

## 관찰·추천제에 의한 수학영재 선발 시 사용되는 교사추천서와 자기소개서 평가에 대한 다변량 일반화가능도 이론의 활용

김 성 연

Vanderbilt University

한 기 순

인천대학교

이 연구는 관찰·추천제에 의한 수학영재 선발 시 사용되는 자기소개서와 교사추천서를 활용하여 측정상황에서 발생하는 오차요인들의 상대적인 영향력을 살펴보고, 이를 바탕으로 채점 영역 내의 문항 수의 조건을 변화시킴으로써 일반화가능도계수를 최대화할 수 있는 효율적인 측정 조건을 탐색하는 방법을 제시하였다. 이를 위해 2011학년도 수도권에 소재하고 있는 대학부설 과학영재교육원 수학분야에 지원한 90명의 자기소개서와 교사추천서를 2명의 교사가 분석적 방법으로 채점한 자료에 대해 다변량 일반화가능도 분석을 적용하였다. 이 연구 결과 산출된 최적의 측정 조건은 특정 영재교육원에서 제작한 루브릭을 바탕으로 수학영재를 선발하기 위해 사용된 선발도구에만 국한되지만, 이 연구에서 적용한 방법론적 틀은 시·도 교육청 영재교육원, 영재학급 그리고 대학의 영재교육원 등 각 영재교육 현장에서 직접 제작한 선발 도구에 측정학적 특성을 바탕으로 최적의 측정 조건을 탐색하는 데 적용 가능하다.

주제어: 교사추천서, 수학영재 선발, 다변량 일반화가능도 이론, 자기소개서

### I. 서 론

최근 현 정부는 창의와 융합을 근간으로 하는 창조경제 시대에 필요한 창의성과 끊임없이 도전하는 열정을 갖춘 창의인재를 양성하기 위해 ‘창조경제를 견인할 창의인재 육성 방안’을 발표하면서 영재교육기관을 융합교육의 선도모델로 육성하여 창의적 융합영재를 양성할 것을 강조하고 있다. 구체적으로 연구 지원 활동 강화를 위해 대학·연구소와 교류하는 R&E (Research & Education, 연구·교육), I&D (Imagination & Development, 상상·개발) 프로젝트 학습을 일반고까지 확대할 예정이며, 과학기술특성화대학 내 융합형 AP (Advanced Placement, 대학과목 선이수제) 과정을 확대하여 영재학교·과학고와의 교육 연계를 활성화 시키며, 융합영재교육원 및 정보보호 영재교육원을 신설하고 확대할 것을 추진하고 있다. 또

한 학생 스스로 분야별 주제를 선택하여 수강할 수 있는 자율선택형 교육과정을 시·도 교육청 영재교육원과 중·고등학교의 영재학급에 도입할 예정이며, URP(Undergraduate Research Program, 이공계 학부생 연구지원 프로그램)를 운영하여 고등학교에서 대학으로 이어지는 영재 지원체계를 강화할 계획이라고 발표하였다(미래창조과학부, 2013). 이와 더불어 교육부(2013)는 영재교육 최적화를 통한 창조적 인재 양성을 핵심으로 하는 ‘제3차 영재교육진흥종합계획’에서 교사가 장기간 학생을 관찰하여 이들의 특성 및 재능을 발굴·추천하는 선발방식인 관찰·추천제 실시기관을 확대하겠다고 했으며, 다문화 학생 등 영재교육 소외자를 고려한 맞춤형 진단검사를 지원하는 등 영재선발 방법의 타당성을 제고하여 꿈·끼를 키우는 영재교육의 기회를 확대하겠다고 발표했다. 국가차원에서 국민행복 실현을 위한 영재교육이 정립될 수 있도록 구체적인 실현 방안을 제시하고 있는 현 시점에서 가장 우선시 논의되어야 할 부분은 영재선발과 관련된 부분일 것이다. 이는 단순한 선발이 아닌 제대로 된 판별을 통한 선발이 되지 않을 경우에는 학생의 부적응을 야기할 수 있으며, 오히려 영재교육의 발전을 저해할 수 있기 때문이다. 따라서 영재학생을 어떻게 선발하는가는 영재분야에서 가장 중요한 이슈일 수밖에 없다(김숙영, 2012; 신윤주, 2011; 이인호, 한기순, 2009; Baum, 1994; Renzulli & Reis, 1986; VanTassel-Baska et al., 2002).

‘제1차 영재교육진흥종합계획’ 기간(2003-2007)에는 대부분의 영재관련 기관에서 1단계는 학교장이나 교사에 의한 추천, 2단계에서는 논리적 사고력 검사나 창의적 문제해결력 검사를 실시, 그리고 이후 단계에서 기타 심층면접이나 구술검사를 통해 다단계 영재선발방법을 사용하였다. 그러나 1단계나 마지막 단계는 다분히 주관적인 성격 때문에 전체 선발 과정에 큰 영향을 미치지 못하였으며, 주로 2-3단계에서 치러지는 지필검사에 많이 의존했던 것으로 나타났다. 이러한 지필평가 결과에 의존함으로써 사교육 선행 학습으로 다져진 현재의 학업 성취 우수자를 선발하는 경향이 있었으며(김미숙, 2007), 영재교육기관에서 영재교육을 받고 있는 학생들의 가정환경적 측면이 일반 학생들보다 월등히 높다는 연구(김미숙 외, 2007)를 통해, 경제적 뒷받침이 되는 학생들이 다른 학생들보다 성장할 수 있는 기회를 더 많이 가졌을 가능성을 입증하였다. 따라서 ‘제1차 영재교육진흥종합계획’ 기간(2003-2007)동안의 영재선발은 가정의 경제적 상황에 영향을 많이 받은 학생들의 현재의 우수한 성취에 지나치게 의존하여, 상대적으로 잠재적 능력이 경시되었다는 지적을 받았다. 이에 ‘제2차 영재교육진흥종합계획’ 기간(2008-2012)에는 사교육이나 선행학습의 효과를 최소화하고 소외계층의 영재성 발굴을 위해 교사 추천의 강화 및 영재성 측정의 강화를 통해 잠재능력을 지닌 학생들을 선발할 수 있도록 1단계에서는 교사추천, 2단계에서는 영재성 검사, 3단계에서는 학문적성 검사 그리고 4단계에서는 면접을 활용하는 방안을 제시하였으며, 2009년 교육과학기술부는 전국 영재교육원의 교육대상자 선발방식을 일부 시범지역을 중심으로 지필고사 방식에서 관찰·추천제로 전환한다고 발표하였다(교육부, 2010). 관찰·추천제를 실시해서 영재를 선발한 기관의 비율은 2009년 3.44%, 2010년 12.95%, 2011년 42.16%, 그리고 2012년 48.35%까지 증가하였다(교육부, 2013).

‘제3차 영재교육진흥종합계획’ 기간(2013-2017)동안 교육부(2013)는 관찰·추천제를 실시

하는 기관의 비율을 2017년에는 70%로 확대하며, 영재선발의 신뢰성 강화를 위해 관찰·추천제와 관련해 영재행동 체크리스트, 창의적 문제 해결력 수행 관찰 도구 등을 활용해 나가겠다고 발표했다. 이와 같은 정책적 결정은 영재가 국가의 가장 중요한 인적 자원으로 인식되면서 별도의 사교육이나 선행학습을 통해 만들어진 영재가 아니라 타고난 영재성 및 잠재가능성을 지닌 영재교육대상자를 선발하겠다는 취지에서 비롯된 것이다. 이처럼 선발 방식을 변화시키으로써 현장에서의 영재 교육을 개선하고자 하는 측정주도 교수(measurement-driven instruction)는 교육 현장을 단기간에 변화시킬 수 있다는 점에서 효과적인 수 있다(성태제, 2010; 이현숙, 2012). 실제로 관찰·추천제의 도입으로 잠재적 영재에 대한 영재교육 기회를 확대하는 성과를 얻을 수 있었으며 동시에 대학부설 영재교육원 선발단계의 경쟁률도 2009년 6.94:1에서 2012년 2.69:1로 완화되는 추이를 보이는 것으로 나타났다. 그러나 선발 도구의 측정학적 측면에 대한 경험적 검증 없이 현장에 적용하도록 하는 것은 평가의 타당성을 위협할 소지가 있다. 교육부(2013)의 제2차 영재교육진흥종합계획의 한계를 보더라도 영재교육 평가체제 미흡으로 질적 수준 담보가 미흡하였으며, 영재교육기관의 특성과 현장을 반영한 합리적인 평가체제 미흡으로 컨설팅 및 평가 일관성에서 어려움을 겪었던 것들을 예로 들면서 관찰·추천제의 타당성과 신뢰도에 대한 미흡을 지적하였다. 또한 관찰·추천제와 관련하여 많은 선행연구들은 이 제도가 충분한 연구가 이루어지지 않은 상태에서 급작스럽게 도입되었기 때문에(신운주, 2011), 선발 도구의 개발 및 적용에 대한 자료와 선발에 대한 이해가 많이 부족함을 지적하며, 연구자들별로 관찰·추천제에 의한 영재선발에 사용된 선발도구에 대해 상관관계를 바탕으로 한 타당도, 그리고 고전검사이론을 바탕으로 한 채점자간 신뢰도나 Cronbach  $\alpha$ 를 분석하기도 하고(남윤모, 2012; 박해정, 2010; 유미현 외, 2011; 윤은정, 박윤배, 2012; 최은주, 2012), 그 결과를 토대로 수정된 선발 도구를 제안하기도 하였다(박해정, 2010; 신운주, 2011; VanTassel-Baska et al., 2002).

그러나 고전검사이론에서 측정의 신뢰도는 단순히 관찰대상의 검사결과 또는 관찰과정이 얼마나 안정적으로 일관성 있게 기록되는가에 초점을 제한하고 있기 때문에 측정상황에서 발생할 수 있는 여러 오차 요인에 대한 설명이 불충분하다는 지적을 받아왔다(김성숙, 김양분, 2001). 즉 관찰·추천제에 사용되는 선발도구를 평가함에 있어 학생의 영재성 이외에도 채점 방법, 채점자 특성, 채점 영역, 채점 영역 내 세부영역(문항), 점수 배분 등과 같은 요인들이 개별 학생의 점수에 영향을 미치는 것을 간과하고 있다. 이를 극복하기 위해서는 측정상황에서 발생할 수 있는 다중 오차요인(국면)을 동시에 분석하고, 평가점수에 대한 요인의 상대적 영향력을 산출하여, 일반화가능도계수와 함께 의사결정자에게 안정적인 점수를 얻기 위한 측정조건을 제시함으로써 신뢰도 추정과정을 향상시킬 수 있는 일반화가능도 이론을 적용하는 방법이 있다. 이를 분석한 연구로 김성찬 외(2012)는 관찰·추천제에 의해 수학 영재를 선발하는 데 사용되는 교사추천서와 자기소개서를 총체적 채점과 분석적 채점으로 각각 평가하여, 영재 학생들의 평가점수에 영향을 미치는 채점자, 채점 영역 그리고 문항으로 설정한 오차요인들의 상대적인 영향력을 살펴보았다. 또한 채점자 수, 채점 영역 수, 그리고 문항 수를 변화시키거나가면서 적정 수준의 일반화가능도계수를 얻을 수 있는 최적의 측정 조

건을 탐색하였다. 이처럼 측정학적 특성의 변화를 바탕으로 적정 수준 이상의 신뢰도를 확보하는 데 필요한 합리적이고 타당한 근거를 제시하는 방법을 연구하는 것은 매우 중요하다. 구체적으로 김성찬 외(2012)의 연구에서 사용된 교사추천서와 자기소개서의 루브릭은 채점 영역을 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역으로 구분하여 총체적 채점의 경우는 두 명의 채점자가 각 채점 영역의 점수를 20점으로, 분석적 채점의 경우는 두 명의 채점자가 각 채점 영역마다 5점 배점으로 채점하는 4개의 문항이 내재되어 있는 균형설계를 적용하였다. 그러나 각 영재기관들은 이미 제작되어 있는 각 채점 영역마다 문항 수가 다르거나, 각 문항마다 채점 배점이 다른 루브릭을 사용하거나(남윤모, 2012; 박혜정, 2010; 신윤주, 2011), 이미 제작되어 있는 루브릭을 바탕으로 연구자들이 속해있는 영재기관에서 받은 교사추천서와 자기소개서의 일부에 대해 파일럿 조사를 하여 각 기관에 맞는 루브릭을 제작하는 과정에서 불균형 자료를 사용하게 된다(류지영, 정현철, 2010; 우미란 외, 2012; 최은주, 2012). 또한 김성찬 외(2012)의 연구에서는 각 요인들에 대해 임의효과 모형을 적용함으로써 채점자, 채점 영역 그리고 문항에 대한 연구 결과를 일반화시키는 정도를 강하게 할 수 있었다. 그러나 실제 문항들은 전체 채점 영역이 아니라 각 문항들이 속해있는 채점 영역과만 관계가 있는 것으로 보는 것이 더 타당할 것이다(Brennan, 2001a). 즉 4개의 채점 영역으로 구성된 루브릭이 있고, 앞으로 이 루브릭을 구성하는 채점 영역에는 변화가 없다고 가정하면, 채점 영역은 측정절차를 반복해도 변하지 않을 것이므로 고정효과 국면으로, 반면에 루브릭에 포함된 각 채점 영역 별 내 문항, 학생 그리고 채점자는 무한 전집을 대표하여 표집된 것으로 가정하여 임의효과 국면으로 정의하는 혼합모형을 활용한 분석이 필요하다. 이 연구에서는 다변량 일반화가능도 분석(Brennan, 2001b)을 실시하여 측정점수의 오차점수에 기여하는 다중 오차요인(국면)들의 상대적 영향력을, 그리고 합성점수를 바탕으로 일반화가능도 계수, 의존도 계수를 산출함으로써 최적의 측정조건을 제시하는 방법을 제시하였다.

이 연구의 목적은 관찰·추천제에 사용되는 선발도구에 대해 특정 영재교육원이 제작한 루브릭에서 산출된 신뢰도 분석 결과를 다른 영재교육기관의 선발도구에 일반화시키기 위한 것이 아니다. 무엇보다 중요하게 이 연구는 현재 각 영재교육기관 현장에서 사용되고 있는 선발도구를 평가함에 있어 사용되는 채점자 수, 루브릭의 채점 영역 수, 문항 수 그리고 배점에 따른 신뢰도의 변화를 바탕으로 해당 선발도구에 대한 최적의 측정조건을 산출하는 방법을 제시하는 데 그 목적이 있다. 따라서 이 연구에서는 영재선발에 있어서 메타분석을 실시한 결과, 학생들의 평가 점수에 가장 큰 영향을 미치는 요인은 학생과 채점 영역 내 문항이었으며, 거의 영향을 미치지 않는 요인은 채점자라는 점(Jiang, et al., 1997), 분석적 채점 방법이 총체적 채점 방법보다 신뢰도가 높았다는 점, 그리고 분석적 채점에서 채점자 효과는 총체적 채점에 비해 상대적으로 작다(김성찬 외, 2012)는 점을 고려하였다. 이를 바탕으로 실제 분석에서는 분석적 채점 방법을 사용한 자료를 활용하고, 현실적인 여건상 채점자 수를 늘리는 것은 수용하기 힘든 조건이라는 점을 감안하여 채점자는 두 명으로 하고, 문항 수를 변화시키면서 일반화가능도 분석을 실시하였다. 이 연구에서 이용한 방법과 절차는 각 영재교육기관에서 사용하고 있는 선발도구를 측정하는 루브릭에 적용하여 최적의 채점 구

조, 즉 채점 영역 수, 문항 수 그리고 채점자 수 조합을 탐색하기 위한 방법으로 응용될 수 있을 것으로 기대한다.

예시를 통하여 살펴볼 수 있는 구체적인 연구문제는 다음과 같다.

첫째, 관찰·추천제에 의한 수학영재 선발 시 사용된 자기소개서와 교사추천서 평가에서 채점 영역별로 피험자 점수의 오차점수에 영향을 주는 요인들의 상대적 영향력은 어느 정도인가?

둘째, 관찰·추천제에 의한 수학영재 선발 시 사용된 16개의 문항으로 구성된 자기소개서와 교사추천서 평가에서 합성점수(composite score)를 바탕으로 최적의 일반화가능도계수를 얻을 효율적인 측정 조건 구성은 어떻게 이루어지는가?

셋째, 관찰·추천제에 의한 수학영재 선발 시 자기소개서와 교사추천서 평가에서 합성점수(composite score)를 바탕으로 현 수준의 일반화가능도계수를 얻을 효율적인 측정 조건 구성은 어떻게 이루어지는가?

## II. 이론적 배경

### 1. 교사추천서와 자기소개서

2011학년도부터 전국의 대학부설 영재교육원에서는 관찰·추천제로 영재 선발을 하였으며, 1차 전형에서 자기소개서, 교사추천서, 학생생활기록부, 행동특성척도, 학업계획서, 학부 모기록지, 그리고 영재교육기관에 따라 영재성을 입증하는 추가 자료를 사용하는 것으로 나타났다. 이 중 교사추천서와 자기소개서, 그리고 학생생활기록부는 영재 선발 시 거의 공통적으로 사용되는 자료이다(김성찬 외, 2012). 교사추천서는 현재 담임이나 교과 담당 교사 혹은 전 학년 담임이나 전 교과 담당교사가 작성할 수 있다. 교사들은 학생들을 학습장면에서 다른 학생들과 비교하면서 지켜볼 수 있는 사람으로 특정 학생의 학습 이해도, 응용력, 창의성 등에 대해 객관적으로 평가할 수 있는 위치에 있으므로 신뢰할 수 있다(Pfeiffer & Blei, 2008). 그러나 교사들은 자신에게 협조적이고 모범생인 학생들에게 보다 긍정적인 평가를 하고, 교사에게 비협조적이고 교실에서 문제를 일으키는 학생들에 대해 부정적인 견해를 보일 수 있다는 우려가 있다(최은주, 2012). 자기소개서는 학생이 스스로를 영재프로그램에 추천하는 형태를 띠게 되며, 다른 사람에 의해 얻을 수 없는 정보를 얻을 수 있다는 장점이 있다. 그러나 자아 개념, 흥미, 가치, 학교 내, 외에서의 활동과 성취 등에 대한 자기 설명서로써 주관적 개념이 개입될 소지가 있다(남윤모, 2012). 학생생활기록부는 학생의 전반적인 학교 생활과 학생의 동기화, 과제 집착력, 인내력, 집중력, 호기심과 같은 정의적인 특성을 알 수 있다. 이 중 교사추천서와 자기소개서 평가에 사용된 루브릭에 대해 구체적으로 살펴보면 다음과 같다(신윤주, 2011).

교사추천서와 자기소개서 평가에서 같은 루브릭을 사용한 연구들은 다음과 같다. 박혜정(2010)은 인지적 영역, 정의적 영역, 그리고 사회적 영역으로 구분하여 각각 부족, 보통, 우수, 그리고 매우우수의 4개의 수준으로 채점하였다(Gronlund & Waugh, 2009). 이 중 인지적 특성과 정의적 특성은 각각 15점, 20점, 25점 그리고 30점으로 사회적 영역은 5점, 10점, 15

점, 그리고 20점으로 구분하여 총 80점으로, 영재교육관련 석사학위를 가지고 있으며, 영재 학생을 지도한 경험이 3년인 교사 1인과 영재학생을 2년간 지도하고 있는 연구자 본인이 채점에 참여하였다. 김성찬 외(2012)는 영역별로 4문항으로 구성된 인지적 영역, 수학적 영역, 정의적 영역 그리고 창의적 영역을 5개의 수준으로 나누어 영재교육대학원생이면서 단위학교 영재학급 및 대학부설 영재교육원에서 지도경력이 있는 두 명의 교사가 평가에 참여하였다. 신운주(2011)는 12항목으로 구성된 학습특성, 각각 6항목으로 구성된 창의성과 과제집착력, 그리고 7항목으로 구성된 리더십으로 총 31개의 항목의 체크리스트를 루브릭으로 활용하였다. 또한 이 결과를 바탕으로 교사추천서의 루브릭으로 12개 항목의 학습특성, 3개 항목의 창의성과 과제집착력, 그리고 5개 항목의 리더십으로 구성된 루브릭을 추천하였다. 최은주(2012)는 인지적 특성, 정의적 특성, 사회적 특성, 그리고 창의적 특성으로 구성된 박혜정(2010)의 루브릭을 수정하여 영재교육을 지도하고 있는 교사 1인과 영재교육을 전공하고 있는 연구자 본인이 평가하였다. 이중 인지적 특성과 정의적 특성은 부족을 15점, 보통을 20점, 우수를 25점, 그리고 매우우수를 30점으로, 사회적 특성과 창의적 특성은 부족을 5점, 보통을 10점, 우수를 15점, 그리고 매우우수를 20점으로 하여 총 100점으로 채점하였다.

교사추천서의 루브릭으로 남윤도(2012)는 한국교육개발원 영재교육연구센터에서 만든 영재교육종합데이터베이스의 온라인 교사추천시스템을 활용하여, 20문항으로 구성된 영재행동 특성검사, 31문항으로 구성된 창의적 인성검사 그리고 9문항으로 구성된 수학 학문적성검사 체크리스트에 교사들의 관찰을 토대로 5점 평정척도에 입력한 값을 수치화하여 점수화하였다. 윤은정, 박운배(2012)는 교사추천서를 기술형 추천서와 체크리스트형 추천서로 구분하여 각각 교과교육학 교수, 교과내용학 교수, 그리고 초등교사 3명이 채점하도록 하였다. 류지영, 정현철(2010)은 KAIST 과학영재교육원에서의 교사들이 사용할 수 있는 체크리스트를 소개하였는데, 8개의 세부영역으로 구성된 인지적 특성, 6개의 세부영역으로 구성된 정의적 특성, 13개의 세부영역으로 구성된 창의력, 13개의 세부영역으로 구성된 수학영재 특성에 대해 각각 7점 척도로 측정하게 하였다. Pfeiffer와 Jarosewich(2003)는 초등학생을 위한 Gifted Rating Scales-School Form에서 지적인 능력, 학문적 능력, 창의성, 예술적 재능, 리더십, 동기화의 6개 영역을 각각 12문항으로 구성하여 총 72문항에 9점 척도를 사용하였다. 자기소개서의 루브릭으로 윤은정, 박운배(2012)는 3개의 문항으로 구성된 자기소개서를 각 문항별로 교과교육학 교수, 교과내용학 교수, 그리고 초등교사 3명이 채점하도록 하였다. 또한 신운주(2011)는 12항목으로 구성된 학습특성, 각각 6항목으로 구성된 창의성과 과제집착력, 그리고 7항목으로 구성된 리더십을 바탕으로 한 체크리스트를 루브릭으로 활용하여 채점한 결과를 바탕으로 자기소개서의 루브릭으로 7개 항목의 학습특성, 4개 항목의 창의성, 5개 항목의 과제집착력, 그리고 6개 항목의 리더십으로 구성된 루브릭을 추천하였다.

이상의 선행 연구 분석 결과, 관찰·추천제에서 사용되는 자기 소개서와 교사 추천서는 같은 루브릭을 사용하고 있는 경우가 적지 않으며, 연구에 사용된 루브릭의 채점 영역 수, 문항 수, 채점자 수 그리고 점수 배분에 대한 근거는 측정학적 특성의 변화에 대한 경험적인 근거 없이 선택적인 판단에 근거하여 경험적으로 제시되어 있는 경우가 대부분이었다. 또한

평가의 신뢰도에 대한 연구로는 문항만을 오차요인으로 갖는 기존의 Cronbach  $\alpha$ 가 대부분이며, 일반화가능도이론을 적용한 연구는 김성찬 외(2012) 뿐이었다. 그러나 김성찬 외(2012)의 연구는 균형 설계라는 한계가 있어 여전히 다양한 연구 설계를 적용해 평가의 신뢰성에 대한 객관성을 확보하는 것이 필요하다. 따라서 이 연구에서는 이러한 점들을 바탕으로 각 채점 영역별로 채점 결과에 영향을 미칠 수 있는 채점자, 채점 영역 그리고 문항 요인 뿐만 아니라 채점 영역 간의 공분산을 동시에 고려한 상태에서 자기소개서와 교사추천서 평가의 적정 신뢰도를 유지할 수 있는 최적의 측정조건을 탐색하는 방법을 제시하였다.

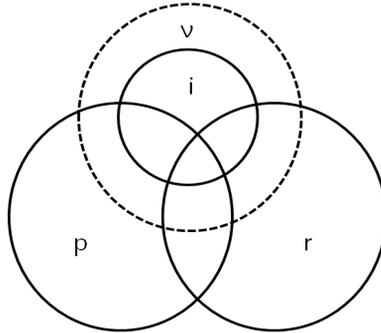
## 2. 다변량 일반화가능도 이론

일반화가능도 이론의 여러 설계 중 모든 학생들이 고정효과 국면의 각 채점 영역에 해당하는 여러 개의 점수를 가지고 있으면서, 고정효과 국면 각각과 관련된 임의효과 분산 성분을 가지도록 설계하여 분석하는 것을 다변량 일반화가능도 분석이라고 하며, 일변량 일반화가능도 분석은 다변량 일반화가능도 분석의 특수한 형태이다. 일변량 일반화가능도 분석에 비해 다변량 일반화가능도 분석을 사용하는 이유는 다음과 같다. 첫째, 고정국면과 임의국면이 혼합된 모형이면서 불균형 설계인 경우에 일변량 일반화가능도 분석을 실시하면 매우 복잡하지만, 다변량 일반화가능도 분석은 모든 채점 영역의 검사 결과를 동시에 분석할 수 있어 상대적으로 쉽게 적용할 수 있다. 둘째, 일변량 일반화가능도 분석에서 제공해주지 못하는 각 고정효과 국면의 채점 영역별로 관련되어 있는 임의효과 국면의 분산성분에 대한 추정 가능성이 뿐만 아니라, 공분산행렬을 제공해주며, 이를 바탕으로 고정효과 국면의 채점 영역별 전집점수에 가중치를 준 합성점수에 대해 오차 분산 및 신뢰도를 제공해준다(Brennan, 2001a). 특히 학생 효과의 공분산 행렬은 각 채점 영역들 간의 측정오차를 고려한 상관관계 정보를 알려주며, 이 정보는 각 채점 영역들이 측정하고 있는 구인들 사이의 관련성을 보여주기 때문에 수렴적 빛 변별적 타당도를 판단하는 자료로 사용된다(이영식, 신상근, 2004). 또한 오차요인들 사이의 관련성을 바탕으로 가장 합리적인 합성점수를 계산하는데 필요한 최적의 가중치를 결정할 수 있다(Brennan, 2001a; Cronbach et al., 1972; Webb et al., 1983).

### 가. G-연구 설계

이 연구의 G-연구는 관찰·추천에 의한 수학영재 선발 시 사용되는 교사추천서와 자기소개서를 평가하기 위해 각각 세부영역을 나타내는 4개의 문항으로 구성된 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역에 대해 각 채점 영역마다 두 명의 채점자가 평가 점수를 부여하는  $p \times r \times i$  [학생  $\times$  채점자  $\times$  문항] 모형을 설계하여 적용하였다. 다변량 일반화가능도 분석에서는 4개의 채점 영역( $\nu$ )에 대해 각각 구분하여 분석이 실시되기 때문에 고정국면인  $\nu$  국면은 설계상에 표시하지 않는다. 여기서 단린 원( $\bullet$ )은 해당 국면이 고정국면인 채점 영역 국면과 교차됨을 의미하며, 열린 원( $\circ$ )은 해당 국면이 채점 영역 국면에 내재됨을 의미한다. 각 국면별 상호작용을 이해하기 위해  $p \times r \times i$  설계를 도식

화하면 [그림 1]과 같다. [그림 1]은 일변량 일반화가능도 분석의 체제인  $p \times r \times (i: \nu)$  설계를 기본으로 하여 그려지지만,  $\nu$  국면이 각각 독립적으로 추정된다는 점을 반영하기 위해 점선으로 나타낸다는 점에서 차이가 있다. [그림 1]에서 실선으로 표시한 원이 분할되는 부분에는 분산성분이 존재하며, 실선과 점선이 교차되는 부분에는 공분산이 존재한다.



[그림 1]  $p \times r \times i^\circ$  의 G-연구 설계

$p \times r \times i^\circ$  설계에서  $\nu$  국면의 네 영역( $\nu_1, \nu_2, \nu_3, \nu_4$ )에 대한 평가 점수는 식(1)과 같이 분할된다. 여기에서  $\nu_k$  채점 영역에서의 문항  $i$ 에 대해  $r$ 번째 채점자가 평가한 학생  $p$ 의 점수는  $X_{priv_k}$ , 해당 채점 영역에서 전체 학생, 문항 그리고 채점자에 대한  $X_{priv_k}$ 의 평균은  $\mu_{\nu_k}$ , 그리고  $\omega, \xi, \psi, \zeta$ 는  $p, r, i$ 의 주효과와 상호작용 효과를 나타낸다.

$$X_{priv_1} = \mu_{\nu_1} + \omega_p + \omega_r + \omega_i + \omega_{pr} + \omega_{pi} + \omega_{ri} + \omega_{pri} \tag{1}$$

$$X_{priv_2} = \mu_{\nu_2} + \xi_p + \xi_r + \xi_i + \xi_{pr} + \xi_{pi} + \xi_{ri} + \xi_{pri}$$

$$X_{priv_3} = \mu_{\nu_3} + \psi_p + \psi_r + \psi_i + \psi_{pr} + \psi_{pi} + \psi_{ri} + \psi_{pri}$$

$$X_{priv_4} = \mu_{\nu_4} + \zeta_p + \zeta_r + \zeta_i + \zeta_{pr} + \zeta_{pi} + \zeta_{ri} + \zeta_{pri}$$

$p \times r \times i^\circ$  설계에서 임의국면에 대한  $\nu$  국면의 각 채점 영역별 분산 및 공분산 성분은 식 (2)로 정의된다. 식(2)에서 대각선 성분은 분산을, 대각선 아래 성분은 공분산을 그리고 대각선 위의 괄호로 표시한 성분은 측정의 오차를 고려한 상관계수를 나타낸다. [그림 1]에서와 같이,  $p$ 와  $r$ 은  $\nu$ 와 교차하므로,  $\sum_p, \sum_r$ , 그리고  $\sum_{pr}$ 에서는  $\nu$  국면의 각 영역별로 공분산 성분이 존재한다. 그러나  $i$ 는 내재되어 있으므로,  $\sum_i, \sum_{pi}, \sum_{ri}$ , 그리고  $\sum_{pri}$  행렬에서는  $\nu$  국면의 각 영역별 공분산성분은 존재하지 않는다.

$$\sum_p = \begin{bmatrix} \sigma_{\nu_1}^2(p) & (\rho_{\nu_1\nu_2}(p)) & (\rho_{\nu_1\nu_3}(p)) & (\rho_{\nu_1\nu_4}(p)) \\ \sigma_{\nu_2\nu_1}(p) & \sigma_{\nu_2}^2(p) & (\rho_{\nu_2\nu_3}(p)) & (\rho_{\nu_2\nu_4}(p)) \\ \sigma_{\nu_3\nu_1}(p) & \sigma_{\nu_3\nu_2}(p) & \sigma_{\nu_3}^2(p) & (\rho_{\nu_3\nu_4}(p)) \\ \sigma_{\nu_4\nu_1}(p) & \sigma_{\nu_4\nu_2}(p) & \sigma_{\nu_4\nu_3}(p) & \sigma_{\nu_4}^2(p) \end{bmatrix}, \quad (2)$$

$$\sum_r = \begin{bmatrix} \sigma_{\nu_1}^2(r) \\ \sigma_{\nu_2\nu_1}(r) & \sigma_{\nu_2}^2(r) \\ \sigma_{\nu_3\nu_1}(r) & \sigma_{\nu_3\nu_2}(r) & \sigma_{\nu_3}^2(r) \\ \sigma_{\nu_4\nu_1}(r) & \sigma_{\nu_4\nu_2}(r) & \sigma_{\nu_4\nu_3}(r) & \sigma_{\nu_4}^2(r) \end{bmatrix}, \quad \sum_i = \begin{bmatrix} \sigma_{\nu_1}^2(i) \\ & \sigma_{\nu_2}^2(i) \\ & & \sigma_{\nu_3}^2(i) \\ & & & \sigma_{\nu_4}^2(i) \end{bmatrix}$$

$$\sum_{pr} = \begin{bmatrix} \sigma_{\nu_1}^2(pr) \\ \sigma_{\nu_2\nu_1}(pr) & \sigma_{\nu_2}^2(pr) \\ \sigma_{\nu_3\nu_1}(pr) & \sigma_{\nu_3\nu_2}(pr) & \sigma_{\nu_3}^2(pr) \\ \sigma_{\nu_4\nu_1}(pr) & \sigma_{\nu_4\nu_2}(pr) & \sigma_{\nu_4\nu_3}(pr) & \sigma_{\nu_4}^2(pr) \end{bmatrix}, \quad \sum_{\pi} = \begin{bmatrix} \sigma_{\nu_1}^2(\pi) \\ & \sigma_{\nu_2}^2(\pi) \\ & & \sigma_{\nu_3}^2(\pi) \\ & & & \sigma_{\nu_4}^2(\pi) \end{bmatrix}$$

$$\sum_{ri} = \begin{bmatrix} \sigma_{\nu_1}^2(ri) \\ & \sigma_{\nu_2}^2(ri) \\ & & \sigma_{\nu_3}^2(ri) \\ & & & \sigma_{\nu_4}^2(ri) \end{bmatrix}, \quad \sum_{pri} = \begin{bmatrix} \sigma_{\nu_1}^2(pri) \\ & \sigma_{\nu_2}^2(pri) \\ & & \sigma_{\nu_3}^2(pri) \\ & & & \sigma_{\nu_4}^2(pri) \end{bmatrix}$$

$\nu$ 국면의 각 영역별 분산은 일반량 일반화가능도 분석에서 분산 성분을 추정하는 데 사용되는 공식과 같은 방법으로 추정되며,  $\nu$ 국면의 각 영역  $\nu_1$ 과  $\nu_2$ ,  $\nu_1$ 과  $\nu_3$ ,  $\nu_1$ 과  $\nu_4$ ,  $\nu_2$ 와  $\nu_3$ ,  $\nu_2$ 와  $\nu_4$ , 그리고  $\nu_3$ 와  $\nu_4$ 의 공분산 성분은 식(3)을 이용하여 각각 추정된다. 또한 학생 효과에서  $\nu$ 국면의 각 영역별 측정의 오차를 고려한 상관계수는 식(4)를 이용하여 각각 추정된다.

$$\sigma_{\nu_i\nu_j}(t) = \frac{\sigma_{\nu_i}^2(t) + \sigma_{\nu_j}^2(t) - \sigma_{\nu_i - \nu_j}^2(t)}{2} \quad (i \neq j, i = 1, 2, 3, 4, j = 1, 2, 3, 4) \quad (3)$$

$$\rho_{\nu_i\nu_j}(t) = \frac{\sigma_{\nu_i\nu_j}(t)}{\sqrt{\sigma_{\nu_i}^2(t)\sigma_{\nu_j}^2(t)}} \quad (i \neq j, i = 1, 2, 3, 4, j = 1, 2, 3, 4) \quad (4)$$

#### 나. D-연구 설계

D-연구는 G-연구에서  $\nu$ 국면의 각 영역별로 분산 및 공분산 성분이 추정되면, 이를 바탕으로 측정 상황에서 가장 효율적인 측정 절차를 결정하는 데 필요한 정보를 제공하는 연구이다(김성숙, 김양분, 2001). 이때 G-연구에서 추정한  $\sum_p$  행렬은 D-연구에서도 동일하게 적용되며, 그 외 행렬들은  $\nu$ 국면의 각 영역별로 정한 D-연구 표본 크기로 각각 나누어 D-연구에 사용할 행렬들을 산출한다. D-연구에서 각 오차국면은 일반화 전집에서의 평균에 대한 의미가 포함되어 있기 때문에 대문자로 표시한다. 식(2)의 분산 공분산 행렬을 이용한 D-연

구의 분산성분은 식(5)로, 공분산 성분은 식(6)과 같이 나타낼 수 있다.

$$\sigma_{\nu_i}^2(\bar{\alpha}) = \frac{\sigma_{\nu_i}^2(\alpha)}{\pi_{\nu_i}(\bar{\alpha}|\tau)} \tag{5}$$

$$\sigma_{\nu_i\nu_j}^2(\bar{\alpha}) = \frac{\sigma_{\nu_i\nu_j}^2(\alpha)}{\pi_{\nu_i\nu_j}(\bar{\alpha}|\tau)} \tag{6}$$

식(5)와 식(6)에서  $\bar{\alpha}$ 는 D-연구 설계의  $p, R, I, pR, pI, RI, pRI$  효과 중 하나를 나타내며,  $\sigma_{\nu_i}^2(\alpha)$ 는 G-연구에서의 분산성분을 의미하며,  $\tau$ 는 측정대상으로  $p$ 를 나타낸다.  $\sigma_{\nu_i}^2(\bar{\alpha})$ 는 고정효과 국면인  $\nu$ 국면의 각 영역에서 추정한 분산성분을  $\bar{\alpha}$ 에 표시된 각 효과에 대한 D-연구의 표본 크기를 나타내는  $\pi_{\nu_i}(\bar{\alpha}|\tau)$ 으로 나눔으로써 얻을 수 있다. 단, 측정대상인 학생의 표본크기는 제외한다.

일반화가능도 이론에서는 검사의 성격에 따라 오차 분산을 두 가지로 구분한다. 검사의 성격이 피험자의 상대적인 서열을 파악하는 데 주요 관심을 둔 검사라면 전집점수 편차와 전집점수 편차의 추정치인 관찰점수 편차와의 차이를 나타내는 상대오차 분산을 활용한다. 반면에 분할 점수에 의해 피험자를 여러 범주 중 하나로 구분하는 것이 주요 목적인 검사라면 전집점수와 전집점수의 추정치인 관찰점수와의 차이를 나타내는 절대오차분산을 활용한다.  $p \cdot r \cdot i$  설계에서  $\sum_p$ 행렬은 전집점수가 이루는 분산 행렬인  $\sum_r$ 행렬이 되며,  $\sum_{pR}$ 행렬,  $\sum_{pI}$ 행렬 그리고  $\sum_{pRI}$ 행렬을 더한 행렬은 상대오차 분산 행렬인  $\sum_{\delta}$ 행렬, 그리고  $\sum_{\delta}$ 행렬,  $\sum_R$ 행렬 그리고  $\sum_I$ 행렬을 더한 행렬은 절대오차 분산 행렬인  $\sum_{\Delta}$ 행렬이 된다. 위의 절차에 따라 상대오차 분산과 절대오차 분산이 구해지면, 고전검사이론의 진점수 분산과 유사한 전집점수 분산과 오차점수 분산의 합에 대한 전집점수 분산의 비로 신뢰도를 구한다. 이 때, 상대오차분산을 이용하면 식(7)과 같은 일반화가능도 계수를 얻을 수 있으며, 절대오차 분산을 이용하면 식(8)과 같은 의존도 계수를 얻을 수 있다(김성숙 외, 2012; 이현숙, 2012; Brennan, 2001a). 현 시점에서 수학영재를 선발하는 데는 상대평가가 이용되므로 일반화가능도계수가 적용되어야 하지만, 실제 분석에서는 추가 정보를 위해 의존도계수도 함께 제시하였다.

$$E\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pR}^2 + \sigma_{pI}^2 + \sigma_{pRI}^2} \tag{7}$$

$$\Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_R^2 + \sigma_I^2 + \sigma_{pR}^2 + \sigma_{pI}^2 + \sigma_{RI}^2 + \sigma_{pRI}^2} \tag{8}$$

### III. 연구방법

#### 1. 연구 자료

이 연구는 김성찬 외(2012)에서 사용한 자료 중 일부 자료로, 2011학년도 수도권에 소재하고 있는 대학부설 과학영재교육원에서 관찰·추천제로 수학영재를 선발하는 데 지원한 90명의 자기소개서와 교사추천서를, 2명의 채점자가 분석적 방법으로 각각 4개의 문항으로 구성된 4개의 영역에 5점 척도로 채점한 자료를 사용하였다. 교사 추천서에는 학생이 영재성을 보인 사례, 우수한 인성이 드러난 사례, 리더십이 드러난 사례, 종합 의견, 그리고 자기소개서에는 지원한 동기, 교육원에서 자신을 뽑아야 하는 이유, 수학에 관심을 갖게 된 동기, 장래 희망이 기록되어 있다. 4개의 영역은 각각 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역으로 구성되어 있으며, 각 영역별 채점 기준은 <표 1>과 같다. 또한 채점자들은 영재교육대학원생으로 단위학교 영재학급 및 대학부설 영재교육원에서 지도경력이 있는 교사들로서 평가에 전문성을 확보하도록 고려하였다. 또한 채점 시에는 시간 간격을 어느 정도 유지함과 동시에 동일 순서로 학생이 채점되지 않도록 임의로 순서를 바꾸어 채점하였

<표 1> 교사추천서와 자기소개서의 루브릭

영역	세부영역 (문항)	채점 기준	배점
인지적 특성	지적호기심	정보 습득이 빠르고 호기심과 궁금함이 많음	5
		정보 습득이 매우 느리며 궁금함이 전혀 보이지 않음	1
	집중력	집중하는 능력이 매우 우수함	5
		집중하는 능력이 매우 부족함	1
독서	독서를 즐기며 많이 읽음	독서를 즐기며 많이 읽음	5
		독서에 대한 언급이 전혀 없음	1
	자기주도 학습 능력	스스로 계획을 세워 학습하는 능력이 잘 갖추어져 있음	5
		스스로 계획을 세워 학습하는 모습이 전혀 없음	1
수학적 특성	수학교과 성적	항상 수학 교과 성적이 우수함	5
		수학 교과 성적에 대한 언급이 전혀 없음	1
	수학적 창의성	수학 문제 해결에 있어서 유창성, 융통성, 다양성, 정교성이 뛰어남	5
		수학 문제 해결에 있어서 유창성, 융통성, 다양성, 정교성이 매우 부족함	1
수학적 사고력	수학적 문제해결 능력	수학 문제를 해결하는 능력이 매우 뛰어남	5
		수학 문제를 해결하는 능력이 매우 부족함	1
	수학적 사고력	수학적 문제를 이해하고 해결하는 데 요구되는 사고능력이 매우 뛰어남	5
		수학적 문제를 이해하고 해결하는 데 요구되는 사고능력이 매우 부족함	1

영역	세부영역 (문항)	채점 기준	배점	
정의적 특성	도전정신	어려운 일에도 정면으로 맞서서 해결하려는 태도를 자주 보임	5	
		⋮	⋮	
	적극성	어려운 일을 해결하려는 태도에 대한 언급이 전혀 없음	1	
		어떤 일이 주어졌을 때 자발적으로 나섬	5	
	과제집착력	⋮	⋮	
		어떤 일이 주어졌을 때 거의 나서지 않음	1	
	수학적 적성	해결하기 힘든 문제를 끝까지 노력하여 해결함	5	
		⋮	⋮	
	사회적 특성	리더십	과제집착력과 관련된 언급이 전혀 나타나지 않음	1
			수학을 매우 즐기고 잘 하며 타고난 소질이 있음	5
도덕성		⋮	⋮	
		수학을 싫어하며 소질이 없음	1	
사교성		조직을 이끌어 가는 리더로서의 능력을 가지고 있음	5	
		⋮	⋮	
유머감각		조직을 이끌어 가는 리더로서의 능력이 없음	1	
		배려하고 나누고 봉사하는 태도를 가지고 있음	5	
유머감각		⋮	⋮	
		배려하고 나누고 봉사하는 태도가 없음	1	
유머감각	사회에 잘 적응하고 대인관계가 원만함	5		
	⋮	⋮		
유머감각	사회에 적응하지 못하며 대인관계가 나쁨	1		
	날카로운 유머 감각을 가지고 있고 자기 유머로 타인에게 영향력을 행사하려 함	5		
	⋮	⋮		
	유머 감각이 전혀 없음	1		

다(김성찬 외, 2012).

## 2. 자료 분석 방법

이 연구에서는 채점 영역을 인지적 영역, 수학적 영역, 정의적 영역, 그리고 사회적 영역의 네 영역으로 한정된 고정국면으로 정의하고, 채점 영역 내의 세부 영역을 나타내는 문항, 평가를 받는 학생 그리고 채점자는 무한 전집에서 임의로 표집된 것으로 가정하여 임의국면으로 정의하였다. 학생들은 각 채점 영역에 따라 문항은 다르지만, 각 채점 영역마다 두 명의 평가자가 채점한 평가 점수를 받게 되므로, 다변량 일반화가능도 분석의  $p \times r \times i^{\circ}$  설계를 적용하였다. 이 설계의 분산 성분과 공분산 성분을 추정하기 위한 G-연구, 그리고 이를 바탕으로 오차요인의 표본 크기를 조절하여 효율적인 측정조건을 탐색하기 위한 D-연구를 수행하기 위해 mGENOVA(A multivariate GENERALized Analysis Of VARIance System) 프로그램(Brennan, 2001b)을 사용하였다.

## VI. 연구결과

### 1. 교사추천서 평가와 자기소개서 평가에서의 G-연구 결과

#### 가. 교사추천서 평가와 자기소개서 평가에서의 분산과 공분산 성분 분석 결과

교사추천서 평가에서  $p \times r \times i$  설계의 각 효과에 해당하는 분산 성분 추정치와 전체 분산 중 해당 효과가 차지하는 비율은 <표 2>와 같다. 수학적 영역과 정의적 영역에서 학생효과가 37.9%와 41.0%로 가장 크게 나타났으며, 인지적 영역과 사회적 영역에서도 잔차효과를 제외하면, 학생효과가 24.6%와 25.0%로 가장 크게 나타났다. 잔차 효과가 크게 나타나는 이유는 연구 설계에 포함되지 않은 다른 요인들에 대한 간접적인 설명으로 G-연구 결과에서 흔히 나타나는 결과이다. 또한 학생 효과가 상대적으로 크게 나타났다는 것은 교사추천서 평가 결과에 학생의 영재성 차이가 반영되어 있다는 것을 의미하므로, 교사추천서가 변별력을 지닌 측정절차라고 해석할 수 있다. 인지적 영역에서는 채점자 효과가 13.4%, 그리고 문항 효과는 5.9%로 나타났다. 이는 학생들의 평가점수에 채점자의 엄격성과 관대함의 차이가 문항의 난이도 차이보다 큰 영향을 미친다는 것을 의미한다. 또한 학생과 채점자의 상호작용 효과는 13.0%, 그리고 학생과 문항의 상호작용 효과는 10.2%로 나타났다. 이는 학생들의 상대적 순위가 채점자에 따라 그리고 문항에 따라 달라지는 경우가 있음을 의미한다. 위의 두 결과를 함께 고려하면 인지적 영역에서는 채점자가 문항보다 학생들의 점수에 더 큰 영향을 미치는 변인이라고 해석할 수 있다. 수학적 영역과 정의적 영역에서도 학생과 문항의 상호작용 효과보다 학생과 채점자의 상호작용 효과가 크게 나타났으며, 채점자 효과가 문항 효과보다 같거나 크게 나타났다. 따라서 이 영역들에서도 채점자의 엄격성이 문항의 난이도보다 학생들의 점수에 더 큰 영향을 미치는 변인이라고 해석할 수 있다. 사회적 영역에서는 문항의 난이도가 채점자의 엄격성보다 크게 나타났으며, 학생들의 상대적 순위는 문항의 난이도보다 채점자의 관대함의 차이에 의해 달라지는 것으로 나타났다.

<표 2> 교사추천서 평가에서의 분산성분과 비율

효과	다변량 일반화가능도 이론			
	인지적 영역	수학적 영역	정의적 영역	사회적 영역
$p$ (학생)	0.193(0.246)	0.278(0.379)	0.257(0.410)	0.203(0.250)
$r$ (채점자)	0.105(0.134)	0.044(0.060)	0.003(0.005)	0.053(0.065)
$i$ (문항)	0.046(0.059)	0.002(0.003)	0.003(0.005)	0.082(0.101)
$pr$ (학생*채점자)	0.102(0.130)	0.137(0.187)	0.125(0.199)	0.112(0.138)
$pi$ (학생*문항)	0.080(0.102)	0.037(0.050)	0.038(0.061)	0.082(0.101)
$ri$ (채점자*문항)	0.013(0.017)	0.007(0.010)	0.000(0.000)	0.019(0.023)
$pri$ (학생*채점자*문항,잔차)	0.245(0.313)	0.228(0.311)	0.201(0.321)	0.261(0.321)

주. ( )는 전체 분산 중 해당 효과의 분산이 차지하는 비율임.

각 학생이 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역에서 받은 평가 점

수가 함께 변하는 정도를 나타내는 각 채점 영역 별 공분산 정보는 일변량 일반화가능도 분석에서는 제공되지 않으며, 다변량 일반화가능도 분석만이 제공해주는 정보이다. 이 정보는 각 채점 영역의 평가 점수에 영향을 미친 요인들이 각 채점 영역 간 공분산에 미치는 영향을 나타낸다. 교사추천서의 공분산 결과는 <표 3>에 제시되어 있다.

<표 3>의 학생 요인을 보면 각 채점 영역 간 공분산이 다른 요인에서 보다 크게 나타났음을 알 수 있다. 이러한 결과는 <표 2>에서 보고한 분산성분 분석결과와 비슷하며, 학생 요인에서 채점 영역 간 측정오차를 고려한 상관계수는 0.825부터 0.988로 높게 나타났다. 이는 한 채점 영역에서 높은 평가 점수를 받은 학생은 다른 영역에서도 높은 점수를 받았음을 의미하므로 각 채점 영역이 영재성이라는 구인의 각기 다른 측면을 측정하고 있다는 것을 나타낸다. 따라서 이 점수들의 평균 점수를 학생들의 영재성 점수로 해석해도 무방하다는 점을 알려주는데 이 결과는 교사추천서의 타당도가 높다는 증거를 나타낸다. 학생과 채점자의 상호작용효과에서 인지적 영역과 사회적 영역의 공분산은 0.07로 나타났는데 이는 인지적 영역에서의 학생들의 순위 차이가 사회적 영역에서의 차이와 관련이 없다는 사실을 보여준다. 전체적으로 볼 때 학생 효과에 비해 다른 효과에서의 공분산은 0에 가까울 정도로 낮게 나타나고 있으므로 이는 각종 오차들 사이에 별다른 관계가 존재하지 않는다는 뜻으로(이영식, 신상근, 2004) 교사추천서의 신뢰도가 높다고 해석할 수 있다.

<표 3> 교사추천서 평가에서의 공분산 행렬

효과		인지적 영역	수학적 영역	정의적 영역	사회적 영역
<i>p</i> (학생)	인지적 영역	0.193	(0.825)	(0.948)	(0.818)
	수학적 영역	0.192	0.279	(0.988)	(0.860)
	정의적 영역	0.212	0.265	0.257	(0.845)
	사회적 영역	0.162	0.660	0.759	0.203
<i>r</i> (채점자)	인지적 영역	0.105			
	수학적 영역	0.072	0.044		
	정의적 영역	0.021	0.014	0.003	
	사회적 영역	0.081	0.054	0.016	0.054
<i>pr</i> (학생*채점자)	인지적 영역	0.102			
	수학적 영역	0.066	0.137		
	정의적 영역	0.065	0.059	0.125	
	사회적 영역	0.007	0.006	0.022	0.112

주. ( )는 측정오차를 고려한 상관계수를 나타냄.

자기소개서 평가에서  $p \times r \times i$  설계의 각 효과에 해당하는 분산 성분 추정치와 전체 분산 중 해당 효과가 차지하는 비율은 <표 4>와 같다. 수학적 영역과 정의적 영역에서 학생효과가 29.3%와 37.1%로 가장 크게 나타났으며, 인지적 영역과 사회적 영역에서도 잔차효과를 제외하면 32.3%와 22.4%로 가장 크게 나타났다. 이는 자기소개서 평가 결과에 학생의 영재성 차이가 반영되어 있다는 것을 의미하므로, 자기소개서가 변별력을 지닌 측정절차라고 해석할 수 있다. 다음으로 영역에 상관없이 9.1%에서 13.0%로 나타난 학생과 문항의

상호작용 효과보다 학생과 채점자의 상호작용 효과가 10.1%에서 25.1%로 크게 나타났다. 이는 학생들의 상대적 순위가 문항의 난이도보다는 채점자의 엄격성 정도에 따라 더 달라진다고 해석할 수 있다. 정의적 영역에서만 문항 효과가 채점자 효과보다 크게 나타났으며, 나머지 영역에서는 채점자 효과가 문항 효과보다 크게 나타났다.

자기소개서의 공분산 결과는 <표 5>에 제시되어 있다. <표 5>의 학생 요인을 보면, 각 채점 영역 간 공분산 성분이 다른 요인에서 보다 크게 나타났으며, 채점 영역 간 측정오차를 고려한 상관계수는 0.672부터 0.953으로 높게 나타났다. 또한 전체적으로 볼 때, 학생 효과에 비해 다른 효과에서의 공분산은 0에 가까울 정도로 낮게 나타나고 있으므로 이는 각종 오차들 사이에 별다른 관계가 존재하지 않는다는 뜻으로(이영식, 신상근, 2004) 자기소개서의 신뢰도가 높다고 해석할 수 있다.

<표 4> 자기소개서 평가에서의 분산성분과 비율

효과	다변량 일반화가능도 이론			
	인지적 영역	수학적 영역	정의적 영역	사회적 영역
$p$ (학생)	0.241(0.323)	0.228(0.293)	0.247(0.371)	0.138(0.224)
$r$ (채점자)	0.014(0.019)	0.048(0.062)	0.000(0.000)	0.037(0.060)
$i$ (문항)	0.000(0.000)	0.000(0.000)	0.008(0.012)	0.017(0.028)
$pr$ (학생*채점자)	0.096(0.129)	0.195(0.251)	0.099(0.149)	0.062(0.101)
$pi$ (학생*문항)	0.068(0.091)	0.101(0.130)	0.068(0.102)	0.058(0.094)
$ri$ (채점자*문항)	0.058(0.078)	0.013(0.017)	0.016(0.024)	0.025(0.041)
$pri$ (학생*채점자*문항, 잔차)	0.269(0.361)	0.192(0.247)	0.228(0.342)	0.279(0.453)

주. ( )는 전체 분산 중 해당 효과의 분산이 차지하는 비율임.

<표 5> 자기소개서 평가에서의 공분산 행렬

효과		인지적 영역	수학적 영역	정의적 영역	사회적 영역
$p$ (학생)	인지적 영역	0.241	(0.672)	(0.952)	(0.700)
	수학적 영역	0.158	0.228	(0.953)	(0.665)
	정의적 영역	0.232	0.226	0.247	(0.759)
	사회적 영역	0.128	0.118	0.140	0.138
$r$ (채점자)	인지적 영역	0.014			
	수학적 영역	0.039	0.048		
	정의적 영역	0.000	0.000	0.000	
	사회적 영역	0.036	0.049	0.000	0.037
$pr$ (학생*채점자)	인지적 영역	0.096			
	수학적 영역	0.091	0.195		
	정의적 영역	0.046	0.056	0.099	
	사회적 영역	0.016	0.036	0.046	0.062

주. ( )는 측정오차를 고려한 상관계수를 나타냄.

## 2. 교사추천서 평가와 자기소개서 평가에서의 D-연구 결과

### 가. 합성점수의 타당성

일반적으로 관찰·추천 영재 선발 시 사용되는 선발도구의 경우 각 채점 영역 점수를 합한 뒤 영역 수로 나눈 평균점수를 학생의 평가점수로 보고하는 경우가 많다. 이렇게 산출된 합성점수는 각 채점 영역에서 학생의 영재성이 같은 비율로 반영되어 있음을 전제로 하는데 다변량 일반화가능도 분석은 이 전제가 타당한지에 대한 정보를 제공한다. 교사추천서와 자기소개서 평가는 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역으로 모두 4문항으로 구성되어 있으므로, 각각 25%를 반영한다고 볼 수 있다. 각 학생의 점수 차이에 각 채점 영역 점수가 같은 비율로 반영되어 있는지를 분석한 결과는 <표 6>에 제시되어 있다. 교사추천서의 경우는 평가점수에 수학적 영역과 정의적 영역이 상대적으로 더 많이 반영되었으며, 자기소개서의 경우는 정의적 영역의 점수 차이가 제일 많이 반영되어 있고 사회적 영역에 대한 정보는 상대적으로 덜 반영되는 것으로 나타났다.

<표 6> 각 채점 영역 점수가 합성점수의 분산에 미친 영향

		인지적 영역	수학적 영역	정의적 영역	사회적 영역
합성점수 분산에 기여한 비율	교사추천서	22.39%	27.74%	27.36%	22.50%
	자기소개서	26.52%	25.56%	29.60%	18.32%

### 나. 합성점수를 바탕으로 최적의 신뢰도계수를 얻을 효율적인 측정 조건 구성

교사추천서와 자기소개서 평가에서의 효율적인 측정 조건을 탐색하기 위해 G-연구 분석 결과 산출된 분산 성분과 공분산 성분을 바탕으로 합성점수의 각 채점 영역별 문항의 표본 크기를 조절하여 D-연구를 실시하였다. 교사추천서와 자기소개서 모두 각 채점 영역별로 4문항으로 구성되어 전체는 16문항이다. 따라서 채점자는 2명으로, 그리고 총 문항은 16개로 고정시킨 상태에서 채점 영역별 문항의 수를 조절하여 일반화가능도계수와 의존도계수를 구하였고, 그 결과는 각각 <표 7>, <표 8>과 같다.

<표 7>의 순위는 교사추천서 평가에서 합성점수 일반화가능도계수를 기준으로 정하였다. 일반적으로 신뢰도의 크기가 얼마 이상이 되어야 한다는 절대적인 기준은 제시되어 있지 않고, 그 기준 또한 연구자들마다 다르다. 그러나 Shavelson, Baxter & Gao(1993)와 Dunbar, Koretz & Hoover(1992)는 안정된 신뢰도는 0.80 이상을 주장하며, Brennan(2001a)을 비롯하여 일반화가능도 분석을 적용한 많은 연구들은 일반적으로 일반화가능도계수는 0.80이상을, 그리고 의존도계수는 0.70 이상인 경우를 적정수준의 신뢰도로 판단하고 있다. 이를 기준으로 보면 원자료의 교사추천서인 경우 일반화가능도계수는 0.841, 그리고 의존도계수는 0.765로 나타나서 신뢰도는 양호하다고 판단할 수 있다. 그러나 같은 16개의 문항이라 할지라도 각 채점 영역에 따라 어떻게 배분되는지에 따라 그 결과는 달라진다. 총 454개의 조합 결과 원자료의 교사추천서의 경우 55위로 나타났으며, 인지적 영역에 1개의 문항, 수학적 영역에

<표 7> 교사추천서 평가에서 채점 영역별 문항 수에 따른 일반화가능도계수와 의존도계수

순위	인지적 영역	수학적 영역	정의적 영역	사회적 영역	합성점수 전집점수 분산	합성점수 상대오차 분산	합성점수 절대오차 분산	합성점수 일반화 가능도계수	합성점수 의존도 계수
1	1	5	4	6	0.220	0.039	0.061	0.848	0.783
2	1	5	5	5	0.224	0.040	0.059	0.848	0.791
3	1	4	5	6	0.219	0.040	0.059	0.847	0.787
4	2	5	4	5	0.218	0.040	0.062	0.847	0.779
5	2	5	3	6	0.214	0.039	0.064	0.846	0.770
6	2	4	4	6	0.213	0.039	0.062	0.846	0.775
7	1	4	6	5	0.223	0.041	0.058	0.846	0.794
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
55	4	4	4	4	0.212	0.040	0.065	0.841	0.765
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
448	11	2	2	1	0.198	0.051	0.091	0.796	0.685
449	11	3	1	1	0.197	0.051	0.094	0.793	0.676
450	12	1	1	2	0.190	0.050	0.096	0.793	0.666
451	12	1	2	1	0.196	0.052	0.095	0.789	0.675
452	1	1	1	13	0.199	0.053	0.084	0.789	0.703
453	12	2	1	1	0.195	0.052	0.098	0.788	0.666
454	13	1	1	1	0.193	0.054	0.101	0.781	0.656

주. 음영부분은 G-연구에서의 표본크기인 원자료의 교사추천서를 나타냄.

<표 8> 자기소개서 평가에서 채점 영역별 문항 수에 따른 일반화가능도계수와 의존도계수

순위	인지적 영역	수학적 영역	정의적 영역	사회적 영역	합성점수 전집점수 분산	합성점수 상대오차 분산	합성점수 절대오차 분산	합성점수 일반화 가능도계수	합성점수 의존도 계수
1	6	1	8	1	0.221	0.046	0.047	0.828	0.825
2	7	1	7	1	0.220	0.046	0.048	0.827	0.822
3	5	1	9	1	0.223	0.047	0.047	0.827	0.826
4	6	1	7	2	0.209	0.044	0.046	0.826	0.819
5	8	1	6	1	0.219	0.046	0.049	0.826	0.818
6	7	1	6	2	0.208	0.044	0.047	0.826	0.816
7	5	1	8	2	0.211	0.044	0.046	0.826	0.821
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
213	4	4	4	4	0.179	0.045	0.054	0.800	0.769
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
448	3	11	1	1	0.198	0.077	0.096	0.720	0.673
449	1	6	1	2	0.182	0.072	0.090	0.718	0.669
450	1	9	1	2	0.191	0.076	0.095	0.716	0.667
451	1	12	1	2	0.197	0.079	0.099	0.714	0.665
452	1	9	1	1	0.203	0.084	0.103	0.706	0.662
453	1	6	1	1	0.196	0.082	0.099	0.706	0.664
454	1	13	1	1	0.208	0.087	0.107	0.706	0.660

주. 음영부분은 G-연구에서의 표본크기인 원자료의 자기소개서를 나타냄.

5개의 문항, 정의적 영역에 4개 그리고 사회적 영역에 6개의 문항을 배분하면 일반화가능도 계수는 0.848로, 그리고 의존도계수는 0.783으로 높아지게 된다. 반면에 인지적 영역에 13개의 문항, 수학적 영역에 1개의 문항, 정의적 영역에 1개 그리고 사회적 영역에 1개의 문항을 배분하면 일반화가능도계수는 0.781로, 그리고 의존도계수는 0.656까지 낮아지게 된다. 이는 <표 2>과 <표 3>에서 제시된 것처럼 인지적 영역은 문항의 난이도보다 채점자의 엄격성이, 사회적 영역은 상대적으로 문항의 난이도가 학생의 평가점수에 영향을 주는 것이 반영되었다고 해석 할 수 있다.

<표 8>의 순위는 자기소개서 평가에서 합성점수 일반화가능도계수를 기준으로 정하였다. 일반화가능도계수는 0.80 이상을, 그리고 의존도계수는 0.70 이상인 경우를 적정수준의 신뢰도로 판단하는 기준으로 보면 원자료의 자기소개서인 경우 일반화가능도계수는 0.800, 그리고 의존도계수는 0.769로 나타나서 신뢰도는 양호하다고 판단할 수 있다. 그러나 같은 16개의 문항이라 할지라도 각 채점 영역에 따라 어떻게 배분되는지에 따라 그 결과는 달라진다. 총 454개의 조합 결과 원자료의 자기소개서의 경우 213위로 나타났으며, 인지적 영역에 6개의 문항, 수학적 영역에 1개의 문항, 정의적 영역에 8개의 문항, 그리고 사회적 영역에 1개의 문항을 배분하면 일반화가능도계수는 0.829로 의존도계수는 0.825로 올라가게 된다. 반면에 인지적 영역에 1개의 문항, 수학적 영역에 13개의 문항, 정의적 영역에 1개의 문항, 그리고 사회적 영역에 1개의 문항을 배분하면 일반화가능도계수는 0.706으로 의존도계수는 0.660까지 낮아지게 된다. 이는 <표 4>과 <표 5>에서 제시된 것처럼 정의적 영역은 상대적으로 문항의 난이도가, 수학적 영역의 경우 채점자의 엄격성이 학생의 평가점수에 영향을 주는 것이 반영되었다고 해석 할 수 있다.

#### 다. 합성점수를 바탕으로 현 수준의 신뢰도계수를 얻을 효율적인 측정 조건 구성

교사추천서와 자기소개서 평가에서의 효율적인 측정 조건을 탐색하기 위해 G-연구 분석 결과 나온 분산 성분과 공분산 성분을 바탕으로 산출된 합성점수의 각 채점 영역별 문항의 표본 크기를 조절하여 D-연구를 실시하였다. 교사추천서와 자기소개서 모두 각 채점 영역별로 4문항으로 구성되어 전체는 16문항으로 구성되어 있으며 일반화가능도계수와 의존도계수는 0.841, 0.765로 나타났다. 현 수준의 일반화가능도 계수를 얻을 수 있도록 채점자는 2명으로 고정하고, 교사추천서와 자기소개서의 각 채점 영역별 문항 수를 줄여나가면서 D-연구를 수행한 결과는 <표 9>과 <표 10>에 제시되어 있다.

<표 9>의 순위는 교사추천서 평가에서 합성점수 일반화가능도계수를 기준으로 정하였다. 채점 영역 별 문항 수를 줄여나감으로써 전체 평가 문항수를 줄여도 원자료의 교사추천서에서와 동일한 신뢰도계수를 얻을 수 있는지 살펴보았다. 원자료 교사추천서에서의 일반화가능도계수 이상을 얻을 수 있는 경우는 전체 384개의 분석 결과 중 인지적 영역에서는 3문항, 수학적 영역, 정의적 영역 그리고 사회적 영역에서는 각각 4문항으로, 원자료보다 한 개의 문항을 줄인 경우만이 해당되었다. 전통적인 Cronbach alpha의 경우 문항수가 늘어나면 신뢰도는 증가하게 되지만 일반화가능도 이론의 경우 다양한 오차국면을 고려함으로써 문항수를

줄여도 안정된 신뢰도를 얻을 수 있음을 재확인할 수 있다.

<표 9> 교사추천서 평가에서 채점 영역별 문항 수 감소에 따른 신뢰도계수

순위	인지적 영역	수학적 영역	정의적 영역	사회적 영역	합계	합성점수 전집점수 분산	합성점수 상대오차 분산	합성점수 절대오차 분산	합성점수 일반화 가능도계수	합성점수 의존도 계수
1	3	4	4	4	15	0.215	0.041	0.064	0.841	0.770
2	4	4	4	4	16	0.212	0.040	0.065	0.841	0.765

주. 음영부분은 G-연구에서의 표본크기인 원자료의 교사추천서를 나타냄.

<표 10>의 순위는 자기소개서 평가에서 합성점수 일반화가능도계수를 기준으로 정하였다. 채점 영역 별 문항 수를 줄여나감으로써 전체 평가 문항수를 줄여도 원자료의 자기소개서에서와 동일한 신뢰도계수를 얻을 수 있는지 살펴보았다. 원자료 자기소개서에서의 일반화가능도계수 이상을 얻을 수 있는 경우는 전체 384개의 분석 결과 중 총 12개의 경우가 해당되었다. 인지적 영역, 정의적 영역, 그리고 사회적 영역에 4문항 그리고 수학적 영역에 1문항으로 총 13개의 문항으로 자기소개서를 평가할 때 일반화가능도계수는 0.809, 그리고 의존도계수는 0.785로 나타났다. 또한 원자료의 16개 문항에서 6문항을 줄여 인지적 영역과 정의적 영역에서는 4문항, 수학적 영역과 사회적 영역에서는 1문항으로 10문항으로만 구성하여도 일반화가능도계수는 0.801, 그리고 의존도계수는 0.775로 원자료보다 높은 신뢰도 계수를 얻는 것으로 나타났다.

<표 10> 자기소개서 평가에서 채점 영역별 문항 수 감소에 따른 신뢰도계수

순위	인지적 영역	수학적 영역	정의적 영역	사회적 영역	합계	합성점수 전집점수 분산	합성점수 상대오차 분산	합성점수 절대오차 분산	합성점수 일반화 가능도계수	합성점수 의존도 계수
1	4	1	4	4	13	0.179	0.042	0.049	0.809	0.785
2	4	2	4	4	14	0.178	0.043	0.050	0.807	0.780
3	4	1	4	3	12	0.187	0.045	0.051	0.807	0.787
4	4	2	4	3	13	0.186	0.045	0.052	0.804	0.781
5	4	1	4	2	11	0.197	0.048	0.053	0.804	0.788
6	4	3	4	4	15	0.178	0.044	0.052	0.804	0.774
7	4	1	3	4	12	0.174	0.043	0.052	0.802	0.771
8	3	1	4	4	12	0.176	0.044	0.050	0.802	0.779
9	4	2	4	2	12	0.195	0.048	0.054	0.801	0.782
10	4	3	4	3	14	0.185	0.046	0.054	0.801	0.775
11	4	1	4	1	10	0.211	0.052	0.056	0.801	0.790
12	3	2	4	4	13	0.176	0.044	0.051	0.801	0.775
13	4	4	4	4	16	0.179	0.045	0.054	0.800	0.769

주. 음영부분은 G-연구에서의 표본크기인 원자료의 자기소개서를 나타냄.

## VI. 결론 및 논의

이 연구는 관찰·추천제에 의한 수학영재 선발 시 사용되는 자기소개서와 교사추천서를 활용하여 각 측정상황에서 발생하는 오차요인들의 상대적인 영향력을 살펴보고, 이를 바탕으로 효율적인 측정 조건을 구성하는 방법을 제시함으로써, 선발도구 자체를 개발하고 평가해야 하는 단위학교에 설치된 영재교육원 및 영재학급 등의 각 영재교육기관에서 바람직한 의사결정이 이루어질 수 있는 토대를 제공하고자 하였다. 이를 위하여, 김성찬 외(2012)에서 사용한 자료 중, 분석적 방법으로 작성된 루브릭을 이용하여 평가점수를 부여한 자료에 다변량 일반화가능도 분석을 수행하였다. G-연구 분석 결과, 교사추천서와 자기소개서 평가에서 일부 채점 영역에서의 잔차 효과를 제외하면, 학생 효과가 가장 크게 나타났다. 이는 교사추천서와 자기소개서 평가 결과에 학생의 영재성 차이가 반영되어서 있다는 것을 의미하므로, 각 선발도구는 신뢰도가 높다고 해석할 수 있다. 또한 교사추천서와 자기소개서에서 채점 영역 간 측정오차를 고려한 상관계수를 살펴본 결과 학생 요인에서 가장 높게 나타났다. 이는 한 채점 영역에서 높은 평가 점수를 받은 학생은 다른 영역에서도 높은 점수를 받았음을 의미하므로 각 채점 영역이 영재성이라는 구인의 각기 다른 측면을 측정하는 신뢰로운 선발도구라는 또 다른 증거라고 해석할 수 있다.

D-연구 분석 결과, 교사추천서에서는 수학적 영역과 정의적 영역이 학생의 평가 점수에 더 많이 반영된 것으로 나타났으며, 자기소개서에서는 정의적 영역의 점수 차이가 제일 높게 반영되고, 사회적 영역에 대한 정보는 상대적으로 덜 반영되는 것으로 나타났다. 적정 수준의 일반화가능도계수를 0.80, 그리고 의존도 계수를 0.70으로 했을 때, 원자료의 교사추천서는 각각 0.841, 0.765 그리고 자기소개서는 각각 0.800, 0.769로 모두 신뢰도가 높은 선발도구로 나타났다. 그러나 효율적인 측정 조건 구성을 탐색하고자 원자료와 동일한 16개의 문항으로 고정시키고, 채점 영역 내 문항 수를 조절해가면서 일반화가능도계수를 살펴본 결과 원자료 교사추천서의 일반화가능도계수는 55위, 그리고 원자료 자기소개서의 일반화가능도계수는 213위로 나타났다. 가장 높은 일반화가능도계수를 산출할 조건은 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역에 각각 교사추천서의 경우는 1개, 5개, 4개, 6개, 그리고 자기소개서의 경우는 6개, 1개, 8개, 1개로 나타났다. 또한 원자료의 일반화가능도계수를 산출할 효율적인 측정 조건을 탐색한 결과, 교사추천서에서는 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역에서 각각 3개, 4개, 4개, 4개로 총 15문항을 적용한 단 한 가지 경우만 가능하였다. 그러나 자기소개서에서는 총 12경우가 가능하였으며 인지적 영역, 수학적 영역, 정의적 영역 그리고 사회적 영역에서 각각 4개, 1개, 4개, 4개로 총 13문항을 적용한 경우에 가장 높은 일반화가능도 계수를 얻을 수 있었고, 각각 4개, 1개, 4개, 1개로 최소 10문항을 적용한 경우에도 기존의 일반화가능도계수보다 높은 값을 얻을 수 있었다.

이상의 예시적 분석 결과를 바탕으로 시사점을 논하면 다음과 같다. 첫째, 교사추천서와 자기소개서는 관찰·추천에 의한 수학영재 선발 시에 활용할 수 있는 신뢰로운 선발도구로

나타났다. 이는 관찰·추천에 의한 영재 선발 시 가장 중요한 평가주체인 교사의 추천서와 학생자신이 자신을 추천하는 형태의 자기소개서가 측정학적 시각에서의 신뢰도를 바탕으로 공정하고 객관적인 자료로 활용될 수 있다는 것을 의미한다.

둘째, 교사추천서와 자기소개서의 루브릭은 구별되어야 한다. 선행연구들을 살펴본 결과, 같은 루브릭을 사용하고 있는 경우가 적지 않았지만, 교사추천서에서의 평가 점수와 자기소개서의 평가 점수에 영재성이 반영되는 정도는 각 채점 영역 점수가 합성점수의 분산에 미친 영향을 분석한 결과 다름을 확인할 수 있었다. 따라서 이 연구결과와 함께 질적 연구를 함께 수행하여 학생의 영재성을 측정하는 데 있어 교사추천서와 자기소개서 루브릭이 상보적인 역할을 할 수 있도록 구성되어야 한다.

셋째, 다변량 일반화가능도 이론은 시·도교육청 영재교육원, 영재학급 그리고 대학의 영재교육원 등 각 현장에서 직접 제작한 교사추천서와 자기소개서에 적용하여 최적의 채점구조를 탐색하는 방법론적인 틀로 적용될 수 있다. 먼저 평가 점수에 영향을 미칠 수 있는 요인들을 미리 밝혀내어 측정하고자 하는 영재성 이외의 다른 요인들의 영향을 배제하려는 노력이 이루어져야 할 것이다. 예를 들어 원자료의 교사추천서와 자기소개서의 경우 채점자 혼란은 제대로 되었는지, 문항은 일반화될 수 있는 지, 특정 문항이 특정 학생에게 유리하게 작용할 가능성은 없는지, 채점 영역은 적정한지 등을 하나하나 점검해 본다면 평가의 유용성을 높이는 데 도움이 될 것이다. 이를 바탕으로 각 오차요인들의 표본 수를 다양하게 변화시켜 가면서 선발도구의 신뢰도를 탐색함으로써, 바람직한 의사 결정이 이루어질 수 있는 최적의 측정 조건을 확정하는 데 유용하게 사용될 수 있을 것이다.

이 연구의 제한점과 후속연구를 제시하면 다음과 같다. 첫째, 이 연구에서 산출된 최적의 측정 조건은 특정 영재교육원에서 제작한 루브릭을 바탕으로 수학영재를 선발하기 위해 사용한 교사추천서와 자기소개서에만 국한된다. 따라서 다양한 과목과 다문화 학생 등 영재교육 소외자를 선발 시 사용하는 교사추천서와 자기소개서뿐만 아니라, 영재특성 체크리스트, 학생생활기록부, 포트폴리오, 그리고 심층면접에 이르기까지 다양한 자료를 활용한 연구가 수행될 필요가 있다. 둘째, 이 연구는 4개의 채점 영역별로 모두 같은 4개의 문항이 포함되고, 이를 같은 5점 척도로 채점한 자료를 사용하였다. 김성찬 외(2013)의 연구에서는 GENOVA(A GENERALIZED Analysis Of Variance System) 프로그램(Crick & Brennan, 1983)를 활용하여 일변량 일반화가능도 분석을 적용하였기 때문에 이러한 제한점을 수반할 수밖에 없었다. 그러나 다변량 일반화가능도 분석은 혼합모형이면서, 각 채점 영역별로 다른 문항 수가 포함되거나, 각각이 다른 척도로 채점되는 불균형 설계에도 쉽게 적용할 수 있다. 따라서 현장에서의 적용가능성을 모색한다면 선행연구를 바탕으로 나름대로 가중치가 주어져 있는 불균형 자료를 활용한 연구가 수행될 필요가 있다. 셋째, 이 연구에서는 각 채점 영역 점수가 합성점수의 분산에 미친 영향을 분석함으로써 평가 도구를 제작한 후 최종점수로 보통 각 채점 영역 점수를 더하거나 평균을 내는 경우가 많은데, 실제로는 각 채점 영역의 분산이나 각 채점 영역 간의 공분산에 따라 반영되는 비율이 달라짐으로써 현실에서 당연하다고 생각하며 사용하고 있는 점수에 문제가 있을 수 있다는 것에 초점을 두었다. 따라서 이러한

합성점수의 분산에 미친 비율을 고려하여 최적의 점수 산출 방식을 확정하기 위한 상대가중치에 대한 연구가 필요하다. 넷째, 이 연구에서 사용한 일반화가능도 분석은 적정수준의 일반화가능도계수를 얻기 위하여 오차 요인의 조건을 조정·조합함으로써 최적의 측정 조건을 제시하여 전체 검사에 대한 의사결정의 준거를 제공한다(Keeves, 1988). 그러나 이처럼 적정수준의 일반화가능도계수가 산출되었다고 해서 조정된 요인의 내용과 질까지 보장되는 것은 아니다. 따라서 이 연구결과와 함께 개별 문항이나 적용된 모형에 따른 질적인 검토가 필요하다.

## 참 고 문 헌

- 교육부 (2010). **올해부터 시험 없이 과학영재 발굴**. 교육과학기술부.
- 교육부 (2013). **“영재교육 최적화를 통한 창조적 인재육성”을 위한 제 3차 영재교육진흥종합계획**. 교육부(창의인재정책관).
- 김미숙 (2007). **한국의 영재 선발 현황과 개선 방향**. 영재선발의 세계적인 동향과 개선 방향을 위한 국제 심포지엄. 연구자료 RM 2007-45. 교육인적자원부.
- 김미숙, 이정규, 이희권, 김언주, 맹희주, 이상천, 정경아, 최호진, 한수연 (2007). **제1차 영재교육진흥종합계획 평가 및 중장기 전망에 관한 연구**. 한국교육개발원.
- 김성숙, 김양분 (2001). **일반화가능도이론**. 파주: 교육과학사.
- 김성숙, 송미영, 박인용 (2012). 다변량 일반화가능도 이론을 적용한 성취수준 설정에서의 오차분석과 최적 조건탐색, **교육평가연구**, 25(4), 679-700.
- 김성찬, 김성연, 한기순 (2012). 관찰, 추천에 의한 수학영재 선발 시 사용되는 자기소개서와 교사추천서 평가에 대한 일반화가능도 이론의 활용. **수학교육논문집**, 26(3), 251-271.
- 김숙영 (2012). **관찰·추천제를 통한 수학 영재의 선발 과정에 관한 연구**. 석사학위논문. 숙명여자대학교.
- 남윤모 (2012). **GED 기반 수학영재 관찰·추천 선발 방식의 타당성 분석**. 석사학위논문. 대구교육대학교.
- 류지영, 정현철 (2010). 영재학급 대상자 선발을 위한 관찰·추천 영재판별모형 개발 연구, **영재교육연구**, 20(1), 257-287.
- 미래창조과학부 (2013). **창조경제를 견인할 창의인재 육성 방안**. 미래창조과학부.
- 박혜정 (2010). **관찰·추천 영재 선발 시 사용되는 자기소개서와 교사추천서의 평가자간 신뢰도분석**. 석사학위논문. 인천대학교.
- 성태제 (2010). **현대교육평가(3판)**. 서울: 학지사.
- 신윤주 (2011). **관찰 추천에서 나타난 과학영재행동특성의 경향: 교사추천서와 학생 자기소개서 분석을 중심으로**. 박사학위논문. 서울대학교.
- 우미란, 김선자, 박중욱 (2012). 관찰·추천 전형으로 선발된 학생들의 교사추천서와 프로

- 그림 수행의 관련성 분석. **영재교육연구**, 22(1), 173-196.
- 유미현, 강윤희, 예홍진 (2011). 과학영재교육원 관찰·추천 방식의 영재선발 분석. **과학영재교육**, 3(2), 27-38.
- 윤은정, 박윤배 (2012). 관찰·추천제에 의한 영재교육대상자 선발에서 심사결과의 신뢰도 및 타당도 분석. **영재교육연구**, 22(4), 929-942.
- 이영식, 신상근 (2004). 다변량 일반화가능도 이론에 의한 말하기 시험의 타당도와 신뢰도에 관한 연구. **Foreign Language Education**, 11(2), 249-265.
- 이인호, 한기순 (2009). 영재교육 대상자 선발에서 교사 추천서의 효용성 분석. **영재교육연구**, 19(2), 381-404.
- 이현숙 (2012). 혼합형 검사의 문항 유형별 가중치에 따른 신뢰도 및 다변량 일반화가능도 분석. **교육평가연구**, 25(1), 95-116.
- 최은주 (2012). **과학영재 관찰, 추천 선발에 사용되는 교사추천서와 자기소개서 분석 및 실제 교육 결과의 관계**. 석사학위논문. 아주대학교.
- Baum, S. M. (1994). Meeting the need of gifted/learning disabled students. *The Journal of Secondary Gifted Education*, 3(5), 6-16.
- Brennan, R. L. (2001a). *Generalizability Theory*. New York: Springer.
- Brennan, R. L. (2001b). *Manual for mGENOVA Version 2.1*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Crick, J. E., & Brennan, R. L. (1983). *Manual for GENOVA: A GENeralized Analysis Of Variance System*. ACT Technical Bulletin, 43, The American College Testing Program.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1992). Quality control in the development and use of performance assessment. *Applied Measurement in Education*, 4, 289-303.
- Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of student achievement, 9<sup>th</sup> ed.* Upper Saddle River, NJ: Pearson.
- Jiang, Y. H., Smith, P., & Nichols, P. (1997). *Error sources influencing performance assessment reliability or generalizability: A meta-analysis*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Keeves, J. P. (1988). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon.
- Pfeiffer, S., & Blei, S. (2008). Gifted identification beyond the IQ test. In Pfeiffer, S. (Ed.), *Handbook of giftedness in children* (pp. 177-196). NY: Springer Science and Business Media, LLC.
- Pfeiffer, S., & Jarosewich, T. (2003). Gifted rating scales. TX: Psych Corp.

- Renzulli, J. S., & Reis, S. M. (1986). The enrichment triad/revolving door model: A school wide plan for the development of creative productivity. In Renzulli, J. S. (Ed.), *Systems and models for developing programs for the gifted and talented*. TX: Prufrock press.
- Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30, 215-235.
- VanTassel-Baska, J., Johnson, D., & Avery, L. D. (2002). Using performance tasks in the identification of economically disadvantaged and minority gifted learners: Finding from project STAR. *Gifted Child Quarterly*, 46(2), 110-123.
- Webb, N. M., Shavelson, R. J., & Maddahian, E. (1983). Multivariate generalizability theory. In L. J. Pyans, Jr. (Ed.), *Generalizability theory: Inferences and practical applications* (pp. 49-66). San Francisco, CA: Jossey-Bass.

= Abstract =

## An Application of Multivariate Generalizability Theory to Teacher Recommendation Letters and Self-introduction Letters Used in Selection of Mathematically Gifted Students by Observation and Nomination

Sung Yeun Kim

*Vanderbilt University*

Ki Soon Han

*Incheon National University*

This study provides an illustrative example of using the multivariate generalizability theory. Specifically, it investigates relative effects of each error source, and finds optimal measurement conditions for the number of items within each content domain that maximizes the reliability-like coefficients, such as a generalizability coefficient and an index of dependability. The method is based on teacher recommendation letters and self-introduction letters, using an analytic scoring method in the context of selection of mathematically gifted students by observation and nomination. This study analyzed data from the 2011 academic year in the science education institute for the gifted, which is attached to the university located in the Seoul metropolitan area. It should be noted that the optimal scoring structures of this study are not generalizable to other selection instruments. However, the methodology applied in this study can be utilized to find optimal measurement conditions for the number of raters, the number of content domains, and the number of items in other selection instruments self-developed by many institutions including: the education institutes for the gifted at provincial offices of education, gifted classes, and the science education institutes for the gifted attached to universities in general. In addition, the methodology will provide bases for making informed decisions in selection instruments of the gifted based on measurement traits.

**Key Words:** Selection of mathematically gifted students, Multivariate generalizability theory, Teacher recommendation letters, Self-Introduction letters

1차 원고접수: 2013년 10월 9일
수정원고접수: 2013년 10월 25일
최종게재결정: 2013년 10월 25일