韓國數學敎育學會誌 시리즈 A <數學敎育>
1998. 5. 제 37권, 제 1호, 1-13.

J. Korea Soc. Math. Ed. Ser. A: The Mathematical Education
May. 1998, Vol. 37, No. 1, 1-13.

수학교과의 동형고사 문항에서 양호도 향상에 유효한 최적정답율 산정에 관한 연구

홍 석 강 (동국대학교)

I. 서 론

학습 평가에 있어서 하나의 평가도구가 양호한 것이 되기 위해서는 이를 구성하고 있는 시험문제 혹은 검사문항 하나 하나가 양질의 것이어야 하며 이 양호도를 판단하는 측도에는 일반적으로 문항의 난이도, 문항의 변별도 및 문항 반응 분포 등이 있다. 즉 입학 시험의 경우이거나 학생들 사이에 등급을 매겨야 하는 경우 수험생들의 능력차를 엄밀히 가려내기 위해서는 문항의 난이도가 주요관심의 대상이 되며 이때 난이도 계수는 0.05이하이거나 0.95이상이면 일반적으로 부적격한 것으로 판단하게 되고 또 수험생의 능력을 변별하는 힘의 크기를 측정할 경우 문항변별도 계수는 보통 0.4 이면 좋은 문항으로 인정할 수 있다. 이외에도 좋은 문항의 선택을 위한 측도로써 타당도와 신뢰도의 개념도 고려되어야하고 또 그 검사도구가 측정하고자 하는 준거에 따라 그 검사에 대한 이론적인 설명과 기술보고서가 반드시 제시되어져서 검사도구 사용자들이 양호한 문항의 선택에 있어서 만전을 기할 수 있도록 해야 할 것이다.

일반적으로 문항의 난이도와 변별력을 향상시키고자 한다면 난이도는 낮아지게 되는 상호 역상관인 관계가 성립하는 것으로 알려져 있는데 그 예로써 근년에 시행한 대학수학능력시험에서 문제들을 약간 쉽게 출제한 결과 고득점에는 동점자가 많아 지게되어 논술과 면접 등이 중요합격 요인으로써 비중이 커지는 경향들이 나타나고 있다. 그러므로 학력평가에 있어서 이 두 가지의 중요한 측도를 효과적으로 반영시키기 위한 기술적인 조정이 요구되고 있는바 그러기 위해서는 문항의 난이도와 변별력 계수의 조정 및 문항 분포형의 추정 등을 통하여 수험자의 능력 변별 판정에 최적인 문항출제 기준을 수립하는 것이 중요한 과제라 할 수 있다. 이런 관점에 비추어 본 연구에서는 앞에서 논한 최적인 문항정답율을 산정하기 위한 실험과제로써 우리 나라 수학교과에서 상호 동형인 2회의 고사를 시행 할 경우 그 검사 문항의 난이도를 기초로 수험자들의 진점수(True Score)를 계산하고 난이도 계수의 이하인 원점수와 이상인 원점수를 분리하여 정답율의 변별력을 더욱 향상시킨 후 이 두 고사의 진점수들을 회귀방정식화 하여 문항의 난이도 조정을 통한 최적정답율의 산정 및 그 분포형을 추정함으로써 문항의 양호도 향상에 유효한 연구결과를 제시하고자 한다.[1)]

II. 연 구 내 용

1. 문항의 양호도 조정을 위한 진능력 점수 산정법

지금 2회에 걸쳐 시행한 동형고사에서 수험자들의 정답율에 대한 원점수를 각각 x, y 라

1) 이것은 진점수 자료를 얻기 위한 선행연구 내용에 따라 자료처리 법에 대한 기술을 한 것으로 일반적인 진능력 점수의 자료가 있으면 반드시 문항을 분리할 필요가 없음.

하고 그들의 진능력이 반영된 점수를 x^*, y^*로 표현하면 그 두 평가치 들은 모두 동형고사로 평가된 것이므로 각각 동형이든지 아니면 그 평가치들에 적절한 수학적 변환을 시행하여 각각의 두 평가치간의 차이를 같은 능력을 표현하는 일정한 상수로 표시할 수 있을 것이다. 즉 다시 말해서 제1회에 시행한 고사에서 관찰된 원점수의 정답율은 제2회에 시행한 고사에서도 역시 그와 동형인 정답율을 나타낼 것이므로 원점수 x, y와 진능력 점수 x^*, y^*의 편차를 $e_x = x - x^*$, $e_y = y - y^*$라 하면 그들은 모두 오차로 정의되고 이 오차들은 진능력점수의 변별을 어렵게 하는 요인들이므로 그러한 오차들을 최소화시키고 수험자들의 진능력을 최적으로 변별하기 위한 계산법이 개발되어지고 있는데 최근에 교육통계학의 회귀해석 분야에서 연구개발되고 있는 여러 종류의 통계적 적합법들이 오늘날 현대 문항 반응 분포론에 주요 응용분야의 연구대상으로 주목받고 있다. 이 분야의 연구는 Levine R.(1955)이 비동형인 수험자 군에 각각 동형인 문항을 출제한 평가고사에서 관측된 원점수로써 그들의 진능력 점수를 회귀방정식화에 의하여 최적인 점수를 계산하는 법을 처음으로 제시한 것을 시작으로 최근에는 Kolen M. J.와 Brennan R. L.(1987) 등이 2회의 동형고사 평가치로써 진능력 점수를 계산하는 경우 회귀함수를 이용한 적합법의 연구와 동류(Congeneric) 분산에 의한 분석법으로써 수험자의 원점수로 다루는 외부적 판별력 판정(External Anchor)에 의한 계산과 Bennet R. E. etal(1991)이 제시한 잠재력이 반영된 진능력점수로 다루는 내부적 판별력 판정(Internal Anchor)에 의한 계산법을 다루고 있는데 본 연구에서는 그 두가지 판정법을 동시에 처리할 수 있는 2변량오차의 회귀모형을 가정하여 그 모형에서 모수들을 추정하고 또 원점수와 진능력점수의 각 회귀모형을 각각 독립인 1변량 오차의 경우로 분리하여 그 회귀계수 및 모수들을 추정한 다음 2변량 오차의 회귀모형에서 추정한 추정치들과의 관계를 규명하고 또 각 표본 평균치(원점수 또는 진능력점수)와의 차이를 계산한 후 그 차이를 중심으로 추정된 회귀자료들을 평행이동 시켜 새로운 진능력점수의 문항정답율을 산정함으로써 문항의 양호도를 향상시키는데 조금이나마 도움이 되는 연구내용을 제시하고자 한다. 이러한 연구들은 대개 여러 학자들이 모두 다변량 해석이론으로 난해한 이론적 표현과 통계적 해석법을 제시하고 있으나 본 연구에서는 그러한 어려운 이론적 계산법을 제시하는 것 보다 누구나 수학교육 현장에서 용이하게 이용할 수 있는 쉬운 이론 및 계산법을 보이고져 단순회귀직선모형의 경우로 한정하여 다음의 주 연구 내용으로 전개하였다.

2. 주연구 내용

Longford N. T.(1995)는 전절에서 논한 2변량 오차 e_x와 e_y를 가지는 다변량 회귀모형에서 회귀계수 및 모평균벡터와 모분산, 공분산행렬을 추정하기 위해서 다음과 같은 계산과정과 단순회귀직선의 경우 [가정 1]과 [가정 2] 및 [정리 1]의 증명을 제시하고 있는데 그 계산 과정과 증명을 요약하면 다음과 같다.

지금 원점수 $(x, y)^T$ (단, T는 전치행렬)가 다변량 정규분포 $N(\mu, \sum)$에 따른다 할 때 원점수 $\{x_i, y_i\}$, $i=1, 2, \cdots, I$.(단 I는 문항수)의 우도함수는

$$l = -\frac{I}{2}\left\{I\log(2\pi) + I det\sum + \sum_{i=1}^{I} e_i^T \sum^{-1} e_i\right\} \cdots \quad (1)$$

이다. 단 $e_i = \{(x_i, y_i)^T - \mu\}$ 여기서 최우도법에 의하여 모수 θ를 추정하고져 할 때 Fisher Scoring 알고리즘을 이용하여 다음의 계산과정을 실행한다.

가) $h = \frac{\partial l}{\partial \theta}$

$$H = -\frac{\partial^2 l}{(\partial\theta\partial\theta^T)} \quad \text{단, } H\text{는 정보행렬}$$

$$\hat{\theta}_{\text{신평가치}} = \hat{\theta}_{\text{구평가치}} + H^{-1}_{\text{구평가치}}\, h_{\text{구평가치}}$$

나) 계산과정 가)에 의하여 H^{-1}구평가치 h 구평가치의 값들이 점점 작아지게 되면 $\widehat{\theta}$ 신평가치와 $\widehat{\theta}$ 구평가치는 같은 값으로 수렴하게 된다.

다) $\frac{\partial l}{\partial \theta} = -\frac{I}{2} t_r\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \right) + \frac{1}{2} t_r\left(\Sigma^{-1} S \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \right) + \sum_{i=1}^{I} e_i \Sigma^{-1} \frac{\partial e_i}{\partial \theta}$ … (2)

(단, $S = e_1 e_1^T + \cdots + e_I e_I^T$는 2변량 오차벡터의 직적을 표시함)

$$\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} = \frac{I}{2} t_r\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_1} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2} \right) - t_r\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_1} \Sigma^{-1} S \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2} \right) - \sum_{i=1}^{I} e_i^T \sum_{i=1}^{I} \left(\frac{\partial e_i}{\partial \theta_i} \right)^T \Sigma^{-1} \frac{\partial e_i}{\partial \theta_2} - \Sigma^{-1} \frac{\partial^2 e_i}{\partial \theta_1 \partial \theta_2} \quad \cdots (3)$$

이고 이 식에서 회귀직선의 경우로 가정하면 분산의 모수들은 오차 e_i 내에서는

$\frac{\partial^2 e_i}{\partial \theta_1 \partial \theta_2} = 0$, $\frac{\partial e_i}{\partial \theta} = -\frac{\partial \mu}{\partial \theta}$ 이므로 정보행렬은

$$-E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} \right) = \frac{I}{2} t_r\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_1} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2} \right) + I \frac{\partial \mu^T}{\partial \theta_1} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_2} \quad \cdots (4)$$

로 된다. 이 식은 (3)식보다 약간 간단하게 표현된 식이다. 이 다변량 모형에서 2변량 오차 $e_x = x - x^*$와 $e_y = y - y^*$를 가진 $y^* = a + bx^* + \delta$ (단 δ는 $N(0.\ \sigma_R^2)$에 따름)의 경우를 가정하고 그 모수 추정량을 계산하였는데 이 때 추정하고져하는 모수 θ에 관하여 분산 및 공분산 행렬 Σ을 각각 편미분하여

$$\frac{\partial \Sigma}{\partial a} = 0, \quad \frac{\partial \Sigma}{\partial \mu_x} = 0$$

$$\frac{\partial \Sigma}{\partial b} = \sigma_x^2 \begin{pmatrix} 0 & 1 \\ 1 & 2b \end{pmatrix}, \quad \frac{\partial \Sigma}{\partial \sigma_x^2} = \begin{pmatrix} 1 & b \\ b & b^2 \end{pmatrix}$$

$\frac{\partial \Sigma}{\partial \sigma_R^2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$을 구하고

또, 회귀계수에 관하여 평균벡터 μ를 편미분하여

$\frac{\partial \mu}{\partial a} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\frac{\partial \mu}{\partial b} = \begin{pmatrix} 0 \\ \mu_x \end{pmatrix}$, $\frac{\partial \mu}{\partial \mu_x} = \begin{pmatrix} 1 \\ b \end{pmatrix}$를 구하였다. 여기서 평균벡터의 추정양은

$y^* = a + bx^* + \delta$ 인 모형에서

$$x = x^* + e_x, \quad e_x \sim N(0,\ \sigma_{e_x}^2)$$

$$y = y^* + e_y, \quad e_y \sim N(0,\ \sigma_{e_y}^2)$$

$$\delta \sim N(0,\ \sigma_R^2)$$

이므로 $E(y^*) = E[a + bx^* + \delta]$ 는 $E(y - e_y) = E[a + b(x - e_x) + \delta]$로써 $\mu_y = a + b\mu_x$ 이다.

그러므로 진능력 점수는 위의 계산 결과에 따라 $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$는 $\left[\begin{pmatrix} \mu_x \\ a + b\mu_x \end{pmatrix} \begin{pmatrix} \sigma_x^2 & b\,\sigma_x^2 \\ b\,\sigma_x^2 & b^2 \sigma_x^2 + \sigma_R^2 \end{pmatrix} \right]$에 따름을 알 수 있고 다음 원점수의 경우도 위의 결과를 기초로 하여 간단히 증명할 수 있는데 즉

$x = x^* + e_x$에서

$Var(x) = Var(x^*) + Var(e_x) = \sigma_x^2 + \sigma_{e_x}^2$ 이고

$y^* = a + bx^* + \delta$에서

$Var(y^*) = Var(bx^*) + Var(\delta)$은

$\sigma_y^2 - \sigma_{e_y}^2 = b^2 \sigma_x^2 + \sigma_R^2$로 되고

따라서 $\sigma_y^2 = b^2 \sigma_x^2 + \sigma_{e_y}^2 + \sigma_R^2$이다.

같은 방법으로 $Cov(x,\ y) = b\,\sigma_x^2$도 구할 수 있으므로 원점수의 분산 공분산 행렬은

$$\begin{pmatrix} \sigma_x^2 + \sigma_{e_x}^2 & b\,\sigma_x^2 \\ b\,\sigma_x^2 & b^2 \sigma_x^2 + \sigma_R^2 + \sigma_{e_y}^2 \end{pmatrix}$$

이고, 여기서 진능력 점수의 회귀직선에서 회귀계수의 추정량은

$$\widehat{b} = \frac{b\,\sigma_x^2}{\sigma_x^2} = \frac{Cov(x^*,\ y^*)}{Var(x^*)} = \frac{\sum_{i=1}^{I} (x_i^* - \mu_{x^*})(y_i^* - \mu_{y^*})}{\sum_{i=1}^{I} (x_i^* - \mu_{x^*})^2}$$

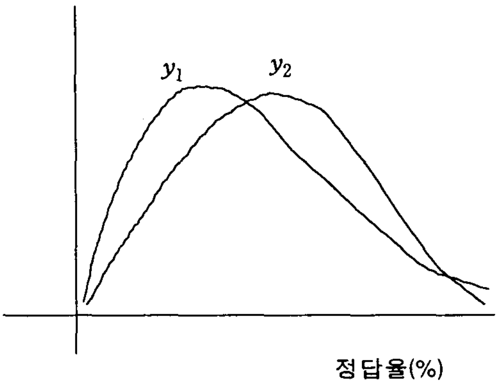
$\widehat{a} = \mu_y - \widehat{b}\mu_x$ 이며

원점수의 평균벡터의 경우도 마찬가지로 $\mu_y = a + b\mu_x$이므로 원점수 $\begin{pmatrix} x \\ y \end{pmatrix}$는

$$N\left[\begin{pmatrix} \mu_x \\ a+b\mu_x \end{pmatrix}\begin{pmatrix} \sigma_x{}^2+\sigma_{e_x}{}^2 & b\,\sigma_x{}^2 \\ b\,\sigma_x{}^2 & b^2\,\sigma_x{}^2+\sigma_R{}^2+\sigma_{e_y}{}^2 \end{pmatrix}\right]$$

에 따름을 증명할 수 있으므로 이러한 증명을 기초로하여 본 연구 내용의 주요 가정들과 정리를 제시할 수 있다.

[가정 1] 2회의 동형고사에서 원점수(Raw score)와 진능력 점수(True score)간에는 $x = x^* + e_x$, $y = y^* + e_y$ 일 때 각각의 측정오차 e_x와 e_y가 존재한다는 가정 하에서 원점수와 진능력 점수의 자료들을 각각 $y = a' + b'x + e$와 $y^* = a + bx^* + \delta$인 선형모형에 적합 시킨다.



<그림 1> 문항 y_1과 y_2의 분포형에 의한 난이도 비교

[가정 2] [가정 1]에서 진능력 점수의 오차 δ는 $N(0, \sigma_R{}^2)$에 따른다

[정리 1] 원점수 $\begin{pmatrix} x \\ y \end{pmatrix}$는

$$N\left[\begin{pmatrix} \mu_x \\ a+b\mu_x \end{pmatrix}\begin{pmatrix} \sigma_x^2+\sigma_{e_x}{}^2 & b\sigma_x^2 \\ b\sigma_x^2 & b^2\sigma_x^2+\sigma_R^2+\sigma_{e_y}{}^2 \end{pmatrix}\right]$$

에 따르고 진능력 점수 $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$는

$$\left[\begin{pmatrix} \mu_x \\ a+b\mu_x \end{pmatrix}\begin{pmatrix} \sigma_x^2 & b\sigma_x^2 \\ b\sigma_x^2 & b^2\sigma_x^2+\sigma_R^2 \end{pmatrix}\right] \quad \cdots(5)$$ [2)]

에 따른다

(단, $\sigma_{e_x}{}^2$, $\sigma_{e_y}{}^2$는 원점수 x, y의 각 측정오차들의 분산. $\sigma_x{}^2$은 x의 표본분산. $\sigma_R{}^2$은 진능력회귀직선의 오차분산.)

지금 위에서 가정한 [정리 1]의 내용을 자세히 관찰하면 2회의 동형고사 점수들(원점수 또는 진능력 점수)의 평균벡터는 모두 동일하게 원점수의 추정치인 μ_x와 진능력 점수의 회귀계수 a와 b로써 이루어진 $\begin{pmatrix} \mu_x \\ a+b\mu_x \end{pmatrix}$로 표기되어 있고 각 점수들의 분산 공분산 행렬은 각 오차들의 분산의 크기를 포함한 각각 다른 모양의 행렬로 표기되고 있는데 이때 평균 벡터 $\begin{pmatrix} \mu_x \\ a+b\mu_x \end{pmatrix}$를 중심으로 표본에서 얻은 두 동형고사 점수들의 표본평균들과의 차이를 계산(즉, 너무 높게 평가된 문항 또는 낮게 평가된 문항들의 점수를 고려하여)한 다음 그 문항의 점수 차이만큼 x축과 y축으로 이동하여 점수의 분포를 조정하고 또 동형고사 점수들의 분산 공분산 행렬의 각 원소들은 문항점수 분포의 분산을 표시하고 있으므로 분산들을 각 문항들의 변별력 크기를 측정할 수 있는 추정치로 간주하여 문항점수의 조정과 변별력 크기의 효과적인 측정을 통하여 최적의 문항분포를 추정하고 그 문항분포를 기준으로 최적의 문항정답율을 계산 하고자 한다.

일반적으로 문항의 분포형은 문항의 종류에 따라 여러 가지 형의 분포형을 예상할 수 있고 또 문항의 출제 형식 즉 주관식, 객관식, 규준지향평가, 목표지향평가 등에 따라 그 분포가 좌 또는 우 비대칭형일 경우가 많으나 여기서

2) 이 논문에서 추정치의 표기를 모집단 기호로 쓰고 있는 것은 여러 종류의 오차가 많으므로 그 구분이 편리하도록 하기 위한 것임.

는 출제자가 거의 정규분포에 근접하는 문항의 난이도를 고려하여 문항을 출제한 것으로 전제하여 다음을 가정한다.

[가정 3] 문항의 변별력을 측정하기 위한 출제기준인 난이도는 근사적으로 정규 분포형에 따른다.

다음에는 Bennet R. E. et al(1991)가 제시한 타당도(validity)의 크기를 측정하기 위한 추정법, 즉 2회의 동형고사를 시행하여 얻은 원점수와 진능력 점수에는 모두 수험자들의 개개인의 동일한 능력의 크기(Unidimensionality)가 반영되어져 있을 것으로 가정하고(동일한 교과내용 범위에서 양호한 문항의 평가고사가 제작된 경우 그 평가시행의 결과 진능력의 크기는 불변인 것으로 가정할 경우) 수험자들의 진능력의 크기를 변별해 내는데 유효한 정의를 다음과 같이 기술할 수 있다.

[정의] 고사 출제자가 지정된 출제 영역 내에서 바람직한 평가고사를 수험자들에게 시행하여 그 결과로 얻은 진능력 점수를 x^*와 y^*라 하면 수험자 개개인의 잠재력의 크기(Trait Score)를 $t = y' = \frac{y^* - a}{b}$ … (6)로 정의하고 그 크기를 재평가 할 수 있다.

<표 1> 원점수 x, y와 진능력점수 x^*, y^*

문항번호	Z_1	Z_2	Z_3	x	y	x^*	y^*
1	1	9.48	33.43	9.48	52.22	17.6988	40.7121
2	1	19.78	78.06	19.78	46.73	28.2825	32.8043
3	1	21.72	36.05	21.72	51.60	30.2760	39.8192
4	1	4.55	24.27	4.55	50.80	12.6330	38.6668
5	1	40.30	31.62	40.30	57.52	49.3677	48.3465
6	1	12.02	45.11	12.02	48.82	20.3088	35.8148
7	1	15.55	43.12	15.55	53.95	23.9360	43.2042
8	1	41.20	39.93	41.20	62.75	50.2925	55.8799
9	1	41.92	33.47	41.92	74.40	51.0323	72.6609
10	1	23.20	50.12	23.20	46.04	31.7967	32.8104
11	1	39.30	27.95	39.30	73.30	48.3402	71.0764
12	1	35.59	40.41	35.59	65.36	44.5280	59.6395
13	1	33.02	42.04	33.02	54.26	41.8872	43.6507
14	1	52.22	49.49	36.82	78.06	45.7919	77.7329
15	1	36.82	54.61	32.81	45.11	41.6714	30.4728
16	1	32.81	53.50	35.71	43.12	44.6513	27.6043
17	1	46.73	42.60	40.77	50.12	49.8507	37.6873
18	1	51.60	30.92	32.26	42.04	41.1063	26.0487
19	1	35.71	41.69	30.26	49.49	39.2772	36.7799
20	1	50.80	79.59	18.09	54.61	26.5460	44.1549
21	1	57.52	34.54	33.43	53.50	42.3085	42.5560
22	1	48.82	56.76	40.05	42.60	55.2761	26.8553
23	1	53.95	61.70	24.27	79.59	32.8962	80.1367
24	1	62.75	34.32	31.62	56.76	40.4486	47.2568
25	1	40.77	59.39	39.93	61.70	48.9875	54.3675
26	1	32.26	63.88	33.47	59.39	42.3496	51.0401
27	1	74.40	53.99	27.95	63.88	36.6775	57.5076
28	1	46.04	62.42	40.14	53.99	49.2033	43.2618
29	1	30.48	28.90	30.92	62.42	39.7294	55.4046
30	1	18.09	75.50	41.69	75.50	50.7960	74.2454
31	1	73.30	63.02	34.54	63.02	43.4491	56.2689
32	1	65.36	66.31	34.32	66.31	43.2230	61.0079
33	1	54.26	64.04	28.90	64.04	37.6537	57.7381

이때 $y' \equiv t$ 라는 가정 하에서 같은 방법으로 원점수에 의하여 변별된 잠재력의 크기를 $y^{+} = \frac{y-a'}{b'}$ … (7)로 정의하면 y와 t의 평균제곱 오차(Mean squared Error)는 $E(y^{+}-t)^2 = \frac{\sigma_{e_y}^{\ 2}}{b^2}$ 로 계산되어진다.

Ⅲ. 계산 예 및 검토

다음에는 홍석강(1994)의 연구[3] 결과에 수록된 자료로 원점수 x, y와 진능력 점수 x^*, y^* (위의 <표 1> 참조)로써 최적문항 정답율 산정과 문항의 난이도 조정을 통한 분포형의 변화를 관찰하고 그 경우 수험생 개개인의 잠재능력치(Trait Score)에 대한 재평가를 시행하기 위하여 Ⅱ절에서 논한 정의와 정리를 이용한 계산 예 및 그 자료 처리의 결과를 제시한 것이다. 이 자료는 어느 대학 입시평가 연구소에서 시행한 수학Ⅱ의 전국 정답율을 계산한 것으로 그 정답율의 난이도로써 그들의 수학문제 해결능력을 변별하기 위하여 그 평가치 Z_1(상수) Z_2(1회 고사) Z_3(2회 고사)의 기초 고사와 X(정답율 42%미만의 문항평가치) Y(정답율 42%이상의 문항평가치)로 1회 고사와 2회 고사에서 학력수준이 저득점인 문항평가치와 고득점인 문항평가치를 기초고사 $Z = (Z_1, Z_2, Z_3)$에서의 문제해결능력이 반영된 최종평가치인 진능력 점수 x^*, y^*를 계산한 것인데 여기서 문항정답율을 42%로 기준으로 한 것은 당시 대학입학학력고사에서 중위권 학생 수학Ⅱ 교과 성적의 평균 문항 정답율을 기초로 한 것이다. 이 진능력 점수 x^*, y^*를 계산하기 위한 연구내용에 관해서는 홍석강(1994)을 자세히 참고하기 바라며 특히 z_1(상수)를 1로 둔 것은 그 연구 내용에서 평균화시킨 교적행렬 계산식에서 필요로 하는 값으로 본 연구의 계산에서는 특별한 의미는 없는 것으로 이해하면 좋을 것이다.

이 자료로써 Ⅱ절의 [가정 1]에 의하여 원점수 x, y와 진능력 점수 x^*, y^*의 회귀직선의 그래프와 분산분석표를 각각 다음과 같이 얻었다. (<표2>, <표3> 참고) 지금 <표2>와 <표3>에서 나타난 바와 같이 원점수 x와 y, 진능력 점수 x^*, y^*의 사이에는 각각 회귀직선상에서 회귀계수 b'와 b가 모두 0이라는 가설을 검정하고자 할 때 F^1_{31}(원점수)=3.570, F^1_{31}(진능력 점수)=2.80으로써 즉, $F^1_{31}(0.90)$=2.80과 거의 같거나 큰 값을 보이고 있는데 여기서 각각의 회귀계수 b'와 b에 대하여 상관성의 유의성이 조금 작은 것으로 나타나고 있는 것은 문항의 원점수를 간에 정답율이 매우 큰 차이를 보이고 있고 그 자료들로써 회귀직선 경향이 크게 두드러지지 않기 때문인 것 같다. 그러나 불확실성을 많이 내포하고 있는 교육통계 자료들에서 이런 작은 유의성의 값으로나마 회귀직선의 경향을 가진 것으로 인정하여 확실성의 통계적 법칙을 유도하고자 하는 오늘날의 교육통계 연구 추세에 따라 10% 정도 크기의 유의성은 불만족스럽지만 유의성을 인정할 만한 값으로 두고 각 점수들 간에는 상관관계가 있는 것으로 인정하여 다음과 같이 Ⅱ절의 [가정 1]과 [가정 2]를 전제로 하고 [정리 1]을 적용하였다[4].

<표 4>는 원점수 x, y, 진능력 점수 x^*, y^*들의 표본평균과 표본분산, 또 원점수와 진능력 점수의 차인 각각의 오차 e_x와 e_y의 평균과 분산을 구한 것이다.

지금 Ⅱ절 [정리 1]에 의하여 원점수 x와 y는 각각 평균치

3) 우리 나라 교육통계 자료에서는 진능력 점수에 관한 자료가 별로 없으므로 부득이 본인의 연구 결과를 연구자료로 삼았음.

4) 이 내용에 대한 언급은 Longford N. T.(1995) 148페이지 참고할 것.

$$\begin{bmatrix} \mu_x \\ a+b\mu_x \end{bmatrix} = \begin{bmatrix} 30.47879 \\ 44.842574 \end{bmatrix}$$ 와 분산 및 공분산 행렬

$$\begin{bmatrix} \sigma_x^{\,2}+\sigma_{e_x}^{\,2} & b^2\,\sigma_x^{\,2} \\ b^2\,\sigma_x^{\,2} & b^2\,\sigma_x^{\,2}+\sigma_R^{\,2}+\sigma_{e_y}^{\,2} \end{bmatrix}$$

$$= \begin{bmatrix} 101.55085 & 41.444655 \\ 41.444655 & 248.18113 \end{bmatrix}$$ 인 정규분포에 따르고 진능력 점수 x^*와 y^*는 각각 평균치

$$\begin{bmatrix} \mu_x \\ a+b\mu_x \end{bmatrix} = \begin{bmatrix} 30.47879 \\ 44.842574 \end{bmatrix}$$ 와 분산 및 공분산 행렬

$$\begin{bmatrix} \sigma_x^2 & b\sigma_x^2 \\ b\sigma_x^2 & b^2\sigma_x^2+\sigma_R^2 \end{bmatrix} = \begin{bmatrix} 100.2214 & 41.444655 \\ 41.444655 & 226.80829 \end{bmatrix}$$

인 정규분포에 따른다. 이 정리에 의한 계산결과를 검토하면 앞의 <그림 2>의 결과에서 관찰한 바와 같이 원점수의 경우 표본평균은 $\mu_x = 30.47879$, $\mu_y = 57.66667$인데 $\mu_y = a + b\mu_x$는 회귀직선에 적합된 결과에 의하면 42%이상의 고득점의 문항의 경우 즉, 문항의 난이도가 쉬운 문항들의 평가치가 상당히 높게 평가되어 있으며(즉 쉬운 문항의 문항정답율이 너무 높게 출제되어 있음) μ_y와 [정리 1]에 의한 원점

<표 2> 원점수의 분산분석표

분산의 요인	자유도	평방합(SS)	평균평방(MS)	분신비	P-값
처 리	1	361.13984	364.13984	3.570	0.0682
오 차	31	3161.70649	101.99053		
전 체	32	3525.84633			

MSE = 101.99053

<표 3> 진능력 점수의 분산분석표

분산의 요인	자유도	평방합(SS)	평균평방(MS)	분신비	P-값
처 리	1	606.10729	606.10725	2.800	0.1043
오 차	31	6709.42883	216.43319		
전 체	32	7315.53608			

MSE = 216.43319

<표 4> 각 점수들의 표본평균과 표본분산 및 오차들의 평균과 분산

통계치 / 점수 및 오차	평 균	분 산
x	30.47879	100.2214
y	57.66667	110.1827
x^*	39.46282	110.7599
y^*	48.55777	228.6105
e_x	-8.98403	1.329453
e_y	9.1028	21.37284
ε	6.91702×10^{-15}	98.8033278
δ	$-9.52773\mathrm{E}\times 10^{-15}$	209.6696510

단, $\sigma_R^{\,2} = 209.6696510$

수 y의 추정치 $a+b\mu_x$를 비교하면 그 두 값 y 평가치의 차가

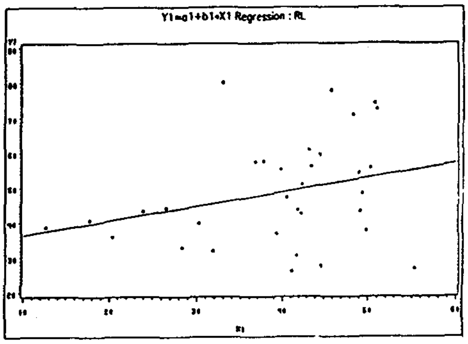
$$\mu_y-(a+b\mu_x)=57.66667-44.842574=12.824096$$

이므로 그 차를 참고하여 쉬운 문항의 정답율을 하향 조정할 필요가 있을 것이다. 또 분산과 회귀직선에 적합된 분산공분산 행렬에서 계산된 분산의 경우에는 표본분산 $\sigma_x^2=100.2214$, $\sigma_x^2+\sigma_{e_x}{}^2=101.550585$로써 두 분산의 차이는 매우 작으므로 즉 원점수 x 에서는 정답율 조정을 하지 않았으므로 $\sigma_{e_x}{}^2=1.329453$으로 난이도와 변별력에는 큰 차이가 없고 원점수 y의 경우 정답율을 하향조정 하면(즉 조금 어려운 문항을 출제하면) 표본분산은 $\sigma_y{}^2=110.1827$인데 [정리 1]에 의한 회귀직선에서 구한 원점수 y의 분산은 248.18113이므로 그 차이 즉 137.817만큼 분산의 크기가 커지게 되고 문항의 난이도는 높아지게 되며 문항 정답율 간의 변별의 크기도 크게되는 것을 알 수 있다. 마찬가지로 진능력 점수의 경우에도 [정리 1]에 의하면 x^*와 y^*는 각각 평균치

$$\begin{bmatrix} \mu_x \\ a+b\mu_x \end{bmatrix}=\begin{bmatrix} 30.47879 \\ 44.842574 \end{bmatrix}$$ 와 분산 및 공분산 행렬은

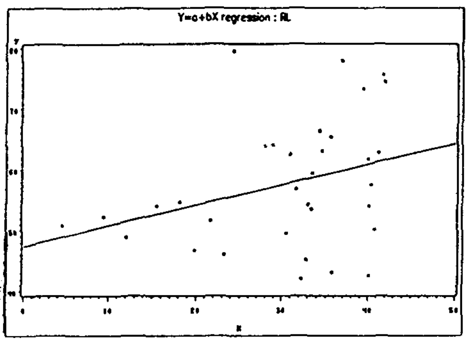
$$\begin{bmatrix} \sigma_x^2 & b\sigma_x^2 \\ b\sigma_x^2 & b^2\sigma_x^2+\sigma_R^2 \end{bmatrix}=\begin{bmatrix} 100.2214 & 41.444655 \\ 41.444655 & 226.80829 \end{bmatrix}$$

인 정규분포에 따르므로 <그림 3>의 결과에서 관찰한 바와 같이 회귀직선 추정치인 x^*의 평균치는 $\mu_x=30.47879$인데 반하여 x^*의 표본평균은 $\mu_{x^*}=39.46282$이며 그 차이는 8.98403이므로 진능력 점수 x^*는 그 차이만큼 하향조정 하면 좋겠고 또 진능력 점수 y^*를 회귀직선에 적합시킨 경우를 참고하면 y^*의 표본평균치는 48.55777인데 $\mu_y=a+b\mu_x=44.842574$이므로 이 두 값의 차이는 3.715196으로써 진능력 점수 y^*의 경우에도 그 차이만큼 문항정답율을 하향조정하면 좋을 것이다. 다음 분산의 경우들을 검토해보면 진능력 점수 x^*의 경우 표본분산은 $\sigma_{x^*}{}^2=110.7599$, 회귀직선에서 구한 분산은 100.2214로써 그 차이는 -10.5385이고 진능력 점수 y^*의 경우 표본분산은 $\sigma_{y^*}{}^2=228.6105$, 회귀직선에서 구한 분산은 226.80829로 그 차이는 -1.80221로써 각 분산 크기의 차들이 약간 작아진 것으로 보아 진능력 점수 x^*, y^*사이에서는, 회귀직선에 적합시킨 결과 원래의 진능력 점수들보다 수험자들의 진능력이 반영된 크기를 측정하고저 할 때 그 분산의 차이만큼(즉 2회의 동형고사 평가치 x^*와 y^*에서 문항정답율의 분포보다 조금 더 첨예하게 되어질 가능성이 있음) 더 조정 할 필요가 있음을 보이고 있다. 즉 다시 말해서 2회째 시행한 동형고사에서 보다 수



<그림 2> 원점수 x, y 회귀직선의 그래프

$$y=47.396519+0.336960x$$

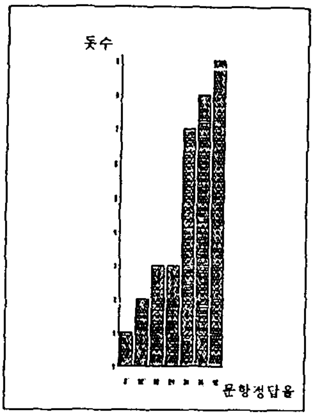


<그림 3> 진능력 점수 x^*, y^*의 회귀직선

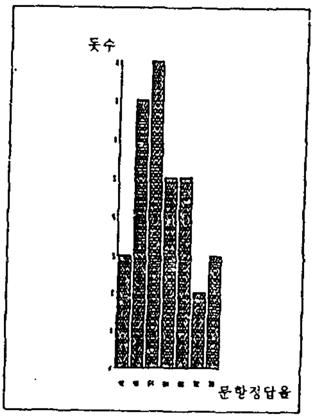
$$y^*=32.238650+0.414531x^*$$

험자들의 진능력이 잘 반영될 수 있도록 양호한 문항이 출제되었더라면 더욱 바람직했을 것임을 뜻하고 있다.

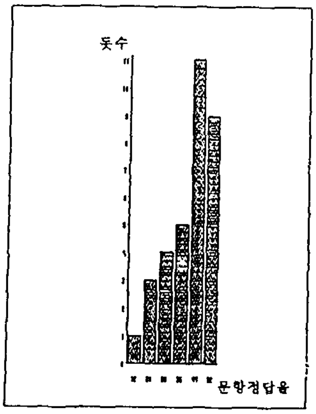
다음 <그림 4>에서는 원점수 x와 y, 진능력 점수 x^*, y^*를 히스토그램으로 그려 그 분포형을 도시하였는데 이 그래프를 관찰하면 문항정답율을 42%를 기준으로 하여 재편하여 그린 결과로 x와 x^*, y와 y^*의 막대그래프가 모두 유사한 분포를 형성하고 있으나 진능력의 점수의 경우를 보면 x^*와 y^*의 분포는 x와 y의 분포보다 문항정답율이 상당히 높게 도시되어 있다. 이제 최종적으로 출제하여야 할 최적인 진능력점수의 문항정답율을 산정하기 위해서는 앞에서 지적한데로 진능력 점수 x^*와 y^*의 표



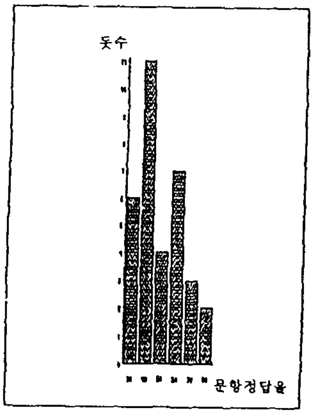
x의 막대분포형



y의 막대분포형



x^*의 막대분포형



y^*의 막대분포형

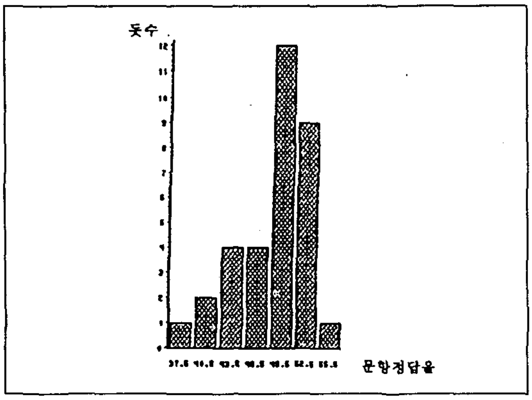
<그림 4> 원점수와 진능력 점수의 분포형

본의 평균치와 회귀직선에 의한 평균치의 추정값과의 차의 값을 고려하고 진능력 점수의 오차들을 최소화하기 위하여 x^*와 y^*의 최적의 회귀직선을 구하고 x^*가 대입된 경우 새로운 회귀직선에 의한 추정치 값 y'^*을 계산해야 하므로 그렇게 하기 위하여 앞 절의 [정리 1]을 이용하고 계산한 결과를 <표 5>에 수록하였으며 그 값을 <그림 5>와 같이 히스토그램으로 그 분포형을 도시하였다.

지금 <그림 5>에서 진능력 점수 y'^*의 분포

<표 5> 진능력 점수의 최적문항정답율

문 항	y'^*	문 항	y'^*	문 항	y'^*
1	39.5577	12	50.6524	23	45.8423
2	43.9344	13	49.5603	24	48.9654
3	44.7587	14	51.1750	25	52.4965
4	37.4628	15	49.4711	26	49.7515
5	52.6537	16	50.7034	27	47.4059
6	40.6370	17	52.8535	28	52.5858
7	42.1369	18	49.2374	29	48.6680
8	53.0362	19	48.4810	30	53.2444
9	53.3421	20	43.2163	31	50.2062
10	45.3876	21	49.7346	32	50.1127
11	52.2288	22	55.0971	33	47.8096



<그림 5> 최적문항정답율 y'^*의 분포형

를 도시할 때 고려되어야 할 것은 그 자료들이 평행이동 되어 계산되어져야 한다는 것인데 즉 x^*의 표본평균치와 μ_x의 차이 즉 8.98403과 y^*의 표본평균치와 $\mu_y = a + b\mu_x$의 차이 즉 3.715196을 중심으로 x^*축과 y^*축에 대략 음의 방향으로 그 차이만큼 평행이동 한 후 그 자료로써 새로운 문항정답율 y'^*를 구하고 그 분포를 도시하여 보면 그 그래프는 근사적으로 정규분포에 가까운 형으로 수험자들의 정답율 반응분포를 나타내고 있음을 알 수 있다. 따라서 이 자료들이 최적인 문항정답율이며 그 분포가 문항의 양호도가 향상된 분포형이고 수험자들의 문항반응율을 평가자의 목표에 따라 기술할 수 있으므로 평가자들은 이 문항정답율을 최종 평가출제의 기준으로 삼을 수 있을 것이다.

끝으로 동일한 출제영역 내에서 2회의 동형고사 시행을 통하여 수험자들 개개인의 진능력, 즉 잠재력의 크기는 불변이라는 가정 위에 각 수험자들이 그 고사 영역 내에서 어느 정도 그들의 진능력을 발휘하고 있는지를 출제자가 추정하고져 하는 경우 그 잠재력을 재평가하기 위하여 II절의 정의를 이용할 수 있고 식(6)에 의하여 자료 처리된 계산결과로써 다음 <표 6>을 얻었다.

<표 6> 잠재력의 크기 $y' = \dfrac{y^* - a}{b}$

문 항	y'	문 항	y'	문 항	y'
1	20.491	12	66.261	23	115.827
2	1.368	13	27.597	24	36.305
3	18.331	14	110.498	25	53.512
4	15.545	15	-4.275	26	45.466
5	38.952	16	-11.207	27	61.105
6	8.648	17	13.176	28	26.656
7	26.517	18	-14.969	29	56.020
8	57.169	19	10.982	30	101.581
9	97.749	20	28.816	31	58.110
10	-1.036	21	24.949	32	69.570
11	93.917	22	-13.018	33	61.663

이 결과에서 보는 바와 같이 잠재력 크기의 추정치 $t = y' = \frac{y^* - a}{b}$도 역시 회귀계수 a와 b를 중심으로 y^*를 일차변환시킨 결과로써 이 추정치는 y^*와는 전혀 다른 새로운 scale을 나타내고 있음을 <표 5>와 <표 6>으로 비교하여 보면 이해할 수 있는데 이 자료에서 y'을 살펴보면 예를 들어 10번, 15번, 16번, 18번, 22번의 문항의 경우는 y'의 값이 모두 음수로 표시되어져 있고 그 값들은 그런 문항들이 모두 잠재력의 크기에 대한 변별이 약한 문항 즉 대단히 어려운 문항임을 의미하고 있다. 다음 9번, 11번, 14번, 23번, 30번 등의 문항은 모두 잠재력의 반영이 강한 문항 즉 수험자들이 모두 문제를 잘 해결할 수 있는 쉬운 문항임을 보이고 있으며 다시 원점수 y로 $y^+ = \frac{y - a'}{b'}$로 변환시킨 값으로 원점수의 잠재력 크기를 추정하고 그 y^+와 y'의 평균제곱오차인 $E(y^+ - t)^2 = 124.98166$의 제곱근 11.175119로 새로운 잠재력의 크기에 대한 편차로써의 한 표준단위로 삼을 수 있는데 이 측도는 원점수와 진능력 점수와의 잠재력 크기의 차이를 기술할 때 각 분포간의 잠재력 크기 차이의 평균치와 더불어 편차로써 표준단위로 삼을 수 있으므로 잠재력 크기 차의 분포의 추정에 유익한 측정 단위로 간주된다.

Ⅳ. 결 론

본고에서는 수학교과에서 2회의 동형고사 $Z = (Z_1, Z_2, Z_3)$와 고학력 능력테스트 X를 시행한 후 그 X 문항들은 저득점 문항 x, 고득점 문항 y로 나누어 원점수 x, y로 두고 기초고사의 능력이 반영된 2회의 고학력 능력테스트 Y 문항도 저득점 문항 x^*, 고득점문항 y^*인 진능력 점수로써 그 평가치들을 회귀직선모형에 적합시켜 문항의 난이도를 조정하기 위하여 최적문항정답률을 산정하고 변별력의 크기를 측정함으로써 문항의 양호도 향상에 유효한 통계적 해석법을 논하였는데 이 논문의 주요 연구결과와 기대효과 및 활용 방안을 요약하면 다음과 같다.

1. 문항의 난이도를 조정하기 위하여 원점수와 진능력점수를 회귀직선모형에 적합시킬 때 2변량 오차를 가진 회귀모형의 경우 모수를 추정할 수 있는 통계적 해석법을 Ⅱ2, 주연구 내용에서 기술하였다.
2. 문항의 난이도 조정을 통하여 문항의 양호도를 향상시키는 과정을 보이기 위해서 문항정답률의 자료가 비대칭인 분포에서 정규분포화하는 과정을 기술하고 그 분포형의 변화를 예로써 도시하였다.
3. 문항 정답률의 크기를 조정하기 위하여 [가정1], [가정 2], [가정 3]에서 세운 가정을 전제로 하여 진능력 점수의 회귀직선을 구하고 그 자료의 평균치들 차이만큼 근사적으로 X^*축과 Y^*축으로 평행이동함으로써 최적문항 정답율 산정에 유효한 진능력 점수의 변환법과 그 계산 예들을 Ⅲ절에서 제시하였다.
4. 잠재력의 크기를 측정할 수 있는 잠재력 점수라는 새로운 점수를 소개하고 두 평가고사에서 잠재력 점수차의 편차를 측정단위로 하여 그 잠재력 점수들 차이의 분포를 정의하는 계산 예를 Ⅲ절에서 보였다.
5. 위의 결과로써 최적문항정답률을 산정하여 변별력의 크기와의 관계를 규명하고 문항의 양호도 향상에 유효한 최적의 문항 분포형을 도시함으로써 문항출제기법의 개선에 유익한 결과를 제시하였으며 이 연구 결과의 개선을 위하여 다음과 같은 점들이 더 연구되어지면 좋을 것이다.

첫째, 이 연구자료에서 추출한 문항의 크기가 작고 각 학력평가연구소의 출제형식에 따라 약간 상이한 결과가 나올 것이 예상되므로 그 출제형식의 과학화 및 다양화로 각 문항

의 정답율이 보다 정규분포형에 수렴하는 출제에 관한 연구가 있어야 할 것이다.

둘째, 문항 배점의 세분화와 평가결과의 산포도를 크게 함으로써 수험생의 개인차를 엄격히 변별할 수 있는 능력심화판별 검정을 위한 평가자료를 출제하는 연구가 더 있어야하며

셋째, 위의 연구 결과에서 제시한 진능력 점수의 평행이동법은 회귀직선의 회귀계수들을 중심으로 하고 또 문항분포를 반분한 경우이므로 이 경우 원점수와 진능력 점수의 평균치간의 차이를 동시에 고려한 복합 변수의 평행 이동을 시행하는 연구가 더 시도되어져야 할 것이다.

결론적으로 이 연구 결과를 활용하기 위한 제언으로써 우리 평가자들은 어떤 평가를 시행하기 전에 먼저 시험적인 예비고사를 수회 시행하여 수험생들의 잠재력 점수의 크기를 정의하고 그것에 기초한 최종문항 정답율과 변별력의 크기 및 심화의 정도를 추정한 후 평가자들이 구하고자하는 양호한 분포형을 도출하고 그 문항들을 양호도의 준거에 비추어 타당한지 여부와 동시에 그것이 평가자들의 목적한 바에 만족스러운 정도인가를 검정한 후 그 양호도를 기초로 최종 평가고사 출제를 위한 기준을 세우는 것이 바람직한 것으로 생각된다.

참 고 문 헌

홍석강 (1988). 신구두고사 평가치변환에 의한 진분포와 모수추정에 관한 연구, 한국수학교육학회지 시리즈A <수학교육>, 제29권 제2호 79-93.

홍석강 (1994). 결측치를 가진 목표지향형 평가모델에서 수학학습능력의 평가에 관한 연구, 한국수학교육학회지 시리즈A <수학교육>, 제33권 제2호 167-175.

Bennet, R. E. Rock D. A. & Wang, M. (1991). Equivalence of free response and multiple choice item. *Journ. of Educational Measurement*, 28, 77-92.

Kolen, M. J. & Brennan, R. L. (1987). Linear equating models for common item nonequivalent population design, *Applied Psychological Measurement* Vol 11, No 3, 263-277.

Kolen, M. J. & Brennan, R. L. (1995). *Test equating, methods and practices*, Springer, NY.

Levin, R. (1955). *Equating the score scale of alternative forms administered to sample of different ability* (Research Bulletin 55-23), Princeton, N. J. Educational. Testing Service.

Longford, N. T. (1995). *Models for uncertainty in educational testing*, Springer, NY.

Study on Estimating the Optimal Number-right Score in Two Equivalent Mathematics-test by Linear Score Equating

Hong, Suk-Kang

Dept. of Mathematics Education, Dong Guk University, 26, Pil-dong, 3-ka Choong-ku, Seoul, 100-715, Korea; email: skhong@kra.dongguk.ac.kr

In this paper, we have represented the efficient way how to enumerate the optimal number-right scores to adjust the item difficulty and to improve item discrimination.

To estimate the optimal number-right scores in two equivalent math-tests by linear score equating

a measurement error model was applied to the true scores observed from a pair of equivalent math-tests assumed to measure same trait.

The model specification for true scores which is represented by the bivariate model is a simple regression model to inference the optimal number-right scores and we assume again that the two simple regression lines of raw scores and true scores are independent each other in their error models.

We enumerated the difference between mean value of x^* and μ_x and the difference between the mean value of y^* and $a+b\mu_x$ by making an inference the estimates from 2 error variable regression model.

Furthermore, so as to distinguish from the original score points, the estimated number-right scores y'^* as the estimated regression values of true scores with the same coordinate were moved to center points that were composed of such difference values with result of such parallel score moving procedure as above mentioned. We got the asymptotically normal distribution in Figure 5 that was represented as the optimal distribution of the optimal number-right scores so that we could decide the optimal proportion of number-right score in each item.

Also by assumption that equivalence of two tests is closely connected to unidimensionality of a student's ability. we introduce new definition of trait score to evaluate such ability in each item.

In this study there are much limitations in getting the real true scores and in analyzing data of the bivariate error model. However, even with these limitations we believe that this study indicates that the estimation of optimal number right scores by using this enumeration procedure could be easily achieved.