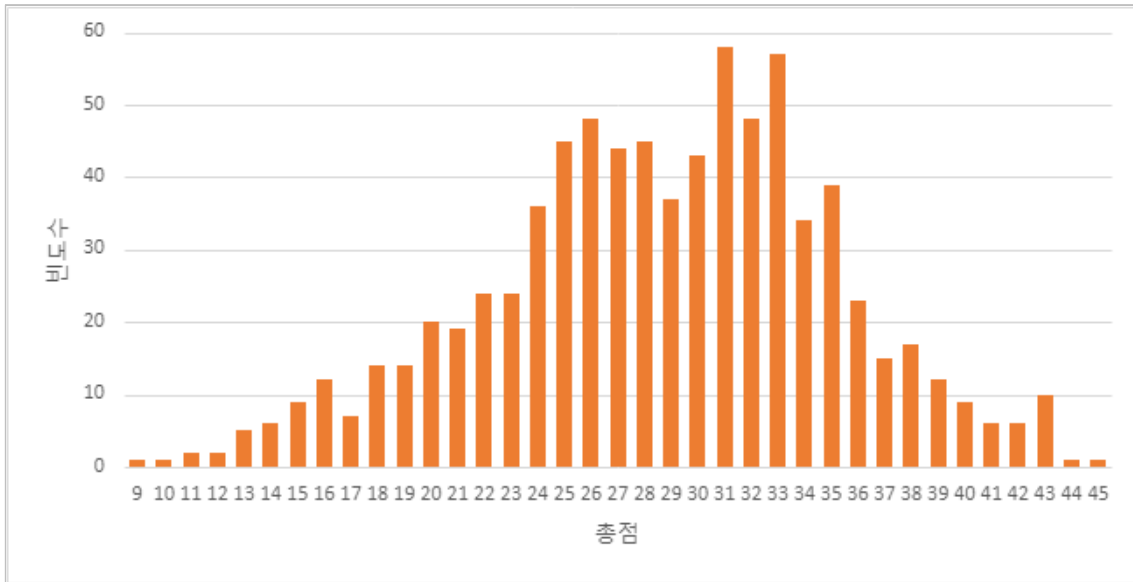
다음으로 문항에 따른 총점 분포를 살펴보았다. 문항별 총점 평균과 표준편차, 첨도와 왜도를 분석한 결과는 <표 VI-3>과 같다.

<표 VI-3> 총점 기준 기술 통계 결과

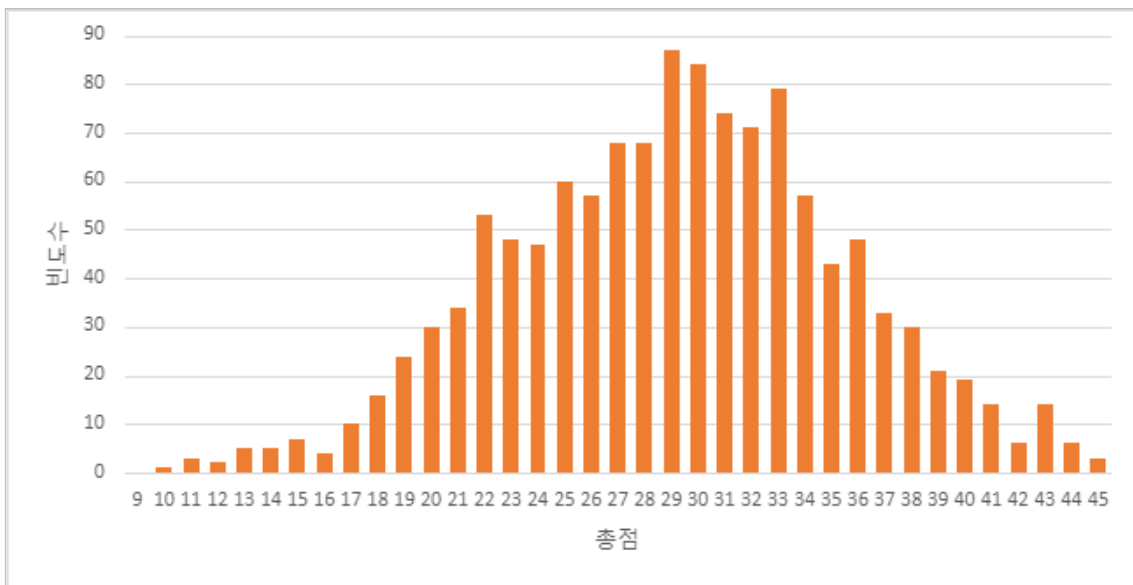
문항	1	2	3	4	5	6
평균	28.61	29.14	29.04	27.70	27.23	28.38
표준편차	6.50	6.28	5.89	5.81	5.91	6.01
첨도	-0.11	-0.17	0.03	-0.02	-0.05	-0.22
왜도	-0.24	-0.09	-0.01	-0.15	0.24	-0.19

총점은 9개 준거, 5점씩 45점을 만점으로 하며 6개 문항 전체의 총점 평균은 28.35였다. 그리고 모든 문항에서 첨도와 왜도의 절대값이 0에 가깝게 나타나 대칭을 이루고 있

는 것으로 나타났다.¹⁰⁾ 문항별 총점의 분포를 도표로 나타내면 [그림 IV-1]-[그림 IV-6]과 같다.

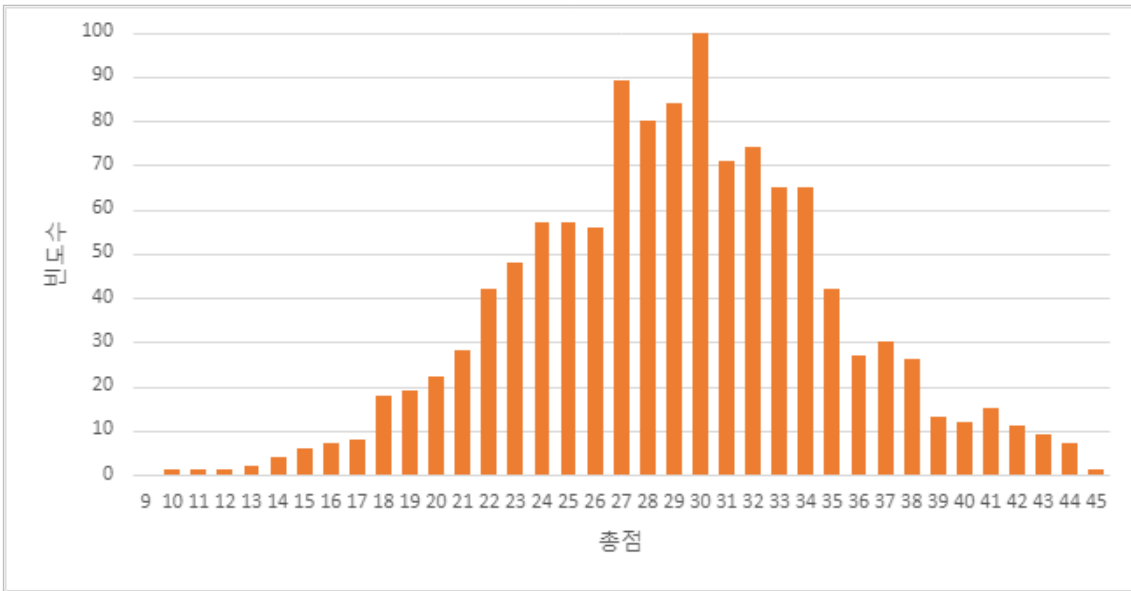


[그림 VI-1] 문항 1 총점 분포

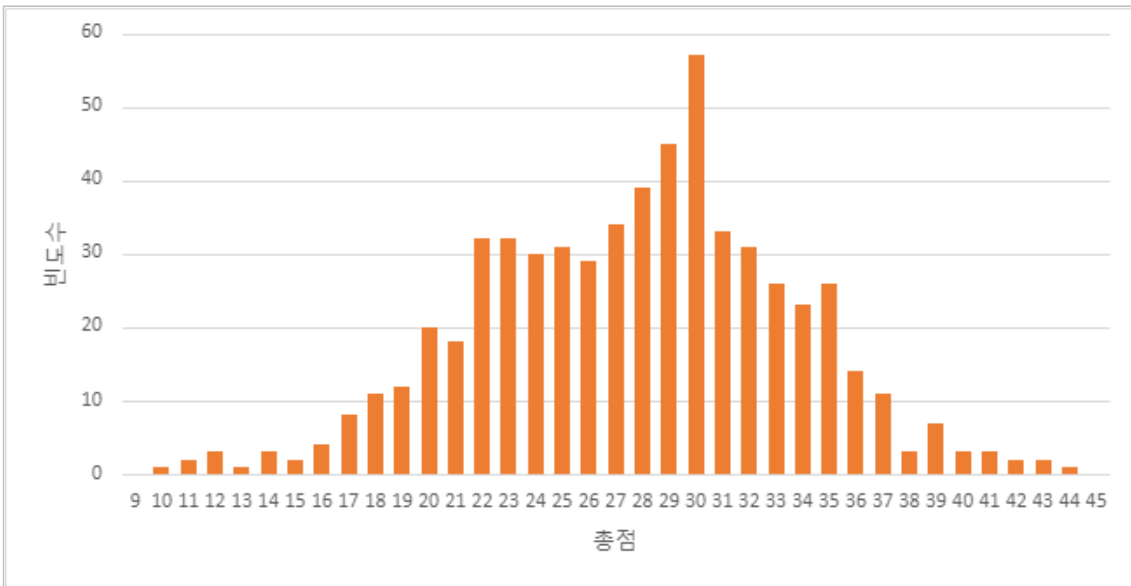


[그림 VI-2] 문항 2 총점 분포

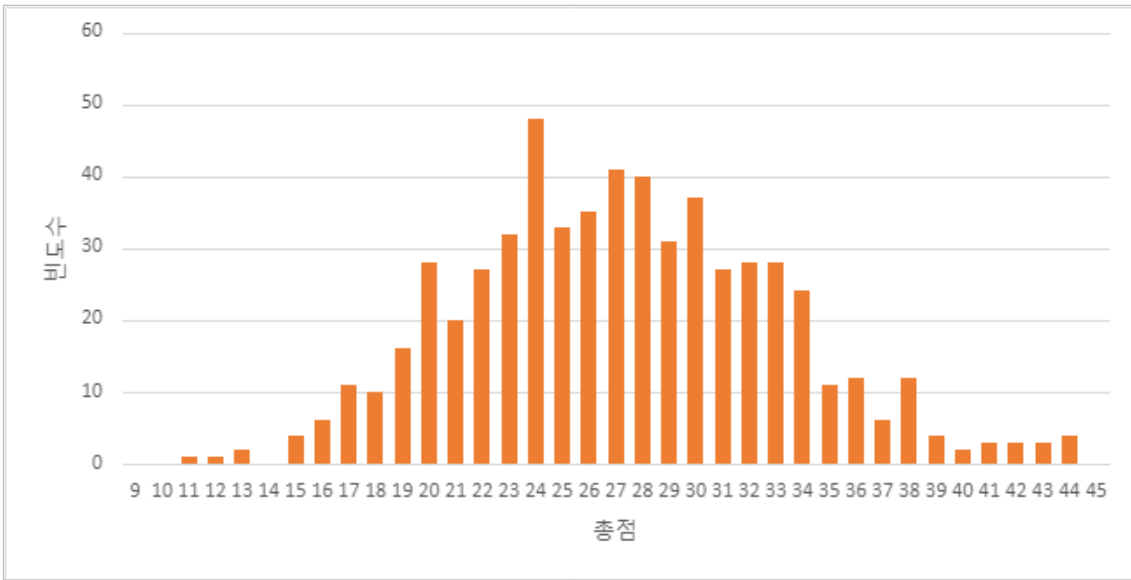
10) 척도는 자료의 분포가 평균을 중심으로 얼마나 뾰족한지 혹은 완만한지를 나타내는 것으로 0을 기준으로 척도가 0보다 작을 경우 정규 분포보다 완만한 것을 의미한다. 반면 척도가 0보다 클 경우 정규 분포보다 뾰족한 모양을 나타낸다. 왜도는 정규분포가 중심축을 기준으로 어느 한쪽으로 치우쳐진 정도를 나타내는 것으로 왜도가 음수인 경우 데이터가 평균보다 오른쪽에 즉, 평균 점수보다 높은 점수에 몰려있음을 의미하며, 왜도가 양수인 경우에는 데이터가 평균보다 왼쪽에 몰려있음을 의미한다.



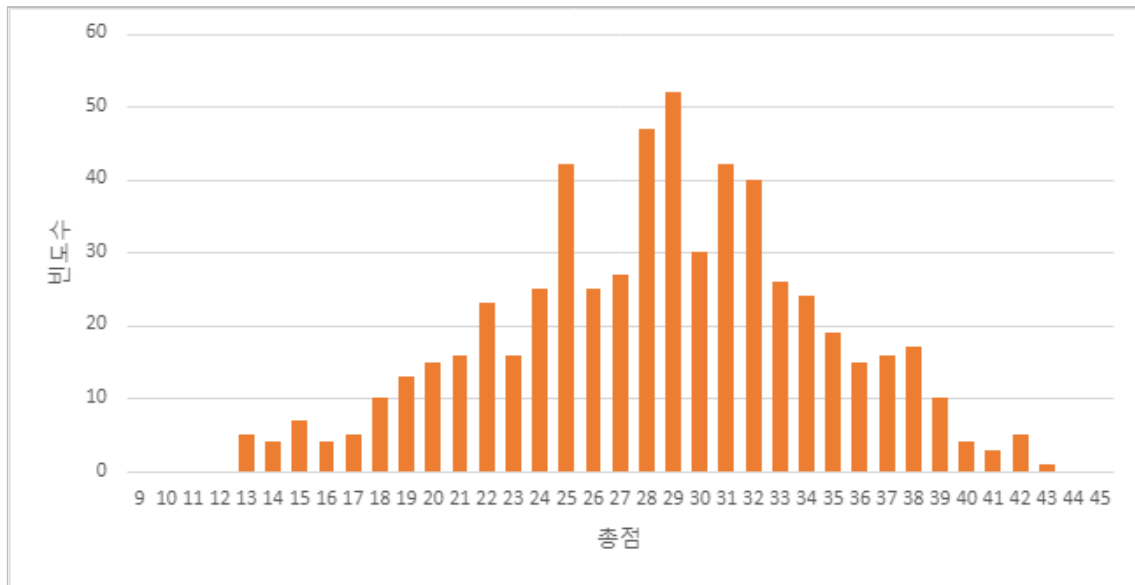
[그림 VI-3] 문항 3 총점 분포



[그림 VI-4] 문항 4 총점 분포



[그림 VI-5] 문항 5 총점 분포



[그림 VI-6] 문항 6 총점 분포

문항 1, 2, 3, 4, 6은 모두 왜도가 음수로 나타나 전체적으로 평균보다 최빈값이 우측에 있었으나 문항 5의 경우 왜도가 양수로 최빈값이 좌측에 위치하는 것으로 나타났다. 이는 평균이나 중앙값보다 낮은 점수에 사례가 집중되어 있음을 의미하는 것으로, 문항 5의 곤란도가 가장 높았던 것과 관련된 것으로 볼 수 있다.

② 신뢰도

본 연구에 참여한 88명의 채점자들은 채점을 담당하는 문항에 따라 4개의 그룹으로 나누어졌다. 이에 따라 채점자 간 신뢰도는 동일한 그룹 내에서 같은 글쓰기 자료를 채점하는 가교 채점 결과에 대한 급내 상관(Intraclass correlation) 분석을 통해 알아보았다. 분석 결과, 각 그룹별 급내 상관은 0.955 - 0.965로 모두 매우 높은 것으로 나타났다. 채점자 그룹 중에서는 1, 3, 6번 문항을 채점했던 3그룹이 0.965로 가장 높고, 2그룹은 0.955로 가장 낮은 수준이었으나, 그 차이는 크지 않았다. 이러한 결과는 채점자들이 일관된 교육과 훈련을 이수한 점, 그리고 공통된 채점 척도를 사용하며 채점 플랫폼을 이용한 점, 채점 피드백 등 지원을 받은 것의 영향으로 볼 수 있다.

<표 VI-4> 채점 그룹별 급내 상관 평균

그룹	1	2	3	4	5(1)	5(3)	5(4)
문항	1, 2, 5	1, 2, 3	1, 3, 6	2, 3, 4	2, 5	6	2, 4
채점자 수	24	21	20	23	16	14	14
급내 상관 평균	0.960	0.955	0.965	0.956	0.940	0.933	0.901

문항별로 채점자들의 신뢰도를 분석한 결과, 각 문항별 급내 상관은 0.901-0.988로 매우 높은 수준으로 나타났으며, 전체 문항 중에서는 3번 문항(고령 운전자 면허 제한)이 0.988로 가장 높았으며, 4번 문항(디지털 세계의 잊힐 권리)의 급내 상관이 0.926으로 상대적으로 낮게 나타났다.

<표 VI-5> 문항별 채점자 간 급내 상관 평균

문항	1	2	3	4	5	6
채점자(명)	65	68	64	23	24	23
급내 상관	0.984	0.986	0.988	0.926	0.959	0.972

③ 상관 분석

글쓰기 자료 채점 결과를 바탕으로 채점에 사용한 9개 채점 준거 사이의 관련성을 확인하기 위하여 상관 분석을 실시하였다. 상관 분석은 SPSS 25.0을 통하여 전체 11,786회의 1-5차 채점의 결과에 대한 이변량 상관 분석으로 이루어졌다. 분석 결과, 9개 채점

준거 중 조직 1(글 전체 구성)과 조직 2(문단 구성 요건) 채점 준거의 상관성이 가장 높은 것으로 나타났으며($r=.805$), 내용 5(다른 입장 고려)와 표현 2(어문 규범과 관습)의 상관성이 가장 낮은 것으로 나타났다($r=.180$). 9개 채점 준거와 총점 간의 상관을 보았을 때 내용 3(이유/근거의 적절성)과 총점의 상관성이 가장 높은 수준으로 나타났으며($r=.825$), 내용 5(다른 입장 고려)의 상관성이 가장 낮은 수준($r=.459$)으로 나타났다.

<표 VI-6> 전체 채점 결과의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.489**	1								
내용3	.540**	.689**	1							
내용4	.468**	.641**	.779**	1						
내용5	.287**	.197**	.289**	.230**	1					
조직1	.470**	.522**	.543**	.547**	.242**	1				
조직2	.491**	.548**	.600**	.595**	.258**	.805**	1			
표현1	.434**	.461**	.526**	.482**	.242**	.526**	.572**	1		
표현2	.346**	.379**	.432**	.401**	.180**	.464**	.489**	.644**	1	
총점	.694**	.754**	.825**	.787**	.459**	.789**	.823**	.741**	.656**	1

** : $p < .01$

다음으로 문항별로 각 채점 준거와 총점과의 관계를 살펴보기 위해 상관 분석을 실시하였다. 분석 결과, 전체 문항 중에서는 2번 문항(혐오 표현 규제)의 준거별 상관의 평균($r=0.744$)이 가장 높았으며, 상대적으로 3번 문항(고령 운전자 면허 제한)이 가장 낮게 나타났으나($r=0.712$), 모두 0.7 수준을 웃도는 양호한 상관을 나타내었다. 채점 준거 중에서 가장 높은 상관을 나타낸 채점 준거는 ‘이유/근거의 적절성’에 관한 내용 3($r=0.824$)이었다. ‘다른 입장에 대한 고려’에 관한 내용 5(SD: 0.464)에서는 상대적으로 낮은 보통 수준의 상관성이 나타났다.

<표 VI-7> 문항별 총점-준거 상관 계수

문항	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	평균
1	0.723**	0.776**	0.827**	0.795**	0.470**	0.790**	0.823**	0.754**	0.645**	0.734
2	0.714**	0.795**	0.845**	0.801**	0.505**	0.809**	0.837**	0.738**	0.648**	0.744
3	0.657**	0.745**	0.806**	0.775**	0.400**	0.793**	0.820**	0.744**	0.666**	0.712
4	0.700**	0.731**	0.815**	0.772**	0.458**	0.785**	0.811**	0.700**	0.661**	0.715
5	0.655**	0.748**	0.850**	0.793**	0.490**	0.774**	0.809**	0.742**	0.617**	0.720
6	0.695**	0.694**	0.802**	0.768**	0.460**	0.755**	0.824**	0.753**	0.693**	0.716
평균	0.691	0.748	0.824	0.784	0.464	0.784	0.821	0.739	0.655	0.723

**: p<.01

상관 분석 결과를 문항별로 나누어 보면, 전반적으로 조직 1(글 전체 구성)과 조직 2(문단 구성 요건)의 상관이 가장 높은 수준으로 나타났으며, 내용 2(주장)와 내용 5(다른 입장 고려), 내용 5와 표현 2(어문 규범과 관습)의 상관은 매우 낮은 수준으로 나타났다. 전체 문항별 채점 준거 간 상관은 <표 VI-8>-<표 VI-13>과 같다.

<표 VI-8> 문항 1의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.551**	1								
내용3	.559**	.700**	1							
내용4	.491**	.666**	.774**	1						
내용5	.312**	.199**	.325**	.251**	1					
조직1	.505**	.542**	.548**	.544**	.261**	1				
조직2	.521**	.569**	.595**	.594**	.273**	.798**	1			
표현1	.470**	.482**	.543**	.517**	.258**	.542**	.583**	1		
표현2	.348**	.405**	.418**	.412**	.152**	.457**	.494**	.641**	1	
총점	.723**	.776**	.827**	.795**	.470**	.790**	.823**	.754**	.645**	1

** : p<.01

<표 VI-9> 문항 2의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.554**	1								
내용3	.580**	.737**	1							
내용4	.492**	.693**	.799**	1						
내용5	.341**	.262**	.330**	.268**	1					
조직1	.508**	.570**	.592**	.581**	.319**	1				
조직2	.519**	.592**	.637**	.620**	.331**	.813**	1			
표현1	.437**	.504**	.539**	.500**	.256**	.540**	.589**	1		
표현2	.355**	.408**	.445**	.410**	.183**	.478**	.497**	.633**	1	
총점	.714**	.795**	.845**	.801**	.505**	.809**	.837**	.738**	.648**	1

** : p<.01

<표 VI-10> 문항 3의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.439**	1								
내용3	.484**	.687**	1							
내용4	.422**	.619**	.756**	1						
내용5	.198**	.182**	.227**	.194**	1					
조직1	.435**	.510**	.539**	.557**	.157**	1				
조직2	.467**	.545**	.580**	.584**	.179**	.801**	1			
표현1	.440**	.442**	.502**	.464**	.191**	.553**	.576**	1		
표현2	.341**	.458**	.409**	.387**	.168**	.497**	.496**	.638**	1	
총점	.657**	.745**	.806**	.775**	.400**	.793**	.820**	.744**	.666**	1

**: p<.01

<표 VI-11> 문항 4의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.483**	1								
내용3	.533**	.698**	1							
내용4	.464**	.610**	.748**	1						
내용5	.315**	.126**	.227**	.187**	1					
조직1	.468**	.503**	.538**	.547**	.254**	1				
조직2	.467**	.555**	.599**	.586**	.273**	.768**	1			
표현1	.395**	.399**	.501**	.448**	.209**	.480**	.516**	1		
표현2	.357**	.363**	.434**	.417**	.188**	.476**	.464**	.627**	1	
총점	.700**	.731**	.815**	.772**	.458**	.785**	.811**	.700**	.661**	1

**: p<.01

<표 VI-12> 문항 5의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.444**	1								
내용3	.508**	.694**	1							
내용4	.436**	.656**	.802**	1						
내용5	.311**	.227**	.380**	.273**	1					
조직1	.419**	.503**	.548**	.530**	.262**	1				
조직2	.423**	.504**	.610**	.593**	.262**	.843**	1			
표현1	.379**	.483**	.546**	.483**	.263**	.499**	.553**	1		
표현2	.278**	.351**	.426**	.366**	.188**	.365**	.423**	.662**	1	
총점	.655**	.748**	.850**	.793**	.490**	.774**	.809**	.742**	.617**	1

** : p<.01

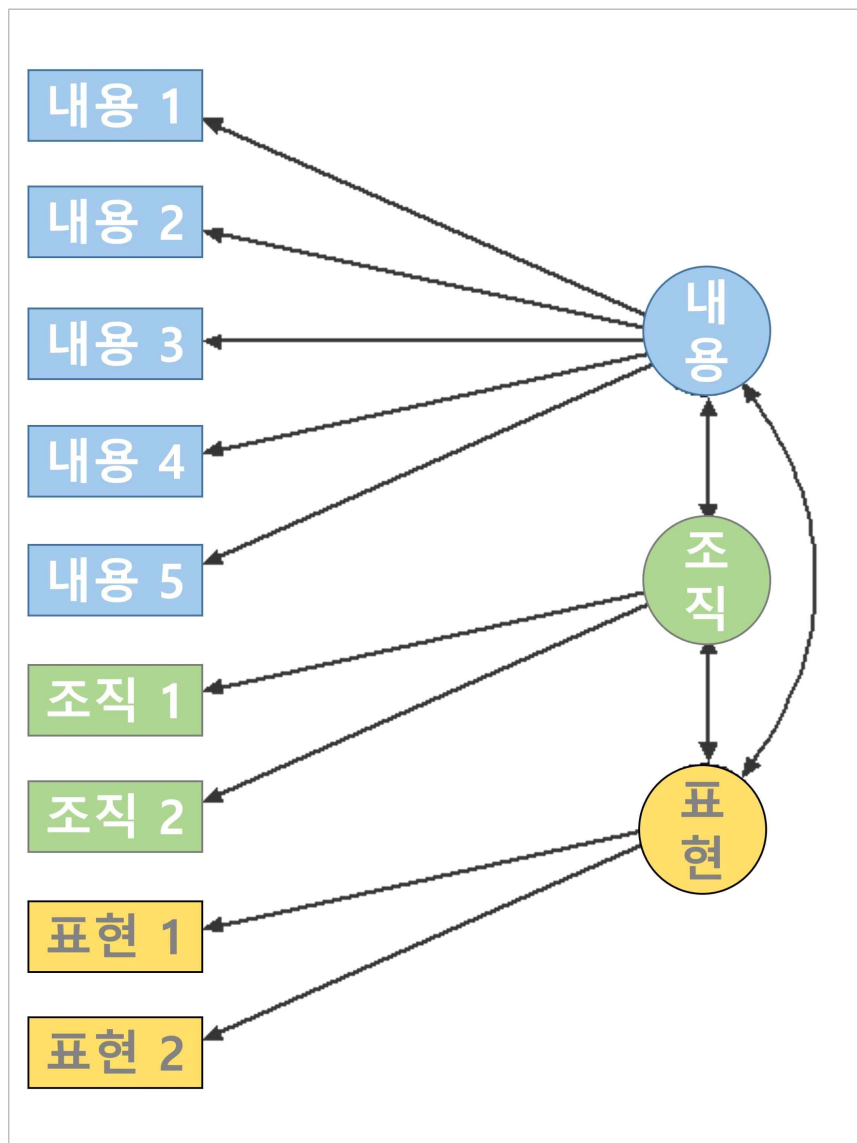
<표 VI-13> 문항 6의 준거 간 상관 계수

준거	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2	총점
내용1	1									
내용2	.433**	1								
내용3	.548**	.606**	1							
내용4	.482**	.653**	.787**	1						
내용5	.290**	.168**	.266**	.233**	1					
조직1	.442**	.454**	.443**	.468**	.234**	1				
조직2	.499**	.481**	.554**	.563**	.269**	.790**	1			
표현1	.450**	.426**	.517**	.459**	.287**	.497**	.579**	1		
표현2	.357**	.379**	.462**	.405**	.226**	.461**	.518**	.671**	1	
총점	.695**	.694**	.802**	.768**	.460**	.755**	.824**	.753**	.693**	1

** : p<.01

④ 요인 분석

본 연구에서 수행한 글쓰기 자료의 채점은 내용, 조직 표현 범주에 해당하는 전체 9개의 채점 준거가 사용되었다. 관련하여 채점에 사용된 3개 범주의 9개 채점 준거가 논증적 글 채점에 적합하였는지를 검증하기 위하여 확인적 요인 분석을 실시하였다. 검증 대상 기본 모형은 본 연구에서 사용된 논증형 글 채점 범주인 내용, 조직, 표현을 각각 하위 5개, 2개, 2개 요인으로 나누어 구성한 것이다.



[그림 VI-7] 확인적 요인 분석의 경로 모형 다이어그램

전체 3개 평가 영역을 요인으로 하위 9개 준거에 대한 확인적 요인 분석 결과, 내용에

해당하는 하위 5개 채점 준거, 조직에 해당하는 하위 2개 채점 준거, 표현에 해당하는 하위 2개 채점 준거의 추정치(estimate)가 모두 유의한 것으로 나타났다. 하위 채점 준거별로 살펴보면 내용 요인에서는 내용 3(이유/근거의 적절성)의 성분 추정치가 가장 높고(0.904) 내용 5(다른 입장 고려)의 성분 추정치가 가장 낮은(0.347) 것으로 나타났다. 조직과 표현 요인에서는 하위 준거들의 성분 추정치가 모두 높은 수준으로 나타났으며, 특히 조직 2(문단 구성 요건)의 경우 가장 높은 성분 추정치(0.980)를 나타냈다.

<표 VI-14> 채점 준거 모델 분석 결과

요인	하위 준거	추정치	표준 오차	Z 점수	p
내용	내용1	0.625	0.009	70.947	< .001
	내용2	0.782	0.008	95.987	< .001
	내용3	0.904	0.007	122.239	< .001
	내용4	0.856	0.008	111.271	< .001
	내용5	0.347	0.010	33.473	< .001
조직	조직1	0.941	0.008	112.279	< .001
	조직2	0.980	0.008	125.973	< .001
표현	표현1	0.826	0.008	101.627	< .001
	표현2	0.714	0.008	84.141	< .001

내용, 조직, 표현의 3개 범주의 크기와 방향을 확인하기 위하여 대한 요인 간 상관 행렬을 확인하였다. 분석 결과, 세 범주 간에는 양호한 상관이 나타났으며, 그중 내용-조직의 상관 추정치($r=.744$)는 내용-표현($r=.674$)과 조직-표현($r=.703$)보다 높게 나타났다.

<표 VI-15> 채점 요인 간 상관 분석 결과

요인	요인	추정치	표준 오차	Z	p
내용	내용	1			
	조직	0.744	0.005	139.763	< .001
	표현	0.674	0.007	97.654	< .001
조직	조직	1			
	표현	0.703	0.007	105.557	< .001
표현	표현	1			

채점 요인 모형에 대한 적합도 분석은 채점 대상 표본의 크기와 통계적 검정력을 고려하여 비교적합지수(CFI)와 터커-루이스지수(TLI), 근사오차평균자승(RMSEA)을 확인하였다. 적합도 분석 결과, CFI와 TLI 모두 0.9 이상으로 양호한 것으로 나타났으며, RMSEA도 0.08 이하로 수용 가능한 범주(0.05-0.08)인 것으로 나타났다.

<표 VI-16> 채점 요인 모형 적합도 분석 결과

CFI	TLI	RMSEA	RMSEA 90% CI	
			Lower	Upper
0.980	0.971	0.063	0.060	0.06629

이상으로 실시한 요인 분석의 결과를 통하여 본 연구에서 설정한 논증적 글의 채점 범주인 내용, 조직, 표현과 하위 채점 준거 9개가 타당하게 작동하였음을 확인하였다.

(2) 문항반응이론 분석

본 연구에서 수집한 채점 결과에 나타난 채점자의 채점 특성과 문항별 준거 특성에 관한 객관적인 비교를 목적으로 문항반응이론 분석을 실시하였다. 분석은 연구 참여자가 단일 문항에 대해서만 글을 작성한 글쓰기 자료 수집과 그에 따른 채점 설계의 특성을 고려하여 각 채점 그룹 및 공통 문항별로 나누어 이루어졌다.

① 채점 특성

채점 특성 분석은 본 연구에 참여한 전체 4개 그룹 88명의 채점자들에게 채점 진행 중 자신의 채점 특성을 점검하고, 채점을 마무리하기까지 안정적으로 채점 특성을 유지 및 개선할 수 있도록 피드백을 제공하기 위한 목적으로 진행되었다. 채점 특성의 분석은 다국면 라쉬 모형(MFRM) 기반의 Facets를 통해 이루어졌으며, 분석 결과 중 채점자 국면에서 제공하는 채점의 엄격성 정보와 척도 사용 적합도 정보를 확인하였다.

채점 특성 분석 결과, 1-4그룹과 5차 채점에 참여한 5(1), 5(3), 5(4) 그룹의 채점 엄격성은 척도상에서 평균적으로 로짓(logit) 값이 모두 음수로 나타나 다소 관대한 경향을 나타내었다. 채점 척도 사용의 특성을 나타내는 내적합 평균 제곱(Infit MnSq)의 경우 척도 사용의 문제가 없음을 의미하는 1을 기준으로 평균적으로는 균형적인 채점이 이루

어진 것으로 나타났다. 그룹별로 보면 척도 사용의 일관성이 떨어져 부적합으로 판정되는 1.5를 넘는 경우와 반대로 제한적으로 척도를 사용하는 과적합 경향을 의미하는 0.5에 가까운 채점 척도 사용의 특성을 나타낸 채점자가 있음을 확인하였다.

<표 VI-17> 그룹별 채점 특성에 대한 다국면 라쉬 모형 분석 결과

그룹	채점자 수	문항	엄격성(logit)			척도 사용(Infit MnSq)		
			최대	최소	평균	최대	최소	평균
1	24	1, 2, 5	0.61	-1.06	-0.29	1.66	0.52	0.99
2	21	1, 2, 3	0.44	-1.42	-0.54	1.68	0.52	1.01
3	20	1, 3, 6	0.31	-0.85	-0.31	1.54	0.62	0.99
4	23	2, 3, 4	-0.26	-1.45	-0.68	1.75	0.47	1.02
5(1)	16	2, 5	0.49	-1.26	-0.18	1.81	0.54	1.01
5(3)	14	6	0.68	-1.61	-0.23	1.75	0.65	1.00
5(4)	14	2, 4	0.36	-1.54	-0.55	1.77	0.6	1.00

다음으로 전체 채점 결과 수집 후 실시한 문항별 파스케일 분석에서 채점자 모수를 확인한 결과, 전반적으로 문항 1-3에서 채점자 효과의 크기가 균형을 의미하는 0에 수렴하는 것으로 나타났다. 그에 반해 문항 4-6의 경우에는 채점자 효과의 값이 크게 나타나는 경향이 있었다. 관련하여 문항 1-3은 채점자 교육 및 훈련 과정에서 반복적으로 교육 및 연습이 이루어진 문항인 것에 반해 문항 4-6은 본 채점 워크숍을 통하여 채점자들이 처음 경험한 문항이었다는 점에서 채점자 효과에 영향을 주었을 것으로 판단된다.

<표 VI-18> 문항별 채점자 효과 분석 결과

항목	문항 1	문항 2	문항 3	문항 4	문항 5	문항 6
최대	.116	.116	.080	.598	.877	.305
최소	-.087	-.073	-.080	-.566	-.729	-.606
평균	.000	.000	.000	.000	.000	.000
중앙값	-.004	.000	.005	.005	-.052	.073

② 문항별 준거 특성

전체 채점 결과로부터 문항별로 9개 준거 사이의 비교 가능한 곤란도와 변별도 정보를 얻기 위하여 파스케일(Parscale) 분석을 통해 각 준거의 척도 점수를 산출하였다.¹¹⁾

분석 결과, 각 문항별로 준거의 곤란도를 살펴보면 공통적으로 전체 문항에서 가장 곤란도가 높은 것은 내용 5(다른 입장 고려)였으며, 나머지 준거들과의 격차가 현저한 것으로 나타났다. 곤란도가 가장 낮은 준거는 문항에 따라 차이가 있었는데, 문항 1-4의 경우에는 표현 2(어문 규범 및 관습)가 가장 곤란도가 낮은 준거였으나, 문항 5와 6에서는 내용 2(주장)의 곤란도가 가장 낮은 것으로 나타났다.

<표 VI-19> 문항별 곤란도 분석 결과

준거	문항 1	문항 2	문항 3	문항 4	문항 5	문항 6
내용 1	-.080	-.111	-.126	.111	.314	.095
내용 2	-.120	-.160	-.079	.120	-.155	-.305
내용 3	.272	.136	.135	.339	.368	.357
내용 4	.294	.231	.304	.542	.452	.421
내용 5	2.170	1.843	2.883	1.923	2.431	2.209
조직 1	.211	.116	.257	.460	.427	.314
조직 2	.304	.207	.293	.614	.528	.356
표현 1	.107	-.031	-.050	.073	.231	-.034
표현 2	-.203	-.228	-.240	-.199	.173	-.208

다음으로 각 문항별로 사용한 준거의 변별도 분석 결과, 변별도가 가장 높은 준거는 문항 1, 3, 6에서는 조직 2(문단 구성 요건)였으며, 문항 2, 4, 5에서는 내용 3(이유/근거의 적절성)이었다. 반대로 변별도가 가장 낮은 준거는 전체 문항에서 내용 5(다른 입장

11) 파스케일 분석 과정에서 본 연구에 참여한 채점자의 수(88명)가 많고, 문항의 다양성(6개 문항)이 큼으로 인해 분석 모형이 작동하지 않는 현상이 나타났다. 이는 올해 연구에서는 조사 참여자가 단일 문항에 대해서만 글을 작성하였기 때문에 복수의 문항을 동시에 비교하기 위한 가교 문항이 없었기 때문이다. 이에 따라 문항별로 척도화 하여 채점 준거의 곤란도와 변별도를 산출하였으며, 각각의 문항 간 수치를 비교하는 방식이 아닌 경향(pattern)을 통해 상대적으로 해석을 해야 하며, 문항 내에서의 해석을 우선적으로 할 필요가 있다. 아울러 본 분석 과정에서 채점 설계 특성(채점자×문항×채점 준거)의 복잡도가 높아 통계 패키지의 구동 문제가 발생하여 점수 척도상의 반응이 약한 1점의 경우 2점과 통합하여 분석을 실시하였다.

고려)로 나타났으며, 이는 해당 준거의 곤란도가 전체 문항에서 매우 높아 참여자들이 글에 잘 담아내기 어려운 준거였고, 이로 인하여 채점자가 글쓰기 자료의 수준을 변별하는데에도 활용하기 어려웠던 것으로 볼 수 있다.

<표 VI-20> 문항별 변별도 분석 결과

준거	문항 1	문항 2	문항 3	문항 4	문항 5	문항 6
내용 1	.807	.864	.665	.751	.570	.759
내용 2	1.227	1.483	1.051	1.018	1.045	.744
내용 3	1.590	2.077	1.516	1.686	1.849	1.280
내용 4	1.374	1.593	1.272	1.347	1.376	1.097
내용 5	.299	.323	.206	.236	.290	.277
조직 1	1.218	1.352	1.433	1.131	1.027	.907
조직 2	1.603	1.723	1.808	1.458	1.358	1.425
표현 1	.978	.936	.931	.736	.834	.865
표현 2	.607	.633	.641	.635	.486	.628

2. 채점 결과 선정

1) 선정 방법

본 연구에서 실시한 5,000편의 글쓰기 자료에 대한 채점 결과의 활용 목적 중 하나는 한국어 글쓰기 자동 채점(AES) 기술의 개발이다. 관련하여 글쓰기 자료와 함께 채점 결과 자료를 기계가 학습하는 방법 중에 사람 채점자의 채점 결과와 같은 산출물을 제공하는 방식으로 해당 글쓰기 자료를 대표 채점자가 채점하고, 해당 결과를 함께 사용하는 방법과 하나의 글쓰기 자료를 복수의 채점자가 채점한 후 대푯값을 사용하는 방법 등이 있다. 그중에서 본 연구에서는 한 명의 채점자가 단독으로 채점한 결과를 사용하는 것으로 인해 발생할 수 있는 채점 결과의 편향성 및 비신뢰성 문제를 확인 및 보완하기 위하여 하나의 글쓰기 자료를 복수의 채점자가 채점하는 방식으로 적용하여 채점 결과를 수집하였다. 이에 따라 본 과제에서는 복수의 채점자가 채점한 결과 중 해당 글에 대한 대표 채점 결과로 선정하고, 이를 학습 데이터로 사용하여 자동 글쓰기 채점 모형 구축에 활용하고자 하였다.

본 연구의 채점 설계를 따라 해당 채점자 그룹 전체가 공통으로 하는 가교 채점과 임의로 배정된 두 명의 채점자가 공통으로 하는 짝 채점 결과 각각에 대하여 공통으로 채점한 채점자들의 채점 특성(엄격성, 척도 사용)을 알아보기 위하여 Facets을 통해 다국면 라쉬 모형 분석을 실시하였다. 이러한 채점 특성 기반 접근은 문항반응이론에 따른 단일 척도화된 정보를 활용하여 비교 가능한 결과를 확보하고, 복수의 채점자 중 한 명의 우수한 채점 특성을 갖고 있는 채점자의 채점 결과를 대표 채점 결과로 선정하기 위하여 적용되었다.

(1) 채점 특성 기반 접근

본 연구에서 채점자들의 채점 특성을 기반으로 글쓰기 자료에 대한 대표 채점 결과를 선정하기 위하여 복수 채점자 간의 채점 엄격성과 채점 척도 사용 특성을 비교하였다. 대표 채점 결과 선정 과정에서 고려한 채점 특성은 채점의 엄격성과 척도 사용의 특성이 다. 채점 엄격성은 해당 채점자가 채점을 할 때 얼마나 엄격하거나 또는 관대하였는지에

관한 정보로서 엄격성의 영향이 없음을 의미하는 0 로짓(logit)에 가까울수록 글쓰기 자료의 절대적 수준에 대해 채점자가 엄격하게 판단하여 점수를 낮추거나 혹은 관대하게 판단하여 점수를 높여 주는 경향이 없음을 의미한다. 따라서 채점 엄격성에 대한 판단에 있어서는 복수의 채점자 엄격성 중 0 로짓에 가까운 것을 선정하는 것이 기본적인 방법이 될 수 있다. 다음으로 척도 사용 특성은 전체 9개 준거에서 사용한 1-5점 척도를 각 준거별 수준에 적합하게 사용하였는지에 관한 정보로서 내적합 평균 제곱(Infit MnSq) 정보를 활용한다. 척도 사용의 특성 정보에서는 균형적인 사용을 의미하는 1.0을 기준으로 1.5 이상은 부적합, 0.5 이하는 과적합으로 판정하며, 1.0에 가까운 척도 사용 특성을 보인 채점자의 채점 결과를 선정하는 것을 기본으로 하였다.

<표 VI-21> 채점 자료 선정을 위한 채점 특성 활용 예시

채점자	채점 엄격성	채점 척도 사용	선정 여부
23700	0.16	1.66	
23701	-0.27	0.88	
23702	-0.69	0.96	
23703	-0.95	0.70	
23704	-0.66	0.54	
23705	-0.83	1.00	
23706	-1.06	1.26	
23707	0.30	1.18	선정(가교)
23708	-0.97	1.11	
23709	-0.97	1.08	선정(짝)

채점 특성을 활용한 대표 채점 결과 선정의 과정은 다음과 같이 이루어졌다. 먼저 채점 엄격성의 영향이 적어 0 로짓에 가까운 채점자를 찾고, 다음으로 그 대상의 채점 척도 사용 특성도 우수한지의 여부를 따져 두 가지 조건을 모두 충족한 대상의 채점 결과를 대표 채점 결과로 선정하였다. <표 VI-21>을 예로 들면, 먼저 가교 채점에서는 전체 8명의 채점자 중 23700은 채점 엄격성이 0.16 로짓으로 0에 가장 가까운 것으로 나타났다. 이어서 채점 척도 사용 특성을 확인한 결과 1.66으로 채점 척도를 부적합하게 사용한 것으로 나타나 채점 척도 사용 특성의 기준을 충족하지 못하였으므로 다시 채점 엄격성을 기준으로 0 로짓에 가까운 차순위 채점자인 23707의 채점 척도 사용 특성을 확인한다. 그 결과 척도 사용 특성이 1.5 미만으로 적합 범주에 해당하는 것으로 확인이 되

었으므로 23707의 가교 채점 결과를 대표 채점 결과로 선정할 수 있다. 다음으로 2명의 채점자가 채점한 짝 채점 글쓰기 자료의 경우 가교 채점과 마찬가지로 두 명의 채점자 사이의 채점 엄격성과 척도 사용 특성을 비교하고, 둘 중 우수한 채점 특성을 갖고 있는 채점자를 선정하여 대표 채점 결과를 선정하였다. 예를 들어 위의 표에서 채점자 23708과 23709가 짝을 이루어 채점을 한 글쓰기 자료의 경우, 두 채점자의 채점 엄격성이 같으나 채점 척도 사용 특성의 측면에서 채점자 23709가 안정적인 척도 사용을 의미하는 1에 더 가까우므로 23709의 채점 결과를 대표 채점 결과로 선정한다.

(2) 선정 결과 보정

한 편의 글쓰기 자료에 대한 두 채점자 간의 판단이 달라지는 주요 원인으로는 해당 글쓰기 자료 자체의 특성의 영향과 기타 문항 및 채점 척도의 영향, 그리고 채점자의 채점 수행의 영향 등이 있다. 본 연구에서는 채점자 간의 채점 특성을 기반으로 대표 채점 결과를 선정하는 방식이 적용되기 어려운 글쓰기 자료로서 짝 채점 사례 중 두 채점자 간 표준편차 평균이 0.9 이상인 글 사례를 전체 채점 자료에서 확인하였다. 해당 글쓰기 자료에 대해서는 이미 1-5차 채점이 진행되어 채점 결과 선정까지 마친 상황이지만, 두 채점자 간의 의견 차이가 상대적으로 크다고 볼 수 있기 때문에 그 결과에 대한 신뢰가 부족하다는 점, 그리고 이러한 격차를 해소하는 것이 전체 채점 결과의 타당성을 높인다는 점 등을 고려하여 재채점을 실시할 필요가 있다. 재채점을 통해 수집된 해당 글쓰기 자료에 대한 새로운 채점 결과는 기존의 짝 채점 결과와 비교하여 어떤 특성이 있으며, 문제적인 부분은 없는지 점검한 후에 최종 채점 결과에 반영하여 활용할 수 있다.

2) 선정 결과

(1) 1-5차 채점

본 연구의 채점 설계를 따라 1-5차 채점의 가교 채점 대상 글쓰기 자료는 각 그룹에 속한 채점자 전체에게 해당 문항별로 할당이 되었으며, 짝 채점에서는 특정 채점자에게 특정 집단의 글쓰기 자료가 할당됨으로 발생하는 채점 결과의 편향 문제를 방지하기 위

하여 채점자들을 채점 신뢰도를 따라 높거나 낮은 사람이 연속으로 배치되지 않도록 정렬한 후 연쇄형으로 글쓰기 자료를 할당하였다. 이에 따라 전체 1-5차 채점에서 가교 채점의 결과는 각 그룹별로 채점 특성 분석을 한 후 그룹별로 가장 우수한 채점 특성을 나타낸 채점자의 채점 결과를 대푯값으로 선정하였다. 짝 채점의 경우 짝을 이룬 두 사람의 채점 특성을 비교하여 더 우수한 채점 특성을 나타낸 채점자의 채점 결과를 확인하였다. 전체 132개의 채점자 쌍의 채점 특성을 비교하였으며, 그 결과 채점 특성이 안정적인 채점자 66명의 짝 채점 결과를 대표 채점 결과로 선정하였다.

(2) 재채점

전체 5,000건 중 92건의 글에 대한 채점 결과에서 채점자 간 편차가 평균 0.9 이상으로 벌어진 것을 확인하였다. 이에 따라 해당 글에 대한 채점 결과 선정을 보류하고, 재채점을 실시하여 짝 채점 결과를 보완하고자 하였다. 재채점 결과에 대한 분석에서는 먼저 기존의 1-5차 채점의 짝 채점 결과와 재채점 결과의 원점수를 비교하였으며, 이를 통하여 재채점 결과의 유형을 파악하고, 문제 여부를 확인하여 최종 채점 결과로 추가할 것인지의 여부를 판단하였다.

재채점 결과에 대한 분석 결과, 재채점 결과와 기존의 짝 채점 결과 사이의 차이는 크게 네 개의 유형으로 나타났으며, 주로는 기존의 채점 결과 중 대표 채점 결과로 선정된 채점 결과와 같거나 비슷한 결과를 나타낸 경우가 92건 중 52건으로 전체의 약 57%를 차지하였다. 다음으로는 반대로 미선정 결과와 유사한 결과를 나타낸 경우가 32건으로 전체의 약 35% 수준이었으며, 나머지는 선정 및 미선정 채점 결과와 총점이 같은 경우(3건, 약 3%), 선정 및 미선정 채점 결과의 총점 범위를 벗어난 경우(5건, 약 5%)였다.

<표 VI-22> 재채점 자료-기존 채점 자료 특성

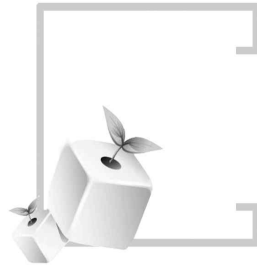
유형	문항 1	문항 2	문항 3	문항 4	문항 5	문항 6	합계
선정 결과 유사	11	14	9	1	9	8	52
미선정 결과 유사	9	10	2	3	7	1	32
선정 및 미선정 결과 중앙	0	2	0	0	1	0	3
선정 및 미선정 점수 초과	2	0	2	0	1	0	5

재채점을 실시한 글쓰기 자료의 채점 결과 중에는 <표 VI-23>과 같이 엄격성이나 채점 척도 사용 특성이 일정 수준 이상으로 채점 특성의 문제가 있었던 경우가 전체의 73%로 나타났다.

<표 VI-23> 문항별 재채점 자료 특성

유형	문항 1	문항 2	문항 3	문항 4	문항 5	문항 6	합계
특이 사항 없음	6	7	6	3	1	2	25
채점 특성 문제	16	19	7	1	17	7	67

이상의 재채점 결과에 대한 분석 결과를 바탕으로 올해 수집한 채점 결과 자료에 재채점 결과를 포함시키고, 대표 채점 결과 선정과 데이터 구축에도 이를 반영하였다.



제 VII 장

인공지능 기반 채점 지원 도구 개발



1. 인공지능 학습 데이터 구축

1) 데이터 관리 체계 수립

본 연구에서는 인간 채점자의 채점을 지원하는 자동 채점 인공지능 개발을 위한 인공지능 학습 데이터를 구축하기 위해 9천여 건의 글쓰기 원자료를 수집하고, 5천여 건의 채점 자료를 생성하였다. 자동 채점 인공지능 개발을 위한 학습용 데이터 세트 구축은 글쓰기 능력 진단 도구를 개발하여 글쓰기 원자료를 획득하는 단계, 진단의 목적에 부합하도록 비유효 데이터를 걸러내고, 개인정보를 비식별화하는 등 글쓰기 원자료를 정제하는 단계, 신뢰할 수 있는 채점 결과를 생성하는 단계, 정제된 글쓰기 자료에 글쓰기 참여자 정보, 글쓰기 과제 정보, 채점 결과 정보 등 주요 정보를 주석화하여 가공하고 최종 검수하는 단계로 진행되어야 한다. 이에 [그림 VII-1]과 같이 인공지능 학습용 글쓰기 평가 데이터 세트 구축을 위한 과정을 설계하고, 데이터 관리 체계를 수립하였다.



[그림 VII-1] 인공지능 학습용 글쓰기 평가 데이터 세트 구축 과정

글쓰기 자동 채점을 위한 학습 데이터는 기존에 생성되어 있는 데이터를 가공하는 방식이 아닌 글쓰기 평가를 통해 직접 데이터를 획득하는 방식으로 구축된다. 따라서 데이터 획득 단계에서 가장 많은 자원이 소요되고, 데이터의 질은 채점 자료 생성 단계에 가

장 크게 좌우된다. 원시 데이터로서 글쓰기 원자료가 어느 정도 통제된 글쓰기 환경에서 주어진 쓰기 과제에 부합하도록 생성되었기에 정제 과정이 매우 복잡하지는 않다. 그러나 정제 단계는 글쓰기 원자료를 검수하여 글쓰기 참여자에게 사례를 지급하기 위한 검수 단계와 이어지고, 이후 채점 자료를 생성하는 단계와 유기적으로 연결되고 최종적으로 채점 결과와 통합하여 데이터가 가공되어야 하므로, 이러한 과정을 종합적으로 통제하고 관리할 수 있는 체계를 구축하는 것이 필요하다. 정제를 거치지 않은 원시 데이터, 정제를 거친 원천 데이터를 분리해서 저장하고, 원천 데이터를 채점자들에게 자동으로 배분할 수 있는 기능까지 갖추어 데이터 획득, 정제, 채점, 가공이 정확하게 이루어질 수 있는 종합 관리 체계가 필요한 것이다.

이에 본 연구에서는 [그림 VII-2]에서와 같이 플랫폼상에 데이터 관리 페이지를 구축하여 학습 데이터 구축을 위한 단계별 과업을 체계적으로 수행하고, 각 단계별 데이터 수를 실시간으로 모니터링하며 저장·관리할 수 있도록 하였다. 화면의 상단에서는 글쓰기 원자료 생성 수량(글 제출 완료)과 모든 정제 단계를 거쳐 채점을 위한 원천 데이터로 활용 가능한 글쓰기 자료 수량(검수 완료)을 실시간으로 확인 가능하도록 하였다. 또한 개별 글쓰기 원자료 정보에 수동 정제가 진행된 여부를 기록하게 하여 원시 데이터에서 모든 정제 단계를 마치기까지의 진행 상황을 모니터링하고, 단계별로 정확한 수량을 파악할 수 있도록 하였다.



[그림 VII-2] 플랫폼 데이터 관리 화면

학습용 데이터 세트 구축의 첫 단계인 글쓰기 원자료 획득의 방법과 절차에 대해서는 3장에서, 채점 자료 생성의 절차와 결과는 5장에서 상세히 기술하였다. 본 절에서는 글쓰기 원자료를 정제하고, 가공하는 단계에서의 과업 수행을 상술하고, 최종적으로 구축된 학습 데이터를 제시하고자 한다.

2) 데이터 정제

(1) 데이터 정제 기준 및 정제 방안

데이터 정제(data cleaning) 과정은 데이터 주석(data annotation) 단계에 들어가기 전에 학습용 데이터로 적합한 데이터를 선별하고 처리하는 작업이다. 데이터 정제는 구축 목적이나 유형 등에 따라 그 절차와 기준이 달라진다. 본 과업에서는 ㉠인공지능 학습 데이터로서 기능을 갖추기 어려운, 질이 낮은 글쓰기 원자료를 선별하여 채점에서 제외하고, ㉡글의 수준에 대해 부정적인 인상을 남겨 채점 일관성에 영향을 줄 수 있는 형식 요소들을 감지하여 편집하며, ㉢개인정보나 학습 데이터로서 윤리적으로 문제가 될 수 있는 내용에 대한 필터링이나 비식별화가 이루어질 필요가 있다.

2022년 301건의 논증적 글쓰기 자료를 수집한 기초 연구에서는 ㉠에 해당하는 질이 낮은 사례로 글자 수 미충족, 논제를 이탈한 내용, 표절이 될 수 있는 외부 자료 활용 등이 발견되었다. 채점 전 문서 편집이 필요했던 ㉡의 사례로는 문장마다 단락 구분을 하여 단락 경계가 드러나지 않는 답안, 제목을 부가한 경우, 서론/본론/결론이라는 구분 항목을 덧붙인 경우와 같이 불필요한 형식 항목이 포함된 경우 등이었다. 2022년도에 획득한 글쓰기 원자료에서 ㉢에 해당하는 사례는 발견되지 않았다.

2023년에는 수집 건수가 확대되었으므로, 데이터 특성이 더 다양하게 나타날 가능성이 있어 5월까지 수집된 2,000여 건의 글쓰기 원자료를 검토하여, 데이터 정제 기준을 재설정하였다. 그 결과 ‘인공지능 학습 데이터로서 기능을 갖추기 어려운 낮은 질의 데이터’로 2022년 글자료에서 나타난 심각한 논제 이탈, 심각한 표절, 글자 수 기준 800자 미만, 1,200자 이상의 글이 동일하게 발견되었고, 이 외에 학술 논문의 장르 양식을 따르는 등 글의 장르성이 다른 글이 발견되어 정제 대상으로 포함하였다.

‘채점에 영향을 줄 수 있는 불필요한 요소’로 2022년 연구에서 발견된 제목, 목차 제시 외에 인사말, 자기소개, 글쓰기 상황 소개, 주제문 제시 등이 포함되었다. 글쓰기 원자료 수집이 대학의 글쓰기 강의를 기반으로 이루어짐으로써, 강의를 위한 과제 수행의 맥락에서 포함될 수 있는 표현들이 나타난 것이다. 본 과업에서 수집하는 글쓰기 원자료는 기본적으로 진단 평가 상황을 전제하기에 채점 일관성을 높이기 위하여 수동 정제 과정에서 이러한 불필요한 요소들은 직접 삭제하도록 설정하였다.

글자 수를 충족해야 글쓰기 참여 사례비가 지급되기에, 2023년 수집 글쓰기 자료에서는 분량을 채우기 위해 글 후반부에 무의미한 기호나 문장을 넣어 분량을 채운 글도 일부 있었다. 이 경우 해당 부분을 삭제한 후 글자 수를 충족 여부를 확인하고 글자 수 기준 미충족 시 데이터 세트에서 제외하였다.

마지막으로 ‘개인정보’에 해당하는 이름, 학교명, 거주지 등에 대한 기술이 포함되어 있어 비식별화하는 기준을 수립하였다. 글쓰기 참여자가 글의 도입부에 자기를 소개하거나 글을 쓰는 배경을 설명하면서 이러한 개인 식별 정보가 일부 포함되게 된 것이다.

인공지능 학습 모델이 인간의 비윤리 또는 편견을 학습하지 않고 사회적 윤리를 준수할 수 있도록 비윤리적 내용이나 혐오나 사회적 갈등을 자극하는 내용 또한 채점 단계에서 발견하여 비식별화할 필요가 있다. 민감 정보¹²⁾와 차별과 혐오 등 비윤리적이라고 할 수 있는 내용은 2023년에 수집된 글쓰기 자료에서는 발견되지 않았으나, 이후 자료에서 나타날 것을 대비하여 기준에는 포함하였다.

2023년 글쓰기 자료에서는 단락 간 여백이 2줄 이상인 글, 들여쓰기가 없거나 불규칙한 글이 다수 발견되었다. 문서 작성 프로그램을 활용한 글쓰기 상황에서 글쓰기 참여자는 장르 관습에 대한 이해를 바탕으로 형식적 요소인 단락 구분, 띄어쓰기 간격 등을 규범에 맞게 작성하는 것이 중요하다. 따라서 단락 간 간격이나 띄어쓰기 간격 등 자체도 평가의 대상이 될 수 있지만, 대규모 채점 상황에서 9개의 채점 준거를 활용할 때 이러한 글의 형식적 요소가 내용 준거의 채점에 큰 영향을 주는 것은 바람직하지 않다. 채점자가 불필요하게 형식적 관습에 벗어난 것에 주의 집중하게 함으로써 채점의 대상이 되는 내용 정보에 대한 집중도를 떨어뜨리고, 글 내용에 대한 부정적인 인상을 남겨 채점에 영향을 미칠 수 있다. 이렇게 편집 양식의 차이가 실제 채점에 미치는 영향이 크므로 글쓰기 자료의 형식을 ‘단락 간 1줄 간격, 단락 들여쓰기 1칸, 자간 간격 1자라는 기준’을 적용하여 글 형식을 전체적으로 통일할 필요가 있었다.

이에 최종적으로 [그림 VII-3]에서와 같이 유형별로 데이터 정제 방안을 수립하였다.

12) 민감 정보는 2011년 3월 「개인정보 보호법」 및 그 시행령이 제정된 이래 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활에 관한 정보, 유전자 검사 등의 결과로 얻어진 유전정보, 「형의 실효 등에 관한 법률」 제2조 제5호에 따른 범죄경력자료에 해당하는 정보로만 한정되어 왔으나, 최근(2020년 8월) 개정된 개인정보 보호법 시행령(이하 ‘시행령’)에서 ‘개인의 신체적, 생리적, 행동적 특징에 관한 정보로서 특정 개인을 알아볼 목적으로 일정한 기술적 수단을 통해 생성한 정보’, ‘인종이나 민족에 관한 정보’가 추가되었다.

1) 데이터 정제 기준 설정		2) 데이터 정제 항목과 처리 방안	
정제 유형	정제 기준	정제 세부 항목	처리 방안
1. 부적합 데이터 선별	내용 요건 미충족	심각한 논제 이탈 심각한 표절 과제 장르 이탈	데이터 세트에서 제외
	형식 요건 미충족	글자 수 기준 미만/초과	
2. 불필요 항목 삭제	채점에 영향을 줄 수 있는 형식적 요소	제목, 주제문 포함 목차 등 구성 표지 인사말, 글쓰기 상황 표현	해당 부분 삭제
	분량 채우기 위한 무의미 표현	글 전개와 무관한 내용 글 전개와 무관한 기호	
3. 민감 정보 삭제	개인 정보	개인 식별 정보	해당 부분 비식별화
	비윤리적 내용	혐오, 차별 조장 내용	해당 부분 삭제
4. 글 형식 편집	단락 형식 이상	단락 간 여백 넓음 들여쓰기 없음	1줄 간격 통일 들여쓰기 1칸 통일
	글자 간 간격 이상	자간 지나치게 넓거나 모두 붙음	자간 간격 통일

[그림 VII-3] 데이터 정제 방안

데이터 정제 과정에서 인력 활용을 최소화하고, 글쓰기 원자료의 특성을 최대한 그대로 보존하는 방식의 자료 수집을 위해서는 글쓰기 원자료 제공자에게 정제의 대상이 되는 항목에 대해 사전에 정확히 알리는 것이 필요하다. 이에 글쓰기 원자료 제공자에게 개인을 식별할 수 있는 정보를 포함하지 않도록 안내하는 유의 사항을 글쓰기 단계 전에 확인하도록 글 자료 수집 플랫폼상에 안내 페이지를 개발하였다.

그럼에도 1학기에 수집한 글쓰기 원자료에서 내용 및 요건을 충족하지 못한 부적합 데이터, 분량을 채우기 위한 불필요 항목이 포함된 데이터, 개인 식별 정보를 포함한 데이터가 적지 않았기에 2학기에 진행할 글쓰기 자료 수집 과정에서는 정제 기준에 해당하는 글 자료가 생성되지 않도록 보다 적극적으로 안내하였다. 그 결과 1학기에 비해 데이터 손실을 낳는 부적합 데이터가 줄었고, 나머지 유형에서도 해당 수량을 줄일 수 있었다.

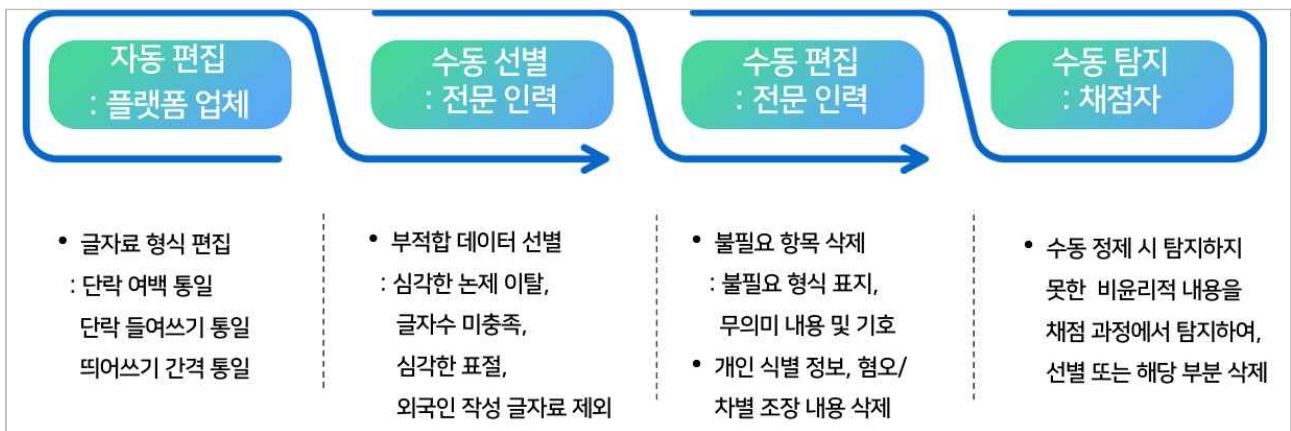
(2) 단계별 데이터 정제

9천 건 이상의 대규모 글쓰기 원자료를 수집하여 데이터화해야 하므로, 모든 정제 과정을 수동으로 해결 시 비용이 증가하고, 오류도 피하기 어렵다. 따라서 정제 유형과 기준, 요구되는 기능에 따라 시간과 비용, 기술 등을 고려하여 부분적으로 데이터 자동 정제 도구를 개발하여 활용하는 것이 필요하다.

본 연구에서는 글자 수를 자동으로 탐지하는 프로그램을 통해 분량 미달의 글을 데이

터 세트에서 제외할 수 있게 하였다. 나아가 글 형식 편집과 관련된 항목만 프로그램 개발을 통해 자동 정제의 대상으로 설정하였다. 내용이 부적합한 데이터를 선별하거나 불필요한 내용을 삭제하는 정제 작업은 모두 글쓰기 과제에 대한 이해도가 높은 연구진과 전문 인력이 수동으로 직접 수행하는 방식을 취하였다. 탐지 기능을 갖춘 프로그램을 개발하는 데 오랜 시간이 걸리고, 그 정확성 또한 보장하기 어렵기 때문이다.

이에 데이터 정제는 글 자료의 형식을 자동으로 편집한 뒤, 편집된 글 자료에서 수동으로 부적합 데이터를 선별하고, 불필요 항목을 삭제하는 단계로 진행되었다. 마지막으로 혹 정제 전문 인력이 탐지하지 못한 글 본문상의 비윤리적인 내용에 대해 채점자가 탐지하는 단계를 포함하여 정제 모델을 수립하였다. 데이터 정제는 [그림 VII-4]에서와 같이 단계를 구분하고, 전문성 영역별로 그 역할을 나누었다.



[그림 VII-4] 데이터 정제 단계별 수행 주체와 정제 항목

○ 1단계 - 자동 정제: 글쓰기 원자료 형식 편집

플랫폼 업체인 엔에스데블에서 플랫폼 내 데이터베이스화 과정에서 단락 간격, 글자/단어/어절/문장 간 간격 등 글의 형식적 요소를 자동으로 편집할 수 있는 정제 프로그램을 개발하고 적용하였다. 이후 연구진에서 결과를 확인하고, 오류를 피드백함으로써 자동 정제가 정확하게, 일관성 있게 이루어지도록 하였다.

글자 및 단어 간 간격 편집은 플랫폼상에서는 800자 미충족 시 글 제출이 불가능하므로 분량을 채우기 위해 아래 제시한 예와 같이 글 중간에 띄어쓰기를 여러 차례 한 경우에 적용되었다.

예)
혐오 표현은 법적으로 적정한 규제가 필요한 행위 양식에 속한다.
그러나 혐오 표현이 자유로운 표현의 지위를 갖기 위해서는
다른 행위자의 자유를 침해하지 않아야 한다는 원칙이 지켜져야 한다.
 (뒤 생략)

대학교의 글쓰기 강좌를 기반으로 대규모 글쓰기 자료 수집 과정에서 플랫폼에 접속하여 직접 글을 쓰는 방식 외에 참여자가 사전에 외부 문서 편집기로 작성한 글을 복사하는 것을 허용함으로써, 어절 및 문장 간 간격 이상 문제가 발생하였다. 이러한 문제는 주로 글쓰기 참여자가 사전에 외부 문서 프로그램으로 작성한 글을 복사하여 플랫폼 화면에 붙여 넣는 과정에서 문서 프로그램 간 호환이 적절하게 이루어지지 않는 데 기인한다. [그림 VII-5]의 왼쪽 ‘원본’ 창에서와 같이 단락의 시작 부분이 달라지는 등 실제 수행 결과와 다른 형태로 플랫폼상에 기입이 되는 문제가 발생한 것이다.

글 검사

다음글 >

ID	K0191	소속	충북대
마지막 저장 시간	2023-05-03 15:25:40	단과대학(계열)	기타계열
문항 번호	Q3	상태	검수 대기
주제	고령 운전자 면허 제한에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오.		

원본 - 100% +

만으로 60세가 지나고 부터는 면허를 일시정지한 후 여러가지 검사를 통해 지금까지 운전을 꾸준히 해왔는지 검사하고

면허시험을 다시 봐야 면허를 사용할 수 있게 해야한다. 수 많은 교통사고 사례를 보면 고령 운전자가 상당히 많다.

이를 봤을 때 나이가 들면 들수록 시력적인 부분이나 청각적인 부분, 순간의 반응속도가 노화되어 사고가 날 위험이 크고

고령운전자들로 인해 사고가 나면 부딪힌 다른 차량과 그 차량안에 타고 있는 통승자들도 위험에 빠지게 될 수 있기 때문에

도로 위에서는 이런 것들이 엄격해야 한다고 생각한다. 교통사고는 사고를 낸 운전자 한 사람만 다치는 것이 아니기 때문에 이러한 부분에서는 나라에서 엄격한 제도를 만들어 교통사고의 위험을 덜어내야 한다고 생각한다. 나이가 차면 면허를 정지 시켜 놓고 그 다음 여러가지 검사를 통해 운전경력이 몇년인지 그 나이까지 꾸준히 운전을 해왔는지 도로에 대한 이해가 어느정도인지를 검사하고 파악한 후에 점수를 매겨 면허를 취득할 수 있도록 하는 법이 필요하다고 생각한다.

824 / 1200

정제본 - 100% +

만으로 60세가 지나고 부터는 면허를 일시정지한 후 여러가지 검사를 통해 지금까지 운전을 꾸준히 해왔는지 검사하고

면허시험을 다시 봐야 면허를 사용할 수 있게 해야한다. 수 많은 교통사고 사례를 보면 고령 운전자가 상당히 많다.

이를 봤을 때 나이가 들면 들수록 시력적인 부분이나 청각적인 부분, 순간의 반응속도가 노화되어 사고가 날 위험이 크고

고령운전자들로 인해 사고가 나면 부딪힌 다른 차량과 그 차량안에 타고 있는 통승자들도 위험에 빠지게 될 수 있기 때문에

도로 위에서는 이런 것들이 엄격해야 한다고 생각한다. 교통사고는 사고를 낸 운전자 한 사람만 다치는 것이 아니기 때문에 이러한 부분에서는 나라에서 엄격한 제도를 만들어 교통사고의 위험을 덜어내야 한다고 생각한다. 나이가 차면 면허를 정지 시켜 놓고 그 다음 여러가지 검사를 통해 운전경력이 몇년인지 그 나이까지 꾸준히 운전을 해왔는지 도로에 대한 이해가 어느정도인지를 검사하고 파악한 후에 점수를 매겨 면허를 취득할 수 있도록 하는 법이 필요하다고 생각한다.

665 / 1200

☰ 목록으로 돌아가기

[그림 VII-5] 데이터 자동 편집 결과 글자 수 변경 사례

그런데 이렇게 글자/단어/어절/문장 간 간격 이상 문제를 해소하기 위해 자동 편집이 이루어지는 프로그램을 개발하여 적용하였을 때, [그림 VII-5]의 오른쪽 ‘정제본’ 창에서 확인할 수 있듯이 정제된 글쓰기 자료의 글자 수는 글쓰기 원자료상의 글자 수보다 줄어 들게 된다. 글쓰기 자료 형식 자동 편집을 적용하면 불필요한 띄어쓰기는 삭제되기 때문이다.

그런데 [그림 VII-5]의 오른쪽 ‘정제본’과 같이 자동 편집 이후 최종 글쓰기 자료 분량이 800자 미만이 되면 채점을 위한 원천 데이터로 사용할 수 없게 된다. 이에 [그림 VII-5]에서 글쓰기 자료 K0191이나 [그림 VII-6]의 글쓰기 자료 E0547과 같이 자동 편집 후 각각 665와 493자로 줄어든 경우 데이터 세트에서 제외하였다. 그러나 [그림 VII-6]에서 알 수 있듯이 형식 자동 편집 후 글자 수가 800자 기준에 미치지 못하는 자료가 적지 않게 발생하였기에, 이에 기준을 750자 이상으로 기준을 완화하여 채점을 위한 원천 데이터를 구축하였다.

ID	Last Updated	Code	Original answer	No. of character for Original answer	Refined Answer	No. of character for Refined answer	Progress
H0754	2023-05-16 16:46	Q2	인터넷이 발달	800	인터넷이 발달한	798	검수 대기
H0591	2023-05-16 18:19	Q3	여러분 유튜브	1290	여러분 유튜브 많	1288	검수 대기
E0547	2023-05-12 16:13	Q1		988		493	검수 대기
G0784	2023-05-12 2:16	Q6	블라인드 채용C	801	블라인드 채용이	799	검수 대기
G0704	2023-05-12 11:35	Q4	디지털 세계에서	800	디지털 세계에서	787	검수 대기
H0408	2023-05-12 1:40	Q3	최근 고령자의	800	최근 고령자의 디	796	검수 대기
H0717	2023-05-15 15:02	Q2	혐오 표현은	803	혐오 표현은 표	797	검수 대기
E0974	2023-05-16 11:06	Q3	고령 운전자 면	800	고령 운전자 면허	796	검수 대기
H0385	2023-05-16 19:53	Q6	블라인드 채용은	802	블라인드 채용은	798	검수 대기
E0988	2023-05-16 14:20	Q5	조력존엄사는 형	805	조력존엄사는 형	788	검수 대기

[그림 VII-6] 자동 정제 후 글자 수 변경 사례

○ 2단계 - 수동 정제: 부적합 데이터 선별

정제 2단계는 부적합 데이터를 선별하는 단계이다. 심각한 표절이나 논제 이탈이 있는 글을 학습 데이터로서 가치가 매우 낮다고 판단하여 최종 데이터 세트에서 제외하였다. 플랫폼상의 데이터를 관리하는 연구 인력과 외부의 정제 전문 인력을 활용하여 자동으로 산출된 글자 수를 확인하고, 나아가 개별적으로 글쓰기 자료의 내용을 정확하게 검토하였다.

① 논제 이탈 사례

[그림 VII-7]은 글쓰기 문항에 없는 ‘피의사실 공표’에 대한 글을 작성한 사례로 논제 이탈로 보아 원천 데이터 세트에 포함하지 않았다.

원본	- 100% +	정제본	- 100% +
<p>피의사실 공표 풍조 방지, 이대로 괜찮은가?</p> <p>우리나라에서는 현재 피의사실 공표죄를 형법 제126조에서 규정하고 있다. 피의사실 공표죄란, 검찰, 경찰, 기타 범죄 수사에 관한 직무를 행하는 자 또는 이를 감독하거나 보조하는 자가 수사 과정에서 알게 된 피의사실을 기소 전에 공표한 경우 성립하는 죄를 말한다. 아직 기소가 되지 않았거나 유죄 판결을 받지 않은 자의 조사받은 사실이나 그 내용을 유출했을 때의 벌금형 규정이 존재하지 않고 있어서 더욱 엄중한 사안이 된다. 예를 들어 죄질이 동일한 범죄자 두 명이 있다고 가정하자. 만약 이 범죄자 중 한 명의 사건만이 언론에 보도될 경우, 해당 언론 및 여론이 죄질과 형량을 판결하는 판사의 선택에 영향을 끼칠 수 있다. 또한 피의자의 기본권 침해도 문제로 제기될 수 있기에 이 같은 상황을 방지하기 위해 피의사실 공표죄를 법으로 제정한 것이다. 그럼에도 지난 수십년간 피의사실 공표죄에 대한 처벌은 ‘국민의 알 권리’를 명분으로 이행되지 않아 왔다. 지난 5월 법무부 산하 과거사위원회에 따르면 2008~18년 피의사실 공표죄로 접수된 사건은 총 347건에 이르지만 기소된 사례는 한 건도 없었다. 피의사실 공표가 사문화되고 있다는 의미이다. 형법에서 피의사실 공표죄를 죄로 분명명시하고 있으나, 이 범죄의 구성요건을 충족하는 그동안의 많은 행위가 그동안 많았지만 제대로 처벌되지 않고 있는 것이다. 피의사실 공표죄를 둘러싼 논란이 끊이지 않고 있는 것도 이 때문이라 하겠다. 한편 이 죄로 처벌된 전례가 없는 이유는 아마도 죄를 저지르는 주체도 검찰이고, 죄를 엄단하는 주체도 검찰이기 때문일 것이다. 방치되기 쉬운 시스템이다. 또한 피의사실은 수사기관이 국민으로부터 위임받은 권력을 행사해 취득한 정보이다. ‘공공의 자산’에 해당해 알 권리의 대상이 되기 때문에 기본권과 충돌하는 것도 처벌 전례가 없는 이유 중 하나일 테다. 그렇다고 완전히 폐지하는 것도 쉽지 않다. 결국 이러한 사실들을 종합했을 때 국민의 알 권리와 피의자의 기본권이 둘 다 보장될 수 있는 절충안을 모색해내는 것이 필요해 보인다. 검찰이 죄를 짓고 검찰이 처단하기 때문에 처벌을 방지하게 되는 것을 막기 위한 대책도 수립해야 할 듯하다. 우리나라의 국가권력 작용 구조상 어쩔 수 없어 보이는 것이 현실이지만, 그렇다고 손 놓고 있을 순 없을 터이니..</p>	1167 / 1200	<p>수 있다. 또한 피의자의 기본권 침해도 문제로 제기될 수 있기에 이 같은 상황을 방지하기 위해 피의사실 공표죄를 법으로 제정한 것이다.</p> <p>그럼에도 지난 수십년간 피의사실 공표죄에 대한 처벌은 ‘국민의 알 권리’를 명분으로 이행되지 않아 왔다. 지난 5월 법무부 산하 과거사위원회에 따르면 2008~18년 피의사실 공표죄로 접수된 사건은 총 347건에 이르지만 기소된 사례는 한 건도 없었다. 피의사실 공표가 사문화되고 있다는 의미이다. 형법에서 피의사실 공표죄를 죄로 분명명시하고 있으나, 이 범죄의 구성요건을 충족하는 그동안의 많은 행위가 그동안 많았지만 제대로 처벌되지 않고 있는 것이다. 피의사실 공표죄를 둘러싼 논란이 끊이지 않고 있는 것도 이 때문이라 하겠다. 한편 이 죄로 처벌된 전례가 없는 이유는 아마도 죄를 저지르는 주체도 검찰이고, 죄를 엄단하는 주체도 검찰이기 때문일 것이다. 방치되기 쉬운 시스템이다. 또한 피의사실은 수사기관이 국민으로부터 위임받은 권력을 행사해 취득한 정보이다. ‘공공의 자산’에 해당해 알 권리의 대상이 되기 때문에 기본권과 충돌하는 것도 처벌 전례가 없는 이유 중 하나일 테다. 그렇다고 완전히 폐지하는 것도 쉽지 않다. 결국 이러한 사실들을 종합했을 때 국민의 알 권리와 피의자의 기본권이 둘 다 보장될 수 있는 절충안을 모색해내는 것이 필요해 보인다. 검찰이 죄를 짓고 검찰이 처단하기 때문에 처벌을 방지하게 되는 것을 막기 위한 대책도 수립해야 할 듯하다. 우리나라의 국가권력 작용 구조상 어쩔 수 없어 보이는 것이 현실이지만, 그렇다고 손 놓고 있을 순 없을 터이니..</p>	1166 / 1200

[그림 VII-7] 논제 이탈 글쓰기 자료 사례

② 외부 자료를 활용한 심각한 표절 의심 사례

본 과업에서는 글 작성 외 목적의 디지털 도구 이용이 통제되는 환경에서 글쓰기를 수행하는 것을 원칙으로 하지만, 개별 과제 부여 등의 방식으로 작성된 자료의 경우 표절 가능성이 잠재되어 있다. 심각한 표절이 의심되는 경우는 ‘카피킬러’와 같은 표절 검증 프로그램을 활용하여 표절 비율을 확인하고, 쓰기 윤리 위반 규정에 따라 해당 글쓰기 자료는 원천 데이터 세트에서 제외하고, 글쓰기 참여 사례비 지급 대상에서도 제외하였다. ‘로봇세’에 대한 다음의 글은 전체 문장에서 표절률이 높게 나타난 사례이다.

<p>문장표절률 : 56%</p> <p>더불어 로봇세 도입의 사회적 배경을 함께 파악해보고자 한다. 로봇세란 로봇의 노동으로 생산하는 경제적 가치에 부과하는 세금을 의미한다.</p> <p>문장표절률 : 93%</p> <p>2016년 세계적인 대결 이전부터 인간과 기계 사이의 게임 대결에서 이미 기계의 우위가 확립되어가고 있었다.</p> <p>문장표절률 : 33%</p> <p>예를 들면 1997년 IBM의 게임 소프트웨어인 Deep blue와 체스 세계 챔피언 Garry Kasparov 간의 체스 시합, 2011년 미국 퀴즈 프로그램에서 IBM의 슈퍼컴퓨터와 최고 성적을 거뒀던 2명의 출연자의</p>
--

퀴즈 대결이 있다.

문장표절률 : 53%

이와 관련하여 해외에서는 로봇세에 관한 연구가 진행되어 왔다. Gerreiro et al. (2017)은 로봇세는 비 전문직 노동자의 임금을 올리는데 도움이 됨으로써 정부가 소득 불평등도 줄이는데 기여하고, 비 전문직과 전문직으로 이루어진 노동시장에서 로봇세는 부분적인 자동화가 이루어질 때 가장 최적 상태임을 밝혔다.

문장표절률 : 50%

결과로 로봇 사용에 대한 세율이 30%까지 올라가게 되면 전면적인 자동화는 발생하지 않았으며, 따라서 비 전문직 노동자는 일자리를 잃지 않게 되었다.

문장표절률 : 73%

그러나 이러한 해법은 효율성이나 재분배 측면에서 낮은 성과를 보였다. 이후 Heathcote et al.(2014)의 모델을 약간 수정하여 조세 제도가 일괄적인 환급을 통하여 노동자들에게 기본소득을 제공하는 경우 환급이 없는 전 연구보다 효율성과 재분배 측면에서 훨씬 나은 결과를 얻을 수 있었다.

문장표절률 : 33%

따라서 이같은 로봇세 도입의 가능성을 파악하는 동시에 노동력만을 생산 요소로 가지고 있는 계층이 일자리를 잃게 되었을 때 로봇세 도입 필요성에 관해 확인할 수 있을 것으로 기대한다.

③ 수집된 글쓰기 자료 간 표절 사례

본 연구에서는 한 글쓰기 참여자가 둘 이상의 다른 문항에 대해 글쓰기 원자료를 제출하는 것을 허용하고 있다. 그런데 이 경우 플랫폼상에서 문항 번호만 다르게 선택하고, 실제 글은 이전에 제출한 글과 동일한 내용을 복사하여 붙여 넣었을 때 연구진이 이러한 자료에 대해 자신의 글에 대한 표절인지 판단하기가 어렵다. 이전에 작성한 글을 개별적으로 확인하지 않는 한, 표절하여 작성한 글에 대해서 문항 번호를 잘못 선택한 실수로만 볼 수 있기 때문이다. 또 동료 학습자의 글쓰기 자료를 받아서 표절하여 제출하는 경우에도 수동으로 검토하는 과정에서 표절 여부를 판단하는 것은 거의 불가능하다.

구축한 데이터 간 중복을 방지하고 표절 글쓰기 자료를 선별하기 위해, 플랫폼 업체에 요청하여 플랫폼 내에서 제출된 글의 유사도를 판별할 수 있는 프로그램을 개발하고 적용하였다. [그림 VII-8]의 오른쪽을 보면 유사한 내용을 쓴 글 자료의 번호와 유사도(%)가 자동으로 제시되고 있는 것을 알 수 있다.

유사도 검토는 플랫폼상에서 자동으로 이루어지지만, 실제 데이터 세트에서 제외할 자

료를 판정하는 단계는 수동으로 이루어져야 한다. 판별 결과 유사도가 50% 이상인 글은 상호 표절 글로 간주하고, 작성 시기를 고려하여 후행 작성 글은 데이터에서 삭제하였다. 유사도가 30-50% 사이의 글은 유사하다고 판별된 글 사이의 유사도를 직접 글을 읽어 표절 가능성을 점검하여 삭제 여부를 판단하였다.


	A	B	C	D	E	F	G	H
1	No.	ID	Refined Answer Edited	Edited Refined Answer	character for Refined answer	Progress	Similar Answer To	Similarity Percentage
2	5	D0841	2023-10-18 16:00	최근, 스타쉽 엔터테인먼트에서 C	822	검수 완료	F0311	14.24
3	22	B0830	2023-10-18 16:00	저는 21세기 대한민국에서 교육	1216	검수 완료	K0058	11.37
4	24	A2038	2023-11-16 16:19	최근 인공지능의 빠른 발달로 cha	891	검수 완료	K1172	15.75
5	38	E1313	2023-11-16 16:19	현대 사회에서는 디지털 미디어를	1054	검수 완료	B1441	13.64
6	46	D1492	2023-10-18 16:00	현대의 AI기술은 딥러닝을 바탕의	917	검수 완료	A1974	12.73
7	53	K0989	2023-10-18 16:00	인공지능(AI) 기술의 발전으로 우	1213	검수 완료	D1005	13.49
8	62	D0965	2023-10-18 16:00	디지털 미디어에서의 잊힐 권리는	803	검수 완료	D0667	16.15
9	68	A1978	2023-10-18 16:01	AI를 활용하여 창작물을 만든 사	1043	검수 완료	D1066	15.38
10	73	H1324	2023-11-16 16:19	창작물이란 독창적으로 지어낸 여	958	검수 완료	D1358	17.33
11	100	D1400	2023-10-18 16:00	디지털 미디어가 발달되지 않았던	808	검수 완료	G0567	13.57
12	142	K0917	2023-10-18 16:00	AI 기술이 발전함에 따라 최근부터	1183	검수 완료	K0942	11.71
13	145	D0894	2023-10-18 16:00	인공지능(AI) 기술이 발전함에 따	1040	검수 완료	K0944	13.63
14	156	B0898	2023-10-18 16:00	하루건 이틀이건, 병상에 누워본	1202	검수 완료	B0710	11.37
15	160	H1389	2023-11-16 16:19	오늘날 AI와 같은 기술의 발전으로	940	검수 완료	K0795	16.95
16	172	A1629	2023-10-18 16:00	동물 실험이란 교육, 실험, 연구 및	824	검수 완료	D1349	30.93
17	173	H1572	2023-11-16 16:19	잘못된 행위를 모르고 하는 것과	941	검수 완료	D0650	12.97
18	206	F0845	2023-11-16 16:19	동물 실험은 고대 그리스부터 사	991	검수 완료	D0381	11.97
19	211	D1145	2023-10-18 16:00	AI가 만들어내는 창작물은 궁극적	836	검수 완료	D1495	12.84
20	217	D1476	2023-10-18 16:00	블라인드 채용은 공정하게 사람들	973	검수 완료	H0268	14.33
21	218	H1566	2023-11-16 16:19	요즘 텔레비전을 보다 보면 미성	1207	검수 완료	D1333	11.16
22	233	B0672	2023-10-18 16:00	디지털 세계 속에 사는 우리는 많	936	검수 완료	D1568	16.66

[그림 VII-8] 데이터 유사도 검사 결과 예시

④ 문서 생성 인공지능 활용 사례

플랫폼상에서 글을 쓰기 전에 윤리 서약을 하도록 했음에도 불구하고 챗지피티(chatGPT) 등 문서 생성 인공지능을 활용한 표절 사례가 확인되었다. [그림 VII-9]는 문서 생성 인공지능을 활용한 글로 의심되는 사례를 인공지능 생성 문서 표절 검사기를 통해 표절로 적발한 사례이다.

****Please input more text for a more accurate result***
Your Text is AI/GPT Generated



100.01%
AI GPT*

축법소년의 연령 제한에 대해 반대하는 입장은 사회적 관점, 인권 및 범죄 예방에 관한 여러 이유로 나타
 납니다. 아래에서는 축법소년의 연령 제한에 반대하는 몇 가지 주요 이유와 그 이유를 뒷받침하는 논증
 을 제시하겠습니다.

인권과 공평성:

연령 제한은 미성년자와 성인 사이의 법적 구분을 나타내며, 이것이 종종 미성년자에게 보다 관대한 처
 우를 제공하거나 성인과 비교했을 때 미성년자에게 보다 가벼운 형벌을 적용하는 방식으로 나타납니다.
 그러나 이러한 접근은 미성년자와 성인 간의 범죄 행위의 중요한 차이를 간과할 수 있습니다. 범죄 행위
 의 심각성은 나이에 따라 달라질 수 있으므로, 이러한 경우에 연령 제한을 통한 법적 구분은 불합리할 수
 있습니다.

범죄 예방 및 재발 방지:

범죄 예방 측면에서 연령 제한은 효과적이지 않을 수 있습니다. 미성년자가 범죄를 저지를 가능성이 있
 는 경우, 그들을 재활(rehabilitation)하는 대신에 범죄 형벌을 부과하는 것은 그들을 범죄 세계에 끌어들이 수 있
 습니다. 범죄 예방을 위해 교육, 상담 및 사회 지원 프로그램에 투자하는 것이 미성년자의 긍정적인 재사
 회화를 촉진하는 데 더 효과적일 수 있습니다.

범죄 행위는 맥락에 따라 다양한 이유로 발생할 수 있습니다. 미성년자가 범죄 행위를 저지를 때, 그들
 의 가정 환경, 사회적 상황 및 미성년자 자신의 생활 상황을 고려하는 것이 중요합니다. 연령 제한을 통
 한 법적 구분은 종종 이러한 맥락을 간과하고 있으며, 미성년자에게 특별한 상황을 고려한 대안을 제공
 하지 않을 수 있습니다.

미성년자는 아직 교육과 성장의 단계에 있는 경우가 많습니다. 범죄 행위가 이들의 미래에 치명적인 영
 향을 미칠 수 있으므로, 교육 및 재사회화가 무시되어서는 안 됩니다. 연령 제한이 너무 엄격하게 시행되
 면, 미성년자의 기회와 성장 가능성을 제한할 수 있습니다.

범죄 행위는 각각 다르며, 미성년자 사이에서도 범죄의 성격과 심각성이 다를 수 있습니다. 법적 구분
 대신 각 사례를 개별적으로 판단하고 사건의 맥락, 범행자의 성격 및 동기, 피해자의 상황 등을 고려해야
 합니다.

국가는 모든 시민의 인권과 복지를 보호하고 증진하는 책임이 있습니다. 미성년자가 범죄 행위를 저지
 른 경우, 국가는 그들에게 교육, 상담 및 사회 지원을 제공하여 범죄 예방 및 재사회화를 촉진하는 데 노
 려해야 합니다.

이러한 변화에는 교육, 사회 지원 및 범죄 예방 프로그램에 대한 투자와 함께 신중한 고려와 감독이 필
 요합니다.

■ Highlighted text is suspected to be most likely generated by AI*
 1,224 Characters

[그림 VII-9] 지피티(chatGPT) 표절 사례

② 제목, 인사말, 목차, 자기소개 등 삭제 및 비식별화

본 연구에서 설계한 논증적 글쓰기 과제는 제목을 포함하지 않도록 하고 있고, 인사말, 목차, 자기소개 등은 장르 관습에서 완전히 벗어나기 때문에 삭제하였다. 이 밖에 논증 전개 과정에서도 이름이나 학교명 등 개인 식별 정보를 포함하는 사례가 다수 발견되어 수동으로 비식별화 처리를 하였다.

예)

제목: 혐오 표현의 법적 규제의 필요성

안녕하세요, 저는 00대학교의 00과 000입니다. 목차는 다음과 같습니다.

1. 서론
2. 본론
3. 결론

1. 서론

혐오 표현은 법적으로 적정한 규제가 필요한 행위 양식에 속한다. 이러한 혐오 표현의 규제 문제를 다루는 데에는 온라인 공간과 오프라인 공간의 문제를 분리하여 논의하는 것이 현재 이 문제가 속해 있는 논의의 지형을 세밀하게 파악하는 데 도움이 될 것이다. (뒤 생략)

○ 4단계 - 수동 정제: 비윤리적 내용 탐지

3단계에서 인공지능 학습 데이터 윤리성 기준을 심각하게 위반한 자료가 탐지되지는 않았지만 9천여 건이 넘는 글쓰기 자료를 모두 꼼꼼하게 읽는 것이 현실적으로 어렵기에, 채점 과정에서 3단계에서 미처 발견하지 못한 비윤리적 내용에 대해 탐지가 이루어지도록 채점자에게 안내하였다. 그 결과 일부 편향적인 시각을 담은 내용이 발견되었으나, 논리적 의견을 진술하는 맥락을 고려하였을 때 허용될 수 있는 정도였기에, 2023년도 수집된 글쓰기 자료 중 최종적으로 비윤리적이라고 판단하여 제외된 사례는 없었다.

3) 데이터 주석

(1) 데이터 주석 절차

글쓰기 자료 자동 채점 프로토타입을 개발하기 위해서는 정제된 글쓰기 자료와 채점

결과를 자동 채점을 위한 인공지능 학습용 제이슨(JSON) 파일로 가공해야 한다. 데이터 가공 단계에서는 원천 데이터에 대한 정교한 라벨링 작업보다는 데이터에 대한 메타 정보로서 데이터 세트 파일 정보, 쓰기 과제 정보, 글쓰기 자료 제공자 정보, 글쓰기 자료 정보를 부착하는 작업, 즉 데이터 주석(Data Annotation)이 핵심 과업이 된다. 데이터 주석 작업을 위한 세부 과업은 메타 데이터를 확정하는 시작 단계, 자동 주석을 지원하는 프로그램을 개발하는 단계, 제이슨 파일로 변환하여 학습 데이터를 구축하고 검수하는 최종 단계로 진행된다. 자동 주석 프로그램은 글쓰기 자료 채점 종료 후 플랫폼 업체와 협력하여 개발하였고, 제이슨 파일 변환 후에 각 데이터별로 주석이 정확하게 이루어졌는지 검수하였다.



[그림 VII-10] 데이터 주석 절차

(2) 데이터 주석 포맷

<표 VII-1>과 <표 VII-2>는 본 과업의 수행 결과 산출되는 제이슨 포맷에 포함되는 주석 정보 전체에 대한 구조이다. <표 VII-1>은 본 연구에서 수집한 글쓰기 자료 9,371건 전체에 대한 주석 정보로, 글쓰기 자료의 특성과 문항 정보, 글쓰기 자료 작성자에 대한 정보를 포함하고 있다. <표 VII-2>는 채점이 수행된 5,000건의 글쓰기 자료에 대한 주석 정보로, 글쓰기 자료에 대한 정보 외에 각 채점자별 점수와 최종 확정 점수, 세부 준거별 점수까지 포함하고 있다.

<표 VII-1> 글쓰기 자료 데이터 세트 제이슨 파일 구조

Level 1	Level 2	Level 3	Level 4	Type	Description	Note
id				str	* 말뭉치 파일 ID	GWRW2300#####0
metadata				obj		
	title			str	* 원시 말뭉치: 국립국어원 [말뭉치 유형 구분] [파일 ID]	국립국어원 글쓰기 말뭉치 GWRW2300#####0
	creator			str	* 생성자: 국립국어원	
	distributor			str	* 배포자: 국립국어원	
	year			str	* 말뭉치 구축 연도	
	category			str	* 비출판물>논술문	
	length _type			num	* 글자 수 범위: 750-1200	
	annotation _level			str	* 분석 층위: 원시	
	sampling			str	* 샘플링 방식: 참여자 모집 후 온라인 글쓰기 자료 수집	
document				arr (obj)	* 문서 정보	
	id			str	* 문서 ID * '원시말뭉치파일ID. 파일 내 문서번호'로구성	GWRW2300#####0.1
	metadata			obj	* 문서의 메타 정보	
		author		obj		
			author_id	str	* 작성자 ID	Writer ID
			author_age	num	* 출생연도 4자리, 모를 경우 "NA"	Birth date 4 digit
			author _occupation	str	* 대학생_(대학교 코드)	대학생_N대학
			author _grade	num	* 학년: 1-4	
			author_sex	str	* 성별: 남성/여성	
		submission		str	* 웹페이지	
		publisher		str	* 미출판, 출판사, 신문사,	미출판

Level 1	Level 2	Level 3	Level 4	Type	Description	Note
					방송사 등	
		date		str	* 수집 일시	형식:YYYYMMDD
		prompt		obj		
			prompt_type	str	* 과제	추후 과제와 배경, 과제와 제시문 등 추가
			prompt_num	str	* 문제 번호	Q1-Q9
			prompt_con	str	* 문제 내용	프롬프트 전체
		written_stat		obj	* 문서 통계 정보	
			written_length	num	* 개별 글의 글자 수	
			paragraph_num	str	* 문단 수	
			sentence_num	str	* 문장 수	
	paragraph			arr (obj)	* 문단	
		id		str	* 문단 ID	GWRW2300#####0.1.1
		form		str	* 정제된 형태	
		sentence		arr (obj)	* 문장	
			id	str	* 문장 ID	GWRW2300#####0.1.1
			form	str	* 정제된 형태	

<표 VII-2> 글쓰기 채점 자료 데이터 세트 제이슨 파일 구조

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
id					str	* 말뭉치 파일 ID	GWGR2300#####0
metadata					obj		
	title				str	* 원시 말뭉치: 국립국어원 [말뭉치 유형 구분] [파일 ID]	국립국어원 글쓰기 채점 말뭉치 GWGR2300#####0
	creator				str	* 생성자: 국립국어원	

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
	distributor				str	* 배포자: 국립국어원	
	year				num	* 말뭉치 구축 연도	
	category				str	* 비출판물>논술채점자료 --> 비출판물>논술 채점 자료	비출판물>논술 채점 자료
	length _type				num	* 글자 수 범위: 750-1200	
	annotation _level				str	* 분석 층위: 글쓰기 채점	
	sampling				str	* 샘플링 방식: 참여자 모집 후 온라인 글쓰기 자료 수집과 채점	
document					arr (obj)	* 문서 정보	
	id				str	* 문서 ID * '일시말뭉치파일ID. 파일 내 문서번호'로구성	GWGR2300#####0. 1
	metadata				obj	* 문서의 메타 정보	
		author			obj		
			author_id		str	* 작성자 ID	Writer ID
			author_age		num	* 출생연도 4자리, 모를 경우 "NA"	Birth date 4 digit
			author_occupation		str	* 대학생_(대학교 코드)	대학생_N대학
			author_grade		num	* 학년: 1-4	
			author_sex		str	* 성별: 남성/여성	
		submission			str	* 웹페이지	
		publisher			str	* 미출판, 출판사, 신문사, 방송사 등	
		date			str	* 수집 일시	형식:YYYYMMDD
		prompt			obj		
			prompt_type		str	* 과제	추후 과제와 배경, 과제와 제시문 등 추가

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
			prompt_num		str	* 문제 번호	Q1-Q9
			prompt_con		str	* 문제 내용	프롬프트 전체
		written_stat			obj	* 문서 통계 정보	
			written_length		num	* 개별 글의 글자 수	
			paragraph_num		str	* 문단 수	
			sentence_num		str	* 문장 수	
	paragraph				arr (obj)	* 문단	
		id			str	* 문단 ID	GWGR2300#####0.1.1
		form			str	* 정제된 형태	
		sentence			arr (obj)	* 문장	
			id		str	* 문장 ID	GWGR2300#####0.1.1.1
			form		str	* 정제된 형태	
	evaluation						
		evaluator			arr (obj)	* 채점자 정보	
			evaluator1_ID		str	* 채점자1_ID	
			evaluator2_ID		str	* 채점자2_ID	
			evaluatorN_ID		str	* 채점자N_ID	가교 답안 복수 채점 경우
			re-evaluator_ID		str	* 재채점자 ID	92건에 한정됨
			final_evaluator_ID		str	* 최종 확정 점수 부여자로 선택된 채점자 ID	1-N 사이에 있는 채점자 1인 1-N과 재채점자 중에서 1인
		evaluation_data			arr (obj)	* 평가 정보	

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
			eva_score_con		arr (obj)	* 내용 평가 정보	
				evaluator1_score_con1	num	채점자 1의 내용 총점	
				evaluator1_score_con1	num	채점자 1, 내용1의 점수	
				evaluator1_score_con2	num	채점자 1, 내용2의 점수	
				evaluator1_score_con3	num	채점자 1, 내용3의 점수	
				evaluator1_score_con4	num	채점자 1, 내용4의 점수	
				evaluator1_score_con5	num	채점자 1, 내용5의 점수	
				evaluator2_score_total_con	num	채점자 2의 내용 총점	
				evaluator2_score_con1	num	채점자 2, 내용1의 점수	
				evaluator2_score_con2	num	채점자 2, 내용2의 점수	
				evaluator2_score_con3	num	채점자 2, 내용3의 점수	
				evaluator2_score_con4	num	채점자 2, 내용4의 점수	
				evaluator2_score_con5	num	채점자 2, 내용5의 점수	
				evaluatorN_score_total_con	num	채점자 N의 내용 총점	가교 답안 복수 채점 경우
				evaluatorN_score_con1	num	채점자 N, 내용1의 점수	가교 답안 복수 채점 경우
				evaluatorN_score_con2	num	채점자 N, 내용2의 점수	가교 답안 복수 채점 경우
				evaluatorN_score_con3	num	채점자 N, 내용3의 점수	가교 답안 복수 채점 경우
				evaluatorN_score_con4	num	채점자 N, 내용4의 점수	가교 답안 복수 채점 경우
				evaluatorN_score_con5	num	채점자 N, 내용5의 점수	가교 답안 복수 채점 경우
				re-evaluator_score	num	재채점자의 내용 총점	

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
				ore_total_con			
				re-evaluator_score_con1	num	재채점자, 내용1의 점수	92건에 한정
				re-evaluator_score_con2	num	재채점자, 내용2의 점수	92건에 한정
				re-evaluator_score_con3	num	재채점자, 내용3의 점수	92건에 한정
				re-evaluator_score_con4	num	재채점자, 내용4의 점수	92건에 한정
				re-evaluator_score_con5	num	재채점자, 내용5의 점수	92건에 한정
			eva_score_org		arr (obj)	* 조직 평가 정보	
				evaluator1_score_total_org	num	채점자 1의 조직 총점	
				evaluator1_score_org1	num	채점자 1, 조직1의 점수	
				evaluator1_score_org2	num	채점자 1, 조직2의 점수	
				evaluator2_score_total_org	num	채점자 2의 조직 총점	
				evaluator2_score_org1	num	채점자 2, 조직1의 점수	
				evaluator2_score_org2	num	채점자 2, 조직2의 점수	
				evaluatorN_score_total_org	num	채점자 N의 조직 총점	가고 답안에 한정
				evaluatorN_score_org1	num	채점자 N, 조직1의 점수	
				evaluatorN_score_org2	num	채점자 N, 조직2의 점수	
				re-evaluator_score_total_org	num	재채점자의 조직 총점	92건에 한정
				re-evaluator	num	재채점자, 조직1의 점수	

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
				or_score_org1			
				re-evaluator_score_org2	num	재채점자, 조직2의 점수	
			eva_score_exp		arr (obj)	* 표현 평가 정보	
				evaluator1_score_total_exp	num	채점자 1, 표현 총점	
				evaluator1_score_exp1	num	채점자 1, 표현1 점수	
				evaluator1_score_exp2	num	채점자 1, 표현2 점수	
				evaluator2_score_total_exp	num	채점자 2, 표현 총점	
				evaluator2_score_exp1	num	채점자 2 표현1 점수	
				evaluator2_score_exp2	num	채점자 2, 표현2 점수	
				evaluatorN_score_total_exp	num	채점자 N의 표현 총점	가교 답안에 한정
				evaluatorN_score_exp1	num	채점자 N, 표현1 점수	
				evaluatorN_score_exp2	num	채점자 N, 표현2 점수	
				re-evaluator_score_total_exp	num	재채점자의 표현 총점	92건에 한정
				re-evaluator_score_exp1	num	재채점자, 표현1 점수	
				re-evaluator_score_exp2	num	재채점자, 표현2 점수	
			evaluator1_total_score		num	채점자 1 총점(9-45)	
			evaluator2_total_score		num	채점자 2 총점(9-45)	
			evaluatorN		num	채점자 N 총점(9-45)	가교 답안 복수 채점

Level 1	Level 2	Level 3	Level 4	Level 5	Type	Description	Note
			_total_score				경우
			re-evaluator_total_score		num	재채점자 총점(9-45)	

(3) 데이터 주석 예시

글쓰기 자료 데이터 주석의 예는 [그림 VII-11]에, 글쓰기 채점 자료 데이터 주석의 예는 [그림 VII-12]에 제시하였다.

```

{
  "id": "GWRW2300001050",
  "metadata": {
    "title": "국립국어원 글쓰기 말뭉치 GWRW2300001050",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": 2023,
    "category": "비출판물>논술문",
    "length_type": "750-1200",
    "annotation_level": "원시",
    "sampling": "참여자 모집 후 온라인 글쓰기 자료 수집"
  },
  "document": [
    {
      "id": "GWRW2300001050.1",
      "metadata": {
        "author": {
          "author_id": "F0005",
          "author_age": 2004,
          "author_occupation": "대학생_F대학",
          "author_grade": 1,
          "author_sex": "남성"
        },
        "submission": "웹페이지",
        "publisher": "미출판",
        "date": "20230419",
        "prompt": {

```

```

        "prompt_type": "과제",
        "prompt_num": "Q2",
        "prompt_con": "혐오 표현이란 정당화될 수 없는 혐오 감정에 근거하여 특정 집
단이나 개인에 대해 공개적으로 모욕, 비하, 멸시, 위협을 하거나 차별과 폭력을 선동하는 언어적, 비
언어적 행위를 말한다. 혐오 표현은 혐오의 대상이 되는 개인 또는 집단의 존엄성을 부정하고 사회 전
체에 미치는 해악이 크기 때문에 혐오 표현을 법적으로 규제해야 한다는 주장이 제기되고 있다. 혐오
표현의 법적 규제에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. [유의 사항] △ 서론, 본론,
결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간:
90분"
    },
    "written_stat": {
        "written_length": 1000,
        "paragraph_num": 5,
        "sentence_num": 13
    }
},
"paragraph": [
    {
        "id": "GWRW2300001050.1.1",
        "form": " 우리는 이 사회를 살아감에 있어, 한 번쯤 유명 연예인의 극단적 선택,
집단 내 사이버 폭력으로 인한 외상적 장애를 입은 이들의 소식을 접해본 적이 있을 것이다. 무엇이
도대체 그들을 사지로 내몬 것일까?",
        "sentence": [
            {
                "id": "GWRW2300001050.1.1.1",
                "form": " 우리는 이 사회를 살아감에 있어, 한 번쯤 유명 연예인의 극
단적 선택, 집단 내 사이버 폭력으로 인한 외상적 장애를 입은 이들의 소식을 접해본 적이 있을 것이
다."
            },
            {
                "id": "GWRW2300001050.1.1.2",
                "form": " 무엇이 도대체 그들을 사지로 내몬 것일까?"
            }
        ]
    }
},
{
    "id": "GWRW2300001050.1.2",
    "form": " 그 이유를 살펴보기 위해선, 먼저 우리는 한국 사회에서 나타나는 혐
오의 양상에 대해 살펴볼 필요가 있다. 과거, 우리나라에서 혐오는 주로 신분의 영역에서 상위계층이
하위계층을 무시하는 양상으로 주로 나타났고, 후에 시대적 변화를 겪으며, 지역 간의 혐오, 성별에 대
한 혐오 등의 다양한 양상으로 현대사회까지 이어져 오게 되었다. 특히 그중에서도 과거와 달리, 인터

```

넷이 발달함에 따라 지금의 현대사회에서는 ‘사이버 혐오’가 주로 나타나고 있다. 그렇다면 사이버 혐오란 무엇일까?”,

```

    "sentence": [
      {
        "id": "GWRW2300001050.1.2.1",
        "form": " 그 이유를 살펴보기 위해선, 먼저 우리는 한국 사회에서 나타나는 혐오의 양상에 대해 살펴볼 필요가 있다."
      },

```

```

      {
        "id": "GWRW2300001050.1.2.2",
        "form": " 과거, 우리나라에서 혐오는 주로 신분의 영역에서 상위계층이 하위계층을 무시하는 양상으로 주로 나타났고, 후에 시대적 변화를 겪으며, 지역 간의 혐오, 성별에 대한 혐오 등의 다양한 양상으로 현대사회까지 이어져 오게 되었다."
      },

```

```

      {
        "id": "GWRW2300001050.1.2.3",
        "form": " 특히 그중에서도 과거와 달리, 인터넷이 발달함에 따라 지금의 현대사회에서는 ‘사이버 혐오’가 주로 나타나고 있다."
      },

```

```

      {
        "id": "GWRW2300001050.1.2.4",
        "form": " 그렇다면 사이버 혐오란 무엇일까?"
      }
    ]
  },

```

```

  {
    "id": "GWRW2300001050.1.3",
    "form": " 사이버 혐오란 인터넷 등의 사이버 공간에서 이루어지는 타인을 비방하는 목적의 모든 혐오를 일컫는 말이다. 특히 이러한 사이버 혐오는 인터넷이 광범위하게 퍼지고 누구나 쉽게 접할 수 있게 되며 과거의 혐오와 달리, 앞서 소개한 유명 연예인의 극단적 선택 등의 양상을 띄며, 보다 확장된 형태를 취하고 있다. 그렇다면 우리는 이러한 사이버 혐오가 만연한 사회에서 어떠한 태도를 갖추어야 할까?",

```

```

    "sentence": [
      {
        "id": "GWRW2300001050.1.3.1",
        "form": " 사이버 혐오란 인터넷 등의 사이버 공간에서 이루어지는 타인을 비방하는 목적의 모든 혐오를 일컫는 말이다."
      },

```

```

      {
        "id": "GWRW2300001050.1.3.2",
        "form": " 특히 이러한 사이버 혐오는 인터넷이 광범위하게 퍼지고 누구

```

나 쉽게 접할 수 있게 되며 과거의 혐오와 달리, 앞서 소개한 유명 연예인의 극단적 선택 등의 양상을 띠며, 보다 확장된 형태를 취하고 있다."

```

        },
        {
            "id": "GWRW2300001050.1.3.3",
            "form": " 그렇다면 우리는 이러한 사이버 혐오가 만연한 사회에서 어떠한 태도를 갖추어야 할까?"
        }
    ]
},
{

```

```

            "id": "GWRW2300001050.1.4",
            "form": " 먼저, 우리는 사이버 공간상의 개개인도 하나의 인격체란 사실을 잊지 않아야 하며 또한 우리는 올바른 인터넷 생활을 위해 ‘미디어 리터러시’의 태도를 함양해야 한다. 여기서 미디어 리터러시란 정보 수용자가 미디어를 접하고 활용하는 과정에서 사용자 스스로가 비판적으로 정보를 수용 하하는 등의 윤리적인 태도를 갖추는 능력으로 현대인의 사이버 활동에 있어 필수적으로 갖추어야 할 덕목이다. ",

```

```

            "sentence": [
                {
                    "id": "GWRW2300001050.1.4.1",
                    "form": " 먼저, 우리는 사이버 공간상의 개개인도 하나의 인격체란 사실을 잊지 않아야 하며 또한 우리는 올바른 인터넷 생활을 위해 ‘미디어 리터러시’의 태도를 함양해야 한다."
                },
                {

```

```

                    "id": "GWRW2300001050.1.4.2",
                    "form": " 여기서 미디어 리터러시란 정보 수용자가 미디어를 접하고 활용하는 과정에서 사용자 스스로가 비판적으로 정보를 수용 하하는 등의 윤리적인 태도를 갖추는 능력으로 현대인의 사이버 활동에 있어 필수적으로 갖추어야 할 덕목이다."
                }
            ]
        },
        {

```

```

            "id": "GWRW2300001050.1.5",
            "form": " 이처럼 우리는 사회에서 일어나고 있는 다양한 혐오의 양상을 이해하는 태도를 가져야 하며 특히 현대사회에서 주로 나타나는 사이버 혐오에 대해 경각심을 가지고, 미디어 리터러시의 태도를 갖추어야 한다. 또한 사이버 상에서도 서로를 배려하는 태도를 가지며, 보다 깨끗하고 건전한, 혐오가 없는 사회를 만들기 위해 노력해야 할 것이다.",

```

```

            "sentence": [
                {
                    "id": "GWRW2300001050.1.5.1",

```



```

"publisher": "미출판",
"date": "20230419",
"prompt": {
  "prompt_type": "과제",
  "prompt_num": "Q2",
  "prompt_con": "혐오 표현이란 정당화될 수 없는 혐오 감정에 근거하여 특정 집
단이나 개인에 대해 공개적으로 모욕, 비하, 멸시, 위협을 하거나 차별과 폭력을 선동하는 언어적, 비
언어적 행위를 말한다. 혐오 표현은 혐오의 대상이 되는 개인 또는 집단의 존엄성을 부정하고 사회 전
체에 미치는 해악이 크기 때문에 혐오 표현을 법적으로 규제해야 한다는 주장이 제기되고 있다. 혐오
표현의 법적 규제에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. [유의 사항] △ 서론, 본론,
결론을 갖춘 완결된 글을 쓸 것(제목을 쓰지 말 것). △ 분량: 1,000자 내외(±200자, 공백 포함), 시간:
90분"
},
"written_stat": {
  "written_length": 1000,
  "paragraph_num": 5,
  "sentence_num": 13
}
},
"paragraph": [
  {
    "id": "GWGR2300001050.1.1",
    "form": " 우리는 이 사회를 살아감에 있어, 한 번쯤 유명 연예인의 극단적 선택,
집단 내 사이버 폭력으로 인한 외상적 장애를 입은 이들의 소식을 접해본 적이 있을 것이다. 무엇이
도대체 그들을 사지로 내몬 것일까?",
    "sentence": [
      {
        "id": "GWGR2300001050.1.1.1",
        "form": " 우리는 이 사회를 살아감에 있어, 한 번쯤 유명 연예인의 극
단적 선택, 집단 내 사이버 폭력으로 인한 외상적 장애를 입은 이들의 소식을 접해본 적이 있을 것이
다."
      },
      {
        "id": "GWGR2300001050.1.1.2",
        "form": " 무엇이 도대체 그들을 사지로 내몬 것일까?"
      }
    ]
  }
}
{
  "id": "GWGR2300001050.1.2",
  "form": " 그 이유를 살펴보기 위해선, 먼저 우리는 한국 사회에서 나타나는 혐

```

오의 양상에 대해 살펴볼 필요가 있다. 과거, 우리나라에서 혐오는 주로 신분의 영역에서 상위계층이 하위계층을 무시하는 양상으로 주로 나타났고, 후에 시대적 변화를 겪으며, 지역 간의 혐오, 성별에 대한 혐오 등의 다양한 양상으로 현대사회까지 이어져 오게 되었다. 특히 그중에서도 과거와 달리, 인터넷이 발달함에 따라 지금의 현대사회에서는 '사이버 혐오'가 주로 나타나고 있다. 그렇다면 사이버 혐오란 무엇일까?",

```

    "sentence": [
      {
        "id": "GWGR2300001050.1.2.1",
        "form": " 그 이유를 살펴보기 위해선, 먼저 우리는 한국 사회에서 나타나는 혐오의 양상에 대해 살펴볼 필요가 있다."
      },

```

```

      {
        "id": "GWGR2300001050.1.2.2",
        "form": " 과거, 우리나라에서 혐오는 주로 신분의 영역에서 상위계층이 하위계층을 무시하는 양상으로 주로 나타났고, 후에 시대적 변화를 겪으며, 지역 간의 혐오, 성별에 대한 혐오 등의 다양한 양상으로 현대사회까지 이어져 오게 되었다."
      },

```

```

      {
        "id": "GWGR2300001050.1.2.3",
        "form": " 특히 그중에서도 과거와 달리, 인터넷이 발달함에 따라 지금의 현대사회에서는 '사이버 혐오'가 주로 나타나고 있다."
      },

```

```

      {
        "id": "GWGR2300001050.1.2.4",
        "form": " 그렇다면 사이버 혐오란 무엇일까?"
      }
    ]
  },

```

```

  {
    "id": "GWGR2300001050.1.3",
    "form": " 사이버 혐오란 인터넷 등의 사이버 공간에서 이루어지는 타인을 비방하는 목적의 모든 혐오를 일컫는 말이다. 특히 이러한 사이버 혐오는 인터넷이 광범위하게 퍼지고 누구나 쉽게 접할 수 있게 되며 과거의 혐오와 달리, 앞서 소개한 유명 연예인의 극단적 선택 등의 양상을 띄며, 보다 확장된 형태를 취하고 있다. 그렇다면 우리는 이러한 사이버 혐오가 만연한 사회에서 어떠한 태도를 갖추어야 할까?",

```

```

    "sentence": [
      {
        "id": "GWGR2300001050.1.3.1",
        "form": " 사이버 혐오란 인터넷 등의 사이버 공간에서 이루어지는 타인을 비방하는 목적의 모든 혐오를 일컫는 말이다."
      },

```

```

        {
            "id": "GWGR2300001050.1.3.2",
            "form": " 특히 이러한 사이버 혐오는 인터넷이 광범위하게 퍼지고 누구나 쉽게 접할 수 있게 되며 과거의 혐오와 달리, 앞서 소개한 유명 연예인의 극단적 선택 등의 양상을 띄며, 보다 확장된 형태를 취하고 있다."
        },
        {
            "id": "GWGR2300001050.1.3.3",
            "form": " 그렇다면 우리는 이러한 사이버 혐오가 만연한 사회에서 어떠한 태도를 갖추어야 할까?"
        }
    ]
},
{
    "id": "GWGR2300001050.1.4",
    "form": " 먼저, 우리는 사이버 공간상의 개개인도 하나의 인격체란 사실을 잊지 않아야 하며 또한 우리는 올바른 인터넷 생활을 위해 ‘미디어 리터러시’의 태도를 함양해야 한다. 여기서 미디어 리터러시란 정보 수용자가 미디어를 접하고 활용하는 과정에서 사용자 스스로가 비판적으로 정보를 수용 하하는 등의 윤리적인 태도를 갖추는 능력으로 현대인의 사이버 활동에 있어 필수적으로 갖추어야 할 덕목이다. ",
    "sentence": [
        {
            "id": "GWGR2300001050.1.4.1",
            "form": " 먼저, 우리는 사이버 공간상의 개개인도 하나의 인격체란 사실을 잊지 않아야 하며 또한 우리는 올바른 인터넷 생활을 위해 ‘미디어 리터러시’의 태도를 함양해야 한다."
        },
        {
            "id": "GWGR2300001050.1.4.2",
            "form": " 여기서 미디어 리터러시란 정보 수용자가 미디어를 접하고 활용하는 과정에서 사용자 스스로가 비판적으로 정보를 수용 하하는 등의 윤리적인 태도를 갖추는 능력으로 현대인의 사이버 활동에 있어 필수적으로 갖추어야 할 덕목이다."
        }
    ]
},
{
    "id": "GWGR2300001050.1.5",
    "form": " 이처럼 우리는 사회에서 일어나고 있는 다양한 혐오의 양상을 이해하는 태도를 가져야 하며 특히 현대사회에서 주로 나타나는 사이버 혐오에 대해 경각심을 가지고, 미디어 리터러시의 태도를 갖추어야 한다. 또한 사이버 상에서도 서로를 배려하는 태도를 가지며, 보다 깨끗하고 건전한, 혐오가 없는 사회를 만들기 위해 노력해야 할 것이다."
}

```

```

        "sentence": [
            {
                "id": "GWGR2300001050.1.5.1",
                "form": " 이처럼 우리는 사회에서 일어나고 있는 다양한 혐오의 양상을
이해하는 태도를 가져야 하며 특히 현대사회에서 주로 나타나는 사이버 혐오에 대해 경각심을 가지고,
미디어 리터러시의 태도를 갖추어야 한다."
            },
            {
                "id": "GWGR2300001050.1.5.2",
                "form": " 또한 사이버 상에서도 서로를 배려하는 태도를 가지며, 보다
깨끗하고 건전한, 혐오가 없는 사회를 만들기 위해 노력해야 할 것이다."
            }
        ]
    },
    "evaluation": {
        "evaluator": [
            {
                "evaluator1_ID": "23211"
            },
            {
                "evaluator2_ID": "23212"
            },
            {
                "final_evaluator_ID": "23212"
            }
        ],
        "evaluation_data": {
            "eva_score_con": {
                "evaluator1_score_total_con": 11,
                "evaluator1_score_con1": 2,
                "evaluator1_score_con2": 2,
                "evaluator1_score_con3": 2,
                "evaluator1_score_con4": 2,
                "evaluator1_score_con5": 3,
                "evaluator2_score_total_con": 12,
                "evaluator2_score_con1": 4,
                "evaluator2_score_con2": 2,
                "evaluator2_score_con3": 2,
                "evaluator2_score_con4": 2,
                "evaluator2_score_con5": 2
            }
        }
    }
}

```

```

        },
        "eva_score_org": {
            "evaluator1_score_total_org": 8,
            "evaluator1_score_org1": 4,
            "evaluator1_score_org2": 4,
            "evaluator2_score_total_org": 6,
            "evaluator2_score_org1": 3,
            "evaluator2_score_org2": 3
        },
        "eva_score_exp": {
            "evaluator1_score_total_exp": 7,
            "evaluator1_score_exp1": 3,
            "evaluator1_score_exp2": 4,
            "evaluator2_score_total_exp": 6,
            "evaluator2_score_exp1": 3,
            "evaluator2_score_exp2": 3
        },
        "evaluator1_total_score": 26,
        "evaluator2_total_score": 24
    }
}
]
}

```

[그림 VII-12] 글쓰기 채점 데이터 주석 예시

3) 최종 데이터 세트

1학기과 2학기에 각각 수집된 글쓰기 자료는 자동 편집, 수동 선별, 수동 편집의 과정을 거쳐 채점을 위한 최종 글쓰기 자료 데이터 세트를 구축하였다. 채점을 위한 글쓰기 자료는 9개 문항에 대한 글쓰기 자료 9,371건으로 확정되었다.

<표 VII-3> 최종 데이터 세트 문항별 건수

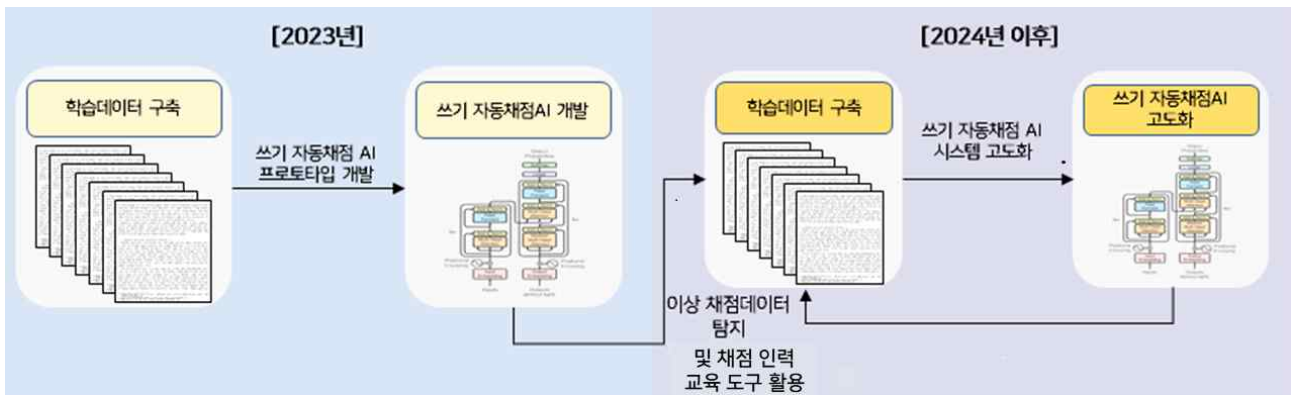
구분	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	총합
채점	794	1,231	1,198	554	585	638	0	0	0	5,000건
미채점	0	0	0	480	444	571	932	1,110	834	4,371건
총합	794	1,231	1,198	1,034	1,029	1,209	932	1,110	834	9,371건

문항별로 1천 건 수집을 목표로 하였는데, <표 VII-3>에서 알 수 있듯이 Q1, Q7, Q9을 제외한 6개 문항에서 목표에 도달하였다. 전체 9,371건은 글쓰기 자료 정보에 대한 주석을 거쳐 논술문 데이터 세트로 구성하였고, 이 가운데 5,000건은 채점 후 데이터 주석 단계를 거쳐 자동 채점 인공지능 학습용 데이터로 가공되었다.

2. 자동 채점 프로토타입 개발

1) 자동 채점 프로토타입 개발 목표

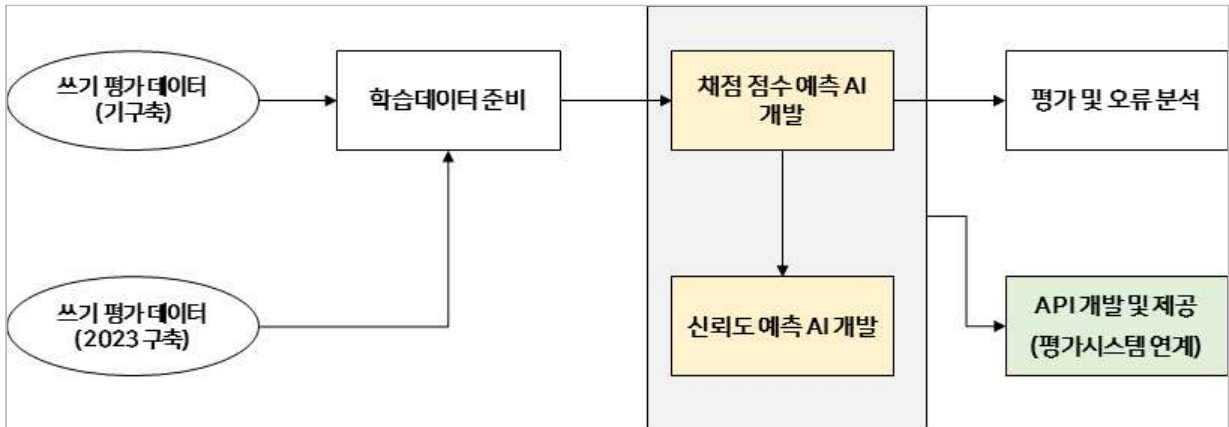
자동 채점 인공지능 SW 개발을 위하여는 다양한 글쓰기 원자료와 일관성 높은 채점 데이터의 확보가 필수적이다. 본 과업에서는 [그림 VII-13]과 같이 2023년 인공지능 기반 채점 지원 도구를 개발하고, 2024년 이후부터 개발된 도구를 활용하여 이상 채점 데이터의 탐지 및 채점 인력 교육 도구로 활용하고자 한다.



[그림 VII-13] 자동 채점 프로토타입 기술 활용 프로세스

2) 자동 채점 프로토타입 개발 내용

자동 채점 프로토타입 개발을 위하여 1) 채점 준거별 점수 예측 인공지능 프로그램과 2) 채점 지원 도구로의 활용을 위한 자동 채점 점수의 신뢰도 예측 프로그램, 3) 평가 시스템과 연계를 위한 API 서비스를 개발하였다. 개발 내용은 [그림 VII-14]와 같다.



[그림 VII-14] 자동 채점 프로토타입 개발 내용

(1) 채점 준거별 점수 예측 인공지능 프로그램 개발

채점 준거별 점수 예측 인공지능 기술 개발을 위하여, 한국어 사전학습 언어모델 (BERT 계열)을 사후학습(fine-tuning) 하는 접근 방법을 활용한다. 한국어 사전학습 모델은 한국어의 일반적인 문법 및 의미를 미리 학습한 모델로, 사전학습 모델 기반 사후 학습 접근 방법은 최근 자연어 처리 분야의 다양한 태스크에서 우수한 성능을 보이는 개발 방법이다.

채점 준거별 점수 예측 인공지능 프로그램 개발을 위하여, 2가지 관점에서 성능 개선 방안을 모색하였다. 첫 번째는 점수 예측 방법 간 비교를 통한 성능 개선이고, 두 번째는 사전학습 언어모델 간 비교를 통한 성능 개선이다.

가. 인공지능 점수 예측 방법 간 비교

인공지능 점수 예측 방법 사이의 비교 실험을 위하여, 1) 베이스라인 모델, 2) 분류 및 점수 맞추기 멀티태스크 모델, 3) 채점 기준 기반 분류 및 맞추기 멀티태스크 모델을 개발하였다.

1) 베이스라인 모델은 지시문과 글 내용을 입력으로 받아서, 점수 클래스를 분류하는 인공지능 프로그램이다. ‘점수 클래스 분류(classification)’는 자연어 처리의 대표적인 문

제 유형으로, 영어권 에세이 자동 채점(automatic essay scoring) 연구에서 자주 활용되는 접근 방법이다. 본 과제에서의 점수 클래스 분류는 1점부터 5점까지 각 점수별로 클래스를 생성하고, 개별 쓰기 데이터가 어느 클래스에 해당하는지 분류하는 방법이다. 비유하면 뉴스 기사가 주어졌을 때, 해당 기사가 정치 기사인지, 경제 기사인지, 사회 기사인지 분류하는 접근 방법을 에세이 평가에 적용한 방법으로, 뉴스 기사 대신 글쓰기 데이터가 주어지고, 경제, 정치, 사회 카테고리 대신 1점, 5점 등 점수 카테고리를 주고 학습하는 방법이다.

2) 멀티태스크 모델은 점수 클래스 분류(classification)와 점수 맞추기(regression)를 동시에 학습하는 모델이다. 점수 클래스 분류 모델은 5개의 점수 클래스 중, 정답 점수의 클래스가 가장 높은 확률을 가지도록 학습하는 방법이다. 이는 자연어 처리에서 자주 활용되는 대표적인 접근 방법이나, 사람이 생각하는 것과 같은 클래스의 레이블(1점, 5점 등)을 고려하지 못하는 한계를 가진다. 예를 들어, 4점 클래스의 쓰기 데이터를 5점으로 잘못 분류하는 경우 대비 1점으로 잘못 분류하는 경우에 더 큰 불이익(penalty)을 주어야 하지만, 그렇지 못하다는 한계가 있다. 반면, ‘점수 맞추기(regression)’ 문제는 클래스 분류가 아닌 점수를 실수 자체(1.0, 5.0 등)로 학습하는 방법으로, 정답 실수 값과 예측 실수 값 사이의 차이를 학습한다. 즉, 4점의 쓰기 데이터를 5.0점으로 잘못 평가하면 1.0의 불이익, 1.0으로 잘못 평가하면 4.0의 더 큰 불이익을 반영하는 모델이다. 세 번째 모델은 지시문, 글 내용, 채점 기준을 활용하여, 첫 번째 및 두 번째 모델과 동일하게 점수 클래스도 분류하고, 잘못 분류된 점수와의 차이 정도도 동시에 같이 학습하는 방법이다.

3) 채점 기준 기반 분류 및 맞추기 멀티태스크 모델은 채점 기준, 지시문, 글 내용을 입력으로 받아서, 점수 클래스를 분류하고 맞추는 멀티태스크 모델이다. 이는 첫 번째, 두 번째 방법 대비, 채점 기준을 입력으로 제시하여, 자동 채점 인공지능 기술이 점수별 채점 기준을 파악하고, 채점 기준과 쓰기 데이터 사이의 관련성을 학습할 수 있도록 하는 것을 목표로 한다.

각 프로그램 유형에 대한 입력 및 출력을 정리하면 <표 VII-4>와 같고, 예시 쓰기 데이터 기반 인공지능 입력 예시는 <표 VII-5>와 같다.

<표 VII-4> 자동 채점 인공지능 프로그램 유형별 입력 및 출력

유형	인공지능 입력	인공지능 출력(목표)
(모델1) 베이스라인 모델	[분류토큰] + (지시문) + [구분자] + (글 내용)	(분류) 3점
(모델2) 분류 및 점수 맞히기 멀티태스크 모델	[분류토큰] + (지시문) + [구분자] + (글 내용)	(분류) 3점 + (점수 맞히기) 3.0
(모델3) 채점 기준 기반 분류 및 맞히기 멀티태스크 모델	[분류토큰] + (채점 기준) + [구분자] + (지시문) + (글 내용)	(분류) 3점 + (점수 맞히기) 3.0

<표 VII-5> 자동 채점 인공지능 프로그램 입력 예시

예시 쓰기 데이터	<p>나는 로봇세의 도입에 반대한다. 이유는 세가지이다. 첫째, 로봇 관련 기술 및 산업의 발전을 저해한다. 둘째, 로봇세의 도입은 결국 물가 상승을 초래한다. 셋째, 그 적용 기준이 모호하다.</p> <p>공장 혹은 무인점포 등에서 사용하는 '로봇'은 기계공학의 정수라 할 수 있다. ...</p> <p>위와 같은 이유에서 나는 로봇세의 도입을 반대한다. 국내 산업을 약화시키며 기업의 경쟁력을 깎아먹는, 혹은 허울뿐인 악법이 될 것이라는 점에서이다.</p>
입력 예시 - (모델1) 베이스라인 모델	<p>[분류토큰, 이후 지시문 영역] 로봇의 발달로 일자리가 줄어들 것이라는 사람들의 불안이 커지면서 최근 로봇세 도입에 대한 논의가 활발하다. 로봇세는 로봇의 노동으로 생산하는 경제적 가치에 부과하는 세금이다. 로봇 기술의 발달로 인해 일자리를 잃는 사람들이 갈수록 많아질 수 있기 때문에, 그런 사람들을 지원하거나 사회 안전망을 구축하기 위해 예산을 마련하자는 것이 로봇세 도입의 목적이다. \n 로봇세 도입에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. [구분자, 이후 글내용 영역] 나는 로봇세의 도입에 반대한다. 이유는 세가지이다. 첫째, 로봇 관련 기술 및 산업의 발전을 저해한다. 둘째, 로봇세의 도입은 결국 물가 상승을 초래한다. 셋째, 그 적용 기준이 모호하다. \n\n 공장 혹은 무인점포 등에서 사용하는 '로봇'은 기계공학의 정수라 할 수 있다. ... \n\n 위와 같은 이유에서 나는 로봇세의 도입을 반대한다. 국내 산업을 약화시키며 기업의 경쟁력을 깎아먹는, 혹은 허울뿐인 악법이 될 것이라는 점에서이다.</p>

<p>입력 예시 - (모델3) 채점 기준 기반 분류 및 맞히기 멀티태스크 모델</p> <p>(분류토큰 이후, 채점 기준은 내용2에 해당)</p>	<p>[분류토큰, 이후 채점 기준 영역] ▶ 주장의 적절성 : 주장이 논제에 부합하는가? \n ▶ 주장의 명료성 : 주장이 글 전체에 일관되고 뚜렷한가? (점수) 5점: 주장이 논제에 부합하며, 글 전체에 걸쳐 매우 일관되고 뚜렷함. \n 4점: 주장이 논제에 부합하며, 글 전체에 걸쳐 대체로 일관되고 뚜렷함. ... [구분자, 이후 지시문 및 글 내용 영역] 로봇의 발달로 일자리가 줄어들 것이라는 사람들의 불안이 커지면서 최근 로봇세 도입에 대한 논의가 활발하다. 로봇세는 로봇의 노동으로 생산하는 경제적 가치에 부과하는 세금이다. 로봇 기술의 발달로 인해 일자리를 잃는 사람들이 갈수록 많아질 수 있기 때문에, 그런 사람들을 지원하거나 사회 안전망을 구축하기 위해 예산을 마련하자는 것이 로봇세 도입의 목적이다. \n 로봇세 도입에 대한 자신의 의견을 논리적으로 제시하는 글을 쓰시오. (내용) 나는 로봇세의 도입에 반대한다. 이유는 세가지이다. 첫째, 로봇 관련 기술 및 산업의 발전을 저해한다. 둘째, 로봇세의 도입은 결국 물가 상승을 초래한다. 셋째, 그 적용 기준이 모호하다. \n\n 공장 혹은 무인점포 등에서 사용하는 '로봇'은 기계공학의 정수라 할 수 있다. ... \n\n 위와 같은 이유에서 나는 로봇세의 도입을 반대한다. 국내 산업을 약화시키며 기업의 경쟁력을 깎아먹는, 혹은 허울뿐인 악법이 될 것이라는 점에서이다.</p>
---	---

나. 사전학습 언어모델 간 비교

한국어 버트(BERT) 계열 사전학습 언어모델은 한국전자통신연구원(ETRI)에서 공개한 코리안버트(KorBERT), 에스케이티(SKT)에서 공개한 코버트(KoBERT), 개인 연구자가 공개한 코일렉트라(KoElectra) 및 코빅버드(KoBigBird), 클루(KLUE) 프로젝트 팀에서 공개한 클루-로버타(KLUE-RoBerta) 등 여러 모델이 있다. 사전학습 언어모델은 사전학습 시에 사용할 단어 집합(사전, vocabulary)과 최대 길이(max length)가 미리 결정되며, 사후학습 단계에서는 일반적으로 사전학습과 동일한 단어 집합 및 최대 길이를 활용하여 사후학습을 수행한다. 대다수의 한국어 버트(BERT) 계열 언어모델은 그 개발자가 처음 제안한 연구와 동일하게 512개의 토큰을 최대 길이로 설정하여 학습한 모델이고, 본 과제에서는 약 1천 자 이상의 쓰기 데이터, 지시문, 채점 기준 텍스트 활용이 필요하여 입력 길이 512 토큰을 초과한다는 어려움이 있다. 공개 모델 중, 코빅버드(KoBigBird) 모델은 영어권에서 긴 길이 입력 처리를 위하여 제안된 빅버드(BigBird) 알고리즘에 한국어 데이터를 학습한 모델로, 최대 4천 토큰 길이를 제공하는 특징을 가진다. 따라서, 본 연구에서는 긴 길이의 쓰기 데이터를 입력으로 처리하기 위하여, 코빅버드 베이스(KoBigBird Base) 언어모델을 기본 모델로 활용하여 채점 준거별 점수 예측 인공지능을 개발한다.

또한, 사전학습 언어모델 기술은 모델의 크기가 커질수록 성능이 향상되는 특징을 가진다. 512개 토큰 제한으로 쓰기 데이터 전체를 입력으로 받지 못하더라도, 딥러닝 모델의 크기가 커짐으로 생기는 성능 향상 정도를 비교하기 위하여, 클루-로버타 라지(KLUE-RoBERTa Large) 언어모델과 비교 실험을 수행하였다.

다. 실험 데이터 및 실험 결과

채점 준거별 점수 예측 인공지능을 위한 실험 데이터는 2023년 구축된 데이터를 학습 데이터와 평가 데이터로 구분하여 실험하였다. 2023년 9월까지 채점이 완료된 4,002건에 대해서, 임의 구분(random split) 방법으로 학습 데이터 3,802건, 평가 데이터 200건으로 구분하여 실험하였다.

첫 번째로, 인공지능 점수 예측 방법 간 비교 실험 결과는 <표 VII-6>과 같다. 사전학습 언어모델은 기본 모델인 긴 길이 코빅버드(KoBigBird Base) 언어모델을 활용하였다. 9개 채점 준거별 정확도를 평균한 결과, (모델3) 채점 기준까지 고려한 멀티태스크 모델이 48.1%의 정확도로 가장 높은 성능을 보였다.

<표 VII-6> 인공지능 점수 예측 방법 간 비교 실험 결과

유형	9개 채점기준 평균 정확도
(모델1) 베이스라인 모델	40.8%
(모델2) 분류 및 점수 맞추기 멀티태스크 모델	41.6% (+0.8%)
(모델3) 채점 기준 기반 분류 및 맞추기 멀티태스크 모델	41.8% (+0.2%)

두 번째로, 사전학습 언어모델 간 비교 실험 결과는 <표 VII-7>과 같고, Base 크기의 긴 길이 언어모델인 코빅버드(KoBigBird) 모델과 큰 규모 클루-로버타(KLUE-RoBERTa Large) 모델을 비교 평가하였다. 9개 채점 준거별 정확도를 평균한 결과, 큰 규모 클루-

로버타 사전학습 언어모델에 (모델2) 분류 및 점수 맞추기 멀티태스크 모델이 가장 높은 43.8%의 정확도를 보였다. 참고로, 클루-로버타(KLUE-RoBERTa) 모델은 512개의 토큰 길이 제한으로 채점 기준 텍스트를 추가로 활용하는 (모델3) 방법은 실험하지 못하였다.

<표 VII-7> 사전학습 언어모델 간 비교 실험 결과

유형	긴 길이 언어모델 (KoBigBird, base 크기)	큰 규모 언어모델 (KLUE_RoBERTa, large 크기)
(모델1) 베이스라인 모델	40.8%	41.6% (+0.8%)
(모델2) 분류 및 점수 맞추기 멀티태스크 모델	41.6%	43.8% (+2.2%)

(2) 채점 점수의 신뢰도 예측 인공지능 프로그램 개발

두 번째로는 자동 채점 인공지능 기술이 채점 지원 도구로 활용될 수 있도록, 자동 채점 점수에 대한 신뢰도 점수 제공 기술을 개발하였다.

신뢰도 점수는 인공지능 모델이 예측한 점수에 대한 확률값을 의미한다. 예를 들어, 자동 채점 인공지능이 두 개의 에세이에 대해 같은 4점을 예측하였더라도, 90%의 신뢰도를 보이는 에세이의 경우 전문 채점자가 4점임을 90% 확률로 믿을 수 있지만, 50%의 신뢰도를 보이는 경우 전문 채점자가 4점이 불확실한 점수라고 알 수 있게 함을 목표로 한다. 이를 위하여 신뢰도 예측 인공지능 프로그램은 예측한 신뢰도 점수(예: 신뢰도 0.8, 0.5 등)가 실제 정확도(예: 정확도 80%, 50% 등)와 일치하는 신뢰도 값 제공을 목표로 한다. 신뢰도 점수 예측 기술은 딥러닝이 기존 기술들보다 높은 정확도를 보이지만, 오답의 경우에도 높은 확률로 오답을 예측한다는 한계를 보완하기 위한 기술이다.

신뢰도 예측 인공지능 개발을 위하여 인공지능의 불확실성(uncertainty) 예측 분야에서 널리 사용되는 온도(temperature) 보정 기반 신뢰도 예측 기술을 개발하였다. 온도 보정 기반 신뢰도 예측 방법에 대한 세부 설명은 다음과 같다. 최근 인공지능 모델은 기본적으로 출력값에 대한 확률 모델이다. 즉, 1점-5점 사이 점수를 분류하는 모델은 1점

부터 5점까지 점수별 확률을 계산(확률 분포 계산)하고, 최고 확률값을 가지는 점수를 최종 예측 점수로 제시하고 해당 최고 확률값을 예측 점수의 신뢰도로 활용한다. 온도 보정 방법은 점수별 확률 분포 계산 시, 미리 정의한 온도 값(temperature)을 활용하여 최고 확률값이 실제 해당 예제의 정확도 일치하도록 확률 분포를 조정하는 기법이다. 온도 보정 기법은 구오 외(Guo et al., 2017)의 연구에서부터 단순하지만 우수한 성능으로 불확실성 예측 분야에서 기본 접근 방법으로 널리 활용되고 있으며, 최근 연구(Shrey Desai & Greg Durrett, 2020)에서 자연어 처리 분야 버트 언어모델 기반 분류 문제에 온도 보정 기법을 적용하여 신뢰도 개선 연구를 수행하였다.

온도(temperature) 보정 기법을 활용한 신뢰도 구간별 정확도 개선 결과는 <표 VII-8>과 같고, 조직2 채점 준거에 대한 신뢰도 개선 결과는 <표 VII-9>와 같다.

<표 VII-8> 온도 보정 기반 신뢰도 예측 개선 결과

예상보정 오차(ECE) 개선	내용1	내용2	내용3	내용4	내용5	조직1	조직2	표현1	표현2
신뢰도 보정 전	0.1265	0.0241	0.3543	0.3156	0.0129	0.3061	0.3141	0.0960	0.2271
신뢰도 보정 후	0.0090	0.0045	0.1549	0.1101	0.0037	0.1047	0.1221	0.0160	0.0556

<표 VII-9> 온도 보정 기반 신뢰도 개선 예시

[채점기준: O2]	신뢰도 보정 전			신뢰도 보정 후		
신뢰도 구간	샘플 개수	평균 신뢰도	평균 정확도	개수	평균 신뢰도	정확도
0.2 - 0.4	-	-	-	1	0.3995	-
0.4 - 0.6	51	0.5298	0.3921	188	0.4969	0.4468
0.6 - 0.8	104	0.6956	0.4519	11	<u>0.6107</u>	<u>0.7272</u>
0.8 - 1.0	45	<u>0.8456</u>	<u>0.5555</u>	-	-	-

(3) 제안: 채점 지원 도구로의 활용 방안

채점 준거별 점수 예측 인공지능을 채점자 지원 도구로 활용하기 위하여, 상위 1개의 단일 점수가 아닌 상위 2개의 점수를 범위(boundary) 형태로 제공하는 방안을 제안한다. 이와 같은 점수 범위 제공이 가능할 경우, 채점자 간 점수 차이가 2점 이상 벌어지는 경우를 방지할 수 있을 것으로 기대된다. 점수의 범위를 예측할 경우, 개발된 인공지능의 채점 준거별 성능은 <표 VII-10>과 같고, 데이터는 최종 5,000건을 학습 데이터 4,800건, 평가 데이터 200건으로 임의 분류하여 평가하였다.

<표 VII-10> 채점 준거별 점수 범위 예측 성능

		C1	C2	C3	C4	C5	O1	O2	L1	L2	평균
단일 점수 예측	정확도	53.00%	36.50%	46.00%	45.50%	40.50%	43.50%	42.00%	40.50%	51.00%	44.28%
	정확도	73.50%	72.50%	76.50%	69.00%	67.50%	74.50%	72.50%	70.50%	78.50%	72.78%
점수 범위 예측	신뢰도	0.8785	0.8842	0.7764	0.8300	0.7684	0.8243	0.8658	0.8822	0.8425	0.8392
	피어슨 상관계수	0.7509	0.7622	0.8144	0.7425	0.8055	0.7583	0.8065	0.7992	0.8415	0.7868
	평균제곱 오차	0.41	0.41	0.295	0.425	0.4	0.465	0.395	0.355	0.275	0.3811

<표 VII-9>를 살펴보면, 점수 범위 예측 정확도는 평균 72.78%이며, 반론 가능성을 고려하는 C5의 정확도가 67.5%로 상대적으로 낮음을 확인할 수 있다. 점수 범위를 예측할 경우의 평균 신뢰도는 0.8392, 피어슨상관계수(PCC)는 0.7868, 정답 점수와의 차이를 계산하는 평균제곱오차는 0.3811이다.

3) 자동 채점 프로토타입 에이피아이(API) 개발

자동 점수 예측 인공지능 및 신뢰도 예측 인공지능 결과를 외부 채점 시스템과 연동하기 위하여, 에이피아이(API) 서비스를 개발하였다.

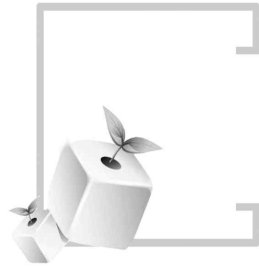
에이피아이(API)란 ‘Application Programming Interface’의 약자로, 컴퓨터 프로그램이 미리 정해진 규칙에 따라 서로 데이터를 주고받는 방법(인터페이스)을 의미한다. 예를 들어, 트위터 에이피아이(Twitter API)는 플랫폼에 접근하여 트윗을 게시하거나 읽을 수 있는 기능을 제공한다.

자동 채점 프로토타입 에이피아이(API)는 채점 시스템과 연계하는 기능을 제공한다. 채점 시스템에서 채점자가 채점 중인 글의 지시문 및 내용을 자동 채점 에이피아이(API)로 전달하면, 자동 채점 에이피아이(API)는 전달된 글을 평가하여 채점 준거별 점수 범위 예측 결과를 채점 시스템에 전달한다. 채점 시스템은 전달받은 채점 준거별 점수 범위 예측 결과를 채점자에게 보여 주어, 자동 채점 결과를 활용할 수 있도록 한다.

자동 채점 프로토타입 에이피아이(API)와 채점 시스템 사이의 데이터 전달 예제는 <표 VII-11>과 같다.

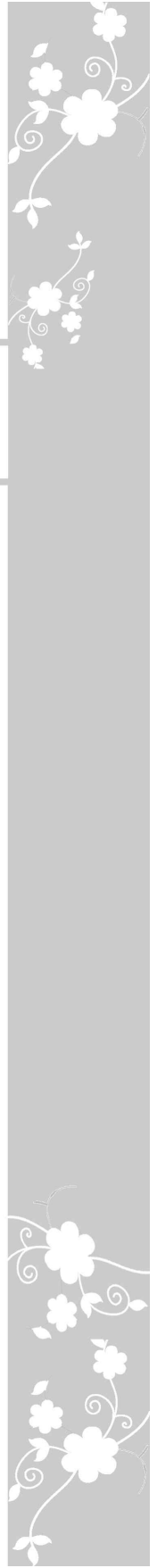
<표 VII-11> 자동 채점 프로토타입 에이피아이(API) 서비스 예시

채점 시스템 → 자동채점 프로토타입 사이 전달 데이터	자동채점 프로토타입 → 채점 시스템 사이 전달 데이터
<p>question=블라인드 채용을 폐지해야 하는가? text=블라인드 채용은 구직자의 개인정보를 감추고, 능력과 경험에 따라 지원자를 평가하는 채용 방식이다. 블라인드 채용은 인종, 성별, 출신학교 등의 인적 요소에 따른 편견이나 차별을 최소화하기 위한 방식으로, 다양한 인재를 공정하게 선발하고자 할 때 적용된다. 일반적으로 회사에서는 직원을 채용하는 방식인 구직자가 제출하는 이력서와 자기소개서 등에서 개인정보를 제거하거나, 제출하지 않도록 하고 지원자의 경력, 교육배경, 기술 등을 검토하여 적합성을 평가하기에 이러한 방식이 공정성을 불러일으킨다는 장점이 주목받고 있다. 하지만 블라인드 채용 방식이 항상 공정하다는 것은 보장되지 않기에 최근에는 블라인드 채용에 대한 비판과 반대 의견도 나오고 있다. 이번 글에서는 블라인드 채용 폐지에 대한 찬성의견에 대해 논의해보고자 한다.\n \n 블라인드 채용은 인적 요소에 따른 차별을 최소화하는 좋은 방법이다. 그러나 이 방식이 공정하다는 것은 보장할 수 없다. 그에 대한 첫 번째 이유로 블라인드 채용에서 지원자의 경력, 학력, 기술 등의 정보를 기반으로 선발하는 운영방식을 예로 들 수 있다. 이로 인해 지원자의 능력 외에도 다른 인적 요소가 고려될 가능성이 있기에 블라인드 채용의 원래 취지에 어긋나는 결과를 초래할 수 있다. 두 번째 이유로 블라인드 채용을 적용하면서 기업이 구체적으로 원하는 인재상을 표현하는 것이 어렵다는 점이다. 블라인드 채용은 기업의 인재상에 대한 명확한 기준을 설정하기 어렵고, 이는 기업에게도 불리한 점입니다. 위에서 말했듯 회사가 직원을 선발할 때 지원자의 경력, 교육배경, 가치관등을 보고 기업의 방향성과 일치하는 지를 알아보지만 블라인드 채용의 방식으로는 이러한 과정을 거칠 수 없기에 이는 큰 단점이 될 수 있다. 마지막으로, 블라인드 채용은 기업과 지원자 간의 상호작용을 감소시킬 수 있다. 지원자들은 자신의 개인적인 경험과 역량을 자랑하는 기회를 상실할 수 있고, 기업들은 인재들의 개인적인 특성과 능력을 파악하는 기회를 놓칠 수 있다.\n \n 블라인드 채용은 인적 요소에 따른 차별을 줄이는 좋은 방법이지만, 이 방식이 공정하다는 것을 보장할 수는 없다. 블라인드 채용이 기업과 지원자 간의 상호작용을 감소시키고, 기업의 인재상 설정에 어려움을 줄 수 있다는 점을 고려해볼 때, 블라인드 채용을 폐지하는 것이 적절한 해결책이 될 수 있다.</p>	<p>* 평가 결과 └ C1 └ 최소 점수: 4 └ 최대 점수: 5 └ 신뢰도: 0.7129 └ C2 └ 최소 점수: 3 └ 최대 점수: 4 └ 신뢰도: 0.7423 └ C3 └ 최소 점수: 4 └ 최대 점수: 5 └ 신뢰도: 0.8292 └ C4 └ 최소 점수: 3 └ 최대 점수: 4 └ 신뢰도: 0.8091 └ C5 └ 최소 점수: 2 └ 최대 점수: 3 └ 신뢰도: 0.6452 └ O1 └ 최소 점수: 3 └ 최대 점수: 4 └ 신뢰도: 0.7960 └ O2 └ 최소 점수: 3 └ 최대 점수: 4 └ 신뢰도: 0.7391 └ L1 └ 최소 점수: 3 └ 최대 점수: 4 └ 신뢰도: 0.8030 └ L2 └ 최소 점수: 3 └ 최대 점수: 4 └ 신뢰도: 0.8204</p>



제 VIII 장

결론



1. 요약

‘2023년 국민의 글쓰기 능력 진단 체계 개발 연구’ 과제는 국민의 글쓰기 능력 향상을 도모할 수 있는 효과적이고 신뢰할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있다. 이를 위해 국민의 글쓰기 능력 진단 지표를 고도화하여, 신뢰도 높은 채점 전문 인력 양성 프로그램을 개발하고, 고품질의 논증적 글쓰기 원자료를 수집하였다. 이를 기반으로 체계적으로 채점 자료를 구축하여, 자동 채점을 지원하기 위한 인공지능 기반 모델 개발을 실행하였다. 세부 연구 결과는 다음과 같다.

□ 글쓰기 능력 진단 지표 고도화

「인공지능 활용 국어능력 진단체계 개발 기초 연구」(국립국어원·서울대학교 국어교육연구소, 2022)의 결과로서 개발된 논증적 글쓰기 문항의 프로토타입을 수정·보완하여 동형의 논증적 글쓰기 문항 15개를 추가 개발하였다. 개발한 문항은 작문 교육 전문가 및 대학 글쓰기 강좌 교강사의 검토, 대학생 참여자의 예비 평가 등의 과정을 거쳐 성인을 대상으로 한 글쓰기 능력 진단 도구로서의 타당성이 검증되었다.

논증적 글쓰기 채점 기준표(rubric)의 고도화를 위해 채점 기준표 상세화, 표현 범주 기준 정량화를 시행하였다. 상세화를 위해 채점 기준표에서 유의 사항 등 기술 내용을 명료화하고, 예시 글쓰기 자료를 선별하여 제시하였다. 표현 범주 정량화를 위해 2022년 기초 연구의 채점 결과를 분석하여 ‘문장과 어휘’, ‘어문 규범과 관습’ 준수를 양적으로 평가할 수 있는 기술문(description)을 마련하였다. 그런데 정량적 기술문은 적용 시 채점 난도가 높고, 글 자료의 분량 차이를 고려한 타당한 적용이 어렵고, 그에 따라 채점 신뢰도를 높이는 데 기여하는 것이 제한적임이 발견되었다. 이에 질적 기준을 추가하는 보완 작업을 통해 기술문을 최종 정련하였다.

본 연구에서 개발한 논증적 글쓰기 채점 기준표(rubric)의 타당성을 검증하기 위해 5 수준(level)을 고루 포함한 총 50건의 글쓰기 자료에 대해 채점 기준표를 적용한 분석적 채점 결과와 외부 전문가 3인의 총체적 채점 결과를 비교 분석하였다. 채점 점수 평균에 대한 차이 검정(t-test) 결과 통계적으로 유의한 차이가 없는 것으로 나타났다.

마지막으로 진단 목적이나 진단 대상의 특성에 따라 내용, 조직, 표현 범주별 또는 하위 준거별로 가중치를 부여하기 위한 방안을 연구하였다. 본 연구에서 개발한 9개의 채점 준거로 구성된 진단 지표는 성인의 일반적인 논증적 글쓰기 능력 전반을 측정하는 총

괄적 모형에 가깝다. 내용 5개, 조직 2개, 표현 2개 채점 준거의 이론적 타당성을 기반으로 범주별 또는 채점 준거 간 가중치를 달리하며 선택적으로 운용할 수 있는 다양한 방안을 검토하였다. 최종적으로 내용 준거에 가중치를 부여하는 내용 중심 모형, 조직과 표현 준거에 가중치를 두는 형식 중심 모형과 이들을 비교적 유사한 수준으로 고려하는 균형적 모형을 제안하였다.

설명적 글쓰기 문항 프로토타입을 개발하고, 채점 준거와 기준표를 개발하는 기초 연구를 수행하였다. 먼저 설명적 글쓰기 및 설명문 평가 관련 국내·외 연구와 사례를 검토하여 진단 지표의 이론적인 논거를 마련하였다. 설명적 글쓰기의 한 유형으로서 비교·대조의 구체적인 대상, 주제, 목적, 분량 등의 기준이 포함된 프로토타입을 개발하였다. 채점 준거는 논증적 글쓰기와 공통 지표로서 내용, 조직, 표현 범주를 설정하고, 내용 범주는 설명문의 특성에 부합하도록 ‘맥락에 대한 이해’, ‘중심 내용’, ‘뒷받침 내용의 적절성’, ‘뒷받침 내용의 충분성’을 하위 준거로 설정하였다.

□ 글쓰기 원자료 수집

대학 협의체 운영을 통해 글쓰기 참여자를 안정적으로 확보하여 대규모 글쓰기 자료를 수집하였다. 대학 협의회와 실무자 회의를 각 3회 진행하였으며, 상시 소통 체계를 구축하여 이메일 및 SNS를 활용하여 자료 수집 방안을 협의했다. 또 학교별 홍보 자료 수요를 조사하고 맞춤형 디지털 홍보 자료를 개발하여 지원하는 등 실무 및 교강사 활동을 지원하였다.

대규모 데이터의 효율적인 수집과 안정적인 관리를 위해 글쓰기 자료 수집 플랫폼을 개발·활용하였다. 플랫폼 작성자 페이지는 글 작성 외에 글쓰기 참여자 정보 입력 및 개인정보 활용 동의, 저작권 이용 허락 계약서 작성, 작성 글 다운로드 기능을 갖추었다. 또 관리자 페이지에는 글쓰기 자료 수집 현황 모니터링, 글쓰기 자료 다운로드 기능을 구축하고, 나아가 글쓰기 자료 간 유사도 검사, 글쓰기 자료 자동 정제 등 인공지능 학습 데이터 구축을 위해 필요한 일련의 데이터를 수집·관리하는 기능을 갖추도록 하였다.

2023년 1학기와 2학기에 걸쳐 총 9,439건의 글쓰기 원자료를 수집하였다. 글쓰기 능력 진단을 위한 글 내용 및 형식 요건을 충족하지 않았거나, 외국인 학생이 작성한 글쓰기 원자료를 제외하는 검수를 진행하였다. 최종적으로 글쓰기 자료 요건을 충족하는 9,371건의 유효 글쓰기 자료를 구축하여, 목표로 한 9천여 건을 충족하였다. 6개의 문항별로는 최소 794건부터 최대 1,231건으로 목표로 한 문항별 1천여 건에 근접하였다.

□ 글쓰기 채점 전문 인력 양성

글쓰기 채점 전문 인력을 양성하기 위해 실무 중심 논증적 글 채점 교육 프로그램을 개발·운영하였다. 우선 프로그램의 설계 단계에서는 국내외 문헌 및 사례를 검토하여 20차시의 블렌디드 교육과정을 설계하였다. 온라인 교육과정(12차시)은 쓰기 채점을 수행하는 데 필요한 기초적인 지식과 다양한 사례를 바탕으로 그 원리를 체득하는 것을 목표로 하였다. 채점 워크숍 형태로 진행한 오프라인 교육과정(8차시)은 채점 과정에서 직면할 수 있는 여러 문제 상황들을 점검하고 동료 채점자들과 협의하여 채점 기준을 익히는 것을 목표로 하였다. 교육 프로그램의 이수 요건은 온라인 강의 수강률 100%, 오프라인 강의 출석률 100%, 강의 참여도로 확정하였다.

프로그램 개발 단계에서는 교육 플랫폼 구성, 교수학습 자료 개발, 학습 진단을 위한 과제 개발이 이루어졌다. 온라인 교육을 운영할 플랫폼으로서 네이버 웨일클래스를 채택하였다. 강의 영상을 포함한 온라인 강의 자료 및 스크립트는 2022년 기초 연구 자료의 사례를 토대로 제작되었다. 강의 자료, 퀴즈 문항, 온-오프라인 교육 간 연계성을 위한 사전 채점 과제 9편을 포함한 온라인과 오프라인 교육용 자료는 자료집으로 제작하여 배부하였다.

프로그램 운영 단계에서는 대학의 글쓰기 교강사를 대상으로 전국의 약 22개 대학에서 총 109명의 교육 참여자를 모집하였다. 온라인 교육과정은 6월-7월에 실행하였으며, 오프라인 교육은 7월 중 4개 권역(수도권, 충청권, 호남권, 경상권)에서 각각 진행되었다.

프로그램 평가 단계에서는 이수 사정을 실시하여 최종적으로 96명이 '2023년 채점 전문 인력 양성 프로그램 1기'를 공식 수료한 것으로 확정하였다. 또 프로그램 평가를 위해 온라인 교육 전과 오프라인 교육 후 채점자 교육용 사전 설문 조사와 사후 설문 조사를 각각 시행하였다. 설문 결과 프로그램 전반에 대한 만족도가 90%를 상회하는 것으로 나타났으며 채점자 교육에 대한 필요성이 사전 설문 대비 높아진 것을 확인할 수 있었다. 프로그램 운영 과정에서 수집된 자료를 바탕으로 온라인 교육과 오프라인 교육 내용의 긴밀한 연계, 온라인 교육의 효과를 극대화하기 위해 보다 적극적인 수강 장치를 마련, 프로그램의 지속적 운영을 위한 플랫폼 안정화를 개선 방안으로 도출하였다.

글쓰기 진단 지표 고도화 연구 결과와 글쓰기 채점 전문 인력 양성을 위한 교육 프로그램 운영 결과를 바탕으로 채점자 매뉴얼을 개발하였다. 매뉴얼 전반부에서는 논증적 글쓰기 평가와 관련된 이론적 배경을 설명하고, 후반부에서는 본 연구에서 개발한 논증적 글쓰기 문항의 특성과 채점 준거를 소개하고, 이를 실제 글쓰기 자료 채점에 적용한 사례와 채점 근거에 대한 해설을 제시하였다.

□ 글쓰기 자료 채점 실행

체계적이고 일관된 채점 실행을 위한 채점 모형과 자동 채점 인공지능 학습 데이터로 활용하기 채점 자료 선정 모형을 제시하였다. 채점 모형은 글쓰기 자료 목록 확정, 채점 방법 결정, 채점 할당 설계, 채점 자료의 관리로 구성되었다. 채점 자료 선정 모형은 원점수 기반의 접근과 척도 점수 기반의 접근으로 구성된다.

다음으로 채점자들에게 일관된 채점 환경을 제공할 수 있는 채점 플랫폼을 개발하였다. 채점 플랫폼은 채점 신뢰도 제고, 채점 피로도 완화, 채점 진행에 대한 소통 강화에 기여함으로써 채점 데이터를 효율적으로 수집하고 관리하는 데 그 목적이 있다. 채점자 페이지에는 기본 정보 입력 화면, 채점 자료 목록 확인 화면, 채점 실행 화면으로 구성하였다. 채점 관리자 페이지는 채점자별 정보 확인 화면, 채점 대상 글쓰기 자료 할당을 위한 화면, 채점 진행 상황 모니터링 화면으로 구성하였다.

본 채점은 채점의 효율적인 관리를 위해 1-5차로 나누어 진행이 되었으며, 채점자 간 편차가 큰 사례에 대한 재채점까지 총 6회에 걸쳐 시행하였다. 채점은 채점 인력 양성 프로그램을 이수하고, 본 연구의 채점에 참여할 의사를 밝힌 88명의 채점자가 회차별로 선별하여 진행되었다.

1-4차 채점에서는 상반기에 수집한 글쓰기 자료 4,002편을 대상으로 가교 채점과 짝 채점을 하는 글쓰기 자료를 편성하고, 개인별로 총 100-108편을 채점하도록 하였다. 하반기에 진행한 5차 채점에서는 개인별로 42-56편씩, 총 998편을 채점하였다. 짝 채점 사례 중 두 채점자 간 평균 표준편차가 0.9 이상인 92편의 글에 대해서는 재채점이 이루어졌다.

채점자들은 채점 수행 직전 온라인 채점자 워크숍에 참석하여 사전 채점 연습, 실시간 채점 연습, 소집단별 토의 활동에 참여하여 채점 기준표 사용과 문항 특성, 응답 특성에 관한 인식을 공유하였다. 이후 채점 일관성 유지를 위하여 가교 채점 결과에 대한 개인별 채점 피드백을 제공하였고, 채점자들은 제공받은 피드백 내용을 숙지한 후 2인 이하의 글쓰기 자료를 채점하는 짝 채점에 참여하였다. 재채점은 우수한 채점 특성을 갖고 있는 채점자들을 선별하여 글쓰기 자료별로 1인이 채점하도록 하였다.

□ 글쓰기 자료 채점 결과 분석 및 선정

글쓰기 능력 평가 문항 및 채점 도구, 채점의 특징을 파악하고, 나아가 채점 자료에 대한 대표 채점 결과를 선정하기 위해 5,000편의 글쓰기 자료에 대한 채점 결과를 분석하

였다. 결과 분석은 원점수 기반의 고전검사이론 분석과 척도 점수 기반의 문항반응이론 분석을 중심으로 이루어졌다.

먼저 고전검사이론에 따라 기술통계, 신뢰도, 상관 분석, 요인 분석을 진행하였다. 기술통계 분석 결과 문항 5(조력 존엄사 허용)의 곤란도가 가장 높고, 문항 2(혐오 표현 규제)가 가장 낮았으며, 준거 중에는 내용 5(다른 입장 고려)가 가장 높고, 표현 2(어문 규범과 관습)의 곤란도가 가장 낮았다. 가교 채점에 대한 채점자 신뢰도 분석 결과 전체 그룹에서 급내 상관이 0.9 이상으로 매우 높은 수준이었다. 9개 채점 준거에 대한 상관분석 결과, 조직 1(글 전체 구성)과 조직 2(문단 구성 요건)의 상관이 0.805로 가장 높았고, 반대로 내용 5(다른 입장 고려)와 표현 2(어문 규범과 관습)의 상관이 0.180으로 가장 낮았다. 총점과의 상관에서는 내용 3(이유/근거의 적절성)의 상관이 0.825로 가장 높았으며, 내용 5(다른 입장 고려)의 상관이 0.459로 가장 낮았다. 각 문항별로 총점-준거의 상관을 비교하였을 때는 전반적으로 평균 0.7 정도의 양호한 상관을 갖고 있는 것으로 나타났다. 다음으로 9개의 채점 준거를 대상으로 확인적 요인 분석을 실시한 결과 9개 채점 준거 모두 모형을 구성하고 있는 것으로 확인이 되었다. 그중 조직 2(문단 구성 요건)의 성분 추정치가 가장 높고, 내용 5(다른 입장 고려)가 가장 낮았다. 분석 모형의 구성적 타당성에 대해서도 모두 양호한 결과가 나타났다.

문항반응이론에 따라 다국면 라쉬 모형을 기반으로 채점자의 채점 특성을 분석하였다. 채점자들은 채점 척도를 안정적으로 사용하고 있었으나, 일부는 높은 수준의 관대성을 보였고, 채점 척도를 부적합하게 사용한 사례로 발견하였다. 문항별로는 채점자들이 전체적으로 균형 있게 채점하고 있었지만, 문항 4-6의 경우에 문항 1-3에 비해 채점자 효과의 값이 커지는 경향이 나타났다. 다음으로 문항별 채점 준거 특성에 대한 분석 결과에서는 전체 문항에서 내용 5(다른 입장 고려)의 곤란도가 가장 높은 것으로 나타났으며, 문항 1, 3, 6의 경우 조직 2(문단 구성 요건)의 변별도가 높았으며, 문항 2, 4, 5에서는 내용 3(이유/근거의 적절성)의 변별도가 높았다. 변별도가 가장 낮은 채점 준거는 공통적으로 내용 5(다른 입장 고려)였다.

자동 채점 인공지능 학습용 데이터를 구축하기 위해 복수의 채점 결과 중 대표 채점 결과를 선정하였다. 선정은 채점자의 채점 특성을 분석하여 보다 균형적이고 안정적인 결과를 나타낸 채점자의 채점 결과를 대푯값으로 선정하는 방식이었다. 다만 짝 채점 사례 중 두 채점자 간 의견 불일치 정도가 큰 92편의 경우 우수한 채점자인 제3자의 재채점 결과를 대푯값으로 채택하였다.

□ 글쓰기 자료의 정제 및 주석화

자동 채점 인공지능 개발용 학습 데이터 구축을 위해 비유효 데이터를 걸러내고, 개인 정보를 비식별화하는 등 글쓰기 자료를 정제하는 단계를 진행하였다. 데이터 정제는 플랫폼 내 데이터베이스화 과정에서 단락 간격, 글자 간 간격 등 ‘형식 요소를 자동 편집’하는 1단계, 심각한 논제 이탈, 심각한 표절, 과제 장르 이탈, 글자 수 기준 미충족 등의 ‘부적합 데이터를 선별’하는 2단계, 제목이나 주제문, 목차와 같은 구성 표지, 인사말과 자기소개와 같은 글쓰기 상황 표현, 글 전개와 무관한 내용과 기호 등 ‘불필요 항목을 삭제’하는 3단계, 개인 식별 정보, 혐오와 차별 조장 내용 등 ‘개인정보나 비윤리적 내용을 삭제’하는 4단계로 진행되었다.

그다음으로 정제된 글쓰기 자료에 파일 메타 정보, 글쓰기 자료 정보, 채점 결과 정보 등 주요 정보를 주석화하고 최종 검수하는 단계를 진행하였다. 파일 메타 정보에는 생성자, 배포자, 말뭉치 구축 연도, 글 유형, 길이, 분석 층위, 샘플링 방식이 포함되었다. 글 자료 정보에는 직업, 성별 등 글쓰기 참여자 정보, 수집 방식, 글 유형, 수집 일시, 글쓰기 과제 발문, 문서 통계가 포함되었다. 채점 정보는 채점자 정보와 채점자별/채점 영역별 세부 점수가 포함되었다.

□ 인공지능 기반 채점 지원 도구 개발

이상 채점 데이터의 탐지 및 채점 인력 교육 도구로 활용할 수 있도록, 채점 지표별 점수 예측 인공지능 프로그램과 채점 지원 도구로의 활용을 위한 자동 채점 점수의 신뢰도 예측 프로그램, 평가 시스템과 연계를 위한 API 서비스를 개발하였다.

먼저 채점 지표별 점수 예측 인공지능 프로그램을 개발하기 위해 베이스라인 모델, 분류 및 점수 맞추기 멀티태스크 모델, 채점 기준 기반 분류 및 맞추기 멀티태스크 모델을 설계하고 모델 간 점수 예측 성능을 비교하였다. 그 결과 채점 기준까지 고려한 멀티태스크 모델이 48.1%의 정확도로 가장 높은 성능을 보였다. 나아가 딥러닝 모델의 크기가 커짐으로 생기는 성능 향상 정도를 비교하였는데, 큰 규모 클루-로버타(KLUE-RoBERTa Large) 사전학습 언어모델을 활용한 분류 및 점수 맞추기 멀티태스크 모델에서 가장 높은 정확도(43.8%)를 보였다. 다음으로 채점 점수의 신뢰도 예측 AI 프로그램을 개발하고, 온도(temperature) 보정 기법을 활용하여 신뢰도 구간별 정확도를 전반적으로 개선할 수 있음을 확인하였다.

채점 지표별 점수 예측 AI를 채점자 지원 도구로 활용하기 위하여, 상위 1개의 점수가

아닌 상위 2개의 점수를 범위(boundary) 형태로 제공하는 방안을 제안하였다. 상위 1개의 점수를 예측하는 정확도는 모든 채점 지표에 대해서 50% 미만이었지만, 상위 2개의 점수를 제시할 경우, 개발된 AI의 채점 지표별 점수 범위 예측 성능은 30% 이상 향상되었다. 채점 지표별 정확도는 C5(다른 입장 고려)에서 61.50%로 가장 낮았고, C1(문제 상황 이해)과 C2(주장)에서 75%로 가장 높았다. 최종적으로 자동점수 예측 AI 및 신뢰도 예측 AI 결과를 외부 채점 시스템과 연동하기 위하여, API 서비스를 개발하여 데이터 전달 예제를 제시하였다.

2. 후속 과제 제언

1) 글쓰기 능력 진단 도구 고도화

(1) 설명적 글쓰기 진단 도구 타당화

본 연구에서는 다양한 설명 방법 중 개념, 사실, 절차, 사건 등 설명의 대상 전반에 폭넓게 활용될 수 있는 비교-대조를 토대로 대상에 대해 설명하는 과제를 개발하였다. 그런데 설명적 글쓰기는 설명의 대상과 목적에 따라 내용 구성의 방법이 달라질 수 있기에(이정찬, 2013), 진단 평가의 목적을 고려하여 설명적 글쓰기의 다양한 문항을 추가로 개발하는 것도 검토할 필요가 있다. 기존의 대규모 평가에서도 설명 대상을 달리함에 따라 나열 구조, 비교-대조 구조, 원인-결과 구조, 문제-해결 구조 등 다양한 조직 방식을 고려한 글쓰기 문항 유형을 활용하고 있다. 특히 설명문을 대규모로 수집하여 자동 채점 인공지능 학습 데이터로 활용하기 위해서는 대학 글쓰기 강좌와 같이 실제 성인의 글쓰기 원자료를 수집하게 되는 글쓰기 맥락과 환경에서의 요구를 반영함으로써 자료 수집의 용이성과 결과의 활용 가능성을 높일 수 있다.

본 연구에서는 개발한 설명적 글쓰기 문항 및 채점 준거의 내용 타당도를 검증하기 위해 전문가 검토를 단계적으로 진행하였다. 향후 ‘예비 평가’를 통해 평가의 신뢰도에 영향을 미칠 수 있는 발문, 난이도, 시간의 충분성 등을 파악할 필요가 있다. 또한 실제 평가를 시행하여 채점 결과 자료를 활용해 문항반응이론에 따라 문항 곤란도 및 변별도 분석, 문항 및 준거 적합도 분석, 총점 및 준거별 점수 상관 분석 등을 분석하여 문항 및 준거의 타당성을 종합적으로 검토하는 연구가 진행될 필요가 있다.

(2) 정교한 문항 분석을 위한 글쓰기 능력 평가 시행 고도화

본 연구에서는 한 명의 학생이 하나의 주제를 선택하여 글을 작성하도록 하여 글쓰기 자료를 수집하였다. 개별 문항에 대해 각기 다른 피험자가 반응하는 이러한 방식은 각기 다른 평가를 실시한 것과 마찬가지로, 문항 간 곤란도나 변별도 차이의 의미를 타당

하게 해석할 수 없게 된다. 또한 대학의 글쓰기 강좌를 기반으로 글 자료를 수집함으로써 동일한 평가 환경에서 글쓰기가 수행되지 않았기에, 같은 문항에 대한 글 자료일지라도 이상치에 해당하는 글쓰기 자료가 수집되는 등 채점 전문성을 흔들며 결과적으로 채점 신뢰도가 하락될 소지가 있었다.

따라서 향후 문항별 글쓰기 자료 특성에 대해 타당하게 비교할 수 있도록 글쓰기 능력 평가 시행을 고도화할 필요가 있다. 자료 수집 환경의 일관성과 안정성을 확보하고, 특정 대학이나 집단에 한정하여 글쓰기 참여자를 표집하여 복수의 문항에 대한 글을 작성하게 하는 평가가 이루어져야 한다. 이를 통해 채점 신뢰도를 높이고, 문항의 영향이나 채점자 효과를 보다 정밀하게 분석하여 조사 도구의 타당화에 기여할 수 있을 것이다.

(3) 논증적 글쓰기의 다양한 평가 상황과 목적에서의 루브릭 활용 방안 마련

본 연구에서 개발한 논증적 글쓰기의 채점 준거는 궁극적으로 국민 다수가 활용하는 것을 목표로 한다. 현재 개발된 진단 지표는 내용 준거가 5개, 조직 준거가 2개, 표현 준거가 2개로서, 총 9개의 채점 준거가 논증적 글쓰기 능력을 구성하는 것으로 설정된 것이다. 준거 개수를 기준으로 삼으면 내용 요소가 상대적으로 조직이나 표현 요소보다 강조된 진단 지표로 볼 수 있고, 준거의 총 개수를 기준으로 삼으면 논술문 평가를 위한 주요 준거를 망라한 총괄형 진단 지표로 볼 수 있다.

그런데 논증적 글쓰기는 과제의 성격이나 상황, 분량의 제한, 응시자의 수준, 문항의 특성 등에 따라 진단의 초점을 두는 부면이 조금씩 상이하다. 가령, 한국의 대학생에게 요구되는 논증적 글쓰기 능력은 논제에 관한 자신의 의견과 생각을 논리적으로 정리하고 풀어내는 것이 중요하다는 점에서 내용이나 조직 범주가 진단의 중심이 되어야 한다. 반면 한국어를 학습하는 외국인 학습자에게 요구되는 논증적 글쓰기 능력은 한국어 소통을 위해 생각을 자연스럽게 구성하고 적절하게 표현하는 것이 주요 학습 목표가 되므로 조직이나 표현 범주가 강조될 필요가 있다. 이렇게 여러 상황과 목적에 따라 채점 준거상의 가중치 비중을 달리 적용한 채점 기준표를 설계함으로써 필요에 맞게 사용할 수 있도록 안내하여야 한다. 여러 종류의 채점 기준표에 대한 상세한 설명과 안내를 통해 국민이 각자의 목적과 상황에 맞는 것을 선택하고, 자신의 논증적 글쓰기 능력을 점검할 수 있도록 하는 일련의 지침이 필요할 것이다.

2) 양질의 글쓰기 자료 추가 구축 및 관리 체계 강화

(1) 다양한 수준의 학습 데이터 구축

정확도와 신뢰도가 모두 높은 고성능의 자동 채점 인공지능을 개발하기 위해서는 양질의 쓰기 채점 데이터의 추가 구축이 필요하다. 1천 자 이상의 논증적 글쓰기는 글쓰기 참여자나 채점자에게 쉽지 않은 과제인데, 마찬가지로 자동 채점 인공지능에게도 본 연구에서 수집한 글쓰기 자료에 대한 채점은 고난도 과제이다. 형태소 분석, 개체명 인식, 구문 분석 등 기존의 다른 자연어 처리 과제가 일반적으로 수만 건 이상의 학습데이터를 학습한 것에 견주면, 1천 자 내외의 논증적 글쓰기 자료에 대한 자동 채점 인공지능을 학습시키기 위해서도 대규모 학습 데이터가 추가적으로 필요하다.

또한 높은 수준의 잘 쓴 글뿐 아니라 논리 전개가 빈약하고, 구성이 미흡한 글 등 다양한 수준의 글쓰기 원자료가 필요하다. 대학생이 쓴 글쓰기 자료를 학습한 인공지능은 대학생의 글을 잘 평가할 수 있지만, 오히려 문해력이 낮은 성인의 글은 평가하기 어려워할 수 있다. 추후 자동채점 인공지능이 적용되는 진단 평가 환경과 목적에 맞는 다양한 수준의 쓰기 데이터를 학습할 수 있어야 한다.

(2) 통제된 조건에서의 글쓰기 자료 수집

본 연구에서 수집하는 글쓰기 원자료는 채점 과정을 거쳐 인공지능 학습용 데이터로 활용될 수 있는 요건을 갖추어야 한다. 글쓰기 과제 요건으로서 인터넷 검색 및 표절을 제한하는 평가 환경에서 작성하여야 하며, 글 분량, 작성 시간, 논제를 준수하고, 제목 및 목차 등 글 본문 외 형식 요소를 포함하지 않는 것을 상정하였다.

그런데 평가 환경의 측면에서 협약 대학의 글쓰기 강의를 기반으로 자료를 수집하는 상황에서, 강의별 커리큘럼에 따라 온라인 자료 활용을 통제하기 어려운 개별 과제 글쓰기의 비율이 절반 이상이 되었다. 결과적으로 인터넷상의 자료를 참고하거나 표절 수준에서 포함하는 글이 적지 않아 채점 일관성 유지를 저해하고, 표절이나 챗지피티 등의 생성 인공지능을 활용한 글을 감별해 내야 한다는 점에서 글쓰기 자료 정제에 대한 부담을 높였다.

향후 대학과의 협약을 통해 글쓰기 자료를 수집하는 안정성과 효율성을 담보하되, 평가 환경 통제에 대해 사전에 충분히 협의할 필요가 있다. 협조가 가능한 강의를 중심으로 자료 수집을 한정하였을 때 수량 확보가 어렵다면 협약 대학을 점차 확대하는 방안도 적극 검토해야 할 것이다. 또 글쓰기 자료 요건 미충족으로 최종 데이터에서 제외되는 수량을 줄이기 위해서도, 미충족의 구체적인 사례를 포함한 매뉴얼을 제작하고 홍보하는 등 대학과의 협력을 보다 강화할 필요가 있다.

(3) 문항별 균형 있는 자료 수집을 위한 시스템 보완

본 연구에서는 인공지능 학습용 데이터로 활용하기 위해서 문항별로 글쓰기 자료를 고르게 수집하는 것을 목표로 하였다. 그런데 문항별로 주제에 대한 선호도가 달라 문항별 글쓰기 자료 수집에서 쏠림 현상이 나타났다. 글쓰기 참여가 원칙적으로 자율이라는 점에서 학습자들의 참여도를 예측하는 것이 어려웠다. 상대적으로 어렵거나 낮은 주제의 문항을 배정받은 글쓰기 참여자들의 참여율이 떨어지는 것을 관찰할 수가 있었다. 대학별 수집 문항이 달랐기 때문에 대학별 글쓰기 참여자의 참여율 편차가 문항별 수집 불균형으로 이어지기도 하였다.

따라서 후속 과제로서 플랫폼상에서 문항별 수집 통제 시스템을 갖추는 것이 필요하다. 특정 문항의 초과분이 발생하는 경우 플랫폼 시스템상 해당 문항을 숨김 처리하여 해당 문항의 참여를 중단하는 것이 가능하다. 또 대학별 수집 상황 모니터링을 위한 협력 시스템을 강화하는 것이 필요하고, 글쓰기 강좌 기반의 수집 경로를 학기 전반부에 운영하며, 학기 후반부에는 부족분만 다양한 통로를 통해 추가 모집하여 수집하는 이원화 방안을 고려해 볼 수 있다. 이는 수집 수량의 예측 가능성을 담보하기 위해서도 필요한 조치이다.

3) 채점자 교육과정 및 운영의 개선

(1) 채점 전문 인력풀 확보를 위한 교육 대상 확대

본 연구에서는 대학생을 대상으로 논증적 글쓰기 자료를 수집하여 채점자 교육 프로그램 참여자 또한 대학의 글쓰기 교강사로 한정하여 모집하였다. 그러나 향후 국민의 글쓰기 능력 진단 체계의 글쓰기 장르 및 글쓰기 참여자 집단의 다양화를 고려하여 채점 전문 인력풀을 확대할 필요가 있다. 대학의 글쓰기 교강사뿐만 아니라 쓰기 및 쓰기 교육 분야의 신진 인력, 국어 교사 등 다양한 집단을 대상으로 채점자 교육 프로그램 참여자를 모집할 수 있다. 그리고 다양한 교육 수요 집단이 이해 가능한 수준에서 이들의 채점 문식성과 채점 실무 전문성을 향상시킬 수 있는 교육 내용을 고안해야 한다.

다양한 인력이 본 프로그램에 참여하도록 하기 위해서는 프로그램을 이수한 후 활용 가능한 영역을 확대하는 것도 고려되어야 한다. 가령 대학의 기초글쓰기 강좌 조교 교육 프로그램 또는 교사 연수 프로그램의 일부로서 본 채점 전문 인력 양성 프로그램을 활용할 수 있도록 하여 채점자 교육에 대한 동기부여와 함께 참여도를 높일 수 있을 것이다. 또한 장기적으로 인증제를 도입하여 글쓰기 채점 전문 인력을 체계적으로 양성 및 관리함으로써 다양한 유형의 채점 과제에 투입 및 활용하는 방안도 모색해 볼 수 있다.

(2) 채점 실무 능력 향상을 위한 교육 내용의 체계화

본 프로그램은 온라인 교육과정을 통해 쓰기 및 쓰기 채점에 관한 문식성을 확보하고, 오프라인 교육과정을 통해 채점 실무 능력을 향상시킬 수 있도록 블렌디드 교육과정을 설계하였다. 온-오프라인 교육 전반에서 채점 실재를 경험할 수 있도록 실제 글쓰기 자료 및 채점 결과를 예시 사례로서 활용하였으나, 채점 전문 인력의 실무 능력 향상을 위해서는 교육 내용에서 채점 실무 부분을 보다 강화하여 체계화할 필요가 있다.

온라인 교육 단계에서부터 채점 실습 과제를 적극 활용하여 참여자별 채점 특성에 대한 피드백을 제공하고, 참여자 간 채점 결과를 비교할 수 있도록 하는 등 채점 실무 중심의 온라인 교육 내용을 설계할 수 있을 것이다. 이를 통해 온라인 교육을 이수한 후 채점 실무 능력의 기준치 이상을 갖추는 것을 기대할 수 있다. 한편 협의를 통한 채점자의 인식 조정이 실제 채점 신뢰도 향상에 미치는 긍정적인 영향을 간과할 수 없는바 집단 워크숍 형태의 오프라인 교육 또한 채점 실무 능력 향상을 위해 필수적이다. 이에 대하여서는 교육 참여자의 피드백을 반영하여 이론과 지식 차원의 교육은 지양하고 채점자 간에 충분하고 적극적인 협의가 이루어질 수 있도록 채점 실제 적용에 교육 시간을 더 할애해야 할 것이다.

(3) 온라인 교육 플랫폼의 지속성과 안정화

프로그램의 개발 차원에서 운영 플랫폼의 지속성과 안정화를 위한 노력이 요구된다. 온라인 교육 플랫폼 이용에 관한 본 프로그램 참여자의 만족도는 매우 높게 나타났으나 네이버 웨일클래스는 본 과업 수행을 위해 올해까지만 한시적으로 이용한 것으로 향후 채점자 교육 프로그램의 지속적 운영을 위해서는 본 프로그램의 이수 관리(동영상 강의 재생률 확인, 퀴즈 풀이 및 평가, 과제 제출 등)에 적합한 플랫폼을 마련하는 것이 우선되어야 한다.

4) 채점 자료의 체계적 구축 및 채점 지원 도구 고도화

(1) 채점 플랫폼 기능 개선을 통한 채점 효율성과 안정성 재고

본 연구를 통해 개발한 글쓰기 자료 수집 및 채점 플랫폼 중 채점 플랫폼은 당면한 대규모 채점 자료 수집에 맞추어 필요한 기능을 최소화하여 개발을 진행하였으며, 이에 따라 실제 채점자의 채점 활동과 직간접적으로 연관성이 높은 활동 중 온라인 채점자 교육 콘텐츠 제공, 채점 워크숍, 연습 채점(사전, 실시간), 채점 진행 알림 및 확인 등의 사항은 기타 온라인 저작 도구 및 플랫폼을 활용하여 진행이 되었다. 또한 채점 플랫폼의 필수적인 기능 중 채점자별 글쓰기 자료 할당이나 채점 그룹 편성, 가교 글쓰기 자료 할당 등의 부분도 플랫폼상 자동화 기능이 없어 수동으로 작업이 이루어졌다. 이와 같은 플랫폼 기능이 수동으로 조작될 경우 대규모로 채점이 진행되는 상황에서 채점 자료가 잘못 할당되거나 채점자 편성의 문제가 발생하는 등 전체 채점 체계 운영에 큰 문제가 발생할 소지가 있다. 따라서 대규모 채점 자료 수집에 관련하여 올해 개발한 채점 플랫폼의 부족한 기능을 개선 및 확장하여 채점 운영 전반의 효율성을 끌어 올릴 필요가 있다.

(2) 자동 채점 인공지능 개발과 연계한 채점 결과 선정 방식 연구

본 연구에서 적용한 채점 결과 자료의 선정 방법은 다국면 라쉬 모형을 통해 각 그룹

별로 가교 채점과 짝 채점으로 나누어 채점 특성을 분석한 결과를 활용하여 이루어졌다. 이 과정에서 같은 그룹이나 짝을 이룬 채점자 사이의 채점 엄격성이나 채점 척도 사용에 대한 비교 가능성을 확보하여 대표 채점 결과를 선정하였으며, 동시에 재채점 결과를 반영하여 일부 이상치에 해당하는 글을 대상으로 채점 결과 선정에 대한 보완을 진행한 바 있다.

그러나 한 편의 글에 대해 한 명의 채점자가 입력한 채점값이 대표 결과로 선정되기 때문에 개별 글쓰기 사례에 대한 특수성이 반영되기 어렵다는 한계가 있었다. 관련하여 채점 자료 선정에 추가로 고려할 수 있는 문항별 특성이나 기타 글 수준 변별, 신뢰도 등 추가적인 채점 특성 관련 정보를 활용하여 선정 작업을 시도하고, 그 결과의 의미를 자동 채점 모형 구축 등을 통하여 확인할 수 있다면 글쓰기 평가 연구 분야에 의미 있는 성과가 될 것으로 예상된다.

(3) 거대 언어모델 기반 자동 채점 인공지능 개발

사람이 논증적 글을 평가할 때, 주장과 근거 사이의 관계가 적절한지, 충분한지 등을 생각(추론)하여 채점하듯이, 추론하는 능력을 갖춘 인공지능 기술 개발이 필요하고, 최근 거대 언어모델(Large Language Model)은 이와 같은 추론 능력이 있다고 평가받고 있다. 물론 아직 거대 언어모델의 추론 능력이 부족하여, 단기적으로는 기존 사전학습 언어 모델 대비 성능이 우수하지 못할 수 있다. 하지만 최근 인공지능 연구가 거대 언어모델을 중심으로 이뤄지는 동향을 고려하면 오픈소스 거대 언어모델 기반의 자동 채점 인공지능도 머지않은 시기에 높은 성능을 달성할 수 있을 것으로 기대한다.

<참고 문헌>

- 과학기술정보통신부·한국지능정보 사회진흥원(2023), 인공지능 학습용 데이터 품질관리 가이드라인 v3.0, <https://aihub.or.kr/aihubnews/qlityguidance/view.do?currMenu=131&topMenu=103&nttSn=10125>
- 국립국어원·서울대학교 국어교육연구소(민병곤, 김종철, 구본관, 김혜정, 손원숙, 박재현, 김승현, 서보영, 차경미, 장지혜, 이효정, 허모아, 강지영, 김동섭)(2018), 2018 국민의 국어능력 실태 조사, 서울: 국립국어원
- 국립국어원·서울대학교 국어교육연구소(민병곤, 남가영, 김선희, 장성민, 권은선, 이은석, 이수진, 김승정, 정현선)(2022), 인공지능 활용 국어능력 진단 체계 개발 기초 연구, 서울: 국립국어원
- 김봉군(2002), 문장기술론(제5판), 서울: 삼영사
- 노은희, 심재호, 김명화, 김재훈(2012), 대규모 평가를 위한 서답형 문항 자동채점 방안 연구, 한국교육과정평가원 연구 보고 RRE 2012-6, 진천군: 한국교육과정평가원
- 노은희, 김명화, 성경희, 김학수, 천가연(2013), 대규모 평가를 위한 서답형 문항 자동채점 프로그램 정교화 및 시범 적용, 한국교육과정평가원 연구 보고 RRE 2013-5, 진천군: 한국교육과정평가원
- 노은희, 이상하, 임은영, 성경희, 박소영(2014), 한국어 서답형 문항 자동채점 프로그램 개발 및 실용성 검증, 한국교육과정평가원 연구 보고 RRE 2014-6, 진천군: 한국교육과정평가원
- 노은희, 송미영, 성경희, 박소영(2015), 한국어 문장 수준 서답형 문항 자동채점 프로그램 개발 및 적용, 한국교육과정평가원 연구 보고 RRE 2015-9, 진천군: 한국교육과정평가원
- 노은희, 송미영, 박종임, 김유향, 이도길(2016), 한국어 문장 수준 서답형 문항 자동채점 프로그램 고도화 개발 및 적용, 한국교육과정평가원 연구 보고 RRE 2016-11, 진천군: 한국교육과정평가원
- 박종임(2013), 쓰기 평가 전문성 신장을 위한 온라인 채점자 훈련 프로그램 개발 방향 연구, 작문연구, 17, 413-440, 한국작문학회
- 박혜영, 이상하, 이명진, 이미영, 주현우, 백은진, 서지희(2021), 평가를 활용한 초·중등학생 글쓰기 능력 신장 방안 연구(Ⅱ), 한국교육과정평가원 연구 보고 RRE 2021-7, 진천군: 한국교육과정평가원
- 이영식(2000), 영어작문 평가에 대한 채점자 훈련의 원리, 영어교육, 55(2), 201-217, 한국영어교육학회
- 이용상, 신동광, 김현정(2022), 한국어 쓰기 평가를 위한 자동채점의 가능성 탐색, 이중언어학, 86, 171-191, 이중언어학회
- 이정찬(2013), 설명적 글쓰기 내용 구성 연구, 서울: 서울대학교
- 장지혜(2016), 대화주의 관점에서의 논증적 글쓰기 교육 연구, 작문연구, 31, 203-228, 한국작문학회
- 최윤곤, 허유리(2013), 한국어 교사의 채점자 훈련 모형 개발, 한국언어문화학, 10(2), 295-316, 국제한국언어문화학회
- 과학기술정보통신부(2022), 인공지능 학습용 데이터 품질관리 가이드라인 v2.0 -품질관리 안내서, 서울: 진한엠앤비
- 홍은실, 오예림, 이성준, 조수진, 안현기, 이승원, 민병곤(2020), 한국어 말하기 평가의 채점 타당도 연구-워크숍 기반의 채점자 훈련을 중심으로, 국제어문, 84, 513-544, 국제어문학회

- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press, 최인철 외 역, 언어테스팅의 설계와 개발, 범문사, 2004, 252-253.
- Bejar, I. I., Williamson, D. M., & Mislevy, R. J. (2006). Human scoring. In I. I. Bejar, D. M. Williamson, & R. J. Mislevy. (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 49-82). Routledge.
- Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine*, 25(3), 27-36. do: 10.1609/aimag.v25i3.1774
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford, UK: Oxford University Press, 145-148.
- Desai, S. & Durrett, G. (2020). Calibration of Pre-trained Transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 295-302). Association for Computational Linguistics.
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155-185.
- Eckes, T. (2012). Operational rater types in writing assessment: Linking rater cognition to rater behavior. *Language Assessment Quarterly*, 9(3), 270-292.
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K.Q. (2017). On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*.
- Haberman, S. J. (2008). When can subscores have value?. *Journal of Educational and Behavioral Statistics*, 33(2), 204-229. do:10.3102/1076998607302636
- Haberman, S. J., & Qian, J. (2007). Linear prediction of a true score from a direct estimate and several derived estimates. *Journal of Educational and Behavioral Statistics*, 32(1), 6-23. do:10.3102/1076998606298036
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.) *Educational Measurement (4th ed)* (pp. 65-110). Westport, CT: Praeger Publishers.
- Herman, J. L. Aschbacher, P. R., & Winters, L. (1992). *A Practical guide to alternative assessment*. Association for Supervision and Curriculum Development, 김경자 역 (2000), 수행평가 과제 제작의 원리와 실제. 이화여자대학교 출판부, 141-162.
- Meng, X.-L. (2020). Reproducibility, Replicability, and Reliability. *Harvard Data Science Review*, 2(4). <https://doi.org/10.1162/99608f92.dbf927f9>
- Mughal, A. Z., & Samreen, A. (2018). Comprehensive Study of Textual Processing and Proposed Automatic Essay Evaluation System. *Journal of Independent Studies and Research Computing*, 16(1), 21-29.
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.

- Purves, A. C. (1984). In search of an internationally valid scheme for scoring compositions. *College Composition And Communication*, 35(4), 426-438.
- Wang, X., Lee, Y., & Park, J. (2022). Automated Evaluation for Student Argumentative Writing: A Survey. *arXiv preprint arXiv:2205.04083*.
- Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 157-180.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge, MA: Cambridge University Press, 정희모 외 역 (2017), 쓰기평가, 글로벌콘텐츠.
- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading?. *Language Testing*, 27(3), 291-300.
- Xi, X., Higgins, D., Zechner, K., & Williamson, D. (2012). A comparison of two scoring methods for an automated speech scoring system. *Language Testing*, 29(3), 371-39.
- Yao, L., Haberman, S. J., & Zhang, M. (2019a). Penalized best linear prediction of true test scores. *Psychometrika*, 1-26. do:10.1007/11336-018-9636-7
- Yao, L., Haberman, S. J., & Zhang, M. (2019b). Prediction of writing true scores in automated scoring of essays by best linear predictors and penalized best linear predictors. *ETS Research Report Series*, 2019(1), 1-27.
- Yoon, S., Evanini, K., & Zechner, K. (2011). Non-scorable response detection for automated speaking proficiency. In *Proceedings of the sixth workshop on innovative use of NLP for building educational applications* (pp. 152-160). Portland, OR: Association for Computational Linguistics.
- Zhang, M. (2013). Contrasting automated and human scoring of essays. *R & D Connections*, 21(2), 1-11.

[부록 1] 채점 매뉴얼 개요

이 연구에서는 글쓰기 진단 지표 고도화 및 글쓰기 채점 전문 인력 양성을 위한 교육 프로그램 운영을 바탕으로 채점 매뉴얼을 개발하였다.

매뉴얼은 1부와 2부로 나누었다. 1부에서는 논증적 글쓰기 평가와 관련된 이론적인 내용을 중심으로 설명하였다. 논증적 글쓰기의 특성과 관련된 평가 요소를 기존의 사례들을 참조하여 서술하였다. 평가에 영향을 미치는 여러 요인들도 함께 제시하였고, 마지막으로 평가 요소별 가중치 부여 방식에 따른 준거별 가중치 모형을 제안하였다. 2부에서는 최종 채점 기준표(<표 II-22>)의 각 준거별 1-5점에 해당하는 글쓰기 자료를 분석한 내용을 중심으로 구성하였다. 본 과업에서 개발한 논증적 글쓰기 채점 기준표의 채점 준거와 특성을 소개하고, 이를 실제 글쓰기 자료에 적용한 사례를 해설과 함께 제시하였다.

1부는 채점자 교육 프로그램의 개발과 운영에 참여한 교수자 가운데 1인을 섭외하여, 2부는 채점 도구의 정교화(채점자 교육용 자료 제작을 위한 예시 글쓰기 자료 선정, 표현 범주 채점 준거 정량화)에 참여한 전문가 가운데 3인을 섭외하여 개발을 의뢰하였다. 채점 매뉴얼의 목차와 주요 내용은 <표>와 같다.

<표> 채점 매뉴얼의 목차와 주요 내용

목차	주요 내용
1부: 논증적 글쓰기 평가	
1. 논증적 글쓰기의 정의 및 특성	· 논증적 글쓰기에 대한 설명 · 대규모 쓰기 평가 문항의 특징
2. 논증적 글쓰기의 평가 요소	· ‘내용/조직/표현’ 범주와 하위 평가 요소 · 국내외 유관 평가 사례 참조 및 정리
3. 논증적 글쓰기 문항의 과제 요구에 따른 영향	· 논제 유형(사실/가치/정책 논제)에 따른 평가의 강조점 차이 · 글쓰기 문항의 자료 제시 유형에 따른 차이 · 글쓰기 분량에 따른 평가의 강조점 차이
4. 평가 범주별 가중치 부여 방식	· 균형적 채점 방식, 내용 중시 채점 모형, 표현 중시 채점 모형

2부: 논증적 글쓰기 채점의 실제

1. 채점 기준표 소개

- <표 II-22>의 9개 준거 및 유의 사항에 대한 설명
- 본 과업 기준표의 특수성과 제한점에 대한 안내(논제 특성, 분량, 응시 환경, 대상자 등)
- 2023년 시행한 문항 제시

2. 글쓰기 답안 채점 사례

- 내용, 조직, 표현의 총 9개 준거에 대하여 2명의 채점자가 동일한 점수를 부여한 준거별, 점수대별 글쓰기 자료 선정(1, 5점은 1편씩, 2-4점은 2편씩 / 총 72편)
- 점수대별 채점을 위한 글쓰기 자료 해설

[부록 2] 저작권 이용 허락 동의서 양식

연구참여자용 설명문(대학생용)

연구 과제명 : 2023년 국민의 글쓰기 능력 진단 체계 개발

연구 책임자명 : 민병곤(서울대학교 사범대학 국어교육과 교수)

이 연구는 국민의 글쓰기 능력 향상을 도모할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있습니다. 귀하는 국립국어원과 업무 협약을 맺은 9개 국립대학과 인근 대학의 재학생(휴학생 포함)으로, 이 연구에 참여하도록 권유 받았습니니다. 이 연구를 수행하는 서울대학교 소속의 정은선 연구원(010-4210-8642)이 귀하에게 이 연구에 대해 설명해 줄 것입니다. 이 연구는 자발적으로 참여 의사를 밝히신 분에 한하여 수행될 것이며, 귀하께서는 참여 의사를 결정하기 전에 본 연구가 왜 수행되는지 그리고 연구의 내용이 무엇과 관련 있는지 이해하는 것이 중요합니다. 다음 내용을 신중히 읽어보신 후 참여 의사를 밝혀 주시길 바라며, 필요하다면 가족이나 친구들과 의논해 보십시오. 만일 어떠한 질문이 있다면 담당 연구원이 자세하게 설명해 줄 것입니다.

1. 이 연구는 왜 실시합니까?

이 연구의 목적은 국민의 글쓰기 능력 향상을 도모할 수 있는 효과적이고 신뢰할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있습니다. 제출한 자료는 글쓰기 능력 진단 체계를 고도화하고, 인공지능 활용 자동 채점 시스템을 개발하는 데 활용될 것입니다.

2. 얼마나 많은 사람이 참여합니까?

글쓰기 자료 수집 참여자 최대 9천 명과 채점자 최대 120 명의 사람이 참여할 것입니다.

3. 만일 연구에 참여하면 어떤 과정이 진행됩니까?

귀하는 90분 동안 1000자 내외의 글을 작성하는 글쓰기를 수행하게 됩니다. 해당 온라인 플랫폼을 통해 쓰기 문항과 유의사항을 확인하고, 주어진 시간 내에 한 편의 완결된 글을 작성하여 제출하면 됩니다.

만약 귀하가 소속 대학의 글쓰기 강좌의 교육과정의 일환으로 국립국어원에서 제공한 글쓰기 능력 진단 체계 개발용 문항에 대한 글을 사전에 작성하였다면, 해당 글을 제출하면 됩니다. 다른 목적으로 작성한 글은 제출할 수 없습니다.

4. 연구 참여 기간은 얼마나 됩니까?

글쓰기 자료 수집 기간은 IRB 승인 이후부터 12월 15일까지 지속됩니다. 귀하는 이 기간 중 귀하가 속한 대학과 국립국어원의 업무 협약에 따라 결정된 참여 기간 또는 연구진이 안내한 참여 기간에 따라, 귀하의 일정을 고려하여 연구에 참여할 수 있습니다. 참여 기간은 귀하가 속한 대학에 게시될 홍보물(포스터 및 리플릿)에서 확인할 수 있고, 귀하가 속한 대학의 글쓰기 자료 수집 코디네이터 또는 본 과제의 연구원 정은선을 통해서도 확인할 수 있습니다.

5. 참여 도중 그만두어도 됩니까?

예, 귀하는 언제든지 어떠한 불이익 없이 참여 도중에 그만둘 수 있습니다. 만일 귀하가 연구에 참여하는 것을 그만두고 싶다면 글을 작성하는 도중이라도 중단하고 제출하지 않으시면 됩니다.

6. 부작용이나 위험 요소는 없습니까?

본 연구에 참여함으로써 예상되는 부작용이나 위험 요소는 없습니다. 본 연구의 결과는 연구 목적으로만 활용될 것이며, 개별 결과는 공개되지 않습니다. 제출한 글쓰기 자료는 쓰기 자동 채점 인공지능 시스템을 개발하기 위하여 활용됩니다. 인공지능 학습 데이터로 가공하기 전에 제출한 글쓰기 자료에서 개인을 식별할 수 있는 정보를 탐지하여 삭제하는 단계를 수행하게 되고, 글은 자동 채점 기술을 개발하는 목적으로만 사용되므로 내용이 외부로 노출되지 않습니다. 다만 예상하지 못한 위험을 대비하기 위해 개인을 식별할 수 있는 정보를 글 내용에 포함하지 않도록 합니다. 만일 연구 참여 도중 발생할 수 있는 부작용이나 위험 요소에 대한 질문이 있으시면 담당 연구원에게 즉시 문의해 주십시오.

7. 이 연구에 참여 시 참여자에게 이득이 있습니까?

귀하가 이 연구에 참여하는 데 있어서 직접적인 이득은 없습니다. 다만 귀하가 제출한 글은 국민의 글쓰기 능력을 증진하는 데 기여할 진단 체계를 개발하는 데 활용되는 귀중한 자료가 될 것입니다.

8. 만일 이 연구에 참여하지 않는다면 불이익이 있습니까?

귀하는 본 연구에 참여하지 않을 자유가 있습니다. 또한, 귀하가 본 연구에 참여하지 않아도 귀하에게는 어떠한 불이익도 없습니다.

9. 연구에서 얻은 모든 개인 정보의 비밀은 보장됩니까?

개인정보관리책임자는 서울대학교의 사범대학 국어교육과의 민병곤 교수(연구책임자,

02-880-7665)입니다. 본 연구에서 수집되는 개인정보는 이름, 이메일, 전화번호, 직업(대학생), 성별, 연령(출생연도), 학년, 소속 학교, 전공 계열, 한국 국적 여부입니다. 모든 개인정보는 암호화된 파일로 관리될 것입니다. 이 개인정보 중 귀하의 이름, 이메일과 전화번호는 사례비 지급을 위한 목적으로 수집하는 정보이며, 사례비 지급이 완료된 이후에 파기됩니다. 귀하의 이름, 이메일, 전화번호에 접근할 수 있는 사람은 연구책임자와, 예산 처리를 위한 연구담당자 3명(정은선, 이수진, 민소연 연구원)으로 한정됩니다.

이름, 이메일, 전화번호를 제외한 직업(대학생), 성별, 연령(출생연도), 학년, 소속 학교, 전공 계열, 한국 국적 여부 정보는 글쓰기 자료를 인공지능의 학습 데이터로 가공하는 단계에서 데이터 특성을 설명하는 메타 정보로 데이터에 부착됩니다. 이에 따라 메타 정보에 해당되는 정보는 공용 드라이브(드롭박스)에 암호화된 개별 파일로 저장되며, 해당 드라이브에는 개인정보 관리책임자가 권한을 부여한 연구진(민병곤, 남가영, 김선희, 장성민, 이성준, 권은선, 정은선)만 접근할 수 있습니다. 이 연구 자료는 서울대 연구윤리지침(2020. 9. 21. 일부 개정)에 따라 최소 5년 이상, 가능한 한 영구 보관할 예정입니다(민병곤 교수 연구실 10동 419호).

저희는 이 연구를 통해 얻은 모든 개인 정보의 비밀 보장을 위해 최선을 다할 것입니다. 이 연구에서 얻어진 개인 정보가 학회지나 학회에 공개될 때(수집되는 자료를 학술대회 발표 및 국제·국내 학술지 게재 혹은 학위논문 게재 등) 귀하의 이름 및 기타 개인정보는 사용되지 않을 것입니다. 그러나 만일 법이 요구하면 귀하의 개인정보는 제공될 수도 있습니다. 또한 모니터 요원, 점검 요원, 생명윤리위원회는 연구 참여자의 개인 정보에 대한 비밀 보장을 침해하지 않고 관련 규정이 정하는 범위 안에서 본 연구의 실시 절차와 자료의 신뢰성을 검증하기 위해 연구 결과를 직접 열람할 수 있습니다. 귀하가 본 동의서에 서명하는 것은, 이러한 사항에 대하여 사전에 알고 있었으며 이를 허용한다는 동의로 간주될 것입니다. 동의서는 관련 법령에 따라 3년을 보관한 후 폐기할 예정이며, 연구 자료의 경우는 서울대학교 연구윤리지침에 따라 가능한 한 영구 보관할 예정입니다.

10. 이 연구에 참가하면 사례가 지급됩니까?

귀하의 연구 참여 시 감사의 뜻으로, 연구비를 통해 2만 원이 지급됩니다.

11. 연구에 대한 문의는 어떻게 해야 됩니까?

본 연구에 대해 질문이 있거나 연구 중간에 문제가 생길 시 다음 연구 담당자에게 연락하십시오.

이름: _____ 전화번호: _____

만일 어느 때라도 연구참여자로서 귀하의 권리에 대한 질문이 있다면 다음의 서울대학교 생명윤리위원회에 연락하십시오.

서울대학교 생명윤리위원회 (SNUIRB) 전화번호: 02-880-5153 이메일: irb@snu.ac.kr

동 의 서

연구 과제명 : 2023년 국민의 글쓰기 능력 진단 체계 개발

연구 책임자명 : 민병곤(서울대학교 사범대학 국어교육과 교수)

1. 나는 이 설명서를 읽었으며 담당 연구원과 이에 대하여 의논하였습니다.
2. 나는 위험과 이득에 관하여 들었으며 나의 질문에 만족할 만한 답변을 얻었습니다.
3. 나는 이 연구에 참여하는 것에 대하여 자발적으로 동의합니다.
4. 나는 이 연구에서 얻어진 나에 대한 정보를 현행 법률과 생명윤리위원회 규정이 허용하는 범위 내에서 연구자가 수집하고 처리하는 데 동의합니다.
5. 나는 담당 연구자나 위임 받은 대리인이 연구를 진행하거나 결과 관리를 하는 경우와 법률이 규정한 국가 기관 및 서울대학교 생명윤리위원회가 실태 조사를 하는 경우에는 비밀로 유지되는 나의 개인 신상 정보를 확인하는 것에 동의합니다.
6. 나는 언제라도 이 연구의 참여를 철회할 수 있고 이러한 결정이 나에게 어떠한 해도 되지 않을 것이라는 것을 압니다.
7. 나는 수집되는 자료가 본 연구 이외에 연구책임자 및 다른 연구자의 연구 목적(수집되는 자료를 학술대회 발표 및 국제·국내 학술지 게재 혹은 학위논문 게재 등)으로 사용되는 것에 동의합니다.

동의함 동의하지 않음

본 연구에 참여하시겠습니까?

동의함 동의하지 않음

연구참여사용 설명문(채점 참여사용)

연구 과제명 : 2023년 국민의 글쓰기 능력 진단 체계 개발

연구 책임자명 : 민병곤(서울대학교 사범대학 국어교육과 교수)

이 연구는 국민의 글쓰기 능력 향상을 도모할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있습니다. 귀하는 국립국어원 업무 협약 대학 또는 권역별 인근 대학에서 글쓰기 교육 강좌의 교강사로 재직하고 있고, 채점자 교육 프로그램을 이수한 전문가이기에 이 연구에 쓰기 자료 채점자로 참여하도록 권유받았습니다. 이 연구를 수행하는 서울대학교 소속의 오예림 연구원(010-4511-7419)이 귀하에게 이 연구에 대해 설명해 줄 것입니다. 이 연구는 자발적으로 참여 의사를 밝히신 분에 한하여 수행될 것이며, 귀하께서는 참여 의사를 결정하기 전에 본 연구가 왜 수행되는지 그리고 연구의 내용이 무엇과 관련 있는지 이해하는 것이 중요합니다. 다음 내용을 신중히 읽어보신 후 참여 의사를 밝혀 주시길 바랍니다. 만일 어떠한 질문이 있다면 담당 연구원을 통해 자세한 설명을 들을 수 있습니다.

1. 이 연구는 왜 실시합니까?

이 연구의 목적은 국민의 글쓰기 능력 향상을 도모할 수 있는 효과적이고 신뢰할 수 있는 글쓰기 능력 진단 체계를 개발하는 데 그 목적이 있습니다. 채점 자료는 글쓰기 능력 진단 체계를 고도화하고, 인공지능 활용 자동 채점 시스템을 개발하는 데 활용될 것입니다.

2. 얼마나 많은 사람이 참여합니까?

최대 120 명이 참여할 것입니다.

3. 만일 연구에 참여하면 어떤 과정이 진행됩니까?

본 연구에 참여하는 채점자는 채점 신뢰도를 높이기 위한 온라인 워크숍(1일, 6시간 예정)에 참여하도록 요청을 받습니다. 워크숍에서는 채점 도구 정교화 및 채점자 교육 프로그램 개선을 위한 자료로 활용하기 위해 활동 과정을 녹화 후 전사하여 문서로 저장하고, 온라인 채팅 기록도 저장됩니다. 이후 채점 플랫폼에 접속하여 개별적으로 할당된 1000자 내외의 글 자료 101-102건을 채점하게 됩니다. 채점자별 채점 건수는 수집된 글쓰기 자료의 수와 채점자 모집 인원 수에 따라 일부 변경될 수 있습니다. 개별 채점 과정은 온라인 채점 플랫폼에서 진행되며, 연구진으로부터 제공 받은 채점 척도를 활용하여 채점을 수행합니다. 채점 진행 중에 협의가 필요한 사항은 플랫폼의 문의 기능을 이용하여 연구진과 협의할 수 있습니다.

4. 연구 참여 기간은 얼마나 됩니까?

글쓰기 자료 채점은 1차 채점(8월), 2차 채점(10월)의 두 기간 동안 진행되므로, 귀하의 일정을 고려하여 가능한 기간에 참여하면 됩니다. 채점에 참여하기 전에, 채점의 신뢰도를 높이기 위하여 1일(6시간) 온라인 워크숍에 참여하게 됩니다. 연구에 참여하는 경우 귀하가 채점할 글 자료는 채점자 100명 모집 기준 101-102건입니다.

5. 참여 도중 그만두어도 됩니까?

예, 귀하는 언제든지 어떠한 불이익 없이 참여 도중에 그만둘 수 있습니다. 그만두는 경우 귀하가 입력한 점수는 폐기되며, 연구 자료로 사용되지 않습니다. 다만 추가 채점자를 충원해야 하므로 연구 담당자에게 해당 사실을 알려 주시면 감사하겠습니다.

6. 부작용이나 위험 요소는 없습니까?

본 연구의 채점에 참여함으로써 예상되는 부작용이나 위험 요소는 없습니다. 제출한 채점 결과는 쓰기 자동 채점 인공지능 시스템을 개발하기 위하여 활용됩니다. 인공지능 학습 데이터에는 귀하가 제공하는 채점 결과 점수만 포함되므로, 인공지능 시스템 개발 과정에서 귀하에게 미칠 부작용이나 위험 요소는 없습니다.

7. 이 연구에 참여 시 참여자에게 이득이 있습니까?

귀하가 이 연구에 채점자로 참여하는 데 있어서 직접적인 이득은 없습니다. 다만 귀하의 채점 결과는 국민의 국어능력을 증진하는 데 기여할 진단 체계와 쓰기 자동 채점 프로그램을 개발하는 데 활용되는 귀중한 자료가 될 것입니다.

8. 만일 이 연구에 참여하지 않는다면 불이익이 있습니까?

귀하는 본 연구에 참여하지 않을 자유가 있습니다. 또한, 귀하가 본 연구에 참여하지 않아도 귀하에게는 어떠한 불이익도 없습니다.

9. 연구에서 얻은 모든 개인 정보의 비밀은 보장됩니까?

개인정보관리책임자는 서울대학교의 사범대학 국어교육과의 민병곤 교수(연구책임자, 02-880-7665)입니다. 본 연구에서는 채점자 선정 및 채점 자료 할당의 목적으로 소속, 전공, 교육 경력, 평가 경력, 평가 교육 이력을 수집하고, 사례비 지급을 위한 목적으로 성명, 주민등록번호와 계좌번호를 수집하며, 귀하와 연락을 취하기 위한 목적으로 이메일과 전화번호를 수집합니다. 모든 개인정보는 암호화된 파일로 관리될 것입니다. 채점자 선정 및 채점 자료 할당 목적으로 수집한 정보는 채점 수행 종료 이후에 파기하며, 사례비 지급을 위한 목적으로 수집하는 개인정보는 사례비 지급 이후 모두 파기됩니다. 연락을 취하기 위한 목적으로 수집하는 개인 정보는 수집 목적을 달성한 이후 파기됩니다. 이 모든 개인정보에 접근할 수

있는 사람은 연구책임자와 담당 연구진(이성준, 오예림)으로 한정됩니다.

채점 데이터를 관리하기 위한 목적의 개인 정보와 채점 과정에서 발생한 자료, 즉 채점 결과 자료, 채점자 워크숍 녹화 및 전사 자료, 채점자 온라인 채팅 기록 자료는 자동 채점 인공지능을 개발하는 연구 과정에 사용됩니다. 이 자료들은 공용 드라이브(드롭박스)에 암호화된 개별 파일로 저장되며, 해당 드라이브에는 개인정보 관리책임자가 권한을 부여한 연구진(민병곤, 남가영, 김선희, 장성민, 이성준, 권은선, 오예림)만 접근할 수 있습니다. 개인 식별 정보가 제외된 연구 자료는 서울대 연구윤리지침(2020. 9. 21. 일부 개정)에 따라 최소 5년 이상, 가능한 한 영구 보관할 예정입니다(민병곤 교수 연구실 10동 419호).

저희는 이 연구를 통해 얻은 모든 개인 정보의 비밀 보장을 위해 최선을 다할 것입니다. 이 연구에서 얻어진 개인 정보가 학회지나 학회에 공개될 때(수집되는 자료를 학술대회 발표 및 국제·국내 학술지에 게재 혹은 학위논문 게재 등) 귀하의 이름 및 기타 개인정보는 사용되지 않을 것입니다. 그러나 만일 법이 요구하면 귀하의 개인정보는 제공될 수도 있습니다. 또한 모니터 요원, 점검 요원, 생명윤리위원회는 연구 참여자의 개인 정보에 대한 비밀 보장을 침해하지 않고 관련 규정이 정하는 범위 안에서 본 연구의 실시 절차와 자료의 신뢰성을 검증하기 위해 연구 결과를 직접 열람할 수 있습니다. 귀하가 본 동의서에 서명하는 것은, 이러한 사항에 대하여 사전에 알고 있었으며 이를 허용한다는 동의로 간주될 것입니다. 동의서는 관련 법령에 따라 3년을 보관한 후 폐기할 예정이며, 연구 자료의 경우는 서울대학교 연구윤리 지침에 따라 가능한 한 영구 보관할 예정입니다.

10. 이 연구에 참가하면 사례가 지급됩니까?

귀하의 연구 참여 시 감사의 뜻으로 연구비를 통해, 온라인 워크숍 참여 수당 20만 원, 채점 건당 1만 2천 원의 사례비가 지급됩니다.

11. 연구에 대한 문의는 어떻게 해야 됩니까?

본 연구에 대해 질문이 있거나 연구 중간에 문제가 생길 시 다음 연구 담당자에게 연락하십시오.

이름: _____ 전화번호: _____

만일 어느 때라도 연구참여자로서 귀하의 권리에 대한 질문이 있다면 다음의 서울대학교 생명윤리위원회에 연락하십시오.

서울대학교 생명윤리위원회 (SNUIRB) 전화번호: 02-880-5153 이메일: irb@snu.ac.kr

동 의 서

연구 과제명 : 2023년 국민의 글쓰기 능력 진단 체계 개발

연구 책임자명 : 민병곤(서울대학교 사범대학 국어교육과 교수)

1. 나는 이 설명서를 읽었으며 담당 연구원과 이에 대하여 의논하였습니다.
2. 나는 위험과 이득에 관하여 들었으며 나의 질문에 만족할 만한 답변을 얻었습니다.
3. 나는 이 연구에 참여하는 것에 대하여 자발적으로 동의합니다.
4. 나는 이 연구에서 얻어진 나에 대한 정보를 현행 법률과 생명윤리위원회 규정이 허용하는 범위 내에서 연구자가 수집하고 처리하는 데 동의합니다.
5. 나는 담당 연구자나 위임 받은 대리인이 연구를 진행하거나 결과 관리를 하는 경우와 법률이 규정한 국가 기관 및 서울대학교 생명윤리위원회가 실태 조사를 하는 경우에는 비밀로 유지되는 나의 개인 신상 정보를 확인하는 것에 동의합니다.
6. 나는 언제라도 이 연구의 참여를 철회할 수 있고 이러한 결정이 나에게 어떠한 해도 되지 않을 것이라는 것을 압니다.
7. 나는 연구를 수행하는 중에 온라인 채점 워크숍을 녹화하여 전사한 문서를 저장하고, 채점자 간 협의를 위한 채팅 기록을 저장하는 것에 동의합니다.
8. 나는 수집되는 자료가 본 연구 이외에 연구책임자 및 다른 연구자의 연구 목적(수집되는 자료를 학술대회 발표 및 국제·국내 학술지에 게재 혹은 학위논문 게재 등)으로 사용되는 것에 동의합니다.

동의함 동의하지 않음

본 연구에 참여하시겠습니까?

동의함 동의하지 않음

<기획·연구>

국립국어원 황용주 학예연구관
박미영 학예연구사

<연구 참여자>

연구책임자 민병곤(서울대학교)
공동연구원 남가영(서울대학교)
김선희(서울대학교)
장성민(인하대학교)
이성준(서울대학교)
권은선(서울대학교)
연구보조원 오예림(서울대학교)
이수진(서울대학교)
정은선(서울대학교)
김동섭(서울대학교)
송규민(서울대학교)
민소연(서울대학교)
홍태경(서울대학교)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2023년 12월 15일

발행일: 2023년 12월 15일

인 쇄: 가람문화사

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘2023년 국민의 글쓰기 능력 진단 체계 개발’ 사업의 결과물을 발간한 것입니다.