

# 수능 사탐 및 과탐 동등화 가산점 산정방법 비교

남 보 우\*

## Ⅰ 요약

대입수학능력시험의 사회탐구영역 응시자와 과학탐구영역 응시자가 동일한 모집단위에서 경쟁을 하도록 많은 대학이 교차지원을 허용하는 상황에서 각 탐구영역의 점수를 비교 가능하도록 동등화 가산점을 산정하여 조정할 필요가 있다.

본 연구는 한국교육과정평가원이 2010년에 제공한 수능 점수를 활용하여 전통적인 점수 조정 방법인 연쇄 선형방법, 터커의 합집단 선형방법, 연쇄 동백분위방법, 빈도추정 동백분위방법을 적용하여 가산점을 산정하고, 언어, 수리, 외국어 영역의 종합등급 집단별로 평균 방법을 적용하여 가산점을 산정한 결과와 비교한 것이다.

비교결과 등급별 평균방법으로 가산점을 산정하는 것이 조정의 근거가 명확하고 간편하게 설명되고, 빈도추정 동백분위 방법을 적용한 결과와 비교할 때 평균자승편차가 가장 적어서 바람직한 것으로 나타났다. 따라서 대입전형 현실에서 가산점 산정에 등급별 평균방법의 적용이 가능한 것으로 분석된다.

핵심주제어 : 수능시험, 대입전형, 교차지원, 검사동등화, 가산점, 등급별 평균방법

## I. 서 론

### 1. 연구배경 및 목적

대학수학능력시험(이하 “수능”이라 함)은 대입전형에서 합격여부를 결정하는 중요한 전형 요소이다. 2010학년도에 90% 이상의 대학이 교차지원을 허용하고, 수능의 서로 다른 선택과

\* 단국대학교 경상대학 경영학부 교수, bwnam@dankook.ac.kr

〈논문 투고일〉 2012-01-17

〈논문 수정일〉 2012-02-22

〈게재 확정일〉 2012-03-05

목 점수를 합한 총점을 비교하여 학생을 선발하는 것으로 나타났다(남보우 2010). 이러한 선발방식은 서로 다른 영역이나 선택과목의 수능점수로 수학능력을 비교할 수 있다는 것을 가정한 것이다. 그러나 현행 수능점수는 영역별, 과목별로 표준화한 것이므로 검사 간 난이도 차이는 반영된 것이지만 응시집단 간의 능력 차이가 반영되지 못한 것이므로 직접 비교하는 것은 잘못일 수 있다. 예를 들어 세계사 과목의 표준점수 70점은 수치상으로는 물리 과목의 표준점수 60점보다 높지만, 세계사 응시집단의 수학능력이 물리 응시집단의 수학능력보다 현저하게 낮으면 표준화한 세계사 70점이 물리 60점보다 수학능력이 높다고 할 수 없다. 따라서 대입전형에서 수학능력을 비교할 수 있도록 응시집단의 능력 차이를 가산점 등으로 점수를 조정할 필요가 있다.

검사 간 점수를 조정하는 문제는 교육평가(educational evaluation)분야의 측정이론 중 검사 동등화(test equating)이론에 기초하여 연구되어 왔다. 검사동등화는 공통과목이나 공통항목을 가교검사(anchor test)로 활용하여 점수를 조정하거나, 공통응시자를 가교집단(anchor group)으로 활용하여 점수를 조정하는 것이다. 2002학년도부터 2004학년도까지 수능에서는 각 선택과목 간 공통문항이 있어서 검사동등화 방법으로 점수를 조정하였다. 2005학년도부터 수능 탐구영역의 선택과목에는 가교검사로 활용할 공통과목이나 공통문항이 없고, 모든 과목을 응시한 가교집단으로서의 공통응시자도 없어서 측정이론을 적용하기 어려웠다. 또한 전체 수능 점수가 제공되지 않았으므로 수능 점수조정 분야의 연구는 활발하게 이루어지지 못하였다.

2010년에 수능점수를 연구자에 한하여 공개할 수 있도록 대법원 판결(공2010상, 663) 및 국가정책이 정해졌다. 이에 따라 남보우(2010)는 한국교육과정평가원이 2010년에 제공한 수능 점수를 활용하여 공통영역인 언어, 수리, 외국어 영역(이하 “언수외”라 함)의 종합점수를 구하고 이를 공변인으로 활용하여 측정이론의 점수조정 방법인 평균방법, 선형방법 및 동백분위방법으로 탐구영역 점수를 조정하였다. 그러나 전체 응시자를 대상으로 평균방법을 적용하여 가산점을 산정하였으므로 특정 수학능력 범위의 지원자가 경쟁하는 대입전형에 그 결과를 적용하기 어렵고, 선형방법이나 동백분위방법은 정교하지만 선형식을 구하거나 백분함수를 구해야 하는 등 복잡하였다. 본 연구는 2010 수능자료를 등급별로 집단화한 후 각 집단에 대하여 평균방법을 적용한 결과와 전체 자료에 대하여 선형방법과 동백분위방법을 적용한 결과를 비교하여 대입전형에서 활용할 수 있는 실용적인 점수조정 방법을 제시하고자 한다. 즉, 대학이 교차지원을 허용하는 상황에서 학생선발의 공정성 확보와 수학능력이 높은 학생을 선발할 수 있도록 탐구영역별 응시집단의 능력 차이를 반영하여 가산점을 산정할 수 있는 바람직한 점수조정 방법을 제시하고자 한다.

## 2. 연구방법 및 범위

검사동등화 이론에 근거한 전통적인 점수조정 방법으로는 평균방법(mean method), 선형방법(linear method), 동백분위방법(equipercentile method)이 있다. 이외에도 문항반응이론(item response theory) 등이 있으나 문항별 수능자료가 제공되지 않으므로 전통적인 점수조정 방법을 대상으로 검토하고자 한다. 선형방법과 동백분위방법은 연쇄방법(chained method)으로 적용하는 것이 간편하지만, 합집단방법이 안정적인 결과를 도출하므로(남현우 2004), 합집단 방법도 적용하고자 한다. 합집단에 의한 선형방법은 터커(Tucker)방법을 적용하고, 합집단에 의한 동백분위방법은 빈도추정방법(frequency estimation method)을 적용하고자 한다.

직업탐구영역(이하 “직탐”이라 함) 응시자는 사회탐구영역(이하 “사탐”이라 함)이나 과학탐구영역(이하 “과탐”이라 함) 응시자에 비하여 현저하게 수학능력이 낮은 것으로 보고되었고(남보우 2010), 현실적으로 사탐과 과탐의 교차지원에 대한 가산점을 정하는 것이 중요한 과제이므로 직탐 응시자의 수능점수는 분석대상에서 제외한다. 수학능력별 응시집단의 가산점을 산정하기 위하여 인수의 수능등급을 평균한 종합등급이 1.00~1.99, 2.00~2.99, 3.00~3.99, 4.00~4.99, 5.00~5.99, 6.00~6.99, 7.00~7.99, 8.00~9.00의 8개 집단으로 구분한다. 그리고 각 집단의 표준점수 및 백분위 평균에 대하여 평균방법, 연쇄 선형방법, 터커의 합집단 선형방법, 연쇄 동백분위방법, 빈도추정 합집단 동백분위방법으로 가산점을 산정하고 비교하고자 한다.

평균방법은 검사간의 분포가 평균만 다르다는 강한 가정을 필요로 하고, 선형방법은 평균과 분산만 다르고 선형의 일차식을 전제로 조정하는데 비하여, 동백분위방법은 이러한 가정을 하지 않으며 비교적 정확한 것으로 알려져 있다. 또한 합집단에 의한 방법이 복잡하지만 바람직한 것으로 알려져 있다. 따라서 각 등급별 평균점수에 대하여 조정방법별 가산점을 산정한 후 빈도추정에 의한 합집단 동백분위방법의 조정결과를 기준으로 평균자승편차(Root Mean Square Difference)를 비교하고자 한다.

현행 수능의 탐구영역 간에는 동형의 공통검사도 없고, 공통 응시집단도 없다. 따라서 탐구영역 간 점수를 조정하려면 공변인을 선정하여 집단 간 능력 차이를 반영하는 근거로 활용해야 한다. 조정하려는 선택과목과 공변인의 상관이 높을수록 더 타당한 조정 점수를 얻을 수 있기 때문에 특정 영역만을 공변인으로 하는 것보다는 상관이 높은 또 다른 영역을 함께 고려하는 것이 더 공정한 결과를 얻을 수 있다(남현우 2003). 남보우(2010)는 사탐 점수와 인수의 점수의 상관계수가 0.84, 과탐 점수와 인수의 점수의 상관계수가 0.80으로 높으므로 인수의와 탐구영역 점수는 모두 수학능력이라는 동일 구인을 측정한다고 논의하였다. 또한 인수의

의 분포모양이 단일영역보다 부드러워서 안정적 점수조정을 위한 완곡화 조건(smoothing)을 비교적 만족시키며, 왜도(skewness)를 비교할 때 탐구영역 점수의 분포와 유사하므로 연수의 점수를 공변인으로 활용하는 것이 바람직한 것으로 분석하였다. 따라서 본 연구에서도 연수의 점수를 공변인으로 활용하여 탐구영역 간 가산점을 산정하고 비교하고자 한다.

## II. 선행연구 및 점수조정 방법 검토

### 1. 선행연구 검토

사람의 점수와 과탐의 점수를 비교할 수 있도록 가산점을 산정하는 것과 같이 서로 다른 두 검사의 결과를 비교하기 위한 점수조정 방법인 동등화(equating), 연계화(linking), 척도화(scaling)에 대하여 많은 문헌이 소개하고 있다(김현철 2003: 남현우 2001, 2003, 2004: 박정 2005: 부재울 1999: 성태제 2000: 최성열·이규민·박아청 2007: Kolen and Brennan 2004). 우리나라에서 검사점수의 조정 또는 동등화에 대한 논의와 연구는 1998년도부터 수능에 표준점수를 도입하면서 상당한 진전을 보였다(박정 2005).

성태제(2000)는 연계화와 동등화에 대한 개념을 소개하고, 공통검사와 학생집단 특성을 결합한 새로운 선형변환식으로 표준화하는 통계적 조절방법을 제시하였다. 김현철(2001)은 동등화 방식에 따라 동등화점수에 상당한 차이가 있음을 밝히고 빈도추정에 의한 동등화의 장점을 제시하였다. 또한 김현철(2003)은 모의자료를 생성하여 연쇄방법과 합집단 방법에 의하여 선형방법과 동백분위방법을 비교한 결과, 선형방법이 동백분위방법보다 전 영역의 실험조건에서 일관성 있는 결과를 산출함을 보였다. 그러나 2005학년도 이후에는 가교검사로 활용할 공통과목이나 공통문항이 없어서 이전의 연구결과를 점수조정에 활용하기 어려웠다.

2005학년도 수능의 점수보고 방식을 검토하기 위하여 다양한 논의와 연구가 진행되었고(남명호 등 2002), 탐구영역 간은 물론 선택과목 간에도 점수를 조정하지 않도록 국가의 수능 점수 보고체제가 결정되었다. 그 이유는 첫째, 선택과목은 독립적인 특성을 가지며, 둘째, 각 선택과목은 각 영역의 공통문항이 없어서 직접 비교하기 어려운 검사체제를 구성하고 있고, 셋째, 각 영역 점수의 상호상관을 과거 자료로 구해보면 동등화시키기에는 상관계수가 너무 낮으며, 넷째, 조정에 대하여는 각 대학이 선택할 문제이며, 조정이 필요한 경우 선발상황을 고려하여 대학 나름대로 조정하는 것이 바람직하다는 것이었다. 그러나 각 대학이 표준점수를 그대로 활용하면 엄청난 점수 왜곡 현상이 생기고 대학별로 표준점수를 다시 가공할

경우 그 방법이 천차만별이어서 혼란이 불가피 할 전망이다(남현우 2004), 최선의 조정방법을 개발하는 것이 필요했다.

남현우(2003)는 외국어 점수를 공변인으로 활용하여 합집단 방법으로 2005학년도 수능 선택과목 점수조정을 할 수 있음을 밝혔다. 또한 박정(2005)은 공통집단을 활용한 동시추정방법으로 능력모수 추정치를 산출하여 사용하면 문항반응이론으로 선택과목 점수조정이 가능함을 보여 주었으나, 문항별 자료가 있어야 분석이 가능하고, 원점수가 동일하여도 산출되는 점수의 순서가 달라질 수 있어 대입전형에서 받아들이기 어려운 면이 있었다. 남보우(2002)는 각 선택과목을 동시에 고려하여 점수왜곡의 불공정성을 최소화할 수 있도록 선형계획모형(linear programming)을 개발하여 평균동등화를 시도하였으나, 수능 점수보고체제에 적용하기는 어려웠다. 남보우(2011)는 이 연구결과를 보완하여 선형계획모형 방법으로 탐구영역 내 선택과목 점수를 조정하는 방법을 제시하였다. 한편 황형태(2005)는 각 과목별 원점수 평균을 활용하여 가산점수를 간단하게 더해주는 가산점수제(additive scoring system)를 제시하였지만, 적절한 보정상수 도입 및 원점수 환산 등 적용하는데 어려움이 있었다. 수능의 점수조정 이외에도 최성열·이규민·박아청(2007)은 성인용 직업적성검사 점수를 동백분위방법, 선형방법, 연계화 평균방법을 적용하여 산출하고 각 방법의 적절성을 평가하였다.

선행연구는 대부분 수능의 선택과목 간 점수조정 방법에 관한 것이었다. 대입전형에서 선택과목 간 점수조정도 중요하지만, 탐구영역 간 점수조정이 우선되어야 한다. 즉, 서로 다른 탐구영역 응시자가 대입전형에 지원하도록 교차지원을 허용하는 현실에서 탐구영역 간 점수를 서로 비교할 수 있도록 가산점의 크기를 산정하는 점수조정이 필요한 실정이다. 따라서 본 연구는 수능 종합등급별 8개 집단에 대하여 각각 평균방법을 적용하여 가산점을 산정하고 선행연구(남보우 2010)의 결과와 비교하고자 하는 것이다.

남보우(2010)의 연구는 전통적인 점수조정 방법인 평균방법, 선형방법, 동백분위방법을 적용하여 가산점을 산출하였고, 본 연구는 종합등급 집단별로 평균방법을 기존의 연구결과와 비교하여 집단별 평균방법의 적용 가능성을 입증한 것이다. 또한 남보우(2011)의 연구는 전통적인 방법으로 해결하지 못한 다수 선택과목의 점수조정 문제를 선형계획모형을 활용하여 해결하는 방법을 제시한 것이지만 본 연구는 전통적인 점수조정 방법을 종합등급 집단별로 단순하고 간편한 평균방법을 적용할 수 있는지 분석한 것이다.

## 2. 검사점수의 조정방법

유형간 비교 가능성을 찾는 경우는 검사 동등화이고, 다른 검사 간 비교가능성을 찾는 경

우는 연계화 또는 척도화이며, 이들은 개념상 다르기는 하지만 통계적인 절차는 다르지 않다 (남현우 2004). 검사 간 비교가 가능하도록 점수를 조정할 수 있는 통계적 절차는 여러 가지가 있지만 본 연구에서 적용하고자 하는 평균방법(mean method), 연쇄 선형방법(chained or two-stage linear method), 터커의 합집단 선형방법(Tucker method), 연쇄 동백분위방법(chained or two-stage equipercentile method), 빈도추정 합집단 동백분위방법(frequency estimation equipercentile method)을 간략하게 요약하고자 한다. Kolen and Brennan(2004)은 각 방법에 대한 내용을 자세히 설명하고 있다.

### (1) 평균방법

두 검사의 분포가 같고 평균만 다를 때는 두 검사의 평균차이로 점수를 조정하는 평균방법을 활용한다. 이 방법은 대학들이 가산점수제 방식으로 도입하기가 간편하다. 검사 X의 평균을  $\mu(X)$ , 검사 Y의 평균을  $\mu(Y)$ 라고 하면 검사 X의 특정점수  $x$ 는 검사 Y의 점수척도로 표시한  $y$  값으로 식(1)에 의하여 조정된다.

$$y = x - \mu(X) + \mu(Y) \tag{1}$$

검사 X를 받은 집단을 1, 검사 Y를 받은 집단을 2라 하고 두 집단이 모두 공통검사 V를 받았다고 하자. 이때는 검사 X의 점수  $x$ 를 공통검사 V의 점수척도  $v$ 로 조정하고, 공통검사 V의 점수  $v$ 를 검사 Y의 점수척도  $y$ 로 2단계에 걸쳐 조정하는 연쇄방법을 활용할 수 있다. 검사 X의 평균을  $\mu_1(X)$ , 검사 Y의 평균을  $\mu_2(Y)$ , 집단1의 공통검사 V의 평균을  $\mu_1(V)$ , 집단2의 공통검사 V의 평균을  $\mu_2(V)$ 라 하자. 식(1)을 두 번 적용하면 검사 X의 특정점수  $x$ 는 공통검사 V의 집단 간 평균의 차이를 반영하여 검사 Y의 점수척도로 표시한  $y$  값으로 식(2)와 같이 조정된다.

$$y = x - \mu_1(X) + [\mu_1(V) - \mu_2(V)] + \mu_2(Y) \tag{2}$$

식(2)는 두 집단의 공통검사 평균차이로 집단의 능력 차이를 보정해 주는 것이다. 예를 들어 사탐과 과탐은 검사가 다르고 응시한 두 집단이 다르므로 직접 비교가 불가능하지만, 언어 등 공통검사가 있으므로 응시집단 간 언어 점수 차이를 더하여 조정하는 것이다.

### (2) 선형방법

두 검사의 분포가 평균과 분산만 다를 경우 선형방법으로 점수를 조정한다. 검사 X의 평

균을  $\mu(X)$ , 표준편차를  $\sigma(X)$ 라 하고 검사 Y의 평균과 표준편차를  $\mu(Y)$ ,  $\sigma(Y)$ 라 하면 검사 X의 특정점수  $x$ 는 식 (3)에 의하여 검사 Y의 척도로 표시된  $y$ 점수로 조정된다.

$$y = \sigma(Y)/\sigma(X)[x - \mu(X)] + \mu(Y) \tag{3}$$

식(3)은 검사 X의 점수를 검사 Y의 척도로 조정한 점수가 검사 Y의 평균과 표준편차와 동일하도록 조정한 것이다. 두 검사를 받은 집단이 서로 다르면, 공통검사를 공변인으로 하여 두 검사의 점수를 2단계로 조정하는 다음의 연쇄방법을 적용할 수 있다. 식(2)의 기호를 그대로 쓰고, 검사 X와 Y의 표준편차를 각각  $\sigma_1(X)$ ,  $\sigma_2(Y)$ 라 하고, 공통검사 V의 집단1과 집단2의 표준편차를 각각  $\sigma_1(V)$ ,  $\sigma_2(V)$ 라 하자. 그리고  $a = \sigma_1(V)/\sigma_1(X)$ ,  $b = \sigma_2(Y)/\sigma_2(V)$ 라 하자. 식 (3)을 두 번 적용하면 다음 식이 된다.

$$y = ab [x - \mu_1(X)] + b[\mu_1(V) - \mu_2(V)] + \mu_2(Y) \tag{4}$$

연쇄방법은 결합분포를 필요로 하지 않고, 두 집단이 매우 유사하다는 가정을 필요로 하지 않기 때문에 간편하다. 다만, 직접적으로 모집단을 포함하여 분석하지 않기 때문에 조정결과를 해석할 때 의미가 애매하다. 이러한 문제를 해결하기 위하여 두 검사와 가교검사의 회귀식이 두 집단에서 동일하다면 터커의 합집단 방법을 다음과 같이 적용할 수 있다.  $\mu_s(X)$ 를 합집단의 검사 X의 평균,  $\mu_s(Y)$ 를 합집단의 검사 Y의 평균,  $\sigma_s(X)$ 를 합집단의 검사 X의 표준편차,  $\sigma_s(Y)$ 를 합집단의 검사 Y의 표준편차라 하자. 검사 X에 응시한 집단을 1, 검사 Y에 응시한 집단을 2로 하고, 응시자 수의 비율을  $w_1$ ,  $w_2$ 라 하며,  $\sigma_1(X,V)$ 를 검사 X와 공통검사 V의 공분산,  $\sigma_1^2(V)$ 을 집단1의 공통검사의 분산이라 하자. 그러면 터커의 합집단 선형방법은 다음의 식들로 구성된다.

$$y = \sigma_s(Y)/\sigma_s(X)[x - \mu_s(X)] + \mu_s(Y) \tag{5}$$

$$\mu_s(X) = \mu_1(X) - w_2r_1[\mu_1(V) - \mu_2(V)] \tag{6}$$

$$\mu_s(Y) = \mu_2(Y) + w_1r_2[\mu_1(V) - \mu_2(V)] \tag{7}$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2r_1^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2r_1^2[\mu_1(V) - \mu_2(V)]^2 \tag{8}$$

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1r_2^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2r_2^2[\mu_1(V) - \mu_2(V)]^2 \tag{9}$$

$$r_1 = \sigma_1(X,V)/\sigma_1^2(V) \tag{10}$$

$$r_2 = \sigma_2(Y,V)/\sigma_2^2(V) \tag{11}$$

### (3) 동백분위방법

동백분위방법은 누적분포를 일치시키도록 조정하는 방법이다. 무선집단이거나 단일집단 설계의 경우 두 검사 X와 Y가 있을 때, 각 검사의 누적분포함수로부터 각 점수의 백분위 함수를 도출할 수 있다. P를 x의 백분위 함수, Q를 y의 백분위 함수라고 하고  $Q^{-1}$ 를 Q의 역함수라고 하면, 검사 X의 특정점수 x는 검사 Y의 척도로 변환된 점수  $e_Y(x)$ 로 식(12)와 같이 조정된다.

$$e_Y(x) = y = Q^{-1}[P(x)] \tag{12}$$

응시집단 1은 검사 X를 치르고, 응시집단 2는 검사 Y를 치러서 서로 직접 비교할 수 없지만, 두 집단 모두 공통검사 V를 받은 경우 V를 통하여 간접적으로 비교할 수 있다. 이 경우의 동백분위방법은 연쇄방법과 합집단에 의한 방법이 있다. 연쇄방법은 응시집단 1의 검사 X에서 특정점수 x의 백분위와 동일한 백분위를 갖는 검사 V의 v점수를 계산하고, 응시집단 2의 v점수의 백분위에 해당하는 검사 Y의 y점수를 계산하면 된다. 응시집단 1의 공통검사 V의 백분위 함수를  $R_1$ 라 하고, 응시집단 2의 공통검사 V의 백분위 함수를  $R_2$ 라 하면, X의 특정점수 x는 검사 Y의 척도로 변환된 점수  $e_Y(x)$ 로 식(13)을 거쳐 식(14)로 조정된다.

$$e_V(x) = v = R_1^{-1}[P(x)] \tag{13}$$

$$e_Y(x) = y = Q^{-1}[R_2(v)] \tag{14}$$

동백분위방법 중 빈도추정방법은 합집단의 빈도분포를 추정하고, 백분위 함수를 구한 후 역함수를 구하여 조정한다. 공통검사의 점수 v가 주어졌을 때 검사 X의 점수와 검사 Y 점수의 조건분포가 집단 1과 집단2에서 같으면, 다음의 식(15)와 식(16)으로 합집단의 빈도분포를 계산할 수 있다.

$$f_s(x) = w_1 f_1(x) + w_2 \sum v f_1(x|v) h_2(v) \tag{15}$$

$$g_s(y) = w_1 \sum v g_2(y|v) h_1(v) + w_2 g_2(y) \tag{16}$$

식(15) 및 식(16)에서  $s$ 는 합집단을 의미하며, 아래첨자 1과 2는 각각 검사 X와 검사 Y를 치른 모집단을 의미한다.  $f$ 와  $g$ 는 각각 검사 X와 검사 Y의 분포를 나타내며,  $h$ 는 주변분포이고  $w_1$ 과  $w_2$ 는 모집단 1과 모집단 2의 가중치이다. 합집단의 빈도분포를 구하면, 누적빈도분포를 구하고 합집단의 백분위 함수를 구한 후 식(12)와 같이 동백분위에 해당하는 동등한 점수를 찾을 수 있다.



### Ⅲ. 수능 사탐과 과탐의 동등화 가산점 조정 비교

현행 수능은 각 영역별 및 과목별로 표준점수를 산출하고 있다. 즉, 언어, 외국어, 수리 영역은 영역별 평균 100점, 표준편차 20점, 탐구영역의 각 선택과목은 과목별 평균 50점, 표준편차 10점을 갖도록 변환한 표준점수 분포에서 개인이 획득한 원점수가 어느 위치에 해당하는지 표준점수를 산출하여 제공한다. 또한 각 영역별 및 과목별 응시학생 전체에 대하여 특정 표준점수보다 낮은 표준점수를 획득한 학생의 비율로 표준점수 백분위를 계산하며, 표준점수를 9등급(stanine scale)으로 나누고 특정 표준점수가 속해 있는 등급을 표시한다.

본 연구는 각 영역별 및 과목별 표준점수, 백분위, 등급으로 산출된 2010학년도 수능 자료를 대상으로 분석하고자 한다. 다만, 수리영역은 ‘가’형과 ‘나’형으로 나뉘고 가형의 경우 선택 과목으로 나누어서 자료가 구성되어 있지만, 본 연구는 수리영역 점수 조정을 하려는 것이 아니므로 이를 구분하지 않고 하나의 수리영역 점수로 한다. 또한 공변인수로 활용하는 인수의 표준점수와 백분위는 언어, 수리, 외국어의 표준점수 평균과 백분위 평균으로 한다. 특정 응시자가 응시한 선택과목 표준점수의 평균에 두 배를 하여 탐구영역 표준점수로 산출함으로써 언어, 수리, 외국어 영역의 척도와 동일하게 하고, 선택과목 백분위를 평균하여 탐구영역 백분위를 산출하여 분석하고자 한다.

#### 1. 평균방법

수학능력 수준별 집단을 구분하기 위하여 언어, 수리, 외국어 등급을 평균하여 수능 종합 등급을 계산하고, 종합등급에 따라 8개 집단으로 구분하였다. <표 1>에 종합등급 집단별 응시자 수, 사탐 및 과탐의 표준점수와 백분위, 인수의 표준점수 및 백분위를 나타냈다. 예를 들어 수능 종합등급이 1.00~1.99 등급에 속하는 사탐응시자는 21,339명이고, 이들의 사탐 표준점수 평균은 127.18점, 백분위 평균 86.54점, 인수의 표준점수 평균 132.35점, 백분위 평균 95.13점을 표시한 것이다.

특정대학의 모집단위 또는 전형유형에 수능 종합등급이 1.00~1.99인 지원자들이 경쟁한다면, 수능 전체 응시자에 대한 가산점을 계산하는 것보다 특정대학의 입학환경인 최상위권 지원자들의 경쟁에 적합한 가산점을 산정하는 것이 바람직하다. 종합등급 1.00~1.99인 경우 사탐응시자의 인수의 표준점수 평균은 132.35이고 과탐 응시자의 인수의 표준점수 평균은 131.09이므로 사탐응시자의 인수의 점수가  $132.35 - 131.09 = 1.26$ 점 높다. 그런데 사탐의 표준점수 평균은 127.18점이지만 과탐 표준점수 평균은 128.95점으로 1.77점 높다. 따라서 사탐

의 표준점수에  $1.26+1.77=3.03$ 점, 또는 과탐 표준점수에  $-3.03$ 점의 가산점을 부여해야 동일한 수학능력을 나타낸다.

〈표 1〉 종합등급별 탐구영역 응시집단의 표준점수 평균 및 백분위 평균

종합등급	사탐응시자					과탐응시자				
	응시자 수	탐구영역		언수외		응시자 수	탐구영역		언수외	
		표준점수	백분위	표준점수	백분위		표준점수	백분위	표준점수	백분위
1.00~1.99	21,339	127.18	86.54	132.35	95.13	8,261	128.95	88.69	131.09	95.04
2.00~2.99	34,464	119.01	77.33	125.43	87.17	19,675	120.10	78.45	124.17	86.82
3.00~3.99	51,019	111.62	67.49	117.09	75.90	35,490	110.84	66.03	116.54	75.19
4.00~4.99	70,393	103.93	56.38	106.83	60.80	50,402	102.06	53.29	107.57	60.49
5.00~5.99	78,815	95.83	44.26	95.99	43.45	49,475	93.52	40.53	96.85	43.89
6.00~6.99	63,552	87.89	32.32	85.86	27.26	29,978	85.78	29.09	85.76	28.20
7.00~7.99	39,647	80.88	22.06	76.10	14.57	13,886	79.48	20.19	75.42	15.27
8.00~9.00	17,860	76.00	15.27	65.27	5.71	5,570	75.08	14.46	64.41	5.80
합계	377,089					212,737				

사탐을 X, 과탐을 Y, 사탐 응시집단을 집단1, 과탐 응시집단을 집단2라고 하고, 언수외를 V라고 하여 식(2)를 적용하면  $y=x-127.18+(132.35-131.09)+128.95=x+3.03$ 이 된다. 이와 같은 방법으로 각 집단별 표준점수와 백분위에 대하여 과탐 가산점을 계산하여 〈표 2〉에 나타냈다. 일반적으로 이과계열의 모집단위에 사탐응시자들이 지원하도록 교차지원을 허용하는 경우는 과탐 응시자에게 가산점을 주고, 문과계열 모집단위에 과탐 응시자들이 지원하도록 교차지원을 허용하는 경우는 사탐 응시자에게 가산점을 주는 것이 보통이다. 다만, 과탐 가산점과 사탐 가산점은 크기가 같고 부호가 다르다.

〈표 2〉 등급별 평균방법을 적용한 과탐 가산점

등급 점수	1.00~1.99	2.00~2.99	3.00~3.99	4.00~4.99	5.00~5.99	6.00~6.99	7.00~7.99	8.00~9.00
표준점수	-3.03	-2.35	0.23	2.61	3.16	2.01	0.73	0.06
백분위	-2.24	-1.48	0.75	2.77	4.17	4.19	2.58	0.90

남보우(2010)의 연구결과에 의하면 언수외를 공변인으로 전체집단에 대하여 평균방법을 적용하면 과탐에 표준점수 2.14점, 백분위 3.34점의 가산점이 계산되어 과탐응시자의 수학능력이 전체적으로 사탐응시자의 수학능력보다 높은 것으로 보고하고 있다. 그러나 <표 2>에서 나타난 바와 같이 상위등급의 경우에는 과탐응시자가 사탐응시자보다 수학능력이 낮고, 중하위 등급의 경우 과탐응시자가 사탐응시자보다 수학능력이 높은 것으로 나타난다. 따라서 전체응시자를 대상으로 평균방법으로 가산점을 산정하는 것은 특정대학의 입학전형에 적합하지 않을 수 있다. 반면, 지원자들의 수학능력 집단별로 평균방법을 적용하여 가산점을 산정하는 것은 집단별 경쟁상황을 반영하고, 이해하기 쉬우며, 집단별 언수외 종합점수를 기준으로 조정하였다는 근거가 명확하므로 대입전형의 실제에서 적용하기 용이할 것이다.

## 2. 선형방법

선형방법은 점수 수준별로 가산점을 계산하는 선형식을 구하는 것이다. 따라서 전체응시자를 대상으로 선형식을 구하고 점수수준별 가산점을 산정하는 것이 필요하다. 연쇄 선형방법으로 식(4)를 적용하기 위해서는 각 탐구영역의 점수평균과 분산, 공변인의 점수평균과 분산이 필요하다. 터커의 합집단에 의한 선형방법으로 식(5)를 적용하기 위해서는 합집단의 평균과 분산을 구해야 하고, 이를 위해서 각 탐구영역의 평균 및 분산과 각 집단이 응시한 공변인의 평균과 분산, 탐구영역과 공변인의 공분산 및 응시자 수의 비율이 필요하다.

<표 3> 선형방법 적용을 위한 모수 추정치

구 분		사탐(X) : 집단1			과탐(Y) : 집단2		
		모수	추정치		모수	추정치	
			표준점수	백분위		표준점수	백분위
응시자 수(비율)		$w_1$	0.639		$w_2$	0.361	
공변인 (언수외 : V)	평균	$\mu_1(V)$	100.36	50.45	$\mu_2(V)$	102.72	53.93
	분산	$\sigma_1^2(V)$	334.35	698.02	$\sigma_2^2(V)$	258.97	549.90
해당 탐구영역	평균	$\mu_1(X)$	99.76	49.71	$\mu_2(Y)$	99.98	49.85
	분산	$\sigma_1^2(X)$	272.23	563.11	$\sigma_2^2(Y)$	271.36	559.80
공분산(탐구영역, 언수외)		$\sigma_1(X,V)$	250.26	522.59	$\sigma_1(Y,V)$	208.28	439.23

사탐을 X, 과탐을 Y라 하고, 사탐 응시집단을 집단1, 과탐 응시집단을 집단2라고 하면, 선형방법을 적용하기 위해서 모형에 활용되는 모수들을 다음의 <표 3>으로 정리하였다. 사탐 응시자의 연수의 표준점수 평균이 100.36이고, 분산은 334.35이며, 사탐응시자의 사탐 표준점수 평균은 99.76이고 분산은 272.23임을 보여준다. 또한 사탐응시자의 연수의 표준점수와 사탐점수의 공분산은 250.26임을 보여준다.

연쇄방법으로 식(4)에 <표 3>의 모수를 적용하면, 과탐의 표준점수를 y라고 할 때 사탐의 척도 점수 x로 조정하는 식은  $x = 0.88149y + 13.75777$ 이 된다. 1.00~1.99등급 과탐 표준점수 평균은 128.95점이고, 이를 조정 선형식에 대입하면 사탐척도로 127.43점이된다. 따라서 가산점은 -1.52점이다. 이와 같이 과탐 상위 4개 집단의 표준점수 평균점과 백분위 평균점에 대하여 가산점을 계산한 결과를 <표 4>에 제시하였다. 또한 터커의 합집단 방법으로 식(5) 내지 식(11)에 <표 3>의 모수를 적용하여 과탐 상위 4개 집단의 표준점수 평균점과 백분위 평균점에 대하여 가산점을 계산한 결과도 제시하였다. 터커의 합집단에 의한 조정의 폭이 연쇄방법에 의한 조정의 폭보다 작은 것으로 나타나지만 조정의 방향은 일치하는 것으로 나타난다. 이러한 조정차이는 김현철(2003)의 연구에서 분석되었는데 공변인과 탐구영역의 상관 계수가 1보다 작기 때문인 것으로 해석될 수 있다.

<표 4> 선형방법을 적용한 과탐 가산점 계산

방법	지표	조정 선형식	수준별 과탐 가산점(y-x)							
			과탐 표준점수 수준(y)				과탐 백분위 수준(y)			
			128.95	120.10	110.84	102.06	88.69	78.45	66.03	53.29
연쇄방법	표준점수	$x = 0.881y + 13.758$	-1.52	-0.47	0.62	1.66				
	백분위	$x = 0.890y + 8.459$					-1.28	0.15	1.21	2.61
터커의 합집단 방법	표준점수	$x = 0.922y + 9.364$	-0.73	-0.04	0.69	1.37				
	백분위	$x = 0.927y + 6.076$					-0.38	0.37	1.27	2.20

### 3. 동백분위법

동백분위법은 백분위가 동일하도록 점수를 조정하는 것이기 때문에 분포의 형태가 불규칙적인 경우 해당부분에서 불안정한 조정이 나타나게 된다. 이러한 경우에 분포의 완곡화(smoothing)를 한다. 본 연구는 언어, 수리, 외국어 점수를 합하여 공변인으로 정하였고, 탐

구영역의 점수는 선택과목의 점수를 모두 합하였으며, 응시자의 수가 대단히 많기 때문에 과목별 불규칙성이 상당부분 완화되어 동백분위법을 적용하는 것이 용이하였다.

〈표 5〉에 각 집단별 연수의 표준점수와 탐구영역 표준점수를 순서대로 정리하여 동일한 백분위에 해당하는 표준점수의 일부를 제시하였다. 동백분위법은 백분위가 동일한 점수로 조정하는 것이므로 연수외를 공통검사로 활용하면 과탐의 표준점수를 사탐의 척도로 변환하여 가산점을 계산할 수 있다. 예를 들어 과탐의 표준점수 119.00점의 백분위는 85이고 동일한 백분위의 과탐 응시집단의 연수의 표준점수는 120.00점이다. 반면 연수의 표준점수 120.00은 사탐 응시집단에서 비례하여 계산하면 백분위가 82.50이고, 동백분위의 사탐 표준점수는 116.75이다. 동백분위를 연쇄적으로 적용한 결과 과탐 표준점수 119.00점은 사탐 표준점수 116.75점과 동등하므로 같은 점수로 조정하기 위해서는 과탐 가산점은 -2.25점이 된다. 이러한 연쇄방법을 적용하여 과탐의 각 점수에 대하여 가산점을 계산할 수 있다.

〈표 5〉 동백분위 집단별 표준점수 및 검사별 합집단 백분위 예시

백분위	동백분위 집단별 검사별 표준점수				동표준점수 검사별 합집단 백분위		
	사탐집단		과탐집단		표준점수	사탐	과탐
	연외수	사탐	연외수	과탐		Ps(x)	Ps(y)
86	122.67	119.00	120.33	119.50	120	87.94	86.26
85	122.00	118.50	120.00	119.00	119	86.61	85.08
84	121.33	118.00	119.33	118.00	118	85.18	83.86
83	120.33	117.00	118.67	117.50	117	83.68	82.58
82	119.67	116.50	118.33	116.50	116	82.10	81.24

빈도추정에 의한 합집단 동백분위 방법을 적용하기 위하여 합집단 백분위를 계산한 결과의 일부도 〈표 5〉에 제시하였다. 식(15)와 (16)을 적용하여 합집단 백분위를 계산하기 위해서는 조금 복잡한 과정을 거쳤다. 우선 공통검사 연수외의 특정점수와 탐구영역 특정점수를 동시에 획득한 빈도를 계산하고, 이를 각 집단의 총인원으로 나누어 결합확률분포와 주변확률분포를 구하였다. 이러한 과정으로 식(15)와 (16)에서 각 점수  $x$ 와  $y$ 에 해당하는  $f_1(x)$ 와  $g_2(y)$ 를 구하고, 사탐과 과탐 응시집단의 공통검사 연수외 점수의 주변확률분포  $h_1(v)$ 와  $h_2(v)$ 를 구하였다. 다시 각 집단별 결합확률분포를 연수외 주변확률분포로 나누어  $f_1(x|v)$ 와  $g_2(y|v)$ 를 구하고, 식 (15)와 (16)을 적용하여 합집단의 빈도분포  $f_s(x)$ 와  $g_s(y)$ 를 구하였다.

합집단의 빈도분포로부터 누적빈도분포를 구한 후 각 표준점수에 해당하는 백분위를 구한 것이다. 합집단의 백분위가 주어지면 동백분위에 해당하는 점수로 조정할 수 있다. 예를 들어 과탐 응시집단의 표준점수 119점은 과탐 검사를 합집단으로 추정했을 때 백분위 85.08에 해당하고, 사탐 검사 합집단의 백분위 85.08은 사탐 표준점수 117.93점에 해당한다. 따라서 과탐 점수에 -1.07점을 가산해야 동등한 점수가 된다. 이와 같이 빈도추정방법에 의한 합집단으로 동백분위 방법을 적용하여 과탐의 각 점수에 대하여 가산점을 계산할 수 있다.

상위등급 4개 집단의 과탐 표준점수 평균에 대한 가산점을 연쇄방법에 의한 동백분위법과 빈도추정 합집단에 의한 동백분위법으로 계산한 결과를 <표 6>에 나타냈다. 연수의 1.00~1.99등급 과탐 응시집단의 표준점수 평균은 128.95점이고 백분위는 95.30이며, 동백분위 연수의 표준점수는 127.30점이다. 사탐 응시집단에서 연수의 표준점수 127.30은 백분위 92.02이며, 동백분위 사탐 표준점수는 124.52으로 연쇄적으로 계산된다. 따라서 과탐 128.95점에 대한 가산점은 -4.43점으로 계산된다. 빈도추정 합집단 방법의 경우 동백분위에 해당하는 과탐 표준점수와 사탐 표준점수를 비교하여 과탐 가산점을 산출하였다.

<표 6> 동백분위방법을 적용한 과탐 표준점수 가산점 계산예

연쇄방법에 의한 조정						빈도추정 합집단 방법에 의한 조정					
과탐응시자			사탐응시자			과탐 가산점	과탐 합집단		사탐 합집단		과탐 가산점
과탐	연수외	표준 점수	연수외	사탐	표준점수 (y)		백분위 Ps(y)	백분위 Qs(x)	표준점수 (x)		
표준 점수	백분위		표준 점수	백분위						표준 점수	
128.95	95.30	127.30	127.30	92.02	124.52	-4.43	128.95	94.94	94.94	126.63	-2.32
120.10	86.60	120.93	120.93	83.36	117.36	-2.74	120.10	86.37	86.37	118.77	-1.33
110.84	72.68	114.35	114.35	73.47	111.24	0.40	110.84	73.48	73.48	111.06	0.22
102.06	56.11	106.78	106.78	63.56	105.78	3.25	102.06	57.88	57.88	103.43	1.37

#### 4. 조정방법의 비교

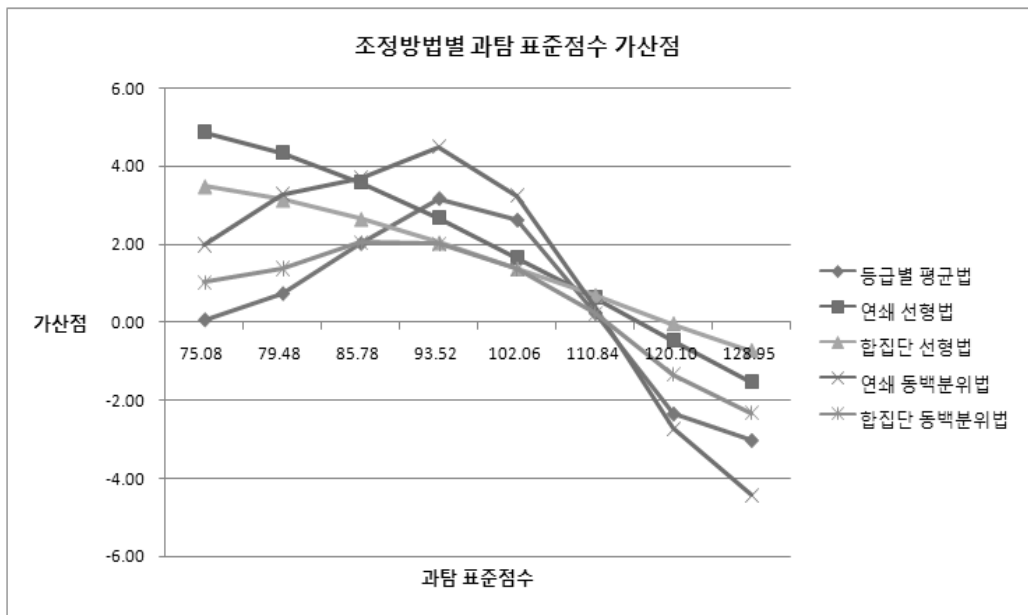
다섯 가지 방법으로 각 집단의 과탐 표준점수 평균점 및 백분위 평균점수에 대하여 가산점을 계산한 결과를 <표 7>에 나타냈고, 표준점수 가산점을 [그림 1]에 표시하였다. 전체 응시자를 기준할 때 과탐 응시자가 사탐응시자보다 연수의 평균점수가 높아서 과탐에 가산점을 주어야 한다고 제안할 수 있다. 그러나 수능 등급별로 분석하면 다섯 가지 조정방법 적용결

과 모두 1.00~2.99등급까지의 상위 점수대에서는 과탐을 감점하고, 3.00등급 이하에서는 가점을 하는 것으로 나타난다. 따라서 특정대학의 모집단위나 전형유형 지원자의 수험능력에 따라 가산점이 달리 적용되어야 한다.

〈표 7〉 각 조정방법에 따른 과탐 가산점

종합등급	과탐 표준점수	과탐 표준점수 가산점						과탐 백분위	과탐 백분위 가산점				
		등급별 평균법	선형법 적용		동백분위법		등급별 평균법		연쇄법	합집단	연쇄법	합집단	
			연쇄법	합집단	연쇄법	합집단							
1.00~1.99	128.95	-3.03	-1.52	-0.73	-4.43	-2.32	88.69	-2.24	-1.28	-0.38	-3.05	-1.60	
2.00~2.99	120.10	-2.35	-0.47	-0.04	-2.74	-1.33	78.45	-1.48	-0.15	0.37	-2.18	-0.95	
3.00~3.99	110.84	0.23	0.62	0.69	0.40	0.22	66.03	0.75	1.21	1.27	0.54	0.60	
4.00~4.99	102.06	2.61	1.66	1.37	3.25	1.37	53.29	2.77	2.61	2.20	3.60	2.07	
5.00~5.99	93.52	3.16	2.67	2.04	4.49	2.01	40.53	4.17	4.01	3.13	6.25	3.25	
6.00~6.99	85.78	2.01	3.59	2.65	3.71	2.04	29.09	4.19	5.27	3.96	6.62	3.40	
7.00~7.99	79.48	0.73	4.34	3.14	3.28	1.38	20.19	2.58	6.24	4.61	6.49	2.44	
8.00~9.00	75.08	0.06	4.86	3.49	1.98	1.03	14.46	0.90	6.87	5.02	4.25	1.49	

[그림 1] 각 조정방법에 따른 과탐 표준점수 가산점 비교



연쇄 선형방법을 적용한 조정보다 합집단 선형방법으로 조정한 결과가 대체로 조정의 폭이 적은 것으로 나타나며, 동백분위 방법의 경우에도 연쇄방법의 조정결과보다 합집단 방법을 적용한 결과가 조정의 폭이 적은 것으로 나타난다. 그러나 모두 유사한 경향을 나타낸다. 선형방법은 선형식을 적용하므로 그래프에서 직선으로 나타나서 최저점과 최고점 근처에서 조정의 폭이 커지기 때문에 중간점수대 이외에서 적용하는 것은 어려운 것으로 해석된다.

조정방법 간의 비교는 준거가 있어야 가능하다. 남현우(2003)의 연구에 의하면 두 검사를 치른 집단을 묶은 합집단의 분포를 통해 동등 점수의 짝을 찾는 방법은 동등화가 한번만 이루어지기 때문에 소위 2단계 동등화인 연계화 방법보다 오차가 적은 결과를 가져오는 것으로 지적하고 수능 선택과목 점수조정에 합집단 방법을 적용할 수 있는 것으로 논의하였다. 또한 빈도추정 방법은 검사 점수 간 관계가 선형이어야 한다는 가정을 필요로 하지 않고, 점수분포의 형태를 고려하기 때문에 더 정확한 결과를 가져오는 것으로 지적하였다. Schmitt et al.(1990)도 Tucker의 선형방법과 빈도추정의 동백분위방법이 표본의 능력 차이에 가장 적게 영향을 받는 것으로 밝혔다. 이러한 논의를 종합할 때 빈도추정에 의한 합집단 동백분위방법이 비교적 정확한 조정 결과를 도출하는 것으로 볼 수 있다.

따라서 본 연구에서는 빈도추정 합집단 동백분위방법으로 조정한 결과를 준거로 각 방법에 의한 결과에 대하여 평균자승편차(Root Mean Square Difference)을 계산하여 <표 8>에 제시하였다. 표준점수 및 백분위를 조정한 결과에 대한 평균자승편차는 등급별 평균방법이 가장 작은 것으로 나타나서 빈도추정 합집단 동백분위 방법의 결과와 가장 근접하는 것으로 나타난다.

<표 8> 빈도추정 합집단 동백분위방법의 조정결과에 대한 평균자승편차

지표 \ 적용방법	등급별 평균방법	연쇄 선형방법	터커의 선형방법	연쇄 동백분위방법
표준점수 조정결과	0.85	1.87	1.32	1.71
백분위 조정결과	2.97	4.78	3.70	5.03

빈도추정에 의한 합집단 동백분위방법은 비록 바람직한 것으로 밝혀졌지만, 조건분포가 동일하여야 한다는 통계적 가정이 필요하다. 또한 빈도추정 및 결합분포 등 계산과정이 매우 복잡하여 대입전형 실무에서 적용하는데 부담이 크며, 조정의 결과를 간단하게 설명하기가 어렵다. 반면, 등급별 평균방법은 계산이 간편하고, 조정결과를 명확하게 설명할 수 있으며, 빈도추정에 의한 합집단 동백분위방법으로 조정한 결과와 가장 근접한 것으로 나타나서 대입전형에 현실적으로 적용할 수 있는 실용적인 방법으로 판단된다.



## IV. 결 론

현행 수능은 영역별, 선택과목별로 표준점수를 산출하므로 시험의 난이도 문제는 해결되지 만 탐구영역 응시집단 간의 능력 차이는 반영되지 못한다. 능력 차이가 반영되지 못한 서로 다른 수능 점수를 비교하여 학생을 선발하면 공정성 문제가 대두되고 또한 수학능력이 높은 학생을 선발하기 어렵게 한다. 우리나라 대학의 90%이상이 탐구영역 간 점수를 비교하여 학생을 선발하는 것으로 나타나고 있다. 이러한 상황에서 공정한 대입전형과 수학능력이 높은 학생을 선발할 수 있도록 탐구영역 응시집단 간의 수학능력 차이를 반영하여 각 탐구영역의 점수를 비교할 수 있도록 조정할 필요가 있다.

교육평가분야의 측정이론에서 점수조정을 위한 다양한 동등화 방법들이 제시되었고, 수능 점수조정을 위한 전통적인 방법들을 비교한 연구들이 있었다. 그러나 이러한 연구들은 2005 학년도 이전의 자료를 분석하였고, 이후에는 수능 점수가 제공되지 않아서 바람직한 점수조정 방법을 찾는 연구가 없었다. 2010년에 한국교육과정평가원은 수능 점수를 연구자에게 제공하였으며, 많은 대학들이 교차지원을 허용하고 있으므로 탐구영역 간 가산점을 산정할 필요가 있었다. 대입전형 현장에서는 가산점 산정을 위하여 복잡한 점수조정 방법보다 간편하고 이해가 용이한 방법이 필요하였다. 특히 특정 대입전형에 일정한 범위의 수학능력을 가진 지원자들이 경쟁한다는 점을 감안하여 본 연구는 평균방법을 수능 등급별로 적용하고 검토하였다. 수능 등급별 평균방법의 현실적용 가능성을 검토하기 위하여 측정이론의 연쇄 선형방법, 터커의 합집단 선형방법, 연쇄 동백분위방법, 빈도추정 동백분위방법을 2010 수능 자료에 적용하고 비교하였다. 공변인은 선행연구 결과에 따라 연수의 종합점수로 선정하여 적용하였다.

분석결과 등급별 평균방법으로 점수를 조정하는 것이 바람직한 특성을 갖는 것으로 나타났다. 즉, 등급별 평균방법은 조정의 근거가 명확하고 간편하게 설명되며, 선행연구에서 가장 바람직한 방법으로 제시되고 있는 빈도추정 동백분위 방법과 평균자승편차가 가장 적게 나타나는 것으로 분석되었다. 등급별 평균방법으로 점수를 조정한 결과, 연수의 종합등급 1.00~1.99 집단의 과탐 표준점수 가산점은 -3.03점이 되고, 2.00~2.99 집단의 가산점은 -2.35점이 된다. 반면 4.00~4.99 집단의 과탐 표준점수 가산점은 2.61점, 5.00~5.99 집단의 가산점은 3.16점이 된다. 이러한 결과는 등급별 평균방법이 정교한 점수조정 방법인 빈도추정 동백분위방법의 결과와 가장 근접하는 것으로 대입전형에 실용적으로 적용할 수 있음을 나타낸다.

본 연구에서 제시한 등급별 평균방법은 남보우(2010)의 연구에서 도입한 전통적인 선형방법, 동백분위방법보다 간편하고 해석이 용이하며 산정결과가 근사하여 실용적인 방법이다.

또한 등급별 평균방법은 남보우(2011)가 최적화 모형을 도입하여 선택과목 간 가산점을 산정하는 방법보다 탐구영역 간 가산점을 용이하게 산정하는 방법이다. 따라서 탐구영역 간 가산점 산정문제에 실용적인 등급별 평균방법을 제시한 점이 본 연구의 특징이다.

본 연구는 교차지원을 허용할 때 탐구영역 간 가산점 산정에 수능 등급별로 평균방법을 적용함으로써 실용적인 접근방법을 제시하였다. 그러나 대입전형에서 등급별 평균방법을 활용하여 가산점을 산정하려면 특정대학 전형의 특성을 면밀하게 고려해야 할 것이다. 예를 들어 특정 탐구영역의 능력이 중요하게 요구되는 모집단위는 가능하면 가산점을 크게 하여야 할 것이고, 특정 탐구영역 능력보다 보편능력이 중시되는 영역에서는 가산점의 크기를 적게 하는 것이 필요할 것이다.

## 참고문헌

- 김현철. 2003. 모의실험에 의한 대학수학능력시험 선택과목 점수 동등화 방법들의 적합성 비교(1): 원점수와 동등화점수의 차이분석, 교육학연구, 41(1), 363-384.
- 남명호, 김신영, 남현우, 권순달, 박정, 조지민. 2002. 2005학년도 대학수학능력시험 점수 체제 연구, 연구보고서 대수능 CAT 2002-32, 한국교육과정평가원.
- 남보우. 2002. 선형계획모형에 의한 점수조정, 2005학년도 대학수학능력시험 점수체제 세미나 연구자료 ORM 2002-16(2002.12.13), 한국교육과정평가원, 44-59.
- 남보우. 2010. 수능 선택과목의 표준점수 심층분석 및 대입전형 활용방안, 국가수준 학업성취도 평가와 대학수학능력시험 자료 분석 연구, 연구보고 RRE 2010-12-2, 한국교육과정평가원, 443-531.
- 남보우. 2011. 대입전형을 위한 수능 탐구영역 점수조정 연구, 산업연구, 단국대학교 부설 미래산업연구소, 51-69.
- 남보우. 2011. 대입수능 선택과목 점수조정을 위한 선형계획모형 개발 및 활용, 경영과학, 28(1), 141-158.
- 남현우. 2001. 검사동등화 방법, 교육과학사.
- 남현우. 2003. 2005수능의 탐구영역 내 과목 간 점수조정에서 합집단 연계화방법의 적용가능성, 교육평가연구, 16(1), 163-182.
- 남현우. 2004. 대학수학능력시험 영역별선택과목의 연계화 가능성탐색, 교육평가연구, 17(2), 93-105.
- 박정. 2005. 문항반응이론에 의한 2005 대학수학능력시험 영역별 선택과목 점수조정의 타당성 검증, 교육학연구, 43(2), 135-154.
- 부재율. 1999. 검사동등화 시의 고려사항, 교육평가연구, 12(1), 43-64.
- 성태제. 2000. 다른 교과와 점수를 연결하는 통계적 조절 방법의 타당성, 교육평가연구, 13(1), 63-99.
- 최성열, 이규민, 박아청. 2007. 검사점수 연계화 방법의 상대적 적절성 비교, 교육심리연구, 21(1), 1-25.
- 황형태. 2005. 대학수학능력시험에서 표준점수제의 개선방안에 대한 연구, 응용통계연구, 18(3), 521-532.
- M.J. Kolen and R.L Brennan. 2004. Test Equating, Scaling, and Linking, 2nd Ed.,

Springer.

A.P. Schmitt, L.L. Cook, N.J. Dorans and D.R. Eignor. 1990. Sensitivity of equating results to different sampling strategies, Applied Measurement in Education, 3(1), 53-71.

## A Comparison of Equating/Linking Methods for the CSAT Score Adjustment between Optional Parts of Social Studies and Science

Nam, Bo-Woo\*

### ABSTRACT

To reflect the ability gap between applicant groups in the college entrance competition, it needs to adjust the standard score of CSAT(College Scholastic Aptitude Test) optional parts, such as Social Studies and Science for the college admission decision. But it is difficult to choose a practical equating method.

Five equating or linking methods were analyzed by using the 2010 CSAT data of Korea Institute of Curriculum and Evaluation. This study developed a mean method for each 8 ranked groups. Other four methods of classical test theory are two-stage linear method, Tucker's linear method, chained equi-percentile method and frequency estimation equi-percentile method. We use a combination of Korean Language, Mathematics, and Foreign Language parts as a co-variate to the analysis.

We got three evidences that the mean method for each ranked groups can be used practically for the college admission decision. First, the ranked group mean method has a comprehensible basis for the score adjustment. Second, the method is simple and easy to apply. Third, the results of the mean method have a minimum RMSD in reference to the frequency estimation equi-percentile method which is relatively correct but complex.

**Key Words** : College Admission, Test Equating, Linking, CSAT, Score Adjustment

\* Professor of Management, Dankook University

