컴퓨터교육학회 논문지 2026년 제29권 제8호
https://doi.org/10.32431/kace.2026.29.8.006

초등학생의 주체적 AI 리터러시 함양을 위한 SPARK-L 교수학습 모델 및 주체적 AI 리터러시 검사도구 개발과 내용타당화*

Development and Content Validation of the SPARK-L Teaching-Learning Model and a Self-directed AI Literacy Assessment Tool for Cultivating Self-directed AI Literacy in Elementary Students

김성록† · 허경††
Seongrok Kim† · Kyeong Hur††

요약

생성형 인공지능의 일상화는 초등학생에게 사고의 외주화와 AI 의존성이라는 교육적 위협을 제기한다. 본 연구는 초등 5~6학년 국어과 AI 융합 수업 맥락에서 학습자의 주체적 AI 리터러시를 함양하기 위한 SPARK-L 교수학습 모델과 주체적 AI 리터러시 검사도구를 개발하고, 전문가 델파이 조사를 통해 그 내용타당도를 검증하는 데 목적이 있다. SPARK-L은 AI Literacy(문해력) 향상을 위한 Situation(상황), Posing Question(질문 구체화), AI-assisted Inquiry(AI 협력 탐구), Reasoning(논리적 구조화), Knowledge Production(지식 창출)의 5단계로 구성되고, R(Reasoning) 단계의 주도권 전환 기제를 통해 인지 처리를 외부 도구에 위임하는 인지적 오프로딩을 통제하도록 설계되었다. 9명 전문가 패널을 대상으로 4차 델파이 조사를 수행하고 Lawshe의 CVR(≥.78)을 합의 기준으로 적용한 결과, 5단계 흐름과 6개 평가 체계 구성요소, 검사도구 10개 문항이 모두 합의에 도달하여 내용타당도가 확인되었다. 본 연구는 AI 협력과 학생 주도성의 균형 및 평가 책무성을 통합한 AI 리터러시 교육의 기초 자료를 제공하고자 한다.

주제어 주체적 AI 리터러시, 질문 기반 탐구학습, 델파이 조사, 인지적 오프로딩, SPARK-L 모델

ABSTRACT

The pervasive integration of generative AI into elementary education raises concerns about the outsourcing of thinking and AI dependency among young learners. This study develops and examines the content validity of the SPARK-L teaching-learning model and a self-directed AI literacy assessment tool for cultivating self-directed AI literacy among 5th- and 6th-grade students in Korean language arts. To improve AI literacy, SPARK-L comprises five stages — Situation, Posing Question, AI-assisted Inquiry, Reasoning, and Knowledge Production — and incorporates an agency-transfer mechanism at the Reasoning stage to counteract cognitive offloading. A four-round Delphi study with nine experts was conducted using Lawshe's Content Validity Ratio (CVR ≥ .78) as the consensus threshold. All five instructional stages, six assessment-system components, and ten assessment items reached consensus. The study provides foundational material for AI literacy education that embeds the balance between AI collaboration and student agency and integrates assessment accountability as a core operating principle.

Keywords Self-directed AI Literacy, Question-Based Inquiry Learning, Delphi Study, Cognitive Offloading, SPARK-L Model

†정회원 경인교육대학교 교육전문대학원 컴퓨터교육과 박사과정
††종신회원 경인교육대학교 컴퓨터교육과 교수 (교신저자)

논문투고 2026년 05월 11일
심사완료 2026년 06월 14일
게재확정 2026년 06월 30일
발행일자 2026년 08월 13일

* 본 논문은 2026년 경기도교육청 AI디지털 활용 연구학교의 지원을 받아 수행된 연구임.

1. 서론

생성형 인공지능의 급격한 발달은 사회 전반을 넘어 교육 생태계에 근본적인 변화를 촉발하고 있다[1]. 챗GPT(ChatGPT)로 대표되는 생성형 AI는 단순한 정보 검색 도구를 넘어, 방대한 데이터를 바탕으로 텍스트, 이미지, 코드 등 새로운 콘텐츠를 창출해 내는 지적 파트너의 위치로 격상되었다. 초등 교육 현장에서도 에듀테크를 활용한 개별화 맞춤형 학습과 AI 융합 교육이 가속화되고 있으며, 이는 학습자의 학업 성취도와 흥미를 높이는 긍정적인 기회를 제공하고 있다[2].

그러나 이러한 AI의 일상화 이면에는 교육적 위협이 공존한다. 생성형 AI가 제공하는 완성도 높은 결과물은 학습자가 스스로 수행해야 할 사고 활동을 AI에 떠넘기는 사고의 외주화라는 교육적 문제 상황을 초래할 수 있다. 이러한 문제 상황은 인지 처리를 외부 도구에 위임하는 현상을 설명하는 인지 과학의 인지적 오프로딩(cognitive offloading) 개념으로 이해할 수 있다[3]. 특히 초등학생은 추상적 사고와 비판적 정보 판단 능력이 발달하는 과도기에 있어, AI가 생성한 결과물에 내재된 환각 현상(hallucination)이나 편향성을 검증하지 못한 채 그대로 수용할 가능성이 상대적으로 높다[4-6].

이는 초등 교육 맥락에서 AI 활용 교육이 기술적 사용법을 넘어 학습자의 주체적 사고를 보호하는 방향으로 설계되어야 함을 시사한다. 따라서 현재 초등 교육의 시급한 과제는 인공지능에 종속되지 않고 이를 주도적으로 통제하며 자신의 학습 목적에 맞게 활용하는 주체적 AI 리터러시를 함양하는 것이다[7, 8]. 본 연구에서 주체적 AI 리터러시는 AI를 학습 목적에 맞게 활용하고, AI가 제시한 결과를 비판적으로 검토하며, 이를 자신의 논리로 재구성하는 역량으로 정의한다.

이와 같은 교육적 요구에 부응하기 위해 본 연구는 질문에 주목하였다. 질문 기반 탐구학습을 핵심 접근으로 설정한 까닭은, 학습자가 스스로 생성한 질문이 탐구의 전 과정에서 주도성을 유지하는 출발점이 되기 때문이다. 학습자는 질문을 통해 해결할 문제를 인식하고, 그 질문을 기준으로 AI에 요청할 정보의 탐색 방향을 설정한다. 이후 AI가 제시한 답변이 질문에 부합하는지 비판적으로 검증하고, 검증한 정보를 자신의 논리로 재구성하여 새로운 지식을 구성한다[9]. 즉 질문은 AI 활용의 각 국면에서 학습자가 사고의 주체로 남도록 하는 장치로 기능한다.

이에 본 연구는 초등학생이 주도적으로 문제를 발견하고 AI와 상호작용하며 결과물을 도출할 수 있도록 돕는 SPARK-L(Situation-Posing question-AI assisted inquiry -Reasoning-Knowledge production for Literacy) 교수학습 모델을 개발하고 그 내용타당도를 검증하고자 한다. 본 모델은 2026학년도 경기도교육청 지정 A 초등학교 AI디지털 연구학교 운영에서 국어과 적용을 위해 구안된 초기 모델이다. 본 모델이 학술적으로 의미 있는 교수학습 모델로 자리 잡기 위해서는 모델 자체의 타당성을 확보하는 동시에 이를 통해 길러지는 학습자의 역량을 측정할 도구가 함께 요구된다.

최근 AI 리터러시를 정량적으로 측정하기 위한 척도 개발 연구가 활발히 진행되고 있으나[10-13], 초등 학습자의 수업 맥락과 SPARK-L 모델의 특성에 특화된 측정도구는 여전히 부재한 실정이다. 따라서 본 연구의 목적은 SPARK-L 교수학습 모델과 이를 통해 발현되는 학습자의 변화를 진단할 주체적 AI 리터러시 검사도구를 개발하고, 전문가 델파이(Delphi) 조사를 통해 두 산출물의 내용타당도를 검증하는 데 있다. 이를 통해 이론적 정합성과 현장 적용 가능성을 고려한 인공지능 융합 교육의 수업 설계 및 평가를 위한 기초 자료를 제공하고자 한다.

2. 이론적 배경

2.1 AI융합 교수학습 모델의 동향

탐구학습은 학습자가 능동적으로 질문하고 정보를 탐색하며 지식을 구성하는 과정으로, 20세기 후반부터 다양한 교수학습 모델로 발전해 왔다[14]. 그러나 생성형 AI의 등장은 탐구의 전제를 변화시키고 있다. 종래의 탐구학습은 학습자가 탐구의 인지 과정을 대부분 직접 수행한다는 전제 위에 설계되었으나, 생성형 AI는 정보 탐색과 요약, 분석 등 탐구의 상당 부분을 대신 수행하거나 보조할 수 있다. 이에 따라 기존 탐구학습 모델이 AI를 학습 과정에 어떻게 위치시킬 것인지를 재검토할 필요가 제기되며, 이에 대한 합의된 답은 아직 마련되지 않은 상태이다.

본 연구는 SPARK-L 모델의 설계 방향을 도출하기 위해 탐구학습과 AI 교육 분야에서 널리 참조되는 네 가지 사례를 분석 대상으로 선정하였다. 구체적으로 단계적 탐구 구조의 정교함을 보여 주는 5E(Engage, Explore, Explain, Elaborate, Evaluate) 학습순환[15], 학생 주도성과 실제 문제 해결을 강조하는 문제기반학습[16], AI를 학습 내용으로 다루는 AI4K12 빅 아이디어[17] 그리고 본 모델의 직접적 모태가 된 경기도교육청 질문기반 탐구수업이다[14].

다만 5E와 PBL, 경기도교육청 질문기반 탐구수업이 단계 구조를 갖춘 교수학습 모델인 반면, AI4K12는 단계별 수업 설계가 아니라 AI 교육의 내용 체계를 제시하는 프레임워크라는 점에서 층위가 다르다. 본 연구는 이러한 차이를 전제로, AI를 학습 과정에 통합하는 관점을 검토하기 위해 AI4K12를 분석에 포함하였다. 분석은 단계 구조, AI와 도구의 역할, 학생 주도성, 평가 방식, 교과 적용 맥락의 다섯 개 축으로 수행하였다.

분석 결과는 다음과 같다. 5E 모델은 Engage부터 Evaluate까지 다섯 단계가 명료하게 구조화되어 있어 교사의 수업 설계에 도식적 틀을 제공한다[15]. 문제기반학습(PBL)은 비구조화된 실생활 문제를 출발점으로 삼아 학생의 자기주도성을 폭넓게 보장하며, 학습자의 인지 및 메타

인지 발달을 다루는 탐구 모델로 평가된다[16]. 경기도교육청 질문기반 탐구수업은 학생 질문 형성을 독립된 영역으로 다루며, 구조화에서 안내 개방형에 이르는 3단계 주도성 스펙트럼을 통해 현장 적용의 유연성을 확보하였다[14]. 한편 AI4K12는 AI 자체를 학습 내용으로 삼아 인식, 표현, 학습, 상호작용, 사회적 영향이라는 다섯 가지 빅 아이디어를 제시함으로써 K-12 AI 교육의 내용 체계를 구축하였다[17].

본 연구에서 검토한 네 사례에서는 AI를 학습 파트너로 명시적으로 통합하면서 동시에 학생 주도성을 단계별로 설계한 경우를 찾기 어려웠다. 5E와 PBL, 경기도교육청 모델은 AI를 모델 안에 포함하지 않거나 일반적 디지털 도구의 한 종류로만 다루며, AI4K12는 AI를 학습 대상으로 삼되 단계별 교수학습 설계가 아닌 내용 프레임워크에 머문다. 이는 AI 시대의 탐구학습 모델이 갖추어야 할 두 요건, 곧 AI와의 협력적 학습 경험 제공과 학생의 주체적 사고 보호를 동시에 반영한 SPARK-L 설계의 필요성을 뒷받침한다. 네 사례의 분석 결과는 Table 1과 같다.

이상의 분석을 바탕으로 본 연구는 SPARK-L을 이 두 요건을 단계 구조에 함께 반영하는 방향으로 구안하였다. 구체적으로 A단계(AI 협력 탐구)에서는 학습자가 AI와 상호작용하며 정보를 수집하고 검증하도록 하고, 이어지는 R단계(논리적 구조화)에서는 AI 사용을 일시 중단하여 학생이 직접 정보를 재구성하도록 설계하였다. 이를 통해 AI 협력과 학생 주도성의 균형을 단계 구조 안에 구현하고자 하였다.

2.2 인지적 오프로딩 이론과 AI 협력 학습

인지적 오프로딩(cognitive offloading)은 인지 과학에서 도구를 사용하여 인지 처리를 외부 자원에 위임하는 현상을 의미한다[3]. 이는 Sweller가 제시한 인지 부하 이론(cognitive load theory)이 학습자 내부의 작업기억 부하를 다루는 것과 달리, 인지 처리의 외재화 자체에 초점을 둔 개념이다. 손가락으로 셈을 하거나 메모지에 일정을 기록하는 것이 전통적 사례이며, 디지털 시대에는 검색 엔진과 스마트폰이 그 역할을 수행해 왔다. 이러한 외재화는 인간의 제한된 작업기억을 효율적으로 활용한다는 점에서 적응적 행동으로 평가된다.

그러나 인지적 오프로딩에는 양면성이 존재한다. 한편으로는 외부 도구가 단순 인지 부담을 덜어줌으로써 고차원적 사고에 자원을 집중시킬 수 있게 한다. 다른 한편으로는 도구에 대한 의존이 일정 수준을 넘으면 학습자 본인의 인지 처리가 약화되어, 도구의 도움 없이 독립적으로 사고를 수행하는 역량이 약화될 가능성이 있다. 이는 학습자의 장기적사고 발달에 부정적 영향을 미칠 수 있다는 점에서 교육적으로 신중히 다루어져야 할 현상이다.

생성형 AI는 정보 검색에 그치지 않고 정보의 통합과 요약, 재구성, 창작까지 외부 도구에 위임할 수 있게 한다는 점에서, 인지적 오프로딩이 적용되는 범위가 기존 디지털 도구보다 넓다. 정보 검색뿐 아니라 정보의 통합·요약·재구성·창작까지 외주 가능하기 때문이다. 신주은 외는 생성형 AI의 교육적 도입에 관한 서사적 문헌 검토를 통해, 생성형 AI가 학습자의 정보 접근성과 학습 효율성을 향상시키는 동시에 비판적 사고력 약화와 학습 의존성 심화의 위험을 동반함을 지적한 바 있다[4]. 또한, 황기석과 차경진은 AI 챗봇과 학습 단계 간 상호작용이 학습 성과에 미치는 영향을 인지부하 이론(cognitive load theory)에 근거하여 실증적으로 분석하였으며, 학습 단계별로 AI 활용 방식을 차별화할 때 학습 성과가 향상됨을 보고하였다[18].

이러한 선행 연구는 AI 융합 교수학습 모델이 AI를 활용하면서도 학습자의 인지 처리를 보호하는 균형 설계를 핵심 과제로 삼아야 함을 시사한다. 이를 단순히 AI 사용을 금지하는 방식으로 해결하기는 어렵다. AI 활용이 교육 현장에서 확대되고 있는 상황에서는 회피보다 비판적이고 전환적

Table 1. Comparison of Inquiry-Based Teaching-Learning Models and an AI Education Framework

Model	Stage Structure	Role of AI / Tools	Student Agency	Assessment	Subject Context
5E Learning Cycle (Bybee et al., 2006)	Engage → Explore → Explain → Elaborate → Evaluate(5 linear stages)	AI not included in original framework; uses general inquiry materials and experimental tools	Emphasized in Explore and Elaborate; presupposes teacher-structured guidance	Evaluate stage explicitly defined; integrates formative and summative assessment	Originated in science; expanded to all subjects
Problem-Based Learning (PBL) (Hmelo-Silver, 2004)	Ill-structured problem → Problem analysis → Information gathering → Solution development → Presentation and reflection(many variants)	AI not included in original framework	Very high; students lead from problem definition through resolution	Product- and process-based assessment; frequent peer assessment	Originated in medical education; centered on real-world problems across subjects
Question-Based Inquiry Instruction (Gyeonggi Provincial Office of Education, 2025)	Preparation → Inquiry → Organization → Practice(4-stage cycle with embedded reflection)	AI not integrated; digital tools positioned as one category of inquiry resources	Three-tier spectrum(structured-guided-open); student question formation treated as independent domain	GRASPS-based performance assessment; focuses on transfer of conceptual understanding	All subjects in elementary education
AI4K12 Big Ideas (Touretzky et al., 2019)	Perception · Representation and Reasoning · Learning · Natural Interaction · Societal Impact(5 big ideas; content framework rather than stages)	AI itself as the object of learning(content, not tool)	Stage-level agency not specified; functions as curriculum guideline	No assessment system included	K-12 AI education in general(subject-agnostic framework)

인 활용이 교육적으로 더 적절하기 때문이다. 이러한 균형 설계의 구체적 구현 원리는 본 연구가 개발하는 SPARK-L 모델의 핵심 설계 원리이며, 3장에서 상세히 다룬다.

2.3 AI 리터러시 측정 동향

AI 융합 교수학습 모델을 통한 학습자의 변화를 진단하기 위해서는 학습자의 AI 리터러시 수준을 측정할 수 있는 도구가 필요하다. 그러나 AI 리터러시 측정 도구 개발은 상대적으로 최근에야 본격화된 분야이다. 황현정 외(2024)가 4년제 대학생을 대상으로 18개 문항 5요인(AI 식별, AI 이해, AI 사용, AI 응용, AI 창조)으로 구성된 척도를 개발하였으며, 탐색적 요인분석(EFA)과 확인적 요인분석(CFA)을 거친 양적 타당화 연구로 평가된다[10].

초등과 중등 학습자를 대상으로 한 측정 도구도 개발되어 왔다. 김성원(2022)은 중학생을 대상으로 6개 하위 영역 30개 문항의 인공지능 리터러시 검사 도구를 구성하였으며[11], 류미영(2023)은 인공지능 리터러시 검사 지표를 3개 영역 30개 문항으로 제시하였다[12]. 윤선호(2022)는 초등 5~6학년을 대상으로 AI 인식, AI 원리 이해 및 활용, AI 사회적 가치, AI 윤리의 4개 영역 16개 문항으로 구성된 측정 도구를 개발하였다[13].

이상에서 검토한 기존 도구들은 학습자가 일상에서 AI를 어떻게 인식하고 활용하는지를 측정하는 데 초점을 둔 일반적 AI 리터러시 측정 도구라는 공통점을 지닌다. 따라서 특정 교수학습 모델을 적용한 이후 학습자에게 나타나는 변화, 특히 그 모델이 핵심 운영 원칙으로 요구하는 구체적 역량을 측정하기에는 정합성이 제한적이다. 예컨대 SPARK-L 모델은 P단계의 질문 정교화, A단계의 정보 교차 검증과 환각 인지, R단계의 주체적 구조화, K단계의 피드백 수용과 수정을 단계별 핵심 역량으로 설정한다. 이와 달리 기존 일반 도구의 문항은 이러한 역량을 문항 수준에서 직접 다루지 않는다. 일반 도구가 'AI를 사용할 수 있다'와 같은 포괄적 활용 수준을 묻는다면, 본 연구의 검사도구는 AI 답변의 교차 검증, AI 정보의 주체적 재구성과 같이 모델의 단계별 역량에 대응하는 문항을 포함한다는 점에서 차별된다.

본 연구는 이러한 차이를 보완하고자, 초등 5~6학년 학습자가 SPARK-L 모델을 통해 함양하는 주체적 AI 리터러시를 측정할 수 있는 검사도구 개발을 모델 개발과 병행하였다. 본 도구는 모델의 교수학습 단계에서 다루는 핵심 역량을 직접 반영함으로써 모델과 도구의 정합성을 확보하고자 하였다.

3. 연구방법

3.1 연구 절차

본 연구는 SPARK-L 교수학습 모델과 주체적 AI 리터러시 검사도구의 내용타당도를 확보하기 위해 2단계 4차에 걸친 델파이 조사를 수행하였다. 델파이 기법은 특정 주제에 대한 전문가 집단의 합의를 구조화된 절차를 통해 도출하는 방법으로, 새로운 교수학습 모델의 타당성 검증에 널리 활용되고 있다[19, 20].

본 연구의 핵심 산출물은 (1) 교수학습 모델과 (2) 주체적 AI 리터러시 검사도구의 두 가지로, 두 도구는 본질적으로 다른 성격을 갖는다. 교수학습 모델은 수업의 단계 구조와 운영 원칙을 규정하는 수업 모델이며, 주체적 AI 리터러시 검사도구는 학습자의 역량 변화를 측정하는 측정 도구이다. 따라서 두 도구는 각각 독립된 타당화 절차를 거치는 것이 학술적으로 정합적이다. 이러한 판단에 근거하여 본 연구는 4차 조사를 다음과 같이 두 단계로 구분하여 설계하였다.

1단계는 SPARK-L 교수학습 모델의 내용타당도를 검토하는 과정(1·2차)이다. 1차 델파이 조사는 연구자가 2장의 이론적 검토에 근거하여 구성한 SPARK-L 모델 초안(5단계 + 평가 영역 5항목, 총 10개 항목)에 대한 전문가 검토로 수행되었다. 1차 조사 결과를 바탕으로 미합의 항목을 수정하고 전문가의 신규 제언을 반영하여 2차 조사를 실시하였으며, 두 차례 조사를 통해 모델 구성요소의 내용타당도를 재검증하였다.

2단계는 SPARK-L 검사도구의 내용타당도를 검토하는 과정(3·4차)이다. 3차 델파이 조사는 황현정 외의 AI 리터러시 측정 도구를 재구조화하여 구성한 검사도구 초안(10문항)에 대한 전문가 검토로 수행되었다. 3차 조사 결과를 바탕으로 미합의 문항의 진술을 정련하여 4차 조사를 실시하였으며, 두 차례 조사를 통해 검사도구 문항의 내용타당도를 재검증하였다.

차수 간 간격은 1주로 운영하였으며, 각 차수는 월요일에 시작하여 금요일에 종료하고 주말에 응답을 분석하여 다음 차수의 수정안을 작성하는 방식으로 진행하였다. 이는 일반적 델파이 연구의 차수 간 간격(2~4주)에 비해 짧다[21]. 그러나 본 연구의 검증 대상은 짧고 독립적으로 판단 가능한 항목들로 구성되어 응답자의 숙고 부담이 상대적으로 낮으며, 패널 모집 단계에서 1주 간격 일정을 사전에 안내하고 동의를 얻었다는 점에서 운영상의 타당성을 확보하고자 하였다. 다만 이러한 간격 운영은 연구학교 일정상의 실용적 제약과도 결부되어 있어 후속 연구에서는 간격을 확장한 재검증이 필요하다.

응답의 안정성 측면에서도, 4차 조사 전 차수에 걸쳐 9명 전원의 응답이 수집되어 100%의 응답률을 유지하였다. 또한, 차수가 진행됨에 따라 동일 항목에 대한 응답의 표준편차가 전반적으로 감소하는 양상을 보였다. 이는 1주 간격 운영이 전문가의 응답 안정성을 저해하지 않았음을 시사한다. 차수별 일정과 검증 내용은 Table 2에 제시된 바와 같다.

Table 2. Schedule and Content of the Four Delphi Rounds

Round	Period	Validation Target	Primary Objective
Round 1	March 2–6	SPARK-L model draft (5 stages + 5 assessment components, 10 items total)	Initial content validation of model components

Round	Period	Validation Target	Primary Objective
Round 2	March 9–13	Revised items from Round 1 (7 items)	Re-validation of revisions and finalization of model
Round 3	March 16–20	Effectiveness assessment tool draft (10 items)	Initial content validation of assessment items
Round 4	March 23–27	Revised items from Round 3 (10 items)	Final consensus stabilization of assessment tool

3.2 델파이 전문가 패널 구성

델파이 전문가 패널의 전문성은 조사 결과의 타당성을 좌우하는 핵심 요소이다. Grant와 Davis는 델파이 전문가 선정 시 주제 관련 전문성, 연구 참여 의지, 적극적 의사소통 가능성, 충분한 시간 할애 가능성의 네 가지 기준을 제시한 바 있다[22]. 본 연구는 이 가운데 주제 관련 전문성을 다음 세 가지 세부 기준으로 구체화하였다. 첫째, AI·디지털 교육 또는 교육평가 분야의 학위나 연구 경력을 보유할 것. 둘째, 초등 또는 중등 교육 현장 경험이나 정책 연구 경험을 보유할 것. 셋째, 본 연구 주제와 관련된 전문성을 검증할 수 있는 학력·소속·연구 이력을 갖출 것. 이 가운데 둘 이상을 충족하는 전문가 9명으로 패널을 구성하였다.

본 연구는 전문가 패널의 규모를 9명으로 설정하였다. 이는 본 연구의 핵심 분석 지표인 Lawshe의 CVR이 패널 규모에 따라 합의 임계값이 달라지는 특성과 관련된다. 전문가 9명에서의 임계값은 .78로, 15명에서의 .49나 20명에서의 .42보다 엄격한 임계값이 적용된다. 또한, 차수 간 간격을 1주로 운영하는 본 연구의 특성상, 응답 수집과 분석의 신속성을 확보하기 위해 패널 규모를 작게 유지할 필요가 있었다.

전문가 패널의 다양성을 확보하기 위해 세 가지 소속 유형(대학 2명, 교육연구원 4명, 초등학교 3명)에 걸쳐 전문가를 모집하였다. 이는 본 연구의 산출물이 학술적 정합성, 정책적 실행 가능성, 현장 적용성의 세 측면에서 검토되도록 의도한 구성이다. 대학 소속 전문가는 학술적 근거와 이론적 정합성을, 교육연구원 소속 전문가는 정책적 실행 가능성과 평가 체계의 적합성을, 초등학교 소속 전문가는 현장 적용의 실현 가능성을 검토하는 위치에 있다. 학력 면에서는 박사 학위 6명, 박사 수료 1명, 석사 학위 2명으로 구성되었다. 전문 분야는 주 전문 분야를 기준으로 AI·디지털 교육 6명, 교육평가 2명, 교육과정 1명으로 분류하였다. 이 가운데 교육평가 전문가 2명은 AI·디지털 교육 현장 경험을 함께 보유하여, 본 연구의 평가 체계를 현장 적용 관점에서 검토할 수 있도록 하였다. 본 연구의 델파이 전문가 패널 9명의 소속, 전공, 학력은 Table 3에 제시된 바와 같다.

Table 3. Composition of the Delphi Panel

Code	Affiliation Type	Specialization	Highest Degree
P1	Elementary School	AI/Digital Education	Doctoral Candidate
P2	University	AI/Digital Education	Ph.D
P3	Educational Research Institute	AI/Digital Education	Ph.D
P4	Elementary School	Educational Assessment; AI/Digital Education	M.A.
P5	University	AI/Digital Education	Ph.D
P6	Elementary School	Educational Assessment; AI/Digital Education	M.A.
P7	Educational Research Institute	Curriculum Studies	Ph.D
P8	Educational Research Institute	AI/Digital Education	Ph.D
P9	Educational Research Institute	AI/Digital Education	Ph.D

한편, 본 전문가 패널에는 국어교육 전문가가 포함되지 않았다. 이는 본 연구가 SPARK-L을 특정 교과 내용에 종속된 모델이 아니라 AI 융합 교수학습 프레임워크로 위치시키고, 모델의 핵심 구성요소인 단계 구조와 평가 체계를 AI 융합 수업 일반의 설계 원리로 보았기 때문이다. 이에 따라 패널은 이러한 설계 원리를 검증할 수 있는 AI·디지털 교육 및 교육평가 전문가를 중심으로 구성하였다. 다만, 국어과 적용 시의 교과 특수성에 대한 자문이 이루어지지 않은 점은 본 연구의 한계이며 이에 대한 보완 방향은 5장에서 논의한다.

3.3 교수학습 모델 초안 및 검사도구 초안 개발

3.3.1 SPARK-L 교수학습 모델 초안

본 연구의 SPARK-L 교수학습 모델 초안은 2.1절의 기존 탐구학습 모델 검토를 바탕으로 초등 5~6학년 국어과 AI 융합 수업 맥락에서 학습자의 주체적 AI 리터러시를 함양하기 위한 5단계 흐름과 평가 체계로 구성되었다. 모델 초안은 다음 세 가지 이론적 근거에 기반하여 설계되었다. 첫째, Bybee 외[15]의 5E 모델이 갖는 단계적 탐구 구조의 정교함을 계승하되, AI를 학습 파트너로 명시적으로 통합하지 못한 한계를 보완하고자 하였다. 둘째, Hmelo-Silver[16]의 PBL 모델이 강조하는 학습자 주도성과 실제 문제 해결 과정의 가치를 수용하되, AI 협력과 학생 주도성 간의 균형을 단계 구조에 직접 반영하고자 하였다. 셋째, Risko와 Gilbert[3]의 인지적 오프로딩(cognitive offloading) 이론에 근거하여, AI 협력 단계와 학습자 주도 단계를 의도적으로 구분함으로써 AI 의존이 학습자의 인지적 책임을 약화시키지 않도록 설계하였다.

SPARK는 Situation(상황 발견), Posing Question(질

문 구체화), AI-assisted Inquiry(AI 협력 탐구), Reasoning(논리적 구조화), Knowledge Production(지식 창출)의 다섯 단계로 구성된다. S단계는 학습자가 삶의 맥락에서 문제를 발견하는 단계로, 교사가 정선한 데이터(영상·자료)를 통해 학습 미션을 수립하고 학급 공감대를 형성하는 활동으로 구성된다. P단계는 브레인스토밍을 통해 문제 해결 아이디어를 도출하고, 교사 및 AI의 도움을 받아 탐구 가능한 핵심 질문으로 정교화하는 단계이며, A단계는 클라우드 기반 생각노트를 활용하여 AI와 상호작용하며 정보를 수집하고 출처 교차 검증을 통한 팩트체크 결과를 누적 기록하는 활동으로 구성된다. R단계는 본 모델의 핵심 차별점이 집약된 단계로, AI와의 상호작용을 일시 중단하는 주도권 전환 기제를 도입하였다. 이는 인지적 오프로딩 이론에 직접 근거한 설계로서, AI 활용 과정에서 외부화된 정보를 학습자가 자신의 논리로 재구성하도록 유도함으로써 정보의 기계적 복제를 방지하고 논리망(Logic-tree) 구축으로 이어지도록 하는 단계이다. K단계는 산출물 초안에 대한 AI의 형성적 피드백과 교사의 심층 지도를 비판적으로 수용하고, 글의 통일성을 집중적으로 고쳐 쓴 후 학급 단위로 공유하는 단계이다. 공유의 구체적 형태는 학급 문집이나 학급 도서로 묶어 학급 구성원이 함께 열람하는 방식 등으로 운영할 수 있다.

본 모델은 단계별 교수학습 흐름과 함께 평가 체계를 통합 설계하였다는 점에서 기존 탐구학습 모델들과 차별된다. 평가 체계 초안은 다음 다섯 항목으로 구성되었다. 첫째, 디지로그 평가 체계는 사고의 확산은 디지털(클라우드 노트)로, 사고의 수렴과 구조화는 아날로그(손글씨 노트)로 이원화하여 평가하는 시스템이다. 이는 학생의 문해력 보호와 AI 의존성 방지를 도모한다. 둘째, 프롬프트 과정 평가는 결과물 자체보다 학생이 AI에게 던진 질문의 구체화 과정과 대화 이력을 평가의 핵심 척도로 삼는다. 셋째, 팩트 체크 및 윤리 평가는 생성형 AI의 답변을 그대로 복사하지 않고 사실 확인 여부와 윤리적 재구성 과정을 평가 항목에 포함한다. 넷째, 고쳐 쓰기 중심 평가는 초안 작성 후 AI 및 교사의 피드백을 반영하여 글을 개선한 고쳐 쓰기의 변화폭을 성장의 지표로 삼는다. 다섯째, 일반화 가능성 측면에서 본 모델의 질문-탐구-글쓰기 과정이 국어과뿐만 아니라 타 교과의 서·논술형 수행평가 모델로 일반화될 수 있도록 설계되었다.

본 모델은 초등 5~6학년 국어과를 적용 교과로 설정하였으며, 국어과 교육과정의 쓰기 영역(특히 고쳐쓰기, 문단 구성)과 매체 영역(디지털 도구를 활용한 표현)을 통합하는 4~8차시 분량의 주제중심 질문기반 프로젝트로 운영되도록 설계하였다. 적용 사례로는 본 연구자가 구안한 SDG 13(기후행동) 주제를 다루는 'AI와 함께하는 생태 전환 프로젝트: 지구를 지키는 스토리텔러'를 상정하였으며, 이는 4차시 R단계와 5~6차시 K단계의 핵심 활동을 구체화하는 맥락으로 활용되었다. 이상의 SPARK-L 모델 초안은 Table 4와 같으며, 1·2차 델파이 조사를 통해 내용타당도를 검증받았다. 그 과정은 4.1절과 4.2절에서 상세히 제시한다.

Table 4. Initial Draft of the SPARK-L Teaching-Learning Model (Submitted to Round 1)

Domain	Stage / Component	Core Strategy
Instructional Flow (SPARK 5 Stages)	S Situation	Through teacher-curated data, learners discover problems situated in life contexts (e.g., AI-applied social phenomena, climate crisis), establishing a learning mission and shared purpose.
	P Posing Question	Students assign roles to AI and refine ideas, with teacher and AI support, into investigable core questions.
	A AI-assisted Inquiry	Students interact with generative AI through cloud thinking notes to collect information and exchange ideas.
	R Reasoning	AI interaction is temporarily suspended: students lead the construction of a Logic-tree using analog notes, critically verifying AI-provided information.
	K Knowledge Production	Students publish AI-collaborated artifacts of social value, transcending mere consumption of AI output.
Assessment Areas	Digilogue Assessment System	A bifurcated system where divergent thinking is assessed via digital tools and convergent/structural thinking via analog tools, aimed at protecting literacy and preventing AI dependency.
	Prompt Process Assessment	Treating the refinement process of student queries to AI and the dialogue history-rather than the output itself-as the core evaluation metric.
	Fact-checking & Ethical Assessment	Including fact-checking and ethical paraphrasing of AI responses (rather than verbatim copying) as evaluation criteria.
	Rewriting-centered Assessment	Treating the magnitude of revision after initial draft, based on AI and teacher feedback, as the indicator of growth.
	Generalizability	Designed so that the model's question-inquiry-writing process can be generalized to written/essay performance assessment models in other subjects beyond Korean language arts.

3.3.2 SPARK-L 주체적 AI 리터러시 검사도구 초안

본 연구의 검사도구는 황현정 외(2024)는 4년제 대학생을 대상으로 탐색적 요인분석과 확인적 요인분석을 거쳐 타당화한 AI 리터러시 측정 도구(5개 요인 18문항)의 이론적 체계를 출발점으로 삼되, 본 연구의 대상(초등 5~6학년)과 적용 맥락(국어과 AI 융합 수업) 및 SPARK-L 모델의 단계 구조에 정합하도록 재구성하여 10개 문항으로 개발하였다.

황현정 외의 척도[10]를 출발점으로 선정한 이유는 다음과 같다. 첫째, 해당 척도는 OECD 역량 프레임워크인 지식·기술·태도 차원에 기반하여 체계적으로 구성되어 있어 초등 교육과정의 핵심 역량 체계와 연결성을 갖는다[2]. 둘째, AI 식별·이해·사용·응용·창조라는 5개 요인 구조가 학습자의 AI 활용 수준을 단계적으로 측정할 수 있어 SPARK-L 모델

의 단계별 학습 목표 진단에 적합하다.

다만 원 척도가 4년제 대학생을 대상으로 개발되어, 본 연구의 학습자(초등 5~6학년) 발달 수준과 적용 맥락(국어과 AI 융합 수업)에 그대로 적용하기에는 재구성이 필요하였다. 첫째, 원 척도의 추상적 진술을 초등 학습자가 이해할 수 있는 구체적 진술로 정련할 필요가 있었다. 둘째, SPARK-L 모델의 핵심 운영 규칙(R단계 주체적 구조화, 팩트체크 원칙, 피드백 수용 중심 평가)을 직접 측정하는 문항이 원 척도에는 부재하였다. 셋째, 본 모델의 적용 교과인 국어과의 K단계 산출 행위(고쳐쓰기, 디지털 시각 표현)를 측정하는 문항이 추가로 요구되었다.

이러한 요구를 반영하여 초등 교육 현장의 수업 맥락에 맞추어 문항을 재구성하였다. 먼저 원 척도의 5개 요인을 SPARK-L 모델의 단계별 핵심 역량과 대응시켜, 모델이 측정하고자 하는 역량과 연결되는 문항을 선별하였다. 선별한 문항은 초등 5~6학년의 발달 수준과 수업 활동 맥락에 맞추어, 추상적 진술을 학습자가 이해할 수 있는 구체적 행동 진술로 재진술하고 한 문항이 하나의 역량만 측정하도록 단순화하였다. 한편 SPARK-L 모델의 핵심 운영 원칙과 국어과 K단계 산출 행위를 측정하는 문항은 원 척도에 대응 항목이 없어 새로 구성하였다.

재구성한 문항을 SPARK 단계와 대응시키면, 학습자가 직접 수행하는 역량을 측정하는 P·A·R·K 단계를 중심으로 문항이 구성된다. 다만 S단계(상황 발견)는 교사가 학습 자료를 정선하여 학습 맥락을 조성하는 도입 단계로, 학습자 자기보고식 문항으로 측정하기보다 평가 체계의 의도된 데이터 스캐폴딩 원칙을 통해 반영하였다. 이에 따라 검사도구는 P·A·R·K 단계의 핵심 역량과 모든 단계의 기초가 되는 AI 이해 및 프로젝트 전반을 지원하는 횡단 역량을 중심으로 구성하였다. 각 문항이 대응하는 SPARK 단계와 원 척도와의 관계는 Table 5에 제시한 바와 같다.

본 검사도구는 5점 Likert 척도 9문항(문항 1~9)과 자유서술형 1문항(문항 10)으로 구성하였다. 정량 문항은 학습자의 단계별 역량 수준을 표준화된 척도로 측정하여 사전과 사후의 변화를 비교할 수 있도록 하는 역할을 한다. 반면 자유서술형 문항은 학습자가 AI의 답변을 자신의 논리로 재구성한 경험을 직접 기술하도록 하여, 정량 문항이 포착하기 어려운 메타인지적 성찰 과정을 질적으로 진단하는 역할을 한다. 두 유형을 함께 둠으로써 역량의 수준과 그 이면의 사고 과정을 다층적으로 확인하고자 하였다. 재구성된 검사도구의 내용타당도는 3·4차 델파이 조사를 통해 검토되었으며, 그 결과는 4.3절과 4.4절에서 제시한다.

Table 5. Measurement Domains of the Initial SPARK-L Self-directed AI Literacy Assessment Tool (Submitted to Round 3)

Item	Competency	Measurement Domain	SPARK Stage	Source
1	AI Understanding	Cognitive understanding of AI operating principles	Foundational	Refined
2	Prompt Design	Ability to design specific queries for AI	P Stage	New
3	AI Tool Application	Use of AI for solving learning tasks	A Stage	Adapted
4	Information Cross-verification	Source verification of AI responses	A Stage	New
5	Hallucination Awareness	Recognition of potential errors and bias in AI responses	A Stage	Adapted
6	Self-directed Structuring	Reorganization of AI information based on one's own logic	R Stage	New
7	Feedback Reception & Revision	Improvement of writing through formative feedback	K Stage	New
8	Digital Creative Expression	Visual expression using digital tools	K Stage	New
9	Collaborative Communication	Collaborative competency in group activities	Project-wide	New
10	Qualitative Reflection	Metacognitive reflection on AI revision experience	Metacognitive	New

3.4 분석 방법

각 차수의 응답은 5점 Likert 척도(1=전혀 타당하지 않음 ~ 5=매우 타당함)로 수집되었으며, 항목별로 평균(M), 표준편차(SD), 내용타당도 비율(Content Validity Ratio, CVR)의 세 지표를 산출하였다. 평균과 표준편차는 응답의 중심 경향과 분산을 보조적으로 검토하기 위한 지표이며, 합의 도달 여부의 직접 판정 기준은 Lawshe[19]가 제시한 CVR이다. CVR의 산출식은 식 (1)과 같다.

$$CVR = \frac{n_e - (\frac{N}{2})}{\frac{N}{2}} \quad (1)$$

여기서 N은 전체 전문가 수, 는는 해당 항목을 타당(5점 척도 중 4점 이상)으로 평정한 전문가 수이다. Lawshe에 따르면 전문가 수 N에 따라 합의 판정의 임계값이 달라지며, 본 연구의 전문가 수 N=9에 해당하는 임계값은 .78이다. 본 연구는 이 .78 이상을 충족한 항목을 타당(합의 도달)으로 판정하였다.

미도달 항목에 대해서는 전문가의 개방형 응답을 함께 수집하여 수정 방향의 근거로 활용하였다. 개방형 응답은 연

구자가 전체 의견을 통독한 뒤 반복적으로 제기되는 쟁점을 중심으로 유사한 의견을 묶어 핵심 수정 방향을 도출하는 방식으로 분석하였으며, 도출된 수정 방향을 다음 차수의 수정안에 반영하여 재검증하였다.

4. 연구 결과

본 장에서는 4차에 걸친 델파이 조사 결과를 차수별로 제시한다. 1·2차 조사는 SPARK-L 교수학습 모델의 구성요소를, 3·4차 조사는 검사도구의 문항을 검증 대상으로 하였다. 각 차수의 결과표에서 N은 5점 척도 중 4점 이상(타당 또는 매우 타당)으로 응답한 전문가 수를 의미하며, 9명 기준 CVR 임계값 .78을 충족하여 내용타당도가 확보된 항목은 합의 여부(Consensus) 열에 ○로 표기하였다.

4.1 1차 델파이 조사 결과 (모델 초안의 타당성 검증)

1차 델파이 조사는 SPARK-L 모델 초안을 두 영역으로 구성하여 검증하였다. 영역 Ⅰ(교수학습 흐름)은 SPARK 5단계, 즉 S단계(상황 발견), P단계(질문 구체화), A단계(AI 협력 탐구), R단계(논리적 구조화), K단계(지식 창출)의 5개 항목으로 구성되었다. 영역 Ⅱ(평가 체계)는 디지로그 평가 체계, 프롬프트 과정 평가, 팩트 체크 및 윤리 평가, 고쳐 쓰기 중심 평가, 일반화 가능성의 5개 항목으로 구성되었다. 두 영역에 걸쳐 총 10개 문항을 검증하였으며, 1차 조사 결과는 Table 6과 같다.

Table 6. Results of the First Delphi Survey

Domain	Item	N	M	SD	CVR	Consensus
Ⅰ. Instructional Logic (SPARK)	S Situation	8	4.11	0.93	.78	○
	P Posing Question	9	4.67	0.50	1.00	○
	A AI-assisted Inquiry	8	4.33	1.00	.78	○
	R Reasoning	5	3.56	1.59	.11	
	K Knowledge Production	9	4.56	0.53	1.00	○
Ⅱ. Assessment System	Digilogue Assessment System	7	4.00	1.00	.56	
	Prompt Process Assessment	8	4.67	0.71	.78	○
	Fact-checking & Ethical	9	4.78	0.44	1.00	○
	Rewriting-centered Assessment	7	4.44	1.13	.56	
	Generalizability	7	4.22	1.09	.56	

1차 델파이 조사 결과는 다음과 같다. 첫째, 영역 Ⅰ(교수학습 흐름)에서는 S·P·A·K 4개 단계가, 영역 Ⅱ(평가 체계)에서는 프롬프트 과정 평가와 팩트 체크 및 윤리 평가 2개 항목이 CVR .78 이상으로 내용타당도를 확보하였다. 둘째, R단계와 평가 체계 3개 항목(디지로그 평가 체계, 고쳐 쓰기 중심 평가, 일반화 가능성)은 합의에 도달하지 못하였다. 특히 R단계는 CVR .11로 임계값에 크게 미달하였고, 평가 체계 3개 항목은 CVR .56으로 임계값에 도달하지 못하였다. 셋째, 이러한 미합의 항목에 대해 수집한 전문가의 개방형 응답을 3.4절에 제시한 절차에 따라 분석한 결과, R단계의 아날로그 활동 강조가 디지털 교육 흐름과 상충하고 논리망 구축의 구체적 방법이 모호하다는 의견, 그리고 디지로그 평가 체계의 학술적 근거가 부족하고 평가 주도권이 AI에 과도하게 의존된다는 의견 등이 공통적으로 도출되었다. 이를 바탕으로 R단계 재설계, 평가 체계 재구성, 평가의 책무성 강화를 핵심 수정 방향으로 도출하였다.

4.2 2차 델파이 조사 결과 (수정 모델의 타당성 재검증)

2차 델파이 조사는 1차에서 미합의된 4개 항목(R단계, 디지로그 평가 체계, 고쳐 쓰기 중심 평가, 일반화 가능성)의 수정본과 더불어, 1차 조사에서 도출된 수정 원칙으로부터 새롭게 도출된 3개 항목을 함께 검증하였다. 추가된 3개 항목은 1차 전문가 의견과 다음과 같이 연결된다. S단계 개선(의도된 데이터 스캐폴딩)은 1차에서 일부 평가 항목이 학습 자료의 사전 정선 과정을 전제로 한다는 의견을 반영하여, 교사의 자료 정선 과정을 별도 항목으로 명시한 것이다. 평가의 책무성은 1차 개방형 응답에서 공통적으로 제기된 평가 주도권이 AI에 과도하게 의존된다는 우려를 반영하여, 총괄 평가의 책임을 교사에게 두는 원칙을 신규 도입한 것이다. 현장 실행 가능성은 1차에서 일반화 가능성 항목이 추상적이라는 지적을 반영하여, 실제 차시 운영 수준의 실행 가능성으로 구체화한 것이다.

1차 미합의 항목의 주요 수정 내용은 다음과 같다. R단계는 아날로그 형식의 강조에서 벗어나 도구 선택의 자율성을 부여하고, 논리망(Logic-tree) 구축 방법을 구체화하는 방향으로 수정하였다. 평가 체계는 1차에서 미합의된 디지로그 평가 체계, 고쳐 쓰기 중심 평가, 일반화 가능성을 제외하고, 전문가 의견에서 도출된 평가의 책무성, 성장 지표의 재정의, 탐구 규칙(팩트체크 3원칙)으로 재구성하였다.

2차 조사 결과는 Table 7과 같다. 7개 항목 모두 CVR .78 이상으로 내용타당도를 확보하였으며, 1차에서 미합의되었던 4개 항목이 모두 합의에 도달하였고 추가로 검증한 3개 항목 역시 합의에 도달함으로써 수정안에 대한 전문가의 합의가 확인되었다. 이로써 SPARK-L 교수학습 모델의 모든 구성요소에 대한 내용타당도가 검토되었으며, 그 최종 구성은 4.5.1절에서 제시한다.

Table 7. Results of the Second Delphi Survey

Item	N	M	SD	CVR	Consensus
1. S Stage Revision: Intended Data Scaffolding	9	4.78	0.44	1.00	○
2. R Stage Revision ①: Flexibility of Analog Format	9	4.67	0.50	1.00	○
3. R Stage Revision ②: Specification of Logic-tree	8	4.33	1.00	.78	○
4. Assessment Accountability: Teacher-led Role	8	4.33	1.00	.78	○
5. Redefinition of Growth Indicator: From Change to Receptivity	9	4.56	0.53	1.00	○
6. Inquiry Rule: Fact-check 3 Principles	8	4.44	0.73	.78	○
7. Field Implementation Feasibility	8	4.44	0.73	.78	○

4.3 3차 델파이 조사 결과 (검사도구 초안의 타당성 검증)

2차 조사를 통해 내용타당도가 검토된 SPARK-L 모델 적용에 따른 학습자의 변화를 측정하기 위해 3.3.2절에서 개발한 10개 문항의 검사도구 초안에 대한 3차 델파이 조사를 실시하였다. 문항은 모델의 단계별 핵심 역량을 반영하여 AI 이해, 프롬프트 설계, AI 도구 응용, 정보 교차 검증, 환각 현상 인지, 주체적 구조화, 피드백 수용 및 수정, 디지털 창의 표현, 협력적 소통, 질적 성찰의 10개 하위 역량을 측정하도록 설계되었다. 3차 조사 결과는 Table 8과 같다.

Table 8. Results of the Third Delphi Survey

Item	Item	N	M	SD	CVR	Consensus
1	AI Understanding	7	4.56	0.88	.56	
2	Prompt Design	8	4.56	0.73	.78	○
3	AI Tool Application	6	3.89	0.78	.33	
4	Information Cross-verification	9	5.00	0.00	1.00	○
5	Hallucination Awareness	9	4.89	0.33	1.00	○
6	Self-directed Structuring	9	4.67	0.50	1.00	○
7	Feedback Reception & Revision	7	4.22	0.83	.56	
8	Digital Creative Expression	6	4.22	0.97	.33	
9	Collaborative Communication	9	4.56	0.53	1.00	○
10	Qualitative Reflection	9	4.89	0.33	1.00	○

10개 문항 중 6개 문항이 CVR .78 이상으로 내용타당도를 확보하였으며, 4개 문항(AI 이해, AI 도구 응용, 피드백 수용 및 수정, 디지털 창의 표현)은 합의에 도달하지 못하였다. 미합의 4개 문항에 대해 수집한 전문가의 개방형 응답을 3.4절에 제시한 절차에 따라 분석한 결과, 측정하고자 하는 역량이 모호하다는 의견, 초등 5~6학년 수준에서 어휘가 어렵다는 의견, 특정 산출물이나 도구에 종속된 표현이라는 의견이 공통적으로 도출되었다. 특히 AI 도구 응용 문항과 디지털 창의 표현 문항은 동화책 만들기, 캔바(Canva) 등 특정 산출물과 도구에 한정된 표현으로 일반성이 떨어진다는 의견이 다수였다.

이러한 의견을 바탕으로 AI 이해 문항은 학술적 표현의 명료화, AI 도구 응용 문항은 특정 산출물 표현의 일반화, 피드백 수용 및 수정 문항은 평가 차원의 재정의, 디지털 창의 표현 문항은 도구 종속성 제거를 수정 방향으로 설정하였다. 미합의 4개 문항의 수정 전후 진술은 Table 9와 같다.

Table 9. Revisions of Non-consensus Items in the Third Delphi Survey

Item	Statement in Round 3 (Before Revision)	Statement in Round 4 (After Revision)	Direction of Revision
1. AI Understanding	I know that AI does not think on its own but generates answers by learning from large amounts of data.	I understand the principle that AI generates answers fitting the question by learning from large volumes of data.	Clarification of academic expression
3. AI Tool Application	I can use AI tools at my own level to achieve my goal of making a storybook.	I can use AI tools fit-for-purpose to solve given learning tasks or problems.	Generalization of product-specific wording
7. Feedback Reception & Revision	I can critically judge the advice given by AI, teachers, and peers, and revise my writing to be more natural and coherent.	I can revise the weaker parts of my writing toward improvement based on feedback from AI, teachers, and peers.	Redefinition of the measured competency
8. Digital Creative Expression	I can creatively express images and designs that best suit the mood and content of my story using Canva.	I can use digital tools to visually express the mood and content of my story to communicate it to readers.	Removal of tool dependency

4.4 4차 델파이 조사 결과 (검사도구의 최종 합의)

3차 조사에서 미합의된 4개 문항(AI 이해, AI 도구 응용, 피드백 수용 및 수정, 디지털 창의 표현)을 수정 방향에 따라 정련하여 4차 델파이 조사로 재검증하였다. 4차 조사 결과는 Table 10와 같다.

Table 10. Results of the Fourth Delphi Survey

Item	Item	N	M	SD	CVR	Consensus
1	AI Understanding (Revised)	9	4.67	0.50	1.00	○
2	Prompt Design (Validated)	9	4.89	0.33	1.00	○
3	AI Tool Application (Revised)	9	4.78	0.44	1.00	○

4	Information Cross-verification (Validated)	9	5.00	0.00	1.00	○
5	Hallucination Awareness (Validated)	9	5.00	0.00	1.00	○
6	Self-directed Structuring (Validated)	9	5.00	0.00	1.00	○
7	Feedback Reception & Revision (Revised)	9	4.78	0.44	1.00	○
8	Digital Creative Expression (Revised)	9	4.67	0.50	1.00	○
9	Collaborative Communication (Validated)	9	4.78	0.44	1.00	○
10	Qualitative Reflection (Validated)	9	5.00	0.00	1.00	○

4차 조사에서 10개 문항 모두 CVR 1.00에 도달하였으며, 3차에서 미합의되었던 4개 문항과 임계값에 머물렀던 프롬프트 설계 문항이 모두 합의에 도달하였다. 이는 3차에서 도출된 수정 방향, 곧 어휘의 명료화, 측정 역량의 명확화, 도구 종속성의 제거가 전문가 의견에 부합하였음을 보여준다. 이로써 검사도구 문항에 대한 전문가 패널의 내용타당도 합의가 확인되었다. 본 결과는 내용타당도 검토 단계의 합의에 해당하며, 학습자 적용을 통한 효과성과 신뢰도 검증은 후속 과제로 남는다. 이러한 검토를 거쳐 도출된 검사도구의 최종안은 4.5.2절에서 제시한다.

4.5 최종 SPARK-L 교수학습 모델 및 주체적 AI 리터러시 검사도구

4.5.1 최종 SPARK-L 교수학습 모델

최종 SPARK-L 교수학습 모델은 Table 11과 같다. 본 모델은 5단계 교수학습 흐름과 6개 항목으로 구성된 평가 체계로 이루어진다. 5단계 교수학습 흐름은 학습자가 삶의 맥락에서 문제를 발견하는 S단계, 교사와 AI의 도움을 받아 탐구 가능한 핵심 질문으로 정교화하는 P단계, 클라우드 생각노트를 활용해 AI와 협력 탐구하는 A단계, 학생 주도의 논리망(Logic-tree)을 구축하여 정보를 재구성하는 R단계, 산출물의 통일성을 고쳐 쓰고 학급 단위로 공유하는 K단계로 이루어진다. 평가 체계는 의도된 데이터 스캐폴딩, 평가의 책무성, 성장 지표의 재정의, 팩트체크 3원칙, 논리망 구축의 구체화, 현장 실행 가능성의 6개 항목으로 구성된다.

Table 11. Final SPARK-L Teaching-Learning Model

Domain	Stage / Component	Core Strategy
Ⅰ. Instructional Flow (SPARK 5 Stages)	S Situation (Session 1)	Through teacher-curated data (videos, materials), learners discover problems situated in life contexts and establish a mission and shared purpose for solving them.
	P Posing Question (Session 2)	Students generate ideas through brainstorming and refine them, with teacher and AI support, into investigable core questions.
	A AI-assisted Inquiry (Session 3)	Using cloud thinking notes, students interact with AI to collect information and document fact-checking through cross-verification (≥2 sources).
	R Reasoning (Session 4)	AI interaction is temporarily suspended (agency transfer); students lead the construction of a Logic-tree and visual outline. Tool choice is flexible (sticky notes, mind maps, digital typing), and tiered structuring templates are provided for students requiring scaffolding.
	K Knowledge Production (Sessions 5-6)	Drafts uploaded to the learning management system (LMS) incorporate AI's formative feedback and the teacher's in-depth guidance critically; students apply self-monitoring strategies (e.g., reflection notes) to revise for textual coherence and share the completed work as a class outcome.
Ⅱ. Assessment System	Intentional Data Scaffolding	The teacher pre-curates materials aligned with curriculum objectives to safeguard learners' self-directed problem discovery.
	Assessment Accountability	Final achievement determination is conducted via teacher rubric-based summative assessment, while AI is restricted to a formative feedback tool.
	Redefinition of Growth Indicator	The core metric is critical feedback receptivity (portfolio evidence such as reasons for selection/rejection) and product completeness, rather than mere word-count increase or revision magnitude.
	Inquiry Rule (Three Principles of Fact-Checking)	Cross-verification of sources (≥2), explicit grounding in objective evidence, and critical examination are applied as foundational rules of inquiry.
	Specification of Logic-tree Construction	Whether students reconstruct AI information into their own logical structure (e.g., cause-effect) rather than mechanically copying it is assessed through artifacts such as mind maps and outline tables.
	Field Implementation Feasibility	The model is designed for theme-based question-driven projects of 4-8 sessions at grades 5-6, integrating session structure, tools, and assessment procedures for school-level implementation.

본 절에서는 앞선 델파이 조사를 통해 내용타당도가 검토된 최종 SPARK-L 교수학습 모델을 4.5.1절에서, 주체적 AI 리터러시 검사도구를 4.5.2절에서 각각 제시한다.

본 모델의 주요 특징은 두 가지로 정리된다. 첫째, R단계에서 AI 주도권 전환을 명시함으로써 AI 협력과 학생 주도성의 균형을 단계 구조에 반영하였다. 둘째, 평가 체계가 단순한 부속 요소가 아닌 모델의 운영 원칙으로 통합되어 있으며, AI를 형성적 피드백 도구로 한정하고 교사 루브릭을 총괄 평가의 단독 근거로 삼는 평가 책무성 원칙을 핵심에 두고 있다.

본 모델의 평가 체계는 1차 조사의 5개 항목에서 2차 조사를 통해 6개 항목으로 재구성되었다. 이 과정은 단순한 항목의 추가나 삭제가 아니라 통합, 재정의, 신규 도입이 복합적으로 이루어진 결과로, 그 변경 경로를 Table 12에 정리하였다.

Table 12. Revisions of Non-consensus Items in the Third Delphi Survey

Round 1 → Final	Status	Rationale
Digilogue Assessment System → Intentional Data Scaffolding	Redefined (R1 CVR .56)	Reframed as the teacher's pre-curation principle at the S stage, addressing the lack of academic grounding for the term "digilogue."
Prompt Process Assessment → Redefinition of Growth Indicator	Integrated (R1 CVR .78)	Both treat process evidence rather than output as the evaluation metric; merged on grounds of conceptual coherence.
Fact-checking & Ethical Assessment → Inquiry Rule (Three Principles of Fact-Checking)	Elevated (R1 CVR 1.00)	Repositioned from assessment criterion to operating rule, concretized into three principles.
Rewriting-centered Assessment → Specification of Logic-tree Construction	Redefined (R1 CVR .56)	Refocused on R-stage evidence of information reorganization, addressing the limitation of revision-magnitude metrics.
Generalizability → Field Implementation Feasibility	Redefined (R1 CVR .56)	Reframed as concrete implementation feasibility in 4–8 session units.
— → Assessment Accountability	Newly introduced in Round 2	Restricts AI to formative feedback; designates teacher rubric as sole basis for summative judgment.

Table 12의 변경 경로는 다음 세 가지 기준에 따라 처리되었다. 첫째, 1차에서 합의된 두 항목이 평가 철학을 공유하는 경우 통합하였다. 프롬프트 과정 평가와 성장 지표의 재정의는 모두 결과물이 아닌 과정의 증거를 평가의 핵심 척도로 삼는다는 공통 원리를 지녀, 별개 항목으로 병치할 경우 평가 체계의 개념적 중복이 발생하므로 하나로 통합하였다. 둘째, 1차에서 미합의된 세 항목(디지로그 평가 체계, 고쳐 쓰기 중심 평가, 일반화 가능성)은 전문가 개방형 응답에서 공통적으로 지적된 사유, 곧 용어의 학술적 근거 부족, 수정 분량 측정의 한계, 진술의 추상성을 반영하여 본 모델의 단계 구조와 정합하는 항목으로 재정의하였다. 셋째, 1차 미합의 항목들의 응답을 분석하는 과정에서 새롭게 도출된 평가 주도권의 AI 의존 우려는 기존 항목으로 흡수하기 어려운 독립적 원칙이므로, 평가의 책무성 항목을 신규 도입하였다. 이러한 재구성을 거쳐 2차 조사에서 6개 평가 체계 항목 모두 CVR .78 이상으로 합의에 도달하였으며, 내용타당도 검토 기준에 따라 평가 체계의 최종안이 도출되었다.

4.5.2 최종 SPARK-L 주체적 AI 리터러시 검사도구

최종 주체적 AI 리터러시 검사도구는 Table 13과 같다. 본 검사도구는 SPARK-L 모델의 단계별 핵심 역량을 직접 반영하는 10개 문항으로 구성되어, 학습자의 주체적 AI 리터러시를 다층적으로 측정할 수 있도록 설계되었다. 10개 문항 중 7개 문항은 SPARK 5단계와 직접 매핑된다. P단계는 프롬프트 설계(문항 2), A단계는 AI 도구 응용·정보 교차 검증·환각 현상 인지(문항 3, 4, 5), R단계는 주체적 구조화(문항 6), K단계는 피드백 수용 및 수정·디지털 창의 표현(문항 7, 8)이 각각 매핑된다. 한편 AI 이해(문항 1)는 모든 단계의 기초가 되는 토대 역량으로, 협력적 소통(문항 9)과 질적 성찰(문항 10)은 프로젝트 전반에 걸쳐 발휘되는 횡단 역량으로 포괄되어 측정의 다층성을 확보하였다.

Table 13. Final SPARK-L Self-directed AI Literacy Assessment Tool

Item	Competency	SPARK Stage	Item Statement
1	AI Understanding	Foundational	I understand the principle that AI (ChatGPT, Wrtn, etc.) generates answers fitting the question by learning from large volumes of data.
2	Prompt Design	P Stage	I can ask AI specific questions including roles or conditions to obtain desired results.
3	AI Tool Application	P Stage	I can use AI tools fit-for-purpose to solve given learning tasks or problems.
4	Information Cross-verification	A Stage	I verify (fact-check) whether the information AI provides is accurate by consulting other sources (books, internet searches, etc.).
5	Hallucination Awareness	A Stage	I am aware that AI-generated information may sometimes be wrong, biased, or one-sided.
6	Self-directed Structuring	R Stage	I can reorganize information according to my own thinking and logic rather than copying AI's output verbatim.
7	Feedback Reception & Revision	K Stage	I can revise the weaker parts of my writing toward improvement based on feedback from AI, teachers, and peers.
8	Digital Creative Expression	K Stage	I can use digital tools to visually express the mood and content of my work to communicate it to readers.

9	Collaborative Communication	Project-wide	I can divide roles, respect others' opinions, and solve tasks together in group activities.
10	Qualitative Reflection	Metacognitive	If you have experience of revising AI's answer in your own way during the project, please describe specifically which parts you revised and why.

본 검사도구는 SPARK 모델의 단계 구조와 직접 매핑된 문항 구성을 통해 모델과 도구의 정합성을 확보하는 방식으로 설계되었다. 따라서 본 단계에서는 통계적 요인 구조 검증보다 단계별 역량 측정의 적합성 확보에 초점을 두었으며, 학습자 적용을 통한 효과성 검증과 신뢰도 분석은 후속 연구의 과제로 남는다.

5. 결론 및 제언

5.1 결론

본 연구는 초등 5~6학년 국어과 AI 융합 수업 맥락에서 학습자의 주체적 AI 리터러시를 함양하기 위한 SPARK-L 교수학습 모델과 주체적 AI 리터러시 검사도구를 개발하고, 4차에 걸친 델파이 조사를 통해 두 도구의 내용타당도를 검증하였다. 본 연구의 주요 결론은 다음과 같다.

첫째, 1·2차 델파이 조사를 통해 SPARK-L 교수학습 모델의 내용타당도가 검토되어 최종안이 도출되었다. 모델은 Situation, Posing Question, AI-assisted Inquiry, Reasoning, Knowledge Production의 5단계 교수학습 흐름과 의도된 데이터 스캐폴딩, 평가의 책무성, 성장 지표의 재정의, 팩트체크 3원칙, 논리망 구축의 구체화, 현장 실행 가능성의 6개 평가 체계 항목으로 구성된다. 1차 조사에서 R단계의 CVR이 .11로 임계값에 크게 미달하여 AI 활용 중단을 아날로그 활동으로 제한한 초기 설계의 한계가 확인되었으며, 2차 조사에서 아날로그 형식의 유연화와 논리망 구축의 구체화로 분리하여 도구 선택의 자율성을 확보한 결과 모든 항목이 합의에 도달하였다.

둘째, 3·4차 델파이 조사를 통해 SPARK-L 주체적 AI 리터러시 검사도구의 내용타당도가 검토되어 최종안이 도출되었다. 검사도구는 5점 Likert 척도 9문항과 자유서술형 1문항으로 구성된 10개 문항이며, 7개 문항은 SPARK 5단계와 직접 매핑되고, AI 이해 문항은 모든 단계의 기초가 되는 토대 역량으로, 협력적 소통과 질적 성찰 문항은 프로젝트 전반에 걸쳐 발휘되는 횡단 역량으로 포괄된다. 3차 조사에서 미합의된 4개 문항은 도구 종속성 제거와 어휘 명료화를 거쳐 4차 조사에서 모두 CVR 1.00에 도달하였으며, 이로써 검사도구 전 문항이 전문가 패널의 내용타당도 합의를 확보하였다.

셋째, 본 연구는 SPARK-L 모델이 기존 탐구학습 모델 및 AI 교육 프레임워크와 구분되는 설계 특징을 지님을 제시하였다. 본 연구에서 검토한 5E[15], PBL[16], AI4K12[17]와 비교할 때, SPARK-L은 두 가지 측면에서 구별된다. 하나는 R단계의 주도권 전환 기제를 통해 AI 협력과 학생 주도성의 균형을 단계 구조에 반영한 점이며, 다른 하나는 평가 체계가 단계별 교수학습 흐름과 통합 설계되어 평가의 책무성을 모델 운영 원칙으로 명시한 점이다. 또한 본 검사도구는 황현정 외(2024)의 척도를 출발점으로 하되 초등 5~6학년 학습자와 SPARK-L 모델 적용 맥락에 정합하도록 재구성됨으로써, 일반적 AI 리터러시 측정 도구에서 상대적으로 충분히 반영되지 않은 특정 교수학습 모델 적용 후의 변화를 진단할 수 있도록 설계되었다.

5.2 제언

본 연구의 한계와 후속 연구 방향은 다음과 같다.

첫째, 본 연구는 모델과 주체적 AI 리터러시 검사도구의 내용타당도 검증에 초점을 두었으며, 실제 학습자 적용을 통한 효과성 검증과 신뢰도 분석은 수행하지 않았다. 본 검사도구의 내적 일관성 신뢰도(Cronbach's α)는 학습자 응답 자료의 수집을 전제로 하므로, 후속 연구에서는 본 모델을 학습자에게 적용하여 사전 검사와 사후 검사를 통해 효과성을 검증하고, 학습자 응답 자료를 바탕으로 문항 분석과 내적 일관성 신뢰도 산출을 함께 수행함으로써 검사도구의 양적 타당화를 완성할 필요가 있다.

둘째, 국어교육 전문가 자문이 수행되지 않았다. 본 연구는 SPARK-L을 AI 융합 교수학습 프레임워크로 위치시키며 AI와 디지털 교육 및 교육평가 전문가 중심으로 패널을 구성하였으나, 국어과 적용 시의 교과 특수성에 대한 자문은 다루지 않았다. 후속 연구에서는 국어교육 전문가의 자문을 통해 교과 특수적 적용 가이드라인을 보완할 필요가 있다. 또한 본 연구의 전문가 패널은 9명으로 일반적으로 권장되는 델파이 패널 규모보다 작다. 본 연구는 9명 기준의 엄격한 CVR 임계값(.78)을 적용하여 합의 판정의 보수성을 확보하였으나, 후속 연구에서는 보다 큰 규모의 패널을 통한 재검증이 필요하다.

셋째, 본 연구의 모델과 검사도구는 단일 연구학교 맥락에 기반하여 개발되었다. 본 연구의 모델과 검사도구는 2026학년도 경기도교육청 지정 A 초등학교 AI디지털 활용 연구학교를 위해 개발되었으므로, 다른 학교와 학년, 교과 맥락에서의 적용 가능성은 별도의 검증이 요구된다. 후속 연구에서는 여러 학교와의 협력을 통해 모델의 적용 범위를 확장한 검증이 필요하다.

넷째, 본 모델을 다른 교과나 학년으로 확장할 때에는 단계별 조정이 요구된다. 본 모델은 특정 교과에 종속되지 않는 AI 융합 교수학습 프레임워크이나, 일부 구성요소가 초등 5~6학년 국어과 맥락에 구체화되어 있기 때문이다. 특히 S단계의 데이터 스캐폴딩 자료와 K단계의 산출물 양식은 교과 교육과정 목표에 맞추어 재구성되어야 하며, R단계의 주도권 전환 기제는 학년에 따라 스캐폴딩 수준과 어휘, 차시 운영을 발달 단계에 맞게 조정해야 한다. 또한 검사도구의 일부 문항이 특정 교과와 주제 맥락에 기반하므로, 교과 특

수성을 반영한 문항 재진술과 별도의 내용타당도 재검증이 요구된다.

다섯째, 본 모델은 도구 중립적 설계를 지향하나 실제 적용 시에는 도구별 제약이 존재한다. 본 연구의 대상인 초등 5~6학년은 만 11~12세로 일반 소비자용 생성형 AI 도구의 최소 이용 연령(만 13세) 미만에 해당하므로, 현재 시점에서는 엔터프라이즈급 데이터 보호가 적용되고 전 연령이 이용 가능한 Google Workspace for Education의 Gemini for Education을 통해 적용하는 것이 현실적이다. 이 경우 학습자의 대화 자료가 모델 학습에 사용되지 않으며, 관리자 설정과 학부모 동의를 전제로 운영된다. 다만 이미지 생성 등 일부 기능은 만 18세 미만에 제한되고, 도구별 출처 제시 방식과 한국어 처리 성능 차이가 A단계 팩트체크 운영에 영향을 줄 수 있으므로, 현장 적용 시 학교 환경에 맞춘 도구 선택 가이드가 필요하다.

참고문헌

[1] Cho, J. (2024). *The effect of generative artificial intelligence education using problem-based learning (PBL) on elementary students' artificial intelligence literacy* [Master's thesis]. Cheongju National University of Education.

[2] OECD. (2018). *The future of education and skills: Education 2030*. OECD Publishing. https://doi.org/10.1787/9789264293045-en

[3] Risko, E., & Gilbert, S. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

[4] Shin, J., Park, P., & Bae, Y. (2023). Educational introduction of generative AI: A narrative literature review. *Journal of The Korean Association of Information Education, 27*(6), 665-681.

[5] Kim, J. (2023). Generative AI, homo cogitans, and the direction of literacy education. *Ratio et Oratio, 16*(2), 7-31. https://doi.org/10.19042/kstc.2023.16.2.7

[6] Park, S., & Lee, Y. (2025). Educational uses of generative artificial intelligence: On hallucinations and supplementary thinking. *Journal of Advanced Technology Convergence, 4*(2), 213-223. https://doi.org/10.23152/JATC.2025.04.02.213

[7] Choi, C. (2023). Preliminary reflections on converging elementary Korean language and AI literacy education. *Journal of Elementary Korean Education, 76*, 434-458.

[8] Park, J. (2025). Development and effectiveness analysis of elementary artificial intelligence literacy-based education program. *Journal of Practical Engineering Education, 17*(1), 139-149. https://doi.org/10.14702/JPEE.2025.139

[9] Cho, E., & Lee, S. (2024). Study on the patterns and attitudes of elementary school students in utilizing generative artificial intelligence in classroom settings. *The Journal of Korean Association of Computer Education, 27*(4), 21-39. https://doi.org/10.32431/kace.2024.27.4.003

[10] Hwang, H., Hwang, Y., Park, J., Shin, M., & Lee, H. (2024). A study on the development and validity verification of measurement tool for AI literacy. *The Korean Journal of Literacy Research, 15*(2), 247-278. https://doi.org/10.37736/KJLR.2024.04.15.2.09

[11] Kim, S., & Lee, Y. (2022). The artificial intelligence literacy scale for middle school students. *Journal of The Korea Society of Computer and Information, 27*(3), 225-238. https://doi.org/10.9708/jksci.2022.27.03.225

[12] Ryu, M., & Han, S. (2023). The study on test standard for measuring AI literacy. *Journal of The Korea Society of Computer and Information, 28*(7), 39-46. https://doi.org/10.9708/jksci.2023.28.07.039

[13] Yoon, S., & Yang, C. (2022). Development and validity study of an AI literacy assessment tool for elementary students. *Proceedings of the Winter Conference of the Korean Association of Information Education*, Jeju, Korea, 17-22.

[14] Gyeonggi Provincial Office of Education. (2025). *Guide for question-based inquiry instruction.* Gyeonggi Provincial Office of Education. https://www.goe.go.kr/goe/na/ntt/selectNttInfo.do?nttSn=2328737&mi=10961

[15] Bybee, R., Taylor, J., Gardner, A., Van Scotter, P., Powell, J., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins, effectiveness, and applications.* BSCS.

[16] Hmelo-Silver, C. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

[17] Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? P*roceedings of the AAAI Conference on Artificial Intelligence, Honolulu, USA, 33*(1), 9795-9799. https://doi.org/10.1609/aaai.v33i01.33019795

[18] HHwang, K., & Cha, K. (2026). Effects of the interaction between AI chatbots and learning stages on learning performance: A cognitive load theory-based analysis. *Korean Journal of Digital Industry, 31*(1), 179-198.

[19] Lawshe, C. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

[20] Linstone, H., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications. Addison-Wesley*.

[21] Hsu, C., & Sandford, B. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation, 12*(10), 1-8. https://doi.org/10.7275/pdz9-th90

[22] Grant, J., & Davis, L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health, 20*(3), 269-274. https://doi.org/10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

김성록

· 2022년 경인교육대학교 교육전문대학원 컴퓨터교육과(교육학석사)
· 2022년~현재 경인교육대학교 교육대학원 컴퓨터교육과(박사과정)
· 2023년~현재 화성태안초등학교 교사

관심분야 : 피지컬컴퓨팅교육, AI·SW교육, 영재교육, 에듀테크

edurainbow89@gmail.com

허경

· 1998년 고려대학교 전자공학과(공학사)
· 2000년 고려대학교 전자공학과 (공학석사)
· 2004년 고려대학교 전자공학과 (통신공학박사)
· 2004~2005년 삼성종합기술연구원(SAIT) 전문연구원
· 2005년~현재 경인교육대학교 컴퓨터교육과 교수

관심분야 : 피지컬컴퓨팅교육, SW교육, AI교육, 데이터과학교육

khur@ginue.ac.kr

부 록

〈표 1〉 탐구 기반 교수학습 모델과 AI 교육 프레임워크 비교

모델	단계 구조	AI·도구의 역할	학생 주도성	평가 방식	교과 적용 맥락
5E 학습순환 (Bybee 외, 2006)	참여→탐색→설명→정교화 →평가 (5단계 선형)	기존 프레임워크에 AI 미포함, 일반 탐구 자료·실험 도구를 활용	탐색·정교화 단계에서 강조, 교사의 구조화된 안내를 전제	평가 단계를 명시적으로 규정, 형성·총괄 평가를 통합	과학과에서 출발하여 전 교과로 확장
문제기반학습(PBL) (Hmelo-Silver, 2004)	비구조화 문제→문제 분석 →정보 수집→해결안 개발 →발표·성찰 (변형 다수)	기존 프레임워크에 AI 미포함	매우 높음; 문제 정의부터 해결까지 학생이 주도	산출물·과정 중심 평가; 동료 평가 빈번	의학교육에서 출발; 실생활 문제 중심으로 교과 횡단
질문기반 탐구수업 (경기도교육청, 2025)	준비→탐구→정리→실천 (성찰이 내재된 4단계 순환)	AI 미통합; 디지털 도구를 탐구 자원의 한 범주로 위치	3단계 스펙트럼(구조화-안내-개방형), 학생 질문 형성을 독립 영역으로 다룸	GRASPS 기반 수행평가, 개념적 이해의 전이에 초점	초등 전 교과
AI4K12 빅 아이디어 (Touretzky 외, 2019)	인식·표현과 추론·학습·자연적 상호작용·사회적 영향 (단계가 아닌 내용 체계인 5개 빅 아이디어)	AI 자체가 학습 대상(도구가 아닌 내용)	단계별 주도성 미규정, 교육과정 지침으로 기능	평가 체계 미포함	K-12 AI 교육 전반 (교과 비종속 프레임워크)

〈표 4〉 SPARK-L 교수학습 모델 초안 (1차 조사 제출본)

영역	단계 / 구성요소	핵심 전략
교수학습 흐름 (SPARK 5단계)	S 상황 발견	교사가 정선한 데이터를 통해 학습자가 삶의 맥락(AI 적용 사회 현상, 기후 위기 등)에 자리한 문제를 발견하고, 학습 미션과 공동의 목적을 수립한다.
	P 질문 구체화	학생이 AI에 역할을 부여하고, 교사와 AI의 도움을 받아 아이디어를 탐구 가능한 핵심 질문으로 정교화한다.
	A AI 협력 탐구	클라우드 기반 생각노트를 활용해 AI와 상호작용하며 정보를 수집하고 생각을 교환한다.
	R 논리적 구조화	AI 상호작용을 일시 중단(주도권 전환)하고, 학생이 아날로그 노트로 논리망(Logic-tree) 구축을 주도하며 AI가 제공한 정보를 비판적으로 검증한다.
	K 지식 창출	단순한 AI 산출물 소비를 넘어, 사회적 가치를 지닌 AI 협력 산출물을 발간한다.
평가 체계	디지로그 평가 체계	확산적 사고는 디지털 도구로, 수렴·구조적 사고는 아날로그 도구로 이원화하여 평가하는 체계로, 학생의 문해력 보호와 AI 의존성 방지를 지향한다.
	프롬프트 과정 평가	산출물 자체가 아니라 학생이 AI에게 던진 질문의 정교화 과정과 대화 이력을 핵심 평가 척도로 삼는다.
	팩트체크 및 윤리 평가	AI 답변을 그대로 복사하지 않고, 사실 확인 여부와 윤리적 재구성 과정을 평가 항목에 포함한다.
	고쳐 쓰기 중심 평가	초안 작성 후 AI·교사의 피드백을 반영하여 글을 개선한 고쳐 쓰기의 변화 폭을 성장 지표로 삼는다.
	일반화 가능성	모델의 질문-탐구-글쓰기 과정이 국어과를 넘어 타 교과의 서·논술형 수행평가 모델로 일반화되도록 설계한다.

〈표 5〉 SPARK-L 주체적 AI 리터러시 검사도구 초안의 측정 영역 (3차 조사 제출본)

문항	역량	측정 영역	SPARK 단계	출처
1	AI 이해	AI 작동 원리에 대한 인지적 이해	토대	정련
2	프롬프트 설계	AI를 위한 구체적 질문 설계 능력	P단계	신규
3	AI 도구 응용	학습 과제 해결을 위한 AI 활용	A단계	차용
4	정보 교차 검증	AI 답변의 출처 검증	A단계	신규
5	환각 현상 인지	AI 답변에 포함된 오류·편향 가능성 인식	A단계	차용
6	주체적 구조화	자신의 논리에 따른 AI 정보 재구성	R단계	신규
7	피드백 수용 및 수정	형성적 피드백을 통한 글쓰기 개선	K단계	신규
8	디지털 창의 표현	디지털 도구를 활용한 시각적 표현	K단계	신규
9	협력적 소통	모둠 활동에서의 협력 역량	프로젝트 전반	신규
10	질적 성찰	AI 수정 경험에 대한 메타인지적 성찰	메타인지	신규

〈표 9〉 3차 델파이 조사 미합의 문항의 수정 전후 진술

문항	3차 진술 (수정 전)	4차 진술 (수정 후)	수정 방향
1. AI 이해	나는 AI가 스스로 생각하는 것이 아니라 많은 데이터를 학습하여 답을 만들어 낸다는 것을 안다.	나는 AI가 많은 양의 데이터를 학습하여 질문에 맞는 답을 만들어 내는 원리를 이해한다.	학술적 표현의 명료화
3. AI 도구 응용	나는 동화책 만들기라는 목표를 이루기 위해 내 수준에서 AI 도구를 사용할 수 있다.	나는 주어진 학습 과제나 문제를 해결하기 위해 목적에 맞게 AI 도구를 사용할 수 있다.	특정 산출물 표현의 일반화
7. 피드백 수용 및 수정	나는 AI·교사·친구가 준 조언을 비판적으로 판단하여 내 글을 더 자연스럽고 매끄럽게 고쳐 쓸 수 있다.	나는 AI·교사·친구의 피드백을 바탕으로 내 글의 부족한 부분을 더 나은 방향으로 고쳐 쓸 수 있다.	측정 역량의 재정의
8. 디지털 창의 표현	나는 캔바(Canva)를 활용해 내 이야기의 분위기와 내용에 가장 잘 어울리는 이미지와 디자인을 창의적으로 표현할 수 있다.	나는 디지털 도구를 활용해 내 이야기의 분위기와 내용을 시각적으로 표현하여 독자에게 전달할 수 있다.	도구 종속성 제거

〈표 11〉 최종 SPARK-L 교수학습 모델

영역	단계 / 구성요소	핵심 전략
Ⅰ. 교수학습 흐름 (SPARK 5단계)	S 상황 발견 (1차시)	교사가 정선한 자료(영상·자료)를 통해 학습자가 삶의 맥락에 자리한 문제를 발견하고, 이를 해결하기 위한 미션과 공동의 목적을 수립한다.
	P 질문 구체화 (2차시)	학생이 브레인스토밍으로 아이디어를 생성하고, 교사와 AI의 도움을 받아 탐구 가능한 핵심 질문으로 정교화한다.
	A AI 협력 탐구 (3차시)	클라우드 생각노트를 활용해 AI와 상호작용하며 정보를 수집하고, 출처 교차 검증(2개 이상)을 통한 팩트체크를 기록한다.
	R 논리적 구조화 (4차시)	AI 상호작용을 일시 중단(주도권 전환)하고, 학생이 논리망(Logic-tree)과 시각적 개요 구성을 주도한다. 도구 선택은 자유롭게 하며(포스트잇, 마인드맵, 디지털 타이핑 등), 스캐폴딩이 필요한 학생에게는 단계별 구조화 템플릿을 제공한다.
	K 지식 창출 (5~6차시)	학습관리시스템(LMS)에 업로드한 초안에 AI의 형성적 피드백과 교사의 심층 지도를 비판적으로 반영하고, 자기 점검 전략(성찰 노트 등)을 적용해 글의 통일성을 고쳐 쓴 뒤 학급 단위 산출물로 공유한다.
Ⅱ. 평가 체계	의도된 데이터 스캐폴딩	교사가 교육과정 목표에 부합하는 자료를 사전 정선하여 학습자의 주체적 문제 발견을 보호한다.
	평가의 책무성	최종 성취 판정은 교사 루브릭 기반 총괄 평가로 수행하고, AI는 형성적 피드백 도구로 한정한다.
	성장 지표의 재정의	단순한 분량 증가나 수정 폭이 아니라, 비판적 피드백 수용도(선택·기각 사유 등 포트폴리오 증거)와 산출물 완성도를 핵심 지표로 삼는다.
	탐구 규칙 (팩트체크 3원칙)	출처 교차 검증(2개 이상), 객관적 근거에의 명시적 기반, 비판적 검토를 탐구의 기본 규칙으로 적용한다.
	논리망 구축의 구체화	학생이 AI 정보를 기계적으로 복제하지 않고 자신의 논리 구조(예: 인과 관계)로 재구성했는지를 마인드맵·개요표 등 산출물로 평가한다.
	현장 실행 가능성	초등 5~6학년에서 4~8차시의 주제중심 질문기반 프로젝트로 설계하여, 차시 구조·도구·평가 절차를 학교 단위 실행에 통합한다.

〈표 12〉 평가 체계 구성요소 변경 경로

1차 → 최종	처리	사유
디지로그 평가 체계 → 의도된 데이터 스캐폴딩	재정의 (1차 CVR .56)	디지로그 용어의 학술적 근거 부족을 해소하기 위해 S단계의 교사 사전 정선 원칙으로 재구성하였다.
프롬프트 과정 평가 → 성장 지표의 재정의	통합 (1차 CVR .78)	두 항목 모두 산출물이 아닌 과정의 증거를 평가 척도로 삼는 공통 원리를 지녀, 개념적 정합성에 근거하여 통합하였다.
팩트체크 및 윤리 평가 → 탐구 규칙 (팩트체크 3원칙)	격상 (1차 CVR 1.00)	평가 기준에서 운영 규칙으로 위치를 이동하고 3원칙으로 구체화하였다.
고쳐 쓰기 중심 평가 → 논리망 구축의 구체화	재정의 (1차 CVR .56)	수정 분량 측정의 한계를 보완하여 R 단계의 정보 재구성 증거에 초점을 재설정하였다.
고쳐 쓰기 중심 평가 → 논리망 구축의 구체화	재정의 (1차 CVR .56)	4~8차시 단위의 구체적 실행 가능성으로 재구성하였다.
– → 평가의 책무성	2차 신규 도입	AI를 형성적 피드백으로 한정하고, 교사 루브릭을 총괄 판정의 단독 근거로 지정하였다.

〈표 13〉 최종 SPARK-L 주체적 AI 리터러시 검사도구

문항	역량	SPARK 단계	문항 진술
1	AI 이해	토대	나는 AI(챗GPT, 뤼튼 등)가 많은 양의 데이터를 학습하여 질문에 맞는 답을 만들어 내는 원리를 이해한다.
2	프롬프트 설계	P단계	나는 원하는 결과를 얻기 위해 역할이나 조건을 포함한 구체적인 질문을 AI에게 할 수 있다.
3	AI 도구 응용	A단계	나는 주어진 학습 과제나 문제를 해결하기 위해 목적에 맞게 AI 도구를 사용할 수 있다.
4	정보 교차 검증	A단계	나는 AI가 제공한 정보가 정확한지 다른 자료(책, 인터넷 검색 등)를 찾아보며 확인(팩트체크)한다.
5	환각 현상 인지	A단계	나는 AI가 만들어 낸 정보가 때때로 틀리거나 편향되거나 한쪽으로 치우칠 수 있음을 알고 있다.
6	주체적 구조화	R단계	나는 AI의 답변을 그대로 베끼지 않고 내 생각과 논리에 따라 정보를 재구성할 수 있다.
7	피드백 수용 및 수정	K단계	나는 AI·교사·친구의 피드백을 바탕으로 내 글의 부족한 부분을 더 나은 방향으로 고쳐 쓸 수 있다.
8	디지털 창의 표현	K단계	나는 디지털 도구를 활용해 내 작품의 분위기와 내용을 시각적으로 표현하여 독자에게 전달할 수 있다.
9	협력적 소통	프로젝트 전반	나는 모둠 활동에서 역할을 나누고 다른 사람의 의견을 존중하며 함께 과제를 해결할 수 있다.
10	질적 성찰	메타 인지	프로젝트를 진행하며 AI의 답변을 내 방식대로 고쳐 본 경험이 있다면, 어떤 부분을 왜 고쳤는지 구체적으로 적어 주세요.