

교육대학원과 사범대학 예비수학교사의 교직 적성·인성 검사에 대한 측정의 동등성 분석

김성연(인천대학교)

I. 서론

우리나라에서는 미래 사회를 살아갈 역량을 갖춘 인재를 육성하고자 수학과 교육과정에서는 2009개정 교육과정부터 창의적 사고 능력, 문제해결 능력, 정보처리 능력, 의사소통 능력 등을 핵심역량으로 규정하여 미래 사회에 대비한 교과 교육의 변화를 꾀하였다. 또한 2015개정 교육과정에서는 문제해결, 추론, 창의·융합, 의사소통, 정보처리, 태도 및 실천을 6가지 수학 교과 역량으로 규정(교육부, 2015)함으로써 사회 구성원의 역할을 성공적으로 수행할 수 있을 뿐만 아니라 수학적 역량과 더불어 수학에 대한 긍정적인 인식과 태도를 심어 주고자 하고 있다(김래영, 김은현, 2017).

그러나 교육의 변화를 통한 인재 양성이 실효를 거두기 위해서는 교사 교육이 선결되어야 하며, 많은 연구들에서는 이미 교사의 지식과 질이 학생들의 학습과 태도에 많은 영향을 미치고 있음이 밝혀졌다(Darling-Hammond, 2000; Hill et al., 2005; Kukla-Acevedo, 2009; Wayne & Youngs, 2003). 특히 중등교육에서 교사는 학생들에게 직접적이고 많은 영향을 미치는 존재이며, 수학교사의 신념, 교육관, 교수학습 방법 등은 학생들의 수학 성취도를 높일 뿐만 아니라 수학에 대한 관심과 흥미, 자신감 및 가치 인식과 같은 긍정적인 태도를 갖게 한다고 알려져 있다(김성연, 2018a; 김정환, 2004; 김현진, 2013; 박정, 2007; 이현숙, 송미영, 2015; 한혜숙, 최계현, 2011; 황혜정, 2011;

Houchard, 2005; Klieme et al., 2009; Kunter et al., 2008; Thompson, 1992).

한편 학교폭력 문제의 심화, 육설의 일상화 등의 문제를 해결하고 학생들의 인성 및 사회성을 함양할 수 있는 교육적 변화를 위해서도 교사들의 변화가 요구되어진다. 즉, 학생들의 바른 인성 함양을 위해서 올바른 가치관과 인성을 갖춘 교사의 지도가 필요하기 때문이다. 수학교육에서도 인지적 측면에서의 교육과 더불어 품성, 자질, 도덕성과 같은 인성이 최근 강조(신준국 외, 2015; Narvaez & Nucci, 2008)되고 있으며, 중등교육에서 수학을 담당할 예비교사가 이러한 교직 적성 및 인성(이하 “교직 적·인성”이라고 함)을 갖춘 교사를 양성하기 위해서는 교원양성기관에서 예비수학교사를 선발할 때부터 교직 적성·인성을 고려한 평가가 강조되어야 한다. 이에 2013년부터 교육부는 교직 적·인성을 갖춘 교사를 선발하기 위해 모든 교원양성기관에서 2회 이상 교직 적·인성 검사 실시를 의무화하고 있으며, 교원양성기관에서 자율적으로 개발하여 활용할 수 있도록 교직 적성·인성 검사 표준안(김정환 외, 2012, 이하 “교직 적·인성 검사”라고 함)을 보급하였다.

교직 적·인성 검사는 중등교육에서 수학 교과를 담당할 예비수학교사들이 교직 적·인성을 갖추고 있는지를 평가함으로써 예비수학교사의 교육자적 교직 적·인성 수준을 진단하고, 교직 적·인성 하위영역 중 부족한 부분을 도와줄 수 있다는 점에서 이 검사의 결과는 매우 중요한 역할을 한다. 그러나 검사 결과를 활용하기에 앞서 자기보고식 검사로 수집되는 자료에는 피험자의 배경변인 또는 성향에 의해 응답 결과가 왜곡되거나 검사의 목적을 인지하고 의도적으로 좋게 응답하는 등의 문제가 발생할 수 있다. 또한 이와 같은 문제가 체계적인 방식으로 검사점수에 개입될 때 측정 오차가 커지면서 측정의 정확성 또는 신뢰도에 부정적인 영향을

* 접수일(2018년 4월 27일), 수정일(2018년 5월 11일), 게재확정일(2018년 5월 23일)

* ZDM분류 : B55

* MSC2000분류 : 97C40

* 주제어 : 교직 적성·인성 검사, 예비수학교사, 의존도계수, 측정의 동등성

조래할 수 있다. 현재까지 수학 성취도 검사에서는 표준화 과정을 거쳐 대규모로 개발된 검사일지라도 국가, 인종, 성별, 언어 등 피험자의 배경변인에 따라 동일한 검사에 대한 응답양식이 다르게 나타나고 있음을 경험적으로 밝힘으로써 검사 결과에 응답편과 효과가 개입되었음이 밝혀졌다(Abedi & Lord, 2001; Li & Brennan, 2007, Möttus et al., 2012). 따라서 검사 결과 활용에 앞서 피험자의 배경변인 및 특성에 관계없이 정의적 영역에서 측정이 일관적이며 정확하게 이루어졌는지에 대해 확인할 필요가 있다. 또한 교직 적·인성 검사가 교직을 이수하는 모든 학생들에게 동일하게 사용할 수 있는 공정하고 타당한 평가인지에 대해서 의의가 제기되었다. 이는 교직 적·인성 수준이 학교 급별에 따라 다르기 때문이다. 구체적으로 안도희와 김유리(2015)는 교육대학원생이 사범대학 학생들보다 전반적으로 교직 핵심역량이 높으며, 스스로를 교사로서 좋은 자질을 가지고 있고, 좋은 교수법을 알고 있으며, 교직이 자신의 능력에 맞는 직업이라고 지각하며, 교직전문성을 향상시키고자 하는 열망이 높다고 밝혔다. 또한 교육대학원생들과 사범대학 학생들의 교육수준 및 교원자격증 취득에 대한 목적 지향의 관점에 차이가 있다고 밝혔다(이선정, 신혜원, 2008). 이처럼 교육대학원과 사범대학 집단 간의 차이를 밝힌 연구들에서는 동일한 검사가 시행되고 있지만, 집단 별 신뢰도의 차이에 대해서는 많은 연구들이 간과하고 있다. 그러나 낮은 신뢰도는 예측 타당도를 포함한 다른 타당도를 낮추는 잠재적인 원인이 될 수 있으므로 집단 별 신뢰도의 차이를 검증하는 것은 매우 중요하다(Lakin & Lai, 2012). 특히 검사의 신뢰도와 관련하여서 교전검사이론을 바탕으로 한 Cronbach's α 나 채점자간 일치도처럼 오차 요인을 하나로만 간주하는 경우에는 검사점수에 다양한 오차요인을 고려하는 일반화가능도 이론과 달리 과대 추정된 신뢰도가 산출될 수 있다는 지적이 제기되어 왔다(김성숙, 김양분, 2001; 이규민, 황경현, 2007; Brennan, 2001a; Cronbach et al., 1997; Lee & Frisbie, 1999; Shavelson & Webb, 1991).

따라서 본 연구에서는 교육대학원과 사범대학의 예비수학교사를 대상으로 일반화가능도 이론의 신뢰도인 의존도계수를 바탕으로 학교 급별에 따라 교직 적·인성

검사가 교직 적·인성을 측정하는 정확도에 차이가 있는지를 공정성 측면에서 탐색하고자 한다. 또한 교직 적·인성 검사 결과 활용방안으로 학교 급별에 따라 예비수학교사의 교직 적·인성에서 차이가 나는 하위영역을 중심으로 교직 적·인성 수준을 증진시킬 수 있는 방안을 모색하고자 한다. 구체적인 연구문제는 다음과 같다.

첫째, 예비수학교사의 교직 적·인성 수준은 학교 급별에 따라 차이가 있는가?

둘째, 예비수학교사의 교직 적·인성 검사의 하위영역 점수에 영향을 미치는 요인들의 상대적인 영향력은 학교 급별에 따라 차이가 있는가?

셋째, 예비수학교사의 교직 적·인성 검사의 신뢰도는 학교 급별에 따라 차이가 있는가?

넷째, 예비수학교사의 교직 적·인성 검사의 합성점수에 하위영역 별 가중치가 미치는 영향력은 학교 급별에 따라 차이가 있는가?

II. 이론적 배경

1. 교육대학원과 대학에서 예비교사 비교 연구

예비교사를 대상으로 이들이 소속되어 있는 학교 급별에 따라 교육대학원과 사범대학 집단으로 구분하여 수행된 연구들을 살펴보면 다음과 같다.

백인순(2000)은 교육대학원생들 대부분이 교육과정의 수준을 대학생보다 높게 인식한다고 하였으며, 이선정과 신혜원(2008)은 교육대학원생은 자신의 전공을 통해 얻은 지식과 역량을 교사로서의 전문성을 개발하기 위해 교원자격증을 취득한다고 밝혔다. 반면에 안재희와 이숙정(2012)은 대학생의 경우 다양한 교직과정을 이수하면서 교직에 대한 적성과 흥미 등을 점검하고 난 후 최종적으로 교원자격증을 취득한다고 밝혔다. 안도희와 김유리(2015)는 교직이수 학생들을 대상으로 학교 급별에 따라 교직핵심역량, 교수능력, 교직전문성 열망에 차이가 있는지, 이들이 지닌 교직핵심역량과 교수능력이 교직전문성 열망에 영향을 미치는지에 대해 분석하였다. 분석결과, 교육대학원생들이 대학생들에 비해 교직핵심역량과 교수능력, 그리고 교직전문성에 대한 열망이 전문성 개발요인에서 높게 나타났지만, 학교 급별에 상관없이 교직핵심역량은 이들의 교직전문성 열망에 직

접적으로 긍정적인 영향을 미치며, 교수능력이 교직핵심역량과 교직전문성 열망 간의 관계를 매개한다고 밝혔다. 오유진(2016)은 영어교과에서 2014학년도 이후 개정된 임용시험의 서답형 쓰기 답안에 대한 학교 급별에 따른 예비교사들의 인식을 비교하였다. 분석결과, 임용고사의 변경으로 인한 재정적, 심리적인 부담감에 대해 교육대학원 출신의 수험생들이 사범대학 출신의 수험생들보다 더 큰 부담을 느낀다고 밝혔으나 졸업한 학교 급별에 상관없이 3단계에서 2단계로 개정된 임용제에 대해서는 적절하게 느낀다고 밝혔다. 또한 분석결과를 바탕으로 학교 급별에 상관없이 교원양성기관에서는 서답형 영어 쓰기 시험에 대비한 과목을 개설해야 할 것을 제안하였다. 김성연(2018b)은 교직 적·인성 검사를 교육대학원과 사범대학에서 동일하게 실시하는 것이 적절한지를 단변량 일반화가능도 이론을 적용하여 분석하였다. 분석결과, 교직 적·인성 검사는 신뢰도 측면에서 현행대로 학교 급별에 상관없이 동일하게 적용할 수 있다고 밝혔다.

이상에서 살펴본 바와 같이 학교 급별에 따른 예비교사들을 대상으로 일반화가능도 이론을 적용한 연구들은 많지 않다. 특히 김성연(2018b)에서는 교육대학원과 사범대학 집단에 교직 적·인성 검사를 동일하게 적용할 수 있는지를 교직 적·인성 검사의 하위영역이 독립이라는 가정 하에 단변량 일반화가능도 이론을 적용하여 분석하였다는 한계를 지적하면서, 후속연구로 교직 적·인성 검사의 총점을 바탕으로 적격과 부적격을 판정하는 교육 현장의 실태를 감안하여 하위영역 간의 상관관계를 고려한 다변량 일반화가능도 이론을 적용할 것을 강조하였다. 따라서 본 연구에서는 상대적으로 연구가 수행되고 있지 않은 수학 교과에서 학교 급별에 따른 예비교사들을 대상으로 동일한 교직 적·인성 검사를 실시하는 것이 공정한지, 그리고 학교 급별에 따라 예비수학교사의 교직 적·인성 수준은 어떠한지를 탐색하고자 한다. 본 연구의 결과는 예비수학교사들의 교직 적·인성 증진을 위한 구체적인 정책 수립의 기초 자료로 유용하게 활용될 수 있을 것으로 기대한다.

2. 교직 적·인성 검사를 분석한 연구

본 연구에서 다루고 있는 교직 적·인성 검사의 분석

을 통해 교직 적·인성을 비교 및 예측하는 연구와 검사의 개발 및 타당화에 대한 연구로 나누어 살펴보면 다음과 같다.

김진영과 이기중(2013)은 예비유아교사의 교직 적·인성을 셸프리더십의 하위전략 중 행동중심전략과 자연보상전략이 예측할 수 있다고 밝혔다. 구체적으로 행동중심전략은 문제해결력·탐구력, 지도성, 공감·포용력, 봉사·희생·협동, 소명감·교직관율, 그리고 자연보상전략은 심리적 안정성, 공감·포용력을 예측할 수 있다고 밝혔다. 원혜경과 이소정(2015)은 예비영유아교사의 교직 적·인성을 재학기간, 정규보육실습 수강여부, 교직진로 희망여부에 따른 차이를 검정하였다. 분석결과, 정규보육실습 수강집단이 미수강 집단에 비해, 그리고 교직진로 희망집단이 교직외 진로 희망집단에 비해 교직 적·인성이 통계적으로 유의하게 높게 나타났다고 밝혔다. 전경희와 조안나(2017)는 입학전형 유형을 고교 교과 성적 및 수능 위주의 일반전형과 면접 및 적성고사 등 인성적 자질 요소를 평가에 반영한 수시전형으로 나누어 교직 적·인성 특성에 대한 구조적 관계를 탐색하였다. 분석결과, 입학전형 유형과 관계없이 교직 적·인성 하위영역 중 언어·의사소통력과 지도성이 교직 적·인성을 예측하는데 가장 설명력이 높았으며, 독립성·자주성과 심리적 안정성을 제외한 모든 하위영역이 교직 적·인성을 예측한다고 밝혔다.

교직 적·인성 검사의 개발 및 타당화에 대한 연구는 대부분 예비유아교사 및 교육대학원생을 대상으로 수행되었다. 이는 교육부에서 표준안으로 보급한 교직 적·인성 검사의 타당화 과정에 교육대학, 사범대학, 또는 교직과를 이수중인 대학생을 연구대상으로 한정(김정환 외, 2012)함으로써 예비유아교사 및 교육대학원생이 포함되지 않았기 때문이다. 예비유아교사들을 대상으로 조운주(2014)는 14개 하위영역의 총 210문항의 교직 적·인성 검사를 9개 하위영역의 총 140문항으로 축소할 것을 제안하였다. 또한 원혜경과 이소정(2017)은 교직 적·인성 검사의 14개 영역에 민감성을 추가하여 15개 하위영역의 총 150문항으로 축소한 후, 각각 75씩 문항으로 나누어 A세트와 B세트의 복수의 척도를 개발하였다. 교육대학원생을 대상으로 김성연(2017a)은 교직 인성과 관련한 영역만을 선정하여 교직 적·인성 검사의 7개 영

역 총 78문항을 총 72개 문항으로 축소할 것을 제안하였다.

이처럼 교직 적·인성 검사는 2013년 교원양성기관에 배포된 이후에 다양하게 연구가 진행되고 있지만, 집단 간 측정의 동등성을 분석하지 않은 상태에서 교직 적·인성을 비교하고 있는 실정이다. 따라서 본 연구에서는 상대적으로 연구가 적게 수행되어온 예비수학교사들을 대상으로 교육대학원과 사범대학 집단에서 교직 적·인성 검사결과에 대해 측정의 동등성을 분석하고자 한다. 본 연구의 결과는 교직 적·인성 검사를 현행대로 예비수학교사의 학교 급별에 상관없이 동일하게 적용할 수 있는지에 대한 경험적 근거를 제시할 수 있다.

3. 수학 교과를 기반으로 수행된 일반화가능도 분석

수학 교과와 관련하여 일반화가능도 분석을 수행한 연구를 국내와 국외로 나누어 살펴보면 다음과 같다.

국내에서 김명화(2005)는 시험의 교육적 효과를 높이기 위하여 고등학교 공통수학 영역에서 구성형 문항을 컴퓨터로 채점하는 채점 자동화 시스템을 구축하고 이 시스템의 신뢰도를 단변량 일반화가능도 이론을 적용하여 분석하였다. 분석결과, 수학 구성형 채점 자동화 시스템에 영향을 미치는 효과는 잔차요인, 문항요인, 그리고 피험자와 채점자의 상호작용요인 순으로 나타났다고 밝혔다. 또한 수학 구성형 채점 자동화 시스템이 적정 수준의 신뢰도에 도달하기 위해서 문항 수는 8개, 그리고 채점자 수는 4명이 필요하다고 밝혔다. 김성연(2016)은 인지적 측면에 초점을 맞춘 수학적 창의성 검사를 유창성, 융통성, 독자성으로 채점한 결과에 단변량 일반화가능도 분석을 수행하여 신뢰도를 가장 높일 수 있는 최적의 채점 요소별 가중치를 탐색하였으며, 수학적 창의성은 다차원적인 요소로 구성되어 있다고 밝혔다. 이문수와 차동춘(2016)은 수학 창의력 문제해결검사에서 수행형 문항 채점의 신뢰도 향상을 위한 측정조건을 탐색하기 위하여 단변량 일반화가능도 이론을 분석하였다. 분석결과, 수행형 문항 채점에 영향을 미치는 효과로 피험자와 문항의 상호작용요인은 크게 나타난 반면, 채점자 요인과 채점자와 관련된 상호작용요인은 상대적으로 작게 나타났다고 밝혔다. 또한 수학 창의력 문제해결검사가 적정 수준의 신뢰도에 도달하기 위해서는

이 연구에서 고려한 문항 수인 5개를 10개로 늘려야 한다고 밝혔다. 김성연(2017b)은 수학적 창의성 태도 검사에서 수학영재 집단과 일반학생 집단으로 구분하여 다변량 일반화가능도 분석을 수행하였다. 분석결과, 수학적 창의성 태도 검사는 신뢰도를 기준으로는 일반학생 집단에서, 조건부 측정 오차를 기준으로는 수학영재 집단에서 더 적합한 것으로 나타났으며, 신뢰도를 향상시키기 위해서 수학영재 집단에서는 수렴적 태도 영역을 높게, 반면에 일반학생 집단의 경우에는 발산적 태도 영역과 문제해결 태도 영역을 높게 반영하여야 한다고 밝혔다. 또한 수학적 창의성 태도 검사가 적정 수준의 신뢰도에 도달하기 위해서 수학영재 집단에서는 수학적 창의성 태도의 다차원 영역을 반영하여야 하는 반면 일반학생 집단에서는 단일차원의 영역만으로도 가능하다고 밝혔다. 김성숙과 전경희(2018)는 국가수준 학업성취도 평가의 정의적 영역에서의 성취수준별 측정의 동등성을 탐색하기 위해 다변량 일반화가능도 분석을 수행하였다. 분석결과, 자신감, 흥미, 가치인식으로 구성된 수학 교과에 대한 정의적 영역에서 가치인식은 수학 성취수준이 높은 집단과 낮은 집단 간에 측정의 동등성을 담보하기에는 다소 부정적인 결과가 나타났으며, 성취수준에 상관없이 측정 요소별로 현행 4문항에서 5문항으로 증가할 때 신뢰도 향상 정도가 가장 크다고 밝혔다.

국외의 경우 측정 대상이 학생인 경우와 교사인 경우로 나누어서 살펴보면 다음과 같다. 먼저 학생의 경우에 Baxter et al.(1993)는 측정, 자릿값, 그리고 확률로 구성된 수학 수행평가 점수에, 그리고 Lane et al.(1996)는 수학적 문제해결, 추론, 의사소통으로 구성된 수학 수행평가 점수에 크게 영향을 미치는 오차 요인은 채점자가 아니라 과제라고 밝혔으며, 신뢰도를 향상시키기 위해서는 채점자 수보다는 과제 수를 증가시켜야 한다고 밝혔다. 또한 Shavelson et al.(1993)는 수학 수행평가에서 과제, 시행횟수, 채점자, 그리고 측정유형을 고려하여 일반화가능도 분석을 수행한 결과 수학 수행평가 점수의 측정 오차에 가장 큰 영향을 미치는 요인은 학생 수준과 학교 수준에서 모두 과제 요인이었으며, 수학 수행평가가 적정 수준의 신뢰도에 도달하기 위해서는 과제의 수가 많아야 한다고 밝혔다. Klein et

al.(1995)는 주에서 2년간 실시된 대규모의 수학 포트폴리오 평가에서 채점자간 일치도는 낮은 편이라고 밝혔다. 또한 일반화가능도 이론을 적용하여 신뢰도를 분석한 결과, 신뢰도를 향상시키기 위해서는 수학 포트폴리오에 포함되는 주제를 늘리거나 채점자를 늘리는 방법이 있을 수 있으나 이는 현실적이지 않으므로 수학 포트폴리오 평가 시 표준화된 평가 시스템의 결과를 함께 이용할 것을 제안하였다. McBee & Barnes(1998)는 수학 수행평가에서 시간측면에서의 안정성과 과제간 일관성을 일반화가능도 이론을 적용하여 분석하였다. 분석결과, 채점자간 일치도는 높게 나타났지만, 검사-재검사 신뢰도는 낮게 나타났다고 밝혔다. 또한 복잡한 문제해결 영역을 포함하는 수학 수행평가가 적정 수준의 신뢰도에 도달하기 위해서는 비슷한 과제를 사용하는 경우에도 과제 수가 많아야 한다고 밝혔다. Christ et al.(2005)는 수학에서의 연산 능력을 평가하는데 평가 시간에 따른 신뢰도를 단변량 일반화가능도 이론을 적용하여 분석하였다. 분석결과, 기준참조평가에 적합한 신뢰도인 일반화가능도계수가 적정 수준에 도달하기 위한 측정 조건을 개인 중재와 집단 중재가 투입되는 경우로 나누어 살펴보았다. 구체적으로 개인 중재가 투입되는 경우에 인지부하가 낮은 결정에는 1분, 그리고 인지부하가 높은 결정에는 4분이 필요하다고 밝혔다. 반면에 집단 중재가 투입되는 경우에 준거참조평가에 적합한 신뢰도인 의존도계수가 적정 수준에 도달하기 위해서 인지부하가 낮은 결정에는 4분, 그리고 인지부하가 높은 결정에는 13분이 필요하다고 밝혔다. Mastergeorge & Martinez(2010)는 주에서 대규모로 실시하는 수학 수행평가에 교사들이 채점하게 되는 학생들이 일반학생 집단인지 장애학생 집단을 알려주지 않는 경우와 알려주는 경우로 나누어 일반화가능도 이론을 적용하였다. 분석결과, 교사들은 채점하는 학생들의 장애상태를 알고 있어도 평균적으로는 일반학생과 다르게 수학 수행평가를 채점하지는 않는 것으로 나타났지만, 장애학생 집단을 평가하는 경우에 점수의 변동이 크게 나타남으로써 일반학생 집단과 다르게 채점할 수도 있다고 밝혔다.

다음으로 수학교사를 측정 대상으로 하는 경우에 Matsumura et al.(2006)는 수학교사들의 수업을 2번 관

찰한 점수를 단변량 일반화가능도 이론을 적용하여 분석하였다. 분석결과, 관찰횟수에 따라 수학교사들의 관찰점수의 차이는 거의 존재하지 않았으며, 관찰점수가 적정 수준의 신뢰도에 도달하기 위해서는 기준참조평가인지 또는 준거참조평가인지에 관계없이 모두 2번의 관찰만으로도 충분하다고 밝혔다. Hill et al.(2012)는 수학교사를 관찰하는데 있어 수업의 첫 30분만을 내용영역으로 간주하고, 내용영역과 채점자 수를 조절하면서 적정 수준의 신뢰도에 도달할 수 있는 효율적인 측정조건을 탐색하였다. 분석결과, 3개의 내용영역과 2명의 채점자, 또는 4개의 내용영역과 4명의 채점자 등이 필요하다고 밝혔다. Ho & Kane(2013)은 자발적으로 지원한 수학교사에 대해 이들이 60분 동안 수업한 장면을 1명의 채점자가 평가하는 경우와 15분씩 나누어서 4명의 채점자가 평가하는 경우의 신뢰도를 분석하였다. 분석결과, 적정 수준의 신뢰도를 얻기 위한 효율적인 측정조건으로 60분 비디오를 1명의 채점자가 평가하는 것보다 15분씩 나누어 비디오를 4명의 채점자가 평가할 것을 제안하였다. Wilhelm & Kim(2015)은 다변량 일반화가능도 분석을 수행하여 수학교사들의 수업을 관찰하는데 많이 사용되고 있는 교수 질 평가도구가 적정 수준의 신뢰도에 도달하기 위해 필요한 관찰횟수를 탐색한 결과, 최소 3번의 관찰횟수가 필요하다고 밝혔다.

이상에서 살펴본 것처럼 수학 교과와 관련하여 일반화가능도 이론을 적용하여 분석한 연구는 대부분 인지적 영역에 초점을 맞추어 단변량 일반화가능도 분석이 수행되고 있으며, 정의적 영역을 고려한 연구는 현재까지 김성숙과 전경희(2018)를 제외하면 거의 수행되고 있지 않는 실정이다. 또한 기존에 일반화가능도 분석을 수행한 연구들은 대부분 적정 수준의 신뢰도에 도달할 수 있는 측정조건들을 탐색하고 있으며, 서로 다른 집단에서 동일한 검사를 실시해야하는지에 대한 공정성 측면에서 연구가 수행된 경우는 한정적이다. 따라서 본 연구에서는 학교 급별에 따른 교직 적·인성 수준의 차이를 살펴봄과 동시에 신뢰도를 바탕으로 공정성 측면에서 두 집단에 동일한 교직 적·인성 검사를 실시하는 것이 적절한지를 살펴보고자 한다.

III. 연구방법

1. 분석 자료

본 연구는 교직 적·인성 검사가 의무화된 2013년부터 최근 2017년까지 교육부에서 표준안으로 보급한 교직 적·인성 검사를 수도권외의 한 교육대학원과 사범대학의 예비수학교사에게 실시한 자료를 분석 대상으로 정하였다. 교직 적·인성 검사는 1학기에는 5월, 그리고 2학기에는 11월의 마지막 주 목요일에 교육부에서 보급한 김정환 외(2012)에서 제시된 지시방법에 따라 40분간 지필검사로 실시되었으며, 한 번 응답한 답안지는 고치거나 수정할 수 없다고 안내한 후 시행되었다. 채점 역시 김정환 외(2012)에서 제시된 채점방법에 따라 일부 문항들의 응답은 역코딩을 하였고, 정직성과 일관성 척도 문항들은 쌍을 이루는 두 문항의 응답이 동일한지 또는 두 문항의 응답의 합이 5점이 되는지의 여부에 따라 1점부터 4점까지 응답된 원자료를 0점부터 4점까지 재코딩 하였으며, 무응답의 경우는 0점으로 처리하였다. 또한 교직 적·인성 검사는 학생들의 입학 연도에 따라 1회 또는 2회가 실시되었으므로, 본 연구에서는 연습 효과를 배제하기 위해 2회 검사를 수행한 피험자의 경우에는 1회 결과만을 분석 자료로 한정하였다.

[표 1] 학교 급별 시행시기와 성별에 따라 교직 적성·인성 검사가 시행된 학생의 수 및 비율

[Table 1] Frequency and ratio of respondents by genders and school levels from 2013 to 2017

| 년도/집단 | 교육대학원 | 사범대학 | 합계 | |
|-------|-------------|--------------|--------------|------------|
| 2013 | 남 | 6명(16.7%) | 15명(13.5%) | 21명(14.3%) |
| | 여 | 13명(36.1%) | 26명(23.4%) | 39명(26.5%) |
| 2014 | 남 | 2명(5.6%) | 4명(3.6%) | 6명(4.1%) |
| | 여 | 0명(0.0%) | 0명(0.0%) | 0명(0.0%) |
| 2015 | 남 | 2명(5.6%) | 12명(10.8%) | 14명(9.5%) |
| | 여 | 3명(8.3%) | 8명(7.2%) | 11명(7.5%) |
| 2016 | 남 | 2명(5.6%) | 8명(7.2%) | 10명(6.8%) |
| | 여 | 2명(5.6%) | 13명(11.7%) | 15명(10.2%) |
| 2017 | 남 | 6명(16.7%) | 16명(14.4%) | 22명(15.0%) |
| | 여 | 0명(0.0%) | 9명(8.1%) | 9명(6.1%) |
| 합계 | 36명(100.0%) | 111명(100.0%) | 147명(100.0%) | |

학교 급별에 따라 교직 적·인성 검사를 수행한 학생들을 시행시기와 성별에 따라 [표 1]에 빈도 및 비율을 제시하였다. 총 학생 수는 147명으로 이 중 교육대학원

생은 36명인 약 25%, 사범대학 학생은 111명으로 약 75%인 것으로 나타났다. 전체 학생 중 남학생과 여학생은 각각 73명과 74명인 50%로 비슷하게 나타났으며, 이는 학교 급별로 구분해서 분석한 경우에도 마찬가지로 교육대학원의 남학생과 여학생은 동일하게 18명으로, 그리고 사범대학의 남학생과 여학생은 각각 55명과 56명으로 비슷하게 나타났다. 또한 2013년에 교직 적·인성 검사에 응시한 학생은 60명인 약 41%로 다른 해에 비해 가장 높게 나타났다. 이는 2013년이 교직 적·인성 검사가 최초로 보급된 해였으며, 다음 해부터 당해 연도 입학자를 대상으로 검사를 실시했던 것과 달리 2013년에는 재학 중인 학생들 전체를 대상으로 검사가 시행되었기 때문이라고 해석할 수 있다.

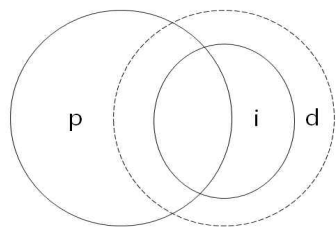
2. 분석 방법

본 연구는 교육대학원과 사범대학 예비수학교사에게 교직 적·인성 검사를 실시한 자료에 학교 급별로 다변량 일반화가능도 분석을 수행하였다. 다변량 일반화가능도 분석은 단변량 일반화가능도 분석의 확장된 형태로 모든 피험자가 2개 이상의 고정된 국면에서 전집점수를 가지고 있으며, 국면들은 서로 상관이 있고, 각 전집점수는 관련이 있는 영역들 중 하나에만 대응될 때 사용하는 분석방법이다. 다변량 일반화가능도 분석은 다변량 일반화가능도 연구(Generalizability study, G-연구)와 다변량 결정 연구(Dependability study, D-연구)로 나누어 수행하였으며, 다변량 G-연구에서는 교직 적·인성 검사점수에 각 국면들이 미치는 상대적인 영향력과 각 하위영역 간 측정의 오차를 고려한 상관계수를 파악하였으며, 다변량 D-연구에서는 신뢰도인 의존도계수와 가중치를 분석하였다. 다변량 일반화가능도 분석의 용어로 전집점수(universe score)는 고전검사이론의 진점수(true score)로, 그리고 국면(facet)과 수준(level)은 분산분석의 요인(factor)과 수준(level)으로 해석할 수 있다. 또한 다변량 일반화가능도 분석은 고전검사이론을 바탕으로 한 신뢰도인 Cronbach's α 가 규준참조 검사에만 적용될 수 있는 것과 달리 규준참조검사와 준거참조검사를 구분하여 신뢰도를 제시한다. 즉, 규준참조검사에서는 상대오차 분산을 활용한 일반화가능도계수, 그리고 준거참조검사에서는 절대오차 분산을 활용

한 의존도계수를 제시한다. 단, 오차 국면이 하나인 경우에 일반화가능도계수는 고전검사이론이 제시하는 Cronbach's α 와 동일하다. 한편 가중치와 관련하여 다변량 일반화가능도 분석은 상대 가중치(relative weight), 명목 가중치(nominal weight), 그리고 실질 가중치(effective weight)를 각각 제시한다. 상대 가중치와 명목 가중치는 문항 수에 비례하여 설정되며, 가중치의 합을 1로 정하는지 또는 검사의 총점으로 정하는지에 따라 분류한다. 반면에 실질 가중치는 고정 국면이 실제 합성 전집점수 분산 및 오차 분산에 기여한 정도를 나타낸다(Brennan, 2001a).

1) 다변량 G-연구 설계

본 연구에서는 교직 적·인성 검사의 하위영역은 고정(fixed) 국면으로, 하위영역 별로 서로 다른 문항은 문항 전집에서 표집된 임의(random) 국면으로, 그리고 피험자는 학교 급별에 따라 교육대학원과 사범대학으로 구분되어 있는 피험자 모집단에서 임의 추출된 것으로 가정하여 두 집단을 구분하여 분석을 수행하였다. 구체적으로 학교 급별에 상관없이 모든 피험자(p)들은 14개 하위영역(d)에 하위영역 별 15개 문항을 포함하는 총 210개의 문항(i)으로 구성된 동일한 교직 적·인성 검사에 응답하였으므로 다변량 G-연구 $p \times i \times d$ 설계를 적용하였으며, 이를 도식화하면 [그림 1]과 같다.



[그림 1] 다변량 G-연구 $p \times i \times d$ 설계
 [Fig. 1] Multivariate G-study for the $p \times i \times d$ design

Brennan (2001a)의 기호에 따라 피험자 위에 검은색으로 채워진 원은 피험자 국면이 고정 국면인 하위영역 국면의 모든 수준과 교차함을 나타내며, 문항 위에 비어있는 원은 문항 국면이 고정 국면인 하위영역의 각 수준에 내재되어 있음을 표시한다. [그림 1]에서 다변량

G-연구 $p \times i \times d$ 설계는 단변량 G-연구 $p \times (i : d)$ 설계를 기본으로 그려지지만, 하위영역(d) 국면이 각 수준에 따라 독립적으로 추정됨을 강조하기 위해 점선으로 표현하였다는 점에서 차이가 있다. [그림 1]에서 실선으로 표시된 원에 의해 분할되는 부분은 분산 성분이, 그리고 실선과 점선이 교차하는 부분에는 공분산 성분과 측정의 오차를 고려한 상관계수가 존재한다. 본 연구에서는 해석의 편의를 위하여 각 국면의 분산 성분과 측정의 오차를 고려한 상관계수만을 제시하였다.

2) 다변량 D-연구 설계

다변량 D-연구는 다변량 G-연구에서와 같은 $p \times I \times d$ 설계로 분석을 수행하였다. 다변량 D-연구에서 각 오차 국면은 다변량 G-연구의 경우 어떤 피험자가 특정 영역의 특정 문항에 응답한 점수를 분석하는 것에 반해 평균에 대한 의미가 포함되어 있기 때문에 피험자를 제외한 국면은 대문자로 표시한다는 점에서 차이가 있다. 다변량 G-연구 분석결과 산출되는 각 국면의 분산과 공분산 성분 추정치를 바탕으로 다변량 D-연구에서는 하위영역 별로 Cronbach's α 와 의존도계수를 산출하였다. 또한 실질 가중치와의 비교를 용이하게 하기 위하여 본 연구에서는 상대 가중치만을 제시하였다. 즉, 교직 적·인성 검사는 하위영역 별로 동일하게 15개의 문항이 포함되어 있으므로 상대 가중치는 0.071로 정하였다. 이상의 모든 분석에는 mGENOVA(Brennan, 2001b)와 R 프로그램을 활용하였다.

VI. 연구 결과

1. 학교 급별 교직 적·인성 검사 기술통계 분석결과
 교직 적·인성 검사의 총점 및 하위영역별 점수에 대한 Cohen's d 와 분산비를 포함한 기술통계 분석결과를 예비수학교사의 소속집단인 교육대학원과 사범대학으로 나누어 [표 2]에 제시하였다. 여기서 Cohen's d 는 효과 크기를 나타내며 교육대학원 집단의 평균에서 사범대학 집단의 평균을 뺀 후, 두 집단의 통합표준편차(pooled standard deviation)로 나눈 값으로, 값이 0.2이하이면 작은 효과크기, 0.2~0.5이면 중간 효과크기, 그리고 0.8

이상이면 큰 효과크기를 나타낸다고 해석한다(Cohen, 1988). 또한 분산비는 사범대학 집단의 분산을 교육대학원 집단의 분산으로 나눈 값으로 1보다 큰 경우, 사범대학 집단의 변동량이 더 크다고 해석한다.

[표 2] 학교 급별 교직 적성·인성 검사의 총점 및 하위 영역별 점수의 기술통계
[Table 2] Descriptive statistics of total and domain scores

| 영역 | 교육대학원 | | 사범대학 | | 문항 수 | Cohen's d | 분산 비 |
|----|---------|--------|---------|--------|------|-----------|--------|
| | 평균 | 표준편차 | 평균 | 표준편차 | | | |
| ① | 48.611 | 5.793 | 46.081 | 5.259 | 15 | 0.439 | 0.824 |
| ② | 46.389 | 5.992 | 46.081 | 5.591 | 15 | 0.052 | 0.871 |
| ③ | 45.667 | 6.071 | 45.135 | 5.715 | 15 | 0.088 | 0.886 |
| ④ | 42.528 | 5.950 | 43.126 | 6.015 | 15 | -0.100 | 1.022 |
| ⑤ | 44.222 | 6.573 | 42.234 | 7.472 | 15 | 0.299 | 1.292 |
| ⑥ | 47.917 | 5.997 | 45.505 | 6.333 | 15 | 0.400 | 1.115 |
| ⑦ | 46.583 | 6.063 | 45.252 | 7.179 | 15 | 0.216 | 1.402 |
| ⑧ | 46.139 | 4.128 | 45.748 | 5.552 | 15 | 0.092 | 1.809 |
| ⑨ | 49.417 | 6.635 | 46.892 | 6.986 | 15 | 0.378 | 1.109 |
| ⑩ | 46.472 | 4.632 | 45.820 | 5.133 | 15 | 0.139 | 1.228 |
| ⑪ | 49.306 | 7.066 | 46.054 | 6.265 | 15 | 0.463 | 0.786 |
| ⑫ | 46.833 | 6.069 | 44.946 | 5.535 | 15 | 0.312 | 0.832 |
| ⑬ | 54.472 | 5.180 | 52.865 | 5.819 | 15 | 0.307 | 1.262 |
| ⑭ | 44.806 | 4.915 | 44.523 | 4.815 | 15 | 0.058 | 0.990 |
| 합계 | 659.361 | 55.511 | 640.261 | 55.233 | 210 | 0.344 | 55.233 |

주. ①문제해결력·탐구력, ②판단력, ③독립성·자주성, ④창의·응용력, ⑤심리적 안정성, ⑥언어·의사소통력, ⑦지도성, ⑧공감·포용력, ⑨지식·정보력, ⑩봉사·희생·협동성, ⑪계획성, ⑫성실·책임감, ⑬소명감·교직관, ⑭열정임.

학교 급별에 상관없이 소명감·교직관, 지식·정보력, 계획성 영역, 문제해결력·탐구력 순으로 평균이 높게 나타났으며, 다음으로 교육대학원에서는 언어·의사소통력과 성실·책임감 영역 순으로, 그리고 사범대학에서는 판단력과 봉사·희생·협동성 영역 순으로 나타났다. 총점의 경우 교육대학원 집단의 평균은 659.361이며 사범대학 집단의 평균은 640.261로 두 집단의 평균 차이는 중간 효과크기를 나타낸 반면, 사범대학 집단이 교육대학원 집단보다 변동성은 1% 큰 것으로 나타났다. 또한 교직 적·인성 검사의 하위영역 중 창의·응용력 영역을 제외 하면 모든 하위영역에서 교육대학원 집단의 평균이 사범대학 집단의 평균보다 높게 나타났다. 구체적으로 계획성 영역에서 교육대학원 집단과 사범대학 집단의 평

균 차이가 가장 크게 나타났으며, 다음으로 문제해결력·탐구력 영역, 언어·의사소통능력 영역, 그리고 지식·정보능력 영역 순으로 나타났다. 반면에 분산비는 공감·포용력 영역에서 1.809로 사범대학 집단이 교육대학원 집단보다 변동성이 80.9% 크게 나타났으며, 계획성 영역에서는 0.786으로 교육대학원 집단이 사범대학 집단보다 변동성이 27.2% 크게 나타났다.

2. 학교 급별 교직 적·인성 검사의 하위영역 점수에 영향을 주는 요인들의 상대적인 영향력 탐색

교육대학원과 사범대학 학교 급별로 예비수학교사의 교직 적·인성 검사점수에 영향을 미치는 요인들의 상대적인 영향력을 탐색하기 위하여 각 하위영역은 고정효과로, 그리고 피험자 및 문항은 무선효과로 가정한다. 변량 $p \times i^*$ 설계를 적용하여 분석한 G-연구 결과는 [표 3]과 같다. [표 3]에는 학교 급별 예비수학교사의 분산 성분 추정치, 해당 분산 성분이 전체 분산에서 차지하는 비율, 그리고 각 교직 적·인성 검사의 하위영역 간 측정 오차를 고려한 상관계수가 제시되어 있다.

분산 성분 추정치는 [표 3]에서 대각선상에 굵게 표시하였으며, 설명력은 소괄호 안에 퍼센트로 표시하였다. 분산 성분의 설명력을 살펴보면, 학교 급별에 상관없이 공감·포용력 영역과 열정 영역에서만 문항 효과가 가장 크게 나타났으며, 나머지 영역에서는 잔차 효과가 가장 크게 나타나는 패턴을 보였다. 잔차 효과는 본 연구 설계에서 고려하지 않은 추가적인 국면의 필요성에 대한 간접적인 설명으로 해석할 수 있다. 반면에 교육대학원에서는 판단력, 계획성, 소명감·교직관 영역에서 피험자 효과가 15%, 24%, 15%로 문항 효과인 12%, 20%, 14%보다 높게 나타났으며, 사범대학에서는 심리적 안정성 영역에서 피험자 효과가 18%로 문항 효과인 15%보다 높게 나타났으며, 교육대학원과 사범대학 모두 지식·정보능력 영역에서는 피험자 효과가 각각 21%와 18%로 문항 효과인 7%와 12%로 높게 나타났다. 이처럼 피험자 효과가 문항 효과보다 높게 나타난 영역에 대해서는 교직 적·인성 검사점수에 피험자 간의 교직 적·인성 차이가 문항 난이도 차이보다 더 큰 영향을 미친다고 해석할 수 있다.

측정의 오차를 고려한 상관계수는 [표 3]의 대각선

[표 3] 학교 급별 다변량 $p^* \times i^*$ 설계의 G-연구 결과

[Table 3] Multivariate G-study for the $p^* \times i^*$ design by school levels

| 효과 | 영역 | 교육대학원 | | | | | | | | | | | | | |
|--------|----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | ① | ② | ③ | ④ | ⑤ | ⑥ | ⑦ | ⑧ | ⑨ | ⑩ | ⑪ | ⑫ | ⑬ | ⑭ |
| p | ① | 0.119 | 0.813 | 0.262 | 0.705 | 0.482 | 0.858 | 0.856 | 0.953 | 0.301 | 0.902 | 0.791 | 0.672 | 0.785 | 0.130 |
| | ② | (15%) | 0.120 | 0.565 | 0.523 | 0.652 | 0.866 | 1.000 | 0.896 | 0.672 | 0.743 | 0.972 | 0.899 | 0.687 | 0.268 |
| | ③ | | (15%) | 0.116 | 0.283 | 0.415 | 0.439 | 0.445 | 0.217 | 0.714 | 0.408 | 0.478 | 0.465 | 0.581 | 0.268 |
| | ④ | | | (12%) | 0.124 | 0.365 | 0.573 | 0.620 | 0.867 | 0.596 | 0.600 | 0.503 | 0.368 | 0.443 | 0.113 |
| | ⑤ | | | | (9%) | 0.145 | 0.449 | 0.837 | 0.718 | 0.373 | 0.185 | 0.423 | 0.454 | 0.261 | -0.374 |
| | ⑥ | | | | | (13%) | 0.117 | 0.724 | 0.782 | 0.653 | 0.412 | 0.730 | 0.561 | 0.780 | 0.272 |
| | ⑦ | | | | | | (12%) | 0.127 | 0.967 | 0.389 | 0.799 | 0.961 | 0.826 | 0.587 | 0.009 |
| | ⑧ | | | | | | | (15%) | 0.041 | 0.634 | 0.712 | 0.710 | 0.814 | 0.828 | 0.180 |
| | ⑨ | | | | | | | | (3%) | 0.159 | 0.164 | 0.475 | 0.486 | 0.598 | 0.310 |
| | ⑩ | | | | | | | | | (21%) | 0.056 | 0.924 | 0.781 | 0.853 | 0.235 |
| | ⑪ | | | | | | | | | | (5%) | 0.192 | 0.976 | 0.738 | 0.387 |
| | ⑫ | | | | | | | | | | | (24%) | 0.123 | 0.793 | 0.148 |
| | ⑬ | | | | | | | | | | | | (13%) | 0.090 | 0.469 |
| | ⑭ | | | | | | | | | | | | | (15%) | 0.073 |
| | | | | | | | | | | | | | | (5%) | |
| d | | 0.197 | 0.096 | 0.171 | 0.693 | 0.284 | 0.222 | 0.156 | 0.887 | 0.054 | 0.489 | 0.155 | 0.249 | 0.089 | 0.966 |
| | | (26%) | (12%) | (17%) | (53%) | (25%) | (23%) | (19%) | (62%) | (7%) | (43%) | (20%) | (25%) | (14%) | (62%) |
| pi,e | | 0.454 | 0.590 | 0.711 | 0.493 | 0.700 | 0.637 | 0.541 | 0.514 | 0.552 | 0.593 | 0.444 | 0.609 | 0.437 | 0.515 |
| | | (59%) | (73%) | (71%) | (38%) | (62%) | (65%) | (66%) | (36%) | (72%) | (52%) | (56%) | (62%) | (71%) | (33%) |
| 효과 | 영역 | 사범대학 | | | | | | | | | | | | | |
| | | ① | ② | ③ | ④ | ⑤ | ⑥ | ⑦ | ⑧ | ⑨ | ⑩ | ⑪ | ⑫ | ⑬ | ⑭ |
| p | ① | 0.088 | 0.883 | 0.631 | 0.567 | 0.590 | 0.569 | 0.426 | 0.663 | 0.489 | 0.517 | 0.627 | 0.521 | 0.557 | 0.234 |
| | ② | (9%) | 0.104 | 0.639 | 0.645 | 0.479 | 0.571 | 0.407 | 0.667 | 0.545 | 0.552 | 0.684 | 0.651 | 0.583 | 0.294 |
| | ③ | | (14%) | 0.098 | 0.522 | 0.625 | 0.547 | 0.602 | 0.905 | 0.594 | 0.575 | 0.465 | 0.698 | 0.575 | 0.433 |
| | ④ | | | (10%) | 0.120 | 0.353 | 0.473 | 0.503 | 0.612 | 0.580 | 0.317 | 0.389 | 0.486 | 0.397 | 0.264 |
| | ⑤ | | | | (10%) | 0.199 | 0.450 | 0.525 | 0.612 | 0.395 | 0.431 | 0.303 | 0.487 | 0.468 | 0.040 |
| | ⑥ | | | | | (18%) | 0.132 | 0.774 | 0.674 | 0.345 | 0.482 | 0.440 | 0.656 | 0.549 | 0.345 |
| | ⑦ | | | | | | (12%) | 0.189 | 0.699 | 0.526 | 0.597 | 0.420 | 0.609 | 0.590 | 0.365 |
| | ⑧ | | | | | | | (19%) | 0.099 | 0.743 | 0.773 | 0.409 | 0.714 | 0.782 | 0.432 |
| | ⑨ | | | | | | | | (8%) | 0.171 | 0.585 | 0.426 | 0.553 | 0.502 | 0.406 |
| | ⑩ | | | | | | | | | (18%) | 0.081 | 0.496 | 0.576 | 0.682 | 0.338 |
| | ⑪ | | | | | | | | | | (7%) | 0.137 | 0.700 | 0.440 | 0.314 |
| | ⑫ | | | | | | | | | | | (14%) | 0.094 | 0.506 | 0.537 |
| | ⑬ | | | | | | | | | | | | (10%) | 0.118 | 0.570 |
| | ⑭ | | | | | | | | | | | | | (16%) | 0.063 |
| | | | | | | | | | | | | | | (5%) | |
| d | | 0.376 | 0.120 | 0.161 | 0.491 | 0.159 | 0.301 | 0.235 | 0.584 | 0.118 | 0.490 | 0.254 | 0.245 | 0.118 | 0.724 |
| | | (38%) | (16%) | (17%) | (40%) | (15%) | (27%) | (23%) | (47%) | (12%) | (44%) | (27%) | (25%) | (16%) | (52%) |
| pi,e | | 0.517 | 0.518 | 0.714 | 0.605 | 0.731 | 0.689 | 0.598 | 0.571 | 0.683 | 0.543 | 0.567 | 0.634 | 0.488 | 0.605 |
| | | (53%) | (70%) | (73%) | (50%) | (67%) | (61%) | (59%) | (46%) | (70%) | (49%) | (59%) | (65%) | (67%) | (43%) |

주1. ①문제해결력·탐구력, ②판단력, ③독립성·자주성, ④창의·응용력, ⑤심리적 안정성, ⑥언어·의사소통능력, ⑦지도성, ⑧공감·포용력, ⑨지식·정보능력, ⑩봉사·희생·협동성, ⑪계획성, ⑫성실성·책임감, ⑬소명감·교직관, ⑭열정임.

주2. ()는 교직 적·인성 검사의 하위영역 별 분산 성분이 전체 분산에서 차지하는 퍼센트임.

위에 표시하였으며, 학교 급별에 상관없이 대부분 정적인 관계를 나타냄으로써 교직 적·인성 검사의 한 하위영역에서 높은 점수를 받은 피험자는 다른 하위영역에서도 높은 점수를 받는 것으로 해석할 수 있다. 또한 측정의 오차를 고려한 상관계수가 가장 높게 나타난 영역으로 교육대학원에서는 지도성 영역과 판단력 영역이었으며, 사범대학에서는 독립성·자주성 영역과 공감·포용력 영역인 것으로 나타났다. 반면에 심리적 안정성 영역과 열정 영역은 측정의 오차를 고려한 상관계수가 교육대학원에서 -0.374로 나타남으로써 심리적 안정성 영역에서 높은 점수를 받은 피험자는 열정 영역에서는 높지 않은 점수를 받는 것으로 나타났으며, 사범대학에서도 0.040으로 두 영역에서 상관관계는 거의 없는 것으로 나타났다.

3. 학교 급별 교직 적성·인성 검사의 영역 별 신뢰도 및 실질 가중치 탐색

다변량 G-연구 결과를 바탕으로 교직 적·인성 검사의 하위영역별 문항 수를 명목 가중치로 정하여 수행한

다변량 D-연구 분석결과는 [표 4]와 같다. 각 하위영역별 문항 수는 모두 동일하므로 명목 가중치는 7.1%로 모두 동일하게 적용하였다. 먼저 준거참조평가에서 적합한 신뢰도인 의존도계수를 살펴보면 교육대학원에서는 계획성과 지식·정보능력 영역에서 각각 0.828과 0.797로, 그리고 사범대학에서는 지도성과 심리적 안정성 영역에서 각각 0.773과 0.771로 가장 높게 나타났다. 일반적으로 적정 수준의 신뢰도에 대한 기준은 연구자들마다 다르지만, 일반화가능도 분석을 적용한 연구들이 보고하고 있는 의존도계수의 범위는 0.60부터 0.96(Marty et al., 2010; Schnitzler et al., 2015; Sluijsmans et al., 2001)이므로 본 연구에서는 0.60을 기준으로 정하였다. 이에 따르면 학교 급별에 상관없이 공감·포용력, 봉사·희생·협동성, 그리고 열정 영역은 적정 수준의 신뢰도에 도달하지 못하는 것으로 나타났으며, 사범대학에서는 문제해결력·탐구력 영역이 적정 수준의 신뢰도에 도달하지 못하는 것으로 나타났다. 따라서 문제해결력·탐구력 영역은 학교 급별에 따라 측정의 동등성을 확보하기에는 다소 부정적인 결과를 보인다고

[표 4] 학교 급별 다변량 $p^* \times I^*$ 설계의 D-연구 결과
 [Table 4] Multivariate D-study for the $p^* \times i^*$ design by school levels

| 집 단 분류 | | 하위영역 | | | | | | | | | | | | | | 합성점수 |
|--------|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | ① | ② | ③ | ④ | ⑤ | ⑥ | ⑦ | ⑧ | ⑨ | ⑩ | ⑪ | ⑫ | ⑬ | ⑭ | |
| 교육대학원 | $\sigma^2(\tau)$ | 0.119 | 0.120 | 0.116 | 0.124 | 0.145 | 0.117 | 0.127 | 0.041 | 0.159 | 0.056 | 0.192 | 0.123 | 0.090 | 0.073 | 0.067 |
| | $\sigma^2(\Delta)$ | 0.043 | 0.046 | 0.059 | 0.079 | 0.066 | 0.057 | 0.046 | 0.093 | 0.040 | 0.072 | 0.040 | 0.057 | 0.035 | 0.099 | 0.004 |
| | Φ | 0.732 | 0.725 | 0.664 | 0.612 | 0.689 | 0.672 | 0.733 | 0.307 | 0.797 | 0.436 | 0.828 | 0.683 | 0.720 | 0.425 | 0.941 |
| | $C(\alpha)$ | 0.797 | 0.753 | 0.711 | 0.791 | 0.757 | 0.734 | 0.779 | 0.548 | 0.812 | 0.585 | 0.867 | 0.752 | 0.756 | 0.680 | 0.962 |
| | $n(\omega)$ | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 |
| 사범대학 | $\sigma^2(\tau)$ | 0.088 | 0.104 | 0.098 | 0.120 | 0.199 | 0.132 | 0.189 | 0.099 | 0.171 | 0.081 | 0.137 | 0.094 | 0.118 | 0.063 | 0.066 |
| | $\sigma^2(\Delta)$ | 0.060 | 0.043 | 0.058 | 0.073 | 0.059 | 0.066 | 0.056 | 0.077 | 0.053 | 0.069 | 0.055 | 0.059 | 0.040 | 0.089 | 0.004 |
| | Φ | 0.598 | 0.710 | 0.626 | 0.623 | 0.771 | 0.667 | 0.773 | 0.562 | 0.763 | 0.540 | 0.714 | 0.616 | 0.745 | 0.414 | 0.938 |
| | $C(\alpha)$ | 0.719 | 0.751 | 0.672 | 0.749 | 0.804 | 0.742 | 0.826 | 0.722 | 0.790 | 0.691 | 0.783 | 0.690 | 0.784 | 0.608 | 0.958 |
| | $n(\omega)$ | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 |
| | $e(\omega)$ | 0.065 | 0.073 | 0.072 | 0.066 | 0.082 | 0.076 | 0.094 | 0.080 | 0.085 | 0.059 | 0.069 | 0.070 | 0.074 | 0.035 | |

주1. ①문제해결력·탐구력, ②판단력, ③독립성·자주성, ④창의·응용력, ⑤심리적 안정성, ⑥언어·의사소통능력, ⑦지도성, ⑧공감·포용력, ⑨지식·정보능력, ⑩봉사·희생·협동성, ⑪계획성, ⑫성실성·책임감, ⑬소명감·교직관, ⑭열정임.
 주2. $\sigma^2(\tau)$ 는 전집점수 분산, $\sigma^2(\Delta)$ 는 절대오차 분산, Φ 는 의존도계수, $C(\alpha)$ 는 Cronbach's α , $n(\omega)$ 는 명목 가중치, $e(\omega)$ 는 실질 가중치를 나타냄.

해석할 수 있다. 그러나 교직 적·인성 검사의 총점에 해당하는 합성점수의 의존도계수와 Cronbach's α 는 학교 급별에 상관없이 모두 0.93이상의 높은 신뢰도가 산출됨으로써 교직 적·인성에 대한 측정은 대체로 양호하다고 해석할 수 있다.

고전검사이론을 바탕으로 한 Cronbach's α 와 의존도계수를 비교한 결과, 학교 급별에 상관없이 모든 영역에서 Cronbach's α 가 의존도계수보다 높게 나타났다. 그러나 상대적 순위가 항상 일치하는 것은 아니었다. 예를 들어 사범대학에서 창의·응용력과 언어·의사소통력 영역의 의존도계수는 각각 0.623과 0.667로 언어·의사소통력 영역에서 더 높은 값이 산출되었지만 Cronbach's α 는 각각 0.749와 0.742로 창의·응용력에서 더 높게 나타났다. 이는 교직 적·인성 검사점수에서 단일 오차요인만을 고려하며, 피험자와 상호작용하는 요인의 분산만을 오차 분산에 포함시키는 Cronbach's α 와 달리 의존도계수는 피험자의 전집점수 분산을 제외한 모든 분산 성분이 오차 분산에 포함되기 때문이다.

또한 명목 가중치가 학교 급별에 상관없이 모든 영역에서 동일하게 7.1%였다면, 교직 적·인성 검사의 총점에 하위영역 점수가 반영된 실질 가중치는 학교 급별에 상관없이 판단력, 언어·의사소통력, 지도성, 그리고 지식·정보능력 영역은 상대적으로 높게, 그리고 창의·응용력, 봉사·희생·협동성, 그리고 열정 영역은 상대적으로 명목 가중치보다 낮게 나타났다. 반면에 교육대학원에서는 문제해결력·탐구력, 계획성, 그리고 성실성·책임감 영역이, 그리고 사범대학에서는 독립성·자주성, 공감·포용력, 그리고 소명감·교직관 영역에서 실질 가중치가 명목 가중치보다 높게 나타났다. 또한 일반적으로 교육대학원에서는 인지적 특성에서 실질 가중치가 높을 때, 그리고 사범대학에서는 정의적 특성에서 실질 가중치가 높을 때 의존도계수가 높아지는 경향이 나타났다.

V. 결론 및 제언

본 연구는 교육대학원과 사범대학 예비수학교사에게 동일하게 실시된 교직 적·인성 검사 결과의 활용방안과 함께 신뢰도를 바탕으로 학교 급별에 상관없이 교직 적·인성 검사의 측정이 일관적이며 정확하게 이루어졌

는지를 살펴보기 위해 김성연(2018b)에서 단변량 일반화가능도 분석의 한계점을 극복한 다변량 일반화가능도 분석을 수행하였다. 주요 분석결과를 바탕으로 본 연구에서 활용한 다변량 일반화가능도 분석방법의 교원양성기관 현장 적용 가능성에 대한 논의 및 시사점을 제시하면 다음과 같다. 먼저 예비수학교사의 교직 적·인성 검사점수 총점을 기준으로 살펴보면 교육대학원 집단이 사범대학 집단보다 높게 나타났으며, 하위영역 점수를 기준으로 살펴보면 창의·응용력을 제외한 모든 하위영역에서 교육대학원 집단이 사범대학 집단보다 높게 나타났다. 또한 학교 급별에 상관없이 소명감·교직관, 지식·정보력, 계획성, 문제해결력·탐구력이 높게 나타났다. 반면에 교육대학원에서는 언어·의사소통력과 성실·책임감 영역이, 그리고 사범대학에서는 판단력과 봉사·희생·협동성 영역이 높게 나타났다. 이는 안도희와 김유리(2015)가 밝힌 교육대학원 집단의 교직핵심역량, 교수영역, 교직 전문성에 대한 열망이 교직을 이수하는 대학 집단보다 높다는 연구 결과와 기존의 선행연구들(백인순, 2000; 오유진, 2016)이 교육대학원 집단의 교육과정 또는 교원양성기관에 대한 인식이 사범대학 집단보다 높게 나타난다는 연구 결과와 부분적으로 일치한다. 또한 교원자격증을 취득하는데 있어 교육대학원생이 교사로서의 전문성을 개발하기 위해서라면, 사범대 학생은 먼저 교직에 대한 적성과 흥미 등을 점검하는 것이 선행되고 있다는 이선정과 신혜원(2008)의 연구 결과와도 부분적으로 일치한다. 즉, 본 연구 결과는 교육대학원생과 사범대학의 예비수학교사들은 미래에 이들이 직업으로 갖게 될 교사로서의 역할과 책임을 포함하는 교사의 자질에 대해 다르게 인식하고 있음을 경험적으로 보여 주었다. 이처럼 학교 급과 학교의 교육연한이 높을수록 교직 적·인성이 높게 나타나고 있음은 교원양성기관에서 교사로서 갖추어야 할 기본적 자질인 교직 적·인성을 선발과정에서부터 교육의 전 과정에 이르기까지 지속적으로 측정하고 지원하는 것이 필요함을 시사한다. 이는 짧은 기간의 교육 프로그램을 통해 오랜 시간동안 자신의 경험 축적에 의해 형성된 교직 적·인성은 쉽게 바꿀 수 있는 성질의 것이 아니기 때문이다(김정환 외, 2012).

또한 교육대학원과 사범대학 집단에 따라 예비수학

교사의 교직 적·인성 수준의 차이를 밝히는 연구 결과들은 교원양성기관 현장에서 집단에 상관없이 교직 적·인성 수준이 높게 나타난 인지적 영역에 대해서는 그 영역을 신장시키기 위한 시간 및 동기화시키기 위한 시간들을 다른 영역을 개발하는데 쓸 수 있음을 시사한다. 이처럼 학교 급에 상관없이 인지적 영역에서 교직 적·인성 수준이 높게 나타났다는 것은 교원양성기관의 교육과정의 즉각적인 해결 방법이 보이지 않는 문제 상황에서 이를 해결하는데 요구되는 적절한 전략을 사용할 수 있게 하며, 교직과 관련된 지식·정보를 수행에 적절히 적용할 수 있게 하며, 그리고 과제를 수행할 때 계획을 세워 철저히 준비할 수 있도록 하고 있다고 해석할 수 있다.

반면에 교직 적·인성 검사의 하위영역 중 교육대학원 집단에서는 언어·의사소통력과 성실·책임감 영역이, 그리고 사범대학에서는 판단력, 봉사·희생·협동성, 창의·응용력 영역이 높게 나타났다. 이를 통해 자신의 전공 분야에서 교과관련 지식을 심화시키기 위해 교육대학원에 진학한 원생들은 대학생들보다 그들 스스로를 미래의 학생들을 교육할 교사로서 갖추어야 할 적성과 인성 역량의 가장 기초가 되는 언어·의사소통력(진경희, 조안나, 2017)을 통해 학급운영, 수업관리, 학습목표를 달성하기 위한 노력을 보다 많이 기울이고 있는 것으로 여겨진다. 반면에 사범대학 학생들은 자신의 교직적성과 흥미를 점검하는데 중점을 두면서 새롭고 독창적이며 유용한 아이디어를 산출해내고 이를 생활과 관련하여 구체적으로 표출하며, 자기의 이익보다 상대방이나 국가 사회의 이익을 우선시하며 양보할 줄 알고 베푸는 생활을 실천하는데 좀 더 주력하고 있는 것으로 여겨진다. 따라서 언어·의사소통력과 성실·책임감 영역은 사범대학에서, 그리고 판단력, 봉사·희생·협동성, 창의·응용력 영역은 교육대학원 집단에서 해당 영역들을 개발하도록 이 영역의 필요성을 더 집중적으로 설명하고 동기를 유발할 수 있도록 좀 더 힘써으로써 교직 적·인성을 향상시킬 수 있는 프로그램 개발의 효율성을 높이는데 기여할 수 있다. 예컨대 사범대학에서는 학생들과 원활히 소통하고, 이념과 사상, 의견이나 태도, 지식이나 사실 등을 전달하고 이해할 수 있는, 그리고 교직의 임무나 의무를 중요하게 여길 수 있도록 반드시 해야만 하

는 일을 촉진하기 위해 스스로 목표를 설정할 수 있도록 하는 자기목표설정, 자기관찰, 자기처벌과 같은 교육 내용(김진영, 이기중, 2013)을 프로그램에 포함할 수 있을 것이다.

다음으로 다변량 G-연구 분석결과, 교직 적·인성 검사점수에 교직 적·인성 수준 차이가 문항 난이도 차이보다 더 많이 반영된 하위영역은 집단에 상관없이 지식·정보능력 영역, 그리고 교육대학원 집단에서는 판단력, 계획성 소명감·교직관 영역, 그리고 사범대학 집단에서는 심리적 안정성 영역으로 나타났다. 또한 학교 급별에 상관없이 대부분 교직 적·인성 검사의 하위영역 간 측정의 오차를 고려한 상관계수는 정적인 관계를 나타냈다. 그러나 심리적 안정성 영역과 열정 영역은 교육대학원 집단에서는 음의 상관관계가, 그리고 사범대학 집단에서는 상관관계가 거의 없는 것으로 나타났다. 또한 다변량 D-연구 분석결과, 대체로 교육대학원에서는 인지적 영역의 계획성과 지식·정보능력 영역이, 그리고 사범대학 집단에서는 정의적 영역의 지도성과 심리적 안정성 영역의 의존도계수가 높게 나타났다. 그러나 교직 적·인성 검사의 하위 영역 중 문제해결력·탐구력 영역이 교육대학원 집단에서는 적정 수준의 신뢰도에 도달하는 것으로 나타났지만, 사범대학 집단에서는 적정 수준의 신뢰도에 도달하지 못하는 것으로 나타났다. 따라서 문제해결력·탐구력 영역은 학교 급별에 따라 측정의 동등성을 확보하기에는 다소 부정적인 결과를 보였다. 또한 학교 급별에 상관없이 Cronbach's α 가 의존도계수보다 높게 나타났지만, 상대적 순위가 항상 일치하지는 않는 것으로 나타났다. 가중치와 관련하여서도 교육대학원에서는 계획성, 판단력, 문제해결력·탐구력, 언어·의사소통력과 같은 인지적 영역이, 그리고 사범대학에서는 심리적 안정성, 독립성·자주성, 공감·포용력과 같은 정의적 능력에서 실질 가중치가 명목 가중치보다 높게 나타났다. 그러나 교직 적·인성 검사의 총점에 해당하는 합성점수의 의존도계수와 Cronbach's α 는 학교 급별에 상관없이 모두 0.93이상의 높은 신뢰도가 산출됨으로써 교직 적·인성에 대한 측정은 대체로 양호하다고 해석할 수 있다.

이처럼 교육부에서 교육대학원과 사범대학에 동일하게 실시할 수 있도록 표준안으로 보급하기 위해 정교하

개 교직 적·인성 검사를 개발했다 하더라도 피험자가 어떤 학교 급별에 속해있는지를 비롯해 다양한 검사 요인들에 의해 측정의 정확성 또는 측정 오차가 다르게 나타날 수 있다는 본 연구의 결과는 선행 연구 결과들 (Li & Brennan, 2007; Powers & Brennan, 2009; Yin, 2005)과 일치한다. 따라서 측정 오차가 피험자 집단에 따라 다른 방식으로 개입될 경우에 측정의 동등성이 위배되는 결과가 초래(김성숙, 전경희, 2018; 김성연, 2018b)될 수 있으므로 교직 적·인성 검사를 개발, 수정 및 보완하는 과정에서 교과목 및 학교 급별 특성을 고려한 측정학적 측면에 대한 경험적 검증을 통해 측정의 동등성을 확보하려는 노력이 필요하다. 이러한 경험적 자료를 다변량 일반화가능도 이론을 적용하여 분석한 결과를 기반으로 교직 적·인성 검사의 타당성을 확보함으로써 예비수학교사들의 교직 적·인성 수준을 보다 정확하고 신뢰롭게 파악할 수 있을 것이다. 또한 본 연구의 결과는 기존에 고전검사이론을 바탕으로 한 Cronbach's α 는 측정 구조를 정확하게 반영하지 못함으로써 신뢰도를 과대추정한다는 연구 결과들(Brennan, 2001a; Dunbar et al., 1992; Fyans, 1983; Shavelson et al., 1993)과 일치한다. 따라서 보다 정밀한 신뢰도로 일반화가능도 이론을 바탕으로 준거참조검사에 적합한 신뢰도인 의존도계수를 산출하는 것이 필요하다.

마지막으로 본 연구의 제한점과 후속연구를 제안하면 다음과 같다. 첫째, 본 연구에서는 다변량 일반화가능도 분석을 수행하는 데 있어 교직 적·인성 검사의 하위영역을 고정된 상태에서 학교 급별에 따라 교육대학원과 사범대학 집단으로 나누어서 피험자와 문항만을 검사점수에 영향을 미치는 요인으로 고려하였다. 따라서 실제 교직 적·인성 검사점수에서 교육대학원과 사범대학 간의 집단효과가 어느 정도인지는 분석에서 제외되어 있다. 향후 교직 적·인성 검사를 연구하는 경우에는 검사점수에 영향을 미치는 요인으로 각 학교 급별을 나타내는 집단 효과 뿐만 아니라 각 학교 급별의 특성과 피험자 특성을 반영한 분석이 요구되어진다. 둘째, 본 연구에서는 하위영역 별 문항 수에 비례하게 명목 가중치를 설정한 상태에서 실질 가중치를 탐색하였다. 따라서 교직 적·인성 검사의 총점에 반영되는 명목 가중치를 다르게 설정하여 그에 따른 의존도계수의 변화

를 확인함으로써 교직 적·인성 검사에 대해 보다 구체적인 가이드라인을 제공할 수 있는 최적의 측정조건과 실질 가중치를 탐색하는 연구를 수행할 필요가 있다. 셋째, 본 연구에서는 동일한 교직 적·인성 검사를 학교 급별에 따라 교육대학원과 사범대학 예비수학교사에게 실시하는 것이 적절한지를 측정의 동등성 측면에서 분석한 것으로, 교직 적·인성 검사 자체의 신뢰성을 담보하는 것은 아니다. 따라서 교직 적·인성 검사가 수행되어야 할 전체 피험자를 대표할 수 있는 연구대상을 표집하여 다변량 일반화가능도 분석을 수행할 필요가 있다.

참 고 문 헌

- 교육부 (2015). 2015 문·이과 통합형 교육과정 총론 주요 사항, 세종: 교육부.
- Ministry of Education. (2015). *The main particular of 2015 National curriculum draft for an integration of arts and sciences*, Sejong: Ministry of Education.
- 김래영, 김은현 (2017). 교원양성교육과정과 교원임용후보자 선정경쟁시험에 나타난 중등수학교사에게 요구되는 지식 분석: 교수를 위한 내용 지식을 중심으로, 교과교육학연구 21(5), 610-623.
- Kim, R. Y. & Kim, E. H. (2017). An analysis of the expected content knowledge for teaching in teacher education programs and teacher employment tests for secondary mathematics, *Journal of Research in Curriculum Instruction* 21(5), 610-623.
- 김명화 (2005). 채점 자동화 시스템 구축을 위한 수학 구성형 문항 채점의 일반화 가능성 연구, 교육문제연구 22, 205-222.
- Kim, M. H. (2005). An application of the generalizability to constructed-response items in mathematics of computer automated scoring, *Journal of Research in Education* 22, 205-222.
- 김성숙, 김양분 (2001). 일반화가능도 이론, 서울: 교육과학사.
- Kim, S. S. & Kim, Y. B. (2001). *Generalizability theory*, Seoul: Kyoyookgwahaksa.
- 김성숙, 전경희 (2018). 정의적 성취특성 측정의 오차오인 및 성취수준별 척도의 동등성 분석, 교육과정평가

- 연구 21(1), 153-172.
- Kim, S. S. & Chon, K. H. (2018). An Analysis of measurement errors and invariance properties by proficiency level in the non-cognitive measures, *The Journal of Curriculum and Evaluation* 21(1), 153-172.
- 김성연 (2016). 수학적 창의성 검사의 채점 영역별 가중치 분석, 수학교육 55(2), 147-169.
- Kim, S. Y. (2016). Analysis of weights depending on scoring domains of the mathematical creativity test, *The Mathematical Education* 55(2), 147-169.
- 김성연 (2017a). 교직 인성 검사에서의 문항 프로파일 분석, 중등교육연구 65(4), 705-729.
- Kim, S. Y. (2017a). Item profile analysis of the personality test for teaching profession, *Second Education Research* 65(4), 705-729.
- 김성연 (2017b). 수학적 창의성 태도 검사에서 수학생제와 일반학생의 다집단 일반화가능도 분석, 수학교육 논문집 31(1), 49-70.
- Kim, S. Y. (2017b). Multigroup generalizability analysis of creativity attitude scale-Korea for mathematically gifted and general students in middle schools, *Communications of Mathematical Education* 31(1), 49-70.
- 김성연(2018a). 예비수학교사의 교직 적성·인성 검사에서 분할점수 변화에 따른 다양한 신뢰도 탐색, 수학교육 60(1), 55-74.
- Kim, S. Y. (2018a). Investigation of various reliability indices of pre-service mathematics teachers' teaching aptitude and personality test based on setting cut scores, *The Mathematical Education* 60(1), 55-74.
- 김성연 (2018b). 교육대학원과 사범대학에서 교직 적성·인성 검사의 일반화가능도 분석, 교원교육 34(2), 1-17.
- Kim, S. Y. (2018b). Generalizability analysis of teaching aptitude and personality test for graduate school of education and college of education students, *Korean Journal of Teacher Education* 34(2), 1-17.
- 김정환 (2004). 초등학교 기간제 교사의 교육능력 관련 요소의 인과관계 분석, 교육평가연구 17(1), 121-139.
- Kim, J. H. (2004). An investigation on casual relationships among educational competence-related factors of period-limit teachers in primary schools, *Journal of Educational Evaluation* 17(1), 121-139.
- 김정환, 남현우, 염시창, 임진영 (2012). 교직 적성·인성 검사 도구 개발 연구, 서울: 교육과학기술부.
- Kim, J., Nam, H., Yeom, S., & Im, J. (2012). *The development of teaching aptitude and personality test*, Seoul: Ministry of Education, Science, and Technology.
- 김진영, 이기종 (2013). 예비유아교사의 셀프리더십과 교직인·적성과의 관계, 한국교육 40(3), 5-25.
- Kim, J. Y. & Lhee, K. J. (2013). The relationship between pre-service kindergarten teachers' self-leadership strategies and teaching personalities and aptitudes, *The Journal of Korean Education* 40(3), 5-25.
- 김현진 (2013). 교사·학교장 신념과 중학생의 자율성 및 자기효능감, 학업성취도의 관계 분석, 교육학연구 51(2), 117-143.
- Kim, H. J. (2013). An analysis of relations among teacher and principal's beliefs and 8th grade students' autonomy, self-efficacy and achievement, *The Journal of Educational Research* 51(2), 117-143.
- 박경 (2007). 우리나라 중학생의 수학에 대한 정의적 특성 변화와 수학 성취에 미치는 영향력 분석, 수학교육 46(1), 19-31.
- Park, J. (2007). The trend in the Korean middle school students' affective variables toward mathematics and its effect on their mathematics achievements, *The Mathematical Education* 46(1), 19-31.
- 백인순 (2000). 교육대학원에 대한 교사의 인식 조사 연구. 석사학위논문, 숙명여자대학교.
- Paek, I. S. (2000). *A study on teacher's understanding of the graduate school of education*. Master's thesis, Seoul: Sookmyung Woman's University.
- 신준국, 부덕훈, 서보익 (2015). 수학교육에서 인성 함양을 위한 중학교 교수·학습 자료 개발 연구, 수학교육 논문집 29(2), 241-265.
- Shin, J. K., Boo, D. H., & Suh, B. E. (2015). A study on the development of teaching and learning materials for character education in middle school, *Communications of Mathematical Education* 29(2), 241-265.
- 안도희, 김유리 (2015). 교직이수 학생들의 교직핵심역량, 교수능력 및 교직전문성 열망 간의 관계, 교원교육 31(1), 203-226.
- Ahn, D. H. & Kim, Y. R. (2015). The relationships between core competencies in teaching, teaching ability and

- professional aspirations for teaching among student teachers, *Korean Journal of Teacher Education* 31(1), 203-226.
- 안재희, 이숙정 (2012). 교직이수 학생들의 진로결정에 대한 인식 분석, *열린교육연구* 20(2), 27-49.
- Ahn, J. & Lee, S. (2012). Analysis of career choices made by student teachers, *The Journal of Yeolin Education* 20(2), 27-49.
- 오유진 (2016). 영어과 임용준비생들의 서답형 영어쓰기 답안에 대한 인식과 이에 대비한 각 교육기관 교육과정의 효율성에 관한 인식, 석사학위논문, 숙명여자대학교,
- Oh, Y. J. (2016). *A study on the awareness of examination for secondary school English teachers' writing answer sheet by preliminary teacher and efficacy of curriculum of each English educational institution*, Master's thesis, Seoul: Sookmyung Woman's University.
- 원혜경, 이소정 (2017). 예비유아교사를 위한 간편형 교직 적·인성 척도 타당화 연구, 유아교육학논집 21(1), 133-165.
- Won, H. K. & Lee, S. J. (2017). Validation of a short form of the aptitude and personality as a teacher scale for preservice early childhood teachers, *Early Childhood Education Research & Review* 21(1), 133-165.
- 이규민, 황경현 (2007). 초등학교 과학과 수행평가의 총체적 채점과 분석적 채점 방식에 대한 일반화가능도 분석, *아동교육* 16(4), 169-184.
- Lee, G. M. & Hwang, K. H. (2007). A generalizability theory approach toward investigating the generalizability of scores from holistic and analytic scoring methods in performance assessments of an elementary school science class, *The Korean Journal of Child Education* 16(4), 169-184.
- 이문수, 차동춘 (2016). 수학 창의력 문제 해결 검사에 서의 일반화가능도 이론과 다국면 라쉬모형의 비교 연구, *교육과정평가연구* 19(2), 251-279.
- Lee, M. & Cha, D. (2016). A comparison of generalizability theory and many facet Rasch measurement in an analysis of mathematics creative problem solving test, *The Journal of Curriculum and Evaluation* 19(2), 251-279.
- 이선정, 신혜원 (2008). 서울소재 교육대학원 가정교육 전공 교육과정에 대한 운영실태와 교육대학원생의 인식, *교원교육* 20(4), 173-186.
- Lee, S. J. & Shin, H. W. (2008). The operation of home economics education course in graduate school of education and the graduate students' perception, *Korean Journal of Teacher Education* 20(4), 173-186.
- 이현숙, 송미영 (2015). PISA 2012 수학 성취도를 설명하는 학생의 정의적 특성 및 교사 특성 분석을 위한 다층 구조방정식모형의 적용, *교과교육학연구* 19(1), 137-158.
- Yi, H. S. & Song, M. Y. (2015). A multi-level SEM approach for the analysis of relationships between math-related educational context variables and math literacy of PISA 2012, *Journal of Research in Curriculum Instruction* 19(1), 137-158.
- 전경희, 조안나 (2017). 예비교사의 입학전형 유형별 교직 적·인성 특성의 구조적 관계 분석, 학습자중심교과교육연구 17(17), 315-334.
- Chon, K. H. & Cho, A. (2017). An analysis of the structural relationship among aptitude, character traits of preservice teachers by college admission types, *Journal of Learner-Centered Curriculum and Instruction* 17(17), 315-334.
- 조운주 (2014). 예비유아교사를 위한 교직적성·인성 검사도구의 타당성 및 개선방안, 육아지원연구 9(2), 101-123.
- Cho, W. (2014). Validation and modification of teaching aptitude test for pre-service early childhood teachers, *Early Childhood Education and Care* 9(2), 101-123.
- 한혜숙, 최계현 (2011). 중등 수학 교사들의 정의적 특성에 대한 인식과 수업 실태 분석, 한국학교수학회논문집 14(4), 491-518.
- Han, H. S. & Choi, K. H. (2011). Secondary mathematics teachers' recognition of the affective domain and analysis of condition in mathematics teaching, *Journal of the Korean School Mathematics Society* 14(4), 491-518.
- 황혜정 (2011). 수학 수업의 교사 지식에 관한 평가 요소 탐색-교수·학습 방법 및 평가를 중심으로, 한국학교수학회논문집 14(3), 241-263.
- Hwang, H. J. (2011). The study on the investigation of the mathematics teaching evaluation standards focused on teaching and learning methods and assessment, *Journal of the Korean School Mathematics Society* 14(3),

- 241-263.
- Abedi, J. & Lord, C. (2001). The language factor in mathematics tests, *Applied Measurement in Education* 14(3), 219-234.
- Baxter, G. P. Shavelson, R. J., Herman, S. J., Brown, K. A., & Valadez, J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact, *Journal for Research in Mathematics Education* 24, 190-216.
- Brennan, R. L. (2001a). *Generalizability Theory*, New York, NY: Springer-Verlag.
- Brennan, R. L. (2001b). *Manual for mGENOVA*, Iowa City: Iowa Testing Programs, University of Iowa.
- Christ, T. J., Johnson Gros, K. N., & Hintze, J. M. (2005). An examination of alternate assessment durations when assessing multiple skill computational fluency: The generalizability and dependability of curriculum based outcomes within the context of educational decisions, *Psychology in the Schools* 42(6), 615-622.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness, *Educational and Psychological Measurement* 57(3), 373-399.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence, *Education Policy Analysis Archives* 8(1), 1-44.
- Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments, *Applied measurement in education* 4(4), 289-303.
- Fyans, L. J. (1983). *Generalizability theory: Inferences and practical applications*, Jossey-Bass Inc Pub.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study, *Educational Researcher* 41(2), 56-64.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement, *American educational research journal* 42(2), 371-406.
- Ho, A. D. & Kane, T. J. (2013). *The reliability of classroom observations by school personnel. Research Paper*, MET Project, Bill & Melinda Gates Foundation.
- Houchard, M. A. (2005). *Principal leadership, teacher morale, and student achievement in seven schools in Mitchell county, North Carolina*, Electronic Theses and Dissertations.
- Klein, S. P., McCaffrey, D., Stecher, B., & Koretz, D. (1995). The reliability of mathematics portfolio scores: lessons from the Vermont experience, *Applied Measurement in Education* 8(3), 243-260.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms, *The power of video studies in investigating teaching and learning in the classroom*, 137-160.
- Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement, *Economics of Education Review* 28(1), 49-57.
- Kunter, M., Tsai, Y. M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction, *Learning and Instruction* 18(5), 468-482.
- Lakin, J. M. & Lai, E. R. (2012). Multigroup generalizability analysis of verbal, quantitative, and nonverbal ability tests for culturally and linguistically diverse students, *Educational and*

- Psychological Measurement* 73(1), 139-158.
- Lane, S., Liu, M., Ankenmann, R. D., & Stone, C. A. (1996). Generalizability and validity of a mathematics performance assessment, *Journal of Educational Measurement* 33(1), 71-92.
- Lee, G. & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets, *Applied Measurement in Education* 12(3), 237-255.
- Li, D. & Brennan, R. (2007). *A multi-group generalizability analysis of a large-scale reading comprehension test*, In annual meeting of the National Council on Measurement in Education. Chicago, IL.
- Marty, M. C., Henning, J. M., & Willse, J. T. (2010). Accuracy and reliability of peer assessment of athletic training psychomotor laboratory skills, *Journal of athletic training* 45(6), 609-614.
- Mastergeorge, A. M. & Martínez, J. F. (2010). Rating performance assessments of students with disabilities: A study of reliability and bias, *Journal of Psychoeducational Assessment* 28(6), 536-550.
- Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment*, CSE Technical Report 681. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- McBee, M. M. & Barnes, L. L. (1998). The generalizability of a performance assessment measuring achievement in eight-grade mathematics, *Applied Measurement in Education* 11(2), 179-194.
- Möttus, R., Allik, J. & Realo, A. (2012). The effect of response style on self-reported conscientiousness across 20 countries, *Personality and Social Psychology Bulletin* 38(11), 1423-1436.
- Narvaez, D. & Nucci, L. P. (2008). *Handbook of moral and character education*, New York, NY: Routledge.
- Powers, S. & Brennan, R. L. (2009). *Multivariate generalizability analyses of mixed-format advanced placement exams*, CASMA Research Report.
- Schnitzler, C., Button, C., Croft, J. L., & Seifert, L. (2015). A new qualitative typology to classify treading water movement patterns, *Journal of sports science & medicine* 14(3), 530.
- Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments, *Journal of Educational Measurement* 30(3), 215-232.
- Shavelson, R. J. & Webb, N. M. (1991). *Generalizability theory: A primer*, London: Sage Publications.
- Sluijsmans, D. (2001). Peer assessment in problem based learning, *Studies in educational evaluation* 27(2), 153-73.
- Thompson, A. G. (1992). *Teacher' belief and conceptions: A synthesis of the research*, New York: Macmillan Publishing Company.
- Wayne, A. J. & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review, *Review of Educational Research* 73(1), 89-122.
- Wilhelm, A. G. & Kim, S. (2015). Generalizing from observations of mathematics teachers' instructional practice using the instructional quality assessment, *Journal for Research in Mathematics Education* 46(3), 270-279.
- Yin, P. (2005). A multivariate generalizability analysis of the multistate bar examination, *Educational and psychological measurement* 63(4), 668-686.

An Analysis of Measurement Equivalence in a Teaching Aptitude and Personality Test for Pre-service Mathematics Teachers between a Graduate School of Education and a College of Education

Kim, Sungyeun

Incheon National University

E-mail : syk@inu.ac.kr

The purpose of this study was to investigate the measurement equivalence and to suggest application ways in teaching aptitude and personality test results for pre-service mathematics teachers between a graduate school of education and a college of education. This study analyzed the scores of the teaching aptitude and personality test of 36 pre-service mathematics teachers enrolled in a graduate school of education and 111 pre-service mathematics teachers in a college of education by performing a multivariate generalizability analysis. The main results were as follows. First, graduate's pre-service mathematics teachers had a higher level of teaching aptitude and personality than that of college's pre-service mathematics teachers based on the total scores. In addition, graduate's pre-service mathematics teachers had higher levels of teaching aptitude and personality than those of college's pre-service mathematics teachers except for a creativity·application domain based on the sub-domain scores. Second, cognitive domains were measured more precisely but affective domains were measured less precisely for graduate's pre-service mathematics teachers than for college's pre-service mathematics teachers. Third, regardless of school levels, Cronbach's α values, which might be overestimated by applying the classical test theory, were higher than dependability coefficients. Fourth, this study showed a somewhat negative result in ensuring the measurement equivalence for a problem solving·exploration domain. However, regardless of school levels, this study indicated that the overall measurement was generally reliable on composite scores. Based on these results, it was confirmed that multivariate generalizability methodologies' approach can be useful for exploring the measurement equivalence issues. Finally, this study suggests how to utilize the results of the test, how to apply a multivariate generalizability analysis for detecting the measurement equivalence, and how to develop future research based on limitations.

* ZDM Classification : B55

* 2000 Mathematics Subject Classification : 97C40

* Key words : teaching aptitude and personality test, pre-service mathematics teachers, dependability coefficients, measurement invariance