



## 수능 영어 빈칸추론 문항은 수험생들의 추론 능력을 측정하는가?\*

김정은 (전북대학교) · 김현진 (아주대학교) · 이선빈 (서울대학교)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: March 8, 2025  
Revised: March 29, 2025  
Accepted: April 18, 2025

Jeong-eun Kim (first author)  
Professor, Department of English  
Language and Literature  
Jeonbuk National University  
567 Baekje-daero, Deokjin-gu,  
Jeonju-si, Jeollabuk-do, 54896  
Republic of Korea  
Email: jek48@jbnu.ac.kr

Hyun-Jin Kim (corresponding author)  
Professor, Dasan University College  
Ajou University  
206 Worldcup-ro, Yeongtong-gu  
Suwon-si, Gyeonggi-do, 16499  
Republic of Korea  
Email: hjinkim@ajou.ac.kr

Seon-bin Lee  
Graduate student, Department of  
English Language Education  
Seoul National University  
1 Gwanak-ro, Gwanak-gu,  
Seoul, 08826 Republic of Korea  
Email: slee24@snu.ac.kr

\* 이 논문은 2024 년도  
한국중등영어교육학회의 Special  
Interest Group(연구팀명:  
국가주도영어시험연구)의 지원을  
받아 수행된 연구임.

### ABSTRACT

Kim, Jeong-eun, Seon-bin Lee and Hyun-Jin Kim. 2025. Can the CSAT English gap-filling questions assess test-takers' inference skills? *Korean Journal of English Language and Linguistics* 25, 562-579.

This study investigates the validity of gap-filling inference questions in the Korean CSAT English section by tracking learners' problem-solving processes. Verbal data were collected from 51 university students, using the think-aloud protocol, as they engaged in gap-filling inference questions. We analyzed how accuracy rates were related to problem-solving time, the frequency and duration of inferencing, and various learner-internal variables such as English proficiency and topic familiarity. The results indicate that topic familiarity was the only significant predictor of accuracy, whereas English proficiency, problem-solving time, the frequency of inferencing occurrences, and the duration of inferencing did not show significant effects on accuracy rates. Additionally, participants reported that vocabulary difficulty, sentence complexity, and challenges in distinguishing answer choices were key factors contributing to the perceived difficulty of gap-filling inference questions. These findings suggest a need to reassess the validity of gap-filling inference questions in the CSAT and raise questions about whether they effectively measure inferential reasoning skills.

### KEYWORDS

수능 영어, 빈칸 추론, 추론 능력, 사고 구술법, 로지스틱 회귀분석 (CSAT English, Gap-filling Questions, Inference, Think-aloud Protocol, Logistic Regression Analysis)

## 1. 서론

대학수학능력시험(이하 수능)의 영어 영역은 전국 단위로 시행되는 고교 졸업자 대상의 대표적인 국가 평가로서, 고등학교 교육과정 성취기준의 달성 여부와 대학 수준의 영어 활용 역량을 동시에 측정하는 것을 목적으로 한다(한국교육과정평가원 2024). 본 시험은 총 45 문항으로 듣기 영역 17 문항과 읽기 영역 28 문항으로 나뉜다. 수능 영어는 수능 과목 중 유일하게 절대평가가 적용되기 때문에 수험생의 영어 실력을 객관적으로 평가하여 정확하게 등급을 판별하는 것이 매우 중요하다. 이에 따라 문항 변별력의 확보는 매년 수능 영어 평가의 핵심 과제이다.

읽기 영역에 포함된 빈칸 추론 문항은 상위 등급 판별에 결정적 영향을 미치는 핵심 유형 중 하나로, 일반적으로 4 문항이 출제된다. 수험생은 이 문항에서 글의 핵심 내용과 논리 전개를 종합적으로 파악한 후, 문맥에 가장 적합한 표현을 추론해야 한다. 이러한 빈칸 추론 문항은 고난이도 문제로 자주 지목되며, 수험생들 사이에서도 높은 인지적 부담을 유발하는 유형으로 인식된다. 예를 들어, EBSi 문항별 정답률 분석에 따르면 2024 학년도 수능 영어 33 번 문항의 정답률은 14.3%로 매우 낮았으며, 오답 선지의 선택률도 10~20%대로 고르게 분포되었다. 이는 수험생들이 정답 선택에 있어 확신을 가지지 못했고, 상위권 학생들조차 정답 도출에 어려움을 겪었음을 시사한다. 또한 해당 문항은 2023 년 기준으로 단일 문항 중 가장 많은 이의 신청이 접수되었고(베리타스알파 2023), 정답의 타당성과 해석 가능성에 대한 사회적 논란을 촉발한 바 있다. 이러한 사례는 빈칸 추론 문항이 과연 학습자의 추론 능력을 타당하게 평가하고 있는지에 대한 근본적 의문을 제기한다.

추론은 단순한 정보 회상을 넘어, 맥락과 배경지식을 활용하여 논리적으로 의미를 구성하는 고차원적 인지 활동으로 정의된다(Graesser et al. 1994). 수능 영어의 빈칸 추론 문항은 제한된 시간 내에 텍스트를 분석하고 문맥적 단서를 바탕으로 정답을 도출해야 하므로, 독해 속도, 배경지식, 인지적 유연성 등이 복합적으로 작용한다. 그러나 지문이 지나치게 복잡하거나 난해할 경우, 수험생은 충분한 이해를 바탕으로 한 추론 수행보다는, 정답을 전략적으로 선택하거나 임의로 답을 고르는 방식으로 대응할 수 있다. 이는 결과적으로 빈칸 추론 문항이 본래 의도한 추론 능력 평가 기능을 제대로 수행하고 있는지에 대한 재검토 필요성을 제기한다.

이에 본 연구는 수능 영어 빈칸 추론 문항이 실제로 수험생의 추론 능력을 평가하는 타당한 도구로 기능하고 있는지를 실증적으로 검토하는 것을 목적으로 한다. 이를 위해 학습자들이 해당 문항을 해결하는 과정에서 소요한 시간과 추론 활동의 발생 여부 및 특성을 정량적·정성적으로 분석하였다. 또한, 영어 능력, 주제 친숙도, 문제 풀이 시간, 추론 발생 빈도 및 지속 시간과 같은 학습자 변인이 정답률에 미치는 영향을 탐색하였다. 아울러, 수험생들이 빈칸 추론 문항을 고난도 문항으로 인식하게 되는 주요 원인을 규명함으로써, 이 유형의 문항이 평가 도구로서 가지는 한계와 개선 필요성을 함께 진단하고자 한다.

본 연구는 빈칸 추론 문항이 수험생에게 요구하는 인지적 부담을 실증적으로 평가하고, 수능 영어 평가 문항의 타당성과 신뢰성을 제고하는 데 기초 자료를 제공함으로써, 향후 공정하고 효과적인 영어 평가 설계에 기여할 수 있을 것이다.

## 2. 연구 배경

### 2.1 수능 영어 추론 문항 운영 실태

수능 영어에서 빈칸 추론 문항은 본래 학습자의 추론 능력을 평가하기 위해 도입되었다. 이 유형의 문제는 글의 논리적 흐름을 파악하고 문맥적 단서를 활용하여 적절한 표현을 추론하는 능력을 측정하는데 초점을 둔다. 일반적으로 빈칸은 지문의 중심 내용이나 핵심 세부 정보와 밀접하게 연결되어 있으며, 출제자는 정답과 관련된 단서가 존재하는 문장을 기반으로 문제를 구성한다. 따라서 이 문항은 학생들이 단순한 해석 능력을 넘어 텍스트 전반의 논리 구조를 이해하고 적용할 수 있도록 설계되어야 한다.

그러나 실제 운영에서는 빈칸 추론 문항이 본래의 평가 목적을 넘어, 상위 등급 학생을 추려내는 도구로 활용되고 있다. 이러한 문항은 대체로 3 점 배점으로 설정되며, 수험생들에게 가장 큰 부담을 주는 요소로 작용하고 있다. 높은 난이도와 복잡한 문항 구조로 인해 ‘킬러 문항’이라는 별칭이 붙었으며, 정답률이 매우 낮다는 특징이 있다. 예를 들어, EBSi의 분석에 따르면 2025 학년도의 빈칸 추론 문항의 정답률은 14.3%~23%로 20%에 근사하거나 이를 넘지 못하는 수준을 보인다. 이에 따라 절대평가 체제에서도 빈칸 추론 문항은 상위 등급 학생들을 선별하는 핵심 역할을 수행하는 것으로 보이며, 수험생들에게는 난이도가 매우 높지만, 1 등급을 받기 위해서는 반드시 해결해야 하는 문항으로 인식되고 있다(매일경제 2022).

이러한 현상은 빈칸 추론 문항이 본래 의도한 평가 목적을 충실히 수행하고 있는지, 혹은 단순히 상위 등급 학생을 줄이기 위한 도구로 활용되고 있는지에 대한 근본적인 의문을 제기한다. 따라서 문항의 본질적인 역할과 타당성을 점검하기 위해서는 학생들의 실제 문제 풀이 양상을 면밀히 분석할 필요가 있다. 예컨대, 학습자들이 빈칸 추론 문항을 해결하는 과정에서 문항이 본래 의도한 논리적 추론 능력을 활용하는지를 검토하고, 극도로 낮은 정답률이 학습자들의 잘못된 추론에 기인한 것인지, 혹은 애초에 추론이 발생하지 않은 결과인지에 대한 분석이 필요하다. 만약 학생들이 시간 압박이나 지문의 난해함으로 인해 논리적 추론을 시도조차 하지 못하고 있다면, 학습자들이 느끼는 추론 방해 요소는 무엇인지 파악해야 할 것이다. 이를 바탕으로, 빈칸 추론 문항이 본래의 평가 목적을 보다 효과적으로 달성할 수 있도록 개선하는 방안을 모색해야 한다.

### 2.2 읽기에서의 추론적 사고와 과정

추론(inference)은 독자가 명시적으로 제공되지 않은 정보를 기존 지식과 문맥을 활용하여 도출하는 인지적 과정으로, 언어 이해 및 사용에 중요한 역할을 한다(Graesser et al. 1994, van Dijk and Kintsch 1983). Bloom(1956)의 교육 목표 분류에서 추론은 고차원적 사고(higher-order thinking)로 간주되며, 단순한 암기나 기계적 학습을 넘어 분석, 평가, 창의적 사고를 요구하는 과정이다.

독해 과정에서 독자는 텍스트의 단서와 배경 지식을 결합하여 암시적 정보를 도출하고, 문장의 의미를 심층적으로 이해한다. 추론은 일반적으로 국지적(local) 추론과 전반적(global) 추론으로 나뉘며, 전자는 문장 간 관계 형성에, 후자는 전체 의미 통합에 초점을 둔다(Chikalanga 1992, Graesser et al. 1994). 국지적 추론은 정보 간 거리가 가까울수록 쉽게 이루어지지만, 정보 간 거리가 멀어질수록 수행이 어려워진다(Johnston et al. 2008). 따라서 독자의 개별적인 특성이 성공적인 추론 수행에

중요한 역할을 하며, 이는 독해 성취에 직접적인 영향을 미친다.

추론은 독자가 텍스트의 의미를 구성하는 데 필수적인 과정으로, 정보 인식, 의미 통합, 검증 및 조정의 세 단계로 이루어진다(Graesser et al. 1994). 먼저, 정보 인식 단계 (Information Recognition Stage)에서는 독자가 텍스트에서 제공되는 단서(cues)와 명시적 정보를 인식하는 과정이 이루어진다. 이 단계에서 독자는 단어, 문장 구조, 문맥적 단서를 활용하여 정보를 추출하고, 이를 자신의 배경지식과 연결할 준비를 하게 된다(Kintsch 1998). 이어지는 의미 통합 단계 (Meaning Integration Stage)에서는 인식된 정보를 논리적으로 결합하고, 텍스트 내부 및 외부 지식을 활용하여 의미를 확장하는 작업이 이루어진다. 이때 사용되는 주요 전략은 명제적(propositional) 추론과 실용적(pragmatic) 추론이다. 전자는 텍스트 내 논리적 연결을 기반으로 한 필연적 추론이며, 후자는 독자의 배경지식과 경험에 기반한 가능성 있는 해석이다(Chikalanga 1992, Gernsbacher et al. 1990). 독해 과정에서는 문장 간 연결뿐 아니라 텍스트의 논리 구조를 파악하고, 새로운 정보를 기존의 인지적 틀(schema)과 결합하는 과정이 포함된다(Rumelhart 1980). 마지막으로, 검증 및 조정 단계 (Validation and Adjustment Stage)에서는 독자가 생성한 추론의 타당성을 평가하고 필요에 따라 조정하는 과정이 진행된다. 이 단계는 기억 기반 모델(memory-based model)과 구성주의 모델(constructivist model)이라는 두 관점에서 설명될 수 있다(van den Broek et al. 2005). 전자에 따르면, 독자는 관련 정보를 즉각적으로 떠올리고 자동적으로 추론을 수행하며(McKoon and Ratcliff 1992), 후자는 독자가 능동적으로 의미를 구성하고 필요 시 추론을 조정한다고 본다. 이 과정에서 인지적 부담(cognitive load)과 배경 지식은 중요한 영향을 미친다(Kendeou et al. 2003, van den Broek 1994).

요약하면, 추론은 단순한 정보 회상을 넘어서 텍스트의 암시적 의미를 구성하고, 배경지식과 문맥을 종합적으로 활용하는 복잡한 인지 과정이다(Graesser et al. 1994, McKoon and Ratcliff 1992, van den Broek 1994). 효과적인 추론이 이루어지기 위해서는 세 단계가 유기적으로 연결되어야 하며, 독자는 논리적 사고를 통해 새로운 정보를 기존 지식과 통합해야 한다(van den Broek et al. 2001).

이러한 사고 과정은 독해 능력이 낮은 학습자에게 특히 부담이 될 수 있으며(Kendeou et al. 2008), 추론 수행은 언어 능력, 배경지식, 작업 기억 용량, 인지적 유연성, 독해 전략 사용 등 다양한 개인차 변수의 영향을 받는다(Johnston et al. 2008, McKoon and Ratcliff 1992). 즉, 추론의 성공 여부는 단순한 지문 이해도를 넘어서 학습자의 인지 능력과 전략적 사고 방식에 의해 결정될 수 있다.

### 2.3 수능 시험의 추론 능력에 영향을 미치는 잠재적 요인

추론에 영향을 미치는 개인차 변인을 살펴보는 일은 수능 영어의 추론 문항이 공정하게 기능하는지를 정밀하게 이해하는 데 중요한 시사점을 제공한다. 다양한 학습자 변인 중에서도 주제 친숙도와 영어 능숙도는 추론 전략의 활용 및 독해 정확도에 핵심적인 영향을 미치는 요인으로 보고된다(Hall 2016, Hammadou 1991, Kim and Lee 2020). 일반적으로 지문의 주제에 익숙한 학습자는 텍스트를 더 깊이 이해하고, 효과적으로 추론할 가능성이 높다.

Hammadou(1991)의 연구에 따르면, 독해 능숙도가 높은 학습자는 더 정교한 추론을 수행하며, 이 과정에서 주제 친숙도가 독해에 미치는 간접적 영향이 드러난다. 특히, 이러한 효과는 초보 학습자보다 숙련된 학습자에게서 더 뚜렷하게 나타나, 학습자의 사전 지식이 추론 과정과 독해 성취에 중요한

역할을 함을 시사한다.

한편, Kim 과 Lee(2020)의 연구에서는 과학 지문을 읽을 때 주제 친숙도와 영어 능숙도 모두가 독해 이해에 유의미한 영향을 미치는 것으로 나타났다. 반면, 비과학 지문의 경우에는 사전 지식의 영향이 크지 않았고, 영어 능숙도가 독해 성과를 결정짓는 핵심 요인으로 작용했다. 이러한 결과는 학습자가 특정 주제에 익숙할수록 추론 활동이 보다 효과적으로 이루어질 수 있으며, 이는 빈칸 추론 문항에서도 유사한 경향이 반영될 가능성을 시사한다.

선행 연구에 따르면, 주제 친숙도는 다양한 유형의 추론 과정에서 중요한 역할을 하지만, 특히 빈칸 추론(gap-filling inference) 문항에서 그 영향이 두드러지는 것으로 나타난다(Hall 2016). 빈칸 추론은 텍스트의 일관성을 유지하기 위한 결속성 추론(coherence inference)에 비해 더 높은 수준의 의미 구성 능력을 요구하며, 학습자는 단순한 정보 회상을 넘어 배경지식을 활용해 논리적 연결을 도출해야 한다(Perfetti et al. 2005). 이러한 이유로, 배경지식이 부족한 학습자는 빈칸 추론 문항에서 원활하게 추론을 수행하지 못할 가능성이 높으며, 이로 인해 해당 문항을 높은 난이도의 문제로 인식할 수 있다. 수능 영어의 추론 문항이 주로 빈칸 추론 형태로 출제된다는 점을 고려할 때, 주제 친숙도는 이러한 문항 해결에 핵심적인 역할을 할 수 있다.

시간 제한이 있는 시험이라는 특수한 맥락에서, 학습자가 문항 풀이에 소요한 절대적 시간은 추론 성공 여부를 좌우할 수 있는 중요한 변인이다. 효과적인 추론을 위해서는 정보를 빠르게 처리하는 능력뿐 아니라, 충분한 시간을 투입하여 내용을 정밀하게 분석하는 과정 또한 필수적이다. 예를 들어, 독해 시간이 확보된 학습자는 문맥적 단서를 면밀히 검토하고 논리적 연결을 깊이 있게 분석함으로써 추론의 정확성을 높일 수 있다 (Just and Carpenter 1980). 반면, 시간이 부족할 경우 정보 처리 부담이 증가하고 핵심 단서를 놓칠 가능성이 높아져, 추론 과정에서 오류가 발생할 위험이 커진다(Rayner 1998). 따라서 수능과 같은 평가 상황에서는 단순한 독해 속도보다, 학습자가 추론 과정에 투입한 시간과 그 활용 방식이 추론 성과를 결정하는 핵심 요인으로 작용할 수 있다.

아울러, 추론 발생 빈도와 지속 시간 또한 학습자의 추론 능력 및 정답률에 영향을 미칠 수 있다. 비록 이들 변인이 정답률과의 관계를 직접적으로 규명한 연구는 제한적이지만, 기존 연구를 통해 간접적 시사점을 얻을 수 있다. 예컨대, 김희주(2018)는 추론 능력이 독해 성취도의 주요 예측 요인임을 밝혔고, Cromley 와 Azevedo(2007)는 읽기 이해에서 추론의 필수적 역할을 강조하였다. 특히, 빈번하고 지속적으로 추론이 발생할수록 문맥 내 의미 연결이 활발히 이루어지고, 정보 통합이 촉진된다. 이로 인해 학습자는 텍스트의 심층적 의미를 구성하게 되며, 이는 정교한 개념적 이해로 이어질 수 있다.

이처럼 일반적인 학습자 특성 외에도, 수능 영어의 맥락적 특성을 함께 고려하는 것이 중요하다. 수능은 제한된 시간 내 복잡한 추론 과정을 요구하는 고난도 문항을 포함하므로, 풀이 시간, 추론 발생 빈도, 지속 시간은 단순한 전략 사용을 넘어 독해 전반의 의미 구성과 정보 통합을 조절하는 핵심 변인이라 할 수 있다.

이러한 논의를 바탕으로, 본 연구는 수능 영어 빈칸 추론 문항에 대한 학습자들의 실제 문제 풀이 양상을 정량적·정성적으로 분석하고자 한다. 이를 위해 정답률 및 풀이 시간, 지문 주제에 대한 친숙도, 영어 능숙도(수능 등급 기준), 추론 발생 빈도와 지속 시간 등을 분석하였다. 또한, 학습자들이 빈칸 추론 문항을 난이도 높게 인식하는 원인을 탐색함으로써, 해당 문항의 평가 도구로서의 한계를 진단하고 개선 방향을 제시하고자 한다.

본 연구에서 상정한 연구 문제는 다음과 같다.

- 1) 학습자들의 수능 영어 빈칸 추론 문항에 대한 정답률과 문제 풀이 소요 시간은 어떠한가?
- 2) 다양한 학습자 변인(주제 친밀도, 수능 영어 등급, 문제 풀이 소요 시간, 추론 발생 빈도, 추론 지속 시간) 중 빈칸 추론 문항의 정답률에 유의미한 영향을 미치는 요인은 무엇인가?
- 3) 학습자들이 빈칸 추론 문항을 난이도가 높은 것으로 인식하는 주요 요인은 무엇인가?

### 3. 연구 방법

#### 3.1 연구 참여자

본 연구에는 총 51 명의 대학생이 참여하였으며, 이들은 모두 1 학년에 재학 중인 학생들이다. 참여자는 인문 계열 학과 소속 29 명, 자연 계열 학과 소속 22 명으로 구성되었으며, 성별 분포는 남 23 명, 여 28 명이다. 연구 참여 시 이들의 나이는 19 세에서 22 세였으며, 평균 20.28 세( $SD = 0.70$ )였다.

영어 학습 경험 조사에 따르면 참여자들은 최소 4 세에서 최대 16 세 사이에 영어를 배우기 시작하였으며, 평균적으로 8.90 세( $SD = 2.64$ )에 영어 학습을 시작하였다. 영어 학습 기간은 최소 4 년에서 최대 18 년까지 다양하며, 평균적으로 11.04 년( $SD = 2.96$ ) 동안 영어를 학습하였다. 이 중 공교육을 통한 영어 교육은 평균 7.02 년( $SD = 1.65$ )을 차지하였고, 사교육(예: 학원, 과외, 온라인 영어 프로그램 등)은 평균 4.02 년( $SD = 2.10$ )으로 참여자들은 비교적 긴 기간 동안 다양한 형태의 영어 학습에 노출되어 있었음을 보여준다.

참여자들의 수능 영어 등급 분포는 1 등급이 없는 대신, 2 등급 15 명, 3 등급 22 명, 4 등급 11 명, 5 등급 1 명, 그리고 2 명이 수능 영어를 응시하지 않은 것으로 나타났다. 영어 능력 자가 평가 결과, 5 점 리커트 척도로 측정된 평균 점수(1 = 매우 못함, 5 = 매우 잘함)는 말하기 1.96( $SD = 0.78$ ), 듣기 2.78( $SD = 0.68$ ), 읽기 2.90( $SD = 0.895$ ), 쓰기 2.16( $SD = 0.89$ )로 나타났다. 이를 바탕으로 참여자들의 영어 능력은 대체로 중급 이하의 수준으로 평가될 수 있으며, 듣기와 읽기 능력이 상대적으로 강한 반면, 말하기와 쓰기 능력은 다소 낮은 경향을 보였다.

#### 3.2 연구 도구 및 자료

##### 3.2.1 빈칸 추론 문항

본 연구에서는 2025 학년도 대학수학능력시험(수능) 영어 영역의 빈칸 추론 기출 문제 총 4 문항(31 번~34 번)을 활용하였다. 또한, 학습자들의 피로도를 완화하고 자연스러운 문제 풀이 흐름을 유도하기 위해 완충 문항(filler) 2 개를 추가로 포함하였다. 완충 문항은 비교적 난이도가 낮은 과년도 기출 문제인 2024 학년도 19 번(심경 변화) 및 2023 학년도 20 번(주제 찾기) 문항을 활용하였다.

문항의 제시 순서는 준무작위화(semi-randomization) 기법을 활용하여 조정되었다. 구체적으로, 빈칸 추론 4 문항 사이사이에 완충 문항을 삽입함으로써 학습자들이 문제 풀이 과정에서 과도한 인지적 부담을 느끼지 않도록 배치하였다.

### 3.2.2 학습자 구두 발화

본 연구에서는 학습자들의 문제 풀이 과정에서 실시간으로 발생하는 인지적 사고 과정을 탐색하고자 구두 발화(think-aloud) 기법을 적용하였다. 구두 발화는 학습자들이 문제를 해결하는 동안 자신의 사고 과정을 명시적으로 언어화 하도록 유도하는 방법으로, 이들을 통해 문제 해결 전략, 추론 과정, 인지적 부담 등을 실증적으로 분석할 수 있다(Ericsson and Simon 1993).

연구에서는 비지시적(non-directive) 구두 발화 기법을 사용하여 학습자들이 문항을 푸는 동안 자연스럽게 자신의 사고 과정을 진술하도록 하였다. 구체적으로, 실험이 시작되기 전, 학습자들에게 구두 발화 절차를 사전 훈련(pre-training)시키고, 문제 풀이 과정에서 최대한 중단 없이 자신의 생각을 소리 내어 말하도록 안내하였다.

구두 발화 데이터는 음성 녹음 장치를 이용하여 수집되었으며, 실험 종료 후 전사(transcription) 작업을 거쳐 (1) 문제 풀이 시간, (2) 추론 발생 여부, (3) 추론에 소요된 시간 분석에 사용되었다.

### 3.2.3 학습자 정보 및 문항 인식 설문지

학습자의 기본 정보와 각 문항에 대한 학습자 인식을 위해 온라인 설문지를 제작하였다. 기본 정보 설문은 학습자의 학번, 성별, 나이, 학년, 전공 등과 영어 학습 경험(영어 수준 자가 평가, 입학 시 수능 영어 등급 등)을 수집하는 문항으로 이루어졌다.

문항 인식 조사는 크게 두 가지 요소로 구성되었다. 첫째, 학습자가 각 문항의 주제에 대해 얼마나 친숙한지를 평가하는 문항을 포함하였다(주제 친밀도, 예: “3 번 문항의 주제는 본인에게 얼마나 익숙한 내용이었나요?”: 1=전혀 그렇지 않다, 5=매우 그렇다). 둘째, 학습자가 문항 풀이 과정에서 겪은 어려움에 대한 질문이었으며(난이도 인식, 예: “3 번 문항을 풀 때, 특히 어려웠던 점이 있었다면 자세히 적어주세요.”), 개방형으로 질문하여 학습자들이 느끼는 어려움을 구체적으로 파악하였다.

## 3.3 연구 절차

연구 참여자의 사전 학습 효과 및 문제에 대한 사전 노출을 최소화하기 위해 본 연구의 데이터 수집은 2025 학년도 대학수학능력시험이 시행된 바로 다음 날인 2024년 11월 15일에 진행되었다. 데이터 수집은 온라인 설문의 학습자 정보 조사를 시작으로, 이후 학습자들의 문제 풀이 과정까지 체계적으로 이루어졌다.

먼저, 연구자는 학생들에게 연구의 목적(학습자들의 문제 풀이 과정을 관찰하는 것)을 설명한 후, 구두 발화법을 안내하는 오리엔테이션 세션을 제공하였다. 이 과정에서 학생들은 연구자 중 한 명이 빈칸 추론 문항 풀이와 구두 발화법을 시연하는 영상을 시청하였으며, 이후 궁금한 점을 연구자에게 질의하는 시간을 가졌다. 그 후 학생들은 각자 본인이 편안하게 느끼는 조용한 독립된 공간에서 구두 발화법을 동반하여 문제를 풀었으며, 발화는 모두 본인의 핸드폰으로 녹음하였다.

문항은 지필고사(pen-and-paper) 형식으로 제공되었으며, 각 문항을 해결한 후 학생들은 자신이 선택한 정답(5지 선다형)을 기록하고, 해당 문항에 대한 친밀도 문항에 응답하였다. 또한, 각 문항을 풀면서 경험한 난이도와 문제 해결 과정에서의 어려움을 개방형 응답 형식으로 기술하였다.

문제 풀이는 한 회차에 걸쳐 연속적으로 진행되었으며, 중간에 휴식을 취하거나 녹음을 중단한 후 재개하는 사례 없이 일괄적으로 수행되었다. 연구자는 학생들이 개별 녹음한 음성 파일을 후속 분석을 위해 저장하였다.

### 3.4 데이터 코딩 및 분석

본 연구에서는 학습자들의 구두 발화 데이터를 정량적 및 정성적으로 분석하기 위해 다양한 코딩 기법과 분석 절차를 적용하였다. 데이터 코딩 및 분석 과정은 다음과 같다.

#### 3.4.1 학습자 구두 발화 데이터 전사 및 전처리

먼저, 학습자들이 녹음한 mp3 파일을 mp4 형식으로 변환한 후, YouTube의 자동 자막 기능을 활용하여 자동 전사된 자막을 생성하였다. 생성된 자막 파일을 docx 형식으로 변환한 후, 원본 녹음 파일과 함께 검토하며 전사 내용을 수기로 보정하였다. 이후, 파이썬(Python)의 pandas, re, pydub 등의 라이브러리를 사용하여 전사된 데이터를 30 초 단위로 병합하고 분석이 용이하도록 변환하였다.

#### 3.4.2 데이터 코딩과 분석

본 연구에서는 완충 문항 2 개를 제외한 빈칸 추론 문항 4 개를 대상으로 데이터 코딩과 분석을 실시하였으며, 이들을 연구 문제별로 정리하면 다음과 같다.

우선, 수능 영어 빈칸 추론 문항에 대한 정답률과 문제 풀이 소요 시간(연구 문제 1)을 측정하기 위해, 각 문항의 정답 여부를 이분법적으로 코딩하였다. 학습자가 문항을 맞힌 경우 1 점, 틀린 경우 0 점을 부여하였으며, 부분 점수는 제공되지 않았다. 문제 풀이 소요 시간은 구두 발화 데이터를 기반으로 측정하였다. 학습자가 특정 문항의 풀이를 시작한 시점과 종료한 시점을 확인한 후, 종료 시간에서 시작 시간을 차감하는 방식으로 계산하였다.

다음으로, 빈칸 추론 문항의 정답률에 영향을 미치는 주요 학습자 변인(연구 문제 2)을 분석하기 위해, 학습자 변인을 크게 수능 영어 등급, 주제 친밀도, 문항 풀이 소요 시간, 추론 발생 횟수, 추론 소요 시간으로 구분하였다. 수능 영어 등급과 주제 친밀도는 학습자들이 설문 문항에 응답한 값을 그대로 반영하여 코딩하였다. 문항 풀이 소요 시간은 앞서 산출한 수치(연구 문제 1)를 이용하였다. 추론 발생 횟수는 학습자들의 단순 해석(literal interpretation; 문장의 단순 직역 혹은 표면적 의미 파악)이나 막연한 추측(guessing; 맥락 고려 없이 직관적 의미 유추)을 배제하고, 문장 내 개념 간 관계를 고려하거나 문맥적 단서를 기반으로 의미를 구성하려는 논리적 추론(inferencing)이 포함된 발화를 중심으로 계수하였다(아래 예시 참고).

<예시 문장>

*Core language teaching materials must concentrate on how a language operates both as a rule-based system and as a sociosemantic system. (2025 학년도 31 번)*

- 단순 해석: 핵심 언어는 언어가 규칙 기반이고, 사회적으로 작동한다.
- 막연한 추측: 언어는 사회적이다? 사람의 대화하는 방식에 대한 건가?
- 추론: 언어 교육이 sociosemantic 이라고 했으니까, 언어 문법 규칙을 넘어서 맥락이 중요할 수 있다는 거지.

코딩의 객관성과 일관성을 확보하기 위해, 두 명의 독립된 연구자가 각각 추론 발화 여부를 코딩한 후, 상호 일치율(inter-rater reliability)을 산출하였다. Cohen's kappa 계수를 활용하여 평가한 결과, 일치도는 .82로 나타나 신뢰도 기준(.80 이상)을 충족하였다. 이후 불일치 사례는 상호 협의 과정을 통해 합의하여 최종 코딩값을 확정하였다.

또한, 추론 소요 시간은 이러한 추론 발화의 총 지속 시간을 측정하여, 학습자가 특정 문항에서 추론을 수행하는 데 걸린 시간을 분석하였다.

마지막으로, 학습자들의 문항 난이도 인식(연구 문제 3)을 평가하기 위해, 개방형 응답에서 학습자들이 언급한 난이도 관련 서술을 질적 분석하였다. 이를 위해 내용 분석(content analysis) 기법을 적용하여 주요 키워드를 추출하고, 응답 내용을 범주화하였다. 이를 통해 학습자들이 빈칸 추론 문항을 어렵다고 인식하는 주요 요인을 도출하고, 난이도 인식과 관련된 공통적인 패턴을 규명하였다.

### 3.5 통계 분석 방법

각 연구 문제에 대한 구체적인 분석 절차는 다음과 같다. 우선, 연구 문제 1(빈칸 추론 정답률과 소요 시간)의 경우, 각 문항별 정답 여부를 채점하고, 이를 바탕으로 평균 정답률을 산출하였다. 또한, 문제 풀이 소요 시간은 학습자의 사고 발화 전사 자료를 활용하여 추출하였으며, 개별 문항에 대한 풀이 시간을 측정하여 평균값을 계산하였다.

둘째, 연구 문제 2(빈칸 추론 정답률에 영향을 미치는 학습자 변인)는 다중 로지스틱 회귀분석(multiple logistic regression analysis)을 이용하였다. 학습자의 문항 정답 여부(정답=1, 오답=0)를 종속 변인으로 설정하였으며, 학습자의 수능 영어 등급(서열 척도), 주제 친밀도(서열 척도), 문제 풀이에 소요한 시간(비율 척도), 추론 발생 횟수(비율 척도), 추론에 소요한 시간(비율 척도)을 예측 변인으로 입력하였다. 이를 통해 해당 변인들이 빈칸 추론 문항의 정답률을 유의미하게 예측하는지 확인하였다.

셋째, 연구 문제 3(빈칸 추론이 어렵다고 느끼는 이유)에 대한 분석은 학습자들의 “이 문제를 풀 때 특히 어려웠던 점은 무엇입니까?”에 대한 응답을 질적 분석하였다. 이 분석은 내용 분석(content analysis) 기법을 활용하여 이루어졌으며, 빈칸 추론 문항의 난이도를 결정하는 주요 요인을 도출하였다.

데이터 분석에는 JASP 통계 프로그램(JASP Team, 2024)을 활용하였다. JASP는 오픈소스 소프트웨어로서 누구나 자유롭게 접근할 수 있으며, 직관적인 인터페이스를 통해 연구자의 분석 효율성과 활용 편의성을 높일 수 있는 특징을 지닌다.

## 4. 결과 및 논의

#### 4.1 빈칸 추론 문항의 정답률과 소요 시간

우선 첫 번째 연구 문제(학습자들의 수능 영어 빈칸 추론 문항에 대한 정답률과 문제 풀이 소요 시간은 어떠한가?)를 위하여 학습자들이 수능 영어 빈칸 추론 문항을 해결하는 데 있어 정답률과 문제 풀이 소요 시간이 어떠한지를 분석하였다(표 1). 먼저, 문항별 정답률을 살펴보면, 31 번과 34 번 문항의 정답률은 각각 54%로 동일했으며, 33 번 문항이 56%로 가장 높았다. 반면, 32 번 문항의 정답률은 28%로 가장 낮았다. 전체 문항의 평균 정답률은 48%로 나타났다.

각 문항의 정답률이 기회 수준(0.5, chance level)과 비교하여 유의미한지 검증하기 위해 일표본  $t$  검정(one-sample  $t$ -test)을 실시하였다. 분석 결과, 31 번( $t(51) = 0.573, p = .569$ ), 33 번( $t(51) = 0.863, p = .392$ ), 34 번( $t(51) = 0.573, p = .569$ ) 문항은 기회 수준과 유의미한 차이가 없었으며, 학습자들이 해당 문항에서 무작위로 정답을 선택한 것과 유사한 수준의 성과를 보인 것으로 나타났다. 반면, 32 번 문항( $t(51) = -3.510, p = .0009$ )은 기회 수준보다 유의미하게 낮았는데, 이는 학습자들이 해당 문항에서 특히 어려움을 겪었음을 시사한다.

다음으로, 문항별 평균 소요 시간을 살펴보면, 32 번 문항이 509.96 초(8 분 29 초)로 가장 길었으며, 33 번 문항이 398.74 초(6 분 38 초)로 가장 짧았다. 31 번과 34 번 문항의 소요 시간은 각각 456.04 초(7 분 45 초), 463.44 초(7 분 43 초)로 유사한 수준을 보였다. 전체 문항의 평균 소요 시간은 459.30 초(7 분 39 초)로 나타났다. 이는 사실 학원에서 제시하는 적정 풀이 시간인 2 분 30 초를 크게 초과하는 결과로, 학습자들이 빈칸 추론 문항을 해결하는 데 예상보다 훨씬 많은 시간이 필요했음을 의미한다.

표 1. 빈칸 추론 문항별 정답률(%)과 소요시간(초)

문항 번호	정답률		총 소요 시간	
	평균( $M$ )	표준편차( $SD$ )	평균( $M$ )	표준편차( $SD$ )
31 번	54	0.50	465.04	140.04
32 번	28	0.45	509.96	176.79
33 번	56	0.50	398.74	148.28
34 번	54	0.50	463.44	154.30
전체	48	0.50	459.30	159.31

이러한 결과는 단순히 문항의 난이도 때문만이 아니라, 본 연구에서 사용된 생각 말하기(think-aloud) 기법의 특성에도 밀접한 관련이 있을 수 있다. 문제를 해결하는 동시에 사고 과정을 언어적으로 표현하는 활동은 학습자들에게 상당한 인지적 부담을 유발할 수 있다(Ericsson and Simon 1993). 특히, 이러한 방식에 익숙하지 않은 학습자들에게는 문제 해결 속도가 자연스럽게 지연될 가능성이 있으며, 이는 문항 난이도뿐 아니라 사고의 언어화라는 추가적인 과제가 복합적으로 작용한 결과일 수 있다.

그럼에도 불구하고, 학습자들이 빈칸 추론 문항 해결에 과도한 시간을 소요했다는 사실은, 수능 영어 빈칸 추론 문항 자체가 높은 수준의 인지적 부담을 요구하고 있음을 시사한다. 전반적으로 낮은 정답률과 장시간에 걸친 문제 해결 시간은, 해당 문항들이 학습자들에게 복잡한 추론 과정, 지문 내 정보의 복합적 처리, 그리고 배경지식의 활용을 동시에 요구하고 있음을 시사한다. 요구되는 추론

과정의 복잡성, 지문 내 정보 처리의 난이도, 그리고 배경지식 활용의 어려움 등이 주요 요인으로 작용하고 있음을 알 수 있다. 특히 수능 영어 시험이 제한된 시간 내에 모든 문항을 해결해야 하는 시험 환경임을 고려할 때, 빈칸 추론 문항에서의 과도한 시간 소요는 학습자들이 지문을 충분히 숙독하지 못하게 하며, 고차원적 사고(higher-order thinking)에 기반한 추론적 사고에 도달하는 데 방해 요소로 작용할 가능성이 크다(Just and Carpenter 1980, Rayner 1998). 결과적으로, 학습자들은 제한된 시간 안에 문제를 해결해야 한다는 압박으로 인해, 깊이 있는 이해와 추론보다는 피상적인 단서 활용이나 소거법에 의존할 가능성이 높아진다.

#### 4.2 빈칸 추론 문항의 정답률에 대한 예측 변인

두 번째 연구 문제(빈칸 추론 문항의 정답률에 영향을 미치는 주요 학습자 변인은 무엇인가?)의 분석에 앞서 학습자 변인인 수능 영어 등급, 주제 친밀도, 문항 풀이 소요 시간, 추론 발생 횟수 및 추론에 소요한 시간을 바탕으로 기술 통계를 실시하였다. 먼저, 수능 영어 등급 분포를 살펴보면, 본 연구에 참여한 학습자들의 수능 영어 등급은 중상위권에 해당하는 2 등급과 3 등급에 전체 학생의 약 75%가 분포되어 있는 상태였다. 주제 친밀도의 경우, 33 번 문항이 가장 높았으며( $M = 3.56, SD = 1.20$ ), 32 번 문항( $M = 2.54, SD = 1.03$ )이 가장 낮았다. 전체 평균은  $2.95(SD = 1.12)$ 로 학습자들이 문항별로 배경지식의 차이를 경험했을 가능성을 시사한다.

추론 발생 횟수는 문항별로 큰 차이를 보였다. 학습자들은 32 번 문항( $M = 2.00, SD = 1.58$ )에서 가장 많은 추론을 수행한 반면, 33 번 문항( $M = 1.40, SD = 1.07$ )에서 가장 적은 추론을 시도하였다. 전체 평균 추론 발생 횟수는  $1.62(SD = 1.32)$ 였다. 주목할 점은, 빈칸 추론 문항임에도 불구하고 “추론 발생 횟수”의 표준편차가 상대적으로 크다는 점이다. 이는 일부 학습자가 추론을 거의 수행하지 않았을 가능성을 시사하며, 결과적으로 특정 문항에서 추론 과정이 학습자의 문제 해결 과정에서 핵심적인 역할을 하지 않았을 가능성을 제기한다.

추론 소요 시간의 경우, 32 번 문항( $M = 31.57$  초,  $SD = 32.42$ )이 가장 긴 추론 시간이 소요된 반면, 34 번 문항( $M = 21.58$  초,  $SD = 29.19$ )이 가장 짧았다. 전체 평균 추론 소요 시간은  $27.65$  초( $SD = 29.62$ )였다. 문제 풀이에 소요된 총 시간이 평균  $459.30$  초(약 7 분 30 초)임을 고려하였을 때, 추론 과정에 소요된 평균 시간( $27.65$  초)은 전체 풀이 시간 대비 약 6.02%에 불과하였다. 이러한 결과는 학습자들이 문항 해결 과정에서 추론적 사고를 충분히 활용하지 않았음을 시사한다.

표 2. 학습자 변인 관련 기술 통계 결과

문항 번호(번)	31	32	33	34	전체
영어 영어 등급(명)		2 등급 = 15, 3 등급 = 22, 4 등급 = 11, 5 등급 = 1			
주제 친밀도	3 (1.01)	2.54 (1.03)	3.56 (1.20)	2.70 (0.97)	2.95 (1.12)
총 소요 시간(초)	465.04 (140.04)	509.96 (176.79)	398.74 (148.28)	463.44 (154.30)	459.30 (159.31)
추론 발생 횟수(회)	1.64 (1.21)	2.0 (1.58)	1.40 (1.07)	1.42 (0.5)	1.62 (1.32)
추론 소요 시간(초)	29.96 (27.31)	31.57 (32.42)	27.56 (32.83)	21.58 (29.19)	27.65 (29.62)

Note: 괄호 안 수치는 표준편차

다음으로 빈칸 추론 문항의 정답률을 예측하는 주요 변인을 확인하기 위해 다중 로지스틱 회귀 분석을 수행하였다. 회귀 분석 결과, 주제 친밀도가 문항 정답률을 예측하는 데 유의미한 영향을 미치는 것으로 나타났다( $\beta = 0.42$ ,  $SE = 0.139$ ,  $Z = 3.023$ ,  $p = .002$ ). 이는 주제 친밀도가 1 점 증가할 때, 문항을 맞출 승산이 1.52 배(95% CI [1.15, 1.69]) 증가함을 의미한다. 즉, 학습자가 문항의 주제에 대해 더 친숙할수록 해당 문항의 정답을 맞출 가능성이 높아지는 것으로 해석할 수 있다.

반면, 수능 영어 등급( $\beta = 0.086$ ,  $p = .53$ ), 총 소요 시간( $\beta = 0.00$ ,  $p = .68$ ), 추론 발생 횟수( $\beta = -0.114$ ,  $p = .43$ ), 추론 소요 시간( $\beta = 0.008$ ,  $p = .28$ )\*\*은 문항 정답률과 유의미한 관계를 보이지 않았다. 특히, 소요 시간과 추론 관련 변수들이 유의한 예측 변인이 아니라는 점은, 학습자가 문제 풀이에 오랜 시간을 투자하거나 더 많은 추론을 수행하는 것이 반드시 정답률을 높이는 요인으로 작용하지 않음을 시사한다.

표 3. 문항 정답률에 대한 다중 로지스틱 회귀 분석 결과

예측 변인	회귀 계수	표준 오차 계수	Z	p	승산비	95% 신뢰구간	
						하한	상한
수능 영어 등급	0.09	0.136	0.634	0.53	1.09	-0.18	0.35
주제 친밀도	0.42**	0.139	3.023	0.002	1.52	0.15	0.69
총 소요 시간	0.00	0.001	0.41	0.68	1.00	0.00	0.00
추론 발생 횟수	-0.11	0.145	-0.787	0.43	0.89	-0.4	0.17
추론 소요 시간	0.01	0.007	1.088	0.28	1.01	-0.01	0.02

Note: 괄호 안 수치는 표준편차; \*\*  $p < .001$

이러한 결과는 빈칸 추론 문항이 평가하는 능력이 단순한 영어 독해력이 아니라, 학습자의 배경지식과 밀접하게 관련될 수 있음을 의미한다(Alptekin and Erçetin 2011, Cromley and Azevedo 2007). 영어 평가의 기본적인 목적은 학습자의 영어 능력, 즉 어휘 및 문법 지식, 독해력, 추론적 사고력을 평가하는 것이다. 그러나 본 연구 결과에 따르면, 특정 문항에서는 학습자가 영어 능력과 무관하게 배경지식만으로 정답을 도출할 가능성이 존재한다. 이에 따라 동일한 영어 독해 능력을 갖춘 학습자라 할지라도 특정 주제에 대한 친숙도에 따라 정답률이 달라질 수 있으며, 이러한 편향은 시험의 타당성(validity)에 부정적인 영향을 미칠 수 있다.

종합적으로, 매우 낮은 추론 발생 빈도와 제한적인 추론 지속 시간, 그리고 주제 친밀도가 정답률에 미치는 유의미한 영향은 특정 문항에서 학습자들이 실제로 추론 과정을 거치지 않았음에도 불구하고 정답을 도출할 가능성이 있음을 시사한다. 이는 본 문항 평가의 핵심 요소로 간주되는 영어 추론 과정이 학습자의 실제 문제 해결 과정에서 충분히 반영되지 못했을 가능성을 함축한다. 만약 특정 문항에서 학습자가 문맥적 단서에 기반한 추론보다는 배경지식이나 직관적 판단을 통해 정답을 선택하는 경향이 강하다면, 이는 해당 문항이 영어 독해 및 추론 능력을 온전히 평가하는 도구로서의 타당성에 의문을 제기할 수 있는 요인이 될 수 있다(Ahmadjavaheri and Zeraatpishe 2020, Awabdy 2012). 따라서 문항이 학습자의 영어 추론 능력을 측정하는 도구로서 적절히 기능하는지에 대한 더욱 면밀한 검토가 요구된다.

### 4.3 빈칸 추론 문항에 대한 학습자들의 인식

본 연구는 학습자들이 빈칸 추론 문항을 난이도가 높은 문항으로 인식하는 요인을 분석하기 위해 내용 분석을 실시하였다. 총 200 개의 학습자 응답을 분석한 결과, 학습자들이 어려움을 느끼는 주요 요인은 (1) 어휘 난이도(119 건), (2) 문장 복잡성(76 건), (3) 선지 변별 어려움(32 건)으로 분류되었다(그림 1). (총 227 개)

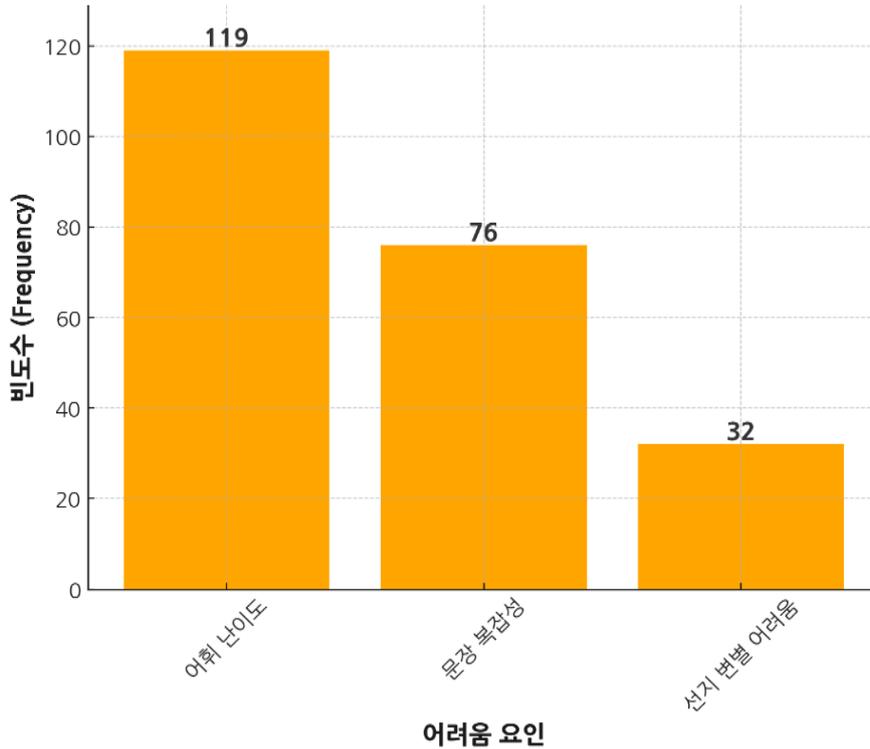


그림 1. 빈칸 추론 문항에서 학습자가 인식한 어려움 내용 분석 결과

첫째, 학습자들은 어려운 어휘가 지문 이해를 가장 크게 방해한다고 응답하였다. prevent, constitute, facilitate, occupy 등의 어휘를 해석하지 못해 문장 이해가 제한되었으며, trustee, impose, negotiable 과 같은 핵심 개념을 이해하지 못해 문맥을 파악하는 데 어려움을 겪었다. 즉, 복잡한 어휘의 등장으로 인해 문장의 의미 파악을 제대로 하지 못하였고, 이는 추론 활동의 좌절로 이어졌다.

<관련 응답 예시>

학습자 1: “concentrate, eager, transparent 의 뜻을 모르겠으며 문장 안에 ‘engaging imaginatively with literature enables learners’라는 부분이 있었는데 단어의 뜻은 모두 알지만 해석하는데 어려움이 있었다. 또 답을 고를 때 perspective 의 뜻을 몰라 어려움이 있었으며 전체적으로 지문이 이해되지 않아 답을 고르기 어려웠다.”

학습자 4: “단어가 어려웠다.”

학습자 12: “모르는 단어가 많아 해석하기도 어려움이 있었고, 본문 내용도 이해하기 어려웠다.”

이와 같은 어휘 난이도에 대한 인식은 빈칸 추론 문항이 추상적이거나 전문적인 개념을 포함할 때 더욱 두드러졌으며, 배경지식 없이 새로운 개념을 빠르게 이해해야 하는 부담이까지 가중되어 지문에 대한 이해를 더욱 어렵게 하였다. 결과적으로, 빈칸 추론 문항이 단순한 독해력을 측정하는 것이 아니라 특정 전문 분야의 이해도를 평가하는 방식으로 작용할 위험이 있으며, 이에 대한 공정성과 타당성에 대한 논의가 필요함을 시사한다.

둘째, 학습자들은 문장 복잡성 또한 주요한 어려움으로 작용한다고 보고하였다. 일부 학습자들은 지나치게 긴 문장이 의미를 명확히 정리하는 것을 방해하며, 문장 간 연결 관계를 분석하는 과정에서 혼란을 느꼈다.

#### <관련 응답 예시>

학습자 13: “전체적으로 뭐라고 하는 건지 모르겠다. 문장구조도 어렵고 단어도 어려웠다.”

학습자 29: “문장을 의미하는 바가 너무 꼬여 있어서 이해하는 데 어려움이 있었다.”

학습자 43: “해석이 너무 어려웠다. 문장이 길어서 끊어서 읽어도 해석이 안되었다.”

학습자 44: “내용 해석을 잘 못해서 무슨 내용인지 잘 몰라서 해맸다.”

빈칸 추론 문항은 단순한 문장 해석을 넘어 문맥 속에서 의미를 추론하는 고차원적 사고를 요구하지만, 어휘 난이도와 문장 복잡성이 결합될 경우 학습자의 해석 과정을 차단하는 요인으로 작용할 수 있음을 시사한다(Zhang and Zhang 2022).

또한, 학습자들은 선지 간 유사성이 높은 경우 정답을 결정하는 과정이 어려웠다고 응답하였다. 연구 결과, 학습자들이 문단의 내용을 전반적으로 이해했음에도 불구하고, 선지의 의미를 명확히 파악하지 못해 문항을 어렵게 평가한 사례가 많았다.

#### <관련 응답 예시>

학습자 3: “1 번이랑 5 번이 답이 좀 헷갈렸다. 선지가 비슷해서.”

학습자 26: “내용은 대충 알겠는데 선지가 헷갈렸다.”

학습자 34: “글의 내용은 잘 파악된 것 같으나 비슷해 보이는 선지가 두 개 있어 둘 중 무엇인지 확실히 결정하지 못했다.”

학습자 47: “지문을 이해하고 선지를 넣는 과정에서 선지가 내가 이해한 내용과 차이가 있었다.”

학습자 49: “(지문은 이해했는데,) 보기에서 모르는 단어가 나오니 어떻게 해석해야 할지 모르겠다.”

네 문항 모두에서 공통적으로 어휘 난이도, 문장 복잡성, 선지 변별 어려움이 주요 난이도 요인으로 언급되었으며, 이러한 경향은 특히 32 번 문항에서 두드러지게 나타났다. 학습자들은 해당 문항에서 ‘being stimulus-driven’과 같이 구체성이 결여된 표현과 생소한 어휘로 인해, 빈칸이 포함된 문장을 해석하는 데 상당한 어려움을 겪었다고 보고하였다. 또한, 정답 선지에 제시된 ‘frees a person from the burden’이라는 표현의 함의를 명확히 이해하지 못한 사례가 다수 확인되었는데, 이는 학습자들이 텍스트 전반의 의미 맥락을 파악하는 데 있어 한계를 느꼈음을 시사한다.

이는 32번 문항의 낮은 정답률이 단지 학습자의 언어 능력 부족 때문이 아니라, 해당 문항이 지닌 과도한 추상성, 생소한 어휘 사용, 본문 및 선지 간 의미 연계의 모호성 등이 복합적으로 작용한 결과임을 보여준다. 특히, 32번 문항과 34번 문항이 주제의 친숙도 면에서는 유사함에도 불구하고, 32번 문항의 정답률이 현저히 낮게 나타난 점은 위와 같은 요인들이 정답 추론 과정에 실질적인 방해 요소로 작용했음을 뒷받침한다. 실제로 32번 문항은 본문과 정답 선지 모두에서 고도의 추상적 표현이 사용되어, 학습자들이 정답 도출에 활용할 수 있는 단서가 충분히 제시되지 않았고, 이로 인해 정답을 유도할 수 있는 인지적 경로 자체가 제한되었을 가능성이 높다. 이와 같은 구조적 제약은 정답 선택률이 28%에 불과한 반면, 특정 오답 선택지에 40%에 달하는 응답이 집중된 양상에서도 확인된다.

요약하면, 학습자들은 지문의 핵심 내용을 어느 정도 파악하였음에도 불구하고, 선지의 의미를 명확히 해석하지 못해 정답 결정에 어려움을 겪었다(Ascalon et al. 2007, Gierl et al. 2017). 특히 일부 응답자들은 선지의 의미가 과도하게 추상적이거나, 선지 간 의미 차이가 지나치게 미세하여 정답을 구분하기 어렵다고 응답하였다. 이는 단순히 학습자의 독해력이나 추론 능력의 결함 때문이 아니라, 문항의 구성 자체가 학습자에게 정답을 파악할 수 있는 기회를 구조적으로 제한하고 있었기 때문일 가능성이 높다. 결국, 빈칸 추론 문항은 학습자의 영어 능력이나 논리적 사고력을 타당하게 측정하기보다는, 언어적 함정을 회피하는 능력을 평가하는 방향으로 기능할 위험이 있음을 시사한다.

## 5. 결론 및 교육적 함의

본 연구는 수능 영어 빈칸 추론 문항이 학습자의 추론 능력을 타당하게 평가하는지 여부를 파악하기 위해 학습자들의 문항 풀이와 관련된 양상(정답률, 소요시간)을 살펴보고, 정답률에 영향을 미치는 주요 학습자 변인(주제 친밀도, 수능 영어 등급, 문항 풀이 소요 시간, 추론 발생 빈도, 추론 지속 시간)을 탐색하였다. 또한 개방형 질문을 통해 학습자들이 각 문항을 어렵게 느끼는 이유에 대해서도 조사하였다.

연구 결과, 수능 영어 빈칸 추론 문항의 정답률이 전반적으로 낮고 문항 풀이 시간이 시험 시간 중 할당된 시간보다 매우 길었으며, 이는 해당 문항들이 학습자들에게 높은 인지적 부담을 초래할 가능성을 시사한다. 따라서 제한된 시간 내 해결해야 하는 시험 환경에서 학습자들은 충분한 추론 과정을 거치기 어려우며, 문항 풀이에 피상적인 단서 활용이나 소거법에 의존할 위험이 높아질 수 있다.

문항 정답률에 영향을 미치는 요인으로는 주제 친밀도만이 유의미한 예측 변수로 나타났다. 즉, 학습자가 해당 문항의 주제와 관련된 배경지식을 보유한 경우 정답 선택 가능성이 유의하게 높았다. 반면, 수능 영어 등급, 문항 풀이 소요 시간, 추론 발생 빈도, 추론 지속 시간은 정답률에 통계적으로 유의미한 영향을 미치지 않았다. 학습자의 구두 발화 분석 결과, 일부 사례( $n = 4$ )에서는 문맥적 단서를 활용한 추론적 사고가 정답 도출에 기여한 정황이 관찰되었으나, 이는 전체 참여자 규모에 비해 극히 제한된 수치였다. 이러한 결과는 수능 영어의 빈칸 추론 문항이 학습자의 영어 능력이나 추론 전략보다는 상대적으로 사전 지식에 크게 영향을 받는 평가 도구로 기능할 가능성을 시사한다.

또한, 학습자들은 어휘 난이도, 문장 복잡성, 유사한 선지 간 변별 어려움을 빈칸 추론 문항의 주요 난이도 요인으로 인식하였다. 이는 학습자들이 제한된 시간 내에 문맥을 기반으로 논리적 추론을 수행하기보다는, 복잡한 어휘와 문장을 해석하고 유사한 선택지 간 미묘한 의미 차이를 구별하는 데

초점을 두게 만들 가능성이 있음을 의미한다. 즉, 빈칸 추론 문항이 학습자의 독해 및 추론 능력을 평가하기보다는 배경지식에 의존하도록 하거나, 선지 간의 미세한 차이를 구별하는 능력을 측정하는 방식으로 출제되었을 가능성이 높다.

위와 같은 결과는 수능 영어 빈칸 추론 문항이 학습자의 실제 추론 능력을 평가하는 도구로서 타당함에 대한 의문을 제기한다. 특히, 일부 문항에서 지문의 난이도가 높아 학습자들이 추론 과정 자체를 수행하지 못하는 경우도 확인되었다. 이는 빈칸 추론 문항이 본래의 목적을 온전히 달성하지 못할 가능성이 있음을 시사하며, 평가 문항을 개발하는 과정에서 보다 근본적인 개선이 필요함을 강조한다.

첫째, 문항의 난이도를 조정하여 학습자가 실제로 추론적 사고를 수행할 수 있는 환경을 조성해야 한다. 이를 위해 (1) 핵심 정보가 포함된 문장을 지나치게 길거나 복잡하게 구성하지 않고, (2) 추론의 근거가 되는 단서는 보다 명확하게 제시하며, (3) 정답과 오답 선택지 간 의미상의 차이를 분명히 하여, 정답 추론이 언어 민감성이 아닌 논리적 판단에 기반할 수 있도록 문항을 설계할 필요가 있다.

둘째, 평가 문항 개발자들은 빈칸 추론 문항의 본래 목적이 추론 능력 평가에 있다는 점을 재확인하고, 배경지식이나 단순 해석에 의존하지 않고도 학습자가 문맥적 단서와 논리 구조를 통해 정답을 도출할 수 있도록 문항을 구성해야 한다. 지나치게 높은 난이도는 오히려 추론 과정 자체를 방해할 수 있다는 점을 고려해야 한다.

셋째, 수능 영어의 평가 방향에 대한 정책적 재검토가 필요하다. 수험생이 제한된 시간 안에 의미를 구성하고 논리적으로 사고할 수 있는 문항이 출제되어야 하며, 이는 평가가 단순 등급 변별보다는 학습자의 사고력과 추론 능력 향상을 유도하는 교육적 기능을 수행할 수 있도록 평가 체계를 재구성해야 함을 시사한다.

본 연구는 수능 영어 빈칸 추론 문항의 평가 방식에 대한 데이터 기반 비평을 통해, 현행 평가 체계의 문제점을 환기하고 개선의 필요성을 제기하는 것을 목표로 하였다. 연구 결과를 바탕으로, 빈칸 추론 문항이 본래의 평가 목적을 충실히 수행하기 위해서는 문항의 난이도를 조정하고, 학습자의 실제 추론 능력을 평가할 수 있도록 평가 체계를 재구성할 필요가 있음을 강조한다. 이는 단순한 평가 방식의 조정에 그치는 것이 아니라, 평가가 학습자의 사고력과 논리적 추론 능력 향상을 촉진하는 방향으로 나아가기 위한 필수적인 과정이며, 향후 평가 문항 설계에 있어 중요한 시사점을 제공할 것이다.

## 참고 문헌

- 김희주(Kim, H.). 2018. 학습장애 위험군 학생의 추론적 이해 능력 분석(*A Study on the Inference Skills in Reading of Students At-Risk for Learning Disabilities*). 박사학위논문(Doctoral Dissertation). 서울대학교(Seoul National University).
- 매일경제. 2022.09.16. [대입 완전정복] 수능영어 1 등급 변수... '빈칸 추론 유형' 대응전략 Available online at <https://www.mk.co.kr/news/economy/10456630>
- 베리타스알파. 2023.11.21. [2024 수능] 이의신청 '종료' 최종 288 건.. 단일문항 영어 33 번 18 건 '최다' Available online at <https://www.veritas-a.com/news/articleView.html?idxno=482337>

- 한국교육과정평가원. 2024.11.14. 2025 학년도 대학수학능력시험 출제방향 보도자료 Available online at <https://kice.re.kr/boardCnts/fileDown.do?fileSeq=64e9311cbbb7654d7a4c6b486104a0eb>
- Ahmadjavaheri, Z. and M. Zeraatpishe. 2020. The impact of construct-irrelevant factors on the validity of reading comprehension tests. *International Journal of Language Testing* 10(1), 1-10.
- Alptekin, C. and G. Erçetin. 2011. Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL quarterly* 45(2), 235-266.
- Ascalon, M. E., L. S. Meyers, B. W. Davis and N. Smits. 2007. Distractor similarity and item-stem structure: Effects on item difficulty. *Applied Measurement in Education* 20(2), 153-170.
- Awabdy, G. W. 2012. *Background knowledge and its effect on standardized reading comprehension test performance*. Doctoral dissertation, University of California.
- Bloom, B. S. 1956. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. David McKay Co, Inc.
- Chikalanga, I. 1992. A suggested taxonomy of inferences for the reading teacher. *Reading in a Foreign Language* 8(2), 697-709.
- Cromley, J. G. and R. Azevedo. 2007. Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology* 99(2), 311-325.
- Ericsson K. A. and H. A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data*. The MIT Press.
- Gernsbacher, M. A., K. R. Varner and M. E. Faust. 1990. Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(3), 430-445.
- Gierl, M. J., O. Bulut, Q. Guo and X. Zhang. 2017. Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of educational research* 87(6), 1082-1116.
- Graesser, A. C., M. Singer and T. Trabasso. 1994. Constructing inferences during narrative text comprehension. *Psychological Review* 101(3), 371-395.
- Hall, C. S. 2016. Inference instruction for struggling readers: A synthesis of intervention research. *Educational Psychology Review* 28(1), 1-22.
- Hammadou, J. 1991. Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *The Modern Language Journal* 75(1), 27-38.
- JASP Team. 2024. JASP (Version 0.19.3)[Computer software]. Available online at <https://jasp-stats.org/>
- Johnston, A. M., M. A. Barnes and A. Desrochers. 2008. Reading comprehension: Developmental processes, individual differences, and interventions. *Canadian Psychology* 49(2), 125-132.

- Just, M. A. and P. A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4), 329–354.
- Kendeou, P., D. N. Rapp and P. van den Broek. 2003. The influence of reader's prior knowledge on text comprehension and learning from text. In R. Nata, ed., *Progress in Education*, 189–209. Nova Science Publishers.
- Kendeou, P., C. M. Bohn-Gettler, M. J. White and P. van den Broek. 2008. Children's inference generation across different media. *Journal of Research in Reading* 31(3), 259–272.
- Kim, J. and B. Lee. 2020. Effects of prior topic knowledge and English language proficiency on EFL reading comprehension of two types of expository text. *Korean Journal of Applied Linguistics* 36(3), 31–53.
- Kintsch, W. 1998. *Comprehension: A paradigm for cognition*. Cambridge University Press.
- McKoon, G. and R. Ratcliff. 1992. Inference during reading. *Psychological Review* 99(3), 440–466.
- Perfetti, C. A., N. Landi and J. Oakhill. 2005. The acquisition of reading comprehension skill. In M.J. Snowling and C. Hulme, eds., *The Science of Reading: A Handbook*, 227–247. Blackwell.
- Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3), 372–422.
- Rumelhart, D. E. 1980. Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce and W. F. Brewer, eds., *Theoretical Issues in Reading Comprehension*, 33–58. Lawrence Erlbaum.
- van den Broek, P. 1994. Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher, ed., *Handbook of Psycholinguistics*, 539–588. Academic Press.
- van den Broek, P., R. F. Lorch, T. Linderholm and M. Gustafson. 2001. The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition* 29(8), 1081–1087.
- van den Broek, P., D. N. Rapp and P. Kendeou. 2005. Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes* 39(2&3), 299–316.
- van Dijk, T. A. and W. Kintsch. 1983. *Strategies of discourse comprehension*. Academic Press.
- Zhang, S. and X. Zhang. 2022. The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research* 26(4), 696–725.

Examples in: English

Applicable Languages: English

Applicable Level: Secondary