

국어과 서술형 평가 문항 개발 및 채점 방법 사례 연구

이영진 한국교원대 강사, 부천여자고등학교 교사

- I. 서론
- II. 이론적 배경
- III. 연구 방법
- IV. 결과 분석 및 논의
- V. 결론

I. 서론

최근 학교에서는 중간 및 기말고사와 같은 지필평가에서 평가 방식 및 평가 문항의 유형이 달라지고 있다. 즉, 기존의 선다형 평가 중심에서 서술형 평가의 비중을 확대시키고 있다. 서울 및 경기 지역을 비롯하여 점차 지역이 확대되는 추세이다.

평가 방식 및 유형이 변화한다는 것은 평가하고자 하는 내용, 방식 등의 변화와 관련을 맺는다. 그리고 근본적으로는 평가의 목적과도 영향을 맺는다. 그러므로 평가의 틀이 변화하게 된 원인이 무엇인지, 이러한 변화를 통해 평가가 의도하는 목적이나 내용이 무엇인지 살펴볼 필요가 있다. 또한 변화된 평가의 유형을 통한 적용이 과연 학교 현장에서 적합한 평가 방식 및 유형인지에 관해서도 살펴볼 필요가 있다.

최근 단위 학교에서 실시되고 있는 주요한 평가 중의 하나인 지필평가에서 서술형 평가 문항을 포함하면서 학생들이 지식을 생성하고 활용하는 과정 및 창의적 사고를 요구하고 있다. 그러나 학생들은 낯선 평가 유형에 대해 어려움을 겪고 있으며 한편 평가자인 교사들은 서술형 평가의 타당도,

신뢰도 확보 및 평가 업무 부담으로 인해 어려움을 겪고 있다.

이 연구에서는 학교 현장에서 실시되고 있는 국어과 서술형 평가의 문항 개발과 채점 방법에 관한 사례를 분석하고자 한다. 또한 국어과 서술형 평가 도입과 관련하여 학교 현장에서 교사들은 문항 개발과 채점 방법에서의 타당도와 신뢰도를 어떻게 확보하고 있으며 서술형 평가에 관한 교사들의 견해가 무엇인지를 살펴봄으로써 서술형 평가의 필요성 및 개선 방안에 관해 고찰하도록 한다.

II. 이론적 배경

1. 평가 방식 및 평가 문항 유형으로서의 국어과 서술형 평가

평가 방식과 문항 유형에 대한 개념 및 국어과 서술형 평가의 도입 배경에 관하여 선행 연구들을 고찰하였다.

박도순(1999)은 선택형 문항에는 진위형, 배합형, 선다형, 완성형, 단답형 등이 포함되며 서술형 문항에는 논문형(서술형)으로 응답 제한 및 응답 자유 서술형으로 제시하고 있다. 선택형과 서술형의 문항 분류는 피험자가 어떻게 반응하는가, 즉 시험의 상황에서 피험자가 제한된 답지에서 선택하여 표기하는가 아니면 답을 구성하여 작성하는가에 따른 것이다.

선택형 문항 유형의 장점은 무엇보다 채점의 객관성 및 신뢰도가 높다는 것이며 일정한 시간에 여러 내용을 많이 다루는 많은 문제를 실시할 수 있어 능력의 표본을 포괄적으로 다룰 수 있다는 것이다. 또한 채점이 쉽고 빠르며 그 결과는 쉽고 의미 있게 통계적으로 처리, 해석될 수 있다. 선택형 문항 유형의 단점은 단순한 기억력 내지 지식의 측정에 치우칠 위험이 있다는 것이다. 또한 문항 제작의 어려움과 함께 추측 요인을 완전히 제거할 수

없어서 학생들에게 정확한 지식을 인식하지 못함에도 불구하고 비슷한 답안을 추측하는 행동을 길러 주고 있다는 비난을 받고 있다. 또한 표현력을 측정하기에 부적당하며 창의성 발휘의 기회가 부족하다는 것이다.

서술형 문항 유형은 학생 자신이 정답을 구성하는 것이 주요 특징이다. 먼저 반응의 자유로 인해 문항이 지시하는 범위 내에서 학생들이 자유롭게 문제를 깊게 또는 넓게 접근할 수 있다는 장점이 있다. 또한 고등정신기능 측정에 효과적이다. 표현력, 사고력, 문제해결력과 같은 능력들을 평가할 수 있다. 이와 함께 비교적 문항제작이 쉽다는 장점이 있다. 그럼에도 불구하고 서술형 문항은 채점자들의 일관성, 채점자 간 또는 채점자 내에서의 신뢰도가 낮다는 단점이 있다. 또한 문항 표집수가 제한된다. 특히 서술형 문항은 현실적으로 몇 개의 제한된 문항만을 출제할 수밖에 없다. 그리고 채점에 시간과 노력이 많이 든다.

선택형과 서술형 문항 유형은 각각의 장단점을 포함하고 있지만 기존 학교 현장에서의 평가는 선택형 문항 유형에 치중되어 있다. 그중에서 주로 선다형 문항이 출제되어 왔다. 그러나 점차 학생들의 다양한 표현력 및 사고의 과정, 창의력을 요구하면서 수행평가뿐만 아니라 학생들의 성적에 민감한 영향을 끼치는 지필평가에서도 서술형 평가의 도입이 확대되었다. 이에 따라 국어과에서는 기존의 간접적으로 평가 요소들에 관한 학생들의 인식 여부를 확인하는 간접 평가의 방식에서, 평가 요소와 관련하여 학생들이 얼마만큼 지식을 생성 및 활용하고, 해당 지식에 관한 자신의 사고력, 창의력을 발휘할 수 있는지 확인할 수 있는 직접 평가 방식을 사용할 수 있게 된 것이다.

서울시(2010/2011) 및 경기도 교육청(2007/2009/2010)은 서술형 평가를 통해 과정 중심의 질적 평가 내실화로 수업 방법의 변화를 촉진하고자 하였다. 구체적으로 학생들의 창의력과 표현력을 향상시키기 위한 방안으로 서술형 평가의 도입을 적극 유도하면서 평가 문항 개발과 관련한 매뉴얼을 일선 학교에 제공하여 서술형 평가의 확대를 적극 추진하였다.

윤지원(2008), 조수현(20011)은 학교 현장에서 접하고 있는 서술형 평

가에 대한 현황을 분석하였다. 주로 중학교 국어과 지필평가에서의 서술형 평가 도입에 대해 문항 분석 및 채점 방법, 서술형 평가에 관한 교사와 학생들의 인식 조사 결과를 통해 서술형 평가의 의도와 목적에 대해서는 인정함에도 불구하고 시행에 있어 여러 문제점들을 제기하였다. 기존의 수행평가가 약화되고, 서술형 평가의 문제 수준이 어렵다는 학생들의 인식 결과와 함께 채점의 공정성 문제로 인한 불신이 일어날 수 있다는 우려의 견해를 설문 조사 결과를 통해 제시하였다. 그러나 그보다 중요한 것은 서술형 평가에 대해 학생들이 접근하고 준비할 수 있는 수업 및 교재와 같은 구체적인 교수·학습 방법이 마련되고 학생들에게 제공되지 않은 상황에서 평가를 실시하는 것의 문제점을 지적하고 있다.

한편 박혜상(2008)은 서술형 평가가 지니고 있는 채점자 간 신뢰도 문제를 확인하기 위해 문항반응이론에 근거하여 통계적인 방법으로 국어과 평가 유형에 따른 학생 반응을 문항 난이도 및 변별도, 신뢰도 등을 통해 확인하고자 하였다.

물론 모든 교과가 지필평가에서 서술형 평가를 의무적으로 도입하는 것은 문제가 있다. 국어의 작문 교과와 같은 경우에는 기존까지 작문 교과가 지닌 학생들의 쓰기 능력을 직접적으로 평가할 수 있는 기회가 없었기 때문에 직접 평가 방식의 서술형 문항 유형, 즉 서술형 평가를 실시하는 것이 긍정적이지만, 다른 교과의 교육 내용 및 평가 요소의 성격을 고려하지 않고 일률적으로 서술형 평가를 도입하는 것은 재고할 점이라 판단된다.

또한 국어과 지필평가에서 서술형 평가를 도입함에 있어서도 고려해야 할 점이 있다. 특히 시험 환경과 관련하여 대체로 50분(고등학교의 경우) 내외의 시간에서 선택형 및 서술형 문항 유형을 병행한다는 것으로 상대적으로 학생들에게 시험에 대한 부담과 압박감을 높이는 것이다. 필수적으로 제한된 시험 시간 동안 학생들이 문제를 해결할 수 있는 능력을 고려하여 적절한 문항 수를 결정하는 것이 적합하다.

서술형 평가를 도입할 때 가장 주의해야 할 점은 채점자 간 신뢰도이다.

채점자들이 일관성을 지니면서 채점을 정확하게 수행하느냐가 관건이 된다. 특히 지필평가와 같이 학생들의 성적에 대해 민감한 영향을 끼치는 시험일 수록 평가의 신뢰도 확보는 중요하다. 그러므로 최대한 평가의 신뢰도를 높이는 방향과 함께, 현실적으로 평가 업무 부담을 줄이는 방향에 대해 모색해야 한다.

2. 국어과 서술형 평가의 문항 개발 절차

일반적인 문항 개발의 절차와 관련하여 이종승(2005), 박도순(1999, 2000)은 절차 모형을 구안하여 제시하였다. 이와 함께 김주환(2010), 박순석 외(2011)는 서술형 평가의 문항 제작을 위한 절차를 제시하였다.

이종승(2005)은 표준화 검사 제작 과정을 세부적으로 제시하였다. 검사 및 평가를 개발할 때 먼저 검사 개발의 계획을 수립하며, 계획 수립의 과정 속에는 구체적인 목적을 설정하도록 한다. 검사의 목적 설정이 끝나면 다음으로는 검사 개발의 전반적인 개요를 수립한다. 이 단계에서는 검사의 하위 요인, 검사 문항의 형식, 배열, 검사의 길이, 검사 대상, 검사 결과의 활용 등과 같은 세부 내용들을 사전 검토한다. 문항 개발에서는 검사 내용을 결정하고, 검사 문항의 형식을 확정하며 이에 따라 문항 작성 및 검토 과정이 이루어진다. 개발한 검사 문항은 예비 검사를 통해 문항 곤란도, 문항 변별도, 답지 반응 분포 등을 통해 문항 분석 및 수정·보완의 과정을 거쳐 최종적인 검사 문항을 선정하도록 한다.

본 검사가 공정하게 실시될 수 있도록 표집 대상 선정 방법 및 검사 실시의 절차와 조건, 지시사항 및 유의사항, 검사 시간, 채점 방법 등을 확정하도록 하며, 신뢰도와 타당도를 통한 검사의 기술적 분석을 수행하도록 한다. 최종적으로 검사의 전반적인 과정을 다루고 있는 검사 요강을 작성하도록 한다.

다음으로 박도순(1999, 2000)은 일반적인 문항제작의 절차에 대해 언급

하였다. 첫째로, 목표의 분석·조직·진술이다. 평가 도구를 제작하기 위해서는 기본적으로 교육을 통해 달성시키려는 교육목표가 무엇인지 분석하는 작업이 필요하다. 이를 바탕으로 하여 구체적인 분류 및 조직, 진술하는 과정이 이어진다. 둘째로, 성취기준의 개발이다. 구체적인 교육 목표에 부합하여 학생들이 달성해야 할 능력과 특성의 형태로 진술하는 과정이 필요하다. 셋째, 평가 대상을 선정한다. 교육 목표 및 성취기준과 관련하여 평가할 구체적인 내용을 선정하는 단계이다. 넷째, 평가 장면의 선정이다. 무엇을 평가하겠다는 목표, 성취기준, 평가 내용이 명확해지면 이러한 요소를 측정하는 도구를 제작하기 위해서 어떠한 평가 장면이 적절한지 결정한다. 다섯째, 검사 문항 형태의 결정이다. 평가 장면에서 어떤 형태의 검사 문항을 사용할 것인지 결정해야 한다. 여섯째, 검사 문항을 작성한다. 평가 목표에 따라 구체적인 문항을 작성한다.

지필평가에서 서술형 평가가 도입되면서 학교 현장에서 서술형 평가 문항 제작에 대한 구체적인 절차들이 논의되기 시작하였다. 김주환(2010), 박순석 외(2011)는 국어과 서술형 평가 문항의 제작 과정을 ①평가 목적의 설정, ②교육과정 분석, ③필수 학습 요소의 추출, ④평가 요소의 선정, ⑤문제 장면의 설정, ⑥목표이원분류표 작성, ⑦문항 초안 작성, ⑧문항 검토 컨설팅, ⑨문항 수정 보완, ⑩최종 문항 완성의 단계로 설명하였다.

①평가 목적의 설정으로부터 ⑤문제 장면(평가 장면)의 설정까지는 일반적인 문항 개발의 절차와 동일하며, 단위학교에서는 이원목적분류표(목표이원분류표)에 따라 출제하고자 하는 문항의 범위, 평가 요소, 행동 영역, 난이도를 고려하여 출제의 방향을 구체적으로 설정하도록 하고 있다. 이를 통해 직접적으로 평가 문항을 개발하는 ⑦문항 초안 작성에서는 출제하고자 하는 문항의 1.5배 정도의 문항을 출제하여 최종적으로 적합한 문항을 선정하여 출제하는 방식을 취하도록 하고 있다.

서술형 평가의 문항 개발과 관련하여 가장 중요한 단계인 ⑧문항 검토 컨설팅은 특히 문항 타당도와 신뢰도 확보를 위한 필수 가정으로 평가 요소

의 타당성, 설정된 평가 장면 및 조건의 압축성, 수험생의 반응 유형 예측, 채점 기준의 타당성, 정오답 경계의 논리적 근거, 평가자의 출제 의도와 문항의 완성도 여부, 문항 구성요소의 효율적 전개, 문항의 정확성에 대해 출제자 및 동료 평가자 간의 협의 과정이다. 이를 통해 문항 수정을 통해 최종적인 문항이 선정되는 단계이다. 이후 전체적인 문항의 최종 검토를 통해 ⑨문항 수정의 보완 및 ⑩최종 문항 완성의 단계로 나아간다.

3. 국어과 서술형 평가의 채점 방법 절차

학교 현장에서는 국어과 지필평가에 서술형 평가 유형이 도입되면서 채점의 공정성을 확보하기 위한 방안들이 모색되었다. 서술형 평가는 특히 채점자 간의 신뢰도 확보가 중요하므로 주로 동 학년 동 교과와 교과협의회를 통해 출제한 평가 문항에 대해 평가 및 채점 일정 및 방법을 정하고 최종적으로 학생들의 성적 확인 및 동의와 최종 성적 입력까지 단위학교별로 구체적인 평가 계획을 수립한다.

김주환(2010), 박순석 외(2011)는 서술형 평가 문항의 채점 과정으로 다음과 같이 제시하고 있다. ①평가 목적의 확인, ②목표이원분류표 확인, ③문항 구성요소 확인(채점 준비), ④시뮬레이션(가채점), ⑤채점 실시, ⑥채점 후 협의(사후 협의), ⑦채점 결과의 학생 확인, ⑧채점 종결의 단계를 언급하고 있다.

①평가 목적의 확인, ②목표이원분류표 확인 단계는 평가 문항 개발에서 계획했던 평가의 내용 및 요소들을 점검하는 단계에 속하며, ③문항 구성요소 확인(채점 준비)은 문항의 평가 장면 및 학생들이 작성해야 할 답안에서 중요한 조건, 학생들이 답해야 할 평가 및 채점의 요소, 인정 답안의 범위를 최종적으로 확인하는 단계이다. 평가 문항 개발 단계에서 평가 및 채점자는 학생들이 작성할 답안을 예측했다 하더라도 예측하지 못한 다양한 답안이 나오는 경우가 많다. 그러므로 ④가채점을 통해 학생들의 답안에 관한 반

응의 유형을 파악하고 인정 답안의 범위를 보다 구체적으로 채점 기준에 적용할 수 있다.

⑤채점 실시에서는 문항별로 채점을 진행하며 채점 결과를 정확하게 확인하도록 한다. 또한 이견이 발생할 수 있는 답안의 채점에 대해서는 교과협의회를 통해 최종적으로 결정한다. ⑥채점 후 협의에서는 최종적으로 완성된 채점 기준을 통해 채점 결과를 재검토하는 과정을 반드시 수행하도록 한다. 채점의 공정성을 확보하기 위해 채점이 완료된 후에는 채점 기준에 대한 설명 및 학생 개개인에게 결과를 확인하도록 하고 이상 없으면 학생 성적일람표에 서명을 통해 채점을 마무리하도록 한다. ⑦채점 종결 이후에는 평가의 주요 목적인 교수·학습의 개선에 주목하여 피드백 과정을 통해 학생들의 언어 능력 향상을 위해 돕도록 한다.

III. 연구 방법

1. 연구 대상

국어과 서술형 평가의 문항 개발 및 채점 방법의 실태와 평가가 일반적으로 갖추어야 할 타당도와 신뢰도에 대해 학교 현장에서는 어떻게 확보하고 있는지 탐색하기 위해 연구 대상으로 경기 부천여고 자연계 학생 140명을 대상으로 하였다. 이 학생들은 2011년 11월 15일에 2학기 기말고사(2차 지필평가)를 실시하였다. 그리고 연구의 목적이라는 전제 하에 학생들의 채점 결과를 활용하겠다는 동의를 얻었다. 표 1은 국어과 지필평가에 응시한 학생 및 평가자인 교사의 인원과 연구 기간의 내용이다.

표 1. 연구 대상 및 인원, 기간

	응시자	문항 개발 및 채점자
연구 대상	경기 부천여고 3학년 자연계 학생	3학년 국어(독서)과 교사
인원	4개 반 140명	4명
연구 기간	2011.10.20~11.27	

표 1에 따라 연구 대상인 응시자는 자연계 4개 반 총 140명을 선정하였다. 위 학교를 선정 한 이유는 1, 2, 3학년 전 학생들을 대상으로 하여 국어, 영어, 수학, 과학, 사회 계열의 해당 과목에 대해 지필평가에서 선택형 문항의 총점과 서술형 문항의 총점을 동일하게 50:50으로 설정하여 문항을 출제 하도록 학업성적관리위원회에서 지침을 정했기 때문이다. 서울시 및 경기도에서 서술형의 비중의 30% 내외로 정한 데 비해서 본교에서는 동일한 비율의 점수로 배정함에 따라 각 교과의 담당 교사들은 서술형 평가의 문항 출제에 대해 비교적 많은 준비를 해왔었다. 또한 동일한 비중은 선택형 문항과 서술형 문항의 채점 결과 간의 상관을 살펴보는 데에도 적절한 사례라고 판단된다.

2. 검사 도구

국어과 서술형 평가 문항 개발 및 채점 방법에 관한 탐색 및 평가의 타당도와 신뢰도 확보 여부를 확인하기 위한 검사 도구로는 고등학교 3학년 독서 과목 2학기 2차 지필평가 문항을 활용하였다. 지필평가 문항은 50분 동안 학생들이 응시하면서 선택형 15문항, 서술형 8문항 총 23개 문항으로 구성되어 있다.

이 중에서 서술형 8개 문항은 표 2와 같은 내용으로 최종 문항 개발하였다.

표 2. 지필평가의 서술형 평가 문항

문항	평가 영역	평가 요소	지시문	행동 영역	난이도	배점 (점)									
1	비문학	부름말의 기능 비교 이해	1. <보기>에서 A, B의 밑줄 친 말이 드러내는 기능의 공통점과 차이점을 위 글을 참조하여 조건에 맞게 쓰시오.	적용	중	7									
2	비문학	글 내용 이해 및 적용	2. 오늘날 '21시 40분'을 96각법의 시각으로 표시하십시오.	적용	중	4									
3	고전문학	시어의 상징적 의미 파악	3. <보기>에서 밑줄 친 시어와 유사한 의미를 가지고 있는 시어를 <다>에서 3개 찾아 쓰시오.	지식	하	6									
4	고전문학	작품 내용 파악 및 작품 비교	4. <보기>의 내용을 참고할 때, ㉠에서 화자가 이루고자 하는 바가 무엇인지 쓰고, 이와 같은 의미로 사용된 시어를 <보기>의 시조에서 찾아 쓰시오.	이해	상	6									
5	현대문학	시어의 의미 비교 해석	5. (나)의 ㉠와 상징적 의미가 같은 시어를 (가)에서 3개 찾아 쓰고, 이들의 공통된 상징적 의미를 쓰시오.	적용	하	9									
6	현대문학	시의 주제 분석	6. ㉠와 관련하여 물음에 답하십시오. (1) ㉠가 '나'에게 전하고 싶은 교훈이 직설적으로 드러난 부분을 (나)에서 찾아 쓰시오. (2) (1)과 <보기>에 공통적으로 드러나는 삶의 자세를 쓰시오.	이해	하	7									
7	현대문학	중심제재의 의미 파악	<table border="1"><tr><td></td><td>아들(창섭)</td><td>아버지</td></tr><tr><td>가치관을 상징적으로 드러내는 단어</td><td>(1)</td><td>돌다리</td></tr><tr><td>가치관의 구체적인 내용</td><td>(2)</td><td>전통적인 가치를 계승</td></tr></table>		아들(창섭)	아버지	가치관을 상징적으로 드러내는 단어	(1)	돌다리	가치관의 구체적인 내용	(2)	전통적인 가치를 계승	지식	상	5
	아들(창섭)	아버지													
가치관을 상징적으로 드러내는 단어	(1)	돌다리													
가치관의 구체적인 내용	(2)	전통적인 가치를 계승													
8	고전문학	중심제재의 기능 파악	8. 인물 간의 관계와 사건 전개 과정을 고려할 때, ㉠의 기능을 서술하십시오.	지식	하	6									

표 2는 2차 지필평가에서 서술형 평가로 제시한 8문항의 평가 영역, 평
가 요소, 평가 문항의 지시문, 내용 영역, 행동 영역, 난이도와 관련한 내용을
정리한 자료이다. 1차적으로 출제자 4명의 교사는 평가 범위 및 관련 영역을
고려하였으며, 평가 요소 추출을 통해 우선적으로 선택형 및 서술형 문항을
1.5배수 출제 원칙에 따라 출제하였으며, 그중에서 적합한 문항을 선정 및

협의를 통한 수정·보완의 과정을 거쳐 서술형 평가를 표 2와 같이 8문항으로 출제하였다.

또한 지필평가를 수행하는 과정 및 채점 과정에서 교과협의회를 수시로 실시하였다. 이에 따라 문항 개발 및 채점에 참여한 교사 4명을 대상으로 하여 면담을 실시하였다. 표 3은 서술형 평가에 관한 교사의 면담 항목을 다른 내용이다.

표 3. 서술형 평가에 관한 면담 항목

항목		주요 면담 내용
서술형 평가 문항 개발	출제 전반	1. 현 형태의 선택형/서술형 평가의 비중이 적절한가? 2. 교육과정 및 수업 내용이 적절히 반영되었는가? 3. 학생들의 수준을 고려하여 문항 간의 난이도와 변별도를 적절히 유지하고 있는가? 4. 시험 시간에 비추어 서술형 평가의 문항 수(소문항 포함)가 적절한가?
	문항 유형	1. (지시문) 지시문이 정답 인정 범위가 모호하지 않도록 조건들을 모두 포함하고 있는가? 2. (자료) 분량의 적절성과 문제 해결에 필요한 내용들을 담고 있는가?
채점 과정 및 방법	채점 과정	1. 현 채점 과정이 적절한가? (예시답안 작성→문항별 채점→인정 답안에 관한 협의→채점 검토) 2. 서술형 평가에서의 채점자간 신뢰도(서술형 평가의 신뢰도)는 어떠한가? 3. 서술형 평가의 채점 업무 부담 정도는?
	답안	1. 기본 답안은 객관적인 입장에서 정답으로 인정되는 보편 타당함을 지니고 있는가? 2. 인정 답안의 범위를 설정할 경우 내용과 절차의 측면에서 적절한가?
서술형 평가의 개선 방안		자유 응답

표 3은 독서 과목 2학기 2차 지필 평가의 문항 개발 및 채점 과정에 참여한 교사들을 대상으로 한 서술형 평가에 관한 면담 항목을 정리한 자료이다. 서술형 평가의 문항 개발 및 채점 과정에 대한 경험을 바탕으로 하여 현재 학교에서 실시되고 있는 서술형 평가의 적절성 및 개선 방안에 대해 면담을 실시하였다.

서술형 평가의 문항 개발과 관련하여 출제 전반의 내용에서는 선택형 및 서술형 평가 문항의 비중의 적절성, 지필평가에서의 교육과정 및 수업 내

용 반영 정도, 학생들의 수준을 고려한 난이도와 문항 변별도 여부, 서술형 평가의 적절한 문항 수에 관한 견해를 얻고자 하였다. 문항 유형에서는 서술형 평가가 갖추어야 할 문항의 구성요소 등을 현재 개발한 문항들이 갖추고 있는가에 관해 질의 응답하였다.

채점 과정 및 방법에 대해서는 현재 수행하고 있는 채점 과정의 적절성 및 채점자 간 신뢰도 여부, 채점 업무의 부담 정도, 채점 기준 및 인정 답안의 설정 과정 및 적절성 여부에 대해 견해를 수렴하였다.

마지막으로 서술형 평가의 개선 방안에 관하여 다양한 견해를 얻고자 하였다.

3. 연구 절차 및 분석 방법

연구 절차는 경기 부천여고 2011학년도 독서 과목 2학기 2차 지필평가 문항 출제 및 채점의 과정을 사례로 활용하였다. 세부적인 절차 및 문항 수에 있어서 학교별로 차이가 있더라도 대체로 학교의 문항 출제(개발) 및 채점 방법의 일반적인 절차는 동일하다.

단위 학교의 성적관리규정에 따라 대체로 지필평가의 출제 및 채점, 성적 처리 일정은 한 달 전후에 공고된다. 이에 따라 각 교과별로 출제(평가) 범위에 대한 학생 안내 및 원안지 제출 기간은 학교별로 차이는 있지만 시험 일로부터 대략 2주 전까지 마감이 된다.

지필평가 문항을 개발함에 있어 출제자인 교사 4명은 선택형 및 서술형 문항의 출제자를 따로 지정하지 않고 공동으로 개발하도록 하였다. 특히 평가 요소의 특성에 따라 선택형 및 서술형 문항을 출제하되, 1.5배수에 근거하여 적절한 출제 문항 수를 고려하여 문제를 개발하도록 하였다. 그러므로 대체로 선택형 및 서술형 문항의 출제 절차는 동일한 절차를 진행하였다. 다만 문항의 검토 및 문항 개발 시 고려할 점, 문항 구성 요소, 채점 기준 및 인정 답안, 학생들의 예상되는 답안 반응 등에 있어서 선택형 문항과는 달리

서술형 문항, 특히 서술형 문항의 평가에서는 고려해야 할 요소들을 추가적으로 포함하고 있으므로 이러한 면을 고려하기 위해서는 선택형 문항에 비해 보다 세밀한 문항 개발 및 채점 방법의 단계를 수행하였다.

3학년 독서 과목의 2차 지필 평가의 서술형 평가의 문항 개발의 절차는 다음과 같이 진행되었다.

첫째, 출제 범위 설정, 지필평가 전체의 문항 수를 고려하였다. 출제자인 교사 4명이 각각 선호하는 평가 범위 내에서의 세부 영역을 정하고, 각각 출제할 문항 수를 배정받았다. 또한 지필평가 문항을 개발하는 과정에서는 시험지를 편집하는 역할 또한 중요하므로 이 업무를 고려하여 출제에 있어서는 상대적으로 적은 문항 수를 개발하도록 배려하였다.

둘째, 각 교사별로 출제할 평가 세부 영역 및 문항 수가 결정되면 최소 원안지 제출 1주일 전 정도에 국어과 교사는 1차적으로 문항을 출제한다. 다른 교과에 비해 교과협의를 통해 문항의 적절성, 문항의 수정·보완 과정의 기간이 오래 소요되므로 1.5배수에 해당하는 문항 출제가 다른 교과에 비해 빠르게 진행되는 편이다.

셋째, 적절한 문항 선정 및 완성을 위한 검토 과정이다. 이는 김주환(2010), 박순석(2011)에서 언급한 ‘문항 검토 컨설팅’ 과정에 해당된다. 먼저 1.5배수 문항들을 검토하면서 각각의 문항에 대한 지시문, 보기, 조건 등 문항 구성 요소를 확인하면서 의도하는 평가 요소가 문항을 통해 적절하게 구현되었는지, 학생들이 정해진 시간 내에서 적절한 답안 반응을 보일지, 학생들이 어떤 답안의 내용을 작성할 것이며 어느 정도 범위까지 인정 답안으로 정할지 세밀하게 협의를 한다. 특히 출제의도에 맞는 답안을 학생들로부터 이끌어내기 위해 조건을 어떻게 지정할 것인가가 중요한 요소이다.

또한 서술형 평가는 학생들이 교과에 대해 알고 있는 정보나 지식 등을 활용하여 평가자가 문항을 통해 출제의도를 파악하고 이와 관련한 지시문 및 관련 자료 등의 문항 구성 요소들을 분석하면서 평가 요소와 관련하여 학생들이 설명, 해석하거나 창의적으로 문제를 해결할 수 있는 능력을 측정하

는 평가 유형이다. 그러므로 기존의 선택형 문항의 평가와는 변별되는 특징이 드러나도록 문항을 구성해야 한다.

넷째, 최종적으로 수정·보완된 문항을 중심으로 하여 이원목적분류표를 작성하였다. 선택형 문항 및 서술형 문항, 특히 서술형 평가의 배치 및 배점을 결정하고, 전체 문항에 대한 평가 요소, 행동 영역, 난이도를 작성하면서 전체 문항의 구성 및 적절성을 확인하는 단계이다.

다섯째, 서술형 평가에 관한 채점 기준을 작성하였다. 각 문항별로 기본 답안 및 학생들의 답안 반응의 유형을 고려하여 인정 답안을 제시하였고, 이와 관련하여 정답 및 오답, 부분 점수를 판별하기 위한 구체적인 채점 기준을 작성하였다. 이러한 과정은 교과협의회를 통해 최종적으로 결정하였다.

그러나 이 단계에서의 채점 기준은 1차적인 자료이다. 학생들의 답안을 전수 채점하는 과정에서 학생들은 채점자가 예측하지 못한 다양한 답안의 반응들을 보인다. 채점자들 간에 이견이 발생할 수 있는 답안에 대해서는 1차 채점 후에 교과협의회를 통해 결정을 내리면서 경우에 따라 채점 기준의 인정 답안 및 채점 기준이 수정되기도 한다. 그러므로 최종적인 서술형 문항의 채점 기준은 모든 채점 및 검토가 완료된 후에 수정하여 작성한다.

3학년 독서 과목 2차 지필 평가의 서술형 평가 채점 과정은 다음과 같이 진행되었다. 첫째, 문항 개발 과정을 통해 평가 목적 및 평가 요소, 채점 기준에 대한 정보를 각각의 채점자들은 숙지하였다. 이를 바탕으로 첫째, 출제자가 작성한 문항별로 채점을 실시하였다. 평가 문항을 개발한 평가자는 평가의 목적 및 의도, 학생들에게 요구되는 답안의 내용을 다른 평가자보다 상세하게 파악하고 있으므로 채점을 실시할 때는 문항을 개발한 평가자의 문항을 본인이 직접 채점하는 과정을 진행하였다. 그리고 채점은 문항별로 채점을 실시하였다. 일반적으로 채점은 문항별로 수행하는 것이, 일관된 기준으로 채점하기에 유리하며 채점 과정에서 상대적으로 서술형 평가 전체 문항에 대해 학생들이 작성한 답안을 한꺼번에 채점하는 것보다 채점자의 오류를 감소시키는 효과가 있다. 이러한 과정은 채점의 첫 단계에 해당되므로 1

차 채점이라 한다.

둘째, 1차 채점이 끝난 뒤에는 교과협의(채점 협의)를 실시하였다. 1차 채점을 통해 각각의 채점자들은 기존에 작성한 채점 기준의 범위에서 벗어나는 다양한 답안 반응 유형들을 수집하여 정리한다. 이를 바탕으로 이원목적분류표 및 채점기준표의 평가 목적, 의도, 평가 요소, 기본 답안 및 인정 답안의 내용들과 대조하여 답안의 적절성 유무를 판단한다. 이를 통해 최종적으로 서술형 평가의 채점 기준을 수정·보완하도록 한다.

셋째, 2차 채점을 실시하였다. 1차 채점에서는 문항별로 각각의 출제자가 채점을 수행했다면, 2차 채점에서는 각 교사들이 담당한 반별로 학생 답안을 다시 분류하여 채점 협의를 통해 최종적으로 수정·보완한 채점 기준에 서술형 평가가 정확하게 채점되었는지 확인하는 절차이다. 이러한 절차를 통해 평가의 신뢰도를 확보할 수 있다. 2차 채점 과정에서 1차 채점한 결과에 다른 견해가 발생할 수 있다. 이런 경우에는 이원목적분류표와 서술형 채점 기준을 바탕으로 하여 1차 채점을 담당한 교사와 협의를 하여 최종적으로 채점의 변동 유무를 확인하도록 한다. 최종적으로 채점의 검토가 완료되었을 경우에는 각각의 학생 답안별로 학생이 획득한 총점을 확인하여 기록하는 과정을 거치도록 한다.

넷째, 채점 결과의 학생 확인 과정을 수행하였다. 문항 개발 및 채점자인 교사들은 담당 학급의 수업 시간을 통해 서술형 평가의 출제 의도, 평가 요소 및 채점 기준을 상세하게 설명하였다. 이러한 과정은 궁극적으로 학생에 대한 피드백 과정으로 학생들의 학습 능력에 대한 개선 및 향상을 유도한다. 그리고 학생 개개인의 채점 결과에 대해 이상이 없는지 확인하는 과정을 거치도록 한다. 이상이 없을 경우에 최종적으로 학생 성적일람표에 개인 서명을 통해 확인을 받도록 한다. 이러한 과정은 서술형 평가의 신뢰도를 확보하기 위한 방법이다.

국어과 서술형 평가의 문항 개발 및 채점 방법에 있어 평가의 타당도와 신뢰도를 확인하기 위해서 통계적인 방법을 사용하였다. 특히 문항 변별도

와 구인 타당도에 주목하여, 선택형 문항과 서술형 문항의 정답률 비교표를 작성하여 빈도분석 및 기술통계의 방법을 통해 정답 비율을 확인하였다. 이는 각 문항별 난이도 및 변별도를 확인하는 과정이다. 또한 선택형 문항 및 서술형 문항의 채점 결과를 바탕으로 하여 상관분석을 실시하였다. 이는 각각의 평가 유형이 독립적으로 의도하는 평가의 목적을 수행하고 있는지 선택형 문항 및 서술형 문항 유형 간에는 어떠한 상관을 가지고 있으며, 이러한 상관계수가 의미하는 바가 무엇인지 확인하기 위한 과정이다.

또한 학생들의 평가 결과를 바탕으로 하여 현재 학업성적관리 규정에 따라 9등급으로 산출되는 것을 고려하여 성적을 분류하였다. 그리하여 상위 1~3등급의 집단과 하위 7~9등급의 집단 간에 선택형 및 서술형 문항 유형의 평가 결과가 어떠한 차이를 보이는지 살펴보고자 t검증을 실시하였다. 전반적인 통계적 분석을 위해 PASW 18.0 프로그램을 사용하였다.

마지막으로 서술형 평가의 문항 개발 및 채점 과정에 참여한 교사들을 대상으로 교과협의를 통해 면담을 실시하고 그 결과를 분석하였다. 특히 서술형 평가 유형의 적절성 및 개선 방안을 중심으로 교사들의 견해에 주목하였다.

IV. 결과 분석 및 논의

1. 국어과 서술형 평가의 타당도 및 문항 변별도, 신뢰도 확보

평가가 일반적으로 갖추어야 할 중요한 요소는 타당도와 신뢰도가 있다. 국어과 서술형 평가 문항 개발 및 채점 방법의 적절성을 확인하기 위해서는 평가의 타당도와 신뢰도가 어느 정도 확보되고 있는지 검증할 필요가 있다. 평가의 타당도와 관련하여 이 연구에서는 내용 타당도 및 구인 타당도

에 주목하였다. 먼저 전문가의 검토를 통해 문항 개발 및 채점 과정의 적절성을 확인하는 과정을 통해 내용 타당도를 확보하였다.

이 연구에서는 일반적인 문항 개발의 단계 및 국어과 서술형 평가를 위한 적절한 문항 개발 및 채점 과정에 비추어 독서과 3학년 2학기 2차 지필평가의 각 과정이 적절하게 실행되었는지 점검하는 데 초점을 두었다. 대체로 교사들은 적합한 문항 개발 및 채점 과정을 수행하였으므로 내용 타당도를 확보하고 있다고 할 수 있다.

그럼에도 불구하고 일반적인 기준에서 벗어난 몇몇 문항 개발의 세부 절차들이 발견되었다. 먼저 최종적으로 개발한 문항을 분석한 결과, 대체로 서술형 평가가 의도하는 목적에 적합한 문항들이 출제되었지만 일부 문항에서는 선택형 문항과 변별을 갖추지 못한 문항들이 발견되었다.

1번 문항은 문법과 관련하여 상대방을 부르는 기능을 수행하는 부름말의 개념을 인식하고 적용하는 문제이다. 지시문에 따르면 <보기>에서 제시한 두 개의 사례에 대해 부름말의 개념에서 볼 때 공통점과 차이점을 서술하도록 요구함으로써 학생들은 개념에 대한 인식 및 사례에 적용을 통해 공통점 및 차이점의 특성을 어떻게 변별하여 기술하는가에 초점을 두고 있다. 이는 지식을 단순히 확인하는 것에서 발전하여 자신이 알고 있는 지식이나 정보를 적절히 분류하고 적용하는 데 평가의 목적을 두고 있으므로 적절하다.

2번 문항은 과거 조선 시대의 시각을 표시하는 방법을 평가 요소로 포함하고 있다. 문항이 의도하는 평가의 목적은 제시된 자료를 읽고 적절하게 주어진 조건이나 상황에 적용할 수 있는 능력을 평가하는 것이나 문항의 유형을 선택형 문항으로 바꾸었을 경우를 고려해 볼 때 변별되는 특징을 갖추고 있지 못하다. 원래 문항 개발 과정에서 협의를 통해 적용하기 위한 근거가 되는 부분을 찾고 이를 적용하는 과정을 출제함으로써 학생들이 제시문을 통해 획득한 지식을 적용하는 과정을 학생 답안에 반영하여 요구하도록 하였으나 학생들의 시험 응시 시간을 고려하여 과정 단계를 생략하도록 최종적으로 결정하였다.

3번 문항은 <보기>의 밑줄 친 시어의 문맥적 의미와 유사한 의미를 갖고 있는 시어를 제시문의 3편에 시 중에서 찾는 문항이다. 이 문항 역시 서술형 평가가 의도하는 목적에 따르면 <보기>에 제시된 시어가 가지고 있는 문맥적 의미가 무엇인지 학생들이 인식한 후에 이와 관련된 의미를 가진 시어를 찾는 과정이 순차적으로 진행되어야 하나 중간 과정이 생략되었다. 이 문항 역시 2번 문항과 동일하게 협의 과정을 통해 응시 시간을 고려하여 중간 단계를 생략하도록 하였다.

4번 문항은 2개의 세부 문항으로 구성되어 있다. 고전 시가의 작품 속에서 화자의 태도와 관련지어 상징적 의미를 가지고 있는 단어의 의미를 파악하여 서술하고, 이러한 태도를 갖추고 있는 단어를 다른 시에 적용하여 해당되는 부분을 찾는 문항이다. 이는 화자의 구체적인 태도 및 화자가 이루고자 하는 소망의 구체적인 내용을 파악한 후에 다른 작품에 적용하는 단계를 거치고 있으므로 서술형 평가에 적합한 문항이다.

5번 문항은 시어의 상징적 의미가 유사한 단어를 찾고 이들의 공통적 의미를 서술하는 방식으로 문항이 구성되어 있다. 그러나 순서가 적절하지 않다고 판단된다. 시어의 상징적 의미가 무엇인지 먼저 파악한 후에 이와 유사한 의미를 지닌 단어를 찾는 방식이 적절하다. 그런 면에서 세부 문항의 순서가 뒤바뀌어 전개되고 있다고 파악할 수 있다.

6번 문항은 작품을 통해 작가가 전달하고자 하는 교훈의 내용을 서술하고, 이를 다른 작품에 적용하여 공통적으로 올바르게 지향하는 삶의 자세에 대해 서술하는 방식으로 문항이 구성되어 있다. 이는 작품이 전달하고자 하는 교훈, 즉 주제를 파악하고 이를 다른 작품에 적용하는 과정을 수행하도록 학생들에게 요구하고 있는 문항이다.

7번 문항은 소설에서 대립적인 가치관을 가지고 있는 인물들의 성격과 가치관의 차이를 상징적으로 드러내는 부분을 조건에 맞게 작성하는 문항이다. 소설에서 갈등을 이루고 있는 인물들 간에 대립을 상징적으로 보여 주는 부분을 찾게 하고 이를 통해 드러나는 인물의 성격 및 가치관을 서술하는 단

계로 구성되어 있다. 7번 문항 역시 사고의 과정을 확인한다는 측면에서 서술형 문항으로 적합하다.

8번 문항은 소설에서 인물 간의 관계와 사건 전개를 고려할 때 주요 용어가 가지고 있는 소설에서의 기능 및 역할을 묻는 문항이다. 그러나 이 문항은 선다형으로 바꾸었을 때 변별되지 않을 가능성이 높다. 처음에 의도했던 문항의 개발 방향은 소설에서의 인물 간의 관계와 사건 전개를 파악하도록 <보기>의 문제 장면을 제공해 주고 이와 관련 있는 주요 용어의 기능과 역할을 제시문의 내용을 참조로 하여 작성하도록 요구하였지만, 다른 문항과 마찬가지로 시험의 상황 및 조건을 고려하여 중간 과정을 생략하였다.

문항 검토의 컨설팅 과정을 통해 서술형 평가의 8개 문항 중에서 일부 문항들은 학생들의 사고의 과정 및 자신의 지식을 생성하거나 활용하는 중간 과정을 생략한 것이 발견되었다. 그러나 대부분의 문항들이 서술형 평가가 의도하는 평가의 목적에서 볼 때 적절한 문항으로 출제되어 평가의 내용 타당도를 확보하고 있다.

한편 서술형 평가 문항의 대부분이 이원목적분류표와 대조해 보았을 때 대체로 ‘적용’ 항목에 속하는 문항들이 많았다. 물론 ‘지식’이나 ‘이해’의 성격을 지닌 문항들이 세부 문항으로 포함되기도 하였지만 적용 중심의 문항이라는 점, 구체적으로 시의 경우에는 시어가 상징하는 의미를 서술하고 이와 관련을 맺을 수 있는 시어를 다른 작품에서 찾는 방식의 문항이, 즉 평가 요소가 비슷한 내용을 담고 있는 문항이 공통적으로 많이 출제되었다는 점은 자칫 문항의 변별도를 저해할 수 있는 요인이라고 판단된다.

상대적으로 선택형 및 서술형 평가 문항에 따른 평가 결과가 각각 50%로 대등한 상황에서 필수적으로 일정 문항 이상의 서술형 평가 문항을 출제하다 보니 서술형 문항의 본래 취지에 맞는 한 편의 글을 요구하는 문항을 출제하기에는 어려움이 있었다. 그리하여 대부분 서술형 평가 문항들은 선택형 문항을 해결하는 데 요구되는 시간보다는 오래 걸릴 것을 예상하면서도 비교적 1~3개 내외의 문장을 작성하는 범위 내에서 학생들의 답안

반응 유형을 유도하여 문항을 출제하였다.

각각의 서술형 평가 문항 분석과 함께 또한 주목할 만한 점으로는 ‘평가 목적의 설정’, ‘교육과정 분석’, ‘필수 학습 요소의 추출’, ‘평가 요소의 선정’, ‘평가 장면의 설정’이 단순히 출제 범위 설정, 지필평가 전체의 문항 수 배정의 절차로 축소화된 것이다. 뿐만 아니라 이원목적분류표의 작성이 문항 초안을 작성하기 전의 단계가 아닌 문항을 최종적으로 보완·수정 및 선정한 후에 이루어졌다는 점이다.

학교 현장에서는 평가 문항을 개발하는 데 있어 간과하고 있는 중요한 문제점의 하나는 평가 요소를 추출하는 단계를 정밀하게 거치지 못하고 있다는 것이다. 단순히 출제 범위를 정하고 이에 따라 출제하는 교사별로 범위를 나눈 후에 각각 출제할 평가 문항의 수를 정하는 수준에 머무르기 때문에 1차적으로 출제한 교사들의 문항들을 검토하다보면 평가 문항에서 의도하는 내용이 비슷한 유형일 경우가 많다. 그러다 보면 문항의 변별도를 낮추게 되는 요인이 될 수 있다. 또한 이는 궁극적으로 문항의 타당도를 저해하는 주요인이 될 수 있다.

실제 학교 현장에서의 이원목적분류표 작성은 최종적으로 개발한 문항이 어떠한 평가 요소 및 내용 및 행동의 특성을 갖추고 있는지 확인하면서 난이도에 따라 최종적으로 배점을 정하는 문항의 적절성을 확인하는 단계의 성격을 지니고 있다. 그러나 엄밀하게 이원목적분류표가 제 기능을 발휘하기 위해서는 평가 요소를 추출하는 과정에서 긴밀히 활용되어야 한다. 그래야만 평가를 통해 의도하는 평가 요소를 추출하고 이에 적합하고 다양한 평가 문항을 개발할 수 있기 때문이다.

다음으로 문항 변별도 및 구인 타당도에 주목하였다. 문항 변별도는 선택형 및 서술형의 정답률 비교를 통해 문항별로 난이도를 확인할 수 있었으며, 구인 타당도는 선택형 및 서술형 문항의 평가 결과에 관한 상관분석과 평가 결과에 따른 상·하위 집단 간에 문항 유형별로 평가 결과가 차이를 보이고 있는가를 통계적으로 검증하는 방법을 활용하였다.

독서 과목 지필평가 결과를 통해 선택형 문항에 대한 반별 및 총 정답률 비교표를 조사하였다. 표 4는 독서 과목 지필평가 선택형 문항의 정답률 비교표이다.

표 4. 선택형 문항의 정답률 비교표

(단위:%)

문항 번호 \ 반	10	11	12	13	평균
1	100	94.74	100	100	98.7
2	88.89	84.21	89.47	86.49	87.3
3	37.04	52.63	44.74	59.46	48.5
4	40.74	44.74	42.11	37.84	41.4
5	77.78	60.53	55.26	75.68	67.3
6	62.96	65.79	63.16	67.57	64.9
7	77.78	86.84	86.84	94.59	86.5
8	77.78	63.16	65.79	56.76	65.9
9	51.85	68.42	81.58	70.27	68.0
10	77.78	76.32	86.84	94.59	83.9
11	77.78	76.32	84.21	91.89	82.6
12	77.78	76.32	86.84	81.08	80.5
13	66.67	86.84	78.95	75.68	77.0
14	96.30	84.21	89.47	94.59	91.1
15	96.30	89.47	94.74	91.89	93.1

표 4는 선택형 15문항에 대해 반별 및 총 평균 정답률 비교표의 자료이다. 학교의 상황에 따라 다르기는 하나 지필평가를 실시한 후에 결과 분석을 교과별로 실시하기도 한다. 결과 분석의 내용 중에서는 각 문항별로 난이도 및 변별도를 고려하여 정답률 10% 미만 및 90% 이상에 해당되는 문항에 대한 원인을 분석하기도 한다. 표 4에 따르면 정답률 90%이상의 문항이 1, 14, 15번 문항으로 총 3문항이 해당된다. 정답률이 높을수록 학생들의 능력을 측정하는 문항 변별도는 낮아진다. 이에 대한 원인으로는 평가 요소, 평가 문

항의 진술 방식, 평가 문항의 구성 요소들을 면밀히 살펴보면서 문제점을 파악할 수 있다.

다음으로 표 5는 서술형 문항의 문항별 정답률 비교표의 자료이다.

표 5. 서술형 평가 문항의 정답률 비교표 (단위:%)

문항 \ 배점(척도)	0	2	3	4	5	6	7	9	정답률
1	5.0	3.6	12.9	2.1	6.4	12.1	57.9	-	57.9
2	97.7	-	-	9.3	-	-	-	-	9.3
3	7.1	7.9	-	15.0	-	70.0	-	-	70.0
4	12.9	-	31.4	-	-	55.7	-	-	55.7
5	12.1	0.7	-	1.4	7.9	24.3	47.9	-	47.9
6	10.0	3.6	25.7	-	0.7	6.4	53.6	-	53.6
7	6.4	35.7	-	-	57.9	-	-	-	57.9
8	28.6	-	-	-	1.4	70.0	-	-	70.0

표 5는 서술형 평가의 정답률 비교표 내용이다. 문항별로 배점에 따라 정답률을 나타내고 있다. 선택형 평가와는 달리 서술형 평가에서는 90% 이상의 문항은 발견되지 않았다. 그러나 10% 미만의 정답률을 나타낸 문항 2가 발견되었다. 이는 학생들의 답안 반응을 고려하지 못한 난이도가 높은 문항이므로 문항의 난이도가 좀 더 고려될 필요가 있다.

주목할 만한 특징은 서술형 평가의 부분 점수에 관한 척도가 세분화될수록 문항의 난이도는 높아졌다. 서술형 평가 1번은 7개의 척도, 2번과 3번은 6개의 척도가 적용됨에 따라 다양한 배점별 분포를 보여 주고 있다. 이는 서술형 평가의 문항 변별도를 높이기 위해서는 문항 구성 요소의 ‘조건’ 항목 및 채점 기준을 엄밀하게 적용하여 척도를 세분화하는 방법이 적합하다는 것을 뒷받침하는 결과이다.

서술형 평가의 구인 타당도 확보 여부를 알아보기 위해 선다형 문항과 서술형 문항 채점 결과 간의 상관분석을 실시하였다. 본래 구인 타당도를 확

인하기 위해서는 평가 요소 및 평가 문항 간의 요인분석을 통해 추출하는 방법이 있으나 이 연구에서는 각각 평가 유형별 채점 결과 간의 상관분석을 통해 선다형 문항과 서술형 문항 간의 어떤 상관을 나타내는지에 주목하였다.

표 6은 선택형 및 서술형 문항의 평가 결과를 상관분석하여 제시한 자료이다.

표 6. 선택형 및 서술형 문항 평가 결과의 상관분석

pearson 상관분석	선택형 총점	서술형 총점	총점
선택형 총점	1		
서술형 총점	.561	1	
총점	.894	.873	1

표 6은 선택형 문항의 평가 결과, 서술형 문항의 평가 결과, 전체 평가 결과의 변인 간의 상관분석을 실시한 결과이다. .01 수준에서 유의확률 안에 속하였으며, 총점 변인을 기준으로 총점과 선택형 총점(.894), 총점수와 서술형 총점(.873) 간의 상관은 매우 높게 나타났으며, 주목되는 선택형 총점과 서술형 총점 간(.561)은 상관이 어느 정도 있게 나타났다.

서술형 문항과 선택형 문항의 평가 결과 간 상관분석 결과를 살펴보면, 어느 정도 상관이 있다는 것은 대체로 동일한 시험 범위와 유사한 평가 요소를 선택형 문항과 서술형 문항이 공유하기 때문이라고 분석할 수 있다. 이는 엄밀하게 말하면 선택형 문항 및 서술형 문항이 각각 변별적이면서 독립적인 평가의 역할을 수행하기 보다는 상호 간에 영향관계에 있다고 말할 수 있다.

이를 문항 개발의 특징과 관련지어 본다면 문항 개발의 초기 단계에서 실제 사례에서는 수행되어야 할 이원목적분류표에 따른 평가 요소 분류가 세밀하게 이루어지 못한 점을 문제점으로 지적할 수 있다. 평가 요소, 행동 영역, 문항의 난이도를 고려하여 선택형 평가 및 서술형 평가가 지필평가에

서 수행해야 할 역할을 구분할 필요가 있다고 판단된다.

그럼에도 불구하고 .561의 상관계수는 선택형 및 서술형 문항이 지필평
가 안에서 나름대로 독립적인 역할을 수행하고 있다고 판단할 수 있는 가능
성의 여지를 주고 있다. 이러한 해석이 가능한 근거로는 각각의 평가 유형에
따른 채점 결과가 상·하위권 학생들 간의 성적에서 유의한 차이를 드러내는
지 여부에 따라 확인이 가능하다. 표 7은 상·하위권 학생들의 선다형 및 서
술형 평가 결과 차이를 분석한 자료이다.

표 7. 선택형 및 서술형 문항의 평가 결과 차이 분석

문항 유형	집단	사례수	평균	표준편차	t	p
선택형	상위 집단	32	46.55	3.76	10.42	.000
	하위 집단	32	24.16	9.23	10.42	.000
서술형	상위 집단	32	42.0	4.22	12.73	.000
	하위 집단	32	24.5	6.52	12.73	.000

표 7은 지필평가 총점을 기준으로 하여 상·하위 집단별 선택형 및 서술
형 문항의 평가 결과 차이를 분석한 독립표본 t검증 결과이다. 대체로 현재
학업성적관리 규정에 따르면 지필평가 결과는 9등급으로 환산하도록 하고
있다. 1등급(4%), 2등급(7%), 3등급(12%), 4등급(17%), 5등급(20%), 6등급
(17%), 7등급(12%), 8등급(7%), 9등급(4%)의 상대평가로 평가 결과를 등급
으로 분류하고 있다. 이 중에서 지필평가 총점을 기준으로 하여 비교적 성적
이 높은 1~3등급에 해당되는 학생들을 상위 집단으로 편성하고, 성적이 낮
은 7~9등급에 해당되는 학생들을 하위 집단으로 편성하였다.

평균 차이를 분석한 결과에 따르면, .05 수준에서 유의한 차이를 보이고
있다. 선택형 문항의 평가에 따르면 상위 및 하위 집단의 평가 결과는 22.39
점의 차이를 나타냈으며, 서술형 문항의 평가에 따르면 상위 및 하위 집단
의 평가 결과는 17.5점의 차이를 보여주고 있어 선택형 및 서술형 문항의 평
가 결과는 모두 상·하위 집단 간에 유의한 차이를 보이고 있었다. 이러한 결

과는 선택형 및 서술형 문항이 비록 평가 요소들을 공유하고 있지만 지필평가에서 어느 정도 독립된 역할을 수행하면서도 상위 및 하위 집단 간의 문항 변별도를 확보하고 있다고 판단할 수 있다.

평가의 신뢰도와 관련하여 최근 채점자간 신뢰도를 확보하기 위한 다양한 통계적 방법들이 활용되고 있다. Crobach α 및 일반화가능도 이론(김성숙 외, 2001), 문항반응이론에 기반한 Rasch 분석(장소영 외, 2009)들은 기존의 고전검사 이론에 다루지 못한 평가 문항, 평가자 변인의 변동에 신뢰도의 변화 여부, 채점자 간의 신뢰도 여부 및 확률에 근거하여 비신뢰성 및 편향을 보이는 사례 추출 등의 진전된 신뢰도 검증 방법을 갖추고 있다.

이 연구에서는 평가의 신뢰도, 특히 채점자 간의 신뢰도 검증을 위한 통계적 분석 단계를 수행하지 않았다. 1차 채점에서 교사들은 학생들 답안에 대한 전수 채점이 아닌 문항별 채점을 실시하였고, 2차 채점에서는 채점 결과를 담당 교사가 협의를 통해 검토하는 과정을 진행하였기 때문에 채점자 간에 점수 결과가 어떠한 차이가 나타나는가를 확인할 수는 없었다. 주로 서술형 평가의 채점 과정에서는 교과협의를 통해 채점자들 간의 평가관의 차이를 조절하고 공통적인 채점 결과를 산출하는 과정에 주목하였기 때문에 개별적으로 채점자 간의 신뢰도 차이가 어떻게 나타나는지 분석하기는 어려웠다.

그럼에도 불구하고 국어과 서술형 평가의 채점 과정은 평가의 신뢰도, 특히 채점자 간 신뢰도를 높이기 위한 방안들이 포함되어 있었다. 출제자이자 채점자인 교사들은 자신들이 각각 개발한 평가 문항별로 1차 채점을 수행하였다. 1차 채점 결과를 바탕으로 하여 교과협의를 통해 채점 기준의 세부 내용을 수정 및 확정하였다. 2차 채점은 담당하는 반별로 검토를 진행하였다. 특히 다양한 학생 답안 반응에 관하여 정답의 인정 여부를 확인하는 교과협의회를 통해 최종적으로 성적을 확정하였다. 그리고 완료된 채점 결과를 각각의 학생들에게 설명하고 채점 결과를 확인시키는 절차를 수행하였다. 이러한 세부 절차들을 통해 평가의 신뢰도를 확보하고 있다고 판단할 수 있었다.

2. 국어과 서술형 평가에 관한 교사들의 인식

국어과 지필평가에서 서술형 평가 도입과 관련하여 지필평가의 평가 문항 개발 및 채점을 수행한 3학년 국어 교사 4명¹과 함께 교과협의회를 실시하고 서술형 평가에 관한 교사들의 인식에 대해 면담을 수행하였다.

서술형 평가의 문항 개발과 관련하여 출제와 관련하여 대체로 교사들은 장기적인 관점에서 지향해야 할 가치라고 판단되나 학생들의 입장에서는 평가 결과의 50%를 차지하는 서술형 평가의 비중에 대해 다소 부담이 될 것이라고 견해를 밝히고 있었다.

지필 평가의 교육과정 및 수업 내용 반영 여부, 문항 개발에 있어 난이도와 문항 변별도 확보 여부에 관해서는 고등학교 3학년인 상황에서 입시를 준비하는 과정에서 단위 학교의 교과 교육과정을 계획할 때 교육과정에서 의도하는 내용과 벗어나는 경우가 많은 것이 사실이라고 언급하고 있다. 그러나 특히 서술형 평가 문항에 대해서는 학생들이 작성해야 하는 답안의 분량, 답안 반응의 유형 등을 고려하면서 난이도와 문항 변별도를 높이고자 문항 개발을 위한 교과협의회를 통해 문항의 수준을 세부적으로 협의하고 있다고 응답하고 있다.

시험 시간에 비추어 서술형 평가 문항 수의 적절성에 관해서는 현재 학교에서 개발하고 있는 서술형 문항의 유형에 따른다면 적절하다고 생각되나 일부 교사(김 교사)는 실제로 학생들이 응시할 때 시험 응시 시간이 부족하다고 학생들이 반응을 보이고 있다는 견해를 제시하였다.

1 면담의 참여한 교사는 4명으로 한 ○○ 교사(경력 15년), 오 ○○ 교사(경력 5년), 김 ○○ 교사(경력 13년), 이 ○○ 교사(경력 11년)이다. 본교는 인문 계열과 자연 계열의 3학년 독서 과목 이수 시간이 다른 관계로 계열을 분리하여 출제하고 있으나 수업 내용이 공통적이므로 공동 출제를 원칙으로 하고 있다. 오 ○○ 교사를 제외한 나머지 교사들은 3학년 담임 및 독서 등의 국어과 심화 과목을 가르친 경험이 있다. 교과협의회는 문항 개발 및 채점 과정에서 수시로 진행되었으며 본 면담지는 최종적으로 학생들의 성적 확인이 끝난 후에 제공하였다.

서술형 평가 문항의 유형과 관련해서 문항 구성 요소의 적절성에 대해서는 특히 서술형 평가는 구체적인 조건을 제시하여 답안 작성 및 채점에서의 모호성을 줄이기 위해 노력하고 있으며, 학생들의 문제 해결 과정이 드러나도록 문항 구성 요소를 고려하여 문항에 반영하고 있다는 견해를 나타내고 있다.

다음으로 서술형 평가의 채점 과정 및 방법과 관련하여 현재 수행하고 있는 채점 과정의 적절성, 채점자 간 신뢰도, 채점자의 평가 업무 부담 정도에 대해서는, 각각 출제한 문항들에 대해 교사들이 기본 답안 및 인정 답안을 작성하며 출제한 문항별로 채점하는 과정을 거치고 있고, 수시로 교과협의를 통해 학생들의 다양한 답안 반응에 대해 정답 인정 여부를 협의하고 있으므로 최대한 채점자 간 신뢰도를 높이고 있다고 밝히고 있다. 그러나 모든 교사들이 이러한 채점 과정에 따른 평가 업무 부담이 높다고 언급하고 있다.

서술형 평가 문항의 답안과 관련하여 기본 답안의 정확성 및 인정 답안의 범위에 대해서는, 교사들은 대체로 보편타당한 수준으로 기본 답안을 구성하고자 노력하고 있다고 밝혔다. 일부 교사의 견해에 따르면(한 교사) 종종 교사 개인마다 평가관의 차이가 있는 관계로 의견의 차이를 보이고, 경우에 따라서 인정 답안의 범위와 관련하여 교사들 간의 갈등을 겪는 경우가 있다고 말하고 있다.

서술형 평가의 필요성과 적절성, 개선 방안에 대해서 대부분의 교사들은 서술형 평가의 필요성은 인정하나 교사의 자율적인 평가권을 확보한다는 측면과 학생들의 수행 과정을 살펴보기 위해서는 서술형 평가보다는 수행평가에서 이러한 서술형 문항 유형이 활용되는 것이 적절하다고 언급하고 있다. 특히 몇몇 교사는(한 교사, 이 교사) 서술형 평가는 단순히 지식을 확인하는 문항 형태에서 벗어나 학생들의 개방적이고 창의적인 사고를 평가할 수 있도록 출제되어야 한다고 언급하고 있다.

IV. 결론

이 연구는 국어과 서술형 평가 도입과 관련하여 학교 현장에서 교사들은 문항 개발과 채점 방법에서의 평가의 타당도와 신뢰도를 어떻게 확보하고 있으며 서술형 평가에 관한 국어과 교사들의 견해가 무엇인지 사례를 통해 살펴봄으로써 필요성 및 개선 방안에 관해 고찰하였다.

국어과 서술형 평가의 이론적 배경과 관련하여 평가 방식과 평가 문항 유형에 관한 이론, 선택형 및 서술형 문항의 특성, 국어과 서술형 평가의 문항 개발 및 채점 방법의 절차에 관한 선행 연구들을 살펴보았다.

이 연구는 2011년 11월 15일 실시한 2학기 독서 과목 지필평가에 주목하면서 학교 현장에서의 국어과 교사들이 실제로 국어과 서술형 문항 개발과 채점 과정의 절차를 사례로 활용하였다.

연구 대상으로는 부천여고 자연계 4개 반 140명을 대상으로 하고, 3학년 독서과 교사 4명을 문항 개발 및 채점 과정의 대상으로 선정하였다. 검사 도구로는 교사들이 개발한 지필평가의 서술형 평가 8문항을 활용하였으며, 국어과 서술형 평가의 필요성 및 개선 방안에 관한 국어과 교사의 견해들을 수렴하기 위해 교과협의회를 통한 면담을 실시하고자 질문지를 작성하였다.

연구 절차로는 국어과 교사들의 문항 개발 및 채점 과정에 주목하였다. 문항 개발과 관련하여 첫째, 출제 범위 설정, 지필평가 전체의 문항 수를 결정하였다. 둘째, 각 교사별로 출제할 평가 세부 영역 및 문항 수가 할당되고 1차적으로 선택형 및 서술형 문항을 1.5배수 범위 내에서 출제하였다. 셋째, 적절한 문항 선정 및 완성을 위한 검토 과정을 수행하였다. 넷째, 최종적으로 수정·보완된 문항을 중심으로 이원목적분류표를 작성하였다. 다섯째, 서술형 평가 채점을 위한 채점 기준을 작성하였다.

채점 과정에서 첫째, 출제자가 개발한 문항별로 채점을 실시하였다. 둘째, 1차 채점이 끝난 뒤에는 교과협의를 실시하였다. 셋째, 2차 채점을 실시

하였다. 1차 채점이 문항별로 각각의 출제자가 자신이 개발한 문항에 대해 채점을 수행하였다면, 2차 채점에서는 각 교사들이 담당한 반별로 학생 답안을 다시 분류하여 교과협의를 통해 최종적으로 수정·보완한 채점 기준에 서술형 평가가 정확하게 채점되었는지 확인하였다. 넷째, 채점 결과의 학생 확인 과정을 수행하였다. 이를 통해 서술형 평가의 신뢰도를 확보하고자 하였다.

국어과 서술형 평가의 타당도 및 문항 변별도, 신뢰도를 파악하고자 하였다. 타당도와 관련하여 내용 타당도에서는 대체로 일반적인 문항 개발 절차에 따라 서술형 평가를 개발하면서 교과협의를 통한 문항 검토 및 수정·보완의 과정을 수행함으로써 확보했다고 판단할 수 있다. 그럼에도 불구하고 일부 문항에서는 학생들의 사고 과정을 요구하는 문항 구성에서 지식을 생성하거나 활용하는 중간 과정이 생략된 문항들이 발견되었다. 또한 연구 대상인 학교가 실시하고 있는 지필평가의 특성상 선택형 및 서술형 문항의 평가 결과가 대등한 상황에서는 학생들의 시험 응시 상황 및 조건을 고려하였을 때 비교적 1~3개 내외의 문장을 작성하는 수준에서 학생들의 답안 작성을 유도하고 있는 것을 발견할 수 있었다.

문항 변별도와 관련하여 선택형 및 서술형 문항의 정오답표를 분석하였다. 그 결과 선택형에 비하여 서술형 문항에서는 난이도를 조절하지 못한 문항은 발견되지 않아 문항의 변별력을 갖추고 있음을 확인할 수 있었다.

이와 함께 선택형 및 서술형 평가 결과 간의 상관분석 및 평가 결과를 바탕으로 하여 상·하위 집단의 서술형 및 선택형 평가 결과의 차이를 분석하였다. 상관분석 결과는 .561로 상관이 있는 것으로 나타났다. 이는 선택형 및 서술형 문항이 공통적인 평가 요소를 공유한다는 점에서 상관이 나타난다고 할 수 있지만 한편으로 각각의 문항 유형들이 지필평가의 틀 안에서 독립적인 역할을 수행하고 있음을 파악할 수 있었다. 이에 대한 근거로서 선택형 및 서술형 문항의 평가 결과가 상·하위 집단 간에 각각 유의한 차이를 보이고 있는 것을 제시할 수 있다. 이를 통해 평가의 구인타당도를 확보하고

있다고 할 수 있다.

평가의 신뢰도와 관련하여 채점자 간 신뢰도 확보 여부를 확인하기 위한 통계적 검증은 실시하지는 않았다. 그러나 채점 과정에 통해 교사들은 각자 출제한 문항에 대해 기본 답안 및 인정 답안을 작성하고 문항별로 1차 채점을 수행하였으며, 2차 채점을 통해 학생들의 다양한 답안 반응에 대해 정답 여부를 확인하는 협의 과정, 최종적으로 완료된 채점 결과를 학생들에게 설명하여 확인하는 절차를 통해 평가의 신뢰도를 확보하고 있다고 할 수 있다.

국어과 서술형 평가에 관한 교사들의 견해를 파악하기 위해 교과협의회 과정에서 면담을 실시하였다. 국어과 서술형 평가에 참여한 교사들은 대체로 서술형 평가의 필요성에 대해 공감하면서 특히 서술형 문항은 학생들이 작성해야 하는 답안의 분량, 답안 반응 유형, 시험 응시 시간의 조건, 문항의 난이도 및 변별도를 고려하여 문항 구성요소들을 마련하고 있다고 언급하고 있다. 채점 과정에서도 답안 작성 및 채점에서의 모호성을 줄이기 위해 구체적인 조건을 마련하고 있으며, 세밀한 채점 과정 및 학생들의 채점 결과 확인의 절차를 통해 신뢰도를 확보하고자 노력하고 있다는 견해를 밝히고 있다.

그러나 평가에 참여한 교사들은 개인마다 평가관의 차이로 인해 답안의 인정 범위와 관련하여 협의회에서 갈등이 일어나는 경우가 있으며, 특히 서술형 평가가 추구하는 학생들의 창의적 사고력을 확인하기 위해서는 지필평가보다는 수행평가를 활용하는 방안에 대해 언급하였다.

실제 학교에서 실시되고 국어과 서술형 평가에서는 세부적인 문항 개발과 채점 절차를 통해 평가의 타당도와 신뢰도를 확보하고 있지만, 그럼에도 불구하고 몇몇 문제점들이 있었다. 실제로 정해진 시험 응시 시간의 제약으로 인해 서술형 평가의 문항 수를 늘린다 하더라도 학생들이 작성해야 하는 답안의 분량 및 채점 기준을 고려하였을 때 논술형과 같은 글쓰기 평가가 실시되기는 한계를 지니고 있다. 이에 대한 대안으로 2012학년도에는 문학 과목에서 시험 응시 시간을 50분에서 60분으로 확대하고 비교적 긴 답안을 작

성하는 서술형 평가 문항을 개발하고 있다.

일부 학교에서는 지필평가 기간 중에서 ‘채점의 날’을 정하여 서술형 평가를 실시하고 있는 교과에 대해서는 채점을 집중적으로 실시하도록 유도하고 있다. 그럼에도 불구하고 성적에 민감한 학생들에게 세밀한 채점 과정은 평가의 신뢰도를 높이지는 하지만, 한편으로는 교사들에게 평가 업무 부담을 높이는 것이 현실이다.

또한 학생들은 비교적 국어과 서술형 문항을 지필평가에서 경험할 기회가 적어서 주어진 문제를 해결하고 조건에 따라 답안을 작성하는 방식에 대해 혼란을 겪기도 한다. 이에 대하여 교사들은 서술형 평가 예시 문항을 학생들에게 미리 공고하여 서술형 문항의 유형 및 답안을 작성하는 방법에 대한 사전 안내가 필요하다고 생각한다.

지식 생성 및 활용, 문제를 해결하는 사고 과정, 창의적 사고력을 함양시키기 위한 방안으로 최근 지필평가에서 서술형 평가가 실시되고 있다. 학교 현장에서는 평가의 타당성 및 필요성에 대해 공감하고 있지만 평가를 시행하는 단계에서 어려움을 겪는 것이 사실이다. 문항 개발 및 채점 방법에 대한 체계적인 이론화와 평가 업무 부담을 경감하는 방법에 대한 연구가 지속적으로 연구될 필요가 있다.

* 본 논문은 2012.10.22. 투고되었으며, 2012.11.11. 심사가 시작되어 2012.11.30. 심사가 종료되었음.

참고문헌

- 경기도 교육청(2007), 『고등학교 서술형·논술형 평가 길라잡이』, 장학자료 2007-11.
- _____ (2009), 『평가 결과 활용의 길잡이』, 장학자료 2009-10.
- _____ (2010), 『서술형 평가 문항의 실제』, 장학자료 2010-7.
- 김성숙·김양분(2001), 『일반화가능도 이론』, 교육과학사.
- 김주환(2010), 『서술형 평가 ROAD VIEW』, 박이정.
- 노명완(1988), 『국어교육론』, 한샘.
- 박순석·박선화·홍석순·류병곤·박서희·박준용·엄내리·박성만·김순호·김주환(2011), 『서술형 평가 ROAD VIEW—국어교과』, 박이정.
- 박영목·한철우·윤희원(2003), 『국어교육학 원론』, 박이정.
- 박도순·홍후조(1999), 『교육과정과 교육평가』, 문음사.
- 박도순(2000), 『문항작성방법론』, 교육과학사.
- 박혜상(2008), 『국어과 평가 문항 유형에 따른 학생 반응 비교—선택형 문항과 서술형 문항을 중심으로』, 고려대학교 교육대학원 석사학위논문.
- 서울특별시 교육청(2010), 『고등학교 서술형 평가 장학자료집』, 서울특별시 교육청.
- _____ (2011), 『2011 학년도 중등 평가 시행 계획』, 서울특별시 교육청.
- 윤지현(2008), 『중학교 국어과 서술형 평가의 현황 분석과 개선 방안 연구』, 한양대학교 석사학위논문.
- 장소영·신동일(2009), 『언어교육평가 연구를 위한 FACETS 프로그램』, 글로벌콘텐츠.
- 조수현(2011), 『중학교 국어과 지필고사의 서술형 문항 분석과 개선방안 연구』, 한양대학교 석사학위논문.
- 최현섭·최명환·노명완·신헌재·박인기·김창원·최영환(1996), 『국어교육학 개론』, 삼지원.
- Amrein, A. L. & Berliner, D. C. (2002), "An Analysis of some unintended and negative consequences of high-stakes Testing," The great lakes center for Education research & practice (Online, Available: <http://www.greatlakescenter.org>).
- _____ (2003), "The effects of high-stakes testing on student motivation and learning," *Educational Leadership* 60(5), pp. 32-38.
- Nichols, S. L., Glass, G. V. & Berliner, D. C. (2006), "High-stakes testing and student achievement: Does Accountability pressure Increase student learning," *Education policy analysis archives* 14(1).
- Shepard, L. A. & Dougherty, K. C. (1991), *Effects of High-stakes Testing on Instruction*, In R. L. Linn (Chair), Effects of High-Stakes Educational Testing on Instruction and Achievement, symposium presented at the annual meeting of the American Educational Research Association and the National Council on Measurement Education, Chicago.

국어과 서술형 평가 문항 개발 및 채점 방법 사례 연구

이영진

이 연구는 국어과 서술형 평가 도입과 관련하여 학교 현장에서 교사들이 어떻게 문항을 개발하고 채점을 하는지 사례를 통해 소개하였다. 평가에서 타당도와 신뢰도를 어떻게 확보하고 있으며 서술형 평가에 관한 교사들의 견해가 무엇인지를 살펴봄으로써 필요성 및 개선 방안에 관해 고찰하였다.

국어과 서술형 평가와 관련하여 평가 유형 및 평가 문항의 특성, 국어과 서술형 평가의 문항 개발 및 채점 방법의 절차를 살펴보았다.

연구 대상은 경기 부천여고 자연계 4개 반 140명으로 2011년 11월 15일 실시한 2학기 독서 과목 지필평가를 통해 학교 현장에서의 3학년 독서과 4명의 국어과 교사들이 실제로 국어과 서술형 문항 개발과 채점 과정의 절차를 사례에 주목하였다.

국어과 서술형 평가의 타당도 및 문항 변별도, 신뢰도를 파악하였다. 타당도와 관련하여 내용 타당도에서는 대체로 일반적인 문항 개발 절차에 따라 서술형 평가를 개발하면서 교과협의를 통한 문항 검토 및 수정·보완의 과정을 수행함으로써 확보하고 있다고 판단할 수 있다. 문항 변별도와 관련하여 선택형 및 서술형 문항의 정오답표를 분석하였다. 그 결과 선택형에 비하여 서술형 문항에서는 난이도를 조절하지 못한 문항은 발견되지 않아 문항의 변별력을 갖추고 있음을 확인할 수 있었다.

이와 함께 선택형 및 서술형 평가 결과 간의 상관분석 및 평가 결과를 바탕으로 하여 상·하위 집단의 서술형 및 선택형 평가 결과의 차이를 분석하였다. 상관분석 결과는 .561로 상관이 있는 것으로 나타났다. 이는 선택

형 및 서술형 문항이 공통적인 평가 요소를 공유한다는 점에서 상관이 나타난다고 할 수 있지만, 한편으로 각각의 문항 유형들이 지필평가의 틀 안에서 독립적인 역할을 수행하고 있음을 파악할 수 있었다. 이에 대한 근거로서 선택형 및 서술형 문항의 평가 결과가 상·하위 집단 간에 각각 유의한 차이를 보이고 있는 것을 제시할 수 있다. 이를 통해 평가의 구인타당도를 확보하고 있다고 할 수 있다.

평가의 신뢰도와 관련하여 교사들은 각자 출제한 문항에 대해 기본 답안 및 인정 답안을 작성하고 문항별로 1차 채점을 수행하였으며, 2차 채점을 통해 학생들의 다양한 답안 반응에 대해 정답 여부를 확인하는 협의 과정, 최종적으로 완료된 채점 결과를 학생들에게 설명하여 확인하는 절차를 통해 평가의 신뢰도를 확보하고 있다고 할 수 있다.

국어과 서술형 평가에 관한 교사들의 견해를 파악하기 위해 교과협의회 과정에서 면담을 실시하였다. 학교 현장에서는 평가의 타당성 및 필요성에 대해 공감하고 있지만 평가를 시행하는 단계에서 어려움을 겪는 것이 사실이다. 문항 개발 및 채점 방법에 대한 체계적인 이론화와 평가 업무 부담을 경감하는 방법에 대한 연구가 지속적으로 이루어질 필요가 있다.

핵심어 국어과 서술형 평가, 문항 개발, 채점 방법, 타당도, 문항 변별도, 신뢰도

ABSTRACT

A Study on the Development of Test Tools and Scoring Methods for Korean Language Essay Test

Lee, Young-jin

This study introduces the development of test tools and scoring methods of Korean art essay assessment at high school. This research investigates how to ensure the validity and reliability in the assessment and teachers' interviews about its necessity and improvement.

140 patients at Bucheon Girls' High School conducted Reading art second semester final exam on Nov. 15, 2011. and 4 Reading art teachers carried out the development of test tools and scoring.

This paper analyzed validity, item discrimination, reliability of Korean language essay assessment. Consequently, the assessment proved to have content validity through experts' reviews and item discrimination through the analysis of correct and incorrect answer table. It also proved to have construct validity through correlation analysis(.561) between multiple-choice test score and open-ended test score. and it had reliability through raters(teachers) performing consultation process about scoring.

Finally, this research conducted the interviews during the teacher council to understand the teachers' opinions about Korean language essay assessment. Teachers agreed to the necessity of the assessment, but they emphasized that the development of test tools and scoring method should be theorized systematically and the study of how to reduce assessment tasks should be conducted consistently.

KEYWORDS Korean Language Essay test, Development of Test Tools, Scoring Methods, Validity, Item Discrimination, Reliability