韓國數學敎育學會誌 시리즈 E
<數學敎育 論文集>
제 13집, 2002. 2. 641-653.

J. Korea Soc. Math. Ed. Ser. E:
Communications of Mathematical Education
Vol. 13, Feb. 2002. 641-653.

수학교육 평가에서 고사 타당도의 측정에 관한 연구

홍 석 강 (동국대학교)

수학 교육평가에서 고사 타당도의 측정은 매우 중요하다. 그것은 평가자가 목적하는 바의 준거에 따라 수험자들의 문제해결력 변별을 위한 고사들 간의 타당도 크기 측정과 또는 고사 문항의 선정, 배치를 통하여 한 고사의 타당도를 높이기 위한 문항별 타당도 계수를 측정 계산하는 것이다. 이 연구에서는 이 두 가지 고사 타당도의 크기 측정법을 이용하여 수학교육 평가 현장에서 직접적으로 적용할 수 있는 고사 타당도의 제고에 유효한 연구 결과를 제시한다.

1. 서 론

타당도(Validity)란 한 평가도구 혹은 검사도구가 측정하고자 하는 것을 어느 정도 충실하게 측정하고 있는가에 관심을 갖는 개념이다. 즉 그것은 무엇을 재고 있는가의 개념 또는 그 증거가 검사점수에서 추론한 내용을 지적해 주는 정도를 의미 한다. 그러므로 여기서 타당화 되는 것은 검사자체가 아니고 검사를 특정 목적으로 사용하는데 따른 추론이 타당화 되는 것이다. AERA(1985)의 검사의 표준에 관한 편람에는 전통적으로 타당도의 증거를 모으는 수단으로써 세 가지 형태의 검사 타당도 즉 내용, 준거, 구인 타당도를 설명하고 있다. 이 표준서에 따르면 내용관련 타당도는 검사 문항의 배열에 요구되는 자료와 과제가 특정 영역이나 학문안에서 기능 혹은 지식 전집의 대표적인 표본이라는 것이 확립될 것을 필요로 함과 또 판단의 준거를 기초로 그 내용의 검사 문항이 주어진 영역의 내용과 기능을 적절하게 표집했는지 여부를 검정함으로써 타당도의 종류를 정의하고 있다. 그 결과로 내용 타당도, 예언 타당도, 공인 타당도, 구인 타당도를 정의하고 있다. 그러나 이들의 타당도를 간단히 요약하면 내용 타당도와 준거 타당도를 기초로 평가 목적의 취지와 기준에 대한 분석과 판단을 근본적으로 분석자의 전문적 소양과 해당 교과에 대한 교과관에 의존하여 평가의 구체적인 내용들이 평가하고자 하는 구체적인 목표 내용들을 대표하고 있는지에 대하여 두 종류 타당도의 복합적 기능을 잘 조화, 이용할 것을 강조하고 있다. 일반적으로 검사에서 신뢰도를 계산하는 일은 비교적 명확하고 기술적인데 반해 이 타당도는 그렇지 못하여 이것을 어떻게 수립할 수 있는가에 대한 노력은 검사이론에 이 타당도의 개념이 도입된 이래 지금까지 이어지고 있다. 이에 관한 연구의 시초로 Cronbach 와 Meehl(1955)이 심리학 검사에 타당도의 개념을 도입한 이래 지난 수십년 동안 이 분야에 수학적 접근이 많이 시도되어 왔으나 아직도 타당도의 개념적 및 방법론적 재 교육의 방향을 제시하지 못하고 불완전한 가운데서 수량화 및 수식화의 개념이 조금씩 발전하고 있는 중에 최근에 이 분야에 관한 연구에 관심이 크게 고조되고 있다.

그러므로 이 연구에서는 수학교육 평가 현장에서 검사 문항별로 타당도 계수를 쉽게 계산 할 수 있도록 Lord(1980)의 타당도 계수 계산법과 또 검사 문항을 평가자가 목적하는 바의 준거에 따라 즉 예를 들어 수험생들의 수학적 고등능력의 변별을 위해서 문제 출제법의 변동에 따른 문제해결력의 전이(Translation)과정에 따른 고사 타당도의 크기를 측정할 수 있는 Wiley(1991)의 연구 결과를 참고로 하여 그들의 이론을 우리 수학교육 평가에 응용하는 예를 보임으로써 평가 개선에 조금이나마 유익한 연구 결과를 제시하고자 한다.

2. 연구의 내용

지금 한 고사 전체 또는 수험자 능력의 타당도를 측정하고자 할 때 여러 학자들 Maguire, Hattie와 Brian(1994), Linn(1993)와 Messick(1995) 등이 제시한 방법들이 있으나 그 내용은 목표지향형 평가나 수행평가에서 타당도를 측정하는 것으로써 일반적인 문제해결력 검정이나 그 해결력의 전이과정에 따라 수험생의 능력을 변별하는데 있어서는 구체적으로 고사 타당도를 수리화 하는 모형을 제시하지 못하고 있으므로 여기서는 Wiley(1991)의 측정법을 이용하기로 하고 그 측정법의 전제와 모형을 다음과 같이 제시한다.

2.1. Wiley의 고사 타당도 측정법

1. 검사의 타당화를 위한 필요조건과 방법

1) 검사과정에서 평가 준거에 입각한 고사내용의 기술적인 구성 및 그 준거의 과정들로 분류된 내용들간에 상관의 크기를 측정 할수 있는 계산적이고 논리적인 네트워크(Nomological Network)를 배열할 것.

2) 평가 결과에서 나타난 수험자의 잠재적 재능(Latent Trait) 요인들간의 상관관계를 계산하면서 준거의 전이(Translation)에 의한 검사 내용의 패턴이나 프로필 등에서 현시된 재능도 탐지하여 수리화 할수 있는 공식을 형식화 하고

3) 1)과 2)에서 논한 두 종류의 내용 및 준거 타당도에서 변별된 잠재능력 분포(Latent Trait Distribution)를 도출함으로써 평가자가 목적하는 바의 타당도를 측정, 계산한다.

2. Wiley의 타당도 측정을 위한 정의 및 정리

여기서 고사 평가치로써 수험자들의 지정된 능력의 변환을 암시하는 모수추정치를 추정하고저 할 때 한 수험자의 고사 평가치 $y_k = \sum_{i=1}^{n} b_i q_i$, $k = 1, 2, \cdots, N$. (단, q_i는 한 수험자의 능력별 요인, b_i는 한 수험자의 능력별 가중치.)는 일반적으로 다차원 분포로 선형가법적 복합분포(Linear

Additive Model)를 형성한다. 그리고 한 고사에서 측정된 기술적(Technical) 또는 문제 해결력의 능력은 수험자 개인이나 또는 테스트 된 그룹에 관계없이 항상 같으므로 문항의 특성과 개인의 능력에 대한 특성으로 분리되어진다. 여기서 $y^* = \sum_{i=1}^{n} b_i^* g_{i\cdot}$ (단, b_i^*는 1차 고사에서 나타난 수험자들 능력의 가중치.)를 1차 고사에서 나타난 수험자들의 능력의 평가치라 할 때, 그 고사에서 나타난 수험자들의 문제해결력을 모수타당도 y^*로 두고 2차고사 y에 대하여 이 타당도 y^*와 차이인 비타당도 $y-y^*$를 고려하여 $y=y^*+(y-y^*)$로써 $\sigma_y^2 = \mathrm{Var}(y) = \mathrm{Var}(y^*) + \mathrm{Var}(y-y^*) + 2\,\mathrm{Cov}(y^*,\ y-y^*)$를 얻을 수 있고 이하 Wiley가 제시한 다음의 정리들을 요약 할 수 있다.

[정리1.] $\mu_y = \sum_{i=1}^{n} b_i \mu_i \quad \mu_{y-y^*} = \sum_{i=1}^{n} (b_i - b_i^*)\mu_i$

$$\sigma_y^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i b_j \mathrm{Cov}(q_i,\ q_j)$$

$$\sigma_{y^*}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i^*\ b_j^*\ \mathrm{Cov}(q_i,\ q_j)$$

$$\sigma_{yy^*} = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i\ b_j^*\ \mathrm{Cov}(q_i,\ q_j)$$

$$\sigma_{y-y^*}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} (b_i - b_i^*)(b_j - b_j^*)\mathrm{Cov}(q_i,\ q_j)$$

$$\mathrm{Cov}(y^*,\ y-y^*) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i^*(b_j - b_j^*)\mathrm{Cov}(q_i,\ q_j)$$

Wiley(1991)는 정리1과 그의 예제에서 수험자 능력별 평균치 μ_i를 모두 같은 값으로 두고 그 모수가중치 b_i^*를 각각 다른 값으로 계산하여 1차고사 평가치의 공분산을 계산하여 정리1을 이용하였으나, 일반적으로 수험자 능력별 평균치들은 능력별로 모두 다른 평균치들을 갖게되므로 그의 계산 과정은 특별한 예에 불과하다. 따라서 일반적인 예로 위의 정리를 풀어서 정리해보면 더욱 쉬운 표현으로써 다음의 정리들을 얻을 수 있다.[1)]

주1) 예로써 σ_y^2의 경우 Wiley의 가중평균치 표현으로

$$\sigma_y^2 = \sum\sum_{i \neq j} b_i b_j \mathrm{Cov}(q_i,\ q_j) = \sum b_i^2\ \sigma_{q_i}^2 + \sum\sum_{i \neq j} b_i b_j\ \mathrm{Cov}(q_i,\ q_j) = E\Big[\sum_{i=1}^{n} b_i(q_i - \overline{q_i})\Big]^2$$ 이고 한 개인의 수험자 평가치를 y_k, $k=1, 2, \cdots, N$라 하면 $\sigma_y^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k - \overline{y})^2$과 일치함을 쉽게 알수 있다.

[정리2.] $\mu_y = \bar{y} = \frac{1}{N}\sum_{j=1}^{N} y_j,\ \mu_{y-y^*} = \bar{y} - \overline{y^*}$

$$\sigma_y^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k - \bar{y})^2,\quad \sigma_{y^*}^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k^* - \overline{y^*})^2$$

$$\mathrm{Cov}(y,\ y^*) = \frac{1}{N}\sum\sum_{k\neq l}(y_k - \bar{y})(y_l^* - \overline{y^*})$$

$$\sigma_{y-y^*}^2 = \frac{1}{N}\sum_{k=1}^{N}[(y_k - y_k^*) - \mu_{(y-y^*)}]^2$$

$$\mathrm{Cov}(y^*,\ y-y^*) = \frac{1}{N}\sum\sum_{k\neq l}[(y_k^* - \overline{y^*})\{(y_l - y^*) - \mu_{y-y^*}\}]$$

[정리3.] 준거의 전이에 따른 두 고사 평가치의 상관계수

$$\rho_{yy^*} = \frac{\mathrm{Cov}(y,\ y^*)}{\sigma_y \sigma_{y^*}} \quad \cdots\cdots\cdots\cdots\cdots (1)$$

는 두 고사를 통한 수험자 능력의 변별에서 일치성의 측정에 관한 타당도 계수이다.

[정리4.] 1차 고사와 2차 고사의 평가치 $y*$, y로써 두 개의 회귀직선과 비타당도 회귀직선은 각각

$$E(y/y^*) = \alpha_0 + \alpha_1 y^* \quad \cdots\cdots\cdots\cdots\cdots (2)$$

단, $\alpha_1 = \frac{\mathrm{Cov}(y,\ y^*)}{\sigma_{y^*}^2}$, $\alpha_0 = \mu_y - \alpha_1 \mu_{y^*}$

$$E(y^*/y) = \gamma_0 + \gamma_1 y \quad \cdots\cdots\cdots\cdots\cdots (3)$$

단, $\gamma_1 = \frac{\mathrm{Cov}(y,\ y^*)}{\sigma_y^2}$, $\gamma_0 = \mu_{y^*} - \gamma_1 \mu_y$

$$E(y - y^*/y^*) = \beta_0 + \beta_1 y^* \quad \cdots\cdots\cdots\cdots\cdots (4)$$

단, $\beta_1 = \frac{\mathrm{Cov}(y - y^*,\ y^*)}{\sigma_{y^*}^2}$, $\beta_0 = \mu_{(y-y^*)} - \beta_1 \mu_{y^*}$

이다.

다음에는 한 고사 전체로써 한 수험자의 능력을 추정하는 모형과 달리 어떤 고사의 문항 하나하나를 단위로 하는 경우 Lord(1980)가 제시한 문항별 타당도 계수를 계산하는 방법을 이용할 수 있다.

2.2. Lord의 문항별 타당도 계수 계산법

<정의1.> 한 개의 문항을 g_i, $i=1, 2, \cdots, n$, 수험생을 a_i, $i=1, 2, \cdots, N$, 고사 점수를 X_{a_i}, 타당도 측정을 위한 준거를 v_i, $i=1, 2, \cdots, n$라 할 때, 타당도 계수는

$$\rho_{Xv} = \frac{\sum_g \sigma_y \rho_{gv}}{\sum_g \sigma_g \rho_{gX}} \quad \cdots\cdots\cdots\cdots\cdots (5)$$

이다. 단 $\rho_{gv} = \dfrac{\sum_g \sum_v (g-\bar{g})(v-\bar{v})}{\sigma_g \sigma_v}$ 임.

<정의2.> Horst의 문항 곤란도지수.

$$P_t = \frac{R_t - \dfrac{R_t - \dfrac{W_i}{K_i - 1}}{K_i - 1}}{N_t - NR_t} \times 100 \quad \cdots\cdots\cdots\cdots\cdots (6)$$

단, P_t : 문항곤란도 지수. R_t : 정답률. N_t : 수험생 총수.

NR_t : 각 문항의 비응답자 수. K_i : 답지의 수. W_i : 오답율.

[정리5.] 준거 v가 정규분포 $N(0, 1)$에 따를 때

$$r_{g_v} = \frac{M^+ - M^-}{S_v} \sqrt{P_g Q_g} \quad \cdots\cdots\cdots\cdots\cdots (7)$$

를 Sample Point Biserial Correlation Coefficient 라 한다. 단, $M^+ = \dfrac{1}{NP_g} \sum {}^+\sqrt{a}$는 문항을 정답으로 표기한 수험생의 ${}^+\sqrt{a}$의[2] 평균치.

이 때, 표준정규분포의 X좌표 수치를 수험생 N의 수가 크면 연속변량으로 작으면 이산변량으로 증분을 정해 실수치를 계산한다.

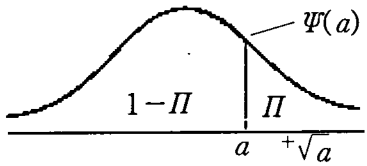
[정리6.] 관측치 X가 정규분포 $N(\mu, \sigma^2)$에 따르고 X상에서 준거 v의 회귀직선이 선형이고 $X < v$일 때, $u=0$, $X > v$일 때, $u=1$이라 하면 객관식 문항평가치인 이항변량 u는

$$\Pi = \Phi(v) = \int_v^\infty \phi(x)dx, \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

주2) ${}^+\sqrt{a}$는 표준정규분포의 X좌표 수치로써 정답을 표시한 수험생 수만큼 선정됨

$$E(x \mid x \geqq v) = \frac{\Psi(v)}{\Pi}, \quad E(X \mid X < v) = -\frac{\Psi(v)}{1-\Pi}$$

이다. 단 Π는 각 문항의 문항곤란도며 $\Psi(v)$는 표준정규분포의 높이(Ordinate)이다.



<그림 1> 문항곤란도 Π와 표준정규분포의 높이 $\Psi(a)$

따라서 정의1과 정리5의 (7)식에서 모타당도 계수(Biserial Correlation Coefficient)는

$$\rho_{g_v} = \frac{\Psi(v_g)\rho'_{g_v}}{\sqrt{\Pi_g(1-\Pi_g)}} \quad \cdots\cdots\cdots\cdots\cdots (8)$$

단, $\rho'_{g_v} = \dfrac{M^+ - M^-}{s_v} \dfrac{\Pi(1-\Pi)}{\Psi(v)}$.로 변환하여 계산한다.

Lord(1980)는 Π의 함수로써 $\dfrac{\Psi(v)}{\sqrt{\Pi(1-\Pi)}}$의 값을 다음과 같이 계산하였다.

[정리7.] Biserial Correlation 과 Sample Point Biserial Correlation 의 관계

<표 1> Biserial Correlation ρ_{g_v} 과 Sample Point Biserial Correlation ρ'_{g_v}의 관계

Π	0.50	0.04 또는 0.86	0.30 또는 0.70	0.20 또는 0.80	0.10 또는 0.90	0.05 또는 0.95
$\dfrac{\Psi(v)}{\sqrt{\Pi(1-\Pi)}}$	0.798	0.79	0.76	0.70	0.58	0.47

3. 계산 실례 및 검토

다음의 문항 분석표는 대입학력 평가 연구소에서 고등학교 3학년 1차 수학 모의고사를 출제하기 위하여 작성한 것인데 그 고사는 모두 24개의 문항들로 구성되었으며 총학생수는 $N=100$명이다.

<표 2> 수험생의 성적자료

수학 I · II-1	2 2 2 2 X 2 2 2 X X 2 2 2 X X X X X X	3 X X X B

수학Ⅰ·Ⅱ-1

<표 3> 1차 수학 모의고사의 문항분석표

문항번호	정 답	유 형	세목(내용영역)	행동영역			예상난이도		
				지식	이해	적용	상	중	하
1	④	집합과 명제	집합의 연산	○					○
2	①	수와 식	나머지정리와 인수정리		○			○	
3	②	수와 식	복소수	○				○	
4	④	수와 식	비와 비례식	○				○	
5	①	방정식과 부등식	이차방정식의 근과 계수와의 관계		○				○
6	②	방정식과 부등식	이차부등식		○			○	
7	③	방정식과 부등식	절대부등식	○					○
8	④	도형의 방정식	원		○			○	
9	②	도형의 방정식	포물선과 직선			○	○		
10	④	함수와 그래프	함수의 합성	○					○
11	①	함수와 그래프	역함수		○			○	
12	③	지수함수와 로그함수	로그의 정의와 성질	○					○
13	②	지수함수와 로그함수	지수함수		○			○	
14	④	삼각함수	삼각함수의 최대·최소	○					○
15	①	수열과 순서도	여러 가지 수열의 합		○			○	
16	①	수열과 순서도	순서도		○			○	
17	④	행렬	행렬의 뜻			○	○		
18	①	행렬	행렬의 연산	○					○
19	②	행렬	일차연립방정식과 행렬			○		○	
주1	40	집합과 명제	조건명제의 합성과 진리집합	○					○
주2	5	방정식과 부등식	삼차방정식의 근과 계수와의 관계	○				○	
주3	$\frac{x^2}{3}-\frac{y^2}{4}=-1$	도형의 방정식	타원과 쌍곡선		○			○	
주4	$\sqrt{5}$	삼각함수	삼각형과 사인·코사인 정리			○	○		
주5	$\begin{pmatrix} -3 & 2 \\ -4 & 3 \end{pmatrix}$	행렬	역행렬		○			○	

여기서 수험생들의 각 단원별 문제해결력의 크기를 변별하기 위하여 표3의 각 단원별 문항 분석표를 재편하여 그 점수 배분율을 다음과 같이 정리하였다.

<표 4> 1차 고사의 단원별 문제 배분율

출제단원	점수	총점 배분율
집합과 명제	5점 (객관식 2점, 주관식 3점)	$\frac{5}{55}$
수와 식	6점 (객관식)	$\frac{6}{55}$
방정식과 부등식	10점 (객관식 6점, 주관식 4점)	$\frac{10}{55}$
도형의 방정식	4점 (객관식)	$\frac{4}{55}$
함수와 그래프	4점 (객관식)	$\frac{4}{55}$
지수함수와 로그함수	4점 (객관식)	$\frac{4}{55}$
삼각함수	7점 (객관식 2점, 주관식 5점)	$\frac{7}{55}$
수열과 순서도	4점 (개관식)	$\frac{4}{55}$
행렬	11점 (객관식 6점, 주관식 5점)	$\frac{11}{55}$
합계	55점	1.00

이때 각 단원별 문제 해결력 즉 한 수험생의 단원별 고사 평가치 X_{11}, …, X_{19}와 그 합계 y_k, $k=1$, …, N를 1차 고사에서 수험생들의 수학적문제 해결력의 평가치 변수로 두고 또 2차 고사에서는 어려운 수학적 고등능력을 테스트하기 위하여 종합적인 주관식 문제 5문제를 출제하여 다음 <표 5>를 만들었다.

<표 5> 2차 고사의 주관식 문제 배분율

주관식 문항	총점 배분율
1	$\frac{10}{55}$
2	$\frac{10}{55}$
3	$\frac{10}{55}$
4	$\frac{11}{55}$
5	$\frac{14}{55}$
합계	1.00

위의 표들로써 다음과 같은 계산 결과를 얻었다.

<표 6> 1차고사의 타당화를 위한 가정(Hypothetical)의 모수치

	총점 배분율	단원별 평균치	단원별 평균치의 공분산 행렬 구조									
				X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}
1	$\frac{5}{55}$	2.38364	X_{11}	1.71844	2.06212	3.43687	1.37475	1.37475	1.37475	2.40581	1.37475	3.78065
2	$\frac{6}{55}$	2.86036	X_{12}	2.06212	2.47455	4.12425	1.64970	1.64970	1.64970	2.88697	1.64970	4.53667
3	$\frac{10}{55}$	4.76727	X_{13}	3.43687	4.12425	6.87375	2.74950	2.74950	2.74950	4.81162	2.74950	7.56112
4	$\frac{4}{55}$	1.90691	X_{14}	1.37475	1.64970	2.74950	1.09980	1.09980	1.09980	1.92465	1.09980	3.02445
5	$\frac{4}{55}$	1.90691	X_{15}	1.37475	1.64970	2.74950	1.09980	1.09980	1.09980	1.92465	1.09980	3.02445
6	$\frac{4}{55}$	1.90691	X_{16}	1.37475	1.64970	2.74950	1.09980	1.09980	1.09980	1.92465	1.09980	3.02445
7	$\frac{7}{55}$	3.33709	X_{17}	2.40581	2.88697	4.81162	1.92465	1.92465	1.92465	3.36814	1.92465	5.29279
8	$\frac{4}{55}$	1.90691	X_{18}	1.37475	1.64970	2.74950	1.09980	0.09980	0.09980	1.92465	1.09980	3.02445
9	$\frac{11}{55}$	5.24400	X_{19}	3.78056	4.53667	7.56112	3.02445	3.02445	3.02445	5.29279	3.02445	8.31724

위의 <표 6>의 1차고사 자료를 수험자 능력의 타당도로 가정하고 그것을 가정적(Hypothetical)모수로 두어 출제 배분율과 난이도가 높은 주관식 문제의 2차고사를 통하여 문제해결력의 전이에 따른 수험자능력의 변별에서 일치성의 크기가 어느 정도 인가를 판별할 수 있는 타당도계수의 계산과 그 요인 분석을 위여 앞절의 정리들을 이용하였다.

여기서 단원별 평균치의 공분산 행렬구조에서 예를 들어 상관계수 ρ_{14}를 구하기 위하여 $\sigma_1{}^2 = 1.71844$, $\sigma_4{}^2 = 1.09980$, $\sigma_{14} = 1.37475$로써 계산하면 ρ_{14}는 1.0으로 완전상관을 이루고 있는 것을 알 수 있으며 마찬가지로 X의 평균치로써 각 단원별간 상관계수를 계산하면 모두 완전상관이 있는 것으로 계산되므로 앞 절의 Wiley의 정리1의 예와 같이 수험자의 능력 및 모평균치를 같은 값으로 두고 그 능력별 가중치만 고려하여 정리1을 전개한 것을 이해할 수 있다.

다음 <표 7>은 앞절의 정리들로 계산된 추정치들이며 특히 두고사 평가치로써 구한 수험자들의 문제해결력의 일치성 변별을 위한 타당도 계수를 계산한 결과 0.9884로써 이 값은 2회의 고사에서 그들의 수학적 문제해결력이 거의 일치하고 있음을 나타내고 있다. 또 비타당도와 타당도의 상관 계수는 -0.0751로써 거의 완전히 무상관 임을 보여주고 있으므로 그들의 수학적 능력의 타당도 검정은

매우 유효함을 알 수가 있다.

<표 7> 2회의 고사를 통한 평가치의 타당도 분석

모수	1회 고사	2회 고사
평균치	26.2200	26.6050
분산	207.9309	207.9863
Cov $(y,\ y^*)$		205.5574
α_0		-0.0743
α_1		0.9883
γ_0		0.6842
γ_1		0.9885
β_0		0.6842
β_1		-0.0114
ρ_{yy^*}		0.9884
$\rho_{(y-y^*)y^*}$		-0.0751

다음 예는 앞의 표2와 유사한 다른 수학 성적표로써 문항수 $n=33$, 수험자수 $N=56$인 경우, 각 문항별 타당도 계수를 계산하기 위하여 2.2절의 Lord의 계산법을 이용하였다. 먼저 정의 2의 (5)식에 의하여 문항곤란도를 계산하면

$$P_{t_{41}}=\frac{R_t-\dfrac{W_t}{K_i-1}}{N_t-NR_t}\times 100=0.64285.$$

즉 Π는 0.64285이고 정의6의 (7)식에 의하여 Sample Point Biserial Correlation Coefficient를 계산하면

$$S_{X_{41}}{}^2=P_{41}Q_{41}=0.237249$$

$$M^+=E[v/u_g=1]=\frac{-0.3700-0.2737-0.1774-\cdots+2.9037+3.000}{36}=1.3150$$

$$S_{M^+}{}^2=1.0048$$

$$\cdot M^-=E[v/ag=0]=\frac{-3.000-2.8685-\cdots-0.5015}{20}=-1.7507$$

$$S_{M^-}{}^2=-0.5749$$

$$M=\frac{-3.0000-\cdots+3.0000}{56}=0.2201$$

$S_M{}^2 = 3.0064$

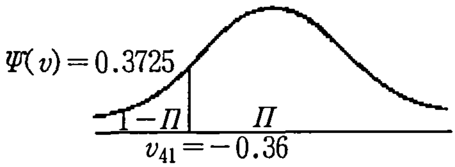


그림2. 문항 41의 문항곤란도와 준거 v_{41}

로써 $\gamma_{gv} = \dfrac{M^+ - M^-}{S_v}\sqrt{P_g Q_g} = 0.84712$이고 $\gamma_{gv} = \rho_{gv}'$이므로

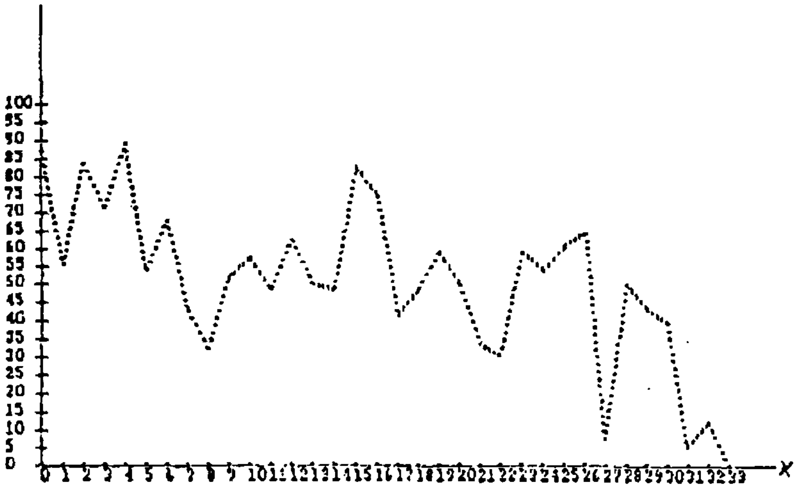
$\rho_{gv} = \dfrac{\Psi(v_g)\rho_{gv}'}{\sqrt{\Pi_g(1-\Pi_g)}} = 0.8476$이다.

즉 문항 41은 타당도가 높은편이며 문항별 표본타당도계수 γ_{gv} 즉 ρ_{gv}'와 모타당도계수(Biserial Correlation Coefficient) ρ_{gv}는 다음 <표 8>과 같다.

<표 8> n=33의 성적표의 각문항별 타당도계수 γ_{gv}와 ρ_{gv}

문항		γ_{gv}	ρ_{gv}	문항		γ_{gv}	ρ_{gv}
1	1	0.666418	0.666417	3	1	0.865731	0.865985
	2	0.863428	0.863335		2	0.842308	0.842188
	3	0.720231	0.720592		3	0.829491	0.829876
	4	0.820700	0.820871		4	0.859129	0.859123
	5	0.632313	0.631792		5	0.864800	0.864762
	6	0.864800	0.867346		6	0.855835	0.855874
	7	0.835725	0.835381				
	8	0.861809	0.861935				
	9	0.835987	0.835712				
	10	0.865712	0.865699				
2	1	0.861579	0.861680	4	1	0.847134	0.847628
	2	0.865739	0.865733		2	0.528949	0.527528
	3	0.851850	0.851851		3	0.865987	0.865986
	4	0.865987	0.865986		4	0.861809	0.861936
	5	0.865029	0.865023		5	0.856130	0.856207
	6	0.739898	0.740545		6	0.467782	0.469840
	7	0.800450	0.800449		7	0.661264	0.661263
	8	0.859358	0.859434				
	9	0.865731	0.865726				
	10	0.859129	03859192				

이 결과에서 문항 g_{24}와 g_{43}의 타당도계수치가 0.865987로 가장 높으며 문항 g_{46}의 타당도 계수치가 0.467782로 타당도가 가장 낮은 것으로 나타났으며 두 계수는 N가 56으로 작은 경우이지만 거의 유효숫자 소숫점 3째자리까지 일치하고 있으므로 표본의 크기 N와 문항수 n가 클때는 모 타당도 계수치를 더욱 정확히 계산할 수 있다.



<그림 3> n=33의 자료의 문항 정답율의 분포

4. 결 론

이 연구에서는 수학교육평가 현장에서 직접적으로 적용할 수 있는 고사 타당도 측정법을 논하였는데 첫째는 수험자들의 수학 문제 해결력 검정을 위하여 수학고사에서 1차와 2차에 걸친 문제 출제의 변동에 의한, 즉 준거의 전이를 기초로 그들의 문제해결력에 대한 고사타당도의 크기를 측정하는 법과 둘째는 평가자가 출제한 고사의 문항별 타당도 계수를 구하는 방법을 논하였는데 그 연구결과와 활용방안을 요약하면 다음과 같다.

1. 수험자들의 수학적고등정신 능력을 검정하고저 할 때 평가자가 목적하는 바의 출제범위를 정한 후 문항 분석표를 작성하고 예비시험으로써 수험생들의 기초 학력을 먼저 테스트한 후 2차고사에서는 고등수학적 문제 해결력을 검정하면서 그 고사 타당도를 계산하는 방법을 구체적으로 3절의 계산 예 및 검토에서 다루었다.

2. 한 고사의 문항들을 문항별로 각각 타당도 계수를 계산하는 방법을 보임으로써 그것을 기초

로 평가자가 목적하는 바의 문항을 선정, 배치할 수 있게 하여 결과적으로 그 고사의 타당도를 더욱 높이게 할 수 있다.

3. 1과 2에서 논한 고사 타당도 측정법을 기초로 수험생 문제 해결력의 변별을 효과적으로 검정할 수 있게 평가자가 목적하는 바의 고사 타당도를 설정하기 위하여 내용타당도와 준거타당도를 조화, 설정하는 법을 구체적으로 다루었다.

참 고 문 헌

홍석강 (1988). 수학교육에서 이해력심도의 측정과 방법 교육문제연구(동국대학교 교육문제연구소) **5**, pp.83-95.

홍석강 (1990). 신구 두 고사 평가치 변환에 의한 진분포와 모수 추정에 관한 연구, 한국수학교육학회지 시리즈 A <수학교육> **29(2)**, pp.79-83.

American Educational Research Association (1985). *Standards for Educational and Psychological testing, Washington* D.C. APA.

Cronbach L.J. & Meehl, P.E. (1955). Construct Validity in Psycholohical Tests. *Psychological Bulletin* **52**, pp.281-302.

Linn, R.L. (1993). Educational Assesment : Expanded Expectations and Challenges. *Educational Evaluation and Policy Analysis* **15**, pp.1-16.

Maguire, T. Hattie, J. & Brian, H. (1994). Construct Validity and Achievement Assessment, *The Alberta Journ. of Educational Research* **XL(2)**, pp.109-126.

Messick, S. (1995). Standards of Validity and Validity of standards in Performance Assessment, *Educational Measurement : Issues and Practice* **14(4)**, pp.5-8.

Lord, F.M. (1980). *Application of Item Response Theory to Practical Testing Problem*, NY. LEA.

Wiley, D.E. (1991). Test Validity and Invalidity Reconsidered, In R.E. Snow & D.E. Wiley (Eds). *Improving Inquiry in the Social Sciences : A Volume in honor of L.J. Crontoach*, pp.75-107, N.J. Erlbaum.