

한국사능력검정시험 문항 풀이 과정 분석을 통한 역사 선다형 문항 개선 방향 탐색¹⁾

이 미 미(홍익대학교)*

홍 선 이(국사편찬위원회)**

<요 약>

본 연구에서는 대학생 연구 참여자 10명이 한국사능력검정시험 12개 문항을 풀이한 120건의 문항 풀이 과정 사례를 분석하여 연구 참여자가 실제 어떤 지식과 사고력을 활용하여 문항을 풀이하는지, 그리고 연구 참여자가 사용한 지식과 사고력이 출제자가 측정하고자 한 지식 및 사고력과 일치하는지를 분석하였다. 분석 결과 연구 참여자는 지식 소환(59건, 49%), 수험전략 활용(34건, 28%), 추론(27건, 23%)이라는 세 가지 문항 풀이 과정을 활용하여 문항의 답을 찾고 있었다. 수험전략 활용과 추론은 한국사능력검정시험 목표 준거에 명시되지 않은 구인이었으나, 문항 풀이에서는 광범위하게 사용되고 있음을 확인할 수 있었다. 반면 목표 준거에 명시된 역사적 사고력을 활용하여 문항을 풀이한 사례는 찾아볼 수 없었다. 세 가지 문항 풀이 과정 중 지식 소환이 가장 많이 활용되었다는 측면에서 한국사능력검정시험 선다형 문항이 연구 참여자의 지식을 측정하고 있음을 파악할 수 있었으나, 연구 참여자가 실제로 문항 풀이에 활용한 지식의 종류와 수준을 분석한 결과 문항의 타당도를 높일 필요성을 확인할 수 있었다. 문항별·연구 참여자별 분석 결과도 타당도를 높일 필요성을 뒷받침했다. 분석 결과에 기초하여 한국사능력검정시험 문항 개선을 위한 3가지 방향을 제언하였다.

주제어 : 한국사능력검정시험, 선다형 평가, 문항 풀이 과정, 역사 지식, 역사적 사고

1) 이 논문은 2018년 국사편찬위원회 한국사연구지원사업 연구용역(3차)의 지원을 받아 수행된 연구의 자료를 수정·보완한 것임.

* 제1저자. mimilee@hongik.ac.kr

** 교신저자. sun2hong@korea.kr

I. 서론

선다형 문항은 학생의 역사 이해 평가에 널리 활용되고 있다. 역사 관련 대표적인 국가 수준 평가인 대학수학능력시험이나 한국사능력검정시험 모두 100% 선다형 문항을 활용하여 피험자의 역사 이해를 평가하고 있다. 국가 수준 평가뿐 아니라, 일선 중·고등학교의 중간고사나 기말고사에서도 선다형 평가는 빈번하게 사용된다. 이처럼 선다형 평가가 널리 활용되는 이유는 상대적으로 짧은 시간에 많은 양의 지식을 측정할 수 있다는 측면에서 효율성이 높고, 채점자에 따라 결과가 달라지지 않는다는 측면에서 객관성이 높기 때문일 것이다(최상훈 외, 2012; Haladyna, 2004).

선다형 평가의 장점으로 지식의 효율적·객관적인 측정이라는 측면이 강조되지만, 선다형 평가가 지식만 측정하는 것은 아니다. 흔히 역사적 사고력이라 일컬어지는 기능(skill)의 측정 역시 목표로 삼기 때문이다. 역사 지식이 역사적 용어, 사실 및 개념에 대한 이해라면, 역사적 사고력은 역사 자료의 분석과 해석에 기초하여 시대적 상황이나 사건에 대한 의미를 파악하고 평가하는 능력, 연속성과 변화 속에서 인과관계를 탐구하는 능력 등을 의미한다. 대학수학능력시험 한국사의 경우, 내용영역과 행동영역이라는 두 가지 영역의 평가목표의 설정을 통해 역사 지식과 역사적 사고력을 평가하고 있다. 내용영역은 ‘I.우리 역사의 형성과 고대 국가의 발전’처럼 고등학교 한국사 교육과정상의 대단원에 해당하는 6개 시대에 포함된 역사 내용 지식이고, 행동영역은 역사 지식의 이해, 연대기적 사고, 역사 상황 및 쟁점의 인식, 역사적 탐구의 설계 및 수행, 역사 자료의 분석 및 해석, 역사적 상상 및 판단으로 역사적 사고력을 포함한다(박진동 외, 2014; 오정현, 박진동, 2015; 최상훈, 2015). 또 다른 대표적인 국가 수준의 역사 평가인 한국사능력검정시험도 역사 지식뿐 아니라 역사적 사고력을 포함하는 6가지 역사교육 목표 준거에 기초하여 문항을 출제, 평가하고 있다(국사편찬위원회, n.d.; 김미선, 2011). 6가지 목표 준거는 역사 지식의 이해, 연대기의 파악, 역사 상황 및 쟁점의 인식, 역사 자료의 분석 및 해석, 역사 탐구의 설계 및 수행, 결론의 도출 및 평가이다(국사편찬위원회, n.d.).

현행 역사 선다형 문항은 이처럼 역사 지식뿐 아니라 역사적 사고력 측정을 목표로 하고 있지만, 선다형 문항을 통해 역사 지식과 역사적 사고력을 어느 정도 평가할 수 있는지에 대해 명확하게 알려진 바는 많지 않다(Reich, 2013; Smith, 2017). 선다형 문항의 정·오답 결과를 통해, 우리는 학생의 역사 이해도에 대해 어떤 주장을 할 수 있는가? 선다형 문항은 출제자가 의도한 지식과 사고력을 피험자로부터 어느 정도 이끌어내는가? 변별을 넘어, 평가 점수가 가지

는 의미를 해석하고 이에 기초하여 역사 교수·학습 향상을 이루어 나가기 위해서는 이와 같은 질문에 대한 답이 필수적임에도 불구하고, 현재 관련 연구는 찾아보기 힘든 실정이다.

본 연구에서 제기하는 이와 같은 문제는 평가의 타당도와 관련이 있다. 최근 타당도는 경험적 증거와 이론이 평가 점수의 해석을 지지하는 정도에 대한 종합적인 판단이라는 관점에서 조망되고 있다(American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, 2014; Braden, 2009). 이런 관점에서의 타당도는 특정 통계를 통해 명시적으로 확인하거나 직관적으로 파악하기 어려운 대상이다. 다양한 경험적 증거와 이론적 근거의 수집과 분석을 통해 판단이 이루어져야 하기 때문이다(Braden, 2009; Messick, 1995). 본 연구에서는 이와 같은 종합적인 관점에서의 타당도를 확인하는데 필요한 여러 측면 중 하나인 피험자의 반응 과정(response processes)에 주목하였다(American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, 2014; Messick, 1995). 피험자가 문항을 풀이하는 과정에서 활용한 지식과 사고력이 출제 의도에 부합하는지, 이런 측면에서 문항이 타당한지를 확인하기 위한 과정이다.

본 연구에서는 대학생 연구 참여자 10명이 한국사능력검정시험 12개 문항을 풀이한 120건의 풀이 과정 사례를 분석하여 연구 참여자가 어떤 지식과 사고력을 활용하는지, 그리고 연구 참여자가 사용한 지식과 사고력이 출제자가 측정하고자 한 지식 및 사고력과 일치하는지, 곧 선다형 문항이 실제로 무엇을, 어떻게 평가하고 있는지를 분석하였다.

이를 토대로 논의 및 제언에서 한국사능력검정시험의 타당도 제고를 위해 고려할 방안을 제안하였다. 한국사능력검정시험은 초·중등 교원임용시험 응시자격 부여, 일부 공기업 및 민간기업의 사원 채용이나 승진 시 반영, 공무원 경력경쟁채용시험 가산점 부여 등 다양한 방면에서 널리 활용되고 있다(국사편찬위원회, n.d.). 나아가 2021년부터는 7급 국가공무원 공개경쟁채용시험에서 한국사 과목을 대체할 예정이다(인사혁신처, 2018). 이처럼 활용도와 응시자의 수가 향후 더욱 증가할 것이라 전망되는 상황에서, 한국사능력검정시험이 명실상부 대표적인 역사과 선다형 평가로서 위상을 확립하기 위해서는, 시험의 평가 방식 및 타당도를 다각도로 검토하고 이를 토대로 문항 출제 및 평가 방식을 개선할 필요가 있다. 본 연구가 이러한 과정에 조금이나마 보탬이 되기를 기대한다.

II. 학생들은 역사 선다형 문항을 어떻게 풀이하는가?:

선행 연구 검토

역사과 선다형 문항 풀이에 관한 선행 연구는 주로 고등학생의 자국사 선다형 문항의 풀이 과정 사례를 중심으로 진행되었으며, 선행 연구는 분석 결과에 기초하여 선다형 문항 개선 방안을 제언한 바 있다(Reich, 2009, 2013; Smith, 2017, 2018; 박진동, 2011; 김수미, 2014). 국내에서는 고등학생을 대상으로 한 연구 2편이 보고되었다. 박진동(2011)은 대학수학능력시험 한국 근현대사 과목에 응시했던 고등학교 3학년 학생과 한국 근현대사 과목을 이수한 2학년 재학생을 대상으로 2009학년도 대학수학능력시험 한국 근·현대사 20개 문항에서 선정한 6개 문항의 풀이 과정을 분석하였다. 연구 참여자가 문항을 풀면서 풀이 과정을 혼자 녹음하고, 연구자는 녹음된 내용을 분석하는 방식으로 연구가 진행되었다. 분석의 초점은 조사에 참여했던 4명의 학생이 6개 문항의 정답을 찾아내는 과정이었다. 분석을 통해 연구 참여자가 문두-제시문-답지를 순서대로 읽는 경우도 그렇지 않은 경우도 있었으며, 풀이과정에서 반복해서 읽기, 느리게 읽기, 침묵하기 등을 활용했음을 확인했다. 또한 재학생에 비해볼 때, 졸업생은 제시문을 중간 중간 생략해가며 읽어 풀이 시간을 단축하였다. 4명의 학생은 모두 상위권 학생이었고, 풀이의 단서를 잘 찾고 있다고 결론지었다. 소수의 학생을 대상으로 한 연구였기에, 이로부터 문항 풀이 과정에 대한 일반화를 이끌어내기는 힘들지만, 연구에 참여한 학생들이 본인이 가지고 있는 역사 지식 외에 다양한 수험전략을 활용하여 문항을 풀이하고 있음을 실증적으로 밝힐 수 있었다.

김수미(2014)는 8명의 고등학생을 대상으로 32개 국사 문항의 풀이 과정을 질문지 조사 방법, 제한된 시간 내에 문항을 풀게 하고 생각을 말하게 하는 면담, 심층면담의 3가지 방법을 사용하여 수집하고 분석하였다. 고등학생의 선다형 문항 풀이에 영향을 미치는 요인이 풀이 순서인지, 지식의 양인지, 자료 유형인지를 확인하려는 목적에서 수행한 연구였다. 곧 학생이 정답과 오답을 선택하는 데 영향을 미치는 요인을 탐구하였다. 연구 결과를 통해 고득점자는 역사 내용 지식이 많고, 리터러시 능력도 좋았으며, 수험요령도 풍부했음을 확인할 수 있었다. 역사지식 없이는 높은 수준의 사고로 나아갈 수 없다는 점을 확인하였고, 학생들은 역사지도 문항을 어렵게 느끼며 답지에 대해서 폭넓고 일반적인 내용과 구체적인 내용을 고려하였음을 밝혔다(김수미, 2014). 이를 토대로 김수미는 선다형 문항 개선 방안으로 답지를 작성함에 있어 역사 지식을 2가지 이상 포함해야 하고, 제시문으로 사용할 새로운 자료를 발굴해야 한다고 제언하였다.

비록 분석의 초점에 차이가 있었지만, 두 편의 선행 연구는 모두 고등학생을 대상으로, 학생이 정답을 찾는지 여부와 어떤 과정을 거쳐 정답을 찾는지에 초점을 두고 있었다. 학생들이 어떤 역사 지식과 역사적 사고력을, 어느 정도로 활용하여 문항을 풀이하고 있는지 그리고 학생이 활용한 지식과 사고력이 문항에서 측정하고자 한 것이었는지, 곧 타당도 측면에 초점을 맞춘 연구는 아니었다. 그럼에도 불구하고 이상의 선행 연구로부터 학생들이 역사 지식과 수험요령, 그리고 수험요령으로 분류하기 어려운 독해와 같은 사고력(김수미, 2014; 박진동, 2011)을 활용하여 선다형 문항을 풀고 있음을 확인할 수 있었다. 대학수학능력시험 문항을 포함하여 연구에 사용된 문항은 모두 역사적 사고력의 측정을 목표로 표방하고 있었으나, 학생들이 선다형 문항 풀이에 역사적 사고력을 활용했다는 사실은 확인할 수 없었다. 선다형 문항으로 역사적 사고력을 측정했음을 보고한 국내 연구는 아직까지 없다.

학생의 선다형 문항 풀이과정에 주목한 국외 선행연구 역시 학생들이 선다형 역사 문항을 푸는 과정에서 역사적 사고력을 활용하지 않는다는 결과를 제시하고 있다. Reich는 두 편의 연구(2009, 2013)에서 고등학생의 선다형 역사 문항 풀이과정을 분석했는데, 학생들의 선다형 문항 풀이과정에서 역사적 사고력의 활용을 확인하지는 못했다. 풀이과정 분석을 통해서 Reich는 학생들이 역사 지식과 수험요령, 그리고 읽기 능력(literacy)을 활용하고 있음을 확인할 수 있었다. 이를 토대로 Reich는 선다형 평가 결과가 지식의 측정이라는 측면에서 학생의 지식 이해도를 과대평가하고 있다고 주장했다. 학생들이 찍기 및 소거법과 같은 다양한 수험요령의 활용하기에, 실제 이해하고 있는 수준에 비해 시험 결과 점수가 높게 나타난다는 기존 선다형 평가 연구 결과를 재차 확인한 것이었다.

Smith(2017)는 3개 학교의 고등학생 27명(26명의 12학년생과 1명의 11학년생)을 대상으로 2010년도 미국 국가 수준 학업성취도 평가인 NAEP 12학년 미국사 기출 4문항을 푸는 과정을 분석하였다. 면담에는 문항을 풀이하면서 머릿속에 떠오르는 생각을 시간 간격을 두지 않고 즉각적으로 말로 표현하는 ‘동시적 생각말하기(concurrent think-aloud)’와 문항 풀이가 끝난 후 면담자의 추가적인 질문에 따라 풀이 과정을 부연 설명하는 ‘회고적 생각말하기(retrospective think-aloud)’ 방식을 활용하였다. 연구 참여자들은 대학 선이수과정인 AP 미국사를 1년 수학하고, 해당 시험에서 3점 이상 취득한 상위권 성적의 학생들이었고, 연구에서 다룬 문항 관련 내용지식을 학습한 상태였다. 면담에 사용한 4개 문항은 NAEP 미국사의 ‘역사적 분석과 해석’ 영역의 문항이었다. NAEP 미국사는 ‘역사 지식과 관점’ 영역으로 30%의 문항을, ‘역사적 분석과 해석’ 영역으로 70%의 문항을 출제한다. ‘역사 지식과 관점’ 영역은 전통적인 인물, 사건, 개념 등을 평가하는 영역이며, ‘역사적 분석과 해석’은 역사적 사고력 측정에 해당하는 영역이다. 역사적 사고력은 인물, 사건, 사료의 중요성을 판단하거나, 역사적 인물 및 역사가, 현재를 사는 사람들의 다양한 관점을 저울질하고 판단하는 등 다양한 사고활동을 포함한다(National Assessment Governing Board, 2011). 그러나 27명의 학생이 ‘역사적

분석과 해석' 측정을 목표로 한 4문항을 풀이한 108건의 사례에서 역사적 사고력의 활용은 찾아볼 수 없었다. 곧 27명의 상위권 학생들이 역사적 사고력을 측정한다고 개발된 문항 4개를 풀면서 역사적 사고력을 활용하지 않았다는 것이다. 그 대신 역사 지식의 소환, 수험전략, 그리고 독해 능력 활용을 확인할 수 있었다. 108건의 문항 풀이과정에서 84건은 역사 지식 인지/ 떠올리기에 해당했다. 나아가 학생들의 정/오답 여부가 학생들의 지식 이해 정도를 정확하게 나타내지도 못했다. 총 108문항 풀이 중 81건에서 학생들은 정답을 맞혔으나, 25%에 해당하는 21건에서는 관련 지식이 부족하거나 심각한 오개념을 가진 경우가 나타났다. Smith는 이를 토대로 NAEP 문항이 의도한 역사적 사고력을 측정하지 않고 있다는 결론에 도달했다. 나아가 역사 지식조차도 잘 측정하고 있다고 보기 어렵다고 부연하였다.

선다형 문항은 역사적 사고력의 측정보다는 지식의 측정에 더 적합한 평가이기 때문에, 문항 풀이 과정에서 역사적 사고력의 활용이 나타나지 않은 연구 결과를 당연하다고 여길 수도 있다. 또는 선다형 평가가 역사 지식이나 역사적 사고력을 정확하게 측정하지는 못하더라도 역사 지식, 독해, 일반적인 사고력 등의 복합적인 활용 능력을 어느 정도 범위 안에서 파악하게 해준다는 점(Haladyna, 2004)에 만족할 수도 있다. 즉 어떤 평가도 완벽한 평가는 있을 수 없기에, 선다형 평가 결과로부터 어느 정도 타당한 학습 성취에 대한 결과를 추론할 수 있다고 여길 수도 있다(Pellegrino, Chudowsky, & Glaser, 2001). 그럼에도 불구하고 선다형 평가가 광범위하게 사용되고 있는 점과, 역사적 사고력의 평가를 공공연하게 표방하고 있다는 점에서 과연 선다형 평가는 실제로 무엇을 어느 정도 측정하며 선다형 평가 결과 점수를 어떻게 해석하여야 하는지를 논의할 필요는 있다. 현재까지 진행된 선행 연구 결과에 기초해 볼 때, 대학수학능력시험이나 한국사능력검정시험 선다형 문항이 출제에서 의도한 역사 지식과 역사적 사고력을 측정하고 있는지에 대한 합리적 의문을 제기할 수 있기 때문이다. 국내 선다형 문항 풀이에 관한 선행 연구는 학생들이 문항 풀이 과정을 분석하였으나, 학생들이 문항 풀이에서 어떤 역사 지식과 역사적 사고력을 사용하는지를 분석의 초점으로 삼지 않았기에 연구 결과를 통해 문항의 타당도를 확인하기는 어려웠다. 역사 평가에 선다형 평가가 널리 사용되고 있고, 평가가 차지하는 중요도가 높음에도 불구하고 선다형 역사 문항을 통해 파악할 수 있는 실체가 무엇인지에 대한 연구는 매우 부족한 상황이다.

III. 연구 방법

본 연구에서는 선다형 평가 문항을 푸는 과정에서 응시자가 어떤 역사 지식과 역사적 사고력을 활용하는지를 파악하고자, 대표적인 선다형 평가인 한국사능력검정시험에서 12문항을 선정하여 10명의 대학생을 대상으로 연구를 진행하였다. 구체적인 연구 방법은 아래와 같다.

1. 연구 참여자

한국사능력검정시험은 대학수학능력시험에 비교해 볼 때 응시자가 다양하다는 특징이 있다. 초등학생에서부터 성인까지 연령도 다양하고, 시험을 치르는 목적도 단순 취미에서부터 진학 및 취업 준비에 이르기까지 다양하다. 한국사능력검정시험 응시자들에게 보이는 이러한 특성은 선행지식이나 기본적인 독해 능력이 응시자에 따라 큰 편차를 보일 수 있음을 의미한다. 연구 참여자를 무작위로 표집할 경우, 풀이 과정에서 나타나는 차이가 어디에서 기인하는 것인지 판단하기 어려울 수 있다. 이에 본 연구에서는 선행지식이나 독해 능력의 수준이 비교적 균질할 것으로 예상되는 대상을 찾고자 노력하였다(Patton, 2015). 이를 위해 중등교사 임용시험을 준비하는 대학생들을 대상으로 10명을 모집하였다. 연구 참여자 모집은 1개 사범대학에서 이루어졌다. 국어교육, 수학교육을 전공하고 있는 3-4학년 학생 중 최근 1년간 한국사능력검정시험에 응시한 적이 있는 학생들을 대상으로 하였다. 전공을 국어교육과 수학교육으로 제한한 것은 참여자의 세부 전공에 따른 반응 편차를 최소화하되, 문과와 이과를 아우르기 위함이었다. 동일 교사 양성 과정에 재학 중이며, 중등교사 임용시험을 목적으로 한국사능력검정시험을 준비했다는 점에서 한국사능력검정시험을 치르는 다른 피험자에 비해 균질성을 갖추고 있다고 판단하였다. 한국사능력검정시험 준비 및 응시 경험을 모집 조건으로 삼았기에, 연구에 참여한 학생 중에는 응시한 시험에서 급수를 취득하지 못한 경우도 있었다. 연구 참여는 자발성에 기초하였으며, 면담 참여에 대한 감사의 표시로 문화상품권 만원권을 증정하였다. 면담은 2018년 8월말과 2018년 9월초에 이루어졌고, 평균 1시간이 소요되었다. 연구 참여자의 기본 정보는 <표 1>과 같다.

<표 1> 연구 참여자 기본 정보

참여자 ID	전공	면담 시기	한국사능력검정시험 최근 응시 시기	한국사능력검정시험 취득급수
S01	수학교육	2018.08	2018.02	3급
S02	국어교육	2018.08	2018.08	-
S03	국어교육	2018.08	2018.08	3급
S04	수학교육	2018.08	2017.11	1급
S05	국어교육	2018.08	2018.08	3급
S06	수학교육	2018.08	2018.02	3급
S07	국어교육	2018.08	2018.05	1급
S08	수학교육	2018.08	2017.08	1급
S09	수학교육	2018.09	2018.08	3급
S10	국어교육	2018.09	2018.08	3급

2. 면담에 활용한 문항

본 연구에서 면담에 활용한 문항은 총 12문항으로, 인증시험으로서의 한국사능력검정시험 문항의 성격을 반영한 문항을 선정하고자 하였다. 한국사능력검정시험은 2006년 11월 처음 시행되었는데, 선발시험이 아닌 피험자의 학습 능력을 인증하는 시험이라는 특성을 지니고 있다. 시행 초기였던 2006년 11월부터 2007년 10월까지 6종의 시험이 실시된 적도 있었으나, 현재는 초/중/고급 3종의 시험을 통해 1급~6급까지의 한국사 능력을 인증하고 있다. 면담 문항은 현재의 초/중/고급 체제로 전환한 2011년 5월 시행 11회 시험부터 면담 수행 이전인 2018년 2월 시행 38회 시험까지를 대상으로 시대, 분야, 역사교육 목표 준거(역사 지식의 이해, 연대기 파악, 역사 상황 및 쟁점의 인식, 역사 자료의 분석 및 해석, 역사 탐구의 설계 및 수행, 결론의 도출 및 평가)를 고려하여 선정하였다. 시대와 분야, 역사교육 목표 준거는 한국사능력검정시험 문항 분석에 자주 사용된 항목이며(김수미, 2007; 박덕우, 2010; 김민정, 김미선, 2011; 김정수, 2011; 은동진, 2015; 박진철, 2016), 제시문의 유형(오정현, 2009)이나 정답률(김정수, 2011; 김덕진, 2012)을 활용하여 문항의 특성을 분석한 경우도 있었다. 본 연구에서는 이상의 요소를 고려하되, 각 요소의 산술적인 분포보다는 인증 시험으로서의 한국사능력검정시험의 대표성을 나타낼 수 있는 문항을 선정하고자 하였다. 이를 위해 한국사능력검정시험 출제 고경력자 2명이 상술한 요소를 고려하여 대표 문항이라고 판단한 80개 문항과 8회 이상 빈출된 42개 내용 주제의 문항을 중심으로 시대, 분야, 역사교육 목표 준거(역사 지식의 이해, 연대기 파악, 역사 상황 및 쟁점의 인식, 역사 자료의 분석 및 해석, 역사 탐구의 설계 및 수행, 결론의 도출 및 평가), 제시문 유형이 가능한 고르게 분포하는 방향에서 최종 문항을 선정하였다. 이렇게 선정된 문항의 기본 정보는 <표 2>와 같고, 정답률은 53%에서 87% 사이에 분포하였다. 면담에 활용한 문항은 부록으로 제시하였다.

<표 2> 12개 문항의 기본 정보

문항 ID	시대	분야	목표준거	문항 출처	주제	제시문 유형	정답률
Q01	선사	사회	역사 지식의 이해	고급 29회 1번	신석기 문화	사진	71%
Q02	고대	문화	결론의 도출 및 평가	고급 26회 8번	백제의 문화	설명문	87%
Q03	고대	사회	역사 상황 및 쟁점의 인식	고급 33회 8번	진대법	가상면담	70%
Q04	중세	문화	역사 자료의 분석 및 해석	고급 30회 14번	동명왕편	사료	74%
Q05	중세	정치	역사 지식의 이해	고급 32회 17번	고려 지방제도	지도	53%
Q06	중세	정치	역사 탐구의 설계 및 수행	고급 34회 23번	사화	사료	73%
Q07	통합	경제	연대기 파악	중급 36회 22번	조선시대 화폐	사진	65%
Q08	근대대동	경제	결론의 도출 및 평가	중급 35회 22번	대동법	학생대화	54%
Q09	근대	경제	역사 상황 및 쟁점의 인식	고급 26회 35번	호포법	사료	86%
Q10	근대	정치	역사 탐구의 설계 및 수행	중급 20회 44번	의열단	사진	63%
Q11	현대	정치	연대기 파악	고급 30회 46번	정부수립 과정	사진	77%
Q12	현대	정치	역사 자료의 분석 및 해석	중급 31회 48번	6월민주항쟁	사료	69%

3. 자료수집방법

면담은 면담자와 피면담자간의 1대 1 면담으로, 생각말하기(think aloud) 방법을 활용하였다. 생각말하기는 피면담자가 문항을 풀이하는 과정에서 드는 생각 모두를 말로 표현하도록 하는 방법(Ericsson & Simon, 1993), 바깥으로 드러나지 않는 피면담자의 풀이 과정을 파악하기 위해 자주 사용되는 방법이다. 이는 선다형 문항 풀이과정 분석을 위해 선행 연구에서도 사용된 바 있다.

생각말하기가 문항 풀이과정을 파악하는 데 효과적인 방법이기기는 하지만, 떠오르는 생각을 모두 표현하는 것은 자연스런 일은 아니다. 따라서 본 면담에 들어가기에 앞서 두 차례에 걸친 생각말하기 연습을 진행하였다. 먼저 면담자는 피면담자에게 뚜렷한 연관성을 가지지 않은 사진 14장을 제시한 후, 이를 자유롭게 분류해 보라는 과제를 부여하였다. 이러한 과제를 수행하는 과정에서 각 사진에서 무엇이 보이는지, 각 사진이 무엇이라 생각했는지, 어떤 기준을 가지고 사진을 분류하고 있는지 등을 말하도록 했다. 정돈되지 않은 생각도, 문장으로 만들기 힘든 생각도 모두 생각이 떠오를 때 바로 말하도록 했다. 피면담자가 말없이 사진을 보고 있거나, 분류를 진행하면, “무슨 생각을 하고 계신가요?”라고 물어 생각을 말하도록 요구했다.

다음으로 12개의 문항 외에 한국사능력검정시험에서 추가로 선정한 1개 문항을 대상으로 두 번째 생각말하기 연습을 진행하였다. 사진 분류하기 과제를 통해 생각말하기를 연습했었기에, 대부분의 참여자들이 문항을 풀면서 떠오르는 생각을 자유롭게 표현하였다. 더러 침묵이 오래 지속되는 경우 면담자가 “지금 무슨 생각을 하고 계신가요?”라고 물어 참여자가 계속해서 지금 어떤 생각을 하는지 말하도록 유도하였다. 이러한 연습 과정은 참여자들에게는 설명만 듣고는 이해하기 어려운 생각말하기 방법에 익숙해질 수 있는 기회를, 면담자에게는 참여자들이 본인들의 인지 과정을 말로 전달하는데 능숙함을 확인할 기회를 제공하였다.

이와 같은 두 차례의 연습을 거친 뒤, 피면담자는 12개 문항을 생각을 말하면서 풀었다. 두 차례의 연습을 했어도 생각말하기 방법에 숙련되지 않은 피면담자가 떠오르는 생각을 모두 말할 수는 없기에, 문항 풀이가 끝난 후 추가적인 질문을 했다. 생각말하기 중 언급이 되지 않았던 답지를 중심으로 “이 답지는 어떻게 읽었나요?” “이 답지에 대해서는 특별히 언급이 없었는데, 추가할 내용이 있나요?” 등의 질문을 통해 문항 풀이 이후 생각을 파악하고자 했다.

4. 자료 분석 방법

중등학생의 역사 선다형 문항 풀이과정을 분석한 선행 연구에 따르면, 학생들은 선다형 문항을 풀면서 3가지 중 한 가지의 인지과정을 거친다. 첫 번째는 학생 본인이 갖고 있는 지식을 소환하여 문항을 풀이하는 것이다. 선행 연구의 ‘역사적 내용’이나 ‘사실 떠올리기/사실 인

지' 코드가 이에 해당한다. 두 번째는 답이 아닌 것을 찾아 제외한 후, 나머지 답지만을 고려하여 답을 찾는 소거법과 같은 수험요령의 활용이다. 세 번째는 독해와 같은 일반적인 읽기 능력을 활용해 답을 찾아내는 것이다. 본 연구에서는 전체 면담 자료의 10%에 해당하는 자료를 가지고 수행한 파일럿 분석에 '지식 소환', '수험전략 활용', '독해'라는 세 가지 코드를 사용했다. 그리고 세 가지 코드로 분석되지 않는 경우는 없는지를 두 명의 저자가 독립적으로 확인하였다. 파일럿 분석 결과, '지식 소환'과 '수험전략 활용' 코드는 그대로 유지할 수 있었으나 세 번째 코드로 분류된 문항 풀이과정은 독해보다는 추론을 활용한 경우가 많음을 확인할 수 있었다. 관련 선행 연구를 수행했던 Smith(2017)는 지식을 소환하지 않고 문항에 제시된 특정 구절의 내용으로부터 답을 추론하는 것을 독해 능력(reading comprehension)이라고 명한 바 있다. 문항 풀이에 있어서 '독해'란 주어진 텍스트로부터 답을 끌어내는 경우에 해당하는데, 본 연구에서는 이와 같은 형태의 '독해'보다는 조금 더 적극적인 형태의 '추론'이 나타났던 관계로 이를 '추론'으로 명명하여 코딩하였다. '추론'에서 피면담자는 텍스트로부터 답을 도출하는 '독해' 이상의 인지 활동을 보여준 경우가 많았으나, 이 때의 '추론'은 문항 출제자가 측정하고자 의도했던 지식이나 사고력과 무관한 양상(construct-irrelevant reasoning)을 보였다. 역사학에 고유한 특성을 보이는 추론이 나타나지 않아 역사적 사고와는 거리가 있었고, 논리성이 결여되는 경우도 종종 나타났기 때문이다. 한국사능력검정시험 목표 준거에 명시된 역사 상황 및 쟁점의 인식, 역사 자료의 분석 및 해석, 역사 탐구의 설계 및 수행 등 역사학에 고유한 사고로 보기 어려운 일반적인 자료 분석과 해석, 일반적인 판단력의 활용은 '추론'으로 분류하였다. 이상의 세 가지 코드에 '익숙도'라는 코드를 추가하였다. 파일럿 분석에서 연구 참여자가 내용이 어느 정도 익숙한지에 기초하여 답을 찾는 경우를 2번 이상 확인하였기고, 이러한 현상이 얼마나 자주 나타나는지를 확인하기 위해 '지식 소환', '수험전략 활용', '추론', '익숙도'라는 4개의 코드를 활용하여 면담을 분석하였다. 그러나 분석 결과 검토 과정에서 '익숙도'는 수험전략의 하나라는 결론에 도달하여, 최종적인 분석틀로는 '지식 소환', '수험전략 활용', '추론'을 사용하였다.

연구 참여자가 문항을 풀면서 지식을 전혀 소환하지 않은 경우는 없었다. 소거법과 같은 수험전략 역시 거의 모든 문항 풀이에 등장했다. 그러나 1문항 풀이-1코드 적용을 원칙으로, 연구 참여자의 정답 선택에 가장 결정적으로 활용된 풀이 과정을 해당 사례의 코드로 분석하였다. 분석 결과의 신뢰도를 확보하기 위해, 두 명의 저자가 독립적으로 20%에 해당하는 자료를 코딩한 후 코헨 카파 계수를 사용하여 신뢰도를 확인하였다. 카파 계수는 '지식 소환', '수험전략 활용', '추론' 코딩 각각에서 모두 0.80 이상이었으며, 서로 다르게 코딩한 자료에 대해서는 논의를 통해 합의에 다다를 수 있었다. 신뢰도 확인 이후 두 번째 저자가 모든 자료를 코딩하였고, 분석 결과는 코드별로, 문항별로, 피면담자별로 비교·대조하여 연구 참여자가 문항 풀이과정에서 어떤 지식과 역사적 사고력을 활용했는지를 확인하였다.

IV. 연구 결과

연구 참여자는 12개 문항을 풀면서 지식 소환, 수험전략 활용, 추론이라는 문항 풀이 과정을 거쳤다. 다음에서는 각각의 문항 풀이 과정별로 연구 참여자가 어떤 지식과 사고력을 사용하였는지를 분석한 후, 문항별 그리고 연구 참여자별 문항 풀이에서 나타난 특성을 분석하였다.

1. 문항 풀이 과정 분석: 지식 소환, 수험전략 활용, 추론

연구 참여자 10명이 12개 문항을 풀이한 120건의 문항 풀이 사례를 분석한 결과, 연구 참여자는 지식 소환을 활용하여 59건(49%), 수험전략을 활용하여 34건(28%), 추론을 통해 27건(23%)의 문항을 풀었다. 이를 통해 연구 참여자는 절반 가까운 사례에서 한국사능력검정시험 목표 준거에 명시되지 않았던 수험전략과 추론을 활용하여 문항을 풀이하고 있음을 확인할 수 있었다. 반면 한국사능력검정시험 목표 준거로 명시되었던 역사적 사고를 활용한 문항 풀이 사례는 찾아보기 어려웠다. 예를 들어 역사적 상황과 쟁점의 인식을 목표 준거로 설계된 문항에 정답을 맞힌 경우에도, 관련한 역사적 사고력의 활용을 찾아보기 어려웠다. 이는 국내 외에서 수행된 선행 연구와도 일치하는 결과이다(Reich, 2009, 2013; Smith, 2017; 박진동, 2011; 김수미, 2014). 지식 소환, 수험전략 활용, 추론을 활용한 대표적인 문항 풀이 양상은 다음과 같다.

지식 소환의 경우, 전체 59건 중 73%에 해당하는 43건이 정답을 맞힌 사례였다. 수험전략 활용(12건, 35%)이나 추론(15건, 56%)에 비해 정답을 맞힌 비율이 높게 나타났다. 그러면 연구 참여자는 어떤 지식을 소환하여 문항을 풀이했을까? 다음은 S04가 신석기 시대에 대한 1번 문항을 풀면서 지식을 소환한 사례로, 지식 소환을 하여 문항을 푼다고 할 때 예상할 수 있는 방식으로 문항을 풀이하고 있다. S04는 우선 제시문에 등장한 ‘고산리 유적’과 ‘갈판’이라는 키워드를 통해, 1번 문항이 신석기 시대에 관한 문제임을 어렵지 않게 파악하였다(1번 문항은 부록 참조).

[제시문을 읽으며] 뭐라 그러죠? 신석기, 간석기 그때인 것 같은데. 고산리 유적이니까, 갈판이 나왔으니까. 아마 신석기겠구나. [문두를 읽으며] 그러면 신석기에 대한 옳은 것을 고르면 되는 거니까 [2번 답지를 읽으며] 군장 없었고 2번 틀렸고. [3번 답지를 읽으며] 3번은 청동기에 대한 내용이고, [4번 답지를 읽으며] 어 4번이 정착 생활 나왔으니까 움집 맞고. [1번 답지를 읽으며] 1번은 독무덤이 아마 청동기였던 것 같고, [5번 답지를 읽으며] 우경은 나오려면 쇠가 있어야 되니까 이건 청동기고. 그러니까 4번.

S04는 본인이 가지고 있는 지식을 바탕으로 독무덤, 군장, 반량전과 명도전, 우경이 신석기 시대와 관련이 없다고 판단한 후, ‘정착 생활이 시작되면서 움집이 나타났다’를 정답으로 선택하였다. 풀이 과정을 볼 때, S04는 신석기 시대와 청동기 시대에 대한 기본적인 지식을 가지고 있으며, 이러한 지식을 활용하여 답을 고를 수 있었다.

그러나 지식을 소환해 문항을 풀이했던 59건(49%)의 사례 중에는 위 사례와는 달리 틀린 지식을 소환하여 정답지를 선택하는 경우도 있었다. 다음 예시는 S07이 6번 사화 문항을 풀이한 사례다(6번 문항은 부록 참조). 이 문항은 제시문에서 (가) 갑자사화와 (나) 을사사화를 확인한 후, 두 사건 사이의 시기에 있었던 기묘사화를 찾기를 요구하고 있다. S07은 다음과 같은 과정을 거쳐서 정답인 2번 ‘위훈 삭제를 주장한 조광조 일파가 축출되었다’를 답으로 선택하였다.

(가)[갑자사화], (나)[을사사화] 사이. [제시문 (가)를 읽으며] 생모 윤씨 폐비. 폐비 윤씨니까, 두 번째 사화. [제시문 (나)를 읽으며] 유인숙. 음모를 꾸미며 몰래 사부를 불러다...이에 윤임, 유관, 유인숙 세 사람에게 사사를 명한다. 이 (나) 사건은 잘 모르겠으니 보기[답지]를 보자. [1번 답지] 서인이 반정을 일으켜 정권을 장악하였다. 동인, 서인이 사화 전에 싸웠으니까 이전 시대니까 답이 아님. [2번 답지] 위훈 삭제를 주장한 조광조 일파가 축출되었다. 조광조가 맨 마지막 기묘사화니까 세모. [3번 답지] 정여립 모반 사건을 계기로 동인이 피해를 입었다. 음 이것도 이전 시대인 것 같은데 일단 세모. [4번 답지] 효종이 죽자는 이후 시대인 것 같으니까 X. [5번 답지] 사림이 이조전랑 임명을 두고 동인과 서인으로 나뉘었다. 보기 중에 제일 이전 시대니까 X. 그래서 2번, 3번 중에 고민을 하는데, 음 2번이 확실히 (가) 시기 이후에 있었던 일이니까, 답은 2번.

S07은 (가)에 등장하는 “폐비 윤씨”라는 키워드를 통해 제시문 (가)가 “두 번째 사화”인 갑자사화임을 파악했지만, (나)가 어떤 사건에 대한 사료인지는 파악하지 못하였다. (나) 사건을 특정하지 못한 상태에서 S07은 답지 1번, 4번, 5번을 제외하고 2번과 3번 중에서 고민을 하다가 2번이 (가) 이후에 있었던 일이 확실하다는 선행 지식에 근거하여 이를 정답으로 골랐다. 정답을 확정하는 데에는 지식을 소환하였으나, S07은 정답에 접근하는 과정에서 대표적인 수험전략인 소거법도 활용하였다. 그런데 문항 풀이 과정을 살펴보면 1번과 5번 답지를 소거한 이유가 틀린 지식에 근거했음을 알 수 있다. 1번과 5번 답지는 봉당정치에 관한 것인데, S07은 봉당정치가 사화보다 앞서 발생하였기 때문에 답이 될 수 없다고 보고 두 답지를 소거했다. 봉당이 사화보다 앞서 발생한 사건이라는 ‘틀린’ 지식을 소환하여 문항을 푼 것이다. 또 S07은 기묘사화가 “맨 마지막 사화”라는 잘못된 지식을 가지고 2번 답지를 정답으로 선택했다. 이러한 과정을 통해 S07은 12개 면담 문항 중 난도가 가장 높았던 문항(정답률 53%)의 정답을 찾을 수 있었다. 이 문항의 5개 답지는 모두 (가) 이후의 사건으로, (가)와 (나) 사건을 모두 특정해야 풀 수 있는 난도가 있는 문항이었다. 이는 지식 소환에 100%에 의존해 풀 수

도 있고, (가), (나), 그리고 5개의 답지의 맥락과 관계를 검토하는 폭넓은 역사적 사고력을 활용하여 풀 수도 있는 문항이었는데, S07은 틀린 지식과 소거법에 기초하여 문항의 정답을 찾았던 것이다.

이와 같은 분석 결과에 기초하여 볼 때, S07이 6번 문항을 맞혔다는 결과를 어떻게 해석해야 하는지의 문제가 생긴다. 내용 주제 측면에서 이 문항은 조선시대 사회에 관한 문항이며, 역사교육 목표 준거는 역사 탐구의 설계 및 수행을 측정하고자 했다. 비록 정답은 맞혔지만, 봉당이 사회 이전에 발생했으며 기묘사회가 네 번째 사회라 인지하고 있는 S07이 조선 시대 사회를 이해하고 있다고 평가하기는 어렵다. “제시된 문제의 성격과 목적을 고려하여 절차와 방법에 따라 역사 탐구를 설계하고 수행하는 능력” 또한 찾아볼 수 없었다(국사편찬위원회, n.d.). 물론 이처럼 틀린 지식을 가지고 선다형 문항의 정답을 맞히는 경우는 예외일 것이라는 합리적 의심을 해 볼 수 있을 것이다.

그러나 본 연구에서 잘못된 지식을 활용하여 정답을 찾아내는 현상은 다른 주제, 다른 시대, 다른 참여자의 문항 풀이에서도 종종 목격되었다. 다음에서도 확인 가능하듯이 S08은 광복 이후 5·10 총선거까지의 현대사 흐름을 묻는 11번 문항을 풀면서 모스크바 3국 외상 회의와 5·10 총선거 실시 사이에 위치한 (가)에 들어갈 수 없는 사건을 골라내었다(11번 문항은 부록 참조). S08은 정답인 ‘반민족 행위 특별 조사 위원회 활동’을 답으로 선택하였으나, 정답을 고른 이유를 반민특위가 광복하자마자 활동했었기 때문에, 즉 (가)보다 앞서 일어난 일인 것 같아서라고 설명했다. 또한 1번 답지로 제시된 좌·우합작 위원회 활동이 (가)에 들어가는지 여부를 확신하지 못하였고, 제1차 미·소 공동 위원회 개최나 유엔 한국 임시 위원단 방한까지 좌·우 갈등을 봉합하는 운동의 일환으로 이해하고 있었다. 이처럼 S08은 모스크바 3국 외상회의와 5·10 총선거 실시 사이에 좌·우 갈등 봉합 운동이 있었다는 대강의 흐름만 알고 있었을 뿐, 답지로 제시되었던 사건이 구체적으로 언제 어떤 맥락에서 발생했는지 이해하고 정답을 고르지는 않았다. 따라서 S08가 정답은 맞혔지만, 이를 통해 S08이 광복에서 5·10 총선거까지의 여러 사건 내용에 대한 확실한 지식을 지니고 있다거나, 사건의 흐름을 시대순으로 정확하게 이해하고 있다고 보기는 어렵다.

지식 소환을 활용한 문항 풀이 중 위에 제시된 S07과 S08의 사례는 예외적이지는 않았으며, 오히려 EBS 문항 풀이 설명 방식과 유사하게 문항을 풀이했던 S04와 같은 사례가 드물었다. 선다형 문항이 지식을 측정하는데 장점이 있다고 하지만, 출제자가 의도하는 지식을 그대로 측정하고 있는 것은 아닐 수 있다는 추정을 가능케 하는 지점이다. 아울러 S07과 S08이 모두 한국사능력검정시험 1급을 취득했었기에, 한국사능력검정시험 1급의 의미를 어떻게 해석해야 하는지에 대해서도 의문을 제기할 수 있다. 지식 소환을 활용했던 문항 풀이 분석 결과는 적어도 선다형 문항을 출제하는 측의 의도와 실제 피험자의 문항 풀이 사이에는 간극이 분명하게 존재하고 있음을 드러내 보여주고 있었다.

지식 소환이 여의치 않을 때, 연구 참여자는 수험전략이나 추론을 활용하여 정답을 찾아나갔다. 전체 문항 풀이의 28%에 해당했던 수험전략 풀이의 경우, 연구 참여자는 앞서 언급한 소거법, 그리고 찍기를 다수 활용하였다. 소거법은 제한적인 지식이나 활용하여 정답이 될 수 없는 답지들을 우선 제거해가면서 최후에 남는 답지를 정답으로 고르는 방식이다. 본 연구에서도 소거법을 이용하여 정답을 찾은 경우가 9번 있었다. 이 중 정답을 맞힌 경우는 5번, 오답을 선택한 경우는 4번으로 비율상 큰 차이는 없었다. 소거법을 활용하였지만, 정답을 찾기 힘들 경우, 연구 참여자는 찍기를 활용했다. 찍기는 문제를 풀 단서가 거의 없을 경우 나머지 다른 문제를 위해 해당 문제에서 시간을 절약하기 위해 많이 활용하는 수험전략의 하나이다(Blackey, 2009). 널리 알려진 소거법이나 찍기 외에 연구 참여자는 자신에게 익숙한 답지를 정답으로 선택하기도 했다. 익숙도는 문항을 풀 단서를 찾지 못해 답지 중 하나를 직감적으로 고른다는 면에서 찍기와 유사하지만, 가장 익숙한 단어나 표현이 들어간 답지를 고른다는 점에서 일반적인 찍기와 차별성을 보이고 있었다.

S10은 의열단에 대한 옳은 설명을 찾는 10번 문항을 읽은 후(10번 문항 부록 참조), 이 주제에 대해서는 “아무것도 모르기 때문에” 찍을 수밖에 없을 거 같다고 밝혔다. S10은 문항에 나오는 단체가 의열단임을 특정하지 못했으나, <보기>에 제시된 조선 혁명 선언이 무력 투쟁을 정당화하는 글이었다는 어렴풋한 기억을 가지고 있었다. 이를 토대로 S10은 ㄴ. ‘조선 혁명 선언을 활동 지침으로 삼았다’가 옳은 설명일 것이라고 보고, ㄴ과 결합된 보기 ㄴ, ㄷ을 탐색하였다. S10은 ㄷ을 골랐는데, 김원봉이라는 이름이 기억에 강렬하게 남아있다는 것이 이유였다. S10외에 4명의 참여자(S01, S02, S05, S08)가 익숙도를 활용하였는데, 예를 들어 S02는 ‘5소경’이라는 단어를 “되게 많이 봤기 때문에” 답으로 택했고, S05는 “많이 들어본 걸 그래도 내지 않을까?”라는 생각에서 잘 모를 경우 제일 익숙한 답지를 고른다고 설명했다. 120개 문항 풀이 사례 중 익숙도에 기초하여 최종 정답을 선택한 경우는 총 7차례로 그 수가 많지는 않았다. 그러나 익숙도는 선행 연구에서 나타나지 않았던 새로운 수험전략이라는 점에서 특기할 만 하다. ‘익숙도’라는 풀이 전략이 인증시험이라는 한국사능력검정시험의 특수성에서 기인하는 것인지 확인하기 위해서는 좀 더 면밀한 연구가 필요하다.

연구 참여자는 지식 소환과 수험전략 외에 추론을 통해서도 문항을 풀이했다. 이는 일반적인 사고력에 가장 가깝다고 볼 수 있는 인지 과정이었다. 추론으로 분류된 인지 과정은 선행 지식을 소환하여 문제의 답을 찾거나, 소거법이나 찍기와 같은 명백한 수험전략이 아닌 사고 활동을 의미하는데, 역사적 사고를 활용한 사고활동은 아니었다. 전체 문항 풀이의 23%에 해당했다. 『동명왕편』 발췌문을 주고 이에 대한 옳은 설명을 찾는 문항에서 S01은 다음과 같이 추론을 활용하였다.

[제시문] 음, 『동명왕편』. [3번 답지]남북국을 최초로 사용했는지는 정확히 모르겠고, [제시문] 귀환, 귀, 성, 귀가 아니라 신이었다, 이들은 장차 무엇을 볼 것인가, 그냥 역사서에 관한 내용인 것 같은데, 후인들이 무엇을 볼 것이냐라고 되어 있으니, 후인들을 중요하게 여기는 거고, 누군가를 기억하게 하려고 한 것이니까, [1번 답지]고구려 계승의식을 기억하게 하려고 하는 것 같다. 끝이요.

S01은 제시된 사료 말미에 ‘후인들이 무엇을 볼 것이냐’는 문구가 있기에 사료의 저자가 후인들을 중요하게 여긴다는 것을 알 수 있고, 후인들을 중요하게 여기므로 누군가를 기억하게 하려고 할 것이라고 추론하였다. 이러한 추론을 바탕으로 S01은 기억과 가장 관련성이 높은 단어인 ‘계승’이 포함된 1번 답지 ‘고구려 계승 의식이 반영되었다’를 답으로 골랐고, 실제로 1번이 정답이었다. ‘후인→기억→계승’으로 이어지는 추론 과정은 일련의 흐름을 갖추고 있어 의식적인 사고 활동에는 해당하지만, 역사학에 고유한 사고를 보여주고 있지는 않으며 확고한 논리를 갖추고 있지도 않다. S01과 마찬가지로 3급을 취득했던 S06도 “‘후인’ 얘기가 나왔으니까, 1번 계승한 거, 그런 게 답일 것 같아서, 1번이 답인 것 같아요.”라며 유사한 추론을 거쳐 정답을 선택하였다. ‘후인’에 주목하지는 않았지만, 역시 추론으로 분류된 S04의 풀이 사례는 다음과 같다.

S04: [제시문을 읽으며] 구삼국사를 얻어 동명왕본기를.. 아 『동명왕편』 이구나. 음.. [제시문] 신기한 사적.. 환이 아니고 성이며 귀가 아니고 신이었다. 약간 추상적인 개념이 많이 있는 것 같아요. 그래서 동명왕이니까, 고구려. 동명왕이니까. 고구려를, 약간 그쪽의 이름이 들어가서, 고구려 쪽의 계승의식이 반영된 것 같아서, 1번 할 것 같고. 2번은 삼국사기가 가장 오래된 역사서였고. [3번 답지] 남북국이라는 용어를 처음 사용한 것도. 어 이것도 이것 같은데... 어... 지금 [3번 답지] 남북국이랑 [1번 답지] 고구려 계승의식이랑 두 개가 헷갈리는데. [4번 답지] 연대순으로 기록하는 편년체는, 이걸 삼국사기였던 것 같고. 아. 2번은 삼국유사구나. [5번 답지] 단군 조선부터 고려까지의 역사는 이걸 『동명왕편』 이후에 나왔던 것 같아요. 1번 아니면 3번인데. 음... 3번으로. 아니 1번 할 것 같아요. 처음 생각했던 걸로. 3번은 확신이 없어 가지고.

면답자: 1번은 확실하게 아는 거라서?

S04: 네. 약간 이름 보고 느낌이. 『동명왕편』이라는 이름 보고, 그런 느낌이 들었어요.

면답자: 동명왕이라는 이름이 왜요?

S04: 동쪽의 밝은 왕이니까 고구려랑 매칭이 되어 가지고요.

S04는 동명왕이라는 이름이 동쪽의 밝은 왕이라는 뜻이므로 고구려와 연결이 된다고 이야기하였다. 전근대 중국 왕조의 입장에서 한반도 왕조를 ‘동쪽의 왕’이라 지칭하긴 하였으나, 그 호칭이 비단 고구려에만 해당되는 것은 아니었다. 그럼에도 불구하고 S04는 동쪽의 밝은

왕이라는 의미를 내포한 호칭이 고구려왕을 지칭한다고 보았고, 이를 근거로 고구려가 들어
가 있는 답지를 골랐다. S04는 동명왕이 누구인지를 설명하지는 못했고, 면담에서 주몽이나
이규보 혹은 『동국이상국집』에 대한 이해를 보여주지 않았다.

2. 문항별·참여자별 분석

문항별 분석에서는 문항별 풀이 과정 코드 분포를 파악한 후, 특정 코드가 많이 나타난 문
항의 특성을 분석하였다(<표 3> 참조). 문항의 특성이 연구 참여자의 풀이 과정에 영향을 미
쳤는지를 확인하기 위한 분석이었다. 지식 소환이 우세했던 문항은 총 5개(1번 문항, 2번 문
항, 7번 문항, 8번 문항, 11번 문항)로 나타났다. 5문항을 풀이하는 데 있어 60%이상의 연구
참여자가 지식 소환을 사용하였다. 수험전략이 우세하게 나타난 문항은 10번 문항 1개로, 총
10명 중 7명의 연구 참여자가 수험전략을 활용하여 답을 찾았다. 추론이 우세했던 문항으로는
4번 문항과 9번 문항, 2개 문항이 있었다. 나머지 4개 문항인 3번 문항, 5번 문항, 6번 문항,
12번 문항에서는 특정 코드의 우세가 나타나지 않았다.

지식 소환이 우세했던 5개 문항의 경우, 특정 패턴이나 특징을 확인하기 어려웠다. 수험전
략 활용이나 추론과는 달리 지식 소환은 ‘역사 지식의 이해’라는 목표 준거와 긴밀한 관련이
있었으나, 해당 목표 준거와 지식 소환 간의 연결성은 찾아보기 힘들었다. ‘역사 지식의 이해’
가 준거였던 문항으로는 1번 문항과 5번 문항이 있었는데, 1번 문항에서는 전체 참여자 10명
중 9명이 지식 소환을 사용하여 지식 소환이 절대적 우세를 보였으나, 5번 문항에서는 지식
소환 4명, 수험전략 4명, 추론 2명으로 특정 방법의 우세가 나타나지 않았다. 반면 ‘역사 지식
의 이해’가 목표 준거로 설정되지 않았던 문항에서 지식 소환이 우세를 나타내기도 했다. ‘연
대기 파악’ 문항 2개(7번 문항, 11번 문항)와 ‘결론의 도출 및 평가’(2번 문항, 8번 문항)에서
지식 소환이 우세를 보였다. 10명의 참여자와 12개 문항 분석 결과로부터 일반화를 도출하기
는 어렵지만, 면담에 활용한 문항이 목표 준거를 측정하고 있다는 증거는 찾을 수 없었다.

다음으로 수험전략이 우세하게 나타났던 문항을 살펴보았다. 수험전략이 우세했던 문항은
Q10 1개로 김상옥, 나석주의 활동을 보여주는 제시문을 통해 의열단을 추론한 후, 의열단에
대한 옳은 설명을 찾는 합답형 문항이었다(10번 문항은 부록 참조). 10명의 참여자 중 1명이
지식 소환, 7명이 수험전략, 2명이 추론으로 문항을 풀이하였다(<표 3> 참조). Q10을 풀며 참
여자들은 “뭔가 단체가 너무 많아서 헷갈”(S06)린다거나 “이 부분은 완전 기억이 날라
갔”(S08)다고 토로하였다. S10은 “아예 아무 것도, 지금은 모르겠어요. 그래서 완전 찍을 것
같아요”라고 말한 후, 김원봉 이름이 익숙하다는 점을 활용하여 답지를 선택하였다. S08은 한
국사능력검정시험에 응시할 때 이 부분은 “2~3일 전에” 급박하게 외운다고 언급하기도 하였
다. 실제 일제 강점기 독립운동은 단체의 수가 많고 명칭도 유사하여 학생들이 학습에 어려움

을 겪곤 한다. 학생들은 이 시기를 맥락적으로 학습하기보다는벼락치기 방식으로 암기하고 있고, 이렇게 암기된 지식은 시험이 끝나면 쉽게 휘발됨을 확인할 수 있다.

추론 활용 비율이 높았던 문항의 경우도 특징이 나타났다. 추론 활용 비율이 높게 나타났던 4번 문항과 9번 문항은 사료를 제시문으로 제공하였다는 공통점이 있었다. 10개 문항 중 사료를 제시한 문항은 모두 4문항(4번 문항, 6번 문항, 9번 문항, 12번 문항)이었다. 비록 추론을 활용한 연구 참여자가 다수를 구성하지는 않았지만, 12번 문항의 경우도 지식 소환을 활용한 참여자가 5명, 추론을 활용한 참여자가 5명으로 추론 활용도가 높은 편이었다(<표 3> 참조). 곧 사료를 제시한 네 문항 중 6번 문항을 제외한 세 문항의 경우, 추론의 비율이 높게 나타났다. 표본이 적어 일반화는 어렵지만, 사료를 제시할 경우 그렇지 않은 경우보다 학생들이 지식 소환 외의 인지과정을 활용하여 문항을 풀 가능성이 높아진다고 추정해 볼 수 있다. 사료를 제시한 문항의 경우, 학생들은 사료의 내용에 기초해 답을 유추하려는 노력을 전개하였고, 이는 추론을 중심으로 답을 선택한 비율이 높게 나타나는 결과로 나타났다. 다음 S08의 문항 풀이 과정에서 사료 내용을 활용한 추론을 쉽게 확인할 수 있다.

S08: [문두] 홍선대원군이 실시한 정책으로 옳은 것은? [제시문 1] 어깨 팔뚝 드러난 적삼 입은 어린 것들 바지 버선 한 번도 못 입었으리. 큰 아이는 다섯 살에 기병에 등록되고, 작은 놈도 세 살에 군적에 올라 있어. [제시문 2]시아버지 삼년상 벌써 지났고 갓난아이 배넛물도 안 말랐는데, 이 집 삼대 이름 군적에 모두 실렸네. 아 네. [2번 답지] 5군영에서 2영으로 군제를 개편하였다 이거인 것 같네요.

면담자: 왜 그렇게 생각해요?

S08: [제시문] 두 개 다 얘기가 군에 관련된 얘긴데 지금 [답지] 1번에서 5번 중에 군에 관련된 얘기는 2번밖에 없는 것 같아요.

S08은 제시문의 사료 발췌문에 공통으로 등장하는 ‘군적’이라는 단어로부터 본 문항이 군과 관련된 내용이라는 것을 추론한 후, 군제 개편을 답으로 골랐다. 군에 관련된 문제에 대한 답을 찾아야 하는데 군을 언급한 것이 2번밖에 없다는 논리에서이다.

그러나 사료를 활용한 문항이 반드시 참여자의 추론을 유도했던 것은 아니다. Q6의 경우 사료 발췌문이 두 개 제시되었으나, 10명의 참여자 중 추론을 활용하여 문항을 풀이한 학생은 단 한 명에 그쳤다. Q6의 제시문에는 ‘존호(尊號)’, ‘중형(重刑)’, ‘가산(家産)’, ‘사부(師傅)’, ‘현부(賢否)’, ‘사사(賜死)’와 같은 난해한 용어가 많이 포함되어 있었는데, 이 중 ‘중형’이나 ‘가산’은 한자어가 병기되지도 않았으며, ‘현부(賢否)’나 ‘사사(賜死)’는 괄호 안에 한자어가 병기되었으나 어휘 수준이 높지 않으면 여전히 뜻을 유추하기 어려운 단어였다. 즉 6번 문항에서 제시된 사료는 참여자가 꼼꼼하게 읽어도 전체적인 내용을 파악하기 힘들었을 뿐 아니라, 정답을 고르기 위해서는 사료를 통해 두 개의 사건(갑자사화와 을사사화)을 각각 정확히 특정해야

했기에, 사료 내용을 활용한 추론이 전개되지 않았다.

지식 소환, 수험전략 활용, 추론의 활용은 연구 참여자에 따라서도 다르게 나타났다. (<표 4> 참조). 연구 참여자들이 12문항을 풀이하면서 가장 많이 활용한 코드 유형에 차이가 보인 것이다. 연구 참여자 중 6명은 지식 소환을 가장 많이 활용하였고, 수험전략과 추론을 많이 활용한 경우는 각 1명이었다. S02나 S08의 문항 풀이에서는 지식 소환, 수험전략 활용, 추론이 다소 고르게 분포하고 있었다.

<표 3> 문항별 분석 결과

문항번호	지식 소환		수험전략		추론		총계		목표 준거	참여자 정답률	실제 정답률
Q01	9	(90%)	1	(10%)	0	(0%)	10	(100%)	역사 지식의 이해	40%	71%
Q02	7	(70%)	2	(20%)	1	(10%)	10	(100%)	결론의 도출 및 평가	80%	87%
Q03	3	(30%)	4	(40%)	3	(30%)	10	(100%)	역사상황 및 쟁점의 인식	60%	70%
Q04	3	(30%)	1	(10%)	6	(60%)	10	(100%)	역사자료의 분석 및 해석	70%	74%
Q05	4	(40%)	4	(40%)	2	(20%)	10	(100%)	역사 지식의 이해	10%	53%
Q06	5	(50%)	4	(40%)	1	(10%)	10	(100%)	역사탐구의 설계 및 수행	30%	73%
Q07	6	(60%)	4	(40%)	0	(0%)	10	(100%)	연대기과약	80%	65%
Q08	6	(60%)	3	(30%)	1	(10%)	10	(100%)	결론의 도출 및 평가	50%	54%
Q09	3	(30%)	2	(20%)	5	(50%)	10	(100%)	역사상황 및 쟁점의 인식	60%	86%
Q10	1	(10%)	7	(70%)	2	(20%)	10	(100%)	역사탐구의 설계 및 수행	60%	63%
Q11	7	(70%)	2	(20%)	1	(10%)	10	(100%)	연대기과약	80%	77%
Q12	5	(50%)	0	(0%)	5	(50%)	10	(100%)	역사자료의 분석 및 해석	90%	69%
총계	59	(49%)	34	(28%)	27	(23%)	120	(100%)	-	-	-

※ 회색은 우세하게 나타난 코딩을 의미하며, 사선은 우세한 코딩이 나타나지 않은 경우를 의미한다.

연구 참여자가 어떤 코드를 가장 많이 활용했는지와 정답률 간에도 일정정도의 관련성을 엿볼 수 있었다. 지식 소환을 가장 많이 활용했던 참여자의 정답률은 58%에서 83%였는데 반해, 수험전략을 가장 많이 활용했던 참여자 2명(S03과 S08)의 정답률은 모두 25%로, 전체 참여자 중 가장 낮게 나타났다. 수험전략에 의존했던 학생들은 문항에서 평가하고 있는 내용을 모르거나, 공부를 하긴 했으나 기억이 나지 않았기에 수험전략을 활용한 경우가 많았다. 이와 같은 결과는 지식 측정이라는 목표에 한국사능력검정시험이 부합하는 측면이 있음을 방증하는 결과라 할 수 있다. 추론을 주로 활용했던 S06은 58%의 정답률을 보여 수험전략에 의존했던 참여자에 비해서는 높은 정답률을 나타냈으나, 지식 소환에 의존했던 참여자들 중에서는 가장 하위에 해당하는 정답률을 보였다. 수험전략과 추론을 주로 활용했던 학생들은 2명과 1명으로 그 숫자가 매우 소수여서 이로부터 의미 있는 일반화를 이끌어내기는 어렵지만, 선형 문항 풀이에 지식이 필요하다는 당연시되는 사실을 뒷받침하는 결과이기도 하다. 물론 선

다형 문항을 통해 확인할 수 있는 지식이 어떤 종류이며, 어떤 수준의 것인지에 대해서는 앞선 결과에서 나왔듯이 면밀한 검토가 필요하다 하겠다.

다음으로 10명의 참여자 중 9명이 지식 소환, 수험전략 활용, 추론을 각각 1회 이상 사용하여 문항을 풀이하였다는 점도 주목할 만 하다. 단 1명의 참여자(S07)만이 지식 소환과 수험전략이라는 2가지 방법만을 활용하였다. 한국사능력검정시험 목표 준거에 명시되지 않은 수험전략이나 추론의 활용은 일부 연구 참여자에 집중되어 활용되기보다는 거의 모든 참여자가 활용하는 양상을 보였던 것이다. 나아가 7명의 참여자는 세 가지 방법을 6번, 4번, 2번 활용하거나, 5번, 4번, 3번 활용하여 어느 한 방법에 편중되지 않은 분포를 보이고 있었다. 이처럼 다수 연구 참여자 풀이 과정에서 세 가지 문항 풀이 과정이 고르게 분포하였으나 목표 준거에 명시된 역사적 사고력은 나타나지 않았다는 점에서 한국사능력검정시험 문항의 타당도를 높일 필요성을 지적할 수 있다.

마지막으로 언급할 점은 참여자가 취득한 급수와 12문항의 정답률 사이의 상관관계를 확인하기 어려웠다는 사실이다. 1급을 취득했던 3명의 참여자는 25%(1명)와 67%(2명)의 정답률을 보였고, 3급을 취득했던 6명의 참여자는 25%(1명), 58%(1명), 67%(2명), 75%(1명), 83%(1명)이었다. 급수를 취득하지 못했던 참여자는 58%의 정답률로 10명 참여자의 평균 정답률을 나타냈다. 면담에 활용했던 문항의 정답률이 53%에서 87% 사이에 분포한다는 점에 비추어볼 때, 급수를 취득했던 연구 참여자에게 이보다 높은 정답률을 기대하는 것은 합리적이다. 그러나 연구 참여자들이 보여준 정답률은 이에 미치지 못했다. 급수를 취득한 9명의 참여자의 면담 정답률을 한국사능력검정시험 급수 인정 기준 점수에 비출 경우 여전히 급수를 취득할 수 있는 학생은 S01(3급, 75%)과 S09(3급 83%) 단 두 명뿐이다.

면담에서의 정답률이 기대치에 크게 못 미친 요인으로 한국사능력검정시험 마지막 응시일과 면담일 간의 격차를 가정해 볼 수 있다. 면담일로부터 1년 전에 응시한 시험에서 1급을 취득하였지만, 면담에서는 25%의 정답률을 보여준 참여자(S08)와 면담이 진행된 달에 시험에 응시하여 3급을 취득하였고, 면담에서도 83%로 10명의 참여자 중 가장 높은 정답률을 보여준 참여자(S09)의 사례를 보면 시험 응시일과 면담일 간의 격차가 클수록 지식이 휘발될 가능성이 더 높아 보인다. 그러나 반대의 사례도 존재한다. 예컨대 S03은 면담이 진행된 달에 시험에 응시하여 3급을 취득하였지만, 면담에서는 25%의 정답률을 보여주었고, S01은 2018년 2월에 시험에 응시하여 3급을 취득하고 6개월이 지나 면담에 응했으나 75%의 정답률을 보였다. 결국 면담에서의 정답률과 한국사능력검정시험 응시일 간의 관계도 확인하기 어렵다. 본 연구가 10명이라는 소수의 연구 참여자를 대상으로 한 연구이기에 이와 같은 결과로부터 일반화를 이끌어내는 어렵지만, 참여자들이 문항을 풀면서 공부했던 지식이 기억나지 않는 현상을 여러 차례 언급하였다는 점은 주목할 필요가 있다. 10명 중 8명의 참여자는 시험 응시일로부터 6개월 이내에 면담을 수행했고 나머지 2명 역시 1년 이내에 면담을 수행하여 취득 급

수가 유효한 상태였기에, 한국사능력검정시험이 6개월도 채 지속되지 못하는 휘발성이 강한 지식을 측정하고 있지는 않은지에 대한 합리적인 의문을 제기할 수 있다. 맥락 없이 암기된 지식이기에 휘발성이 강해졌을 수도, 지엽적인 지식이기에 휘발성이 강해졌을 수도 있다. 한국사능력검정시험의 정교화를 위해서는 문항을 통해 측정하고 있는 지식의 특성이 어떠한지 그리고 이에 기초할 때 시험 결과를 어떻게 해석해야 하는가에 대한 추가 연구가 요구된다 하겠다.

V. 논의 및 제언

본 연구에서는 연구 참여자가 어떤 지식과 사고력을 활용하여 한국사능력검정시험 선다형 문항을 풀이하는지를 분석하였다. 연구 참여자는 지식을 소환하거나, 소거법이나 찍기와 같은 수험전략을 활용하고, 주어진 문항의 내용에 기초하여 일반적인 추론을 수행하면서 문항을 풀이하였다. 세 가지 문항 풀이 과정 중 지식 소환이 가장 많이 활용되었다는 측면에서 한국사능력검정시험 선다형 문항이 연구 참여자의 지식을 측정하고 있다는 점을 확인할 수 있었다. 그러나 본 연구의 분석 결과에 기초하여 선다형 문항으로 측정되는 지식이 무엇인가에 대한 여러 질문을 제기할 수 있었다. 지식 소환을 사용한 다수의 문항 풀이 과정에서 연구 참여자는 출제자가 측정을 의도한 지식 없이도 정답을 맞혔다. 즉 전문가가 문항의 전통적인 내용 타당도를 검증하면서 문항을 풀이하는 과정에서와는 상당히 다른 지식을 활용하여 문항을 풀이하고 있었다. 또 일부 연구 참여자는 출제자가 측정을 의도한 지식을 소환하였으나 오답을 선택하기도 했다. 생각말하기가 아닌 일반적인 시험 방식으로 문항을 풀었다면, 전자는 해당 문항에서 측정하는 지식을 가지고 있다고 평가되었을 것이고 후자는 지식이 없거나 부족하다고 평가되었을 것이다. 이는 고등학생을 대상으로 했던 국내 선행연구에서는 보고되지 않았던 현상이다. 한국사능력검정시험이 고부담시험이 아닌 인증시험이라는 점 때문에 문항 풀이에서 차이가 나타났을 수도 있다. 만일 그러하다면, 피험자의 특성을 정교하게 파악하여 문항을 개발할 필요성이 제기된다. 지식 소환과 관련하여 또 한 가지 주목할 사실은 한국사능력검정시험에서 측정하는 지식의 휘발성이다. 시험에 응시하고 2주도 채 지나지 않아 면담을 수행했던 연구 참여자도 문항을 풀면서 기억이 나지 않는다, 잊어 버려서 문항을 풀 수가 없다는 언급을 자주 했다. 시험을 치르는 동시에 휘발되기 시작하는 지식을 측정한다는 것이 어떤 의미가 있는지에 대한 고민이 필요해 보인다.

<표 4> 연구 참여자별 분석 결과

참여자 ID	전공	지식 소환		수험전략		추론		총계		정답률	취득급수
S01	수학교육	6	(50%)	4	(33%)	2	(17%)	12	(100%)	75%	3급
S02	국어교육	5	(42%)	3	(25%)	4	(33%)	12	(100%)	58%	-
S03	국어교육	4	(33%)	6	(50%)	2	(17%)	12	(100%)	25%	3급
S04	수학교육	6	(50%)	2	(17%)	4	(33%)	12	(100%)	67%	1급
S05	국어교육	6	(50%)	4	(33%)	2	(17%)	12	(100%)	67%	3급
S06	수학교육	2	(17%)	4	(33%)	6	(50%)	12	(100%)	58%	3급
S07	국어교육	9	(75%)	3	(25%)	0	(0%)	12	(100%)	67%	1급
S08	수학교육	4	(33%)	5	(42%)	3	(25%)	12	(100%)	25%	1급
S09	수학교육	8	(67%)	1	(8%)	3	(25%)	12	(100%)	83%	3급
S10	국어교육	9	(75%)	2	(17%)	1	(8%)	12	(100%)	67%	3급
총계		59	(49%)	34	(28%)	27	(23%)	120	(100%)	59%	-

※ 회색은 우세하게 나타난 코딩을 의미하며, 사선은 우세한 코딩이 나타나지 않은 경우를 의미한다.

다음으로 한국사능력검정시험은 역사적 사고력을 목표 준거로 표방하고 있으나, 참여자들의 문항 풀이 과정을 분석하였을 때 역사적 사고력으로 코딩될 수 있는 사례는 단 한 건도 발견되지 않았다. 120건의 문항 풀이에서 여러 역사적 자료의 타당성을 판별한 후, 분석 결과를 종합하여 결론을 도출한다거나, 역사 정보의 분석을 통해 시대적 배경과 사회적 의미를 맥락적으로 파악함으로써 답을 도출하는 경우를 확인할 수 없었다. 이러한 결과는 역사학에 고유한 고차원적 인지 과정을 측정하기 위한 방법으로 현행 선다형 문항의 효과에 대한 의구심을 제기한 Reich(2009, 2013), Smith(2017)의 연구 결과와도 일치한다.

그러나 한국사능력검정시험이 현행 선다형 문항을 고집하는 한 역사적 사고와 같은 행동영역을 측정할 수는 없다고 선부르게 단정내릴 수는 없다. 실제 사료를 활용한 문항의 경우 비록 역사적 사고라고 규정하기는 어려우나 추론 과정을 통해 정답을 탐색하는 모습을 보여주었다. 다만 사료를 활용한다고 해서 항상 이러한 효과를 기대할 수 있는 것은 아니다. Smith 역시 역사 수업이나 평가에 사료를 포함하는 것이 역사적 사고를 유발할 수 있는 일종의 ‘품질보증마크(hallmark)’로 간주되곤 하나, 사료를 포함했다는 것만으로 역사적 사고를 끌어내는 불충분하다고 지적한 바 있다(Smith, 2017). 6번 사화 문항처럼 논리적 연관성이 없이 발생한 역사적 사건의 순서를 묻는 문항에 사료를 활용하는 것은 큰 의미가 없을 수도 있다.

연구 분석 결과에 기초해 볼 때, 한국사능력검정시험이 나아갈 방향을 다음 3가지로 제안할 수 있다. 첫째, 선다형 평가 문항의 장점인 효율적인 지식 측정의 정교화이다. 이를 위해 응시자가 수험전략이나 추론과 같이 평가에서 의도하지 않은 인지 능력을 활용하는 것을 최소화할 방안이 필요하다. 물론 선다형 평가에서 수험전략이나 읽기와 연관된 추론을 원천적으로 배제하기는 불가능하다. 그러나 본 연구에서 확인했던 익숙도의 활용을 방지할 필요는

있다. 익숙한 답지를 답안으로 채택하는 수험전략에 대응하기 위해서는 새로운 내용 주제와 새로운 답지를 활용할 필요가 있다. 또한 시대 순서에 따른 문항 배치도 재고려할 필요가 있다. 문항의 번호를 통해 어느 시대 문항일 것임을 추론하고 문항을 푸는 것은 정확한 지식 측정을 어렵게 하고 있었다.

둘째, 역사적 사고력의 측정을 목표로 한다면, 역사적 사고력을 측정할 수 있는 문항 개발에 노력을 기울여야 할 것이다. 선행연구에서 지속적으로 지적되었고, 본 연구에서도 확인했듯이 현행 선다형 문항은 피험자의 역사적 사고력을 측정하고 있다고 보기 어렵다. 그렇지만 선다형 문항을 통해 역사적 사고력의 측정이 불가능한 것은 아니다. Seixas, Gibson, & Ercikan은 역사적 사고를 측정할 때 잘 고안된 선다형 문항이 도움이 될 수도 있다고 주장한다(Seixas, Gibson, & Ercikan, 2015). 스텐포드 역사 그룹(SHEG) 또한 역사사고평가(HATs)라고 불리는 문항들을 개발한 바 있다. HATs는 원사료를 제공한 후 서답형(2-3문장)으로 작성된 답을 요구하였다. 문두와 자료가 동일한 선다형 문항 역시 개발하여 두 유형의 타당도를 측정한 결과 HATs는 물론 HATs를 변형한 선다형 문항도 역사적 사고를 측정하는 데 타당한 것으로 나타났다. 다만 선다형 문항은 HATs에 비하여 문항 출제자의 의도와 관련 없는 추론이 나타나는 사례가 더 많았다(Smith, Breakstone, & Wineburg, 2019). 전통적인 선다형 문항에 대해 비판적인 반슬레드라이트는 가중 선다형(weighted multiple-choice)이라는 대안적인 문항 형식을 제안하기도 했다. 일반적인 선다형 문항은 정답지가 하나이지만, 가중 선다형 문항은 답지 가운데 오답(0점)은 하나이며, 나머지 답지는 모두 답이 될 수 있으나 점수가중치를 다르게 설정할 수 있다. 예를 들어 가장 설득력 있는 답지는 4점 혹은 3점, 두 번째로 설득력 있는 답지는 2점 등을 부여하는 것이다. 그는 가중 선다형 문항이 전통적인 선다형 문항보다 역사적 사고를 측정하는 데 이점을 줄 수 있다고 주장했다(VanSledright, 2014; 2015). 역사적 사고 측정에서 선다형 문항의 가능성은 추가적인 연구를 필요로 한다.

셋째, 평가 결과를 해석하는 방법을 보강할 필요가 있다. 1개 문항에 정답을 제시했다는 것, 특정 급수를 취득했다는 것과 같은 시험 결과가 무엇을 의미하는지를 명확하게 소통할 필요가 있다. 또한 문항 타당도를 높이기 위해서는 평가 결과의 지속적인 해석이 필요하다. 선다형 평가 문항의 실제적인 효율성과 객관성, 또는 대중과 정책입안자의 인식 속에서 광범위하게 공유되고 있는 선다형 평가의 효율성과 객관성 때문이라도 역사 이해의 평가에 있어 선다형 평가를 완전히 배제하기는 힘들 것이다. 그러나 평가 결과를 지속적으로 분석, 해석, 연구함으로써 문항의 타당도를 높이려는 노력을 기울여야 할 것이다.

참 고 문 헌

- 국사편찬위원회(n.d.) 한국사능력검정시험 웹사이트. <http://www.historyexam.go.kr/pageLink.do?link=examInfo>(검색일: 2019. 6. 28.)
- 김덕진(2012). 한국사능력검정시험을 통해 본 초등학생의 역사의식. **역사학연구**, 47, 1-27.
- 김미선(2011). 「한국사능력검정시험 등급별 평가 수준 위계화 방안」, **역사교육** 118, (2011), 1-49.
- 김민정, 김미선(2011). 한국사능력검정시험 평가 문항 현황과 개선 방안. **역사교육논집**, 47, 463-495.
- 김수미(2007). 제1·2회 한국사능력검정시험의 분석과 문항의 방향. **역사교육연구**, 6, 177-227.
- 김수미(2014). 학생의 선다형 국사 문항 풀이 과정 분석. **역사교육**, 129, 139-195.
- 김정수(2011). 한국사능력검정시험의 목표 준거와 문항 반응 분석-제1,2회 3급 문항을 중심으로-. **역사와 교육**, 12, 69-134.
- 박덕우(2010). 제1-8회 한국사능력검정시험 초급 문항 분석: 초등학교 교과서와의 비교분석을 중심으로. 석사학위논문. 서울시립대학교.
- 박진동(2011). 고등학생의 선다형 문항 풀이 과정과 시험전략 고찰. **역사교육**, 117, 195-117.
- 박진동, 이미미, 김민정, 오정현, 방대광, 남궁원, 김광규, 류성완. (2014). **2017학년도 수능 한국사 필수화에 따른 교사 연수 자료**. 한국교육과정평가원. 수능 CAT 2014-15.
- 박진철(2016). 人物史 教育을 위한 ‘한국사능력검정시험’의 分析 研究. **한국학연구**, 58, 227-257.
- 오정현(2009). 역사과 평가 문항의 위계성 검토. **역사교육연구**, 9, 97-144.
- 오정현, 박진동(2015). 수능 필수 한국사의 평가목표 설정과 적용방안, **역사교육**, 135, 31-68.
- 은동진(2015). 한국사능력검정시험의 중급과 고급 간 위계성에 대한 검토. 석사학위논문. 서강대학교.
- 인사혁신처(2018.12.11). 2021년 7급 공채시험에 공직적격성평가(PSAT)도입-2018.12.11.(화), 공무원임용시험령 개정(안) 국무회의 통과-. http://www.mpm.go.kr/board/board.do?boardId=bbs_0000000000000029&mode=view&cntId=2661(검색일 2019. 5.20.)
- 최상훈(2015). 수능 필수 한국사의 내용타당도 제고 방안. **역사교육**, 135, 1-29.
- 최상훈, 김미선, 김수미, 방지원, 오정현, 신향수, 박진동, 김민정(2012). **역사과평가의 이론과 실제**. 서울: 책과함께.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education(1999). *Standards for educational and psychological testing*. Washington D.C.: American Educational Research Association.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education(2014). *Standards for educational and psychological testing*. Washington D.C.: American Educational Research Association.
- Blackey, R. (2009). So many choices, so little time: Setting strategies for understanding and taking multiple-choice exams in history, *The History Teacher*, 43(1), 53-66.
- Braden, J. (2009). Validity. In E.M. Anderman, & L.H. Anderman (Eds.), *Psychology of classroom learning: An encyclopedia*. Vol. 2. (pp. 961-965), Deteroit, MI: Macmillan Reference USA.
- Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ded.). Hahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' reponses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- National Assessment Governing Board. (2011). *U.S. history framework for the 2010 National Assessment of Educational Progress*. Retrieved from <https://www.nagb.gov/assets/documents/publications/frameworks/history/2010-history-framework.pdf>
- Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Los Angeles, CA: Sage.
- Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington D.C.: National Academy Press. 2001 [kindle book]
- Reich, G. A. (2009). Testing historical knowledge: Standards, multiple-choice questions and student reasoning. *Theory and Research in Social Education*, 37(3), 325-360.
- Reich, G. A. (2013). Imperfect models, imperfect conclusions: An exploratory study of multiple-choice tests and historical knowledge. *The Journal of Social Studies Research*, 37, 3-16.
- Seixas, P., Gibson, L., & Ercikan, K. (2015). A design process for assessing historical thinking: The case of a one-hour test. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 102 - 116). New York, NY: Routledge.
- Smith, M. D. (2017). Cognitive validity: Can multiple-choice items tap historical thinking processes? *American Educational Research Journal*, 54(6), 1256 - 1287.

- Smith, M. D. (2018). New multiple-choice measures of historical thinking: An investigation of cognitive validity. *Theory & Research in Social Education*, 46(1), 1-34.
- Smith, M., Breakstone, J. & Wineburg, S. (2019). History assessments of thinking: A validity study, *Cognition and Instruction*, 37(1), 118-144.
- VanSledright, B. (2014). *Assessing historical thinking and understanding : Innovative designs for new standards*. New York, NY: Taylor & Francis.
- VanSledright, B. (2015). Assessing for learning in the history classroom. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 102 - 116). New York, NY: Routledge

· 논문접수 : 2019.07.04. / 수정본접수 : 2019.08.12. / 게재승인 : 2019.08.14.

ABSTRACT

Analysis of Students' Response Processes in Proficiency Test of Korean History Items

Mimi Lee

Assistant Professor, Hongik University

Sunny Hong

Research Fellow, National Institute of Korean History

This study analyzed how 10 college students solved twelve multiple-choice items from the Proficiency Test of Korean History(PTKS) to elicit the knowledge and thinking that students employed during the response processes. We analyzed how the knowledge and thinking that participants used in solving the multiple-choice items aligned with the knowledge and thinking intended for testing. The analysis results revealed that participants were engaged in three response processes in approaching and solving the test items: fact recall in 59 cases(49%), test taking skills in 34 cases(28%), and general reasoning in 27 cases(23%). Although test taking skills and general reasoning are construct-irrelevant variances, participants used them widely with all participants employing test taking skills and 9 out of 10 participants using general reasoning. Historical thinking skills specified in test purposes, however, did not surface in response processes, which raises the question of validity. The fact that participants used fact recall the most of the three response processes solidifies the claim that the PTKH assesses test takers' historical knowledge, but the analysis of the kinds and levels of knowledge used in participants' response processes called for the need to strengthen the validity of PTKH items. Cross case analysis of response processes across items and across participants also pointed to the need to strengthen the validity of the test items. Based on the analysis results, we discussed three implications of the study for enhancing the PTKH items.

Key Words: Proficiency Test of Korean History, Multiple-choice Test, Response Processes, Historical Knowledge, Historical Thinking

부록: 면담에 활용한 문항

Q01

1. (가) 시대에 대한 설명으로 옳은 것은? [1점]

이곳은 제주도 고산리 유적 발굴 현장입니다. 이 유적의 최하층에서 이른 민무늬 토기가 출토됨에 따라 (가) 시대가 기원전 8000년경부터 시작되었음을 알게 되었습니다. 이 외에도 화살촉, 갈돌, 갈판 등의 석기가 출토되었습니다.



- ① 널무덤과 독무덤을 만들었다.
- ② 권력을 가진 군장이 백성을 다스렸다.
- ③ 반량전, 명도전 등의 화폐를 사용하였다.
- ④ 정착 생활이 시작되면서 읍집이 나타났다.
- ⑤ 우경이 시작되어 깊이갈이가 가능해졌다.

Q03

8. 다음 가상 인터뷰의 (가)에 들어갈 대답으로 적절한 것은? [2점]

[인터뷰]

국왕과의 대화

기자: 인재를 동용하기 위해 각별한 노력을 기울이셨다고 들었습니다. 그렇게 하신 특별한 이유가 있으신가요?

왕 : 반란을 진압한 후, 정치를 바로잡기 위해서는 노력 있는 인재를 필요하다는 것을 절감했습니다. 그래서 널리 인재를 추천받아 안우와 음파소를 발탁하였습니다.

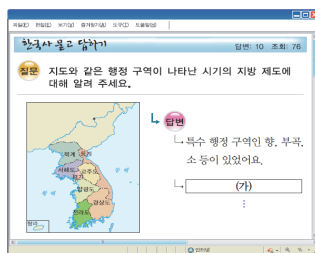
기자: 한편 민생 안정을 위한 세로로 제도도 마련하였는데요, 자세한 설명 부탁드립니다.

왕 : 사냥을 가는 길에 울고 있는 백성의 사연을 듣고 신하들에게 명령을 내려 (가)

- ① 병자에게 약을 지급하는 해민국을 설치하였습니다.
- ② 백성들에게 곡식을 빌려주는 진대법을 마련하였습니다.
- ③ 호구를 정확하게 파악하기 위한 호폐법을 시행하였습니다.
- ④ 강제로 노비가 된 자를 해방시키는 노비안검법을 실시하였습니다.
- ⑤ 거금을 모아 그 아자로 빈민을 구제하는 제위포를 운영하였습니다.

Q05

17. (가)에 들어갈 내용으로 옳은 것은? [2점]



- ① 경제소를 설치하여 유품소를 통제하였어요.
- ② 전국의 모든 군현에 지방관을 파견하였어요.
- ③ 상수리 제도를 실시하여 지방 세력을 견제하였어요.
- ④ 5소경을 설치하여 수도의 편제성을 보완하고자 하였어요.
- ⑤ 국경時代に 병마사를 파견하여 적의 침입에 대비하였어요.

Q02

8. (가)에 들어갈 불상으로 옳은 것은? [1점]

이 불상은 국보 제83호로 머리에 삼삼관을 쓰고 있다. 뚜렷하게 조각된 눈·코·입, 균형 잡힌 신체, 입체적으로 처리된 옷주름 등이 뛰어난 조각 수법으로 표현되어 있다.

일본 교토 사(興隆寺)에 있는 불상과 모습이 매우 비슷하여 일본에 전파된 우리 고대 문화의 흔적을 보여준다.



Q04

14. 다음 역사서에 대한 설명으로 옳은 것은? [2점]

구삼국사(舊三國史)를 얻어 동명왕본기(東明王本紀)를 보니 그 신이한 사적이 세상에 전하는 것보다 더하였다. 그러나 처음에는 믿지 못해 귀환(鬼幻)으로만 여겼는데, 세 번 반복하여 읽어서 점점 그 근원에 들어가니, 환(幻)이 아니고 성(聖)이며 귀(鬼)가 아니고 신(神)이었다. …… 이것을 기술하지 않으면 후인들이 장차 무엇을 볼 것인가.

—「동명왕원」—

- ① 고구려 계승 의식이 반영되었다.
- ② 현존하는 가장 오래된 역사서이다.
- ③ 남북국이라는 용어를 처음 사용하였다.
- ④ 연대순으로 기록하는 편년체로 서술되었다.
- ⑤ 단군 조선부터 고려까지의 역사를 기록하였다.

Q06

23. (가), (나) 사이의 시기에 있었던 사실로 옳은 것은? [2점]

(가) 생도 윤씨를 폐비하는 의논에 참여한 자와 (어머니에게) 존호(尊號)를 올려서는 안 된다고 주장한 자를 모두 중형으로 다스려, 죽은 자는 그 시체를 베고 가산을 몰수하였으며, 그 가족이나 친족은 연좌하였다. 살아 있는 자는 매[拷]로 때리며 심문한 후 벌리 귀양 보냈다.

(나) 유인숙은 윤임과 혼인을 맺고 음모를 꾸미며 속으로 권세를 잃을까 근심하였다. 내가 죽자마자 자기에게 불리하다고 여기고서 물레 사부(側府)를 불러다가 나의 현부(賢否)를 물었으며, 이것은 모두 배반을 떠돌아 끌고 자기의 욕망을 이루려고 회한 것이니 죄가 중해서 관판되어 범으로 용서할 수가 없다.

이에 윤임·유관·유인숙 세 사람에게 사지(賜死)를 명한다.

- ① 서인이 반정을 일으켜 정권을 장악하였다.
- ② 위훈 사재를 주장한 조광조 일파가 축출되었다.
- ③ 정여립 모반 사건을 계기로 동인이 피해를 입었다.
- ④ 효종이 죽자 자의대비의 복상 문제로 예송이 전개되었다.
- ⑤ 사림이 이조전랑 임명을 둘러싸고 동인과 서인으로 나뉘었다.

Q07

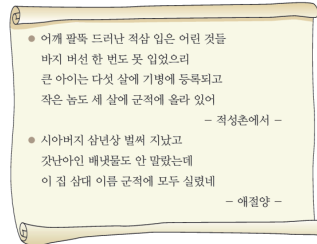
22. (가)~(다) 화폐를 처음 발행된 순서대로 옳게 나열한 것은? [1점]



- ① (가) - (나) - (다) ② (가) - (다) - (나)
 ③ (나) - (가) - (다) ④ (나) - (다) - (가)
 ⑤ (다) - (나) - (가)

Q09

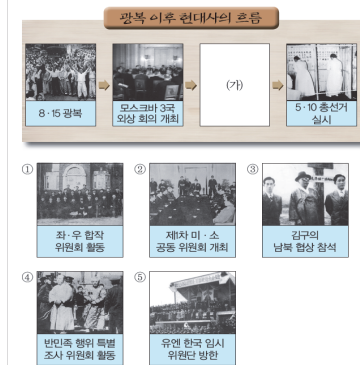
35. 다음 자료에 나타난 문제점을 해결하기 위해 흥선 대원군이 실시한 정책으로 옳은 것은? [1점]



- ① 호포법을 제정하여 군포를 징수하였다.
 ② 5군에서 2영으로 군제를 개편하였다.
 ③ 양전 사업을 실시하여 지계를 발급하였다.
 ④ 속대전을 편찬하여 국가 체제를 정비하였다.
 ⑤ 초계문신제를 실시하여 관리를 재교육하였다.

Q11

46. (가)에 들어갈 사진으로 옳지 않은 것은? [3점]



Q08

22. 밑줄 그은 '이 법'에 대한 설명으로 옳은 것은? [2점]



- ① 양반에게도 군포를 징수하였다.
 ② 결작을 부과하는 계기가 되었다.
 ③ 1결당 쌀 4~6두로 납부액을 고정시켰다.
 ④ 육의전을 제외한 시전 상인의 특권을 폐지하였다.
 ⑤ 관청에 물품을 조달하는 공인이 등장하는 배경이 되었다.

Q10

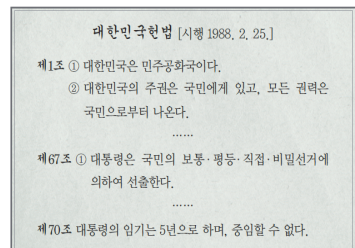
44. 다음 의거를 일으킨 단체에 대한 설명으로 옳은 것을 <보기>에서 고른 것은? [2점]



- ① 가, 나 ② 가, 다 ③ 나, 다
 ④ 나, 라 ⑤ 다, 라

Q12

48. 다음과 같이 헌법이 개정된 계기로 가장 적절한 것은? [3점]



- ① 제헌 국회가 출범하였다.
 ② 장면을 총리로 하는 내각이 들어섰다.
 ③ 대통령 특별 선언으로 10월 유신이 선포되었다.
 ④ 3·15 부정 선거에 항의하는 시위가 확산되었다.
 ⑤ 호헌 철폐와 독재 타도를 외친 6월 민주 항쟁이 일어났다.