

英才教育研究

Journal of Gifted/Talented Education

1997, Vol. 7, No. 1, pp. 77~115

수행평가를 활용한 영재 관별

오 영 주

(한국교육개발원)

I. 서론

영재 교육의 주된 목표는 잠재력이 풍부한 영재 학생을 선발하여 그들의 지적 욕구를 충족시켜 주고 더 나아가 인류를 위해 일할 수 있도록 그들의 능력을 최대한 계발시켜 주는 것이다. 이러한 영재 교육의 성패는 우선 장래 탁월한 영재성을 발휘할 잠재력이 풍부한 인재를 가려내는 것에서 시작된다고 해도 과언이 아니다. 따라서 영재 교육자는 좋은 교육을 제공하기에 앞서서 잠재성이 풍부하여 장래 탁월한 성취를 이룰 인재를 어떻게 관별할 수 있는지 그 방안을 모색하는데 많은 노력을 기울이게 된다.

학생의 학습 수준을 평가하거나 학생을 선발해야 하는 교육 현장에서는 그들을 무엇으로 어떻게 평가해야 가장 바람직하고 정확하게 평가할 수 있는가라는 질문에 부딪힌다. 교육 평가는 교육의 목표와 일치할 때 가장 바람직한 평가를 하는 것이라는 이론을 영재 선발을 위한 평가에 적용시켜 보면, 교육자가 어느 분야의 영재를 길러내고자 하는지에 따라 평가 방식도 결정되어져야 함을 알 수 있다. 따라서 특정 분야의 영재성을 길러내기 위하여 영재를 선발하고자 결정하였다면 그 특정

* 이 내용은 조석희, 오영주, 김홍원, 박경숙 (1996)이 공동으로 개발하고 수행한 사업의 일부이며, 보다 상세하고 전체적인 수행평가의 내용은 한국교육개발원 연구보고서를 참조하기 바람.

영재성을 가장 타당하고 신뢰롭게 가려낼 수 있는 평가의 내용과 수준 및 방법까지 차별화하여야 하겠다.

각 학생이 가진 능력을 최대한 계발시켜 창의적인 사람으로 육성한다는 영재 교육의 목표는 바로 21세기를 바라보며 지향하고 있는 일반적인 교육 목표와 일치하기도 한다. 이러한 창의적인 인재 육성을 위한 영재의 선발이 지금까지 널리 사용되어 오던 일회적 선다형 지필검사나 상대평가(양적 평가) 방식으로 타당하게 이루어진다고 보기 어렵다 (Feldhusen, Asher, & Hoover, 1984). 그러므로 기존에 널리 사용되던 지필검사를 대신하거나 또는 보완할 평가 방법으로 수행평가의 활용 방안이 대두되기 시작하였다.

본 연구에서는 전통적인 지필검사에 대한 하나의 대안으로서 영재 교육이 지향하고 있는 교육 목표와 방향과 궤를 같이하는 수행평가의 이론과 실행 방안을 살펴보고 영재 선발을 위한 대안적 평가로서 그 타당성과 실용 가능성을 검토해 보고자 한다. 구체적으로 본고는 (가) 수행평가의 필요성, (나) 수행평가의 이론과 실제, (다) 수행평가를 활용한 영재 선발의 사례를 살펴보고, 그리고 (라) 영재 선발 도구로서 수행평가가 안고 있는 과제와 이를 극복하기 위한 방안을 제시하고자 한다.

II. 이론적 배경

1. 수행평가의 필요성

기존에 사용되던 선다형 지필검사와는 달리 수행평가는 지식이나 정보를 재구성하는 능력이나 자신에게 의미있는 지식과 정보를 적극적으로 수집하고 전략적으로 활용하는 능력 및 창의적 산출물을 생산해 내는 능력을 타당하게 측정해 주는 대안적 평가 방식으로 널리 고려되고 있다. 더 나아가서 수행평가는 각 개인의 과제집착력이나 끈질김과 내적 동기 유발과 같은 비인지적이고 인성적인 측면까지도 파악할 수 있는 평가 방식이다(Feldhusen, Asher, & Hoover, 1984). 국립교육평가원(1996)과 미국 기술평가국(Office of Technology Assessment, 1992)의 자료를 토대로 영재를 선발하고자 하는 교육 현장에서 수행평가를 널리 활용해야 하는 중요한 이유가 무엇인지 구체적으로 살펴보도록 하자.

가. 전통적인 평가 방식은 영재의 주된 특성인 창의성을 측정하지 못한다.

우리가 살아가면서 배운 지식을 여러 상황에 현명하고 융통성있게 이용할 수 없다면 지식을 완전

히 소화했다거나 이해했다고 말할 수 없다. 지식을 이해한다는 것은 정보를 단순히 소유한다기 보다는 주어진 문제를 해결하기 위해 지식을 알맞게 적용하는 능력이 있다는 의미이다. 창의성은 명제적 지식을 단순히 재생하는 능력이 아니라 지식을 재창조하고 재발견하는 능력이다. 이렇게 본다면, 단편적인 지식을 소유하고 있는지에 초점이 맞추어진 선다형 지필검사로는 무엇을 창조할 능력이 있는지 평가하지 못한다.

나. 전통적인 평가 방식은 각 개인의 특성을 측정하지 못한다.

선다형 지필검사는 학생들이 객관적인 지식이나 정보를 수동적으로 받아들이거나 단순히 재생산하는 존재로만 간주하여 객관적이고 보편화된 지식의 소유 수준만을 평가한다 (백순근, 1997). 이 평가 방식은 각 개인을 지식의 조합자나 능동적인 지식의 조정자로 보지 않으므로 개인의 개성과 특성을 질적으로 평가하는데 초점이 맞추어져 있지 않다. 모든 학생들이 같은 문제에 대해 객관적인 지식을 갖고 있는지 아닌지를 확인하는 이러한 평가는 각 학생이 개성과 특기를 갖도록 요구하는 다원적이고 다양한 영재교육의 목표와 부합하는 평가 방식이 아니다.

다. 전통적인 평가 방식은 실제적이고 총체적인 능력을 측정하지 못한다.

선다형 지필검사는 측정하기가 비교적 쉬운 단순한 인지적 영역에만 평가의 초점을 두어 지식의 조각 조각을 단편적으로는 측정하나 실제로 한 개인이 문제나 과제를 해결할 때 필요한 창의적 문제해결력과 같이 복합적이고 총체적인 능력을 측정하지 못한다. 그리고 선다형 지필검사는 학생의 능력을 종이와 연필을 통해 간접적으로만 측정하기 때문에 학생이 실제 문제에 부딪혔을 때 자기 가진 지식을 활용하여 어떻게 문제를 해결할 수 있는지와 같은 실제적인 능력을 평가하지 못한다.

라. 전통적인 평가 방식은 과정적 지식을 측정하지 못한다.

다양하고 급변하는 사회에서 어떤 주어진 과제를 해결하기에 필요한 정보를 학생에게 모두 가르치는 것은 거의 불가능하다. 그러므로 세계화와 정보화 시대를 대변하는 21세기의 교육은 전문화되고 고도화된 지식을 소유하고 그를 활용할 능력을 가진 인재로 양성하는데 초점을 두고 있는데, 이는 학생에게 과정적 지식을 갖게 한다는 의미이다.

인지 심리학자들은 학습자가 정보나 지식을 자기 나름대로 이해하고 의미를 구성하는 주체이기 때문에 학습자 개인이 소유한 지식의 정도뿐만 아니라 그 지식을 어떻게 습득하는지와 같은 인지적 과정에도 관심을 가져야 한다고 주장한다 (Case, 1978 ; Siegler, 1978). 따라서 인지 심리학자나 영재 교육학자들은 각 개인을 평가할 때 어떤 수준의 사고를 하는지, 문제 해결 과정에서 어떤 전략

을 사용하는지, 과정이 효율적이고 수준이 높은지와 같은 과정적 절차의 능력까지도 고려할 것을 강조하고 있다 (Feldhusen, Asher, & Hoover, 1984 ; Sternberg, 1984). 그러나 지식을 얼마나 소유하고 있는지와 같이 결과적 지식(명제적 지식)의 양에 초점이 맞추어져 있는 선다형 지필검사는 그 지식을 어떻게 습득하는지 또는 그 지식을 어떻게 활용하는지와 같은 과정적 지식을 평가하지 못하고 있다.

마. 전통적인 평가 방식은 과제집착력과 같은 비인지적 특성을 측정하지 못한다.

선다형 지필검사는 대개 문제를 해결해야 하는 시간이 정해져 있고 문제를 해결하느냐 하지 못하느냐는 답지에서 정답을 제대로 찾았는가로만 평가한다. 그러므로 학생이 문제를 해결하는데 답지의 수준을 질적으로 높이려고 끈질기게 노력할 필요가 없고, 노력할 수도 없으며, 노력했다고 하더라도 그 노력의 정도를 평가할 수도 없다. 탁월한 성취 수준을 보이거나 또는 성공한 사람들은 많은 시간과 오랜 기간을 과제를 해결하기 위해 또는 기술을 습득하기 위해 투자한다고 한다 (Bloom, 1985; Howe, 1990; Renzulli, 1982). 또한 그들이 그러한 끈질긴 노력과 집착을 하는 이유가 스스로 성취하고자 하는 내적 동기를 갖고 있기 때문이라고 한다 (Feldhusen, 1985). 선다형 지필검사로써는 탁월한 성취자가 필수적으로 갖고 있는 이러한 비인지적 영재의 특성을 평가하지 못한다.

2. 수행평가의 정의 및 특징

교육 현장에서 의미하는 ‘수행평가’(performance assessment)는 학생에게 행동하거나 산출물을 만들거나 답을 작성하도록 요구하여 구체적인 상황 하에서 실제로 행동하는 ‘과정’과 행동한 ‘결과’를 동시에 평가하는 평가 방식이다(OTA, 1992 ; Wiggins, 1993). 즉, 수행평가는 학생에게 답이 나온 과정을 서술하거나, 실제로 행동하거나, 실연하거나, 산출물을 내도록 하므로써 그들의 지식 수준, 정보 활용 능력, 고차원적 사고 기능, 창의성, 그리고 비인지적 특성을 평가하는 방법이다.

국립교육평가원(1996)과 OTA(1992)는 수행평가가 다음과 같은 특징을 갖는다고 발표하였다.

- (가) 한 가지 정답을 선택하게 하는 선다형 지필검사는 학생이 그 답을 어떻게 선택했는지 또는 어떤 과정을 거쳐 문제를 해결했는지 평가하지 못한다. 그러나 수행평가는 학생에게 문제의 정답을 선택하게 하는 것이 아니라 자기 스스로 정답을 작성하고 구성하게 하여 문제 해결의 과정을 평가한다.

- (나) 수행평가는 가상적이거나 추상적인 상황에서 학생의 능력을 평가하기 보다는 자연스럽고 실제적인 상황에서 문제를 해결하도록 하여 그들의 능력을 평가한다.
- (다) 수행평가는 문제를 해결한 결과뿐만 아니라 문제 해결의 과정까지도 중시하는 평가 방식이다. 즉, 무엇을 얼마나 아는지를 보는 명명적 지식의 습득 수준과 더불어 아는 것을 실제로 적용하여 문제를 해결하는 절차적(과정적) 지식의 수준도 평가한다.
- (라) 수행평가는 학생의 단편적인 지식을 일회적으로 평가하기 보다는 종합적이고 지속적으로 평가하여 개인의 전반적인 발달과 능력을 총체적으로 평가한다.
- (마) 수행평가는 하나의 연구과제를 주고 팀을 이루어서 공동으로 그 과제를 해결하도록 한 후 팀을 단위로 해서 평가하므로 개인의 수행을 평가하기도 하지만 집단의 수행 수준도 평가한다.
- (바) 수행평가는 창의성이나 문제해결력과 같은 인지적 영역은 물론 개인의 행동 발달 상황이나 흥미, 태도, 과제집착력, 체력과 같은 비인지적 영역도 종합적으로 평가한다.

3. 수행평가의 유형

현재 널리 사용되고 있는 수행평가의 종류는 서술형 지필검사, 논술고사, 구두 시험, 면접, 관찰, 찬반 토론, 실기/실연, 실험/실습, 연구보고서, 컴퓨터 시뮬레이션, 포트폴리오 등이 있다 (국립교육평가원, 1996 ; 남명호, 1996 ; 배호순, 1997 ; OTA, 1992). 다음에는 각 수행평가의 의미와 사용 방법에 대해 살펴보겠다.

가. 서술형 지필검사

서술형 지필검사는 주어진 답지 중에서 정답을 선택하게 하는 것이 아니라 답을 직접 서술하게 하는 평가 형태로서 학생들이 개념을 단순히 암기하고 있는 수준인지 아니면 그 개념을 제대로 이해하여 문제를 해결해 나가는 수준인지를 평가해준다. 그러므로 정답을 얻는 과정이 올바른지 아닌지 또는 어느 과정에서 오류가 발생하는지 등을 파악할 수 있어 개인에 대한 평가가 보다 정확하게 이루어진다는 장점을 갖고 있다.

나. 논술고사

논술고사는 일종의 서술형 검사이기는 하지만 기준이 되는 특별한 정답이 없고 개인의 생각이나

주장을 논리적으로 조직하여 설득력있게 표현하는 능력을 평가하는데 초점을 둔 평가라는 점이 서술형 검사와 구별된다. 논술고사는 작문 능력을 평가함과 동시에 글을 조직하는 능력이나 비판력 및 관련 지식의 습득 정도도 파악해주는 장점을 갖고 있어 수행평가에서 자주 사용되는 유형이다.

다. 구두 시험

구두 시험은 학생에게 특정 내용이나 주제에 대해서 자신의 의견이나 생각을 발표하도록 하여 지식 수준뿐만 아니라 이해력, 표현력, 조직력, 의사 소통 능력까지도 평가하는 유형이다. 구두 시험을 실시하는 방법은 학생에게 주제나 질문을 사전에 미리 알려주거나, 특정 영역만 미리 알려주거나, 또는 사전에 알려주지 않고 평가자가 제시한 질문에 그 자리에서 답변하게 하는 등 다양한 형식이 있다. 구두시험을 실시할 때의 인원 수에 따라 한 명의 평가자와 한 명의 학생이 일 대 일로 하는 시험, 여러명의 평가자와 한명의 학생이 하는 다수 대 일로 하는 시험, 반대로 일 대 다수가 하는 시험, 마지막으로 다수 대 다수로 하는 시험이 있다.

라. 면접

면접은 평가자가 학생과 직접 대면하여 질문하면 그에 대해 학생이 대답하게 하여 지필식 검사나 서류만으로는 파악되지 않는 점을 평가하는 유형이다. 면접하는 방법은 한 명의 평가자와 한 명의 학생이 일 대 일로 하는 면접, 여러명의 평가자와 한명의 학생이 하는 다수 대 일로 하는 면접, 반대로 일 대 다수가 하는 면접, 마지막으로 다수 대 다수로 하는 면접이 있다. 면접을 통한 수행평가는 진행상 융통성을 발휘하여 사전에 계획하지 않았던 예상외의 정보나 자료를 얻을 수 있고, 또한 한 개인을 다양한 측면에서 평가할 수 있다는 장점을 갖고 있다.

마. 관찰

관찰은 학생을 이해하고 평가하기 위한 가장 보편적인 방법으로 거의 모든 수행 평가에서 활용되는 평가 방식이다. 인위적이 아닌 자연스런 상황에서 관찰해야 정확한 평가가 되는데, 객관적이고 정확한 관찰을 위해 관찰 대상을 있는 그대로 기술하는 기록법이나 체크리스트 및 평정척도를 이용하기도 하고, 때로는 비디오 녹화를 한 다음 동일한 상황을 여러명의 평가자가 평정하기도 한다.

바. 찬반 토론 / 논쟁

찬반 토론은 개인적, 사회적, 또는 국가적인 이슈에 대해 학생 자신(그룹)의 생각과 주장을 발표하게 하여 그들의 다양한 능력을 평가하는 방식이다. 개인 또는 집단이 사전에 준비하거나 또는 즉

흥적으로 토론하게 할 수 있는데, 대개 전자가 후자보다 더 보편적으로 활용되는 방식이다. 찬성과 반대 의견을 토론하는 과정에서 사전에 준비한 자료나 내용의 다양성, 충실성, 논리성을 평가하면서 동시에 토론을 진행하는 능력, 반대 의견을 수렴하고 비판하는 능력, 자신의 의견을 논리적으로 소신있게 피력하는 능력, 자신의 생각대로 상대를 이끌어가는 능력, 토론을 이끌어 가는 지도력이나 성격까지 평가할 수 있다.

사. 실기 / 실연

수행평가에서 말하는 실기와 지금까지의 실기 시험과의 차이점은 종래의 실기 시험이 상황을 통제하거나 강요하여 평가가 이루어질 때 수행평가에서의 실기는 통제하거나 강요하는 상황이 아니라 자연스러운 상황에서 관찰되는 실제 수행 능력을 평가한다는 점이다. 실기를 통한 평가는 예체능 분야에서 주로 사용되어 오던 방법이나 현재는 말하기 평가에서 발음, 어휘 구사력, 표현력, 청취력, 이해력을 보고자 할 때 많이 활용되고 있다.

아. 실험 · 실습

실험이나 실습은 직접 실험하거나 실습하는 과정과 그 결과로 산출된 결과물을 평가하는 평가 방식이다. 이 평가는 과제와 관련된 개념의 습득 수준은 물론 실험을 설계하고 수행하는 능력, 기구를 조작하는 기능, 관찰하고 기록하는 능력, 습득된 지식을 실제로 활용하는 능력과 같이 과정적 지식이나 실제 문제해결 능력도 평가할 수 있다.

자. 연구 보고서

연구보고서는 학생이 여러 가지 연구 주제 중에서 자신의 능력이나 흥미에 적합한 주제를 선택하여 그 주제에 대하여 자료를 수집하고, 분석하며, 종합한 보고서를 평가하는 방식이다. 연구 주제는 평가자가 정해주거나 또는 학생이 정할 수 있으며, 과제의 수행은 과제의 수준이나 유형에 따라 개인이 수행할 수도 있고 관심있는 학생들이 모여 공동으로 수행할 수도 있다.

차. 컴퓨터 시뮬레이션

실제 실험해보기에는 비용이 많이 들거나, 위험하거나, 불가능하거나, 또는 장시간에 걸쳐 수행해야 하는 경우 실제 상황을 모의하는 프로그램을 컴퓨터에 넣어 학생에게 조작하게 하여서 그들의 문제 해결력을 간접적으로 평가하는 방법이다. 시간이나 공간 및 경제적 제약으로 인해 대규모 실기 평가가 용이하지 않고, 그렇다고 지필검사만으로는 피험자의 능력을 파악하기 어려울 때 효과적인

이다. 예를 들어 다양한 조건에 따라 동물이나 식물을 키우는 능력, 미국에서 서부 개척시대에 많은 장애를 뚫고 무사히 서부까지 도착하는 능력, 복잡한 연구 절차와 많은 실험 기구로 연구를 수행하는 능력, 복잡한 도형을 순차적으로 작도하는 능력을 평가할 수 있다.

카. 포트폴리오

포트폴리오란 특정 과제나 프로젝트 수행에 초점을 두고 의도적으로 수집한 작품집을 평가하는 방식이다. 포트폴리오에 포함되는 작품은 대개 다양한 검사 결과(성적), 차트, 평정표, 일지, 연구보고서, 소견 기록, 스케치, 관찰 기록, 계산 과정을 보여주는 연습지, 미완성/완성 작품, 녹음한 테이프, 읽은 책의 명단, 사진, 비디오 테이프, 사고 과정을 기록한 내용 등이다.

포트폴리오는 결과와 과정 중에서 어디에 중점을 두고 작성하느냐에 따라 “본보기 포트폴리오”, “과정 포트폴리오”, “결합 포트폴리오”로 분류된다(OTA, 1992). 본보기(exemplary) 포트폴리오란 대표적인 작품이나 가장 우수한 작품의 견본을 모은 작품집으로 주로 선발 목적에 활용된다. 과정(process) 포트폴리오는 특정 과제에 관한 완성 작품, 미완성 작품, 초벌 작품 등을 수집한 작품집으로 주로 학생의 발달을 진단하는데 사용된다. 그리고 결합(combined) 포트폴리오란 전술한 두가지 작품이 결합된 형태이다. 포트폴리오는 어떤 방법으로 작품을 수집하느냐에 따라 자연 발생적 산출물을 모은 포트폴리오, 정기적 산출물을 모은 포트폴리오, 그리고 통제되고 계획된 상황의 산출물을 모은 포트폴리오로 분류되는데, 평가에 활용되는 포트폴리오는 대개 통제되고 계획된 상황의 산출물을 모은 포트폴리오이다.

포트폴리오의 장점으로는 학생의 과거와 현재의 상태를 쉽게 파악해주고, 장점이나 약점 및 잠재력에 대해서도 파악해준다는 점이다. 그러므로 포트폴리오를 활용한 평가는 개인을 총체적이고 지속적으로 평가해주는 대표적인 수행평가의 기법으로 간주되어 표준화검사의 대안으로, 학생 선발용 도구로서, 또한 학력검사나 기초능력검사를 대체하는 평가 방식으로 인정되고 있다(OTA, 1992; 배호순, 1997).

III. 수행평가를 활용한 영재의 선발

한국교육개발원의 영재교육팀은 장래 우수한 성취를 이룰 영재를 선발할 수 있는 방안을 다각도로 연구해 보았다. 이 연구는 강원도 횡성군에 있는 민족사관고등학교(이하 ‘민족사관고’로 칭함)가

언어, 수학, 과학 분야에 영재성이 있거나 있을 가능성이 있는 입학생의 선발을 한국교육개발원에 의뢰함으로써 시작되었다. 영재교육팀에서는 이제까지 많이 사용되어 온 일회의 선다형 지필검사로 창 의성과 잠재성이 뛰어난 영재를 선발하는데 한계가 있음을 인지하여 그 대안적 평가 방안으로 다양하고 새로운 유형의 수행평가를 개발하였고, 이를 여름 캠프를 열어 실시해 보았다.

한국교육개발원 영재교육팀에서 1997학년도 민족사관고의 신입생을 선발하기 위하여 개발한 수행평가는 다음의 세 가지 능력을 평가하는데 그 초점을 두었다.

첫째, 지금까지 배우거나 접해보지 않았던 문제에 접하게 하여 학생들의 창의적 문제해결력이나 잠재 능력을 평가하였다.

둘째, 학생들에게 다양한 과제와 학습 활동에 참여하게 하여 지필검사에서 측정하기 힘든 적용력, 실용력, 분석력, 종합력, 평가력 등의 고급 사고 능력을 평가하였다.

셋째, 성공하거나 위대한 업적을 남기는 인물들의 공통점은 우수한 지적 능력뿐 아니라 창의성과 과제에 대한 집착력도 뛰어나다고 하므로 이러한 비지적 측면도 평가하였다.

1. 영재 선발 단계

한국교육개발원에서 민족사관고등학교의 신입생을 선발하기 위해 개발한 평가 과정은 크게 3단계로 나뉘어진다. 제 1단계는 학교 성적 상위 1%인 학생의 중학교 학교장의 추천, 2단계는 지식 습득의 수준을 평가하기 위한 지필검사, 3단계는 창의적 문제 해결력과 잠재 능력을 평가하기 위한 수행평가로 구성되었다.

전국에서 학교장의 추천을 받은 262명의 학생을 대상으로 제 2단계의 검사를 실시하였다. 이 단계에서는 국어, 영어, 수학, 과학 교과에서 선다형과 단답형으로 구성된 지필검사를 개발하고 실시하였는데, 이 검사를 실시한 목적은 학교간 성적차를 해소하기 위한 것이었다. 다시 말해서 학교 성적 상위 1%라 하더라도 '가' 학교에서 상위 1%인 학생과 '나' 학교에서 상위 1%인 학생의 능력에는 차이가 있으므로 일률적인 평가 체제를 적용하여 이러한 학교 차를 해소하고자 하였다.

제 3단계에서의 평가는 학생들이 지금까지 접해보지 않았던 과제와 수시간 수일을 요구하는 과제를 포함시켜 창의적 영재의 요소인 폭넓고 깊이 있는 지식의 양, 창의적 사고력, 고차원적 사고력, 과제집착력, 지구력을 평가하였다.

평가 활동을 구성하는데 적용된 이론적 모델은 Renzulli(1977)의 '삼부 심화학습 모델'로서, 각 평가 활동마다 탐색 활동 단계 → 사고 및 기능 훈련 단계 → 프로젝트 연구 및 해결 단계로 진행되었다. 제 1부의 탐색 단계는 학생들에게 강의나 토론을 통해 다양한 아이디어를 내면화 할 수 있도록

록 해주는 단계이었다. 제 2부의 훈련 단계는 중, 소그룹 활동을 통해 학생에게 특정 프로젝트를 연구할 수 있는 사고능력, 기술, 방법 등을 갖게 해주는 단계이었고, 제 3부의 실제 프로젝트 연구 단계는 개인이나 소그룹이 함께 활동해 독창적인 결과물을 생산해 내도록 하는 단계이었다.

2. 캠프의 운영

- (가) 캠프 대상자 : 캠프에 참가한 학생은 1단계와 2단계의 선발 과정을 거친 216명 중에서 결사생 5명을 제외한 211명의 전국 소재 중학교 3학년 학생이었다. 성별 학생 수는 남학생 144명과 여학생 67명이었으며, 지역별 학생 수는 특별시와 광역시 109명, 시 88명, 읍과 면 14명이었다.
- (나) 인적 사항 : 캠프는 영재교육 전문가 4명, 수행평가 개발 책임자 4명, 강의 및 평가 교사 54명 (국어 13명, 영어 12명, 수학 12명, 과학 17명), 민족사관고 교사 14명으로 총 76명에 의해 진행되었다. 민족사관고의 교사는 캠프 생활을 총괄적으로 지도하는 담임 교사로서의 역할과 캠프 운영이 전반적으로 잘 운영되도록 조절하는 조절관으로서의 역할을 담당하였다.
- (다) 물적 사항 : 캠프는 강원도 횡성군 횡성읍에 위치한 민족사관고등학교에서 운영되었다. 기숙 장소는 학교내에 있는 생활교육관(기숙사)과 가정교육관이며 학생들은 4인 1실 단위로, 교사는 2인 1실 단위로 기숙하였다. 평가 활동을 위해 필요하였던 장소는 30명이 활동할 수 있는 12개의 교실, 30명이 동시에 작업할 수 있는 컴퓨터실, 실험실, 강당, 도서관, 운동장이었다.
- (라) 캠프 일정 : 1차 캠프는 1996년 7월 21일 부터 7월 26일 까지였으며, 2차 캠프는 7월 26일 부터 7월 31일 까지 운영되었다. 각 캠프는 국어와 영어에 1일, 수학에 1일, 과학에 2일로 총 5박 6일간 진행되었다.

다음에는 한국교육개발원 영재교육팀에서 개발하고 실시했던 다양한 유형의 수행평가를 국어, 영어, 수학, 과학 교과 순으로 소개하겠다.

3. 국어과 수행평가

가. 국어과 평가 활동의 운영

국어과 캠프는 27명의 학생을 한 반으로 하여 4개의 반을 편성하였으며, 각 반에 7가지 평가 활동을 돌아가며 실행하였다. 매 교시마다 한 반씩 말하기 평가 활동에 참여하는 동안, 나머지 3개의 반은 그 이외의 활동에 참여하였다 (표1 참조).

〈표 1〉 국어과 캠프 일정표

시 간 08:00 - 08:10	국어 캠프 프로그램에 대한 오리엔테이션 4개반의 학생이 각자 교실에서 청취			
	반 별 활 동			
	1반 (27명)	2반 (27명)	3반 (27명)	4반 (27명)
08:10 - 09:00	활동 7	활동 1&2	활동 1&2	활동 1&2
09:10 - 10:00	활동 3&4	활동 7	활동 3&4	활동 3&4
10:10 - 11:00	활동 5&6	활동 5&6	활동 7	활동 5&6
11:10 - 12:00	활동 1&2	활동 3&4	활동 5&6	활동 7

나. 국어과 평가 활동과 평가 기준

국어 캠프는 학생의 언어 구사 능력과 언어 정보 처리 능력 및 언어 표현력을 측정하고자 다음에 설명되는 7가지 평가 활동으로 구성되었다.

(1) 주제문 작성 및 중요도 평정하기

인문과학 분야의 설명문을 읽고 그 전체 내용의 주제를 한 문장으로 진술하고, 이 주제문에 비추어가며 각각의 내용이 차지하는 중요도를 평정하는 활동이었다. 중요도의 평정은 장문의 글을 의미 있는 단위로 다시 나누어 쓴 것에 5단계로 표시하되 가장 중요한 단계(5단계)와 그 다음으로 중요한 단계(4단계)만을 표시하도록 하였다. 그 중심 내용과 세부내용을 이해하는 능력과 글 내용을 분석하고 비교 평가하는 능력을 평가하였다.

(2) 고전 문학 작품 감상하기

주어진 고전 문학 작품(춘향전)의 일부분을 읽고 그 내용에 출현하는 인물인 '춘향'의 여인상에 대하여 각자의 의견을 적는 논술 활동이었다. 글의 길이는 제한하지 않으며, 문단 구성, 맞춤법, 띄

어쓰기 등에 유의하도록 하지만 글을 감상하고 이해하는 능력과 좋은 내용을 잘 조직하여 표현하는 능력을 평가하는데 초점을 두었다.

(3) 글을 읽고 이어질 내용 예측하기

짧은 수필을 읽고 지금까지 읽은 내용을 기초로 하여 다음에 이어질 내용을 추리하는 활동으로 이어질 내용을 최소한 3문장 이상의 길이로 완성하도록 하였다. 이해한 글을 재조직하여 관련된 아이디어를 표현하는 능력을 평가하였다.

(4) 내용을 읽고 논설문 쓰기

역사상에 나타난 파쟁에 대한 어느 학자의 글(설명문)을 읽고 ‘파쟁’을 소재로 ‘단결과 분열의 교훈’이라는 논제의 글을 쓰는 활동이었다. 어떤 내용을 읽은 후 그 글의 내용이 인간, 사회, 국가에 주는 의미에 대한 학생의 창의적 사고와 아이디어를 평가하기 위한 활동이었다.

(5) 두 편의 이야기 듣고 느낌 쓰기

녹음 테이프를 통해 주제가 비슷한 두 편의 대화체 이야기를 연속해서 듣고, 그 느낌을 하나의 글로 적고 쓴 글의 제목도 붙이는 활동이었다. 이는 청취한 내용에 대한 ‘간단한’ 요약과 그 내용에 대한 ‘나의 생각’을 조화롭게 표현하는 능력을 평가하기 위한 활동이었다. 특히 학생이 주장하고자 하는 내용의 중심 주제를 인지하는지 또한 현대 우리 사회에 대한 올바른 비판과 제언을 어느 수준으로 할 수 있는지를 평가하였다.

(6) 여러 주제들을 통합하여 한 편의 내용 구성하기

서로 관계가 없는 4개의 단편적 글에서 어떤 공통점을 찾아 이들을 가능한 한 많이 통합하여 하나의 내용으로 통합 논술을 쓰고, 쓴 글에 제목도 붙이는 활동이었다. 이는 관련 아이디어를 생성하고, 그 아이디어들을 창의적으로 조직하며, 그를 글로 유창하게 표현할 수 있는 능력을 평가하는 활동이었다.

(7) 주어진 주제에 관하여 3분간 말하기

주어진 주제에 관하여 약 1분간 자신의 의견을 머리 속에 정리하고 3분 정도 그 의견을 말하는 활동으로 학생의 말하는 태도, 억양, 목소리의 크기 보다는 내용의 풍부성과 논리성, 표현의 유창성, 전달의 적합성, 설득력을 평가하는데 초점을 두었다. 평가받을 순서가 된 학생은 시험관이 있는 평가실에 가서 책상 위에 놓인 세 개의 주제 중에서 한 개를 선택하였는데 주제가 쓰여있는 쪽지가

뒤집혀 있으므로 주제를 뽑기전에는 어떤 주제가 선택될지 알 수 없도록 하였다. 선택된 주제가 마음에 들지 않거나 그 주제에 대해 좋은 의견이 없더라도 다른 주제로 변경하지 못하였다.

[말하기 평가의 주제]

1. 중, 고등학교 학생들은 학교에서는 학교가 지정한 교복을 입는 것이 좋다.
2. 요즘 신세대들은 우리의 전통 문화와 사회 규율을 너무 등한시하고 있다.
3. 초등학교에서는 남녀 공학을 하는 것이 좋으나, 중학교 이상에서는 남녀 공학을 하는 것이 좋지 않다.

4. 영어과 수행평가

가. 영어과 평가 활동의 운영

학생들을 두 개 조로 나누어 1조는 독해/영작문 활동에, 2조는 말하기/듣기 활동에 참가시켰다. 말하기/듣기 활동 조는 6개의 반으로 나누어 각 교실에 9명씩 배치하였고, 독해/영작문 활동 조는 3개의 그룹으로 나누어 각 교실에 18명씩 배치하였다 (1그룹은 2개반을 합친 것임). 활동이 끝난 후 조를 교체하여 평가 활동을 계속하였다 (표2 참조).

〈표 2〉 영어과 캠프 일정표

13:10-13:30	영어시험에 대한 오리엔테이션 (학생 각자의 반에서 청취함)	
13:30-15:30	작문/독해 활동반 1조 (1-2반, 3-4반, 5-6반) 9명/1반 × 2반 x 3그룹	말하기/듣기 활동반 2조 (7, 8, 9, 10, 11, 12반) 9명/1반 × 6개반
15:30-15:50	1조와 2조의 활동을 교체시키기 위해 학생을 이동시킴	
15:50-17:50	작문/독해 활동반 2조 (7-8반, 9-10반, 11-12반) 9명/1반 × 2반 x 3그룹	말하기/듣기 활동반 1조 (1, 2, 3, 4, 5, 6반) 9명/1반 × 6개반

나. 영어과 평가 활동과 평가 기준

영어 캠프는 학생들의 영어 능력을 측정하기 위하여 독해, 영작문, 말하기와 듣기의 3가지 평가 활동으로 구성되었다.

(1) 독해 활동

독해는 비교적 짧은 글 내용의 정보처리 능력, 중심 내용의 이해력 및 독해력, 세부 내용의 기억력, 그리고 글 내용의 비교력을 평가하는 활동이었다. 100단어 이내의 글을 1-2분 정도의 제한된 짧은 시간 동안만 OHP로 제시한 후 주어진 질문에 적절한 답을 오지선다형 답지에서 찾는 활동으로 모두 10개의 단문을 제시하였다.

(2) 영작문 활동

영작문은 글 내용의 정보처리 능력, 글 내용의 이해력과 독해력 및 요약력, 그리고 관련 내용의 조직력과 표현력을 평가하는 활동이었다. 영어로 된 긴 문장을 약 3분간 OHP로 보여주고 읽었던 내용을 기억해 20분간 영어로 요약하는 활동으로 모두 2개의 긴 글을 제시하였다.

(3) 말하기와 듣기 활동

말하기 활동은 관련 주제의 이해력, 어휘 및 문장의 생성력, 그리고 관련 내용의 구성력과 조직력 및 표현력을 평가하기 위한 것이었다. 듣기 활동은 여러 관련 내용의 청취력과 이해력, 그리고 분석력과 통합력을 평가하기 위한 것이었다.

말하기와 듣기 활동은 발표 차례가 된 학생이 한 주제에 대해 영어로 3분간 발표하면 나머지 8명은 그 발표와 관련된 질문을 한 가지 이상씩 하고, 발표자는 그 질문에 적절하게 응답하는 형식으로 진행되었다. 체비뽑기로 9명 학생의 발표 순서를 정하였고, 발표 순서가 된 학생은 10개의 주제 중에서 (예: The person I respect best; The place I want to visit; My Hometown 등) 한 개를 뽑도록 하였는데, 그 주제가 어렵거나 마음에 들지 않으면 한 번 더 뽑을 수 있었다.

5. 수학과 수행평가

가. 수학과 평가 활동의 운영

수학 캠프는 27명의 학생을 한 반으로 하여 총 4개 반에서 진행되었다. 캠프 첫날 배부된 수학과

수술형 지필검사를 각자 시간을 조절하여 개별적으로 풀게 하였는데, 이때 각종 참고 자료를 사용할 수 있으나 다른 학생과 함께 풀지 못하게 하였다.

수학과 캠프가 열린 오전에는 학생을 4개 반으로 나누어 단답형 지필검사를 실시하였고, 그 후 모든 학생을 강당으로 옮겨 1단계 스케치패드(이하 'SP'로 칭함) 활동을 실시하였다. 스케치패드의 개념, 활용 목적, 활용 방법 등에 대한 강의를 들은 후 그들이 발견한 수학 정리를 보고서에 써내도록 하였다.

오후에는 다시 학생들을 4개 반으로 나누어 진행하였다. 두 개의 반은 2시간 연속으로 빨대 활동을, 한 반은 SP 활동을, 나머지 한 반은 데빌/매지믹서/구두시험 활동을 하였다. 빨대 활동은 2시간 연속으로 진행되어야 했으므로 1시간이 경과한 후에는 빨대 활동을 하지 않던 두 개반의 학생들만 교체시켰고, 2시간이 끝난 후에는 4개 반의 학생을 모두 교체시켜 새로운 활동을 하게 하였다 (표 3 참조).

데빌/매지믹서/구두시험이 실시된 반의 학생은 60분 동안 3개의 교실을 왕래하며 3가지 활동을 하였다. 27명의 학생 중에서 7명의 학생은 매지믹서 활동실에서 10분간 활동하였고, 매지믹서 활동이 끝난 학생은 3~4명씩 그룹을 형성해 구두시험을 보았으며, 구두 시험이 끝난 학생은 나머지 시간 동안 데빌 활동을 하였다.

〈표 3〉 수학과 캠프 일정표

시 간	활 동 내 용			
8:00 - 9:50	단답형 지필검사 (100분 소요) 27명/반 × 4반 = 108명			
10:10 - 11:50	Sketchpad Demo (강당) 전체 학생 대상			
13:10 - 16:10	반 별 활 동			
	1 반	2 반	3 반	4 반
13:10 - 14:10	빨대	빨대	스케치패드	매지믹서/구두 시험/데빌퍼즐
14:20 - 15:20	빨대	빨대	매지믹서/구두 시험/데빌퍼즐	스케치패드
15:30 - 16:30	스케치패드	매지믹서/구두 시험/데빌퍼즐	빨대	빨대
16:40 - 17:40	매지믹서/구두 시험/데빌퍼즐	스케치패드	빨대	빨대

나. 수학과 평가 활동과 평가 기준

수학 캠프는 수학자들이 실제로 활용하고 있는 일반화와 추상화 능력, 문제 해결력, 수학 정리 발견 능력, 일상 생활에서 수학 정리를 찾는 능력, 공간 지각 능력을 측정하기 위하여 다음에 서술한 6가지 활동으로 구성되었다.

(1) 서술형 지필검사

이 검사는 수학적 지식을 이용하여 창의적이고 확산적인 사고를 하는 능력, 다단계로 사고하는 능력, 그리고 추론하는 능력을 측정할 수 있는 6개의 서술형 문항으로 구성되었다. 한 문항을 푸는데 1시간 반 정도의 시간이 소요되는 문제로서 모두 푸는데 9시간 정도 소요될 것으로 예상하였으며, 학생 각자가 자신의 시간을 조절해 2일간 독자적으로 해결하도록 하는 프로젝트형 지필검사이었다.

(2) 단답형 지필검사

이 검사는 수학자들이 실제로 활용하고 있는 수학적 사고 능력, 즉 일반화와 추상화 능력, 직관력, 두뇌회전력, 응용력을 측정할 수 있는 10개의 단답형 문항으로 구성되었다. 문제의 수준은 한 문항을 푸는데 10분 정도 소요되는 것이었다.

(3) 스케치패드로 동적 기하 정리하기

스케치패드는 컴퓨터에서 점이나 직선 및 원을 이용해 기하학적 구성을 동적으로 조작해보므로서 도형들의 본질적인 관련성을 찾게 하는 컴퓨터 시뮬레이션 프로그램이다. 학생들에게 익숙치 않은 문제를 동적으로 보여주고 점의 자취에 대하여 서술하거나 직선과 원 위를 움직이는 동적인 영상에 대해 추측하는 과제를 부여하였다.

이 활동은 2단계로 나누어 진행되었다. 첫 단계에서는 수학자 Gauss가 발견하고 증명한 정17각형의 작도 과정과 피타고라스 정리와 관련된 여러 기하학적 증명을 화면으로 보면서 수동적으로 수학 법칙을 발견하는 활동을 하였다. 두 번째 단계에서는 학생 스스로 SP 프로그램을 조작하면서 기하학적 성질과 도형의 작도 순서를 발견하는 활동을 하였다.

(4) 빨대로 정다면체 만들기

초보적인 3차원 다면체를 다루는 능력은 미적분학에서 다루는 3차원 그래프를 활용하는 능력의 기초가 된다. 캠프에서는 초보적인 3차원 개념을 설명해주는 물체(모형)를 만들어 공간에 대한 여

러 관계나 성질을 관찰하게 하므로써 학생들의 공간 지각력, 삼차원 개념의 추론 능력, 그리고 적용 능력을 측정하였다. 학생들은 빨대로 제작된 견본품을 보면서 ‘정사면체’, ‘정팔면체’, ‘정육면체에 내접하는 정사면체’ 순으로 3종의 정다면체를 제작하였다.

(5) 데빌(Devils) 퍼즐

이 퍼즐은 악마와 같이 어렵다고 ‘데빌’이라는 이름이 붙여졌다고 하는데, 7 조각을 모두 사용해서 주어진 평면적 형태를 만들도록 되어 있다. 학생들의 평면 지각 능력, 조작 능력, 그리고 추론 능력을 측정하기 위하여 3개의 퍼즐 문제를 제시하였다.

(6) 매지 믹서 (Magi Mixer)

매지믹서는 육각형의 6개 모서리 중에서 다섯 개는 1~6까지의 숫자가 쓰여있는 흰색 주사위가 달려있고, 한 개는 검은색 주사위가 달려있으며, 가운데에는 10, 20, 30, 40, 50, 60 숫자가 쓰여있는 검은색 주사위가 달려있다. 매지믹서를 활용하는 방법은 흰색 주사위에 나타난 5개 숫자를 빠짐없이 한 번씩만 순서에 상관없이 가감승제하여 검은 주사위에 나타난 숫자의 합을 만드는 것이다. 학생들의 계산 속도와 수학적 다양성을 측정하기 위하여 3개의 문제를 제시하였다.

(7) 구두 시험

구두 시험을 통해 학생들의 수학적 표현력, 기억력, 새로운 결론 유도 능력, 창의력, 분석력을 평가할 수 있었지만 짧은 시간으로 학생들의 사고 과정을 충분히 탐색할 수 없었기에 성적에 반영하지는 않았다. 구두 시험은 두 명의 평가자가 학생 3~4명을 대상으로 평가하는 다수 대 다수 형태로 진행되었다.

6. 과학과 수행평가

가. 과학과 평가 활동의 운영

과학 캠프는 각 반 9명 단위로 편성하여 2일간 진행되었다. 구체적으로는 3명을 1조로, 그리고 3조를 한 반으로 만들어 함께 활동하게 하였다 (표4 참조). 각 조의 구성은 남학생 2명과 여학생 1명이었다.

〈표 4〉 과학과 캠프 일정표

일자 시간	과 학 제 1 일	과 학 제 2 일
08:00-09:00	다단계 선다형 지필검사 (9명/교실 × 12개 교실)	주관식 대응 지필검사 (9명/교실 × 12개 교실)
09:20-10:40	생물 실험 및 보고서 (9명/교실 × 12개 교실)	화학 실험 및 보고서 (9명/교실 × 12개 교실)
11:00-12:00	서술형 개방 지필검사 (9명/교실 × 12개 교실)	서술형 장기 지필검사 (9명/교실 × 12개 교실)
12:00-13:00	점 심 식 사 (반 별)	
13:00-16:45	공동탐구 I (조별) 개별탐구 연구보고서 (9명/교실 × 12개 교실)	공동 탐구 연구보고서 발표/집단 토론 (조별) (9명/교실 × 12개 교실)
17:00-18:00	마당놀이 - 세뱃줄 당기기 (반별, 운동장 4 영역)	마당놀이 - 수수깡 다리 만들기 (조별, 운동장)
18:00-19:00	저 녁 식 사 (반 별)	
19:10-21:10	공동탐구 II (조별) (9명/교실 × 12개 교실)	과학 공연 준비- 1시간 (반별) (9명/교실 × 12개 교실)
21:10-22:30	자 유 과학 공연 - 2시간 (전체, 강당)	

나. 과학과 평가 활동과 평가 기준

과학 캠프는 학생들의 과학 개념의 이해와 적용력, 문제해결력, 과학적 탐구력, 그리고 과학 태도와 인성을 측정하기 위하여 모두 12가지 활동을 하였으나 선다형이나 단답형 지필검사를 제외한 9가지 수행평가 활동만을 소개하기로 한다.

(1) 서술형 장기 지필검사

서술형 장기 지필검사는 학생들의 과학 개념의 이해와 적용력, 과학적 탐구력을 측정할 수 있는 9개의 문항으로 구성되었다. 9개 문항 중에서 학생들은 물리, 화학, 생물, 지학 각 분야별로 한 문제씩 즉, 1~3번 중 하나, 4~5번 중 하나, 6~7번 중 하나, 8~9번 중 하나를 선택하여 총 4문제만 풀도록 하였다. 정확한 평가가 될 수 있도록 오답에 대한 감점제가 있음을 사전에 알려주었다.

(2) 개별 탐구 연구보고서

이 활동은 학생 각자가 세 가지 완구를 관찰하고 주어진 질문에 답하는 것으로서 정확한 결론을

내리지 못하더라도 자신의 생각을 모두 기록하도록 하였다. 이 활동으로 과학 개념의 이해와 적용력, 탐구 사고 기능 (가설 설정, 정보 수집과 활용, 추론, 종합), 탐구 능력(창의적 사고), 그리고 과제집착력을 평가하였다. 학생들은 세 가지 완구(‘벽대고돌이’, ‘한없는돌이’, ‘신기루보물단지’ 각 1개씩)를 만지지는 말고 관찰만 하여 각 완구가 작동하는 원리를 연구보고서에 기술하였다.

(3) 화학 실험

화학 실험은 주어진 실험 기구와 과학적 지식을 활용하여 실험을 계획하고 실시하며 결론에 도달하는 능력을 평가하는 활동이었다. 미지의 플라스틱의 밀도를 구하기 위해 플라스틱을 물과 에탄올에 각각 넣어보게 하여 과학적 개념과 원리를 적용하여 문제를 해결하는 과정적 지식이나 실험 기능의 수준을 평가하였다.

교사가 실험 기구의 설치 방법과 조작 방법을 실연해주었지만 준비되어 있는 실험기구 중에서 실험에 필요한 기구는 각 학생이 파악해서 가져가게 하였다. 학생 각자 개별적으로 실험하였고, 그 과정과 결과를 7개의 문항으로 구성된 연구보고서에 기술하였다.

(4) 생물 실험

생물 실험은 주어진 재료와 기구를 이용하여 실험을 설계하고 그에 따라 실험을 실시하고 결론에 도달하는 능력을 평가하는 활동이었다. 학생들에게 아밀라아제의 활성이 온도와 산도에 따라 어떻게 다른지 알아내고, 아밀라아제의 활성화를 위한 최적 조건을 제시하게 하였다. 이 실험이 요구하는 과학적 개념이나 원리는 화학적 소화, 효소의 특징, 효소의 체내 작용, 체내의 생리적 특징들이다. 실험 진행 절차는 화학 실험과 동일하였으며, 학생들은 실험 과정과 결과를 4개의 문항으로 구성된 연구보고서에 기술하였다.

(5) 공동 탐구

공동 탐구는 한 조 3명의 학생이 공동으로 세 종류의 완구를 6시간 동안 탐구하여 그 과정과 결과를 연구보고서에 서술하는 활동으로 학생들의 개방적 탐구력, 공동 탐구 태도, 지구력을 평가할 수 있었다. 각 조에 한 세트의 완구(벽대고돌이, 한없는돌이, 신기루보물단지 각 1개씩)를 나누어 주면서 2일간 계속 탐구할 수 있도록 해주었다. 교내에 있는 도서를 마음대로 참고하고 복사할 수 있도록 허용하였고, 또한 교실이나 도서관 등 어디에서나 활동할 수 있도록 하였다. 세 가지 완구와 관련된 과제를 해결하기 위해 같은 조에 소속된 팀원끼리 서로 아이디어를 교환하도록 권장하였지만 평가가 공정해지도록 보고서를 작성할 때는 1인 1과제 원칙으로 하였다.

(6) 공동 탐구 발표와 집단 토론

공동 탐구의 과정과 결과에 대해 조팀원이 함께 발표하고 토론하는 동안 학생들의 개방적 탐구 능력, 발표 및 토론 능력, 지적 흥미의 정도, 지구력과 몰두, 공동 탐구 태도를 평가하였다. 이 활동은 각 반에서 4시간 동안 진행되었는데, 각 조의 역할은 한가지 완구에 대해 발표하거나('발표자팀'), 반론하거나('반론자팀'), 토론을 진행하는('사회자팀') 것으로서 3회전이 진행되는 동안 각 조의 역할은 계속 바뀌었다. 각 조의 발표 순서나 발표할 완구는 '가위 바위 보'로 정하였으며, 각 회전의 소요 시간은 약 60분이었다.

(7) 마당 놀이 - 세بات줄 당기기

세بات줄 당기기 활동은 남학생 6명과 여학생 3명으로 구성된 한 반의 학생이 운동장에 그려진 세بات줄 당기기의 한 영역에 선 후, 3개의 반이 동시에 줄당기기 하는 활동이었다. 이 활동으로 학생들이 힘의 합이나 마찰력과 같은 물리 분야의 지식을 이해하고 적용하는 능력을 평가하였고, 그리고 지구력과 협동심도 평가하였다. 세 바줄이 묶인 부분이 원의 중심에 있는 상태에서 게임이 시작되었으며, 세바줄의 묶인 부분을 자기 영역으로 끌어온 반이 이기는 것이었다.

(8) 마당 놀이 - 수수깡 다리 만들기

남학생 2명과 여학생 1명을 한 조로 하여 가장 가벼우면서도 튼튼한 수수깡 다리를 만들게 하는 활동으로 다리위에 추나 벽돌을 올려보아 정해진 시간 동안 다리가 부러지지 않으면서 무게가 가장 가벼운 다리가 우승하는 것이었다. 이 활동으로 학생들이 힘의 합, 탄성력, 응력, 압력 등 물리 분야의 지식을 이해하고 적용하는 능력을 평가하였다.

(9) 과학 공연

과학 공연은 학생들이 알고 있는 과학 개념과 지식을 창의적으로 재구성하고 표현하는 활동으로 연극, 만담, 노래와 무용, 과학 합주, 시 짓고 낭송하기 중에서 한 가지 유형을 선택하여 반별로 5분간 공연하였다. 공연을 구상하고 준비할 시간이 짧았기 때문에 학생들에게 공연 주제를 예시해 주었지만 다른 주제를 정하여도 무방하였다.

학생들에게 예시해준 공연 주제는 다음과 같다.

- [연극] ▶ 갈릴레이의 종교 재판
▶ 석유가 고갈된 미래 사회의 생활 모습
▶ 원자력 발전소 건설에 대한 찬반 회의

- ▶ 인공 수정에 대한 찬반 토론
- [만담] ▶ 인체내 각 기관들의 대화 (예: 심장, 간, 대장, 소장, 위....)
- ▶ 동물 대표들의 지구환경 대책 회의(예: 어류, 양서류, 파충류....)
- ▶ 꽃들의 대화 (예: 민들레, 보리, 소나무, 이끼....)
- ▶ 돌들의 대화 (예: 역암, 사암, 화강암, 현무암....)
- [노래/무용] ▶ 전압, 전류, 저항의 관계를 주제로 한 노래(개사) 및 무용
- ▶ 물의 합성과 분해를 주제로 한 노래(개사) 및 무용
- [과학 합주] ▶ 과학 오케스트라 I (물컵, 다섯음 돌리개, 고무줄 파리 이용)
- ▶ 과학 오케스트라 II (PVC 파이프, 빨대 이용)
- [시짓기/낭송] ▶ 지구를 주제로 한 시짓기와 낭송 (예: 지구의 수권, 암권, 대기권)
- ▶ 에너지를 주제로 한 시짓기와 낭송 (예: 에너지 전환과 보존)

IV. 수행평가의 결과 분석

캠프에 참가한 211명의 국어, 영어, 수학, 과학 교과목의 수행 점수를 SPSS로 분석하였다. 각 과목별 점수의 집중경향치 (평균, 표준편차, 최저 점수, 최고 점수)를 지역별로 그리고 성별로 산출하였다. 학생들의 출신 지역에 따른 점수 차이 검증은 일원변량분석으로, 학생들의 성별에 따른 점수 차이 검증은 t-검증으로 하였다. 그 결과 지역간에는 통계적으로 유의한 점수 차가 없었으나 성별 간에는 특정 교과목에서 통계적으로 유의한 점수 차가 있었다. 국어와 영어에서는 여학생이 남학생보다 더 높은 점수를 받았으나 수학에서는 남학생이 여학생보다 더 높은 점수를 받았다.

1. 지역간 점수 차이 검증

학생들의 국어, 영어, 수학, 과학 교과목의 수행 점수가 그들의 출신 지역(특별시/광역시, 시, 읍/면)에 따라 유의한 차이가 있는지 보고자 지역별로 평균과 표준편차를 산출하였고, 그리고 일원변량분석도 하였다. 과목별 일원변량분석 결과는 국어 [F(2, 208) = .1950, p<.82], 영어 [F(2, 208) = .5534, p<.57], 수학 [F(2, 208) = 2.8701, p<.06], 과학 [F(2, 208) = 2.6718, p<.07] 이었다.

분석 결과 학생들의 출신 지역에 따라 어떤 과목에서도 통계적으로 유의한 점수차가 나타나지 않았고 또한 네 과목의 점수를 통합한 총점에 있어서도 점수 차가 없었다 [$F(2, 208) = 2.4350, p < .09$]. 이 결과는 국어, 영어, 수학, 과학 점수에 있어서 학생이 대도시나 시 또는 읍/면에서 교육 받는 것이 수행평가에서 제시한 활동을 수행하는데 어떤 차이를 보이지 않는다고 해석될 수 있다.

그러나 어떤 교과에서도 통계적으로 유의한 지역간 점수 차를 보여주진 않았지만 수학[특별시(119점), 시(117점), 읍면(90점)]과 과학[특별시(202점), 시(206점), 읍면(188점)] 교과에서 대도시에 거주하는 학생의 성적이 읍면에 거주하는 학생의 성적보다 높은 것을 발견할 수 있었다.

2. 성별간 점수 차이 검증

수행평가에서 학생의 성별에 따라 통계적으로 유의한 점수 차가 있는지 보고자 과목별로 t-검증을 하였다(표 5 참조). 검증 결과는 국어 [$t(209) = -6.92, p < .000, 2$ -tailed], 영어 [$t(209) = -2.75, p < .007, 2$ -tailed], 수학 [$t(209) = 3.86, p < .000, 2$ -tailed], 과학 [$t(209) = 1.55, p < .123, 2$ -tailed]이었다.

분석 결과 국어와 영어에서는 여학생이 남학생보다 더 높은 점수를 받았지만 수학에서는 남학생이 여학생보다 더 높은 점수를 받은 것으로 나타났다. 이는 말하고, 듣고, 쓰고, 읽는 것과 같이 언어적인 능력에는 여학생이 남학생보다 더 우수하나 개념을 추상화하고, 공간을 지각하며, 정리를 발견하는 수학적 능력에는 남학생이 여학생보다 더 뛰어난 수행을 한다고 볼 수 있다.

〈표 5〉 수행평가 점수의 성별 평균, 표준 편차, t 값

과 목	성 별	평균 점수	표준 편차	점수 간격		t
				(최저)	(최고)	
국어	남학생	182	20.39	117	241	-6.92**
	여학생	203	21.19	147	241	
영어	남학생	70	11.76	25	92	-2.75*
	여학생	75	9.32	43	94	
수학	남학생	123	44.96	39	252	3.86**
	여학생	100	33.71	38	178	
과학	남학생	205	29.03	78	277	1.55
	여학생	199	21.09	159	250	
총점	남학생	581	77.65	377	773	.41
	여학생	577	55.10	458	718	

주: * $p < .01$ ** $p < .0001$

자료 입력 (남학생 = 1, 여학생 = 2)

N = 211 (남학생 n = 144, 여학생 n = 67)

과목별 최고 점수: 국어 350점, 영어 100점, 수학 300점, 과학 500점 (총점: 1250점)

3. 수행평가 점수의 상관 관계

가. 국어과 점수의 상관 관계

국어과에서 실시한 다양한 활동들의 관련성을 보고자 활동별 점수의 상관계수(r)를 산출하였다. 표6에서 보듯이 활동 2의 점수는 활동 4, 5, 6, 7과 유의한 상관을 보여주어 문학 작품을 감상하는 능력은 다양한 종류의 글을 쓰는 능력과 관계가 있으며, 활동 4는 활동 2, 3, 6과 유의한 상관을 보여주어 논설문을 쓰는 능력은 다양한 글을 읽거나 종합하여 글 쓰는 능력과 관련이 깊다고 해석할 수 있다.

활동 7은 활동 1, 2, 3, 5와 유의한 상관 관계를 보여주어 자신의 의견을 말로 표현하는 능력은 다양한 종류의 글을 이해하는 능력이 뛰어날수록, 또한 자신의 의견을 피력하여 글로 표현하는 능력이 뛰어날수록 더 잘 한다고 해석할 수 있다. 활동 1은 활동 7과만 상관이 있어 글을 이해하고 정보를 처리하는 능력은 말하기 능력과는 비례하지만 글로 표현하는 능력과는 상관이 없다고 말할 수 있다.

〈표 6〉 국어과 활동별 점수의 상관계수(r)

	활동1	활동2	활동3	활동4	활동5	활동6	활동7	총점
활동1	1.00							
활동2	.15	1.00						
활동3	.09	.05	1.00					
활동4	.15	.27**	.20*	1.00				
활동5	.10	.16*	-.03	.01	1.00			
활동6	.05	.22**	.12	.16*	.00	1.00		
활동7	.18*	.28**	.16*	.15	.21**	.15	1.00	
총 점	.44**	.51**	.49**	.60**	.29**	.57**	.61**	1.00

주: * p < .01 ** p < .001

읽기 활동 1 : 주제문 작성 및 중요도 평정하기

- 활동 2 : 고전 문학 작품 감상하기
- 활동 3 : 이어질 내용 예측하기
- 쓰기 활동 4 : 논설문 쓰기
- 활동 5 : 이야기 듣고 자신의 느낌 쓰기
- 활동 6 : 주어진 내용을 통합하여 논술하기
- 말하기 활동 7 : 주어진 주제로 말하기

나. 영어과 점수의 상관 관계

영어과에서 실시한 다양한 활동간의 관련성을 보고자 활동별 점수의 상관계수(r)를 산출한 결과 (표7 참조) 독해력과 영작문 및 말하기/듣기 능력간에는 서로 서로 밀접한 관련성이 있는 것으로 나타났다. 그러므로 영어 교과에서는 어떤 한 유형의 활동에서 우수하면 다른 유형의 활동에도 우수할 것으로 기대할 수 있다.

국어에서는 읽고, 쓰며, 말하는 능력간에 일률적인 상관을 보이지 않았으므로 한 유형에서 우수하여도 다른 유형에서는 잘 못할 수 있겠지만, 영어에서는 독해력이 뛰어나면 영작문도 잘하고 영어 회화에도 능숙할 것이라고 말할 수 있다.

〈표 7〉 영어과 활동별 점수의 상관계수(r)

	독 해	영작문	말하기/듣기	총 점
독 해	1.00			
영 작 문	.42**	1.00		
말하기/듣기	.35**	.33**	1.00	
총 점	.82**	.79**	.65**	1.00

주: * p < .01 ** p < .001

다. 수학과 점수의 상관 관계

수학과에서 실시한 다양한 활동간의 관련성을 보고자 활동별 점수의 상관계수(r)를 산출하였다 (표8 참조). 단답형과 서술형 지필검사간에는 통계적으로 유의한 상관이 나타나 학생들의 수학적 능력을 측정하기 위해서는 단답형과 서술형검사 중에서 어떤 유형을 사용하더라도 무방하다고 할 수 있다. 왜냐하면 한 검사에서 높은 점수를 받으면 다른 검사에서도 높은 점수를 받을 것으로 기대할

수 있기 때문이다. 다양한 수학활동을 통해 측정된 점수도 두 종류의 지필검사 점수와 유의한 상관을 보여 주어 지필검사에서 우수한 점수를 내면 수학적 활동도 잘 수행하며, 그 반대로 수학적 활동을 잘 하면 지필검사 점수도 높을 것이라고 말할 수 있다.

<표 8> 수학과 활동별 점수의 상관계수(r)

	단답형	서술형	수학활동	총 점
단답형 지필검사	1.00			
서술형 지필검사	.39**	1.00		
다양한 수학활동	.40**	.40**	1.00	
총 점	.77**	.74**	.64**	1.00

주: * $p < .01$ ** $p < .001$

라. 과학과 점수의 상관 관계

과학과에서 실시한 다양한 활동간의 관련성을 보고자 활동별 점수의 상관계수(r)를 산출한 결과 (표9 참조) 과학적 지식은 탐구 능력과 상관이 있었지만 과제집착력/태도와는 상관이 없었다. 그러나 탐구 능력은 과학적 지식뿐만 아니라 과제집착력과도 밀접한 관계가 있었다. 이 결과는 과학적 지식과 개념의 양을 풍부하게 소유하고 또한 과제에 집착할수록 탐구 능력은 뛰어나며, 탐구 능력이 뛰어날수록 과제에 더욱 더 집착하게 된다고 해석할 수 있다. 그러므로 과제집착력은 과학적 지식만을 측정하는 지필검사만으로는 평가할 수 없으며 탐구 능력을 측정할 수 있는 수행평가를 통해서만이 평가되는 것이라고 말할 수 있다.

<표 9> 과학과 활동별 점수의 상관계수(r)

	지식/개념	탐구능력	과제집착력	총 점
과학적 지식/개념	1.00			
탐구 능력 / 기능	.35**	1.00		
과제집착력/ 태도	.07	.50**	1.00	
총 점	.82**	.65**	.48**	1.00

주: * $p < .01$ ** $p < .001$

4. 수행평가에 대한 설문 조사 결과

캠프를 통한 수행 평가로 고등학교 신입생을 선발한 것이 우리 나라에서는 처음 시도된 방법이었다. 따라서 이러한 평가 활동이 학생들에게 어떻게 인식되었으며, 장점과 개선해야 할 문제점이 무엇인지 파악하고자 설문 조사를 실시하였다.

1차 캠프에 참석한 108명의 학생들에게 국어, 영어, 수학, 과학 교과의 수행평가에 대한 의견을 묻고자 15개의 문항으로 구성된 설문지를 만들었다. 앞의 13개의 문항은 5단계 척도로 구성되었으며 (5-매우 그렇다, 4-약간 그렇다, 3-보통이다, 2-약간 아니다, 1-매우 아니다), 나머지 2개 문항은 서술형으로 구성되었다. 각 교과별 설문 조사 결과에서 괄호 안에 제시된 수치는 학생들이 응답한 5단계 척도에서의 평균치이다.

15개 문항으로 구성된 학생용 설문지의 내용은 다음과 같다.

- | | |
|--------------|---------------------|
| [문제의 체제] | ① 문제 조건의 명료성 |
| | ② 답안 조건의 명료성 |
| [문제의 난이도] | ③ 문제의 난이도 |
| | ④ 시간 배당의 적절성 |
| [사고력 요구 정도] | ⑤ 호기심 자극 정도 |
| | ⑥ 문제 해결 욕구 자극 정도 |
| | ⑦ 끈기성 요구 정도 |
| | ⑧ 독창적인 사고력 자극 정도 |
| | ⑨ 종합적 사고력 요구 정도, |
| | ⑩ 답안 작성시 사고력 요구 정도 |
| [답안 작성의 만족도] | ⑪ 활동별 답안 작성의 만족도 |
| [종합적인 느낌] | ⑫ 평가 활동에의 참여도 |
| | ⑬ 평가 활동에 대한 종합적인 느낌 |
| [장점과 문제점] | ⑭ 좋았던 점 |
| | ⑮ 개선해야 할 점 |

가. 국어과 수행평가에 대한 설문지 분석

국어 문제에 있어 문제 조건(3.8)과 답안 조건(3.4)의 명료성 및 문제의 난이도(3.1)는 보통 정도이나 문제 해결을 위한 시간(2.5)은 비교적 많이 부족했다고 응답하였다.

국어 문제가 호기심(4.2)과 문제 해결 욕구(4.1)를 자극하였고, 끈기성(4.4)과 깊이있는 사고력도 요구(4.7)했다고 응답했으나, 독창적인 사고력을 요구(3.9)하거나 표현(3.6)하게 하는데는 보통 수준이었다고 평하였다.

많은 학생들이 국어 캠프 활동에 열심히 참여하였고(4.5) 국어 평가가 기존에 치르던 시험보다 내용과 방법이 좋다고(4.6) 응답하였다. 답안 작성의 만족도를 물어 본 결과, 작문에는 보통(3.0)이지만 말하기에는 약간 불만족(2.5)했다고 응답하여 학교에서 경험해 보지 못한 말하기 활동에는 그들 스스로 만족스럽게 수행했다고 느끼지 않는 것을 알 수 있다.

국어과 수행평가에서 좋았던 점을 물어본 결과(표10 참조), 많은 학생들이 '말하기와 같은 표현 능력의 평가', '듣기 평가와 같은 새로운 평가 방식', '작문 능력을 평가하는 문제', '종합적인 사고력을 요구하는 문제', '다양한 내용으로 흥미를 유도하는 활동' 등을 좋게 평가하였다.

개선할 점에 대해서는 '시간 부족', '작문이 너무 많음', '너무 어려움', '말하기 평가시 준비 시간의 부족', '토론식 평가 활동의 부족' 등을 지적하였다 (표11 참조). 학생들은 기존의 시험과 달리 긴 글을 쓰거나 말로 조리있게 표현하거나 스스로 문제를 해결하도록 요구하는 유형의 평가에 익숙치 않아 부담을 느꼈으며, 말하기에서 집단 토론을 통한 평가가 없었던 점에 다소 불만이었던 것으로 보인다. 특히 비슷한 유형의 활동이 많아서 지루했다는 평가도 있었는데 이는 언어 분야의 수행평가가 읽기를 하더라도 이해력이나 비판력 및 종합력을 보기 위해서는 결국 말하거나 쓰는 활동으로 평가되기 때문인 것으로 보인다. 그러므로 언어 능력을 평가하고자 할 때 다양한 평가의 유형과 매체의 활용 방안을 계속 연구해야 할 것이다.

〈표 10〉 국어과 수행평가에서 좋았던 점

장 점	응답자 수
말하기 등의 자기 표현 능력의 시간이 있는 점	38
처음 접하는 시험 방식 (듣기 평가 등)	38
작문 능력을 늘릴 수 있음	26
종합적인 사고력을 요구하는 문제임	25
생각하면서 쓸 수 있는 문제임	22
다양한 내용으로 흥미를 유도함	16
진정한 평가라고 생각함	11

〈표 11〉 국어과 수행평가에서 개선되어야 할 점

개선되어야 할 점	응답자 수
시간이 너무 모자람	55
작문이 너무 많았음	19
너무 어려움	15
문제 설명이 정확했으면	13
말하기 시간이 너무 길었음	10
토론식 평가가 있었으면	7
말하기 평가시 준비 시간을 더 주었으면	6
지문이 너무 길었음	6
말하기에서 주제에 따라 난이도가 차이 남	5
비슷한 유형이 많아서 지루함	5

나. 영어과 수행평가에 대한 설문지 분석

영어 문제에 있어 문제 조건(4.1)과 답안 제시 요령(3.5)의 명료성은 보통 수준이었고, 문제의 난이도(3.6)와 문제 해결을 위한 시간의 적절성(3.6)에도 보통이라고 응답하였다.

영어 문제가 호기심(4.0)과 문제 해결 욕구(4.0)는 자극하였지만 끈기성이나(3.7) 깊이있는 사고력(3.8) 및 독창적인 사고(3.7)의 요구는 보통 수준이라고 반응하였다.

많은 학생들이 영어 활동에 열심히 참여하였고(4.5) 영어 평가 문제가 학교에서 본 시험 문제보다 내용이나 방법이 더 좋다(4.5)고 응답하였다. 답안 작성의 만족도에 있어서 작문(3.4)과 독해(3.8)에는 보통이라고 응답했지만 말하기에는 약간 불만족(2.7)하다고 응답하여 평소 익숙치 않던 영어 회화에 어려움을 느꼈던 것으로 보인다.

학생들에게 영어과 수행평가에서 좋았던 점을 물어 본 결과, ‘이야기를 기억해서 영작하기’, ‘말하기 평가의 경험’, ‘OHP를 사용하는 평가 방식’, ‘다양한 유형의 활동’, ‘흥미로운 독해 지문’, ‘적절한 시간’이라고 응답하였다(표12 참조). 개선할 점은 ‘선명치 못한 OHP’, ‘시간 부족’, ‘너무 쉬운 문제’, ‘듣기 평가의 제외’, ‘사회성이 강한 문제 내용’, ‘짧은 지문’ 등으로 나타났다(표13 참조). 비교적 쉬운 문제의 출제나 적절한 시간 배당이 장점이라는 응답과 단점이라는 응답이 동시에 나타난 것으로 보아 영어과 평가 활동이 우수한 학생들에게는 문제가 충분히 어렵지 않았고 시간도 남았는데 비해 다른 학생들에게는 그 반대였던 것으로 분석된다.

〈표 12〉 영어과 수행평가에서 좋았던 점

장 점	응답자 수
선생님이 친절함	45
이야기를 기억하여 영작한 점	29
말하기 평가의 경험	29
OHP를 이용한 새로운 시험 형식	29
다양한 유형의 평가 활동	29
흥미로운 독해 지문	18
시간이 적절함	11
쉬운 문제의 출제	8
시험 방식의 순발력과 긴장감	8

〈표 13〉 영어과 수행평가에서 개선되어야 할 점

개선되어야 할 점	응답자 수
OHP가 선명하지 못함	21
시간이 부족함	13
듣기 평가의 시간이 없는 점	10
독해 문제가 너무 쉬움	10
말하기 주제가 다양하지 않았음	9
단문이 많고 장문이 적은 점	9
독해 문제 유형이 평이했음	5
외국인과의 대화가 없는 점	5
학생들의 말하기 수준 차가 심한 점	5

다. 수학과 수행평가에 대한 설문지 분석

수학 문제에 있어 문제 조건(3.4)과 답안 제시 요령(3.3)의 명료성은 보통이었으며, 문제의 난이도(3.0) 및 시간 배당의 적절성(3.0)도 보통이라고 응답하였다.

수학 문제가 호기심(4.2)과 문제해결 욕구(4.3)를 자극하였고, 끈기성(4.6), 깊이있는 사고력(4.8), 독창적인 사고력(4.1)도 비교적 많이 요구했다고 응답하였다.

많은 학생들이 수학과 활동에 열심히 참여하였고(4.4), 평가 문제가 학교에서 본 시험 문제보다 내용이나 방법이 좋았다고(4.5) 응답하였다. 답안 작성의 만족도에서는 개별 서술형검사(3.1)와 구두 시험(3.0)에는 보통이라고 응답하였지만 빨대 문제(2.3), 스케치패드(2.5), 퍼즐(2.5)에는 불만족하다고 응답하여 새로운 유형의 활동에는 자신이 만족할만큼 수행하지 못한 것으로 보인다.

학생들에게 수학과 수행평가에서 좋았던 점을 물어 본 결과(표14 참조), 많은 학생들이 '퍼즐이 많아서 재미있었음', '스케치패드를 배운 점', '빨대 구조물 만들기', '사고력을 요하는 문제', '문제 내용의 참신성', '구두 시험으로 자신을 나타낼 수 있었던 점', '다양한 평가 방식'을 좋게 평가하고 있는 것으로 나타났다.

개선할 점에 대해서는 '스케치패드가 너무 어려웠음', '구두 시험이 불필요함', '시간 부족', '문제 풀이가 없는 점', '이동이 너무 잦음' 등을 지적하였다(표15 참조). 학생들은 스케치패드와 같은 컴퓨터 시뮬레이션 수행평가가 어렵기는 하지만 재미있고 유익한 활동으로 생각했으며, 많은 학생들을 대상으로 짧은 시간에 실시했던 구두시험은 일부 학생들에게 불만족했던 것으로 판단된다. 짧은 시간에 자신의 능력을 충분히 발휘하지 못한 것이 아쉽고, 문제 풀이 방법을 알기 원하며, 진행이 너무 빠르고 시간이 모자란다는 점에도 주목할 필요가 있겠다.

〈표 14〉 수학과 수행평가에서 좋았던 점

장 점	응답자 수
퍼즐이 많아 흥미있음	27
스케치패드를 배운 것이 흥미있음	25
빨대 구조물 만들기	22
사고력을 요하는 문제	18
문제 내용이 참신함	16
구술 시험에서 능력을 발휘할 수 있는 점	13
다양한 유형의 평가	9
개별 과제 해결이 재미있음	5

〈표 15〉 수학과 수행평가에서 개선되어야 할 점

개선되어야 할 점	응답자 수
스캐치패드가 너무 어려움	30
시간이 부족하고 진행이 너무 빠름	22
구두시험이 필요하지 않음	18
문제 풀이를 해주었으면	18
이동이 너무 잦음	9
퍼즐이 어려움	7
지필검사 문제가 너무 어려움	5
밤에 문제를 주어 잠을 자지 못한 점	5

라. 과학과 수행평가에 대한 설문지 분석

과학 문제에 있어 문제 조건(3.8)과 답안 제시 요령(3.5)의 명료성은 보통이었고, 문제의 난이도(3.5)나 시간 배당의 적절성(3.5)도 보통이었다고 응답하였다.

문제는 호기심(4.3)과 문제 해결 욕구(4.2)를 자극하였고, 깊이있는 사고력(4.5)과 독창적인 사고력(4.2) 및 끈기성(4.5)도 많이 요구하였다고 응답하였다.

많은 학생들이 과학과 활동에 열심히 참여하였고(4.6) 평가 문제가 학교에서 본 시험 문제보다 내용이나 방법이 좋다고(4.5) 응답하였다. 답안 작성의 만족도에서는 객관식 지필검사(3.3), 주관식 지필검사(3.2), 공동 탐구 및 보고서 작성(3.4), 발표 및 토론(3.5), 마당놀이(3.4), 무대 공연(3.2)에는 보통이라고 응답한 반면, 생물 및 화학 실험(2.6)에는 불만족한 것으로 드러났다. 학생들이 지금까지 많이 경험해 본 지필검사에는 만족스런 수행을 하였지만 개별적으로 또한 스스로 알아서 실험하는 활동에는 자신이 만족할 만큼 수행하지 못한 것으로 보인다.

학생들에게 과학과 수행평가에서 좋았던 점을 물어 본 결과(표16 참조), 많은 학생들이 ‘공동 토론을 통한 문제 해결’, ‘사고력, 창의력, 상상력을 자극하고 요구하는 문제’, ‘재미있고 유익한 과학 공연’, ‘다양하고 재미있는 문제와 실험’, ‘마당 놀이’, ‘실험을 통한 평가’, ‘평소에 보지 못했던 과학 완구를 가지고 실험한 점’, ‘학교에서 배우지 않은 새로운 문제들’, ‘시험이외의 다양한 프로그램’, ‘개인별로 실험한 점’, ‘다양하고 풍부한 실험 재료’ 등을 좋게 평가하고 있는 것으로 나타났다. 학생들은 다양하고 새로운 유형의 활동이나 개인별로 실험한 경험, 그리고 완구로 긴 시간 연구하고 토론한 활동에 매우 만족했던 것으로 보인다.

개선할 점은 '시간 부족' (실험, 시험, 보고서 작성, 수수깡 다리 만들기, 무대 공연 준비), '바쁜 시험 일정', '불명확한 지시 사항' (실험, 답지 작성, 토론 방법, 보고서 작성), '너무 어려움', '실험 기구 부족', '실험을 위한 정보 미비' 등으로 나타났다(표17 참조). 학생들은 2일간의 과학과 캠프가 지속적으로 진행되어 시간의 부족함을 많이 느낀 것으로 보인다. 또한 새롭게 접해본 다양한 유형의 수행평가가 재미있고 유익하지만 어떻게 해야 하는지 스스로 알고 진행하기가 어려웠다는 응답은 그들이 과정적 지식의 평가에 익숙치 않아 심리적인 압박을 받은 것으로 보인다.

〈표 16〉 과학과 수행평가에서 좋았던 점

장 점	응답자 수
공동 토론을 통한 문제 해결	48
한 과제를 장시간 깊게 생각한 점	43
과학 공연이 재미있고 유익했음	35
개인별 실험, 과정을 각자 설계한 실험	35
마당 놀이가 재미있고 배운 것을 활용한 점	24
실험을 통해 평가한 점	22
학교에서 배우지 않은 새로운 유형의 문제 제시	14
시험이외의 다양한 프로그램	8
풍부하고 완벽한 실험 준비 및 재료	6

〈표 17〉 과학과 수행평가에서 개선되어야 할 점

개선되어야 할 점	응답자 수
시간 부족 (실험, 보고서 작성, 무대 공연, 다리 만들기에서)	68
불명료한 지시 사항 (실험, 답안지 작성, 발표와 토론, 보고서 작성)	29
일정이 너무 바쁨	20
너무 어려움, 모르는 개념이 있음, 문제가 너무 추상적임	18
실험을 위한 자료(책)가 부족함	8
불필요한 마당 놀이가 있음	5
보다 많은 실험을 하였으면	6
과학 공연의 주제를 변경하기 어려움	5

5. 수행평가 결과의 종합

캠프에서 다루어진 다양한 활동이 학생들에게 어떻게 느껴졌으며, 또한 영재이거나 영재일 가능성이 있는 학생을 선발하려는 평가의 목적이 어느 정도 달성되었는지 보고자 활동간 상관을 내보았고 설문지로 학생들의 반응도 분석해 보았다. 이러한 분석은 우수한 학생을 최소한의 오류로 선발하기 위해 수행평가를 계속 활용해야 하는지, 그리고 어떤 경우에 어떤 활동으로 측정해야 오류를 최소화할 수 있는지에 대한 정보를 제공해줄 것이다.

활동별 점수를 분석한 결과, 국어에서는 글을 읽고 듣는 것과 같은 이해 능력과 말하고 쓰는 표현 능력간에 일률적인 상관을 보여주지 않았지만 영어에서는 일률적인 상관을 보여주었다. 이는 언어 능력 중에서도 모국어와 외국어를 이해하고 표현하는 능력간에 차이가 있으며, 같은 언어에서도 이해력과 표현력에 차이가 있음을 시사해주므로, 어떠한 영재를 선발하려는가에 따라 수행평가 활동이 달라져야 한다고 말할 수 있다.

수학에서는 선다형검사가 아니라면 어떤 유형의 수행평가를 하여도 학생들의 수학적 능력을 파악하는데 무리가 없다고 볼 수 있지만, 다양한 수학적 활동간에 어떤 활동이 더욱 더 수학적 능력을 측정해주는지는 추수 연구로 밝혀져야 할 것이다. 또한 기하나 대수와 같은 세분화된 수학적 능력을 가장 정확하게 측정해줄 평가 유형이 무엇인지에 대한 추수 연구도 필요하다고 본다.

창의적 산출물을 생산하는 과학자는 과학 분야의 지식이외에 탐구력과 과제집착력과 같은 특성을 지니고 있다. 이 연구에 의하면 과제집착력이 지필검사만으로는 평가될 수 없다는 결과를 보여주어 장래 성취할 가능성이 있는 과학 영재를 선발하려면 다양한 유형의 수행평가를 활용해야 한다고 말할 수 있다.

학생 설문지를 분석한 결과, 학생들은 4개 교과에서 제시된 문제와 활동이 자신들의 호기심과 문제해결 욕구를 자극하였으며, 또 고급 사고력과 창의적인 문제해결력 및 탐구력도 요구했다고 반응하였다. 그리고 그들은 문제를 해결하고 싶은 욕구로 인해 거의 모든 캠프 활동에 열심히 참여하게 되었다고 하였는데, 그 주된 이유가 학교에서 지금까지 경험해 보지 못한 문제와 다양한 유형의 수행평가가 흥미를 유발시켰고 호기심을 자극하였기 때문이라고 한다. 또한 가상적이 아니고 실제 문제를 해결해야 하는 상황이었고 스스로 과정을 통제하면서 해결해 나갈 수 있었기 때문에 더 적극적으로 참여하게 되었다고 하였다.

학생들이 처음 접해보아 흥미가 유발되고 호기심을 자극하며 재미있었다고 평가한 수행평가의 유형은 국어와 영어의 경우 특정 주제에 대한 의견 발표와 질의 응답이었고, 수학은 퍼즐, 빨대 제작, 컴퓨터 시뮬레이션이었고, 과학은 실험, 공동 탐구, 완구에 대한 집단 토론이었다. 이러한 활동

들이 학생들의 호기심을 자극하면서 동시에 평가의 목적을 제대로 달성하는지에 대해서는 지속적인 연구가 따라야 할 것이다. 학생들은 비구조화되고 잘 정의되지 않아서 깊이있게 오랫동안 생각해야 하는 문제가 어렵기는 했지만 자기들이 경험한 어떤 평가 방법보다 자신의 능력을 최대한 발휘하게 하므로 “진정한 평가”가 이루어질 것이라고 반응하였다.

캠프를 통한 수행평가에서 개선되어야 할 점으로는 이런 활동들을 하기에 주어진 시간이 충분치 못했고 해결해야 할 과제가 너무 많았다는 것이다. 그러나 과제 해결을 위한 시간이 충분했다는 반응과 부족했다는 반응이 엇갈려 나타난 것으로 보아 우수한 학생들에게는 주어진 시간이 적절했지만 창의적으로 문제를 해결하는 능력이 부족한 학생들은 그 시간이 부족했던 것으로 판단된다. 그러므로 선발하고자 하는 학생의 능력에 따라 평가 내용의 수준을 결정해야 하며, 문제 해결을 위한 시간도 적절하게 배당할 필요가 있다고 하겠다.

짧은 시간에 많은 정보를 처리하는 능력을 측정하기 위하여 영어과 평가에서는 OHP를 사용하였는데, 모든 학생에게 선명하게 잘 보이도록 세밀한 부분까지 신경써야 공정한 평가가 이루어질 것이라고 판단된다. 이는 실험 실습, 말하기, 질의 응답과 같은 유형의 활동뿐만 아니라 지필검사에서도 표준화된 시험 환경을 제공하는데 각별한 주의가 요구됨을 의미한다.

V. 결 론

21세기의 영재 교육이 지향하고 있는 창의적인 인재의 양성을 위해서는 우선 뛰어난 성취할 잠재력이 있는 학생을 선발하는 일에 초점을 맞추어야 한다. 영재를 선발하기 위한 새로운 방안을 생각해야 하는 이유는 기존의 선다형 지필검사가 지식의 양을 단편적으로 평가하기는 하나 성취를 위해 필수적인 지식 활용의 효율성, 창의적인 사고, 과제집착력과 같은 요소를 평가하는데 많은 한계를 갖고 있기 때문이다. 그러므로 본 연구에서는 기존의 선다형 지필검사를 대신할 평가 방법으로 수행평가를 활용할 방안을 다각도로 모색해 보았다.

본 연구에서 수행평가를 활용해 본 결과, 수행평가가 영재의 주된 특성인 창의성이나 풍부하고 폭넓은 지식과 같은 인지적 측면뿐만 아니라 과제집착력이나 내적 동기와 같은 성격적 측면까지도 평가해주는 장점이 있음을 확인할 수 있었다. 그러나 수행평가가 영재 교육 목표에 부합하는 평가로서 갖는 타당성이나 대안적 평가로서 그 가치가 인정되지만 그에 못지 않게 학생 선발에서나 학교 교육 현장에서 수행평가의 활용을 가로막는 요소들도 존재한다는 것을 발견하였다.

한국교육개발원에서 민족사관고등학교 신입생을 선발하기 위해 개발한 다양한 수행평가 활동을

실행해 본 결과에 더해 영재 선발을 위한 대안적 평가로서 수행평가가 갖는 문제점을 밝히고 그리고 그 문제점을 해소할 수 있는 방안을 제안하고자 한다.

(1) 수행평가는 학생이 습득한 지식의 양을 타당하게 평가해주지만 창의적 문제해결력과 창의성 및 잠재력을 어느 정도 타당하게 측정해주는지 충분하게 검증되지 않았다는 문제점을 갖고 있다. 이 문제는 (가) 수행평가의 수행과 미래에 성취하는 정도에 대한 연구나 (나) 어떤 유형의 수행평가가 성취를 예언하는지에 대한 연구를 지속적으로 진행시켜서 그 타당성을 검증받아야만 근본적으로 해결될 것이다. 이러한 종단적 연구가 수행되기 전까지는 다양한 유형의 평가를 지속적으로, 그리고 다단계로 실시하여 이러한 문제점을 보완해야 할 것으로 보인다. 또한 수행평가와 선다형 지필검사를 병행하는 것도 고려해 볼 수 있는데, 그 이유는 선다형 지필검사가 수행평가보다 더 경제적이고 습득한 지식의 양도 단시간에 측정해주는 장점을 갖고 있기 때문이다.

(2) 수행평가는 평가 과정이나 평가 결과가 선다형 지필검사보다 신뢰롭지 않을 수 있다는 문제점을 갖고 있다. 논술 평가에 있어서의 문제점을 예로 들면, (가) 한 학생의 논술에 평가자마다 다른 점수를 주는 경우(평가자간 신뢰도 문제), (나) 여러 학생의 논술 점수를 한 평가자가 다른 시각과 견해로 평가하는 경우(평가자내 신뢰도 문제), (다) 여러명의 평가자가 여러명의 학생을 각기 달리 평가하는 경우(평가자간과 평가자내 신뢰도 문제)가 있다. '다' 항의 예로는 A 평가자가 'ㄱ' 학생의 수행을 평가할 때 B 평가자는 'ㄴ' 학생을 각자 다른 평가틀로 평가하는 경우이다. 이러한 평가자 신뢰도 문제를 해결하여 공정하고 신뢰로운 평가가 이루어지려면 평가자간의 주관적인 판단이 개입되지 않도록 평가틀을 상세화하거나, 사전에 여러 다양한 경우를 모의하여 훈련하고 협의하는 과정을 거치거나, 또는 평가자간의 평가 결과를 비교 검토하면서 재조정하는 단계를 거칠 수 있다.

(3) 수행평가의 내용과 방법을 개발하고 실행하려면 기존의 지필검사보다 많은 비용이 소요되므로 수행평가를 활용화하고 보편화하는데 제약이 따른다는 문제점이 있다. 또한 장시간이 소요되는 과제를 부여하거나 넓은 공간 및 실제적인 상황의 공간을 활용해야만 바람직한 평가를 할 수 있는 경우가 많다.

경비는 단독 학교가 수행평가를 개발하고 진행하기 보다 같은 지역에 있는 학교나 같은 목적을 가진 학교가 공동으로 진행하면 많이 절감할 수 있다. 장 시간이 요구되어 시간적인 제약이 따르는 수행평가는 컴퓨터 시뮬레이션 기법을 활용하여 실제 상황과 비슷한 문제를 주어 짧은 시간 동안 평가하는 방안이 있다. 그리고 실제 상황의 공간을 활용해서 평가해야 하는 문제는 방학을 이용해서 같은 지역에 소재한 학교나 또는 같은 목적을 가진 학교가 공동으로 실시하여 장소를 확보하는 방안, 대학이나 연구소 및 지역사회의 여러 기관의 협조를 받아 실시하는 방안, 그리고 실제 장소에 캠프를 열어 실시하는 방안을 고려할 수 있다.

(4) 영재를 선발하기에 타당한 수행평가를 개발할 수 있는 전문가의 수가 적을뿐만 아니라, 개발된 수행평가를 활용해 신뢰롭게 평가할 수 있는 전문가의 수도 부족하여 수행평가의 질적 수준을 통제하기 어렵다는 문제점을 갖고 있다. 이 문제를 근본적으로 해결하려면 영재를 어떻게 판별하고 교육해야 하는지를 알고 있는 영재 교육 전문가를 양성하는 일과, 다양한 교과 영역에서 수행평가를 개발할 교과 전문가를 양성하는 일, 그리고 이를 활용해 신뢰롭게 평가할 수 있는 평가자를 양성하는 일이 동시에 이루어져야 하겠다. 우리 나라는 아직 수행평가를 활용해서 영재를 선발할 수 있는 방안과 전략을 갖고 있는 전문가가 부족한 상태이므로 여러 학교가 공동으로 운영하여 소수의 전문가를 최대한으로 공유하는 방안을 고려할 수 있다.

본 연구의 결과, 단편적인 지식의 습득 수준만을 측정해주는 선다형 지필검사의 단점을 보완하여 창의성이나 문제해결력등 고급사고력을 평가하고 과제집착력과 동기 유발과 같은 인성적 측면을 평가하는데 수행평가가 활용될 수 있다는 확신을 갖게 되었다. 그러나 영재 선발을 위한 도구로서 수행평가가 전국적으로 활용되는데는 평가의 타당성이나 신뢰성 문제, 시·공간 및 경제적 제약성, 그리고 전문가의 부족과 같은 문제점이 있는 것도 발견하게 되었다. 창의력과 잠재력이 뛰어난 영재를 최소한의 오류로 그리고 타당하고 신뢰로운 방법으로 선발할 수 있도록 다각도로 또한 지속적으로 수행평가의 활용 방안을 연구해야 할 것이다.

참 고 문 헌

- 국립교육평가원 (1996). 수행평가의 이론과 실제. 서울: 국립교육평가원.
- 남명호 (1996). 초등교육에서 수행평가의 활용: 과제와 전망. 초등교육연구. 제10집, 87-100.
- 배호순 (1997). 포트폴리오 평가에 관한 분석적 고찰. 수행평가의 이론과 실제. 한국교육평가연구회 춘계 학술 세미나 발표 논문집, pp. 81-124.
- 백순근 (1997). 수행평가의 이론적 기초. 수행평가의 이론과 실제. 한국교육평가연구회 춘계 학술 세미나 발표 논문집, pp. 3-44.
- 조석희, 오영주, 김홍원, 박경숙 (1996). 민족사관고등학교 장학생 선발 체제 개발 연구. 한국교육개발원.
- Bloom, B. S. (1985). Developing talent in young people. NY: Ballantine Books.
- Case, R. (1978). Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R. Siegler (Ed.), Children's thinking: What develops? (pp. 37-71). NJ: Erlbaum.
- Feldhusen, J. F. (Ed.). (1985). Toward excellence in gifted education. Denver: Love Publishing Co.
- Feldhusen, J. F., Asher, J. W., & Hoover, S. M. (1984). Problems in the identification of giftedness, talent, or ability. Gifted Child Quarterly, 28, 149-151.
- Howe, M. J. A. (Ed.). (1990). Encouraging the development of exceptional skills and talents. UK: The British Psychological Society.
- Office of Technology Assessment (1992). Testing in American schools: Asking the right questions, OTA-SET-519. DC: U.S. Government Printing Office.
- Renzulli, J. S. (1977). The enrichment triad model: A guide for developing defensible programs for the gifted. CT: Creative Learning.
- Renzulli, J. S. (1982). What makes a giftedness? Reexamining a definition. Phi Delta Kappan, 60, 180-184.
- Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), Children's thinking: What develops? (pp. 109-149). NJ: Erlbaum.
- Sternberg, R. J. (1984). What should intelligence tests test? Implications of a

triarchic theory of intelligence for intelligence testing. *Educational Researcher*, 13, 5-15.

Wiggins, G. (1993). Assessment, authenticity, context and validity. *Phi Delta Kappan*, 75(3). 703-713.

ABSTRACT

An Application of Performance Assessment for the Identification of Gifted and Talented Students

Oh, YoungJoo

(Korean Educational Development Institute)

The paper-and-pencil multiple choice test has been widely used to identify gifted and talented student in Korea. Such test have several drawbacks; such as not being able to assess creative problem solving abilities and high level thinking abilities which are important characteristics of gifted students.

The Korean Educational Development Institute (KEDI) developed a performance assessment which challenged traditional methods of identification. The 5-day summer camp was held to select gifted students for the Korean Minjok Leadership Academy. 211 students were evaluated in the aspects of creative problem solving abilities, high level thinking abilities, task commitments, and cooperativeness with various performance tasks, such as essay tests, conversations, oral examinations, computer simulations, puzzles, experiments, group discussions, debates, research reports, and games.

As a result, it was found that there were several limitations of the performance assessment in terms of low reliability, requiring high costs and many professionals, and taking long times and large spaces. Expanding and continuing research should follow in order for a performance assessment to use widely as an identification methods because it assesses 'true' abilities of each individual student.