

## 일반화가능도 이론을 적용한 유아 기질 평가도구의 신뢰도와 오차요인 분석\*

최연철<sup>1)</sup> 김경철<sup>2)</sup>

---

### 요약

---

본 연구의 대상은 만4세 2학급과 만5세 3학급의 담임교사 5명과 각 유아의 어머니가 유아의 기질을 평가한 자료이다. Buss와 Plomin(1984)의 유아기질검사 결과를 활용하여 평가자, 평가문항, 평가대상 유아 그리고 각 요인간의 상호작용에 대한 분산분석을 실시하였으며 각각의 분산추정치에 대한 분석을 수행하였다. 본 연구를 통하여 유아기질평가에서 부모를 평가자로 활용할 수 있는 가능성을 파악할 수 있었다. 또한 일부 하위 요인 문항의 경우 오차변인의 비율을 줄일 수 있는 방안을 모색해야 할 필요가 있다는 점을 알 수 있었다. 본 연구는 유아기질평가도구가 가지고 있는 문제점에 대한 탐색을 통해 평가도구의 제작 및 활용시 적용할 수 있는 정보를 마련해줄 수 있다는 점에서 의의가 있을 것이라고 판단된다.

**주제어:** 유아기질검사, 일반화가능도이론, 오차요인, 신뢰도추정

---

## I. 서론

### 1. 연구의 목적 및 필요성

유아교육에서 평가는 다양한 역할을 수행한다. 즉 유아교육의 목표를 달성했는지에 대해 판단할 수 있는 자료를 마련해주고(Tyler, 1949), 유아교사 및 행정가들이 바람직한 의사결정을 내릴 수 있도록 중요한 정보를 제공하기도 한다(Stufflebeam, 1971).

---

\* 이 논문은 건국대학교 학술진흥(KU)연구비 지원에 의한 논문임

1) 건국대학교 유아교육과 교수

2) 한국교원대학교 유아교육과 교수, 교신저자 kbrian@knue.ac.kr

그러나 유아교육에서의 평가는 어려운 과제를 안고 있다. 평가의 대상이 유아이기 때문에 평가 상황에 따라 혹은 여러 가지 맥락적 조건에 따라 평가의 결과가 달라질 수 있다. 또한 유아마다 개별적인 특성이 매우 다양하기 때문에 각자의 특성을 제대로 평가한다는 것은 쉬운 일이 아니다.

그동안 유아교육연구에서는 연구자가 직접 제작한 평가도구나 외국의 검사도구를 번역, 수정, 보완한 검사도구를 활용해왔으며 내적신뢰도와 타당도를 산출하거나 검사자간 일치도를 구하는 절차를 거쳐서 도구의 적절성을 검증해왔다. 그런데 일반적으로 가장 많이 사용하는 신뢰도 계수인 Cronbach  $\alpha$ 는 오차의 원인을 오직 하나로 규명하고 있다는 문제를 가지고 있다. 따라서 평가결과에 영향을 미치는 다양한 오차요인을 파악하고 각 오차요인의 상대적 영향력을 산출하기 위해서는 고전검사이론의 한계에서 벗어나야 할 필요가 있다. 특히 유아를 대상으로 한 평가는 평가에 영향을 미치는 제반 요인들, 즉 유아의 능력, 평가자, 문항의 적절성, 그리고 그 모든 요인들의 상호작용의 상대적인 영향에 의해 평가의 결과가 달라질 수 있다는 점을 고려해야 한다. 따라서 평가결과의 신뢰도 전체에서 각 요인의 상대적 영향력을 파악하는 것은 중요한 의미를 갖는다.

고전검사이론에서는 전체 오차 가운데 각 요인의 상대적인 영향력을 파악할 수 없기 때문에 본 연구에서는 이러한 한계를 극복하기 위해 일반화가능도 이론을 활용하여 분석을 수행하고자 한다. 일반화가능도 이론은 평가에서 발생하는 다양한 오차요인을 동시에 분석하고 이를 토대로 평가점수의 신뢰도에 영향을 미치는 여러 요인의 상대적 영향력을 파악하고 신뢰도를 높이기 위한 방안을 모색할 수 있도록 도와주는 분석방법이다(김성숙, 김양분, 2001; 이종성, 1988; Brennan, 2001).

일반화가능도 분석은 G 연구와 D 연구로 구분할 수 있는데 G 연구는 연구자가 설정한 각 국면에 대한 분산의 크기와 전체 분산 가운데 차지하는 비율을 산출할 수 있는 분석방법이다. 한편 D 연구는 G 연구에서 추정된 분산성분을 토대로 하여 가장 효과적인 평가조건에 대해 탐색하는 분석방법이다. 예를 들어, 평가자의 수를 몇 명으로 조정하고 문항을 몇 개로 설정해야 최적의 평가가 이루어질 것인가에 대한 결정을 내리기 위해서는 D 연구의 결과를 토대로 하면 된다. 본 연구에서 활용한 검사도구의 경우 문항의 수가 이미 고정되어 있으며 평가자의 수를 늘리는 것 또한 현실적으로 매우 어렵기 때문에 D 연구는 수행하지 않았다.

G 연구는 허용 가능한 관찰 전집을 추정하기 위하여, 평가결과의 오차에 영향을 주는 요인들을 동시에 분석하고 분산의 크기를 비교하여 오차 요인의 상대적인 영향력을 산출하는 연구이다(Brennan, 2001). G 연구에서는 평가에 영향을 미치는 복합적인 오차요인을 구분하여 동시에 분석하기 때문에 바람직한 신뢰도를 확보하기 위한 정보

를 제공할 수 있다.

그동안 일반화가능도 분석은 동료평가(강애남, 이규민, 2006), 업적평가(김판석, 오성호, 이선우, 2000), 수행평가(남명호, 1996; 유선희, 1998), 교원능력평가(김성연, 2014) 등에서 활용되어왔으며 특히 말하기 평가를 포함한 수행평가에서는 일반화가능도 이론을 적용한 연구가 많이 수행되어왔다(예: 김경선, 이규민, 강승혜, 2010; 이영식, 신상근, 2004; 조재윤, 2008). 이러한 연구에서 다루고 있는 평가의 공통적인 면은 지필고사와는 달리 평가자에 따라 점수가 달라질 수 있으며 더 나아가서 피험자의 능력과 평가 문항에 따라 결과의 차이가 발생할 수 있는 평가라는 점이다. 유아를 대상으로 한 평가 역시 수행평가와 마찬가지로 유아, 평가자, 문항에 따라 평가결과가 달라질 수 있다는 점을 고려한다면 유아 평가에서도 일반화가능도 이론을 적용해야 할 필요가 있다고 판단된다(최연철, 1996).

본 연구에서는 기질검사의 결과를 연구대상으로 삼았는데 기질에 대한 평가는 수행평가와 마찬가지로 여러 가지 요인에 의해 평가결과에서 차이가 나타날 수 있다고 판단하였기 때문이다. 기질 연구에서 가장 선도적인 역할을 한 학자로 Thomas와 Chess를 들 수 있다. 그들은 뉴욕종단 연구에서 장기간 영아를 관찰하여 기질의 구성요인 9가지를 제시하였으며(Thomas & Chess, 1977) 그 후 9가지 기질의 구성요인을 기준으로 영아의 기질을 3가지 유형으로 구분하였다. 그러나 Thomas와 Chess의 기질 분류는 방법론적 문제와 9개 기질 요인간의 상호중복 가능성, 모든 영아를 3가지 기질 유형을 분류하는데 따른 어려움 등으로 비판받고 있으며 그래서 일부 학자들은 9개 구성요인 중 규칙성, 접근-철회, 적응성, 반응강도, 기분 등 5개 차원만 사용하여 분류하기도 한다. 한편 Rothbart와 Derryberry(1981)는 활동수준, 미소와 웃음, 공포, 제한에 대한 불평, 진정성, 주의집중 등 6개 기질 차원으로 구분한 바 있으며 Buss와 Plomin(1984)는 기질 차원을 정서성, 사회성, 활동성, 수줍음 등으로 분류하였고 각 5 문항씩 총 20문항으로 구성된 Emotionality, Activity, and Sociability(EAS) 기질검사를 제작하였다. 본 연구에서는 유아교육연구에서 많이 사용하고 있으며 비교적 최근에 제작된 검사도구인 EAS 기질검사를 활용하였다.

## 2. 연구문제

첫째, 유아의 기질평가에서 발생하는 평가의 오차원의 크기는 어떠한가?

둘째, 유아의 기질평가에서 발생하는 각각의 오차원이 유아평가에 미치는 상대적 영향력은 어떠한가?

## II. 연구방법

### 1. 연구대상

본 연구에는 햇살유치원의 만4세 개나리반(13명)과 진달래반(12명) 2학급, 만5세 민들레반(13명), 개망초반(17명), 들국화반(17명)의 담임교사 5명과 각 유아의 어머니 72명이 참여하였다(모든 명칭은 가명). 교사가 담당학급 유아의 기질을 평가한 결과와 각 어머니가 자기 자녀의 기질을 평가한 결과자료를 연구대상으로 삼았다. 자료 수집은 2013년 6월 3일에서 14일까지 수행하였으며 연구에 참여한 유아에 대한 평균월령(2013년 6월 현재 기준)과 인원수는 표 1과 같다.

〈표 1〉 연구에 참여한 유아의 평균월령과 인원수

	평균월령	남아수	여아수	총인원
만4세 2학급	59.12	8	17	25
만5세 3학급	71.8	32	15	47

### 2. 검사도구

본 연구에서 사용하고자 하는 기질 측정도구는 Buss와 Plomin(1984)이 개발한 아동용 EAS를 이해원(2012)이 수정, 번안한 검사도구이다. 기질이 어떻게 결정되느냐에 대해서는 학자들마다 의견을 달리한다. Buss와 Plomin은 기질이 유전에 의해 선천적으로 결정된다는 입장을 취하고 있으나 Thomas와 Chess는 환경에 의해 후천적으로 형성된다는 이론을 전개한 바 있다(Buss & Plomin, 1984). 본 연구에서 Buss와 Plomin의 EAS를 활용한 이유는 그들의 검사도구가 기질의 안정적인 특성을 반영하고 있기 때문이다.

앞에서 기술한 바와 같이, 본 연구에서 기질검사를 선택한 이유는 기질에 대한 평가는 수행평가와 마찬가지로 여러 가지 요인에 의해 평가결과에서 차이가 나타날 수 있다고 판단하였기 때문이다. 하지만 한편으로는 일반화가능도 연구에서 발생할 수 있는 잔여분산을 최소화해야할 필요가 있다고 판단하여 기질의 항상적인 특성을 강조한 Buss와 Plomin의 입장을 반영한 평가도구를 선택하였다.

EAS의 각 문항은 1점-4점까지 반응하도록 4점 Likert 척도를 사용하고 있으며 점수가 높을수록 각 요인의 특성이 높음을 의미한다. 원래 각 5문항씩 총20문항으로 구

성되어 있으나 각 범주별로 신뢰도 수준을 저해하는 문항(정서성 2문항, 사회성 1문항, 수줍음 1문항)을 제외하고 16문항만 분석에 포함하였다. 본 연구에서 활용한 기질 도구에 대한 간단한 설명과 본 연구의 평가대상인 유아 72명 평가결과 144부(교사와 어머니의 평가결과 각 1부) 전체의 신뢰도는 표 2와 같다.

〈표 2〉 기질검사도구의 문항수 및 신뢰도

척도	문항수	문항번호	신뢰도
정서성	4	2, 6, 11, 15	0.66
활동성	5	4, 7*, 9, 13, 17*	0.71
사회성	3	3, 5, 10	0.68
수줍음	4	1*, 8, 14*, 20	0.74

\* 표시는 역채점 문항

### 3. 분석방법 및 절차

본 연구에서는 GENOVA(a GENeralized Of analysis VAriance system) 3.1을 사용하여  $p \times i \times r$ (유아 $\times$ 문항 $\times$ 평가자) 설계에 의해 G 연구를 실시하였다. G 연구를 통해 평가대상 유아, 검사도구, 두 집단의 평가자 그리고 각 요인간의 상호작용에 대한 분산분석을 실시하였으며 각각의 분산추정치에 대한 분석을 수행하였다.  $p \times i \times r$  설계를 채택한 이유는 평가대상이 되는 유아, 검사도구에 포함된 문항, 교사와 부모로 구성된 평가자 그리고 그 세 가지 요인의 상호작용이 본 연구에서 활용한 유아평가 결과에 영향을 미친다고 볼 수 있기 때문이다.

## III. 연구결과

### 1. 정서성 하위요인의 오차원 크기와 상대적 영향력

만4세 2학년과 만5세 3학년 유아를 대상으로 교사와 부모가 수행한 기질검사 가운데 정서성 하위요인의 일반화연구  $p \times i \times r$  설계의 분석결과는 표 3, 표 4와 같다.

〈표 3〉 만4세 2학급 정서성 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	24	1516.12	1.29	1.88	0.08	13.33	23.52
문항	3	1492.22	2.37	4.57	0.04	6.67	11.77
평가자	1	1489.33	4.21	10.52	0.04	6.67	11.77
유아×문항	72	1561.50	0.53	2.05	0.14	23.33	41.17
유아×평가자	24	1530.25	0.41	1.60	0.04	6.67	11.77
문항×평가자	3	1497.16	0.25	0.95	0.00	0.00	0.00
유아×문항×평가자	72	1595.00	0.26		0.26	43.33	
합계	199	10681.58			0.6	100.00	100.00

만4세의 경우, 표 3에서 볼 수 있는 바와 같이 설명할 수 없는 오차분산인 잔여분산(유아×문항×평가자)의 추정치 백분율이 가장 높게 나타났으며(43.33%), 유아와 문항 상호작용(23.33%), 유아(13.33%), 문항(6.67%), 평가자(6.67%), 유아와 평가자 상호작용(6.67%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산성분이 전체 점수분산에서 차지하는 비중이 큰 것은 일반화가능도 이론 연구에서 나타나는 일반적인 현상으로, 잔여분산성분에는 모형에서 구성한 요인들로 설명되지 않는 오차의 분산성분 부분이 모두 포함되기 때문이다(강애남, 이규민, 2006). 따라서 잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아와 문항 상호작용은 41.17%, 유아는 23.52% 그리고 문항, 평가자, 유아와 평가자 상호작용은 모두 11.77%로 나타났다.

유아와 문항 상호작용의 분산성분 비율이 높게 나타났기 때문에 문항에 따라 유아에 대한 평가가 달라졌음을 알 수 있으며 유아의 분산성분 비율로 보아, 정서성의 경우, 유아의 개인차가 평가결과에 반영되었음을 알 수 있다. 한편 문항, 평가자 그리고 유아와 평가자 상호작용 모두 11.77%의 오차분산 비율을 차지하고 있음을 알 수 있다. 반면 문항과 평가자의 상호작용 오차분산 추정치 백분율은 전혀 나타나지 않았다. 따라서 문항이 바뀌어도 평가자가 동일한 기준을 가지고 평가했음을 알 수 있다.

〈표 4〉 만5세 3학급 정서성 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	46	2926.12	2.39	2.81	0.19	22.35	36.53
문항	3	2855.15	13.03	8.82	0.12	14.12	23.08
평가자	1	2818.00	1.94	1.30	0.00	0.00	0.00
유아×문항	138	3045.50	0.58	1.74	0.12	14.12	23.08
유아×평가자	46	2955.75	0.60	1.80	0.07	8.24	13.47

(표 4 계속)

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
문항×평가자	3	2860.78	1.23	3.67	0.02	2.35	3.84
유아×문항×평가자	138	3125.00	0.33		0.33	38.82	
합계	375	20586.3			0.85	100.00	100.00

만5세의 경우, 표 4에서 볼 수 있는 바와 같이 설명할 수 없는 오차분산인 잔여분산의 추정치 백분율이 가장 높게 나타났으며(38.82%), 유아(22.35%), 문항(14.12%), 유아와 문항 상호작용(14.12%), 유아와 평가자 상호작용(8.24%), 문항과 평가자 상호작용(2.35%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아는 36.53%, 문항은 23.08%, 유아와 문항 상호작용은 23.08%, 유아와 평가자 상호작용은 13.47%, 문항과 평가자 상호작용은 3.84%로 나타났다. 유아의 분산성분 비율이 가장 높게 나타났으므로, 정서성의 경우 유아의 개인차가 평가결과에 반영되었음을 알 수 있다. 그러나 문항 오차변인 비율로 보아 정서성의 4개 문항이 단일국면을 평가하고 있다고 보기 어려우며 유아와 문항 상호작용의 분산성분 비율이 높으므로 문항이 달라지면 유아에 대한 평가가 달라졌음을 알 수 있다. 그러나 평가자 오차변인이 전혀 나타나지 않은 것으로 보아 부모와 교사의 평가점수에서 차이가 없음을 알 수 있다.

## 2. 활동성 하위요인의 오차원 크기와 상대적 영향력

만4세 2학년과 만5세 3학년 유아를 대상으로 교사와 부모가 수행한 기질검사 가운데 활동성 하위요인의 일반화연구  $p \times i \times r$  설계의 분석결과는 표 5, 표 6과 같다.

〈표 5〉 만4세 2학년 활동성 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	24	2110.00	2.00	4.64	0.16	27.59	50.01
문항	4	2079.60	4.38	1.93	0.04	6.90	12.51
평가자	1	2062.11	0.02	0.01	0.00	0.00	0.00
유아×문항	96	2157.00	0.31	1.18	0.02	3.45	6.25
유아×평가자	24	2119.20	0.38	1.47	0.02	3.45	6.25
문항×평가자	4	2088.48	2.22	8.53	0.08	13.79	25.00
유아×문항×평가자	96	2200.00	0.26		0.26	44.83	
합계	249	14816.39			0.58	100.00	100.00

만4세의 경우, 표 5에서 볼 수 있는 바와 같이 설명할 수 없는 오차분산인 잔여분산의 추정치 백분율이 가장 높게 나타났으며(44.83%), 유아(27.59%), 문항과 평가자 상호작용(13.79%), 문항(6.90%), 유아와 문항 상호작용(3.45%), 유아와 평가자 상호작용(3.45%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아는 50.01%, 문항과 평가자 상호작용은 25.00%, 문항은 12.51%, 유아와 문항 상호작용은 6.25%, 유아와 평가자 상호작용은 6.25%로 나타났다.

유아의 분산성분 비율이 높은 것으로 보아, 활동성의 경우, 유아의 개인차가 평가결과에 반영되었음을 알 수 있다. 그러나 문항과 평가자 상호작용의 분산성분 비율이 높은 것으로 보아 평가자가 문항에 따라 다른 기준을 가지고 평가했음을 알 수 있다. 반면 평가자 오차변인이 전혀 나타나지 않은 것으로 보아 부모와 교사의 평가점수에서 차이가 없음을 알 수 있다.

〈표 6〉 만5세 3학년 활동성 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	46	4293.60	2.68	7.78	0.23	34.33	85.19
문항	4	4175.25	1.26	0.54	0.00	0.00	0.00
평가자	1	4170.51	0.31	0.14	0.00	0.00	0.00
유아×문항	184	4372.00	0.40	1.00	0.00	0.00	0.00
유아×평가자	46	4309.60	0.34	0.86	0.00	0.00	0.00
문항×평가자	4	4184.85	2.32	5.88	0.04	5.97	14.81
유아×문항×평가자	184	4470.00	0.40		0.40	59.70	
합계	469	29975.81			0.67	100.00	100.00

표 6에서 볼 수 있는 바와 같이, 만5세 유아에 대한 분석 결과는 만4세 분석결과와 다르게 나타났음을 알 수 있다. 즉 만4세의 경우 평가자에서만 오차변인이 나타나지 않았지만 만5세반에서는 평가자를 포함하여 문항, 유아와 문항 상호작용, 유아와 평가자 상호작용에서도 오차변인이 전혀 나타나지 않았음을 알 수 있다.

### 3. 사회적 하위요인의 오차원 크기와 상대적 영향력

만4세 2학년과 만5세 3학년 유아를 대상으로 교사와 부모가 수행한 기질검사 가운데 사회적 하위요인의 일반화연구 pXiXr 설계의 분석결과는 표 7, 표 8과 같다.

〈표 7〉 만4세 2학급 사회성 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	24	1615.66	0.63	2.86	0.07	24.14	46.67
문항	2	1603.24	1.29	1.39	0.01	3.45	6.67
평가자	1	1600.69	0.03	0.03	0.00	0.00	0.00
유아×문항	48	1627.00	0.18	1.27	0.02	6.90	13.34
유아×평가자	24	1620.00	0.18	1.25	0.02	6.90	13.34
문항×평가자	2	1605.04	0.87	6.17	0.03	10.34	19.99
유아×문항×평가자	48	1640.00	0.14		0.14	48.28	
합계	149	11311.63			0.29	100.00	100.00

만4세의 경우, 표 7에서 볼 수 있는 바와 같이 설명할 수 없는 오차분산인 잔여분산의 추정치 백분율이 가장 높게 나타났으며(48.28%), 유아(24.14%), 문항과 평가자 상호작용(10.34%), 유아와 문항 상호작용(6.90%), 유아와 평가자 상호작용(6.90%), 문항(3.45%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아는 46.67%, 문항과 평가자 상호작용은 19.99%, 유아와 문항 상호작용은 13.34%, 유아와 평가자 상호작용은 13.34%, 문항은 6.67%로 나타났다. 유아의 분산성분 비율이 높은 것으로 보아, 사회성의 경우, 유아의 개인차가 평가결과에 반영되었음을 알 수 있다. 또한 문항과 평가자 상호작용의 분산성분 비율로 보아 문항에 따라 평가자가 다른 기준을 가지고 평가했을 가능성이 있다. 반면 평가자 오차변인이 전혀 나타나지 않은 것으로 보아 부모와 교사의 평가점수에서 차이가 없음을 알 수 있다.

〈표 8〉 만5세 3학급 사회성 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	46	3189.66	0.64	3.62	0.06	16.67	66.68
문항	2	3160.23	0.09	0.06	0.00	0.00	0.00
평가자	1	3160.11	0.06	0.04	0.00	0.00	0.00
유아×문항	92	3212.00	0.24	0.88	0.00	0.00	0.00
유아×평가자	46	3199.33	0.21	0.76	0.00	0.00	0.00
문항×평가자	2	3163.57	1.64	6.03	0.03	8.33	33.32
유아×문항×평가자	92	3250.00	0.27		0.27	75.00	
합계	281	22334.9			0.36	100.00	100.00

만5세의 경우, 표 8에서 볼 수 있는 바와 같이 설명할 수 없는 오차분산인 잔여분산의 추정치 백분율이 가장 높게 나타났으며(75.00%), 유아(16.67%), 문항과 평가자 상

호작용(8.33%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아는 66.68%, 문항과 평가자 상호작용은 33.32%로 나타났다. 유아의 분산성분 비율이 높은 것으로 보아, 사회성의 경우, 유아의 개인차가 평가결과에 반영되었음을 알 수 있다. 또한 문항과 평가자 상호작용의 분산성분 비율로 보아 문항에 따라 평가자가 다른 기준을 가지고 평가했을 가능성이 있다. 반면 평가자 오차변인이 전혀 나타나지 않은 것으로 보아 부모와 교사의 평가점수에서 차이가 없음을 알 수 있다.

#### 4. 수줍음 하위요인의 오차원 크기와 상대적 영향력

만4세 2학급과 만5세 3학급 유아를 대상으로 교사와 부모가 수행한 기질검사 가운데 수줍음 하위요인의 일반화연구 pXiXr 설계의 분석결과는 표 9, 표 10과 같다.

〈표 9〉 만4세 2학급 수줍음 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	24	1340.37	0.16	0.10	0.00	0.00	0.00
문항	3	1345.94	3.17	2.40	0.04	5.48	9.53
평가자	1	1336.85	0.41	0.15	0.00	0.00	0.00
유아×문항	72	1370.50	0.29	0.93	0.00	0.00	0.00
유아×평가자	24	1380.75	1.67	5.42	0.34	46.58	80.97
문항×평가자	3	1350.36	1.34	4.36	0.04	5.48	9.53
유아×문항×평가자	72	1437.00	0.31		0.31	42.47	
합계	199	9561.77			0.73	100.00	100.00

만4세의 경우, 표 9에서 볼 수 있는 바와 같이 유아와 평가자 상호작용의 분산 추정치 백분율이 가장 높게 나타났으며(46.58%), 설명할 수 없는 오차분산인 잔여분산(42.47%), 문항(5.48%), 문항과 평가자 상호작용(5.48%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아와 평가자 상호작용은 80.97%, 문항은 9.53%, 문항과 평가자 상호작용은 9.53%로 나타났다. 유아의 분산성분이 전혀 나타나지 않은 것으로 보아, 수줍음의 경우, 유아의 개인차가 평가결과에 반영되지 않았음을 알 수 있다. 또한 문항과 평가자 상호작용의 분산성분 비율이 80.97%인 것으로 보아 문항에 따라 평가자가 다른 기준을 가지고 평가했을 가능성이 있다. 반면 평가자 오차변인이 전혀 나타나지 않은 것으로

로 보아 부모와 교사의 평가점수에서 차이가 없음을 알 수 있다.

〈표 10〉 만5세 3학급 수줍음 하위요인 분산성분 분석 결과

효과	자유도	제곱합	제곱평균	F	분산성분	비율	제외비율
유아	46	2297.50	0.37	0.22	0.00	0.00	0.00
문항	3	2301.93	7.14	2.46	0.04	4.49	9.29
평가자	1	2286.64	6.13	1.42	0.01	1.12	2.32
유아×문항	138	2368.00	0.36	0.78	0.00	0.00	0.00
유아×평가자	46	2385.50	1.78	3.90	0.33	37.08	76.75
문항×평가자	3	2317.06	3.00	6.57	0.05	5.62	11.63
유아×문항×평가자	138	2528.00	0.46		0.46	51.69	
합계	375	16484.63			0.89	100.00	100.00

만5세의 경우, 표 10에서 볼 수 있는 바와 같이 설명할 수 없는 오차분산인 잔여분산 추정치 백분율이 가장 높게 나타났으며(51.69%), 유아와 평가자 상호작용(37.08%), 문항(4.49%), 문항과 평가자 상호작용(5.62%), 평가자(1.12%)의 순으로 오차분산 추정치 백분율이 나타났다.

잔여분산을 제외한 나머지 분산의 성분비율을 전체(100%)로 간주하고 재계산한 결과, 유아와 평가자 상호작용은 76.75%, 문항은 9.29%, 문항과 평가자 상호작용은 11.63%, 평가자는 2.32%로 나타났다. 유아의 분산성분이 전혀 나타나지 않은 것으로 보아, 수줍음의 경우, 유아의 개인차가 평가결과에 반영되지 않았음을 알 수 있다. 또한 문항과 평가자 상호작용의 분산성분 비율이 76.75%인 것으로 보아 문항에 따라 평가자가 다른 기준을 가지고 평가했을 가능성이 있다.

#### IV. 논의 및 결론

본 연구에서는 유아를 대상으로 한 기질평가에 영향을 주는 오차요인을 분석하고 분산의 크기를 비교하여 오차 요인의 상대적인 영향력을 파악하고자 하였다. 본 연구에서 밝혀진 결과를 바탕으로 논의를 하면 다음과 같다.

첫째, 유아 오차변인의 비율이 낮게 나타난 경우가 있다는 점이다. 수줍음 하위요인은 만4세와 만5세 모두 오차변인의 분산이 전혀 나타나지 않았으며, 정서성 하위요인의 경우 잔여분산을 제외한 나머지 분산의 성분비율이 23.52%, 36.53% 등으로 비교적 낮게 나타났다. 이러한 결과는 수줍음 하위요인의 경우, 유아 기질의 개인차가 평가과

정에서 반영되지 않았고 유아 개인차 요인이 수줍음 평가결과에 전혀 영향을 미치지 않았음을 알 수 있다. 또한 정서성 하위요인의 경우에도 유아 기질의 개인차가 평가결과에 많이 반영되지 않았음을 알 수 있다. 따라서 수줍음과 정서성의 평가문항에 대한 검토가 필요하다고 하겠다.

둘째, 평가자 오차변인의 분산이 거의 나타나지 않았다는 점이다. 즉 만4세의 정서성, 만5세의 수줍음을 제외하면 평가자에 따른 오차변인의 분산이 전혀 나타나지 않았다. 이러한 결과는 부모와 교사의 평가결과가 일치하고 있다는 것을 보여주며 따라서 기질평가의 경우 부모를 평가자로 활용할 수 있다는 시사점을 던져준다. 부모는 유아의 개성, 특성, 장·단점을 직접 파악하고 있으므로 유아 평가자로서의 잠재적 가능성을 지니고 있으며 일상생활 속에서 유아와 만나기 때문에 생태적 타당도를 확보할 수 있는 평가자라는 장점이 있다.

셋째, 문항 오차변인의 비율은 하위변인에 따라 다르게 나타났다. 하위요소별 분석에서 정서성과 수줍음의 경우 만4세와 만5세 모두에서 오차변인 분산이 산출되었다. 따라서 정서성과 수줍음의 하위 평가문항은 각각 정서성과 수줍음이라는 단일국면을 평가하고 있다고 판단하기 어렵다.

넷째, 유아와 문항 상호작용 분산의 경우, 만4세와 만5세 모두 정서성 하위요인에서 오차 분산이 산출되었으며 만4세의 경우 활동성과 사회성 하위요인에서 오차분산이 나타났다. 나머지 경우에는 유아와 문항 상호작용의 분산이 전혀 나타나지 않았으므로 문항이 달라지더라도 유아에 대한 평가가 달라지지 않았음을 알 수 있다.

다섯째, 유아와 평가자 상호작용 분산의 경우 만5세의 활동성 하위요인과 사회성 하위요인을 제외한 나머지 경우에 모두 오차 분산이 나타났다. 특히 수줍음 하위요인의 경우 잔여분산을 제외한 나머지 분산의 성분비율이 80.97%, 76.75%로 나타나 전체 오차분산에 차지하는 비중이 매우 높은 것으로 나타났다. 따라서 몇 경우를 제외하면 평가자에 따라 특정 유아에 대해 편향적 평가를 했을 가능성이 있으며 수줍음 하위요인의 경우 그 정도가 크다는 것을 알 수 있다.

여섯째, 문항과 평가자 상호작용 분산의 경우 만4세 정서성을 제외하면 모든 경우에서 오차 분산이 나타났음을 알 수 있다. 따라서 평가자가 문항에 따라 다른 기준을 가지고 평가했음을 알 수 있다.

본 연구를 통하여 유아기질평가의 경우 부모를 평가자로 활용할 수 가능성을 파악할 수 있었다. 다만 일부 하위 요인 문항의 경우 유아의 개인차가 평가결과에 반영될 수 있는 문항인지, 단일국면을 평가하고 있는지, 하위문항이 적절하게 구성되었는지

그리고 평가자가 일관된 기준을 가지고 평가할 수 있도록 문항이 작성되었는지에 대한 고려를 해야 할 필요가 있다는 알 수 있었다. 예를 들어, 정서성과 수줍음의 경우에는 다양한 오차요인이 발생하였기 때문에 각 문항이 적절하게 제시되었는지에 대한 고려가 필요하다. 본 연구는 유아기질평가도구가 가지고 있는 문제점에 대한 탐색을 통해 평가도구의 제작 및 활용시 적용할 수 있는 정보를 마련해줄 수 있다는 점에서 의의가 있을 것이라고 판단된다.

본 연구에서는 연령별 차이가 있을 수 있다는 가정 하에 만4세와 만5세를 구분하여 분석하였으나 전반적으로는 볼 때 연령별 차이가 크게 나타나지 않았다. 이러한 결과는 본 연구의 성과이기도 하지만 한편으로는 후속연구가 필요한 근거가 되기도 한다. 본 연구는 다섯 학급의 유아만을 대상으로 평가한 결과를 활용했다는 점에서 제한점을 가지고 있으므로 더 많은 유아를 대상으로 한 연구를 수행해야 할 필요가 있다. 또한 연구에 참여한 부모는, 교사와 달리, 평가에 대한 전문적인 교육을 받지 못했다는 점도 본 연구의 제한점 가운데 하나이다.

## 참고문헌

- 강애남·이규민(2006). 학생들의 동료평가를 활용한 수행평가 결과의 일반화가능도 분석. *교육평가연구*, 19(3), 107-121.
- 김경선·이규민·강승혜(2010). 일반화가능도 이론을 적용한 한국어 말하기 성취도 평가의 신뢰도와 오차요인 분석. *한국어 교육*, 21(4), 51-75.
- 김성숙·김양분(2001). *일반화가능도 이론*. 서울: 교육과학사
- 김성연(2014). 미국의 수업관찰평가 분석을 통한 우리나라 교원능력개발평가에서의 다변량 일반화가능도 이론 활용성 탐색. *한국교육*, 41(1), 5-29.
- 김관석·오성호·이선우(2000). 업적평가체계로서의 다면평가제도의 도입과 추진방법에 관한 사례연구, *한국행정학보*, 34(4), 348-349.
- 남명호(1996). 수행평가에 있어서 일반화가능도 이론의 활용. *교육평가연구*, 9(2), 73-93.
- 유선희(1998). 수행평가의 일반화가능도, 채점결과의 신뢰도 및 지필 검사와의 상호관계에 관한 연구. *교육평가연구*, 11(2), 23-41.
- 이영식·신상근(2004). 다변량 일반화 가능도 이론에 의한 말하기 시험의 타당도와 신뢰도에 관한 연구. *외국어교육*, 11(2), 259-265.

- 이종성(1988). 일반화가능도이론. 서울: 연세대학교 출판부.
- 이혜원(2011). 친사회적 행동과 관련된 변인들의 영향력에 관한 연구. 경희대학교 교육대학원 석사학위논문.
- 조재윤(2008). 일반화가능도 이론을 이용한 말하기 평가의 과제와 채점자 요인 최적화 조건에 관한 연구. 고려대학교 박사학위논문.
- 최연철(1996). 유아 평가를 위한 부모교육이 어머니와 교사의 관찰자 신뢰도에 미치는 영향. 중앙대학교 석사학위논문.
- Brennan, R. L.(2001). Generalizability theory. Springer Verlag.
- Buss, A. & Plomin, R. (1984). Temperament: Early developing personality traits. Hillsdale, NJ: Erlbaum.
- Rothbart, M. K., & Derryberry, D.(1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown(Eds.), Advances in developmental psychology (Vol. 1, pp. 37-86). Hillsdale, NJ: Erlbaum.
- Stufflebeam, D.L.(1971). Educational evaluation and decision making. Ithaca, IL: Peacock.
- Thomas, A., & Chess, S.(1977). Temperament and development. New York: Brunner/Mazel.
- Tyler, R.W.(1949), Basic principles of curriculum and instruction. IL, Chicago: University of Chicago Press.

·논문접수 10월 23일 / 수정본 접수 11월 28일 / 게재 승인 12월 22일

·교신저자: 김경철, 한국교원대학교 유아교육교과 교수, 이메일 kbrian@knue.ac.kr

## Abstract

### **An Analysis of Error Sources and Reliability Estimation in Temperament Assessment of Young Children Using Generalizability Theory**

Youchul Choi and Kyungchul Kim

The purpose of this study is to determine error sources and effects of each error source in temperament assessment of young children. Temperament of four- and five-years old children was assessed using the Emotionality, Activity, and Sociability(EAS) developed by Buss and Plomin(1984) and the evaluation results were analyzed using G study of generalizability theory. The results of G study shows that parents can be effective evaluators for temperament assessment of young children. Also strategies to reduce the effects of error were discussed.

Key words: temperament assessment of young children, Generalizability theory, error source, reliability estimation