

교육대학원 예비공학교사의 교직 적성·인성 검사에서 일반화가능도 분석

김성연

인천대학교 교육대학원

Generalizability Analysis of Teaching Aptitude and Personality Test for Pre-service Engineering Teachers in a Graduate School of Education

Sung-Yeun Kim

Graduate School of Education, Incheon National University

요약 본 연구는 측정학적 특성을 바탕으로 교육과학기술부에서 표준안으로 보급한 교직 적성·인성 검사를 교육대학원의 예비공학교사에게 실시하는 것이 적절한지를 살펴보고, 교직 적성·인성 검사가 적정 수준의 신뢰도에 도달할 수 있는 효율적인 측정 조건을 제시하였다. 이를 위해 수도권에 소재한 한 교육대학원에서 2013년부터 2017년까지 공학교육전공의 예비공학교사 99명을 대상으로 실시한 교직 적성·인성 검사자료에 일반화가능도 분석을 수행하였다. 주요 분석결과는 다음과 같다. 첫째, 교직 적성·인성 검사점수에 미치는 요인의 상대적인 영향력은 일반적으로 잔차, 영역 내 문항, 피험자, 피험자와 영역의 상호작용, 영역, 그리고 시행횟수 크기의 순으로 나타났다. 둘째, 고전검사이론을 바탕으로 한 Cronbach's α 는 측정 구조를 적합하게 반영하지 못함으로써 과대추정 될 수 있으므로 준거참조검사에서는 의존도계수와 같은 보다 정밀한 신뢰도를 추정할 필요가 있다. 셋째, 교직 적성·인성 검사 표준안은 의존도계수를 바탕으로 교육대학원의 예비공학교사에게 적용 가능하다는 경험적 근거를 제시하였다. 넷째, 교직 적성·인성 검사가 적정 수준의 신뢰도에 도달할 수 있는 효율적인 측정 조건은 총 문항 수를 기준으로 시행횟수는 2회, 영역은 12개, 그리고 영역 내 문항은 8개로 원자료인 총 210개 문항을 96개로 축소할 수 있는 것으로 나타났다. 마지막으로 본 연구의 제한점과 후속 연구를 위한 방향을 제시하였다.

Abstract This study examined the fairness of the teaching aptitude and personality test assigned to pre-service engineering teachers in a graduate school of education based on measurement traits. For this study, we analyzed the teaching aptitude and personality test scores of 99 students enrolled in engineering education in a graduate school of education located in the Seoul metropolitan area from 2013 to 2017. The main results were as follows. First, the estimated variance due to residual was generally the highest, followed by nesting of items within domains, pre-service engineering teachers, interactions of pre-service engineering teachers with domains, domains, and occasions. Second, dependability coefficients were better indicators than Cronbach's because the latter may have been overestimated by applying the traditional reliability coefficient in inappropriate manners. Third, the teaching aptitude and personality test can be applied to pre-service engineering teachers in a graduate school of education based on empirical evidence considering dependability coefficients. Fourth, a total of 96 items from the original 210 items, with 2 occasions and 12 domains containing 8 items in each domain, were optimal measurement conditions to reach adequate degrees of reliability based on the total number of items. Finally, the results were discussed, the study limitations described and future research directions proposed.

Keywords : Criterion-referenced test, Dependability coefficients, Generalizability analysis, Pre-service engineering Teacher, Teaching aptitude and personality test

*Corresponding Author : Sung-Yeun Kim (Incheon National Univ.)

Tel: +82-32-835-8163 email: syk@inu.ac.kr

Received February 27, 2018

Revised April 16, 2018

Accepted May 4, 2018

Published May 31, 2018

1. 서론

우리나라에서는 2015년 인성교육진흥법 및 그 시행령[1]이 발효됨으로써 각 급 학교의 장은 매년 자체적으로 인성교육 계획을 수립하고 시행하여야 한다[2]. 이처럼 인성교육진흥법 제정을 통해 인성교육을 강조하는 것은 인성교육 인식 제고와 활성화 등을 촉진할 수 있지만, 실질적으로 학생의 인성 함양에 도움이 되는지에 대한 여부는 미지수다[3, 4]. 따라서 성공적인 인성교육이 학교 현장에서 이루어질 수 있도록 다양한 접근과 노력이 필요하다. 그 중 하나는 교원양성 단계에서부터 예비교사의 인성교육 역량을 강화하는 것이다[5]. 특히 특성화고는 특정 전문 분야에 소질과 적성을 가지고 진로를 결정하여 일찍부터 해당 산업분야의 직업 준비교육을 받으려는 학생들을 대상으로 특정 전문분야의 교육을 실시하는 고등학교로 정의[6]함으로써 일반고와는 구별된다. 따라서 특성화고 학생은 졸업 후 일자리 조기진입 비중이 높기 때문에 직장 구성원과 상호작용을 하는데 필요한 능력으로 올바른 의사소통 능력, 근면, 성실, 예절 등을 포함하는 사회성을 갖출 필요가 있다. 그러나 특성화고 졸업 근로자의 경우 학창 시절 겪었던 실패로 인한 부정적 자아개념이 형성되면서 직장에 적응하지 못하고 [7], 직장 내에서 성실성, 조직적응력 및 공격책임감과 같은 인성이 부족하다는 문제가 지적되었다[8]. 또한 특성화고에 재학 중인 학생들은 자신의 감정을 조절하고 건강하게 표현하는 것에 어려움을 겪기도 한다[5, 9, 10]. 따라서 특성화고 교사는 학생들에게 필요한 역량을 가르치기 위해서 교사 스스로가 이러한 역량을 지니고 있어야 한다[2]. 또한 특성화고 교사는 많은 전문분야 중에서 특정산업과 연관된 직업분야에 초점을 맞추어 그 분야의 직업인으로서 필요한 전문지식 및 기술뿐만 아니라 [11] 학생의 인성적 요인을 중요하게 다루어야 한다.

이와 더불어 교육부에서는 교원 적성·인성 강화를 위한 방안으로 교사 양성과정에서 교직 적성·인성 검사를 시행함으로써 예비교사의 자질을 향상시키고자 하였다. 교원양성과정은 향후 교사로 복무할 예비교사를 양성하는 과정이기 때문에 이들의 교직 적성·인성 수준은 장차 복무할 교육현장의 질적 발전을 위해 필수적인 요인이 된다[12]. 구체적으로 교육과학기술부[13]는 각 교원양성기관에서 자율적으로 활용할 수 있도록 교직 적성·인성 검사 표준안[14]을 개발하여 보급하였으며, 교원자격

검정령 제19조의 무시험검정 합격기준에 의해 예비교사들은 교원양성기관 재학 중에 해당 기관의 장이 실시하는 교직 적성·인성 검사에서 2회 이상의 적격 판정을 획득하여야만 교원 자격증을 취득할 수 있게 되었다. 교직 적성·인성 검사 표준안은 타당화 과정을 거쳐 14개 영역의 영역 내 15문항을 포함하는 총 210개 문항으로 구성되어 있으며, 학교급 또는 전공에 상관없이 많은 교원양성기관들에서 활용되고 있다[12, 15]. 그러나 교직 적성·인성 검사 표준안은 개발 과정에서 교대, 사대, 교직과정 교수들로 전문가 그룹을 구성함으로써 유아 및 특수교원 양성기관의 전문가가 포함되지 않았으며, 피험자 대상이 사범대학 및 교육대학의 대학생으로 제한됨으로써 교육대학원 원생들이 포함되지 않았다는 한계가 있다. 일례로 조운주[16]는 예비유아교사들을 대상으로 이들에게 적합한 교직 적성·인성 검사로 기존 14개 영역의 총 210 문항에서 9개 영역의 총 140문항으로 수정할 것을 제안하였으며, 원혜경과 이소정[17]은 영유아의 발달을 고려하여 기존 14개 영역에 민감성 영역을 추가하여 예비유아교사에게 적합한 15개 영역의 각각 75문항을 포함하는 A세트와 B세트를 제안하였다. 또한 교직 적성·인성 검사에 대해 최근 김성호[18]는 검사 결과 부적격으로 판정되는 예비교사의 수가 한 명도 없는 교원양성기관이 전체의 87.2%로 교직 적성·인성 검사가 예비교사를 걸러내는 장치로서의 역할을 하지 못하고 있다는 문제점을 지적하였으며, 한국교육과정평가원[19]은 교직 적성·인성 검사의 타당도와 신뢰도에 대한 제고의 필요성을 강조하였다.

따라서 본 연구는 교직 적성·인성 검사 표준안이 교원양성기관에 보급 된지 5년이 지난 현 시점에서 대학생을 대상으로 타당화된 검사 표준안이 상대적으로 연구가 수행되지 않았던 교육대학원의 공학교육을 전공하는 원생들에게 동일하게 적용할 수 있는지에 대한 실제적인 관심에서 연구의 필요성을 인식하였다. 이에 본 연구에서는 측정학적 특성 중 하나인 신뢰도 산출 시 다중 오차 요인을 고려함과 동시에 준거참조검사에 적합한 신뢰도인 의존도계수를 산출하는 일반화가능도 분석을 수행함으로써 교육대학원의 예비공학교사에게 교직 적성·인성 검사 표준안을 사용할 수 있는지에 대한 경험적인 근거를 제시하고자 하며, 적정 수준의 신뢰도에 도달할 수 있는 효율적인 측정 조건을 바탕으로 교직 적성·인성 검사를 개선 및 향상시킬 수 있는 방안을 제시하고자 한다.

구체적인 연구문제는 다음과 같다.

첫째, 교직 적성·인성 검사에서 교육대학원의 예비공학교사 검사점수에 영향을 미치는 요인들의 상대적인 영향력은 어떠한가?

둘째, 교육대학원의 예비공학교사를 위한 교직 적성·인성 검사가 적정 수준의 신뢰도에 도달하기 위한 최적의 문항 수, 영역 수, 그리고 시행횟수 의 조합은 무엇인가?

2. 연구방법

2.1 분석 자료

본 연구에서는 2013년부터 2017년까지 한 교육대학원에서 교육부에서 표준안으로 보급한 교직 적성·인성 검사[14]를 공학교육전공 원생들인 예비공학교사에게 실시한 결과를 분석 자료로 활용하였다. 이 자료에는 예비공학교사들의 입학 연도에 따라 1회 또는 2회의 결과가 혼합되어 있으므로, 본 연구에서는 연구 설계에 따라 시행횟수를 고려하지 않은 연구 설계는 반복측정되지 않은 집단(No Replication Group, 이후 ‘NR-G’라고 함)으로, 그리고 시행횟수를 고려한 연구 설계는 1회와 2회 모두 검사점수가 있는 반복측정된 집단(Replication Group, 이후 ‘R-G’라고 함)으로 자료를 구분하여 분석 대상으로 정하였다. 단, 1회와 2회 검사를 모두 수행한 학생의 경우에는 첫 번째 결과만을 NR-G의 자료에 포함시켰다. 각각 NR-G에는 99명이, 그리고 R-G에는 38명의 예비공학교사 자료가 포함되어있다. 또한 교직 적성·인성 검사점수는 김정환 외[14]에서 제시한 방법에 따라 정직성과 일관성 척도에 해당하는 문항들에 대해서는 두 문항에 동일한 배점으로, 그리고 무응답은 0점으로 채코딩하였다.

Table 1. Frequencies of respondents by majors and genders from 2013 to 2017

	2013		2014		2015		2016		2017		Total
	F	M	F	M	F	M	F	M	F	M	
①	1	3	0	0	0	0	4	3	1	2	14
②	2	10	0	1	1	5	1	8	1	7	36
③	0	2	0	0	0	1	0	1	2	5	11
④	1	6	0	2	0	2	2	2	1	6	22
⑤	2	2	0	1	1	1	0	2	4	3	16
Total	6	23	0	4	2	9	7	16	9	23	99

Note. F=Female, M=Male, ①Architectural education, ②Mechanical engineering education, ③Electric education, ④Technical education, ⑤ Computer education.

구체적으로 교직 적성·인성 검사를 받은 NR-G에 속하는 예비공학교사들을 시행 시기, 전공, 그리고 성별에 따라 Table 1에 빈도를 제시하였다. 총 교육대학원의 예비공학교사 수는 99명이며, 2013년에 29명을 제외하면 해가 거듭될수록 더 많은 예비공학교사들에게 교직 적성·인성 검사가 시행되었음을 알 수 있다. 2013년에는 이 표준안 검사를 교육부가 처음으로 보급하면서 재학 중인 교육대학원의 예비공학교사 전체를 대상으로 실시했기 때문에 상대적으로 그 수가 많다고 해석할 수 있다. 또한 공학교육 전공 내에서 기계교육전공 예비공학교사가 36명으로 가장 많이 나타났다. 성별 구성은 남자 예비공학교사는 75명이며, 여자 예비공학교사는 24명으로 2016년도 건축교육전공을 제외하면 모든 전공에서 남자 예비공학교사의 수가 여자 예비공학교사의 수보다 많이 나타났다.

2.2 분석방법

교육대학원의 예비공학교사인 피험자(p)에게 14개 영역(d) 별로 서로 다른 15개 문항(i)이 포함되어 있는 총 210개 문항으로 구성된 교직 적성·인성 검사 표준안이 2회(o) 실시되었으므로, 본 연구에서는 3가지 G-연구 설계를 분석하였다. 단, 피험자(p)는 무한한 모집단에서 무작위로 추출되었으며, 영역(d)과 문항(i)은 무한한 영역과 문항 전집에서 표집되었다고 간주하여 모두 무선효과(random effect)로 가정하였다. 교육대학원의 예비공학교사마다 총 210개 문항의 교직 적성·인성 검사점수를 고려하는 경우에는 이를 일반화가능도 분석 체계로 표현한 $p \times i$ 설계를, 예비공학교사마다 14개 영역 별로 서로 다른 15개 문항이 포함되어 있는 총 210개의 교직 적성·인성 검사점수를 고려하는 경우에는 $p \times (i : d)$ 설계를, 그리고 예비공학교사마다 14개 영역 별로 서로 다른 15개 문항이 포함되어 있는 210개의 문항을 2회 실시한 총 420개의 교직 적성·인성 검사점수를 고려하는 경우에는 $p \times o \times (i : d)$ 설계를 적용하였다. 각 설계의 분산성분을 추정하기 위한 G-연구 분석과, 이를 바탕으로 오차국면의 수를 조절함으로써 효율적인 측정 조건을 탐색하는 D-연구 분석에는 GENOVA[20]프로그램과 엑셀을 사용하였다. 단, D-연구 분석을 위해서는 G-연구 설계와 동일한 $p \times I$ 설계, $p \times (I : D)$ 설계, 그리고 $p \times O \times (I : D)$ 설계를 적용하였다.

또한 신뢰도의 적정 수준에 대해서는 연구자들마다

견해가 다르지만 고전검사이론을 바탕으로 한 검사-재검사신뢰도 및 Cronbach's α 보다는 낮은 계수를 산출하는 의존도계수에 대해서는 일반적으로 0.70이상[21-24]을 보고하고 있다. 그러나 Webb 외[25]와 Norcini[26]는 신뢰도 종류에 상관없이 0.80이상을 제시하며, 고부담 검사(high risk test)일수록 신뢰도가 높아야 한다는 점에서는 연구자들의 견해가 일치하고 있다. 따라서 본 연구에서는 0.84를 적정 수준의 신뢰도로 정하였다.

3. 연구결과

3.1 기술통계 분석결과

본 연구에서 활용한 교직 적성·인성 검사 표준안[14]의 하위 영역 점수와 총점에 대해 Cohen's d [27]와 분산비를 포함한 기술통계 분석결과를 NR-G와 R-G로 나누어 Table 2에 제시하였다.

Table 2. Descriptive statistics of domain scores

Dom	NR-G(99)		R-G(38)		# of items	Cohen's d	Var ratio
	Mean	SD	Mean	SD			
①	48.85	6.32	48.53	6.42	15	1.03	0.05
②	48.68	5.32	48.16	5.26	15	0.98	0.10
③	46.17	5.89	45.55	5.41	15	0.84	0.11
④	42.27	6.03	41.55	5.76	15	0.92	0.12
⑤	47.29	7.40	46.08	6.96	15	0.88	0.17
⑥	48.44	6.42	47.32	6.25	15	0.95	0.18
⑦	46.59	6.22	45.34	6.71	15	1.17	0.19
⑧	47.41	6.25	46.76	6.37	15	1.04	0.10
⑨	51.02	6.82	50.50	6.49	15	0.91	0.08
⑩	47.58	5.96	46.42	5.98	15	1.00	0.19
⑪	48.96	7.40	48.11	7.26	15	0.96	0.12
⑫	47.46	5.79	46.84	5.66	15	0.96	0.11
⑬	53.13	5.80	52.26	6.12	15	1.11	0.15
⑭	40.81	4.43	39.32	3.86	15	0.76	0.36
Total	664.67	61.52	652.74	63.42	210	1.06	0.19

Note. ()=# of pre-service teachers; Dom=Domain; Var= Variance; ①Problem Solving-Exploration; ②Judgment; ③Independency-Autonomy; ④Creativity-Application; ⑤Psychological Stability; ⑥Language-Communication; ⑦Leadership; ⑧Sympathy-Tolerance; ⑨Knowledge-Information; ⑩Service-Sacrifice-Corporation; ⑪Plan; ⑫Sincerity-Responsibility; ⑬Vocation-Education Values; ⑭Passion.

Cohen's d 는 NR-G의 평균에서 R-G의 평균을 뺀 후 두 집단의 통합표준편차로 나눈 값으로, 분산비는 R-G의 분산을 NR-G의 분산으로 나눈 값이다. 총점을 기준으로 NR-G의 평균이 R-G의 평균보다 높으며, R-G의 분산이 NR-G의 분산보다 크게 나타났다. 또한 열정 영역

에서 두 집단의 평균 차이가 가장 큰 0.36의 중간 효과 크기가 나타났으며, NR-G가 R-G에 비해 변동성이 24% 더 크게 나타났다.

3.2 요인들의 상대적인 영향력 탐색

3.2.1 다양한 G-연구 설계

교육대학원 예비공학교사의 교직 적성·인성 검사점수에 영향을 미치는 요인들의 상대적인 영향력을 탐색하기 위해 모든 요인들을 무선희효과로 가정한 $p \times i$, $p \times (i : d)$, 그리고 $p \times o \times (i : d)$ 설계를 적용하여 분석한 G-연구 결과를 Table 3-Table 5에 제시하였다. Table 3에서 문항 효과와 피험자 효과는 각각 30.93%와 8.25%로 나타났다. 이 중 문항 효과는 잔차 효과를 제외하면 가장 크게 나타남으로써 문항들의 난이도 차이가 교직 적성·인성 검사점수에 미치는 영향력이 상당히 크다고 해석할 수 있다. 또한 피험자 효과는 피험자의 교직 적성·인성 수준 차이가 교직 적성·인성 검사점수에 반영되어 있다고 해석할 수 있다.

반면에 피험자와 문항의 상호작용을 포함한 잔차 효과는 60.82%로 가장 크게 나타났다. 여기서 피험자와 문항의 상호작용 효과란 교직 적성·인성 검사점수에서 피험자의 상대적 순위가 문항에 따라 크게 달라지고 있다고 해석할 수 있다. 또한 G-연구 분석결과 산출되는 잔차 효과에는 고차의 상호작용 효과가 포함되어있기 때문에 일반적으로 크게 나타날 수 있으며, 동시에 본 연구 설계에서 고려하지 못한 다른 요인들의 필요성에 대한 간접적인 설명으로 해석할 수 있다[21, 28-31].

Table 3. Results for G-Study $p \times i$ design

Effect	df	SS	MS	VAR(SE)	%VAR
p	98	1784.04	18.20	0.08(0.01)	8.25
i	209	6416.46	30.70	0.30(0.03)	30.93
p,i,e	20482	12173.85	0.59	0.59(0.01)	60.82

Note. VAR=Estimated variance component; %VAR= Percentage of total variance.

문항 국면 외에 추가적으로 영역 국면을 고려한 분석 결과를 Table 4에 제시하였다. 잔차 효과를 제외하면 영역 내 문항 효과가 29.95%로 가장 크게 나타남으로써 영역 내 문항들의 난이도 차이가 교직 적성·인성 검사점수에 미치는 영향력은 크다고 해석할 수 있다. 피험자 효과는 8.16%로 잔차 효과를 제외하면 두 번째로 크게 나

타남으로써 교직 적성·인성 검사점수에 교직 적성·인성 수준 차이가 반영되어 있다고 해석할 수 있다. 다음으로 피험자와 영역의 상호작용 효과는 5.52%로 영역 효과인 1.05%보다 크게 나타났다. 이는 전체 피험자의 교직 적성·인성 검사점수는 영역 별로 큰 차이를 보이지 않으나, 교직 적성·인성 검사점수의 피험자 간 상대적 순위는 영역에 따라서 차이가 있다고 해석할 수 있다. 마찬가지로 영역 내 피험자와 문항의 상호작용을 포함한 잔차 효과는 55.31%로 가장 크게 나타났다.

Table 4. Results for G-Study $p \times (i: d)$ design

Effect	df	SS	MS	VAR(SE)	%VAR
<i>p</i>	98	1784.04	18.20	0.08(0.01)	8.16
<i>d</i>	13	596.49	45.88	0.01(0.01)	1.05
<i>i: d</i>	196	5819.97	29.69	0.29(0.03)	29.95
<i>pd</i>	1274	1729.43	1.36	0.05(0.00)	5.52
<i>pi: d, e</i>	19208	10444.43	0.54	0.54(0.01)	55.31

Note. VAR=Estimated variance component; %VAR= Percentage of total variance.

문항, 영역 국면 외에 추가적으로 시행횟수 국면을 고려한 분석 결과를 Table 5에 제시하였다. 잔차 효과를 제외하면 영역 내 문항 효과가 29.26%로 가장 크게 나타남으로써 영역 내 문항들의 난이도 차이가 교직 적성·인성 검사점수에 미치는 영향력은 상당히 크게 나타났다. 다음으로 시행횟수와 영역의 상호작용 효과, 그리고 피험자 효과 순으로 각각 14.71%와 9.29%로 나타났다. 이는 시행횟수에 따라 영역 간에 특정 차이가, 그리고 피험자의 교직 적성·인성 수준 차이가 교직 적성·인성 검사 점수에 반영되어 있다고 해석할 수 있다.

Table 5. Results for G-Study $p \times o \times (i: d)$ Design

Effect	df	SS	MS	VAR(SE)	%VAR
<i>p</i>	37	1388.32	37.52	0.08(0.02)	9.29
<i>o</i>	1	0.00	0.00	0.00(0.00)	0.00
<i>d</i>	13	436.97	33.61	0.01(0.01)	1.32
<i>i: d</i>	196	3892.84	19.86	0.25(0.03)	29.26
<i>po</i>	37	111.21	3.01	0.01(0.00)	1.25
<i>pd</i>	481	687.49	1.43	0.02(0.00)	1.73
<i>od</i>	7252	4331.22	0.60	0.13(0.01)	14.71
<i>pi: d</i>	13	7.51	0.58	0.00(0.00)	0.00
<i>oi: d</i>	196	66.36	0.34	0.00(0.00)	0.00
<i>pod</i>	481	348.28	0.72	0.03(0.00)	2.94
<i>poi: d, e</i>	7252	2482.64	0.34	0.34(0.01)	39.50

Note. VAR=Estimated variance component; %VAR= Percentage of total variance.

반면에 시행 효과와 피험자와 시행의 상호작용 효과는 각각 0.00%와 1.25%로 교직 적성·인성 검사점수에 미치는 영향력은 상대적으로 작게 나타났다. 이는 교직 적성·인성 검사점수는 시행횟수에 따라 차이가 거의 없으며, 검사점수의 상대적 순위도 시행횟수에 따라 변화가 거의 없다고 해석할 수 있다. 또한 영역 내 피험자, 시행횟수 및 문항의 상호작용을 포함한 잔차 효과는 39.50%로 가장 크게 나타났다.

3.2.2 다양한 D-연구 설계 비교

연구 설계들의 비교를 용이하게 하기 위하여, G-연구 설계와 동일하게 D-연구를 수행한 결과를 Table 6-Table 7에 제시하였다.

Table 6. Comparison of default D study results of different designs

	$p \times i$	$p \times (i: d)$	$p \times o \times (i: d)$
$\sigma^2(p)$	0.084	0.080	0.080
$\sigma^2(\Delta)$	0.004	0.009	0.011
Φ	0.951	0.903	0.881

Note. $\sigma^2(p)$ =Estimated universe score variance; $\sigma^2(\Delta)$ =Estimated absolute variance; Φ =Dependability coefficient.

Table 6은 원자료와 같은 표본 크기를 갖는 D-연구의 분석결과이다. 국면이 늘어날수록 절대오차분산은 증가하며, 의존도계수는 감소하는 것으로 나타났다. 특히 고전검사이론을 바탕으로 한 Cronbach's α 는 기준참조검사에 적합한 신뢰도로 $p \times I$ 설계에서의 일반화가능도 계수와 동일하다. $p \times I$ 설계에서 전집점수분산은 0.084, 상대오차분산은 0.003(=0.594/210)으로, 일반화가능도계수는 0.966(=0.084/(0.084+0.003))으로 Table 6에 제시된 신뢰도와 비교할 때 가장 높게 나타남을 확인할 수 있다.

Table 7은 구체적인 영역 효과와 시행횟수 효과를 제시하였다. 각 효과는 연구 설계 중 오직 영역 국면만 또는 시행횟수 국면만 무시하였을 때의 차로 계산하였으며, 영역 효과가 시행횟수 효과보다 크게 나타났다.

Table 7. Domain and occasion effect

	Domain effect		Occasion effect	
	$p \times (I: D) - p \times I$	$p \times O \times (I: D) - p \times (I: D)$		
$\sigma^2(p)$	-0.004		0.000	
$\sigma^2(\Delta)$	0.005		0.002	
Φ	-0.048		-0.022	

Note. $\sigma^2(p)$ =Estimated universe score variance; $\sigma^2(\Delta)$ =Estimated absolute variance; Φ =Dependability coefficient.

3.3 최적의 시행횟수 탐색

국면이 가장 많이 포함됨으로써 이론적으로 가장 적합한 G-연구 설계인 $p \times o \times (i: d)$ 와 동일한 설계를 적용한 D-연구 결과 중 일부를 Table 8에 제시하였다. 원자료인 14개 영역에 영역 별로 15개 문항이 포함되어 있는 교직 적성·인성 검사를 2회 실시하였을 때의 의존도계수는 0.881로 Table 6과 동일하게 나타났다. 따라서 원자료의 경우 시행횟수를 1회로 축소할 경우에 의존도계수는 0.817로 적정 수준의 신뢰도에 도달하기 위해서는 최소 시행횟수는 2회인 것으로 나타났다.

또한 적정 수준의 의존도계수에 도달할 수 있는 효율적인 측정 조건을 탐색하기 위해 시행횟수, 영역, 그리고 문항의 수를 조정한 결과를 해석하면 다음과 같다. Table 8에 제시된 바와 같이 시행횟수를 1회로 하는 경우에는 적정 수준의 의존도계수에 도달하는 경우는 없는 것으로 나타났다. 따라서 최적의 시행횟수를 2회로 정하고 영역 수와 문항 수에 대한 효율적인 측정 조건을 탐색한 결과, 총 문항 수가 가장 작은 경우를 기준으로 원본인 210 문항에서 12개 영역의 영역별 8개 문항을 포함하는 총 96문항으로 축소하여도 적정 수준의 의존도계수에 도달하는 것으로 나타났다. 또는 원자료의 14개 영역을 모두 사용하는 경우를 기준으로 할 때는 문항 수를 7개로 줄여 총 98개 문항으로 축소할 수 있는 것으로 나타났다. 만일 영역과 문항을 개발할 때 상대적으로 영역 내 문항을 개발하는 경우가 더 용이하다는 점을 기준으로 정하면 이 때 효율적인 최적의 측정조건은 총 126개 문항인 9개 영역에 영역 별 14개 문항이 포함될 수 있는 것으로 나타났다.

Table 8. Dependability coefficients using D-Study $p \times O \times (I: D)$ design

(D, I)	Total	1 Occ	2 Occ	3 Occ	4 Occ	5 Occ
(12, 8)	96	0.780	0.848	0.873	0.887	0.895
(14, 7)	98	0.786	0.853	0.878	0.891	0.899
(13, 8)	104	0.787	0.854	0.879	0.892	0.900
(10, 11)	110	0.781	0.849	0.874	0.887	0.895
(11, 10)	110	0.784	0.852	0.877	0.890	0.898
(9, 14)	126	0.783	0.850	0.876	0.889	0.897
(14, 15)	210	0.817	0.881	0.905	0.917	0.925

Note. (D, I)=(# of domains, # of items); Total=# of total items; Occ=Occasion(s)

4. 결론 및 논의

본 연구에서는 교육부에서 표준안으로 보급한 교직 적성·인성 검사[14]결과를 활용하여 교육대학원의 예비공학교사 검사점수에 영향을 미치는 요인들의 상대적인 영향력을 탐색하고, 이를 바탕으로 적정 수준의 신뢰도에 도달할 수 있는 효율적인 측정 조건을 제시하였다. 주요 연구 결과를 바탕으로 일반화가능도 분석이 교육대학원 예비공학교사를 위한 교직 적성·인성 검사 도구를 개선하고 향상시킬 수 있는 방안에 대해 논하면 다음과 같다.

첫째, 교직 적성·인성 검사점수에 G-연구 $p \times i, p \times (i: d)$, 그리고 $p \times o \times (i: d)$ 설계를 적용하여 분석한 결과, 잔차 효과를 제외하면 피험자 효과가 두 번째 또는 세 번째로 큰 영향력을 미치는 것으로 나타났다. 이는 교직 적성·인성 검사점수에 교육대학원 예비공학교사의 교직 적성·인성 수준 차이가 반영되어 있다고 해석할 수 있다. 잔차 효과를 제외하면 G-연구 $p \times i$ 설계에서는 문항 효과가, $p \times (i: d)$ 설계 및 $p \times o \times (i: d)$ 설계에서는 영역 내 문항 효과를 나타내는 문항의 난이도에 따라 검사점수가 차이가 나는 것으로 나타났다. 또한 모든 G-연구 설계에서 본 연구 설계에서 고려하지 못한 국면들의 필요성을 간접적으로 설명하는 잔차 효과가 교직 적성·인성 검사점수에 가장 큰 영향력을 미치는 것으로 나타났다. 따라서 일반화가능도 분석은 교직 적성·인성 검사점수 차이가 교육대학원 예비공학교사의 교직 적성·인성 수준 차이 때문인지, 영역의 특성 때문인지, 영역 내 문항들의 난이도 차이 때문인지, 예비공학교사가 문항에 다르게 응답하기 때문인지 등을 포함하여 본 연구에서 고려하지 못한 추가적인 요인들이 필요한지를 밝히는데 적용될 수 있다. 즉, 예비공학교사를 대상으로 교직 적성·인성 검사를 사용할 때는 G-연구 분석 결과를 바탕으로 본 연구에서 고려하지 못한 피험자의 전공, 성별, 검사 시행시간, 그리고 시행연도 등과 같은 요인들이 고려되어야 함을 알 수 있다.

둘째, 다양한 G-연구 설계의 분석 결과를 비교한 결과에서 시행횟수 효과가 검사점수에 미치는 영향력은 상대적으로 작게 나타났으며, 영역 효과는 시행횟수 효과보다 크게 나타났다. 이처럼 설계에서 고려하는 국면에 따라 그 국면들의 상대적인 영향력이 달라진다는 것은 관찰점수 분산 대비 진점수 분산 비율로 산출되는 고전 검사이론을 바탕으로 한 신뢰도가 과대 추정될 수 있음

을 시사한다. 즉, 고전검사이론의 진점수 분산에 해당하는 피험자간 교육대학원 예비공학교사 분산이 가장 크게 나타났다가더라도 다른 국면들이 포함됨에 따라 다시 영향력이 재조정됨으로써 이 부분이 오차점수분산에 기여할 수 있기 때문이다.

셋째, 다양한 D-연구 설계의 분석 결과를 비교한 결과에서는 연구 설계에 국면이 늘어날수록 오차분산은 증가하며, 신뢰도인 의존도계수는 감소하는 것으로 나타났다. 특히 Cronbach's α 는 본 연구 설계에서 단일 오차요인만을 고려하는 D-연구 $p \times I$ 설계의 규준참조검사에 적합한 일반화가능도계수로 본 연구 설계에서 산출된 어떤 신뢰도보다도 높게 나타났다. 따라서 Cronbach's α 와 같이 측정 구조를 잘 반영하지 못하는 신뢰도를 교직 적성·인성 검사의 신뢰도로 제시하기 보다는 다중 오차요인을 고려하는 일반화가능도 분석을 바탕으로 준거참조검사에 적합한 의존도계수와 같은 보다 정밀한 신뢰도를 추정할 필요가 있다.

넷째, 국면이 많이 포함됨으로써 이론적으로 가장 적합한 D-연구 $p \times O \times (I:D)$ 설계를 분석한 결과 적정 수준의 신뢰도에 도달할 수 있는 교직 적성·인성 검사의 최적의 시행횟수는 2번인 것으로 나타났다. 따라서 일반화가능도 분석은 측정학적 특성 중 하나인 신뢰도를 바탕으로 그동안 교직 적성·인성 검사에 대해 교육부에서 공시한 2회 이상이라는 시행횟수에 대한 적절성에 대한 경험적 근거를 제공하는데 활용될 수 있다. 또한 시행횟수를 2번으로 고정한 상태에서 원자료인 14개 영역의 영역 별로 15문항을 포함하는 총 210개 문항을 영역 수와 문항 수에 대한 조절을 통해 효율적인 측정 조건을 3가지 기준에서 살펴본 결과는 다음과 같다. 최소 문항 수 기준, 원자료의 14개 영역을 모두 사용하는 기준, 그리고 용이성을 기준으로 교직 적성·인성 검사를 각각 12개 영역의 영역 별로 8개 문항을 포함하는 총 96개, 14개 영역의 영역 별로 7개 문항을 포함하는 총 98개, 그리고 9개 영역의 영역 별로 14개 문항을 포함하는 총 126개 문항으로 축소하여도 적정 수준의 신뢰도에 도달하는 것으로 나타났다. 이처럼 일반화가능도 분석은 교육부에서 교직 적성·인성 검사 표준안을 보급한 취지에 따라 각 교원양성기관의 특성을 반영하여 교직 적성·인성 검사를 자율적으로 개발하는데 유용하게 활용할 수 있다.

마지막으로 본 연구는 첫째, 한 교육대학원에 소속된

예비공학교사를 대상으로 분석이 수행되었으며 둘째, 새로운 연구 설계에서 고려해야 할 오차 요인에 대한 필요성만을 언급하였으며 셋째, 단변량 일반화가능도 분석을 수행함으로써 영역 별로 다른 문항들을 포함하는 문항 효과의 상대적인 영향력을 분석하지 못하였으며, 그리고 넷째, 시행횟수를 제외하면 최적의 측정 조건으로 문항 수와 영역 수에 대한 정보만을 제시한다는 한계점을 갖는다. 따라서 향후 예비공학교사 전체를 대표할 수 있는 연구대상을 표집하고, 본 연구 설계에서 제안한 오차 요인을 고려한 일반화가능도 분석을 수행하며, 다변량 일반화가능도 분석을 통해 영역 내 문항 효과를 밝히며, 그리고 다변량 일반화가능도 분석 결과를 바탕으로 문항 프로파일 분석을 수행하여 교육대학원 예비공학교사에 적합한 교직 적성·인성 검사를 구성하는 문항들을 탐색할 수 있는 지속적인 후속 연구들을 제안한다.

References

- [1] Ministry of Education, Character Education Promotion Law. Sejong: Ministry of Education, 2015.
- [2] I. J. Lee, "Systematic Approach of Character Education in School and Competences of Teachers", *Journal of Moral & Ethics Education*, vol. 39, pp. 271-300, 2016. DOI: <https://doi.org/10.18850/JEES.2016.39.11>
- [3] N. Kim, "A Critical Consideration on the Idea and Implication of the 'Character Education Promotion Law'", *Studies on Life and Culture*, vol. 39, pp. 55-102, 2016.
- [4] H. K. Park, "Discussions with the Issue of Character Education Promotion", *The Journal of Law of Education*, vol. 27, no. 3, pp. 77-100, 2015. DOI: <http://doi.org/10.17317/tjle.27.3.201512.77>
- [5] M. W. Kim, J. Y. Park, J. S. Choi, M. J. Kim, T. H. Kim, "The Effect of Social Emotion Learning on Teacher Character of Specialized Pre-service Teacher", *Journal of the Korean Institute of Industrial Educators*, vol. 42, no. 2, pp. 47-66, 2017.
- [6] Ministry of Education, Science and Technology, Enforcement Decree of the Elementary and Secondary Education Act, Seoul: Ministry of Education, Science and Technology, 2009.
- [7] C. R. Go, S. A. Bae, "Development of the Workplace Etiquette Education Program for Strengthen of Employment Capability of Specialized High School Students", *Journal of the Korean Institute of industrial educators*, vol. 36, no. 2, pp. 201-218, 2011.
- [8] K. M. Bae, "The Analysis of the Cognitive Level of Basic Job Skills of Specialized High School Based on NCS", *Journal of Educational Innovation Research*, vol. 27, no. 1, pp. 133-158, 2017.

- DOI: <https://doi.org/10.21024/pnuedi.27.1.201703.133>
- [9] M. W. Kim, T. H. Kim, "Analysis of Personality Level of Students of Industrial-field Specialized High Schools and Meister High Schools", *The Journal of Vocational Education Research*, vol. 36, no. 1, pp. 23-46, 2017. DOI: <http://doi.org/G704-000975.2017.36.1.004>
- [10] H. S. Lim, Y. C. Jyung, "Relationships among Creativity, Ego-identity and Career Decision Status of Vocational High School Students", *Journal of Agricultural Education and Human Resource Development*, vol. 42, no. 4, pp. 53-73, 2010. DOI: <https://doi.org/10.23840/agehrd.2010.42.4.53>
- [11] H. K. Kang, "Issues and Their Improvement of Specialized High School", *The Journal of Educational Research*, vol. 2, pp. 1-25, 2004.
- [12] C. S. Park, "Validation Study of Teaching Aptitude-Personality Test for Alternative Modification of the Test", *Journal of Holistic Convergence Education*, vol. 19, no. 3, pp. 107-124, 2015.
- [13] Ministry of Education, Science, and Technology. Official Approval Provision for Teacher Qualification. Seoul: Ministry of Education, Science, and Technology. 2012.
- [14] J. H. Kim, H. W. Nam, S. C. Yeom, J. Y. Im, The Development of Teaching Aptitude and Personality Test. Seoul: Ministry of Education, Science, and Technology, 2012.
- [15] S. Y. Kim, "Item Profile Analysis of the Personality Test for Teaching Profession", *Second Education Research*, vol. 65, no. 4, pp. 705-729, 2017.
- [16] W. J. Cho, "Validation and Modification of Teaching Aptitude Test for Pre-service Early Childhood Teachers", *Early Childhood Education & Care*, vol. 9, no. 2, pp. 101-123, 2014.
- [17] H. G. Won, S. J. Lee, "Validation of a Short Form of the Aptitude and Personality as a Teacher Scale for Preservice Early Childhood Teachers", *Early Childhood Education Research & Review*, vol. 21, no. 1, pp. 133-165, 2017.
- [18] S. H. Kim, Unqualified Candidates were only 0.6% based on Aptitude and Personality Test for Pre-service Teachers, Available From: <http://www.shinmoongo.net>, (accessed Dec., 15, 2017)
- [19] Korea Institute for Curriculum and Evaluation, A Study on Directions to Improve Official Approval Provisions for Teacher Qualification in a Competency-based Educational Environment. Seoul: Korea Institute for Curriculum and Evaluation, 2016.
- [20] Crick, J. E. Crick, R. L. Brennan, *Manual for GENOVA: A Generalized Analysis of Variance System. Research and Development Division*, American College Testing Program, 1983.
- [21] R. L. Brennan, *Generalizability Theory*, New York, NY: Springer-Verlag, 2001. DOI: <https://doi.org/10.1007/978-1-4757-3456-0>
- [22] S. B. Dunbar, D. M. Koretz, H. D. Hoover, "Quality Control in the Development and Use of Performance Assessments", *Applied Measurement in Education*, vol. 4, no. 4, pp. 289-303, 1992. DOI: https://doi.org/10.1207/s15324818ame0404_3
- [23] L. J. Fyans, *Generalizability theory: Inferences and practical applications*, Jossey-Bass Inc Pub, 1983.
- [24] R. J. Shavelson, G. P. Baxter, X. Gao, "Sampling Variability of Performance Assessments", *Journal of Educational Measurement*, vol. 30, no. 3, pp. 215-232, 1993. DOI: <https://doi.org/10.1111/j.1745-3984.1993.tb00424.x>
- [25] N. M. Webb, R. J. Shavelson, E. H. Haertel, "4 Reliability Coefficients and Generalizability Theory", *Handbook of statistics*, vol. 26, pp. 81-124, 2006. DOI: [https://doi.org/10.1016/S0169-7161\(06\)26004-8](https://doi.org/10.1016/S0169-7161(06)26004-8)
- [26] J. J. Norcini, Measurement Issues in the Use of Simulation for Testing Professionals: Test Development, Test Scoring, Standard Setting. In A. Tekian, C. H. McGuire, W. C. McGaghie, *Innovative Simulations for Assessing Professional Competence*, pp. 77-104, Chicago, IL: University of Illinois, 1999.
- [27] J. Cohen, *Statistical Power Analyses for the Social Sciences*, Hillsdale, NJ, Lawrence Erlbaum Associates, 1988.
- [28] L. J. Cronbach, R. L. Linn, R. L. Brennan, E. H. Haertel, "Generalizability Analysis for Performance Assessments of Student Achievement or School Effectiveness", *Educational and Psychological Measurement*, vol. 57, no. 3, pp. 373-399, 1997. DOI: <https://doi.org/10.1177/0013164497057003001>
- [29] A. Nußbaum, "Multivariate Generalizability Theory in Educational Measurement: An Empirical Study", *Applied Psychological Measurement*, vol. 8, no. 2, pp. 219-230, 1984. DOI: <https://doi.org/10.1177/014662168400800211>
- [30] R. J. Shavelson, N. M. Webb, *Generalizability theory: A primer*, Sage, 1991.
- [31] A. Vallevand, C. Violato, "A Predictive and Construct Validity Study of a High-stakes Objective Clinical Examination for Assessing the Clinical Competence of International Medical Graduates", *Teaching and learning in medicine*, vol. 24, no. 2, pp. 168-176, 2012. DOI: <https://doi.org/10.1080/10401334.2012.664988>

김 성 연(Sung-Yeun Kim)

[정회원]



- 1997년 2월 : 영남대학교 영남대학원 수학과 (이학석사)
- 2003년 8월 : 성균관대학교 성균관대학원 수학과 (통계학박사)
- 2011년 2월 : 연세대학교 연세대학원 교육학과 (교육학박사)
- 2016년 3월 ~ 현재 : 인천대학교 교육대학원 조교수

<관심분야>

교육측정 및 평가, 교원평가, 수학교육, 영재교육