

논증 기반 타당도 접근에 기반한 수능 작문 평가의 개선 방향 탐색

장성민 서울대학교 국어교육과 박사과정 수료

- I. 서론
- II. 논증 기반 타당도 접근의 이론적 배경
- III. 수능 작문 평가에 대한 해석적 논증의 구성
- IV. 수능 작문 평가의 개선 방향 탐색
- V. 결론

I. 서론

이 연구는 국내의 수능 작문 평가 개선을 위해 논증 기반 타당도 접근(the argument-based approach to validation)의 관점에서 작문 평가를 둘러싼 담론의 해석적 논증(interpretative arguments)을 점검하고, 이를 바탕으로 타당도를 갖춘 수능 작문 평가의 대안을 제안하는 데 목적이 있다.

기존의 대학수학능력시험 작문 평가의 간접 측정 방식에 대해서는 지속적으로 문제제기가 이루어져 왔다(장성민·민병곤, 2016 등). 최근에는 대학수학능력시험에 서술형, 논술형 문항의 반영을 적극 검토하는 방안에 대해서도 논의가 이루어지고 있다. 실제로 우리와 유사한 교육 환경을 가지고 있는 일본의 경우에는 2020년부터 대입 작문 평가(대학입학공통시험)에 짧은 분량의 서술형 문항을 도입하기로 발표하고, 2017년 5월 예시 문항을 제시하기도 하였다.

그럼에도 불구하고 작문 평가에서 객관식의 선다형 문항을 완전히 배제하는 것이 가능한가에 대해서는 논의가 더 필요해 보인다. 대학 입시에 대한 높은 수준의 사회적 관심과 이에 따른 엄청난 시간과 비용의 문제, 공정성과

형평성 확보의 어려움 등으로 말미암아 서술형, 논술형 문항만으로 작문 능력을 측정하는 것이 현실적으로 쉽지 않기 때문이다. 실제로 오랜 전통을 가지고 전 과목의 학업성취도를 논술형 문항으로 측정하고 있는 프랑스의 바칼로레아의 경우에도 그 변별성과 사회적 효용에 대해서는 문제 제기가 지속적으로 이루어지고 있으며, 2016년 개편한 미국의 SAT의 경우에는 논술형의 에세이 평가 외에 증거 기반 읽기·쓰기(evidence-based reading & writing)의 선다형 문항을 병행하여 활용하는 등 선다형 문항을 일괄적으로 배제하고 있지는 않다.

이 연구의 주요한 관심사는 간접 측정 방식에 의한 기존 작문 평가의 타당성을 높이는 데 초점이 있다. 작문 능력에 대한 직접 측정의 중요성에 대해서는 당위론적으로 이견의 여지가 없으나, 현실적으로 간접 측정 방식을 완전히 폐기할 수 없다면 그 타당성을 확보하기 위한 방안을 모색하는 것도 작문 교수·학습 평가 담론의 또 하나의 중요한 책무가 될 것이기 때문이다. “선다형 평가 자체가 암기식 교육을 유도하는 잘못된 평가라는 생각은 서술식, 자유기술식만이 올바른 평가라는 고정관념에 가깝다. 어떤 의미에서 현수능과 같은 선다형 평가는 매우 고차원적인 사고 능력을 평가하는 교과 적 성 검사로 볼 수 있다.”(권태현, 2018: 77)라는 선행 담론의 지적은 대학수학 능력시험의 작문 평가가 지금까지 유지되어 온 논리를 대변하는 것이기도 하다.

이 연구는 수능 작문 평가의 타당성을 이론적으로 확보하기 위해 논증 기반 타당도 접근(the argument-based approach to validation)에 주목하였다. 툴민(Toulmin, 1958/2003)의 논증 구조에 기반을 둔 일련의 평가 논증(assessment argument) 담론은 ⑦ 검사 점수의 해석과 사용에 내포된 광범위한 문제에 관심을 가지고, ⑧ 이에 수반되는 이론적, 이데올로기적 가정의 전반적인 역할에 주목하며, ⑨ 대안적 접근의 가능성에 대한 고려를 포함한다는 장점이 있다(Kane, 2006). 이는 구인 타당도를 중심으로 준거 타당도와 내용 타당도의 문제를 포괄하고자 한 교육 측정 담론의 통합적 경향과도

맥락을 공유한다(Messick, 1989). 논증 기반 타당도 접근은 구체적인 증거에 기반한 추론(evidentiary reasoning)과 대안적 접근에 대한 고려를 바탕으로 기존의 수능 작문 평가가 지닌 타당성의 문제를 점검하고, 중등 학습자의 작문 능력과 측정의 목적에 부합하는 대안적 평가 방식의 의미 있는 정보들을 도출하는 데 기여할 수 있을 것으로 기대된다. 케인(Kane, 2013)의 “해석적 논증(interpretative arguments)”과 미스레비(Mislevy et al., 2003)의 “증거 중심 설계(evidence-centered design)”는 그러한 접근이 반영된 담론의 대표적인 예이다.

이상의 문제의식을 바탕으로 이 연구에서는 다음과 같이 두 가지 연구 문제를 설정하였다.

첫째, 수능 작문 평가에 대한 해석적 논증은 어떻게 구성되는가?

둘째, 수능 작문 평가에 대한 해석적 논증의 결과로부터 어떤 시사점을 도출할 수 있는가?

II. 논증 기반 타당도 접근의 이론적 배경

논증 기반 타당도 접근은 미국의 심리 측정학자 케인(Kane, 1992)에 의해 처음으로 제안되었다. 케인은 검사 점수를 해석하는 것의 타당성이 논증의 명료성(clarity), 일관성(coherence), 개연성(plausibility)에 대한 판단으로부터 획득되는 것으로 규정하고, 해석적 논증(interpretative argument)과 타당도 논증(validity argument)에 대한 설명을 통해 이를 뒷받침하였다. 해석적 논증은 검사 점수의 해석과 사용에 관여하는 명시적·암묵적 추론(reasoning)을 총칭하는 개념이다. 케인은 해석적 논증이 검사 점수로부터 어떤 결론이나 결정을 도출하기까지에 이르는 일련의 추론(inferences)과 가정

(assumptions)으로 구성되며, 각각의 추론과 가정은 다시 경험적 증거로부터 정당화되는 타당도 논증을 통해 수용되는 것으로 설명하였다.

논증 기반 타당도 접근은 ‘주장(claim), 자료(data), 전제(warrant), 보증(backing), 반박(rebuttal)’ 등으로 구성되는 툴민(Toulmin, 1958/2003)의 논증 구조에 이론적으로 기대고 있다.¹⁾ 툴민에 따르면, 논증적 추론이 성립하기 위해서는 최소한 ‘자료(D)’와 ‘전제(W)’로부터 ‘주장(C)’을 도출하는 형식을 갖추어야 한다. 주장을 우리가 정당화하고자 하는 결론을 지칭하는 개념으로, 개인의 해석적 논증에 대응한다. 자료는 주장을 뒷받침하기 위해 동원되는 사실이나 정보를 지칭하는 개념으로, 개인의 타당도 논증에 대응한다. 전제는 특정한 자료로부터 어떤 특수한 주장에 도달하는 단계를 확정하는 추론의 규칙을 지칭하는 개념으로, 보증(B)의 지원을 받아 정당화됨으로써 보편적으로 받아들일 수 있는(generally acceptable) 규칙으로 통용된다.

툴민의 이론에서 눈여겨볼 지점은, 논증적 추론이 적용되지 않는 예외적 조건으로서 ‘반박(R)’을 설정하고 있다는 점이다. 반박은 자료로부터 주장으로 이어지는 추론의 과정이 적용되지 않는 대안적 명제(counterclaim)을 지칭하는 개념으로, 이는 대안적 명제의 영향력을 지지하거나 약화시키는 일종의 근거로서 반박 자료(rebuttal data)의 뒷받침을 통해 다시 별개의 논증 구조로 확장될 수 있다(Mislevy et al., 2003). 논증 기반 타당도 접근의 관점에서는 평가의 타당성을 확보하기 위해 대안적 관점을 파악하고 그 개연성을 따져 보는 절차가 필수적으로 수반되어야 한다고 지적하였다. 검사의 점수를 해석하는 것이 그것이 사용되는 맥락에 관계 없이 고정되어 있거나 반드시 당위론적으로 수용되어야 하는 절대적인 것은 아니기 때문이다.

1) 연구자에 따라 claim을 ‘결론’(민병곤, 2000), ‘명제’(장성민, 2016) 등으로 번역하기도 하고, warrant를 ‘보증’(장혜영 역, 2008), backing을 ‘지원’(민병곤, 2000)으로 번역하기도 한다. 이 연구에서는 윌리엄스와 콜럼(Williams & Colomb, 2007)을 번역한 윤영삼 역(2008)의 번역을 따랐다.

선행 담론(Bachman, 2005)에서 평가 사용 논증(assessment utilization argument)의 주요한 전제들로 제시한 ‘관련성(relevance), 유용성(utility), 의도된 결과(intended consequences), 충분성(sufficiency)’ 등은 상이한 관점의 개연성을 판단하기 위해 고려되어야 하는 준거의 목록으로 볼 수 있다.

케인(Kane, 2006)은 ‘채점, 일반화, 외삽, 결정’의 네 가지 개념을 통해 해석적 논증에 수반되는 추론의 과정을 설명하였다. 그는 이 네 가지 추론의 조건이 모두 충족된 경우에만 해석적 논증을 정당화할 수 있으며, 이 가운데 어느 하나라도 충족되지 못할 경우 전체 논증의 타당도를 확보하기 어렵다고 지적하였다.

첫째, ‘채점(scoring)’은 평가(검사)를 통해 관찰된 수행(observed performance)으로부터 관찰 점수(observed score)로 이어지는 추론의 관계에 주목하는 것이다. 이는 채점 준거가 피검사자의 수행을 측정하는 데 적합해야 하며, 정확하고 일관된 방식으로 적용되어야 한다는 가정을 전제로 한다. 채점의 추론에서 대안적 관점과의 논쟁을 촉발하는 지점은 검사가 어떻게 운영되었는가, 피검사자의 반응을 어떻게 수치화하였는가, 어떤 준거를 채점에 활용하였는가 등에 관한 증거를 수집함으로써 파악된다.

둘째, ‘일반화(generalization)’는 관찰 점수로부터 전집 점수(universe score)로 이어지는 추론의 관계에 주목하는 것이다. 이는 관찰된 표집의 결과가 모집단을 대표할 수 있어야 하며, 그 크기가 무선판 오류를 통제할 만큼 충분히 커야 한다는 가정을 전제로 한다. 일반화의 추론에서 대안적 관점과의 논쟁을 촉발하는 지점은 신뢰도(reliability)나 일반화 가능성(generalizability)에 관한 증거들을 검토함으로써 파악된다. 이는 궁극적으로 검사가 이루어지는 특정한 시공간이나 검사의 형식, 평정자 등에 관계없이 상이한 맥락에서도 동일하게 적용될 수 있는가의 문제로 귀결된다.

셋째, ‘외삽(extrapolation)’은 전집 점수로부터 목표 점수(target score)로 이어지는 추론의 관계에 주목하는 것이다. 이는 검사 과제가 요구되는 목표 능력을 충분히 측정하며, 무관련 분산 요인의 영향이 크지 않다는 가정을

전제로 한다. 외삽의 추론에서 대안적 관점과의 논쟁을 촉발하는 지점은 검사 결과와 목표 능력 사이의 관계에 주목하는 경험적 증거들을 탐색함으로써 파악된다.

마지막으로 ‘결정(decision)’은 목표 점수에 대한 소기의 결론으로부터 검사 점수의 사용(the uses made of scores)으로 이어지는 추론의 관계에 주목하는 것이다. 이는 검사의 결과나 가치에 대한 가정을 전제로 한다. 결정의 추론에서 대안적 관점과의 논쟁을 촉발하는 지점은 어떤 결정으로 말미암아 발생하는 긍정적, 부정적 결과나 상이한 결정에 따른 가시적 증거의 비교 등을 통해 파악될 수 있다.

국외의 선행 연구에서 논증 기반 타당도 접근은 언어 평가를 설계하는데 자주 활용되어 왔다(Bachman, 2005; Chapelle et al., 2010; Kane, 2011; Mislevy et al., 2002). 이는 언어 능력을 측정하는 것이 “특정한 범위의 상황에서 자극이나 과제에 반응하여 행동 또는 수행하는 방식에 관한 기질”(Kane, 2006)로서 특성(trait)을 관찰하고 그 결과를 해석, 사용하는 일련의 절차에 전형적으로 부합했기 때문인 것으로 추정할 수 있다. 따라서 논증 기반 타당도 접근에 따라 해석적 논증의 과정을 거치는 것은, 기존의 수능 작문 평가에서 일반적으로 통용되어 온 수행의 기준(performance standard)이 예비 대학생으로서 중등 학습자가 갖추어야 할 것으로 기대하는 작문 능력을 타당하게 측정하는지 체계적으로 점검하도록 도울 수 있다. 논증 기반 타당도 접근이 지향하는 증거 기반 추론과 대안적 관점의 고려는 작문 교육 담론의 이론적 경향에 비추어 현행 수능 작문 평가의 문제점이 무엇이며, 이를 극복하기 위해 어떤 요소들이 대안적으로 더 고려되어야 하는지 탐색하도록 지원할 수 있을 것이다.

III. 수능 작문 평가에 대한 해석적 논증의 구성

이 장에서는 II장의 논의를 바탕으로 ‘채점, 일반화, 외삽, 결정’의 네 범주에서 수능 작문 평가에 대한 해석적 논증을 구성함으로써 개선의 시사점을 도출하기 위한 이론적 토대를 마련하고자 한다.²⁾ 각각의 범주에 대한 추론은 이를 뒷받침하는 가정을 충족해야 하며, 일련의 추론과 가정은 대안적 관점에 대한 고려를 포함하여 경험적 증거로부터 정당화되어야 한다.

첫째, 채점의 추론에 따라 수능 작문 평가에 대한 해석적 논증을 구성하기 위해서는 ‘수능 작문 평가 문항이 정상적인 고등학교 교육과정을 이수한 중등 학습자의 작문 능력을 측정한다.’라는 가정을 충족해야 한다. 대학수학능력시험은 고등학교 교육과정의 내용과 수준에 맞추어 출제하도록 규정되기 때문에 수능 작문 평가 또한 교육과정에서 의도하는 작문 능력의 제 측면을 측정할 수 있는 방향으로 이루어져야 한다. 하지만 이러한 가정에 따라 기존의 수능 작문 평가를 되짚어 보면, ‘지식, 기능, 태도’의 세 범주를 균형 있게 신장시키고자 하는 교육과정의 의도와 달리 기존의 수능 작문 평가는 기능을 측정하는 데 편중되어 왔다. 조금씩 개선하려는 움직임이 학술 담론(최숙기, 2018 등) 및 수능 문항의 실제 출제 경향 모두에서 부분적으로 감지되고는 있으나, 대체로는 ‘계획하기, 내용 생성하기, 내용 조직하기, 표현하기, 고쳐쓰기’로 이어지는 전통적인 인지주의 작문 이론의 쓰기 과정을 관습적으로 연결하던 것으로부터 크게 벗어나지는 못하는 실정이다.

2) 케인은 이 네 개의 범주 외에 목표 점수로부터 이론에 의해 정의된 구인(construct as defined by the theory)으로 이어지는 관계에 주목하는 ‘이론 기반 추론(theory-based inferences)’를 추가로 설정하기도 하고, ‘결정’을 대신하여 구인의 명칭(label)이나 기술(description)의 문제와 관계되는 ‘함의(implication)’를 설정하기도 하였다. 다만 이론과의 관련이나 구인의 명칭, 기술을 다루는 것이 별도의 독립된 논의로 확장될 필요가 있다고 보고, 이 연구에서는 그러한 문제에 대해 주목하지 않았다.

이 연구는 고등학교 교육과정에서 의도하는 작문의 지식, 기능, 태도를 두루 측정할 수 있는 형태로 수능 작문 평가를 개선함으로써 채점 추론의 가정을 충족시킬 수 있다고 판단하였다. 수행으로 직접 연결되는 작문 기능과 별개로 작문 지식을 가르쳐야 하는가에 대해서는 연구자에 따라 의견을 달리할 수 있으나, ‘기능적 지식(functional knowledge)’과 ‘과정적 지식(procedural knowledge)’을 나눈 선행 담론의 논의들을 염두에 둔다면 방법적 지식을 아는 것 외에 사실적, 개념적 지식을 아는 것도 일종의 작문 능력을 갖추는 것으로 규정할 수 있다.³⁾ “일반적 국어 활동에서의 기능적 지식은 투입 요소인 것에 비해 메타 국어 활동에서의 기능적 지식은 수행의 산출물”이라고 설명한 박종훈(2005: 488)의 지적은 이를 뒷받침하는 대표적인 예이다. 설명 텍스트(expository text), 서사 텍스트(narrative text)의 전형적인 상위 구조로서 ‘논증’과 ‘이야기’에 관한 장르 지식의 중요성을 강조하는 것(van Dijk, 1995)이나, 2016년 개편된 미국의 대학 입학 시험(SAT)에 다른 필자가 쓴 글의 수사적 특성을 분석하도록 하는 메타적 쓰기의 에세이 과제를 부과한 것은 작문 능력에서 사실적, 개념적 지식이 갖는 위상을 짐작케 한다.

최근에는 하나 이상의 자료를 동시에 읽고 이를 분석하여 필자의 관점을 수사적으로 뒷받침하는 ‘자료 통합적 글쓰기(writing from sources)’에 대한 사회적 수요가 늘어나면서 인용, 환연 등의 텍스트 차용(text borrowing)에 수반되는 쓰기 윤리의 문제도 교육적으로 대두가 되고 있다(강민경, 2013; 최숙기, 2018 등). 태도를 수능 작문 평가의 문항으로 설계하는 것은 얼핏 보기에도 파편화된 명제적 지식을 형식적으로 측정하는 것처럼 여겨질

3) 기능적 지식은 “상황 맥락에 따른 언어 표현과 관련된 지식”으로서 기본적으로 명제적 지식의 성격을 지닌다(논증의 구성 요소). 기능적 지식에 방법적, 조건적 지식의 성격이 없는 것은, 방법적, 조건적 지식이 본질적으로 수행과 관련된 범주이기 때문이다. 반면 과정적 지식은 “특정 텍스트를 산출하거나 이해하기까지의 과정을 잘 수행하는 방법에 관한 지식”을 지칭한다. 여기에는 명제적, 방법적, 조건적 지식의 성격이 모두 존재한다(논증을 구성하는 방법, 특정한 논증 요소가 활용되는 조건과 그 이유). 이와 관련된 보다 상세한 설명은 박종훈(2005)을 참고할 수 있다.

수 있다. 하지만 수행의 속성이 부각되는 작문 영역에서 태도는 지식, 기능 등 작문 능력을 구성하는 다른 범주와 배타적으로 구분되는 독립된 지점으로 존재하지 않는 경우가 많다. 예컨대 쓰기 윤리의 경우 타인의 아이디어를 무단으로 도용하거나 왜곡하지 말아야 한다는 인식에 있어서는 태도의 측면이 부각되지만, 타인의 아이디어를 무단으로 도용하거나 왜곡하는 것이 구체적으로 어떻게 판단되는가(인용 부호나 출처 표시 없이 몇 단어 이상 그대로 가져오는 경우 등)에 있어서는 지식의 측면이 부각되고, 타인의 아이디어를 가져다 사용해야 하는 구체적인 작문 상황의 맥락에서 이러한 기준에 저촉되지 않으면서 자료를 충분히 활용하고 변형, 통합하는 방법에 있어서는 기능의 측면도 반영된다. 따라서 태도를 독립적으로 측정하는 대신 지식, 기능과 통합하여 수행의 형태로 문항을 구성한다면, 태도의 평가는 단편적 진술을 기계적으로 측정하는 것으로부터 충분히 자유로워질 수 있다.

둘째, 일반화의 추론에 따라 수능 작문 평가에 대한 해석적 논증을 구성하기 위해서는 ‘수능 작문 평가 문항을 통해 관찰된 점수가 상이한 변인의 영향에 관계없이 보편적으로 통용된다.’라는 가정을 충족해야 한다. 이는 수능 작문 평가에서 세운 판단 기준에 따른 결과가 피검사자의 성별, 지역, 종교, 지식 체계 등에 따라 서로 다른 의미를 내포하지 않으며, 적어도 특정 문항에서 측정하고자 하는 구인에 있어서만큼은 모든 평정자가 유사한 해석을 도출할 수 있어야 한다는 점을 함의한다. 하지만 이러한 가정에 따라 기존의 수능 작문 평가를 되짚어 보면, 쓰기 과정별로 평가 요소를 인위적으로 분절하여 문항을 구성하여 왔기 때문에 작문의 실제성과 관련하여 신뢰성과 일반화 가능성을 확보하는 것이 쉽지 않아 보인다. 장성민·민병곤(2016)에 따르면 선다형 문항에 의한 기준 수능 작문 평가의 점수와 실제 작문 수행 간 상호 예언력은 21.5%에 불과하였으며, 이마저도 여학생의 경우에는 12.9%로 더 낮게 나타났다.

이는 일차적으로 평가 문항에서 측정하고자 한 구인과 피검사자가 인식하는 작문 상황의 맥락이 불일치한 데서 그 원인을 찾을 수 있다. 평가 문

항에 포함된 가시적 결과물(텍스트)은 해당 문항에서 측정하고자 하는 특정한 하나의 독립된 구인으로부터 단독적인 영향을 받은 것이 아닐뿐더러 ‘표현하기’, ‘고쳐쓰기’ 등의 경우에는 각각이 지칭하는 외연의 범위가 지나치게 포괄적이기 때문이다. 미국의 대학 입학 시험(SAT)에서 ‘쓰기와 언어(writing & language)’라는 통합된 영역을 제시하고 표현의 문제를 체계적으로 나누어 측정하거나, 완결된 글을 메타적으로 분석하는 에세이 과제를 부과하여 피검사자의 작문 능력을 총체적으로 파악하려는 것은 이러한 문제점을 극복하기 위한 결과의 소산으로 볼 수 있다. SAT의 작문 평가 문항이 대부분 완결된 텍스트로부터 작문 맥락에 부합하는 점검과 조정을 요구하는 형태로 설계되어 있다는 점은, 남학생과 여학생에게서 실제 작문 수행과의 유의미한 상관이 존재하는 구인이 각각 ‘계획하기, 내용 조직하기’와 ‘고쳐쓰기’로 상이하게 나타난 우리의 현실(장성민·민병곤, 2016)과 대조된다.

셋째, 외삽의 추론에 따라 수능 작문 평가에 대한 해석적 논증을 구성하기 위해서는 ‘수능 작문 평가의 결과가 미래의 수행이나 다른 맥락에서의 수행을 예측한다.’라는 가정을 충족해야 한다. 이 연구에서는 고등학교 졸업 자격 시험과 변별된 대학 입학 시험으로서의 속성이 부각되는 우리의 현실에 비추어 이 가정을 되짚어 볼 필요가 있다고 판단하였다. 즉 수능 작문 평가를 통해 관찰된 수행은 대학수학능력시험 본래의 목적에 부합하는 미래의 수행이나 다른 맥락에서의 수행을 예측하는 데 기여할 수 있어야 한다.

이에 따르면 작문 능력을 대학수학능력의 하나로서 내용 지식의 학습을 위한 사고의 도구(writing to learn)로 재규정하고, 수능 작문 평가에서 이를 체계적으로 측정할 수 있는 방안을 모색할 필요가 있다. 이러한 발상은 일상 의사소통에서 통용되는 언어 대신 다양한 내용 영역에서 활용될 수 있는 학습용 언어를 측정하고자 하는 외국의 사례를 통해서도 뒷받침된다.⁴⁾ 그동안

4) 2016년 개편된 미국의 대학 입학 시험(SAT)에서는 고차원적인 학습에 수반되는 적절한 어휘가 일상 의사소통이나 직접적인 어휘 지도만으로 습득되지 않으며, 문식 행위를 통한

대학수학능력시험에서의 화법, 작문 영역이 독서, 문법, 문학 영역과 달리 전문적인 교수·학습을 반드시 거치지 않더라도 해결할 수 있는 문항으로 구성되어 있어 변별력을 확보하기 어렵다는 비판에 직면하곤 했는데, 이는 기존의 수능 담론에서 화법, 작문 능력을 교과 적합도(disciplinary-appropriateness)에 대한 인식을 전제로 성립하는 대학수학능력의 하나로서가 아닌 일상 의사소통 능력과 동일시하여 왔기 때문에 발생한 문제로 볼 수 있다. 일례로 교육과정에서 ‘읽기의 생활화’나 ‘자발적 읽기’가 강조되는 것을 통해 확인할 수 있듯 독서 영역의 경우에는 학습 이외의 읽기 상황이 보편적으로 통용되면서도 수능에서는 인문, 사회, 과학, 예술 등 다양한 내용 지식을 학습하는 상황으로 과제가 설계되어 왔다. 반면 작문 영역의 경우에는 중등 학습자에게 구어적, 자기 표현적 속성이 두드러지는 일부 상황을 제외하고는 대체로 내용 지식의 학습에 통용되어 왔음에도 불구하고 수능에서 이를 적절히 반영하지 못하는 경우가 다반사였다. 직업, 역사/사회, 인문, 과학 등 네 개의 내용 영역으로 나누어 대입 작문 평가를 설계하는 미국의 사례는 우리의 현실과 대조를 이룬다.

마지막으로 결정의 추론에 따라 수능 작문 평가에 대한 해석적 논증을 구성하기 위해서는 ‘수능 작문 평가의 결과가 학습자의 작문 능력에 대한 유용한 정보를 제공한다.’라는 가정을 충족해야 한다.⁵⁾ 대학수학능력시험은 일차적으로 예비 대학생으로서 고등학교 3학년 수험생을 분류하고 그 결과를

간접적인 교수·학습이 수반되어야 하는 것으로 규정하였다(College Board, 2015). 이들은 어휘 분야의 선행 담론을 바탕으로 Tier 1(일상 언어), Tier 2(영역 일반적 학습 언어), Tier 3(영역 특수적 학습 언어)로 세분하고, 대학 입학 시험에서 Tier 2를 측정해야 한다고 주장하였다.

5) 결정의 추론을 뒷받침하는 가정에 대해서는 연구자에 따라 상이한 해석을 보일 가능성이 존재한다. 가령 일반적인 언어 능력 평가와 변별되는 대학수학능력시험의 선발적 기능에 주목하여 온 전통적 관점에서는, 수능 작문 평가의 결과가 학습자의 작문 능력에 대한 유용한 정보를 제공한다는 것이 어떻게 하면 본래의 선발적 기능을 잘 구현할 것인가에 주목한 형태로 해석될 수 있다.

선발에 활용하려는 목적 외에 예비 수험생으로서 고등학교 1~2학년 학습자로 하여금 자신이 부족한 점이 무엇인지 되돌아보고 이를 보완하여 스스로의 실력을 향상시키도록 돋는 기능도 수반하기 때문이다. 따라서 수능 작문 평가의 결과는 학습자나 교사로 하여금 어떤 적절한 결정을 내리는 데 기여하며, 이후의 교수·학습에 유의미한 도움을 줄 수 있어야 한다.

하지만 이러한 가정에 따라 기존의 수능 작문 평가를 되짚어 보면, 현재의 평가 문항은 채점, 일반화, 외삽의 측면에서 타당성을 확보하지 못하기 때문에 반복적인 문제 풀이에 의해 문항 유형에 기계적으로 익숙해지도록 유도할 뿐 작문 능력을 기르는 것과 관련된 실질적인 정보를 제공하는 데에는 이르지 못한다. 간접 측정이 갖는 공정성, 신뢰성의 장점을 유지하면서 실제 작문 능력으로 전이될 수 있는 지점을 찾아, 이를 중심으로 수능 작문 평가를 재구조화하려는 노력이 필요하다.

그러나 평가 문항이 수 년에 걸쳐 특정한 몇 가지 방식으로 고착화된다면, 아무리 실제 작문 능력과의 상관이 높은 우수한 문항이라 하더라도 현실적으로는 반복적인 문제 풀이 기술(skill)의 습득으로 변질될 우려가 높다는 점을 경계할 필요가 있다. 이 연구의 궁극적인 목적 또한 항구적인 수능 작문 평가 방식을 제안하는 데 있는 것이 아니라, 현재의 평가 방식에 내포되어 있는 타당성의 문제를 되짚어 보고 대안적 관점의 모색을 통해 이를 개선 할 수 있는 방안을 마련하는 데 그 초점이 놓여 있다. 따라서 이 연구에서 제안한 대안적 방식들이 상당 부분 반영된다 하더라도 일정한 시간이 경과한다면, 관성화된 적용을 지양하는 대신 새롭게 발굴된 여러 담론을 반영하여 재개선의 절차를 거쳐야 할 것이다.

IV. 수능 작문 평가의 개선 방향 탐색

이 장에서는 III장의 해석적 논증으로부터 도출되는 수능 작문 평가의 개선 방향을 탐색하고자 한다. 국내의 이론적, 경험적 논의만으로 추론과 가정을 뒷받침하는 증거가 뒷받침되지 않는 경우에는, 2016년 개편된 미국의 대학 입학 시험(SAT)이나 2020년 개편 예정인 일본의 대학 입학 시험(대학 입학공통시험)의 사례를 검토하여 반영하였다. 수능 작문 평가의 개선 방향은 다음과 같이 네 가지로 정리하였다.

1. 기능 측정에 편중된 평가 개선

수능 작문 평가에 대한 해석적 논증으로부터 도출되는 첫 번째 개선 방향은, 작문 기능(functions)의 측정에 편중되어 온 현재의 평가 내용을 개선하고 지식, 기능, 태도의 제 측면을 균형 있게 고려하여 교육과정이 의도하는 작문 능력을 두루 측정할 필요가 있다는 점이다. 기존의 수능 작문 평가에서는 대체로 인지주의 작문 이론의 쓰기 과정에 따라 평가 요소를 세분하고, 각각의 요소별로 정형화된 과제 유형에 따라 기능의 수준을 측정하는 형태로 이루어졌다. 이는 교수·학습과 평가 간 괴리라는 일차적 문제 외에 장르 지식, 쓰기 윤리 등 그 중요성이 갈수록 부각되는 작문 교육 담론의 최근 논의들을 반영하지 못한다는 문제점이 있다. 물론 수행의 속성이 부각되는 작문 영역의 평가에서 기능이 배제된 채로 지식이나 태도만을 개별적으로 측정하는 것 또한 문제의 소지가 있으나, 기능과 지식, 기능과 태도를 유기적으로 결합하여 과제를 설계한다면 기능의 거시적 틀 속에서 지식이나 태도의 문제를 고려할 수 있는 가능성도 존재한다.

선행 담론에서 텍스트 유형(text type)을 ‘설명 텍스트’와 ‘서사 텍스트’로 양분하고 ‘논증’과 ‘이야기’를 각각의 전형적인 상위구조로 지목한 것(van

Dijk, 1995)은, 중등 학습자가 작문 능력을 갖추기 위해 학습해야 하는 하나의 대표적인 장르 지식의 예를 보여준다.⁶⁾ 국어과 교육과정에서 듣기·말하기, 읽기, 쓰기 외에 ‘문학’을 별개의 독립된 영역으로 설정하고 있어 서사 텍스트와 관련된 장르 지식이 작문 영역의 교육과정에 직접 반영되지 않는다는 점을 고려할 때, 작문 영역에서 다루어야 할 중핵적인 장르 지식 가운데 한 가지가 ‘논증’이라는 점에 대해서는 이견의 여지가 없어 보인다. 최근에는 논증을 설득 의사소통의 전유물로 국한하지 않으며, ‘정보 전달’과 ‘설득’ 또한 배타적인 의사소통 목적으로 분리하지 않는 관점이 등장하는 추세이다 (장성민, 2016).

이를 바탕으로 작문의 장르 지식을 측정하기 위한 평가 문항의 예를 다음과 같이 생각해 볼 수 있다. 먼저 기존의 평가 문항과 유사한 형태로 구성하면, ① 작문의 상황 맥락(목적, 주제, 독자, 매체)이 담긴 메모와 ② 논증 구조(주장, 자료, 전제, 보증, 반박, 반박 자료)에 대한 전반적인 소개가 담긴 자료를 제시한 다음 ③를 바탕으로 ④의 세부 요소별로 적합한 정보나 아이디어의 유형을 찾도록 하는 문항을 설계할 수 있다. 여기에서 더 나아가 미국의 대학 입학 시험(SAT)의 평가 방식을 적용하여 완결된 텍스트에 대한 메타행위적 속성을 부각시키면, ⑤ 분석의 대상이 되는 학생의 글과 ⑥ 피드백의 주체가 되는 교사 또는 동료 학습자의 논평을 나란히 제시한 다음 논평의 초점이 되는 논증 요소와 이에 대응하는 논평의 타당성을 무엇을 말하는가(내포, intention), 왜 말하는가(의도, intension), 어디에 초점을 두어 말하는가(주의, attention) 등의 기준으로부터 점검해 보도록 하거나, ⑦ 완결된 텍스트로서 개인의 작문 결과물과 ⑧ 개인이 작문을 수행하는 과정에서 산출한 성찰일지 기록을 함께 제시한 다음 ⑨와 ⑩ 사이의 괴리된 지점을 찾아 개

6) 상위구조(superstructure)는 “텍스트의 유형을 표시해 주는 총괄적 구조”(van Dijk, 1995)를 가리킨다. 상위구조는 텍스트를 구성하는 일련의 정보들을 범주화하고 배열하는 추상적인 도식(schema)으로서, 범주들의 조합은 관습적인 규칙에 기반하여 이루어진다.

선 방향을 연결하도록 하는 문항의 형태로도 설계할 수 있다. 메타적 속성이 부각된 후자의 문항 유형은 피검사자로 하여금 외부에 존재하는 객관적 정보의 사실 여부를 판단하는 데 머무르지 않고, 장르 지식을 중심으로 한 의미 구성에 직접 활용하도록 요구한다는 점에서 기존의 문항 유형과 변별된다.

쓰기 윤리는 지식, 기능, 태도의 제 측면이 두루 고려될 수 있는 학습 요소이다. 국어과 교육과정에서는 ‘인식’이라는 태도의 차원에서 이를 반영하고 있으나, 여기에는 무엇을 비윤리적 쓰기로 규정할 것인가에 관한 지식이나 구체적인 작문 상황의 맥락에서 쓰기 윤리를 지키면서 글쓰기를 수행하는 기능의 문제도 포함되기 때문이다. 가령 타인의 아이디어를 가져다 사용해야 하는 구체적인 작문 상황의 맥락을 구성하고 비윤리적인 수행을 판별하는 기준을 제시한 다음 그러한 기준에 저촉되지 않으면서 ‘자료 속 정보 가운데 어느 것을 수용하고 어느 것을 수용하지 않을 것인가?’, ‘수용한 정보 가운데 어느 것을 변형하고 어느 것을 변형하지 않을 것인가?’, ‘필자 스스로 생성한 정보와 자료에 기원을 둔 정보를 서로 어떻게 연결할 것인가?’ 등의 수사적 문제를 해결하는 형태로 문항을 설계한다면, 태도의 문제 또한 타당성을 확보한 형태로 측정할 수 있을 것이다.

2. 쓰기 과정 중심의 인위적 분절 개선

수능 작문 평가에 대한 해석적 논증으로부터 도출되는 두 번째 개선 방향은, 인지주의 작문 이론의 쓰기 과정을 중심으로 하는 평가 요소의 인위적 분절을 개선하고 완결된 텍스트에 대한 실제적인 분석을 지향함으로써 평가 결과의 신뢰성과 일반화 가능성을 확보할 수 있어야 한다는 점이다. 이러한 방향은 다음의 두 가지 측면에서 장점이 있다.

첫째, 그동안 쓰기 과정 중심의 평가 항목에 가려져 있던 작문의 장르적 문제에 주목할 수 있다는 점이다. 수능 작문 평가의 출제 범위에 해당하는 고등학교 선택과목 “화법과 작문”의 교육과정은 장르 일반적 속성을 공유하

는 ‘화법과 작문의 본질’(2015 개정 교육과정의 경우에는 ‘화법과 작문의 태도’도 포함)을 제외하고는 대부분의 성취기준이 의사소통 목적에 따른 장르적 속성이 부각된 형태로 구성되고 있다. 그럼에도 불구하고 기존의 수능 작문 평가에서 쓰기 과정 중심의 평가 요소에 비추어 정형화된 문항을 출제하다 보니, 장르적 속성에 대한 충분한 이해가 선행되지 않더라도 독해력만으로도 과제를 해결할 수 있는 교수·학습과 평가 간 괴리가 반복되곤 하였다. 또한 장르 지식에 대한 기본적인 이해를 염두에 두지 않다 보니, ‘의견 제시형 에세이(opinion essay)’와 같이 정보 전달, 설득, 표현 등의 의사소통 목적으로 복합적으로 중첩되어 나타나는 간장르(inter-genre)적 속성의 함의에 대해서도 주목하지 못했다.

둘째, 문항별로 측정하고자 하는 특정한 세부 쓰기 과정의 구인 외에 상이한 구인으로부터 영향을 받아 야기되던 결과 해석의 신뢰성과 일반화 가능성의 문제를 해결할 수 있다는 점이다. 가령 ‘계획하기’를 측정하기 위해 설계된 가장 최근의 기출 문항에서 세부 문단별로 어떤 정보를 어떻게 조직할 것인지 제시한 것은, 실제로는 그것만으로 이미 미시적 표현을 제외한 모든 쓰기 과정을 거친 결과로 볼 수 있다.⁷⁾ 문항별로 측정하고자 하는 특정한

- 7) 해당 문항은 다음과 같다(2018학년도 대학수학능력시험 국어 영역 훌수형 6번).

△ 다음은 (가)를 반영하여 (나)를 작성하기 위한 민호의 작문 계획이다. (나)에 반영된 내용으로 적절하지 않은 것은?	
1문단	허생의 처가 추구한 행복의 조건이 외적 조건이라고 한 기준의 내 의견과, 토의를 통해 수정된 내 생각을 함께 써야겠어. ①
2문단	허생의 처가 행복하지 않은 이유를 생겨 문제를 중심으로 파악했던 의견에 의문을 제기하고 이에 답하는 식으로 써야겠어. ②
3문단	‘영수’가 허생의 처의 말을 인용하면서 개진한 의견을 포함하여 허생의 처가 행복해지기 위한 조건을 써야겠어. ③
	나와 ‘영수’가 허생의 처의 행복을 가족 간 관계의 측면에서 논의한 내용을 바탕으로, 내가 기준에 갖고 있던 행복에 대한 생각이 편협했음을 깨달았다는 내용을 써야겠어. ④
	허생의 처가 왜 행복하지 않은지에 대해 나와 ‘영수’가 동의했던 두 가지 이유 중 강요된 희생을 주된 이유로, 소원한 관계를 부자적 이유로 구별하고 이에 비추어 나의 삶을 반성하는 내용을 써야겠어. ⑤

세부 쓰기 과정의 평가 요소는 작문의 문제 해결 과정에 본질적으로 내포되어 있는 회귀성으로 말미암아 실제 작문 수행에서 그다지 보편적으로 통용되는 지점이 아닐 가능성도 배제하기 어렵다.

이러한 논의는 국어과 교육의 초점을 기존의 ‘국어 활동’ 중심에서 ‘메타 국어 활동’으로 확장하자는 지식 중심의 국어과 교육 담론의 주장과도 연결된다. 이들의 담론에서 메타 국어 활동에 주목한 배경에는, 이들이 행동(수행) 영역의 교육을 중심에 놓되 여기에 관여하는 과정적, 기능적 지식의 문제를 국어과 교육의 전면에 내세웠기 때문인 것으로 추정된다. 대학 작문 교육에 ‘쓰기에 대한 쓰기(writing about writing, WAW)’ 담론의 바람이 불었었고, 새롭게 개편된 대학 입학 시험(SAT)에 ‘고쳐쓰기’와 관련된 선다형 문항이 높은 비중으로 포함되거나 완결된 텍스트에 대한 메타적 분석의 에세이 쓰기를 요구하는 미국의 사례도 이와 유사한 맥락에서 이해할 수 있다.

지식 중심의 국어과 교육 담론에서는 국어 활동과 메타 국어 활동에 관여하는 지식의 유형을 <표 1>과 같이 2×2 행렬의 매트릭스로 규정하였다(민병곤, 2006; 박종훈, 2005).

<표 1> 국어 활동과 메타 국어 활동에 관여하는 지식의 유형

	국어 활동	메타 국어 활동
과정적 지식	표현과 이해 활동의 과정에 대한 지식(ⓐ)	국어를 대상화하여 관찰하고 분석하는 과정에 대한 지식(ⓑ)
기능적 지식	특정한 상황 맥락에서 언어 표현이 수행하는 기능에 대한 지식(ⓓ)	상황 맥락과 언어 표현의 관계에 대한 지식(ⓔ)

박종훈(2005)은 지금까지 국어과 교육이 주로 Ⓛ에 편중된 형태로 이루어져 왔음을 지적하였다. “‘상황에 맞게 들을 줄 알아야 하며, 읽기를 통해 필자의 의도를 파악해야 한다.’라는 식의 선언적 수준의 논의”에 머물렀다는 것이다. 장르 지식(ⓑ)과 메타인지(ⓒ, ⓔ)의 중요성에 대한 인식이 갈수록 더해가고 있는 오늘날에도 적어도 수능 작문 평가에 있어서만큼은 그러한

지적이 여전히 유효하다. 앞서 제시한 첫 번째 개선 방향이 과정적 지식(ⓐ, ⓒ)과 기능적 지식(ⓑ, ⓑ)의 평가를 균형 있게 고려하자는 주장이었다면, 여기에서 제시하는 두 번째 개선 방향은 국어 활동 중심의 기존 평가 방식(ⓐ, ⓑ)을 메타 국어 활동에 주목한 대안적 평가 방식(Ⓒ, ⓔ)으로까지 확장하자는 주장으로 볼 수 있다.

가령 앞 절에서 제안한 ‘쓰기 윤리’라는 학습 요소에 대해 메타 국어 활동의 형태로 평가 문항을 재구성해 본다면, 구체적인 작문 상황의 맥락에서 마주하는 수사적 문제의 해결 대신 비윤리적 수행이 포함된 텍스트를 완결된 형태로 제시하고 그것이 어떤 점에서 비윤리적이며, 이를 개선하기 위해 무엇을 보완해야 하는가에 대해 구체적으로 생각해 보도록 요구할 수 있다. 메타 국어 활동에 주목하여 평가 문항을 설계하는 것은 기존의 쓰기 과정 중심의 인위적 평가 요소 분절에 따른 타당성의 문제를 해결하고, 실제적인 상황 맥락에서 평가 결과에 대한 해석을 신뢰롭게 일반화하는 데 기여할 것으로 기대된다.

3. 학습 목적 글쓰기 상황 중심의 재구조화

수능 작문 평가에 대한 해석적 논증으로부터 도출되는 세 번째 개선 방향은, 수능 작문 평가에서 측정하고자 하는 작문 능력을 대학수학능력의 하나로 재규정하고, 학습 목적 글쓰기(writing to learn)의 과제 맥락을 상정하여 이를 중심으로 재구조화할 필요가 있다는 점이다. 이는 고등학교 졸업 자격 시험과 변별되는 대학 입학 시험으로서 대학수학능력시험의 지니고 있는 본연의 목적에 천착한 결과이다. 수능 작문 평가를 통해 관찰된 결과는 대학에서의 학습과 밀접하게 관련을 맺는 작문의 수행을 예측하는 데 기여할 수 있어야 한다.

학습 목적 글쓰기 상황에서는 “읽기와 쓰기, 학습이라는 세 가지 행위의 토대 위에 성립”(Spivey, 1997)하는 복합적 속성이 주요하게 부각된다. 수

능 작문 평가가 읽기 자료를 통해 과제에 대한 인지적 반응의 범위를 교과 적합도(disciplinary-appropriateness)에 부합하는 방향으로 일정하게 조정 해 왔다는 점에서, 그것이 자료 통합적 글쓰기와 공유된 속성을 보이거나(장 성민·민병곤, 2016) 텍스트 이해 과정 중심의 평가틀로 설계되어 왔다는 점 (최미숙, 2018)에 대해서는 어느 정도 합의가 도출되고 있다. 대학 입학 시험(SAT)에서 에세이 쓰기를 평가할 때 ‘읽기, 분석, 쓰기’를 고려하도록 하는 미국의 사례는 학습 목적 글쓰기에 수반되는 세 가지 능력의 구인을 두루 측정하고자 한 것으로 볼 수 있으며, 선다형 문항으로 구성된 ‘쓰기와 언어 (writing & language)’ 영역에 직업, 역사/사회, 인문, 과학 등 네 개의 내용 영역별로 지문을 제시하고 지문마다 11개의 하위 문항을 구성한 것 또한 읽기 자료를 통해 교과 적합도를 확보하려는 학습 목적 글쓰기 상황의 특성을 반영한 것으로 해석된다.

이 연구는 기존의 대학수학능력시험에서 독서와 작문 영역의 평가 구도를 통합적으로 접근함으로써 학습 목적 글쓰기 상황에 가까워질 수 있다고 판단하였다.⁸⁾ 자료 통합적 글쓰기에 대한 사회적 수요가 지속적으로 증가하고 있으며 작문 영역의 평가 과제(수능, 대입 논술고사 등)에 읽기 자료를 반드시 포함하고 있음을 염두에 둘 때, ‘읽기 후 과정’과 ‘계획하기’의 본질이 서로 유사하다고 밝힌 선행 담론의 규정(Kennedy, 1985)은 더 이상 새로운 이야기가 아니다. 일본의 대입 평가(대학입학공통시험)에서 “독자들이 텍스트를 이해할 때 텍스트의 구조와 내용을 파악한 후, 텍스트 내용을 정밀하게 조사하고 해석하며, 나아가 텍스트에 대한 생각을 형성하는 ‘과정’을 구조화”(최미숙, 2018: 45)하는 것은 우리의 대학수학능력시험에서 작문 영역과 독서 영역 모두에 거의 동일하게 적용된다. 사실적 이해, 추론적 이해, 비

8) 실질적인 통합으로 보기에는 무리가 있으나, 기존의 대학수학능력시험에도 서로 다른 영역 간 평가를 통합적으로 구성하려는 시도가 발견된다. 가령 2018학년도 대학수학능력시험의 흘수형 4~7번에서는 화법 문항(4~5번)과 작문 문항(6~7번)이 동일한 지문(대화문, 글)의 하위 문항으로 함께 조직되어 있다.

판적 이해, 창의적 이해 등으로 이어지는 독서 수행의 흐름은 수능 작문 평가나 자료 통합적 글쓰기의 문제 해결 과정에 요구되는 사고의 흐름과도 상응한다. 특히 학습 목적 글쓰기 상황에서의 교과 적합도가 읽기 자료의 내용 영역을 통해 확보된다는 점을 고려할 때, 기존의 독서 문항에 제공된 상이한 내용 영역의 지문별로 1~2개의 작문 문항을 출제하는 것은 효율적인 대안이 될 수 있다.

2020년부터 새롭게 개편되는 일본의 대학 입학 시험(대학입학공통시험)에서는 평가 문항의 유형을 <표 2>와 같이 분류하고 있다(최미숙, 2018 재인용).⁹⁾

<표 2> 일본 대학 입학 시험의 평가 문항 유형 분류

텍스트의 내용과 해석을 답하는 문제		생각을 답하는 문제	
① 텍스트 부분의 파악, 정밀 조사·해석(텍스트 부분을 파악, 정밀 조사·해석하여 답하는 문제)	② 텍스트의 전체 파악, 정밀 조사·해석(텍스트의 전체를 파악, 정밀 조사·해석하여 답하는 문제)	③ 텍스트의 정밀 조사·해석에 근거한 생각의 형성(텍스트의 정밀 조사·해석에 근거한 생각을 답하는 문제)	④ 텍스트의 정밀 조사·해석을 근거로 자신의 생각을 형성(텍스트의 정밀 조사·해석을 바탕으로 발전시킨 자신의 생각을 답하는 문제)
선택식·단답식	선택식·조건부 기술식	선택식·조건부 기술식	자유기술식·소논문

<표 2>에 따르면, 피검사자에게 요구되는 인지적 반응의 범위가 ‘필자-자료 간 통합(writer-source integration)’으로까지 열려 있어 채점의 객관성과 공정성을 확보하는 것이 상대적으로 쉽지 않은 ④를 제외하고는 대부

- 9) 다만 이 연구에서 일본의 대학 입학 시험의 모든 특징을 그대로 수용하자는 관점을 취하는 것은 아니다. 대표적으로 인문, 사회, 과학, 예술 등 학습용 제재가 아닌 대화문, 기사문, 홍보문, 규약문, 계약서, 제안서 등의 실용적 제재만을 읽기 자료로 활용한 것에 대해서는 국내의 평가 담론에서도 회의적으로 접근하고 있다(권태현, 2018; 최미숙, 2018). 대학수학능력시험 본래의 목적에 부합하는 방향으로 외삽을 지향하는 이 연구에서도 이러한 점에 대해 비판적인 접근을 견지한다.

분 선다형 문항으로도 측정이 가능하다. 이 가운데 ①, ②가 기존 수능의 독서 영역과 관련을 맺는 유형이라면, ③은 독서 영역의 비판적, 창의적 이해나 작문 영역과 관련을 보이는 유형이라 할 수 있다. 비록 자유기술식에 가까운 형태로 진행되는 미국 SAT의 에세이 쓰기가 ‘읽기, 분석, 쓰기’의 제 측면을 전면적으로 개방하여 평가하는 것에 미치지는 못하지만, 독서 영역과 작문 영역의 결합을 통해 기존 수능 작문 평가의 난점을 해소하는 데 일정하게 기여할 수 있을 것으로 판단된다. 차후 우리의 대학수학능력시험에도 서술형, 논술형 문항이 반영된다면, 기존의 독서 영역에서 다루어지던 전통적인 이해 능력의 평가 요소는 선다형 문항의 형태로, 비판적, 창의적 이해나 작문 능력과 관련된 평가 요소는 서술형, 논술형 문항의 형태로 구체화할 수 있을 것이다.

4. 전이 가능한 능력의 측정

수능 작문 평가에 대한 해석적 논증으로부터 도출되는 네 번째 개선 방향은, 전이 가능한 능력을 타당하게 측정함으로써 분류와 선발의 일차적 목적 외에 학습자의 작문 능력에 대한 유용한 정보를 제공하는 데에도 기여할 수 있어야 한다는 점이다. 결정의 추론으로부터 도출된 이러한 방향은 엄밀히 말해 평가 담론에 국한되는 지점이라기보다 작문 교육 담론 전반에 놓여 있는 과제로 볼 수 있다.

이는 작문 영역에서 교수·학습을 통해 실제적인 능력의 향상으로 이어질 수 있는 중핵적 지식이 무엇인가에 대한 근원적인 고민으로부터 출발할 수 있다. 기존의 수능 작문 평가를 염두에 둘 때, 우리의 작문 교육 담론에서는 전통적인 인지주의의 작문 이론으로부터 도출되는 쓰기 과정에 관한 지식을 그러한 지식의 전형으로 줄곧 규정해 온 것으로 짐작된다. 이는 대학수학능력시험이 처음 도입된 1990년대부터 지금까지도 여전히 유효해 보인다. 하지만 20년 이상 관습적으로 지속되어 온 그러한 지식은, 최근 필자들의 실

제 작문 수행이 ‘계획하기’로부터 순차적으로 이어지는 정형화된 절차로 일 원화되지 않으며 기존의 가정과는 정반대로 ‘표현하기’와 ‘고쳐쓰기’에 의한 아이디어 촉발이 이루어지는 경우도 적지 않음을 보여주는 연구 결과로부터 반박에 직면해 있다(김혜연, 2014). 이는 작문 담론이 지속적으로 새로운 주제를 발굴하고 각각의 주체가 부침을 경험하며 쟁점을 형성하여 왔음에도 불구하고, 정작 교수·학습이나 평가 담론에서는 20여 년 전의 논의에서 탈피하지 못하고 있음을 보여주는 대표적인 예이다.

상황이 그러하다 보니, 작문 교수·학습 담론에서 기존에 교육 내용으로 설정하여 온 지식, 기능, 태도 등을 가르침으로써 학습자의 국어 활동 능력, 메타 국어 활동 능력을 신장시키는 데 실제로 기여할 수 있는가에 대해서는 낙관적인 전망이 이루어지지 못했다. 작문 지식, 작문 태도와 수행 사이의 상관성에 주목한 국내의 여러 선행 연구에서는 이들 사이의 상관관계가 그리 높지 않은 것으로 보고되어 왔다(박영민, 2010 등). 하지만 이들이 상정한 작문 지식은 쓰기의 본질, 과정, 구성, 표현, 실제, 규범 및 어휘 등에 관한 지식을 지칭하는 것이었으며, 작문 태도(쓰기 동기)는 협력적 상호작용, 효능감, 노력, 도전심 등을 지칭하는 것이었다. 지식과 태도의 이러한 규정은 구체적인 수행의 맥락과 직접 관련을 맺는 항목들로 구성되어 있는 것으로 보기 어렵다. 만약 장르 일반적인 지식 대신 장르 특수성이 부각된 지식(설명문의 구성 원리)을 대상으로 삼거나, 수행과 전혀 다른 맥락에서 도출되는 태도 대신 구체적인 상황 맥락을 공유하는 태도(자료 통합적 글쓰기 - 쓰기 윤리)를 측정한 다음 수행과의 상관성에 주목하였다면, 그 결과는 기존의 결과와 상이하게 나타났을 수 있다. 이들이 염두에 둔 지식이나 태도가 현장의 교사나 학습자가 상식적으로 생각하는 범주의 요소임에는 틀림 없으나, 그것이 작문 교수·학습 담론에서 반드시 고려해야 하는 중핵적인 학습 요소들을 제대로 반영하고 있는가에 대해서는 쉽게 답변하기 어렵다. 어쩌면 ‘지식 따로, 수행 따로’라는 식으로 치부되어 온 작문 교육의 오랜 편견은 실제로 지식이 수행의 문제와 별개이기 때문이 아니라, 이를 규정하여

온 관행적 논의가 수행으로의 전이를 타당하게 설명하지 못했기 때문에 발생한 결과로 볼 수 있다.

이러한 점에서 2015 개정 교육과정에 도입된 ‘핵심 개념(core concept)’의 발상은 교과 내용의 영속적 이해(enduring understanding)를 도모할 수 있는 중핵적인 지식의 목록을 알려줄 것이라는 기대를 심어 주었다. 하지만 교육과정 문서에서 “교과의 기초 개념이나 원리”로 정의되는 ‘핵심 개념’이라는 용어는 지식의 한 종류로서 ‘개념’과 동일한 의미로 사용된 것이 아니라 교육과정 내용 체계의 하위 영역(area)을 기계적으로 지칭하는 ‘상위개념’과 유사한 의미로 통용되었다. 핵심 개념의 발상은 이해 중심 교육과정 담론의 ‘빅 아이디어(big idea)’에 기원을 둔 것이었지만, 우리의 교육과정에서는 이마저도 그 취지를 정확하게 반영하지 못하고 있다. 학습자의 작문 능력에 관한 유용한 정보를 제공하기 위해서는 핵심 개념의 의미가 교육과정 내용 체계의 하위 영역을 기계적으로 지칭하는 데 머무르지 않고, 작문 능력을 뒷받침하는 “통합적, 생성적, 영속적 이해”(천경록, 2017)에 기여하는 지식의 목록을 제공하는 방식으로 설정되어야 한다. 여기에는 구체적인 학습 요소의 *값*(value)을 응축하고 “개념의 상이한 속성(attribute)을 가리키는 도식 표상의 요소”(Anderson, 2010)로 기능하는 ‘슬롯(slot)’이 그 중심에 놓여야 한다. 이러한 문제의식은 “국가 수준 교육과정 문서 수준에서는 의사소통과 언어, 표현과 이해, 감상, 텍스트, 맥락 등이 국어과의 핵심 개념으로 제시될 수 있고, 개별 수업 시간에는 보다 구체적 내용으로 제시될 수 있다.”(천경록, 2017: 52)라고 규정한 선행 담론의 대안적 지적과도 상응한다.

V. 결론

이 연구는 논증 기반 타당도 접근에 기반하여 수능 작문 평가를 둘러싼 담론의 해석적 논증을 구성하고, ‘채점, 일반화, 외삽, 결정’의 범주에서 타당도를 갖춘 수능 작문 평가의 대안적 방향을 도출하였다. 논증 기반 타당도 접근은 기존의 평가 방식과 대안적인 평가 방식에 대한 추론과 가정에 주목하고 이론적, 경험적 증거에 대한 뒷받침을 통해 평가의 구체적인 내용과 방법에 관한 명시적 진술들을 이끌어 내기 위한 하나의 유용한 틀로 활용되었다. 이 연구는 상이한 추론의 범주별로 도출되는 여러 가정들을 충족하기 위해 (1) 기능 측정에 편중된 평가 개선, (2) 쓰기 과정 중심의 인위적 분절 개선, (3) 학습 목적 글쓰기 상황 중심의 재구조화, (4) 전이 가능한 능력의 측정 등 의 개선 방향을 탐색하였다.

이 연구는 당위적으로 충족되어야 한다고 여겨지던 여러 가정들이 현실적 여건으로 말미암아, 또는 인식의 부재와 평가 방식의 관습화된 고착으로 말미암아 타당성의 문제로부터 자유롭지 못했던 기존의 수능 작문 평가에 대한 성찰을 요구한다는 점에서 의미가 있다. 해석적 논증을 뒷받침하는 증거로 최근 개편되었거나 개편이 예정되어 있는 국외의 대학 입학 시험 사례를 함께 살펴본 것은, 유사한 맥락을 공유하고 있는 다른 나라의 실태에 대한 이해를 바탕으로 우리의 작문 평가 정책을 결정하는 데에도 하나의 유용한 기초 자료로 활용될 수 있다.

다만 해석적 논증을 거쳐 산출된 연구의 성과가 결실을 맺기 위해서는 실제 평가 상황에서 이를 실천하고, 실증적인 자료 분석을 통해 효과를 검증하며, 파생된 문제점들을 보완하려는 작업이 뒷받침되어야 한다. 가령 수능 작문 평가를 메타 국어 활동에 주목한 형태로 전환하자는 논의나 ‘읽기, 쓰기, 학습’의 복합적 속성이 부각된 학습 목적 글쓰기 상황을 중심으로 재구조화하자는 논의는, 비단 작문 영역뿐 아니라 다른 영역과의 관계 설정에도 영

향을 미칠 수 있다는 점에서 영역 고유의 특수성을 간과한 제안으로 비춰질 가능성도 존재한다. 게다가 최근에는 대학수학능력시험에 서술형, 논술형 문항을 반영하고 이를 채점하는 방안에 대해서도 지속적인 논의가 오고 가는 만큼(노은희, 2017 등) 변화하는 작문 평가 담론의 흐름에 맞추어 유연하고 융통성 있게 적용할 필요가 있다.

* 본 논문은 2018. 4. 30. 투고되었으며, 2018. 5. 8. 심사가 시작되어 2018. 6. 1. 심사가 종료되었음.

참고문헌

- 강민경(2013), 「쓰기 윤리 교육의 내용 연구」, 서울대학교 박사학위논문.
- 권태현(2018), 「해외 대입 작문 평가의 동향 및 시사점」, 『작문연구』 36, 53-82.
- 김혜연(2014), 「쓰기 과정에서 생성하기와 검토하기의 역동적 상호작용」, 서울대학교 박사학위논문.
- 노은희(2017), 「컴퓨터 기반 평가 체계를 대비한 한국어 문장 수준 서답형 문항 자동채점 프로그램의 개발과 전망」, 『국어교육학회(제65회)·한국독서학회(제40회) 연합 전국학술대회 자료집』, 130-155.
- 민병곤(2000), 「신문 사설의 논증 구조 분석」, 『국어국문학』 127, 133-154.
- 민병곤(2006), 「말하기·듣기 교육 내용으로서의 ‘지식’에 대한 고찰」, 『국어교육학연구』 25, 5-38.
- 박영민(2010), 「쓰기 지식과 쓰기 동기가 중학생의 설명문 쓰기 능력에 미치는 영향」, 『새국어 교육』 84, 127-148.
- 박종훈(2005), 「지식 중심의 국어교육 내용 범주 설정 시론」, 『국어교육』 117, 469-492.
- 장성민(2016), 「학습 목적 글쓰기 교육에 대한 반성적 검토」, 『국어교육』 155, 63-105.
- 장성민·민병곤(2016), 「대학수학능력시험 “국어” 작문 영역 평가의 타당성 검토: 선다형 점수와 수행형 점수의 상관관계를 중심으로」, 『작문연구』 29, 163-209.
- 천경록(2017), 「국어과 ‘핵심 개념’의 의미 탐색」, 『청람어문교육』 63, 41-65.
- 최미숙(2018), 「일본 ‘대학입학공통테스트’ 국어 서술형 문항 분석 및 시사점」, 『제4회 국어교육학회 국제학술대회 자료집』, 38-55.
- 최숙기(2018 ㄱ), 「미국 SAT 작문 평가 문항 분석을 통한 대학수학능력시험 국어 영역의 작문 평가 개선 방안 탐색」, 『작문연구』 36, 107-144.
- 최숙기(2018 ㄴ), 「중학생의 자료 통합적 글쓰기에 나타난 자료 글 차용하기 수행 양상 연구: 인용과 환언하기를 중심으로」, 『2018년 한국작문학회·한국교육과정평가원 공동학술대회 자료집』, 173-188.
- Anderson, J. R. (2010), *Cognitive psychology and its implications* (7th ed.), New York, NY: Worth Publishers.
- Bachman, L. F. (2005), "Building and supporting a case for test use", *Language Assessment Quarterly* 2(1), 1-34.
- Breton, P. & Gauthier, G. (2006), 『논증의 역사』, 장혜영 (역), 서울: 커뮤니케이션북스(원서 출판 2000).
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010), "Does an argument-based approach to validity make a difference?", *Educational Measurement: Issues and Practice* 29(1), 3-13.
- College Board (2015), *Test specifications for the redesigned SAT[®]*, New York, NY: The

- College Board.
- Kane, M. (1992), "An argument-based approach to validation", *Psychological Bulletin* 112, 527-535.
- Kane, M. (2006), Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed.), Westport, CT: American Council on Education and Praeger.
- Kane, M. (2011), "Validating score interpretations and uses", *Language Testing* 29(1), 3-17.
- Kane, M. (2013), "The argument-based approach to validation", *School Psychology Review* 42(4), 448-457.
- Kennedy, M. L. (1985), "The composing process of college students writing from sources", *Written Communication* 2(4), 434-456.
- Messick, S. (1989), Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.), New York, NY: American Council on Education and Macmillan.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002), "Design and analysis in task-based language assessment", *Language Testing* 19(4), 477-496.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003), "On the structure of educational assessments", *Measurement: Interdisciplinary Research and Perspectives* 1, 3-62.
- Spivey, N. N. (1997), *The constructivist metaphor: Reading, writing, and the making of meaning*, San Diego, CA: Academic Press.
- Toulmin, S. (1958/2003), *The uses of argument* (update ed.), Cambridge: Cambridge University Press.
- van Dijk, T. A. (1995), 『텍스트학』, 정시호 (역), 서울: 민음사(원서출판 1980).
- Williams, J. M. & Colomb, G. G. (2008), 『논증의 탄생』, 윤영삼 (역), 서울: 흥문관(원서출판 2007).

논증 기반 타당도 접근에 기반한 수능 작문 평가의 개선 방향 탐색

장성민

이 연구는 국내의 수능 작문 평가 개선을 위해 논증 기반 타당도 접근의 관점에서 작문 평가를 둘러싼 담론의 해석적 논증을 점검하고, 타당도를 갖춘 대안적 관점을 제안하는 데 목적이 있다. 논증 기반 타당도 접근은 검사 점수의 해석과 사용에 내포된 광범위한 문제에 관심을 가지고, 이에 수반되는 가정과 증거에 주목하며, 대안적 관점의 가능성에 대한 고려를 포함한다는 장점이 있다. 이 연구는 ‘채점, 일반화, 외삽, 결정’의 네 범주에서 수능 작문 평가에 대한 해석적 논증을 구성하였으며, 이를 바탕으로 (1) 기능 측정에 편중된 평가 개선, (2) 쓰기 과정 중심의 인위적 분절 개선, (3) 학습 목적 글쓰기 상황 중심의 재구조화, (4) 전이 가능한 능력의 측정 등의 개선 방향을 탐색하였다.

핵심어 대학수학능력시험, 논증 기반 타당도 접근, 해석적 논증, 학습 목적 글쓰기, 장르 지식, 쓰기 윤리, 메타 국어 활동, 국어 교육, 작문 교육

ABSTRACT

Study of the Direction of CSAT Writing Assessment Based on the Argument-based Approach to Validation

Chang Sungmin

The purpose of this study is to construct interpretative arguments of the CSAT Writing Assessment based on the argument-based approach to validation and to suggest alternative approaches with validity. The adoption of the argument-based approach to validation has several positive effects, such as focusing attention on a broad array of issues inherent in the interpretations and uses of test scores, emphasizing the role of supporting assumptions and evidence, and allowing for the possibility of alternatives. This study constructed interpretative arguments in the category of “scoring, generalization, extrapolation, decision” and suggested (a) revising biases of measuring functions, (b) modifying artificial segments of writing processes, (c) reconstruction centered on writing-to-learn contexts, and (d) measuring of transferable capabilities based on those inferences.

KEYWORDS College Scholastic Ability Test(CSAT), Argument-based Approach to Validation, Interpretative Arguments, Writing to Learn, Genre Knowledge, Writing Ethics, Meta-language Performance, Korean Language Education, Writing Education