

수학적 창의성 검사의 채점 영역별 가중치 분석

김성연(인천대학교)

I. 서론

최근 교육부(2015)는 배움을 즐기는 수학교육 추진을 위한 제2차 수학교육 종합계획을 발표하면서 이 계획이 완료되는 2019년에는 다양한 사업을 통해 우리나라 학생들이 수학에 대한 흥미와 자신감, 수학교육의 가치 등이 전반적으로 상승하고, 창의적 융합 인재 양성에 수학이 기여하는 바가 더욱 커질 수 있도록 노력하겠다고 밝혔다. 여기서 창의적 융합 인재란 인문학적 상상력, 과학기술 창조력을 갖추고 바른 인성을 겸비하여 새로운 지식을 창조하고 다양한 지식을 융합하여 새로운 가치를 창출할 수 있는 사람으로, 문·이과 통합형 교육과정인 2015개정 교육과정이 추구하는 인재상이다(교육부, 2014). 이처럼 창의성은 학교 교육과정이 추구하는 인간상으로 5차 교육과정에서는 창조적인 사람(문교부, 1987), 6차 교육과정에서는 창의적인 사람(교육부, 1992), 7차와 2007개정 교육과정에서는 기초능력을 토대로 창의적인 능력을 발휘하는 사람(교육부, 1997; 교육인적자원부, 2007), 그리고 2009개정 교육과정에서는 기초 능력의 바탕 위에 새로운 발상과 도전으로 창의성을 발휘하는 사람(교육과학기술부, 2009) 등 꾸준히 강조되어 왔다. 특히 2009년 교육과학기술부가 발표한 창의·인성교육 기본방안에는 창의성이 몇몇 소수에게만 선천적으로 타

고난 특별한 능력이며, 모든 학생들이 교육을 통해 창의성이 개발되기는 어렵다는 오해가 존재하였음을 지적하면서, 학년과 수준에 상관없이 모든 학생을 대상으로 모든 교과에서 창의성 교육의 필요성을 주장하였다.

이에 따라 수학은 창의적인 사고력을 길러줄 수 있는 교과(교육부, 1992)로 일반적인 창의성을 기르는 수단적 의미로 여기거나, 수학과에서 문제해결을 창의성과 유사한 개념으로 해석하였던 것과 달리 2009개정 교육과정에서 수학적 창의성이란 용어가 등장(교육과학기술부, 2009)하게 되었다. 수학적 창의성이란 수학적 과제를 해결하는 과정에서 다양하고 독창적인 해결 방법을 산출하거나 새로운 관점에서 과제를 탐구하고 지식을 구성하는 능력을 의미한다(한혜정 외, 2012). 이러한 수학적 창의성을 신장시키기 위하여 교수·학습에서는 수학적 문제 해결, 추론 능력, 의사소통 능력을 강조하며, 확산적 사고를 촉진시키며, 고차원으로 확장해서 사고할 수 있게 하며, 학생 스스로 개념과 용어의 필요성을 인식하고 정의하게 하고 있다. 또한 평가에서는 수학적 지식과 기능을 바탕으로 창의적으로 사고하는 능력을 강조하고 있다(교육과학기술부, 2009). 일례로 서울특별시교육청(2010)에서는 평가 개선 방안의 하나로 정기고사에서 서답형 문항의 반영 비율을 전체 점수의 50%로 확대할 것을 발표하기도 하였으며, 이후 수행평가와 서술형, 논술형 문항의 반영 비율을 30%이상으로 수정하였고, 대전광역시교육청(2011)에서는 서술형과 서답형을 포함하는 주관식 문항을 배점의 30%이상으로 출제할 것이라는 평가지침을 발표하였다. 또한 2015개정 교육과정에서는 수학과 평가에서 창의융합을 선택형, 단답형, 서술논형 등의 다양한 문항 형태를 활용한 지필평가와 프로젝트 평가 방법을 명시하고 있다.

이처럼 평가 방식을 변화시킴으로써 학교 현장의 교육을 개선하고자 하는 측정주도 교수는 교육 현장을 단기간에 변화시킬 수 있다는 점에서는 효과적(성태제,

* 접수일(2015년 11월 11일), 수정일(2015년 12월 15일), 게재확정일(2015년 12월 30일)

* ZDM분류 : C43

* MSC2000분류 : 97C40

* 주제어 : 가중치, 다변량 일반화가능도 분석, 수학적 창의성

* 본 연구를 위해 자료의 사용을 허락해주신 김용범 선생님과 논문의 발전을 위해 조언을 해주신 익명의 심사위원님들께 감사드립니다.

* 본 연구는 2013년 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 연구입니다(KRF-2013S1A3A2055007).

* 본 연구는 2015 국제수학교육학술대회에서 발표한 원고를 수정·보완한 것입니다.

2010)일 수 있다. 그러나 검사의 측정학적 측면에 대한 경험적 검증 없이 다소 임의적으로 정한 비율을 검사에 적용하도록 하는 것은 평가의 타당성을 위협할 소지가 있다(이현숙, 2012). 마찬가지로 동일한 검사 명세화를 기준으로 수학적 창의성이라는 구인을 측정하도록 검사가 개발되었다 할지라도 채점 영역 및 반영 방식 등에 따라 검사 점수가 달라질 수 있으므로, 결과적으로 신뢰도에도 영향을 주게 된다. 즉, 검사 제작 시 채점 영역에 따라 배점 비율을 어떻게 정할 것인지에 대한 적절하고 합리적인 판단을 하는 것은 매우 중요하다. 수학적 창의성과 관련한 연구들에서도 수학적 창의성 신장 방안(김용대, 2003; 도중훈, 2007; 방승진, 최중오, 2010; 유경선, 2012; 이동희, 김관수, 2010; 정유화, 2011; 황우형, 이유나, 2009), 수학적 창의성과 다른 요인과의 관계(신승윤, 2013; 윤지윤, 2014; 이경화, 고진영, 박숙희, 2007), 수학적 창의성 개념과 정의(김부운, 이지성, 2006; 김홍원, 김명숙, 송상현, 1996; 이종희, 김기연, 2007; 최은선, 2010; Balka, 1974; Ervynck, 1991; Haylock, 1987; Sriraman, 2004)와 더불어 수학적 창의성 측정과 평가에 대한 연구(김부운, 이지성, 2005; 김용범, 2009; 김정은, 2010; 남승인, 2007; 변은진, 전평국, 2001; 손은영, 2005; 유운재, 2002; 2003; 이강섭, 2010; 이대현, 2014; 정유화, 2011; 조석희, 황동주, 2007; 하수현, 이광호, 2014; 홍주연, 한인기, 2014; Leikin, 2009; Leikin, Koichu, & Breman, 2009; Leikin & Lev, 2007)가 활발하게 이루어지고 있다.

그러나 수학적 창의성 검사를 개발한 연구들에서 제시하는 검사 점수는 대부분 각 채점 영역을 더하거나 평균을 계산하여 산출하고 있으며, 검사의 신뢰도와 관련하여서는 고전검사이론을 바탕으로 한 Cronbach α 나 채점자간 일치도를 보고하고 있다. 따라서 검사 개발자의 판단에 의해 미리 결정된 채점 영역별 가중치가 실제 검사 점수에서 반영되는 비율이 달라질 수 있으며, 검사 점수에 영향을 미칠 수 있는 다양한 오차요인들(예: 문항, 검사 양식, 실시 시간, 채점 영역, 장소, 채점 방법, 채점자, 관찰 횟수 등)을 고려하지 못한 과대 추정된 신뢰도가 산출될 수 있다는 지적이 제기되어 왔다(김성숙, 김양분, 2001; 이규민, 황경현, 2007; Brennan, 2001a; Cronbach et al., 1972, 1997; Lee, Brennan, & Frisbie, 2000; Lee & Frisbie, 1999; Shavelson & Webb, 1991).

또한 수학적 창의성 검사의 채점과 관련한 연구들도 과거에 실시된 유사한 성격의 검사에서 적용하고 있는 관례나 전문가들의 선형적인 판단에 의하여 결정되는 경우가 대부분이었다. 일부 연구들은(이강섭, 2010; 이강섭, 심상길, 2005) 후속 연구로 창의성의 하위 영역을 유창성, 융통성, 그리고 독창성으로 채점하여 점수화하는 방법에서 각 채점 영역간의 가중치에 대한 면밀한 검토를 제안하고 있다. 그러나 현재까지도 측정학적 이론을 바탕으로 수학적 창의성 검사에서 채점 영역별 가중치를 탐색하는 방법에 대한 연구는 거의 이루어지고 있지 않은 실정이다. 따라서 본 연구에서는 위에서 제기된 문제점들을 보완하는 방법 중 하나로 일반화가능도 분석 방법을 제시하고자 한다.

일반화가능도 이론에서 검사 점수는 허용가능한 관찰 전집으로부터 무선 표집된 점수이며, 전집점수(universe score)는 일반화전집의 모든 조건에 대한 어느 한 측정 대상의 평균점수를 의미한다. 여기서 전집점수는 고전검사이론의 전점수와 유사한 개념이며, 전집이란 측정대상의 측정조건들에 일반화 과정을 포함한다는 의미에서 표집단위를 측정대상으로 하는 모집단의 의미와는 차이가 있다. 즉, 일반화가능도 이론에서는 관찰된 행위 하나가 행동 전집을 대표하는 일부분이라고 가정하고, 전집을 구성하는 국면(facet)을 바탕으로 표집된 행위가 얼마나 일반화되어질 수 있는가를 추정하는 것이다. 국면은 분산분석 설계에서 효과나 요인과 유사한 개념이며, 분산분석에서 처리나 수준은 조건으로 표현된다. 단, 국면은 허용가능한 변동요인을 의미하므로, 측정 대상은 포함되지 않는다. 반면에 고전검사이론에서는 측정의 신뢰도가 단순히 관찰 대상의 검사 결과 또는 관찰 과정이 얼마나 일관성 있게 기록되는가에 초점을 맞추기 때문에 측정 상황에서 발생할 수 있는 여러 오차 요인에 대한 설명이 불가능하다는 지적을 받아왔다(김성숙, 김양분, 2001; Brennan, 2001a; Cronbach et al., 1972, 1997; Lee et al., 2000; Lee & Frisbie, 1999). 이와 같은 단점을 보완하는 방법으로 일반화가능도 이론은 측정 상황에서 발생하는 다중 오차 요인을 동시에 분석하고, 오차 요인이 검사 점수에서 차지하는 비율이 어느 정도인가를 확인하여, 이러한 오차 분산의 크기를 감소시킴으로써 신뢰도를 향상시킬 수 있는 효율적인 측정 조건을 제시한다.

일반화가능도 이론은 크게 G-연구와 D-연구로 나누어 수행한다. G-연구에서는 허용 가능한 관찰전집과 관련된 요인들의 분산성분 추정값을 산출하며, D-연구에서는 G-연구 결과 산출된 요인들의 분산성분 추정값을 바탕으로 각 오차요인의 수를 늘림으로써 신뢰도인 일반화가능도계수가 적정 수준에 도달하는 최적의 검사조건을 규명한다(Brennan, 2001a; Shavelson & Webb, 1991). 일반화가능도이론에 대한 좀 더 자세한 설명이 필요한 독자는 Brennan(2001a), 김성숙과 김양분(2001), 또는 이규민(2003)을 참조할 수 있다(강애남, 이규민, 2006).

본 연구는 기존에 개발된 수학적 창의성 검사 결과가 공개된 채점 자료에 다변량 일반화가능도 분석을 수행함으로써 각 채점 영역의 점수를 종합한 합성점수의 신뢰도를 바탕으로 최적의 채점 영역별 가중치를 결정하는 방법을 제시하고자 한다. 구체적인 연구문제는 다음과 같다.

첫째, 수학적 창의성 검사에서 합성점수 산출시 채점 영역별로 피험자의 검사 점수에 영향을 미치는 요인들의 상대적인 영향력은 어느 정도인가?

둘째, 수학적 창의성 검사에서 합성점수의 신뢰도는 어떠한가?

셋째, 수학적 창의성 검사에서 합성점수 산출시 채점 영역별 최적의 가중치는 얼마인가?

II. 이론적 배경

1. 심리 요소법에 근거한 수학적 창의성 측정

수학적 창의성을 심리 요소법을 바탕으로 측정하는 접근은 확산적 사고를 창의성으로 간주하는 Guilford(1967)의 제안을 수학 영역에 적용하여 계량화한 것으로, 수학적 창의성과 관련되어 있다고 판단하는 문항을 제시하고 그것으로부터 유창성, 유연성, 독창성(또는 새로움), 그리고 정교성(또는 조직성) 등을 계량화하는 방법이다(유윤재, 2007; Haylock, 1987). 여기서 유창성은 문제 상황에 유의미한 답으로서 여러 가지 반응 또는 아이디어를 낼 수 있는 능력으로, 융통성은 서로 다른 범주의 반응 또는 아이디어를 낼 수 있는 능력으로, 독창성은 다른 사람들과는 다른 참신하며, 질적으로도 수준 높은 반응 또는 아이디어를 낼 수 있는 능력으로, 그리고 정교

성은 산출한 반응 또는 아이디어를 보다 구체화하고, 세밀하게 다듬어 일반화할 수 있는 능력으로 정의한다. 각각의 평가 준거는 의미 있는 반응의 수, 반응의 유형별 가지 수, 반응의 상대적 회귀 빈도와 질적인 참신성, 가치, 그리고 반응의 구체성, 세밀성, 일반화이다(김홍원 외, 1996).

외국의 경우 Hollands(1972), Brandau와 Dossey(1979), Leikin(2009), 그리고 Shriki(2010, 2013)는 수학적 창의성을 유창성, 유연성, 독창성, 그리고 정교성으로 측정하였다. 특히 Shriki(2013)는 영역별로 준거 집단에서 가장 높은 점수를 100으로 상정한 상대적 점수를 산출하고, 이를 바탕으로 창의성에서 각 채점 영역이 중요하다고 생각되는 교사의 신념을 상대 가중치로 반영하여 창의성 검사의 총점을 계산하는 방법을 제시하였다. 단, 독창성의 상대적 점수를 산출하기 위해서는 먼저 사전에 결정된 준거가 있어야 하는데, 예를 들어 33%를 독창성 점수를 산출하기 위한 상한선으로 정한다면 30명의 학생으로 구성되어있는 집단의 경우에는 10명이하의 학생들이 응답한 문제에 대해서만 점수를 부여하였다. 또한 Evans(1964), Jensen(1973), Balka(1974), Tuli(1980), Silver(1997), Imai(2010), Kim, Cho와 Ahn(2004), 그리고 Mann(2005)는 수학적 창의성을 유창성, 융통성, 그리고 독창성으로 측정하였다. Leikin(2009)는 이전 연구인 Leikin과 Lev(2007)에서 활용하였던 유창성, 융통성, 그리고 독창성에 Ervynck(1991)에서 강조한 통찰력을 바탕으로 수학적 창의성을 측정하였다. 한편 Mainville(1972)는 수학적 창의성을 유창성과 독창성으로 측정하였다. 반면에 Haylock(1997)은 수학적 창의성을 융통성과 독창성만으로 측정하였다. 즉 창의성을 측정하는 전형적인 Torrance 검사(Torrance, 1966; 1974)에서는 유창성을 측정하고 있지만 수학 맥락에서 유창성이라는 준거는 창의성을 측정하는데 융통성에 비해 훨씬 덜 유용하게 쓰인다는 것이다. 예를 들어 만일 학생들에게 “답이 4가 되는 문제를 만드시오”라는 문제가 주어진다면, 어떤 학생은 “5 - 1”, “6 - 2”, “7 - 3”,... 를 계속 적어 낼 수 있다. 결국 이 학생은 유창성에서 높은 점수를 받게 되겠지만 어떤 창의성도 보여주고 있지 않다는 것이다. 그러나 Foster(1970), Bauer(1971), 그리고 Maxwell(1974)은 유창성만을 기준으로, Krutetskii(1976)

는 융통성만을 기준으로 수학적 창의성을 측정하였다.

우리나라에서 김홍원 외(1996)는 수학적 창의성 판별도구를 개발하면서 수학적 창의성을 유창성, 유연성, 독창성, 그리고 정교성으로 측정하였다. 황혜정 외(1997), 그리고 이강섭과 심상길(2005)도 수학적 창의성을 유창성, 유연성, 독창성, 그리고 정교성으로 측정하였다. 김홍원 외(1997), 송상현(1998), 권오남 외(2000), 이강섭, 황동주와 서종진(2003a, 2003b), 그리고 Lee, Hwang과 Seo(2003)는 수학적 창의성을 유창성, 융통성, 그리고 독창성으로 측정하였다. 권오남 외(2000)에서는 유창성과 융통성의 경우는 문항에 대한 반응의 개수와 유형의 가지 수에 2배하여 점수를 부여하였으며, 독창성의 경우는 반응의 빈도수를 확인하여 전체 반응의 4%이상~6%이하는 1점, 2%이상~4%미만은 3점, 그리고 2%미만은 5점으로 차등화하여 점수를 산출하고, 세 영역의 총점은 20점으로 제한하였다. 한편 이강섭과 심상길(2005)은 유창성, 융통성, 독창성, 정교성에 각각 5, 5, 6, 2점씩 배점하였다. 점수를 분배한 근거로는 유창성과 융통성은 중복되는 부분이 있고, 창의성에서는 독창성이 보다 중요하다는 연구자들의 판단을 이유로 들었다. 반면에 Lee 외(2003)에서는 독창성 점수의 경우에는 상한선을 부여하지 않았으며, 총점을 산출하지 않았다. 이대현(2014)에서는 초등 영재학생 10명을 대상으로 Leikin(2009)이 제시한 수학적 창의성 채점 방법을 변형하여 유창성, 융통성, 그리고 독창성으로 수학적 창의성을 측정하였다. 이 방법에 의해 산출된 창의성 점수는 심진법의 수에 기초하여 총점수가 개인이나 집단에 의해 산출되는 문제에 대한 해들의 모임인 해결 공간(solution space)과 그 수에 초점을 맞추어 해석 가능하다는 장점을 설명하였다. 구체적으로 독창성 점수는 각 반응이 수학적으로 유의미해야 한다는 활용 가능성(availability)과 검사 대상 학생들의 교육과정에 기반을 둔 교재에 포함되지 않은 해결방법과 특별한 교육 시기나 상황에서 적용될 수 있는 교육 과정에 기반을 둔 해결 방법인 비관습적 해결 방식(unconventional approach)을 근거로 산출하였다. 한편 조석희와 황동주(2007)는 유창성만을 기준으로 수학적 창의성을 채점하는 방식이 효율적이며 타당하고 신뢰롭다고 주장하였다.

이처럼 수학적 창의성은 선행 연구를 수정하거나 전

문가들의 판단에 의하여 연구자들마다 다양한 방법으로 측정되고 있다. 그러나 현재까지 측정학적 특성을 바탕으로 채점 영역별 가중치를 탐색한 연구는 거의 이루어지고 있지 않는 실정이다.

2. 수학 분야에서 일반화가능도 이론의 적용

일반화가능도 이론을 적용한 교과로는 언어(김경선, 이규민, 강승혜, 2010; 김보라, 이규민, 2012; 김성숙, 2011; 신동일, 2001; 이영식, 신상근, 2004; 이향, 2012; 이현숙, 2012; 조재운, 2009; 최연희, 2001; Brown, 2013; Brown, Glasswell, & Harland, 2004; Li & Brennan, 2007; Lynch & McNamara, 1998; Matsumura et al., 2008; Wan & Brennan, 2005), 과학(김양분, 1989; 이규민, 황경현, 2007; 이기영, 김찬중, 2005; 이기영, 안희수, 2005; 이옥수, 2007; 이현숙, 2012; Gao, Shavelson, & Baxter, 1994; Messick, 1989; Powers & Brennan, 2009; Shavelson, Lung, & Lewin, 1993; Webb, Schlackman, & Sugrue, 2000), 그리고 체육(김도연, 허종관, 2002; 안애정, 1997; 이기영, 2004; 이지연, 2005; 정영한, 2009; 조정환, 1991; Godbout & Schutz, 1983; Morrow, Fridye, & Monaghan, 1986; Stamm & Moore, 1980; Stratford, Norman, & McIntosh, 1989; Taylor, 1979) 등이 있으며, 다양한 연구자들에 의해 연구가 수행되고 있다.

외국의 경우, 일반화가능도 이론을 수학에 적용한 연구를 측정 대상에 따라 구분하면 학생과 교사로 나누어 볼 수 있다. 먼저 학생이 측정 대상인 경우로, 이들이 치른 대규모의 표준화된 수학 성취도 검사 결과에 일반화가능도 이론을 적용함으로써 적정 수준의 신뢰도 수준에 도달하는데 필요한 최적의 측정 조건을 탐색하고, 검사의 효율성을 신장시킬 수 있는 방법을 제시하는 연구들이 수행되었다(American College Testing, 1989; Colton, 1993; Jarjoura & Brennan, 1982, 1983). 다음으로 수학 교사가 측정 대상인 경우로, 이들의 수업을 관찰하여 신뢰롭고 효율적인 평가점수를 산출하기 위해 미국의 경우에는 필요한 관찰횟수와 관련하여 일반화가능도 이론이 주(state) 차원에서 비교적 최근에 활발하게 적용되고 있다(김성연, 2014a, 2014b; Hill, Charalambous, & Kraft, 2012; Hill et al., 2008; Hill, Schilling, & Ball, 2004; Ho & Kane, 2013; Kane & Staiger, 2012; Matsumura et

al., 2008; Measures of Effective Teaching, 2009; Wilhelm & Kim, 2015).

우리나라의 경우에는 수학과 관련하여 김명희(2000)는 8명이 채점한 수학 수행평가 결과에 일반화가능도 이론을 활용하여 채점자간 신뢰도를 산출한 결과 피험자에 대한 채점자간 기준이 어느 정도 일정하며, 채점자 판단의 변동이 적음을 밝혔다. 김명희(2005)는 채점 자동화 시스템 구축을 위한 문항 채점 자료에 일반화가능도 이론을 적용하여 적정 수준의 일반화가능도계수를 얻을 수 있는 문항 수와 채점자 수를 탐색하였다. 또한 수학영제 선발과 관련하여 관찰·추천제 이후 공통적으로 사용되는 선발 도구인 교사추천서와 자기소개서를 총체적 채점과 분석적 채점으로 평가한 자료에 일반화가능도 이론을 적용하여 일반화가능도계수를 최대화할 수 있는 효율적인 측정 조건(김성연, 한기순, 2013; 김성찬, 김성연, 한기순, 2012)과 분석적으로 채점한 자료에서 인지적, 수학적, 정의적, 그리고 사회적 영역별로 가중치를 탐색한 연구(김성연, 한기순, 2014)들이 있다.

이상에서 살펴본 연구들은 대부분 단변량 일반화가능도 이론을 적용하고 있으며, 특히 우리나라의 경우에는 수학에 일반화가능도 이론을 적용한 연구는 상대적으로 그 수가 적다. 더욱이 수학적 창의성 검사에 일반화가능도 이론을 적용한 연구는 거의 이루어지지 않고 있는 실정이다.

III. 연구방법

1. 분석 자료

본 연구는 김용범(2009)이 공개한 수학적 창의성 검사의 채점 결과를 예시적 분석 자료로 사용하였다. 이 검사를 선정한 이유는 첫째, 수학적 창의성을 측정하는데 적합하다고 알려진 다양한 해결 방법이 있는 개방형으로 문제가 구성(Becker & Shimada, 1997; Burjan, 1991; Ervynck, 1991; Hoeflinger, 1998; Krutetskii, 1976; Leikin, 2009; Leikin & Lev, 2007; Polya, 1973;

Sheffield, 1994; Silver, 1997)되어 있기 때문이다. 둘째, 일반적으로 수학적 창의성은 확산적 산출물을 측정하는데 고려하는 요인인 유창성, 융통성, 독창성, 정교성 등이 사용되고 있지만, 초등 수학영제들에게는 정교성을 제외하는 것이 더 적합하다고 알려져 있기 때문이다. 즉, 초등학생의 수학적 창의성을 측정하는데 있어서는 풀이 과정에서 논리적인 증명이나 진술의 정교성보다는 아직 다듬어지지 않은 새롭고 기발한 아이디어를 다양하게 제시하는 것이 더 적합하기 때문이다(송상현, 1998; 이대현, 2014; Leikin, 2009). 또한 김홍원 외(1997)와 이강섭과 심상길(2005)에서는 수학적 창의성은 정교성을 측정하지 않아도 충분히 측정할 수 있음을 실증적으로 밝혔다. 셋째, 본 연구의 목적은 분석 결과로 산출된 가중치를 일반화하는 것이 아니라 각 연구자들 및 교육 현장에서 제작한 수학적 창의성 검사에 적용할 수 있는 다변량일반화가능도 분석 방법을 제시하는 것이므로, 채점 자료가 공개되어있는 검사가 필요하기 때문이다.

구체적으로 김용범(2009)은 중학교 수학교과와 수와 연산, 문자와 식, 함수, 확률과 통계, 기하의 5영역에서 6개 문항의 창의적 수학능력 검사 문항을 개발하였으며, 개발된 평가문항의 효과성을 검증하기 위하여 초등 수학영제 30명을 대상으로 총 60분 동안 검사를 실시하였다. 수학적 창의성은 본 검사에서 활용한 창의적 수학능력 검사에서 산출된 점수로 정의하였다. 검사결과에 대한 채점은 유창성, 융통성, 그리고 독창성으로 실시되었다. 유창성은 각 문항에 대한 옳은 반응의 개수로 점수를 부여하였다. 융통성은 다른 학생들과는 다른 관점에서 문제를 바라보고 풀이를 제시하거나, 규칙성을 찾아낸 경우를 추출하여 점수를 부여하였다. 독창성은 다른 학생들과 차별화된 아이디어를 반영한 풀이방법 및 규칙성의 발견에 대하여 차등적으로 점수를 부여하였다. 먼저 검사 개발자인 김용범(2009)이 1차로 채점을 하고, 이를 다른 수학교사가 재검하였으며 그 결과를 공개하였다. 각 검사 문항의 배점 및 채점 영역 분류표는 [표 1]과 같다.

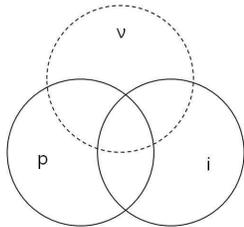
[표 1] 수학적 창의성 검사의 문항 배점 및 채점 영역

[Table 1] Item scores and scoring domains of a mathematical creativity test

문항 번호	문제 이름	관련 내용 영역	측정하고자하는 능력	채점 영역 (점)			시간 (분)
				유창성	융통성	독창성	
1	9 × 9 단표의 확률 계산하기	확률과 통계	수학적 추론능력 (귀납적 사고능력)	10	3	3	10
2	수의 피라미드	규칙성과 함수	규칙성 찾기 및 직관적 통찰능력	10	3	3	10
3	블록의 개수 맞추기	규칙성과 함수	규칙성 찾기 및 문제 만들기	10	3	3	10
4	생년월일 맞추기	문자와 식	정보의 조직화 능력	10	3	3	10
5	곱셈의 재발견	수와 연산	수학적 추상화 능력	10	3	3	10
6	Nim 게임에서 반드시 이기기	논리적 사고	거꾸로 문제해결하기	10	3	3	10

2. 분석 방법

본 연구에서는 채점 영역(ν)을 유창성(ν_1), 융통성(ν_2), 그리고 독창성(ν_3)의 3수준으로 한정된 고정국면으로 정의하고, 채점 영역 내의 문항과 검사를 치르는 학생은 무한 전집에서 임의로 표집된 것으로 가정하여 임의국면으로 정의하였다. 학생들이 치른 검사의 문항들은 모두 동일한 채점 영역에 의해 점수가 부여되므로, 다변량 일반화가능도 분석의 $p \times i$ 설계를 적용하였으며, 이를 그림으로 도식화하면 [그림 1]과 같다.



[그림 1] $p \times i$ 의 G-연구 설계

[Fig. 1] G-study for the $p \times i$ design

[그림 1]은 단변량 일반화가능도 분석의 체제인 $p \times i \times \nu$ 설계를 기본으로 하여 그려지지만, 채점 영역인 ν 국면($\nu_l, l = 1, 2, 3$)이 각각 독립적으로 추정된다는 점을 강조하기 위하여 점선으로 나타낸다는 점에서 차이가 있다. [그림 1]에서 실선으로 원이 분할되는 부분에는 분산성분이 존재하며, 실선과 점선이 교차되는 부분에는 공분산 성분이 존재한다. 분산과 공분산 성분 추

진치를 바탕으로 유창성, 융통성, 독창성의 채점 영역별로, 그리고 채점 영역별 가중치를 고려한 합성점수를 바탕으로 학생 효과, 문항 효과, 그리고 학생과 문항의 상호작용효과를 포함하는 잔차 효과의 크기를 상대적으로 비교할 수 있다.

$p \times i$ 설계에서 3개의 채점 영역인 $\nu_l (l = 1, 2, 3)$ 에서 학생들의 수학적 창의성 검사 점수는 식(1)과 같이 분해된다. 여기서 X_{piv_l} 는 l 번째 채점 영역 ν_l 의 문항 i 에 대한 학생 p 의 점수, μ_{ν_l} 은 해당 채점 영역에서 전체 학생과 문항에 대한 X_{piv_l} 의 평균, 그리고 ω, ξ, ψ 는 각 채점 영역에서 p 와 i 의 주효과와 상호작용을 포함한 잔차 효과를 나타낸다.

$$\begin{aligned}
 X_{piv_1} &= \mu_{\nu_1} + \omega_p + \omega_i + \omega_{pi} \\
 X_{piv_2} &= \mu_{\nu_2} + \xi_p + \xi_i + \xi_{pi} \\
 X_{piv_3} &= \mu_{\nu_3} + \psi_p + \psi_i + \psi_{pi}
 \end{aligned}
 \tag{1}$$

$p \times i$ 설계에서 임의국면에 대한 각 채점 영역을 나타내는 $\nu_l (l = 1, 2, 3)$ 의 분산 및 공분산 성분은 식(2)와 같다. 식(2)에서 대각선 성분은 분산($\sigma_{\nu_l}^2(\gamma), l = 1, 2, 3$)을, 대각선 아래 성분은 공분산($\sigma_{\nu_l \nu_m}(\gamma), l = 1, 2, 3, m = 1, 2, 3, l \neq m$)을, 그리고 대각선 위의 성분은 측정의 오차를 고려한 상관계수($\rho_{\nu_l \nu_m}(\gamma), l = 1, 2, 3, m = 1, 2, 3, l \neq m$)를 나타낸다. [그

림 1]에서와 같이, p 와 i 는 ν 와 교차하므로 \sum_p, \sum_i , 그리고 \sum_{pi} 에서 모두 ν 국면($\nu_l, l = 1, 2, 3$)의 채점 영역간에 공분산 성분이 존재하게 된다.

$$\sum_{\gamma} = \begin{bmatrix} \sigma_{\nu_1}^2(\gamma) & \rho_{\nu_1\nu_2}(\gamma) & \rho_{\nu_1\nu_3}(\gamma) \\ \sigma_{\nu_2\nu_1}(\gamma) & \sigma_{\nu_2}^2(\gamma) & \rho_{\nu_2\nu_3}(\gamma) \\ \sigma_{\nu_3\nu_1}(\gamma) & \sigma_{\nu_3\nu_2}(\gamma) & \sigma_{\nu_3}^2(\gamma) \end{bmatrix} \quad (\gamma = p, i, pi) \quad (2)$$

여기서 ν 국면($\nu_l, l = 1, 2, 3$)의 각 채점 영역별 분산 성분은 단변량 일반화가능도 분석에서 분산 성분을 추정하는데 사용되는 방법과 같은 식(3)으로, ν 국면($\nu_l, l = 1, 2, 3$)의 각 채점 영역간 공분산 성분은 식(4)로, 그리고 ν 국면($\nu_l, l = 1, 2, 3$)의 각 채점 영역간 추정의 오차를 고려한 상관계수는 식(5)를 이용하여 각각 추정된다. 단, MS 는 제곱합을 자유도로 나눈 값인 평균제곱합을 의미하며, n_i, n_p , 그리고 n_{pi} 는 학생 수, 문항 수, 그리고 학생 수와 문항 수의 곱을 나타낸다. 또한 γ 는 p, i, pi 이며, l 과 m 은 1, 2, 3이고, $l \neq m$ 이다.

$$\widehat{\sigma}_{\nu_l}^2(p) = \frac{MS_{\nu_l}(p) - MS_{\nu_l}(pi)}{n_i}, \quad (3)$$

$$\widehat{\sigma}_{\nu_l}^2(i) = \frac{MS_{\nu_l}(i) - MS_{\nu_l}(pi)}{n_p},$$

$$\widehat{\sigma}_{\nu_l}^2(pi) = MS_{\nu_l}(pi),$$

$$\widehat{\sigma}_{\nu_l\nu_m}(\gamma) = \frac{n_{\gamma}}{n_{\gamma} - 1} \left(\frac{\sum_{\gamma} \bar{X}_{p\nu_l} \bar{X}_{p\nu_m}}{n_{\gamma}} - \bar{X}_{p\nu_l} \bar{X}_{p\nu_m} \right) \quad (4)$$

$$\rho_{\nu_l\nu_m}(\gamma) = \frac{\sigma_{\nu_l\nu_m}(\gamma)}{\sqrt{\sigma_{\nu_l}^2(\gamma)\sigma_{\nu_m}^2(\gamma)}} \quad (5)$$

2) D-연구 설계

D-연구에서는 G-연구에서 추정된 \sum_p 행렬은 동일하게 적용되며, 나머지 행렬들은 ν 국면($\nu_l, l = 1, 2, 3$)에서 각 채점 영역별로 정한 D-연구의 표본 수로 나누어 산출된다. D-연구에서 각 오차국면은

일반화 전집에서의 평균에 대한 의미가 포함되어 있기 때문에 대문자로 표시한다. 식(2)의 분산 공분산 행렬을 이용한 D-연구의 분산 성분은 식(6)으로, 공분산 성분은 식(7)과 같이 나타낼 수 있다.

$$\sigma_{\nu_l}^2(\bar{\alpha}) = \frac{\sigma_{\nu_l}^2(\alpha)}{\pi_{\nu_l}(\bar{\alpha}|\tau)} \quad (6)$$

$$\sigma_{\nu_l\nu_j}^2(\bar{\alpha}) = \frac{\sigma_{\nu_l\nu_j}^2(\alpha)}{\pi_{\nu_l\nu_j}(\bar{\alpha}|\tau)} \quad (7)$$

식(6)과 식(7)에서 $\bar{\alpha}$ 는 D-연구 설계의 p, I , 그리고 pI 의 효과 중 하나를 나타내며, $\sigma_{\nu_l}^2(\alpha)$ 는 G-연구에서의 분산성분을 의미하며, τ 는 추정대상으로 p 를 나타낸다. $\sigma_{\nu_l}^2(\bar{\alpha})$ 는 고정효과 국면인 ν 국면($\nu_l, l = 1, 2, 3$)의 각 채점 영역에서 추정된 분산성분을 $\pi_{\nu_l}(\bar{\alpha}|\tau)$ 으로 나눈으로써 얻을 수 있다. 여기서 $\pi_{\nu_l}(\bar{\alpha}|\tau)$ 는 $\bar{\alpha}$ 에 표시된 각 효과에 대한 D-연구의 표본 수를 나타내며, 이때 추정대상을 나타내는 p 의 표본 수는 제외한다.

D-연구에서 식(2)와 같은 ν 국면($\nu_l, l = 1, 2, 3$)의 각 채점 영역별로 분산 및 공분산 성분이 산출되면, 다중 오차요인을 고려한 신뢰도인 일반화가능도계수를 구할 수 있다. 본 연구에서 사용한 수학적 창의성 검사는 학생들의 상대적인 서열을 파악하는데 관심을 둔 검사이다. 따라서 고전검사의 진점수에 해당하는 진집점수의 편차, 그리고 진집점수 편차의 추정치인 관찰점수 편차와의 차이를 나타내는 상대오차 분산을 활용한다. 즉, $p \cdot \times I \cdot$ 설계에서 \sum_p 행렬은 진집점수 분산 행렬인 \sum_{τ} 행렬이 되며, \sum_{pi} 행렬은 상대오차 분산 행렬인 \sum_{δ} 행렬이 된다. 위의 절차에 따라 상대오차 분산이 구해지면, 진집점수 분산을 진집점수 분산과 상대오차 분산의 합으로 나누어 상대평가에 적용 가능한 신뢰도인 일반화가능도계수를 얻을 수 있다.

또한 본 연구는 각 채점 영역별로 가중치를 두어 합산한 합성점수의 신뢰도를 산출하여야하기 때문에 G-연구 결과 ν 국면($\nu_l, l = 1, 2, 3$)의 각 채점 영역별로

구한 분산 및 공분산 성분 추정치에 가중치를 두어 합성한 합성점수의 분산 및 공분산 성분 추정치를 구하여야 한다. 가중치에는 상대 가중치, 명목 가중치, 그리고 실질 가중치가 있다. 상대 가중치(relative weight)는 검사 개발자가 개발 초기 단계에 합성점수에 기여하는 채점 영역의 상대적인 배점 비율을 의미하며, 상대 가중치의 합은 항상 1이다. 본 연구에서 수학적 창의성 검사 원자료의 분석 결과는 채점 영역별 상대 가중치를 문항 수에 비례하게 모두 .333으로 두었다. 또한 검사 개발자는 합성점수의 총점을 먼저 정하고 각 채점 영역의 총점을 합산하는데 필요한 가중치를 정하게 되는데 이를 명목 가중치(nominal weight)라고 한다. 본 연구에서는 총점을 검사 개발자가 정한 96점과 동일하게 적용하여 상대 가중치에 96을 곱한 32를 명목 가중치로 정하였다. 합성점수 C 를 산출하는데 사용된 각 채점 영역별 명목 가중치가 ω_{ν_l} ($l = 1, 2, 3$)일 때, 합성 전집점수 분산 ($\sigma_C^2(p)$), 합성 상대오차 분산($\sigma_C^2(\delta)$), 그리고 합성 일반화가능도 계수(Ep^2)는 각각 (8)-(10)의 식을 이용하여 얻을 수 있다. 단, $n_{i\nu_l}$ ($l = 1, 2, 3$)은 고정된 채점 영역인 ν_l ($l = 1, 2, 3$)에서 i 의 표본 수를 나타낸다.

$$\sigma_C^2(p) = \sum_l \sum_m \omega_{\nu_l} \omega_{\nu_m} \sigma_{\nu_l \nu_m}(p), \quad (l, m = 1, 2, 3) \quad (8)$$

$$\sigma_C^2(\delta) = \sum_l \sum_m \omega_{\nu_l} \omega_{\nu_m} \frac{\sigma_{\nu_l \nu_m}(pi)}{n_{i\nu_l}}, \quad (l, m = 1, 2, 3) \quad (9)$$

$$Ep^2 = \frac{\sigma_C^2(p)}{\sigma_C^2(p) + \sigma_C^2(\delta)} \quad (10)$$

반면에 실질 가중치(effective weight)는 합성 전집점수 분산 및 오차 분산에 고정국면인 3개의 채점 영역이 실제로 기여하는 정도를 타낸다. 즉, 명목 가중치가 채점 영역별로 사전에 검사 개발자의 판단에 의해 결정되는 것이라면, 이러한 명목 가중치에 근거하여 산출된 실질 가중치는 전집점수 분산 및 오차분산에 대한 채점 영역의 통계적인 기여도를 의미한다. 합성 전집점수 분산과 합성 상대오차 분산에 대한 채점 영역별 실질 가중치 ($ew_{\nu_l}(p)$, $ew_{\nu_l}(\delta)$, $l = 1, 2, 3$)는 (11)-(12)의 식에서 얻을 수 있다(Brennan, 2001a; Powers & Brennan;

2009).

$$ew_{\nu_l}(p) = \frac{\omega_{\nu_l} \sum_m \omega_{\nu_m} \sigma_{\nu_l \nu_m}(p)}{\sigma_C^2(p)}, \quad (l = 1, 2, 3) \quad (11)$$

$$ew_{\nu_l}(\delta) = \frac{\omega_{\nu_l} \sum_m \omega_{\nu_m} \sigma_{\nu_l \nu_m}(\delta)}{\sigma_C^2(p)}, \quad (l = 1, 2, 3) \quad (12)$$

$p \cdot \times i \cdot$ 설계의 분산, 공분산 추정치, 그리고 일반화가능도계수는 mGENOVA(A multivariate GENeralized Analysis Of VAriance System) 프로그램(Brennan, 2001b)과 엑셀의 매크로를 사용하여 산출하였다. 또한 모든 분석 결과는 평균점수 척도에서 보고하였다.

VI. 결과 분석 및 논의

1. 채점 영역별로 창의성 검사 점수에 미치는 요인들의 영향력 탐색

$p \cdot \times i \cdot$ 설계의 채점 영역별 분산 성분과 채점 영역간 공분산 성분의 추정치와 해당 성분이 전체 분산에서 차지하는 비율, 그리고 각 채점 영역 간 측정오차를 고려한 상관계수는 [표 2]와 같다. 유창성 영역에서는 학생 분산이 47.65%로 가장 크게 나타났으며, 융통성과 독창성 영역에서도 각각 37.45%와 39.31%로 상대적으로 크게 나타났다. 반면에 문항 분산은 채점 영역과 상관없이 .00%에서 15.01%로 가장 작게 나타났다. 이는 수학적 창의성 검사 점수에 영향을 미치는 요인은 학생 간의 수학적 창의성 차이에서 나타나며, 문항 자체의 특성이 검사 점수에는 거의 영향을 미치지 않고 있음을 의미한다. 학생과 문항의 상호작용이 포함되어 있는 잔차 분산은 유창성 영역을 제외한 나머지 채점 영역에서는 가장 크게 나타났다. 잔차 분산이란 원인을 규명하지 못한 오차 분산으로, 본 연구 설계에 포함되지 못한 다른 국면들의 필요성에 대한 간접적인 설명으로 G-연구 결과에서 흔히 나타나는 결과이다. 학생과 문항의 상호작용 효과란 수학적 창의성 검사에서 학생들의 상대적 순위가 문항에 따라 달라질 수 있음을 의미한다.

[표 2] $p^* \times i^*$ 설계의 G-연구 결과

[Table 2] Results of G-study for the $p^* \times i^*$ design

분산 성분	채점 영역	유창성	융통성	독창성
학생(p)	유창성	2.4313 (47.65)	<.9010>	<.8591>
	융통성	.5866	.1743 (37.45)	<.9954>
	독창성	.9128	.2832	.4643 (39.31)
문항(i)	유창성	.7657 (15.01)	.00000	.00000
	융통성	.0000	.0000 (.00)	.00000
	독창성	.0308	.0000	.0000 (.00)
잔차 (pi, e)	유창성	1.9054 (37.34)	.2912	.7169 (60.69)
	융통성	.2985	.2912 (62.55)	.7169
	독창성	.5236	.3316	.7169 (60.69)
전체 분산		5.1025 (100.00)	.4655 (100.00)	1.1812 (100.00)

주1. <>는 측정오차를 고려한 상관계수 값을 표시함.

주2. ()는 채점 영역별 분산성분이 전체 분산에서 차지하는 퍼센트임.

[표 2]에서 각 학생이 유창성, 융통성, 그리고 독창성 채점 영역에서 받은 점수가 함께 변하는 정도를 나타내는 각 채점 영역 간의 측정오차를 고려한 상관계수는 .8591부터 .9954로 높게 나타났다. 이는 한 채점 영역에서 높은 점수를 받은 학생은 다른 영역에서도 높은 점수를 받았음을 의미하며 세 영역에서의 전집점수들이 거의 선형관계라는 것을 의미한다. 그러나 이것이 그 영역들을 무시해야 한다는 것을 의미하는 것은 아니며 (Brennan, 2001a; Webb, Shavelson, & Maddahian, 1983), 각 채점 영역이 수학적 창의성이라는 구인의 각기 다른 측면을 측정하고 있다는 것을 나타낸다. 또한 학생 요인을 제외한 채점 영역 간의 공분산이 대부분 낮게 나타났으므로 이는 오차들 사이에는 별다른 관계가 존재하지 않는다는 의미로 수학적 창의성 검사의 신뢰도가

가 높다고 해석할 수 있다.

2. 수학적 창의성 검사의 신뢰도 분석

G-연구 분석 결과 산출된 [표 2]의 분산과 공분산 성분을 바탕으로, 원 자료의 표본 수와 같게 D-연구를 수행한 결과 얻은 채점 영역별 신뢰도와 합성점수의 신뢰도 결과는 [표 3]과 같다. 채점 영역별 신뢰도를 살펴본 결과 유창성이 .8845로 가장 높게 나타났으며, 다음으로 독창성, 그리고 융통성의 순으로 나타났다. 문항 수를 상대 가중치로, 그리고 총점을 96점으로 정하여 산출한 합성점수의 신뢰도는 (8)에 의해 .8841인 것으로 나타났다. 또한 수학적 창의성 검사 개발 시 가장 많이 보고된 Cronbach α 를 산출하기 위하여 문항만을 오차요인으로 갖는 단변량 $p \times i$ 설계를 적용하여 분석한 결과 .9181로 나타났다.

일반적으로 신뢰도의 크기가 얼마 이상이 되어야 한다는 절대적인 기준은 제시되어 있지 않고, 그 기준 또한 연구자들마다 다르다. 그러나 일반화가능도 이론을 적용한 연구들은 대체로 일반화가능도계수는 .80이상을 적정수준의 신뢰도로 판단하고 있다(Brennan, 2001a; Dunbar, Koretz, & Hoover, 1991; Shavelson, Baxter, & Gao, 1993). 따라서 본 연구에서 활용한 수학적 창의성 검사의 신뢰도는 양호하다고 판단할 수 있다.

[표 3] $p^* \times I^*$ 설계의 D-연구 결과

[Table 3] Results of D-study for the $p^* \times I^*$ design

	유창성	융통성	독창성	합성점수
전집점수분산	2.4313	.1743	.4643	6794.3424
상대오차분산	.3176	.0485	.1195	891.0234
일반화가능도계수	.8845	.7822	.7953	.8841

D-연구는 G-연구 설계와 같은 설계뿐만 아니라 다른 설계에서도 적정 수준의 신뢰도에 도달하는데 필요한 효율적인 측정 조건을 탐색하는데 활용된다. 따라서 $p^* \times I^*$ 설계와 $p^* \times I^*$ 설계에서 문항 수를 조절하면서 일반화가능도계수의 변화들을 살펴본 결과는 [표 4]와 같다. $p^* \times I^*$ 설계에서는 채점 영역마다 같은 문항이 사용되므로, 문항 수를 1부터 10까지 변화한 결과를 제시하였다. 반면에 $p^* \times I^*$ 설계에서 문항은 채점

영역에 따라 다르므로, 본 연구 설계와의 비교를 용이하게 하기 위하여 총 문항 수가 6개 이하인 경우로 제한한 결과 중 일반화가능도계수가 .80에 도달하는 경우만 제시하였다.

[표 4] $p^* \times I^*$ 설계와 $p^* \times I^\circ$ 설계의 다양한 D-연구 결과

[Table 4] Various results of D-study for the $p^* \times I^*$ design and the $p^* \times I^\circ$ design

$p^* \times I^*$ 설계			
문항수	전집점수분산	상대오차분산	일반화가능도계수
1	6794.3424	5346.1402	.5596
2	6794.3424	2673.0701	.7177
3	6794.3424	1782.0467	.7922
4	6794.3424	1336.5350	.8356
5	6794.3424	1069.2280	.8640
6	6794.3424	891.0234	.8841
7	6794.3424	763.7343	.8990
8	6794.3424	668.2675	.9105
9	6794.3424	594.0156	.9196
10	6794.3424	534.6140	.9271

$p^* \times I^\circ$ 설계			
문항 수	전집점수 분산	상대오차 분산	일반화가능도 계수
2, 1, 2 (5)	8308.5248	2040.7686	.8028
1, 1, 4 (6)	5318.4834	1296.4782	.8040
2, 3, 1 (6)	6181.8802	1382.7531	.8172
2, 2, 2 (6)	6794.3424	1491.7444	.8200
2, 1, 3 (6)	7443.7934	1600.7365	.8230
3, 1, 1 (5)	11827.2138	2478.8992	.8267
3, 2, 1 (6)	9393.2187	1796.0018	.8395
3, 1, 2 (6)	10116.9631	1904.9931	.8415
4, 1, 1 (6)	13337.9923	2209.2506	.8579

주1. 음영 부분은 G-연구 설계에서와 동일한 국면의 조건 수에 해당하는 6문항을 나타냄.

주2. $p^* \times I^\circ$ 설계에서 문항 수는 유창성, 융통성, 독창성 채점 영역별 순서를 나타내며, ()는 총 문항 수를 나타냄.

[표 4]에 제시된 D-연구 결과는 [표 2]의 G-연구 결과를 바탕으로 산출된 것이다. 먼저 상단의 $p^* \times I^*$ 설계 분석 결과, 원래 수학적 창의성 검사와 같은 6개의 문항으로 구성된 경우, 이 검사의 일반화가능도계수는 [표 3]의 결과와 일치하는 .8841이다. 또한 단변량

$p \times i$ 설계를 적용할 때의 일반화가능도계수 .9181에 도달하기 위해서는 적어도 9개의 문항이 필요한 것으로 나타났다. 또한 문항 수가 늘어날수록 일반화가능도계수도 증가하는 것으로 나타났다. 다음으로 하단의 $p^* \times I^\circ$ 설계 분석 결과, 대체로 유창성 영역의 문항 수가 많은 경우에 일반화가능도계수가 증가하는 것으로 나타났다. 특히 유창성 영역의 문항 수가 3, 융통성과 독창성 영역의 문항 수가 1이어서 총 5문항으로 구성된 수학적 창의성 검사의 일반화가능도계수가 유창성 영역의 문항 수가 1이나 2이면서 총 6문항으로 구성된 검사의 일반화가능도계수보다 높은 것으로 나타났다. 그러나 총 문항 수를 6개로 고정하는 경우에는 어떤 경우에도 원래 수학적 창의성 검사의 신뢰도에는 도달하지 못하는 것으로 나타났다.

3. 채점 영역별 최적의 가중치 탐색

본 연구에서 사용된 수학적 창의성 검사는 채점 영역별로 모두 6개의 문항으로 구성되어 있으므로, 문항수에 비례하여 상대 가중치를 .333으로 할당하는 경우, 실질 가중치는 유창성, 독창성, 그리고 융통성의 순으로 각각 .59, .25, 그리고 .16으로 나타났다. 이는 본 연구에서 사용하고 있는 수학적 창의성 검사가 유창성은 10점을 만점으로, 그리고 융통성과 독창성의 경우는 3점을 만점으로 채점 영역에 따라 배점에 차이가 있음이 어느 정도 반영되어 있다고 해석할 수 있다. 그러나 여전히 유창성과 독창성은 처음 검사 개발자가 상정한 의도보다 낮게, 그리고 융통성은 높게 반영되어 있음을 알 수 있다. 실질 가중치를 배점에 비례하게 유창성, 융통성, 그리고 독창성의 순으로 .625, .1875, 그리고 .1875가 되기 위해서는 상대 가중치를 .36, .39, 그리고 .25로 두어야 하며, 실질 가중치를 문항 수에 비례하여 모두 .333이 되기 위해서는 상대 가중치를 .15, .53, 그리고 .32로 정해야 하는 것으로 나타났다. [표 5]는 상대 가중치를 각 채점 영역마다 .1씩 변화시켜 가면서 일반화가능도계수를 분석한 결과에서, 원자료의 일반화가능도계수보다 높은 14가지의 경우를 제시하였다. 채점 영역별 최적의 가중치는 유창성, 융통성, 그리고 독창성을 각각 .5, .4, 그리고 .1로 나타났으며 이 때 일반화가능도계수는 .8907인 것으로 나타났다.

[표 5] 상대 가중치 조합에 따른 실질 가중치와 일반화가능도계수

[Table 5] Effective weights and generalizability indices based on the combination of relative weights

	상대 가중치 조합			실질 가중치			일반화가능도계수	
	유창성	융통성	독창성	유창성	융통성	독창성		
<i>ew(p)</i>	5: 4: 1	.7794	.1583	.0623	.8907			
	6: 3: 1	.8400	.1052	.0551	.8905			
	5: 3: 2	.7608	.1166	.1226	.8902			
	6: 2: 2	.8222	.0691	.1087	.8902			
	7: 2: 1	.8878	.0629	.0493	.8895			
	6: 1: 3	.8052	.0340	.1608	.8892			
	7: 1: 2	.8715	.0311	.0974	.8892			
	5: 2: 3	.7427	.0764	.1809	.8892			
	4: 5: 1	.7021	.2264	.0715	.8890			
	4: 4: 2	.6825	.1773	.1403	.8886			
	8: 1: 1	.9270	.0285	.0445	.8880			
	5: 1: 4	.7252	.0375	.2373	.8875			
	4: 3: 3	.6636	.1301	.2062	.8874			
	4: 2: 4	.6456	.0849	.2695	.8855			
	<i>ew(δ)</i>	5: 4: 1	.7719	.1477	.0640	.8907		
		6: 3: 1	.8478	.0693	.0534	.8905		
5: 3: 2		.7494	.0953	.1319	.8902			
6: 2: 2		.8258	.0388	.1103	.8902			
7: 2: 1		.9016	.0269	.0449	.8895			
6: 1: 3		.8000	.0159	.1691	.8892			
7: 1: 2		.8808	.0107	.0931	.8892			
5: 2: 3		.7227	.0537	.2014	.8892			
4: 5: 1		.6655	.2831	.0772	.8890			
4: 4: 2		.6441	.1973	.1585	.8886			
8: 1: 1		.9397	.0067	.0383	.8880			
5: 1: 4		.6932	.0223	.2706	.8875			
4: 3: 3		.6183	.1271	.2409	.8874			
4: 2: 4		.5897	.0716	.3216	.8855			

주. 상대가중치 조합은 유창성, 융통성, 독창성 순서를 나타냄.

한편 수학적 창의성을 3개의 채점 영역이 아닌 하나의 영역에 대해서만 채점하는 경우를 살펴보았다. 이를 위해 문항 수와 상대 가중치를 원자료에서의 수학적 창의성과 같게 하고, 전체 문항 6개를 하나의 영역인 유창성, 융통성, 그리고 독창성에서만 각각 채점하는 경우로 분석한 결과 합성점수의 상대오차보다 각각 6.74배, 1.53배, 그리고 2.11배 크게 나타남으로써 3개의 채점 영역을

모두 고려하는 것이 더 바람직한 것으로 나타났다.

V. 결론 및 제언

본 연구는 기존에 개발된 수학적 창의성 검사를 활용하여 합성점수 산출시 채점 영역별로 학생들의 창의성 검사 점수에 영향을 미치는 요인들의 상대적인 영향력을 살펴보고, 이를 바탕으로 신뢰도와 채점 영역별 최적의 가중치를 탐색하는 방법을 제시하였다. 이러한 방법론적 틀은 수학적 창의성 검사뿐만 아니라 교사들이 제작하여 학교에서 시행하는 검사에까지 광범위하게 활용될 수 있다. 예시 자료에 대해 다변량 일반화가능도 분석을 수행한 연구 결과는 다음과 같다.

첫째, 채점 영역별로 창의성 검사 점수에 영향을 미치는 요인들의 영향력을 탐색한 결과, 잔차 분산을 제외하면 학생 분산이 가장 크게 나타났으며, 문항 분산은 채점 영역과 상관없이 가장 작게 나타났다. 이는 수학적 창의성 검사 점수에 학생의 수학적 창의성 차이가 잘 반영되어 있으며, 문항의 난이도는 수학적 창의성 검사 점수에는 거의 영향을 미치지 않고 있다고 해석할 수 있다. 또한 학생 요인에서 각 채점 영역간의 측정오차를 고려한 상관계수는 모두 높게 나타났다. 이는 한 채점 영역에서 높은 점수를 받은 학생이 다른 영역에서도 높은 점수를 받았음을 의미한다. 반면에 문항이나 잔차 요인의 공분산은 학생 요인에 비해 대부분 낮게 나타남으로써 채점 영역별 오차들 사이에는 별다른 관계가 존재하지 않는 것으로 나타났다.

둘째, 수학적 창의성 검사의 신뢰도 분석 결과, 채점 영역별로는 유창성, 독창성, 그리고 융통성의 순으로 높게 나타났으며, 문항 수를 상대 가중치로 준 합성점수의 신뢰도는 .8841인 것으로 나타났다. 또한 문항만을 오차 국면으로 고려하여 단변량 일반화가능도 분석을 한 결과 얻게 되는 Cronbach α 는 .9181로 나타났다. 문항과 채점 영역을 모두 고려하여 다변량 일반화가능도 분석을 한 결과에서 .9181 수준의 신뢰도를 얻기 위해서는 9개의 문항이 필요한 것으로 나타났다.

셋째, 검사 개발자의 의도를 반영하여 수학적 창의성 검사의 합성점수를 산출하는데 있어 문항 수에 비례하게 각 채점 영역을 .333으로 할당하는 경우, 실질 가중치는

유창성, 독창성, 그리고 융통성의 순으로 각각 .59, .25, 그리고 .16으로 나타났다. 반면에 실질 가중치가 채점 영역별로 총점에 비례하게 유창성, 융통성, 그리고 독창성의 순으로 .625, .1875, 그리고 .1875가 되기 위해서는 명목 가중치를 .36, .39, 그리고 .25로 할당해야 하는 것으로 나타났다. 또한 본 연구에 활용한 수학적 창의성 검사에서 채점 영역별 최적의 가중치는 유창성, 융통성, 그리고 독창성을 각각 .5, .4, 그리고 .1인 경우로 나타났으며, 이 때 일반화가능도계수는 .8907인 것으로 나타났다. 이 경우 수학적 창의성 검사의 총점을 96점이라고 할 때, 실제 배점은 유창성, 융통성, 그리고 독창성별로 48점, 38점, 그리고 10점으로, 각 문항별 배점은 현재 10점에서 8점으로, 현재 3점에서 6점으로, 그리고 현재 3점에서 1점으로 조정하는 것이 필요하다. 한편 수학적 창의성 검사에서 3개의 채점 영역을 고려하지 않고 하나의 영역으로만 채점을 하는 경우에는 채점 영역과 상관없이 오차 분산이 합성 상태오차 분산보다 크게 나타났다. 이러한 분석 결과는 초등학교를 대상으로 한 이대현(2014)의 연구 결과에서 수학적 창의성은 유창성, 융통성, 그리고 독창성을 고려하여야 한다는 연구 결과와 일치한다. 반면에 수학적 창의성에서 상대빈도와 표준정규분포를 이용하여 독창성에 더 높은 가중치를 부여하여야 한다는 이강섭, 심상길(2005)과 이강섭(2010)의 연구결과와는 일치하지 않는다. 한편, 조석희와 황동주(2007)의 연구 결과와는 부분적으로 일치한다. 본 연구 결과에서도 세 개의 채점 영역 중 유창성이 가장 신뢰로운 채점 영역으로 나타났다지만, 조석희와 황동주(2007)의 연구에서처럼 유창성만으로 창의성을 측정하는 경우에도 상당히 양호한 신뢰도를 확보할 수 있다는 결과와는 상충하는 부분이 있다. 이는 검사 대상 집단이 달라진 것뿐만 아니라 검사의 신뢰도를 Cronbach α 로만 상정하였기 때문임을 하나의 이유로 들 수 있다.

이상의 예시 자료를 활용한 분석 결과를 바탕으로 시사점을 논하면 다음과 같다. 첫째, 다변량 일반화가능도 분석 결과는 다양하게 개발된 수학적 창의성 검사에서 최적의 측정 조건을 탐색하는데 적용할 수 있다. 먼저 검사 결과에 영향을 미칠 수 있는 요인들을 미리 밝혀내어 측정하고자 하는 창의성 이외의 다른 요인들의 영향을 배제할 때 활용될 수 있다. 즉, 수학적 창의성 검사에

서 문항의 난이도, 시행 횟수, 검사 유형에 따라 학생들의 순위가 달라지지는 않는지, 또는 특정 문항이 특정 학생에게 유리하게 작용하고 있지는 않는지 등을 하나하나 점검해 봄으로써 수학적 창의성 검사 점수에서 학생의 수학적 창의성 차이가 잘 반영되도록 검사를 구성할 수 있을 것이다. 이를 바탕으로 각 요인들의 표본 수를 다양하게 변화시켜 가면서 검사의 신뢰도를 탐색함으로써, 적정 수준의 신뢰도를 얻을 수 있는 수학적 창의성 검사를 구성하는 최적의 측정 조건을 확정하는데 유용하게 사용될 수 있다.

둘째, 채점 영역별 최적의 가중치를 탐색하는 방법은 수학적 창의성 검사뿐만 아니라 다양한 검사에서 활용될 수 있다. 즉, 선택형과 서답형으로 구성되어 있는 수학적 성취도 검사, 교사 추천서와 자기소개서, 영재행동특성 체크리스트로 구성되어 있는 수학영재선발도구, 그리고 학습지도와 생활지도영역으로 구성되어 있는 교원능력개발평가에 이르기까지 채점 내용 영역별 가중치 조합 결과를 탐색함으로써 합성점수 산출 방식을 확정하는데 필요한 방법론적인 틀로서 기능할 수 있다.

셋째, 기존에 당연시 여기는 검사에 대한 가정들은 틀릴 수 있음을 보여준다. 먼저 대부분의 검사에서 보고하고 있는 Cronbach α 는 과대추정된 신뢰도일 수 있다. 이는 고전검사이론을 바탕으로 문항만을 단일 오차요인으로 상정하기 때문이다. 다음으로 대부분의 검사 결과시 제공하는 합성점수는 문항들의 합이나 평균으로 보고함으로써 암묵적으로 영역별 동일한 가중치를 부여하고 있다는 것을 가정하고 있다. 그러나 본 연구 결과에서 분석한 바와 같이 각 채점 영역의 분산이나 각 채점 영역간의 공분산에 따라 실제 합성 전집점수 분산에 기여하는 바가 달라짐으로써 실질 가중치는 다르게 된다. 따라서 연구자가 상정한 처음의 가중치를 얻었는지에 대한 검증이 필요하다.

넷째, 수학적 창의성의 측정은 다차원 영역을 반영해야 한다. 경험적 자료를 근거로 창의성이 다차원적 속성을 지녔다는 것은 수학적 창의성 발달을 위해 각 채점 영역별로 학교 수업이나 수학 교과서에서 고려해야 할 방안들을 제시할 수 있다. 실례로 양승민(2012)은 남학생들에게는 유창성과 융통성을 발달시키기 위해서는 수학적 추상화와 일반화 및 적용능력, 독창성을 위해서는 정

보의 조직화에 초점을 둔 수업이 필요하며, 여학생들에게는 유창성과 독창성 발달을 위해 정보의 조직화 능력, 융통성 발달을 위해서는 정보의 조직화 능력과 공간화/시각화 능력에 초점을 둔 수업이 필요하다고 제안하였다. 또한 윤기옥(2008)은 중학교 1학년 수학 교과서에서 수학적 창의성을 활용한 문제가 있는지 분석한 결과, 창의적 문제들이 창의성의 3가지 요인을 모두 만족시키는 다양한 유형의 문제이기 보다는 대부분 유창성의 기준을 만족시키는 비슷한 유형을 문제들로 이루어져 있다고 밝혔다. 따라서 본 연구 결과와 함께 교수법과 교육과정에 대한 질적 분석을 통하여 실제 학교 현장에서 유창성뿐만 아니라 융통성과 독창성도 함께 신장시킬 수 있는 방법이 제시되어야 할 것이다.

마지막으로 본 연구의 제한점은 다음과 같다. 첫째, 본 연구에서 제시한 최적의 채점 영역별 가중치는 예시적 자료로 활용한 특정 연구자가 개발한 수학적 창의성 검사에만 국한된다. 둘째, 본 연구에서는 수학적 창의성을 확산적 사고에 초점을 맞춘 유창성, 융통성, 그리고 독창성의 세 영역으로만 한정하였다. 셋째, 본 연구에서는 공개된 자료를 활용함으로써 수학적 창의성 검사 점수는 한 명의 채점자에 의해서만 평가가 이루어져있다. 따라서 다양한 채점 영역이 있는 공신력 있는 수학적 창의성 검사를 다수의 학생들을 대상으로 실시하여, 여러 명의 채점자에 의해 평가가 이루어진 검사 점수에 다변량 일반화가능도 분석을 수행하는 것이 필요하다.

참 고 문 헌

- 강애남, 이규민 (2006). 학생들의 동료평가를 활용한 수행평가 결과의 일반화가능도 분석. 교육평가연구 19(3), 107-121.
- Kang, A. N., & Lee, G. M. (2006). A generalizability theory approach to investigating the generalizability of performance assessment using student peer reviews. Journal of Educational Evaluation 19(3), 107-121.
- 교육부 (1992). 제6차 교육과정의 개요. 교육부 고시 제 1992-11호.
- Ministry of Education. (1992). *The 6th curriculum*. Notification No. 1992-11 of the Ministry of Education.
- 교육부 (1997). 제7차 교육과정. 교육부 고시 제1997-15호.
- Ministry of Education. (1997). *The 7th curriculum for elementary and middle schools*. Notification No. 1997-15 of the Ministry of Education.
- 교육부 (2014). 2015 문·이과 통합형 교육과정 총론 주요 사항. 세종: 교육부.
- Ministry of Education. (2014). *The main particular of 2015 National Curriculum draft for an integration of Arts and Sciences*. Sejong: Ministry of Education.
- 교육부 (2015). 제2차 수학교육 종합계획. 세종: 교육부
- Ministry of Education. (2015). *The 2nd curriculum for mathematical education*. Sejong: Ministry of Education.
- 교육과학기술부 (2009). 2009개정 교육과정 총론. 교육과학기술부 고시 제2009-41호.
- Ministry of Education, Science, and Technology. (2009). *The 2009 curriculum revision*. Notification No. 2009-41 of the Ministry of Education, Science, and Technology.
- 교육인적자원부 (2007). 2007년 개정 교육과정 총론. 교육인적자원부 고시 제 2007-79호.
- Ministry of Education & Human Resources Development (2007). *The 2007 curriculum revision*. Notification No. 2007-79 of the Ministry of Education & Human Resources Development.
- 권오남, 김정효, 윤희선, 박재희, 태혜경, 윤태숙 (2000). 창의적 문제해결력 중심의 수학 교육과정 적용 및 효과분석. 수학교육 논문집 5, 87-116.
- Kwon, O. N., Kim, J. H., Youn, H. S., Park, J. H., Tae, H. K., & Youn, T. S. (2000). Application and examination the effect of mathematics curriculum to enhance creative problem solving abilities. The Mathematical Education 39(2), 81-99.
- 김경선, 이규민, & 강승혜 (2010). 일반화가능도 이론을 적용한 한국어 말하기 성취도 평가의 신뢰도와 오차 요인 분석. 한국어 교육 21(4), 51-75.
- Kim, K. S., Lee, G. M., & Kang, S. H. (2010). Analysis of error sources and estimation of reliability in a Korean speaking achievement test by applying generalizability theory. Journal of Korean Language Education 21(4), 51-75.
- 김도연, 허종관 (2002). 일반화가능도이론을 적용한 주관적 배구기능검사의 신뢰도 추정. 한국체육측정평가학회지 4(2), 15-28.

- Kim, D. Y., & Heo, J. K. (2002). Generalizability coefficient estimation of subjectively volleyball test. *The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science* 4(2), 15-28.
- 김명화 (2005). 체점 자동화 시스템 구축을 위한 수학 국성형 문항 체점의 일반화가능도 연구. 교육문제연구 22, 205-222.
- Kim, M. H. (2005). An application of the generalizability to constructed response items in mathematics for computer automated scoring. *Journal of Research in Education*, 22, 205-222.
- 김명희(2000). 수행평가의 타당도 검증을 위한 측정학적 접근. 박사학위논문, 이화여자대학교.
- Kim, M. H. (2000). *A psychometric approaches of the validity verifications in performance assessments*. (Unpublished doctoral dissertation, Ewha Woman University, Seoul, Korea).
- 김보라, 이규민 (2012). 일반화가능도 이론을 적용한 초등학교 쓰기 수행평가의 총체적 체점과 분석적 체점 방식 비교. 교육학연구 50(4), 49-76.
- Kim, B. R., & Lee, G. M. (2012). A comparison of holistic and analytic scoring methods for elementary school writing assessment by applying generalizability theory. *Korean Journal of Educational Research* 50(4), 49-76.
- 김부운, 이지성 (2005). 수학적 창의성의 평가방안에 대한 모색. 한국학교수학회논문집 8(3), 327-341.
- Kim, B. Y. & Lee, J. S. (2005). A note on the assesment of mathematical creativity. *Journal of the Korean School Mathematics Society* 8(3), 327-341.
- 김부운, 이지성 (2006). 수학에서의 창의적 태도의 측정 도구 개발과 그 적용. 수학교육 45(1), 25-34.
- Kim, B. Y. & Lee, J. S. (2006). Development and its applications of the CAS-K in Mathematics. *The Mathematical Education* 43(1), 25-34.
- 김성숙 (2011). 학문 목적 한국어 쓰기 능력에 대한 분석적 평가의 일반화가능도 검증. 한국어교육 22(3), 29-48.
- Kim, S. S. (2011). Analysis of error sources and estimation of reliability in an analytic evaluation on writing ability for academic purpose Korean by applying generalizability theory. *Journal of Korean Language Education* 22(3), 29-48.
- 김성숙, 김양분 (2001). 일반화가능도 이론. 서울: 교육과학사.
- Kim, S. S. & Kim, Y. B. (2001). *Generalizability theory*. Seoul: Kyoyookgwahaksa.
- 김성연 (2014a). 미국 수학교사의 교수 질 평가도구 분석을 통한 우리나라 수학 교원능력개발평가에서의 일반화가능도 이론 활용성 탐색. 수학교육 논문집 28(4), 431-455.
- Kim, S. Y. (2014a). Exploring the application of generalizability theory to mathematics teacher evaluation for professional development instructional quality assessment of mathematics teachers in the U.S. *Communications of Mathematical Education* 28(4), 431-455.
- 김성연 (2014b). 미국의 수업관찰평가 분석을 통한 우리나라 교원능력개발평가에서의 다변량 일반화가능도 이론 활용성 탐색. 한국교육 41(1), 5-29.
- Kim, S. Y. (2014b). Exploring the application of multivariate generalizability theory to teacher evaluation for professional development in Korea based on the analysis of classroom observations in the U.S. *The Journal of Korean Education* 41(1), 5-29.
- 김성찬, 김성연, 한기순 (2012). 관찰·추천에 의한 수학영재 선발 시 사용되는 교사추천서와 자기소개서 평가에 대한 일반화가능도 이론의 활용. 수학교육 논문집 26(3), 251-271.
- Kim, S. C., Kim, S. Y., & Han, K. S. (2012). An application of generalizability theory to self-introduction letter and teacher's recommendation letter used in identification of mathematically gifted students by observations and nomination. *Communications of Mathematical Education* 23(3), 251-271.
- 김성연, 한기순 (2013). 관찰·추천제에 의한 수학영재 선발 시 사용되는 교사추천서와 자기소개서 평가에 대한 다변량 일반화가능도 이론의 활용. 영재교육연구 23(5), 671-698.
- Kim, S. Y., & Han, K. S. (2013). An application of multivariate generalizability theory to teacher recommendation letters and self-introduction letters used in selection of mathematically gifted students by observation and nomination. *Journal of Gifted/Talented Education* 23(5), 671-698.
- 김성연, 한기순 (2014). 수학영재 선발에서 교사추천서와

- 자기소개서 채점내용 가중치에 따른 신뢰도 분석. 영재와 영재교육 13(1), 43-65.
- Kim, S. Y., & Han, K. S. (2014). Analysis of reliability coefficients depending on different domain weights in scoring teacher recommendation letters and self-introduction letters used in selection of mathematically gifted students. *The Journal of the Korean Society for the Gifted and Talented* 13(1), 43-65.
- 김양분 (1989). 일반화가능도이론에 의한 과학적 사고기능검사의 신뢰도 추정에 관한 연구. 교육평가연구 3(1), 183-209.
- Kim, Y. B. (1989). A study of reliability estimation of a scientific critical thinking tool based on generalizability theory. *Journal of Educational Evaluation* 3(1), 183-209.
- 김용대 (2003). 문제해결을 통한 수학적 일반성의 발견. 수학교육 논문집 15, 153-159.
- Kim, Y. D. (2003). The discovery of mathematical generalization through problem solving. *Communications of Mathematical Education* 15, 153-159.
- 김용범 (2009). 수학영재판별을 위한 창의적 수학능력 검사 문항 개발. 석사학위논문, 인천대학교.
- Kim, Y. B. (2009). The development of questions to measure the creativity mathematical ability. (Unpublished master's thesis, University of Incheon, Incheon, Korea).
- 김정은 (2010). 초등 수학 수업에서 수학 영재 판별 도구 개발에 관한 연구. 석사학위논문, 광주교육대학교.
- Kim, J. E. (2010). A Study on the developing identifying instrument for math gifted children in elementary mathematics classes. (Unpublished master's thesis, Gwangju National University of Education, Gwangju, Korea).
- 김홍원, 김명숙, 방승진, 황동주 (1997). 수학영재 판별도구 개발연구(II). 서울: 한국교육개발원.
- Kim, H. W., Kim, M. S., & Bang S. H., & Hwang, D. J. (1997). *A study on the development of the identification tools for mathematically gifted students(II)*. Seoul: Korean Educational Development Institute.
- 김홍원, 김명숙, 송상현 (1996). 수학영재 판별도구 개발 연구(I). 서울: 한국교육개발원.
- Kim, H. W., Kim, M. S., & Song, S. H. (1996). *A study on the development of the identification tools for mathematically gifted students(I)*. Seoul: Korean Educational Development Institute.
- 남승인 (2007). 수학 창의성 신장을 위한 평가 문항 개발 방안. 수학교육 논문집 21(2), 271-282.
- Nam, S. I. (2007). A study on the development evaluation item to extend mathematical creativity. *Communications of Mathematical Education*, 21(2), 271-282.
- 대전광역시교육청 (2011). 창의 인성 교육을 위한 평가방법 개선. 대전: 대전광역시교육청.
- Daejeon Metropolitan Office of Education. (2010). *The improvement of evaluation methods for creativity and personality education*. Daejeon: Dejeon Metropolitan Office of Education.
- 도종훈 (2007). 학교수학에서 추측과 문제제기 중심의 수학적 탐구 활동 설계하기. 수학교육 46(1), 69-79.
- Do, J. H. (2007). Designing Mathematical Activities Centered on Conjecture and Problem Posing in School Mathematics. *The Mathematical Education* 46(1), 69-79.
- 문교부 (1987). 제 5차 중학교 교육과정. 문교부 고시 제 1987-7호.
- Ministry of Education(1987). *The 5th curriculum for middle schools*. Notification No. 1987-7 of the Ministry of Education.
- 방승진, 최중오 (2010). 수학 학문적 창의성 신장을 위한 멘토십 프로그램 모형 개발. 영재교육연구 20(1), 205-229.
- Bang, S. J. & Choi, J. O. (2010). Implementations of Mentorship Program Model for the Academic Creativities of Mathematics. *Journal of Gifted/Talented Education* 20(1), 205-229.
- 변은진, 전평국 (2001). 개방형 문제를 활용한 평가가 수학적 창의력에 미치는 효과. 수학교육논문집 11, 259-277.
- Byeon, E. J. & Jeon, P. K. (2011). The effects of evaluation based on open-ended problems on mathematical creativity. *Communications of Mathematical Education* 21(2), 271-282.
- 서울특별시교육청 (2010). 창의성 계발을 위한 평가 개선 기본 계획. 서울: 서울시교육청.
- Seoul Metropolitan Office of Education. (2010). *The master plans of improving evaluation for creativity development*. Seoul: Seoul Metropolitan Office of Education.
- 성태제 (2010). 현대교육평가 (3판). 서울: 학지사.
- Seong, T. J. (2010). *Modern educational evaluation (3rd*

- edition). Seoul: Hakjisa.
- 손은영 (2005). 탐구학습이 초등학교 6학년 학생들의 수학적 창의력에 미치는 효과. 석사학위논문, 한국교원대학교.
- Son, E. Y. (2005). *The effects of inquiry learning on the mathematical creativity of 6th students*. (Unpublished master's thesis, Korea National University of Education, Chung-Buk, Korea).
- 송상헌 (1998). 수학 영재성 측정과 판별에 관한 연구. 박사학위논문, 서울대학교.
- Song, S. H. (1998). *A study on the measurement and discrimination of the mathematical giftedness*. (Unpublished master's thesis, Seoul National University of Education, Seoul, Korea).
- 신동일 (2001). 일반화가능도 이론 적용을 중심으로 한 말하기평가도구 타당도 검증 연구. 응용언어학 17(1), 199-221.
- Shin, D. I. (2001). Validation process of an EFL speaking test on G-theory and other analytical techniques. *The Applied Linguistics Association of Korea* 17(1), 199-221.
- 신승윤 (2013). 초등수학영재의 수학 창의성 문제해결력과 메타인지와의 관계. 석사학위논문, 대구교육대학교.
- Shin, S. Y. (2013). *The relationship between mathematically gifted elementary students' math creative problem-solving ability and metacognition*. (Unpublished master's thesis, Daegu National University of Education, Daegu, Korea).
- 안애정 (1997). 일반화가능도 이론에 의한 체조선수관 관정 오차원 분석. 박사학위논문, 명지대학교.
- Ahn, A. J. (1997). *Application of generalizability in analyzing sources of error variance of gymnastic judgement*. (Unpublished doctoral dissertation, Myongji University, Seoul: Korea).
- 양승민 (2012). 수학 창의성 발달을 위한 수학적 창의성과 수학적 사고력의 상관관계 분석. 석사학위논문, 고려대학교.
- Yang, S. M. (2012). *Analysis of correlation between mathematical creativity and mathematical thinking for the development of mathematical creativity*. (Unpublished master's thesis, Korea University, Seoul, Korea).
- 유경선 (2012). 공간능력과 수학적 창의성 신장을 위한 도형학습 과제의 효과. 석사학위논문, 서울교육대학교.
- Yoo, K. S. (2012). *Effects of figure learning tasks for improvement of spatial ability and mathematical creativity: focusing on a planar figure of a solid figure*. (Unpublished master's thesis, Seoul National University of Education, Seoul, Korea).
- 유윤재 (2002). 수학적 창의성의 검사. 한국수학교육학회 학술발표논문집 1, 1-22.
- Yoo, Y. J. (2002). A mathematical creativity test. *Studies in Mathematical Education*, 1, 1-22.
- 유윤재 (2003). 창의적 수학문제해결력 검사도구의 요소. 수학교육 논문집 17, 159-168.
- Yoo, Y. J. (2003). Components of a mathematical creative problem solving ability test. *Communications of Mathematical Education*, 17, 159-168.
- 유윤재 (2007). 수학 영재 교육. 서울: 교우사.
- You, Y. J. (2007). *Mathematically gifted education*. Seoul: Kyowoosa.
- 윤기욱 (2008). 중학교 1학년 수학 교과서에 나타난 창의성 측정 문항의 분석. 석사학위논문, 단국대학교.
- Yoon, K. O. (2008). *Analysis for creativity factors of problems in 7th-grade mathematics textbooks*. (Unpublished master's thesis, Dankook University, Seoul, Korea).
- 윤지윤 (2014). 초등수학영재와 일반학생의 수학불안과 수학적 창의성의 관계. 석사학위논문, 대구교육대학교.
- Yoon, J. Y. (2014). *The relations between mathematics anxiety and mathematical creativity of elementary mathematically gifted and non-gifted students*. (Unpublished master's thesis, Daegu National University of Education, Daegu, Korea).
- 이강섭 (2010). 수학 창의성 평가에서 독창성의 점수화 방법. 수학교육 49(1), 111-118.
- Lee, K. S. (2010). A scoring system for the originality in evaluation of mathematical creativity. *The Mathematical Education* 49(1), 111-118.
- 이강섭, 심상길 (2005). 창의성 증진을 위한 수학 활동 프로그램 평가 방법의 소개. 수학교육 논문집 19(1), 101-110.
- Lee, K. S., & Sim, S. K. (2005). An introduction of evaluation methods for mathematical activity program to develop creativity. *Communications of Mathematical Education* 19(1), 101-110.
- 이강섭, 황동주, 서종진 (2003a). 개방형 문항에 대한 중

- 학교 영재학생과 일반학생의 반응 연구. 수학교육논문집 17, 181-190.
- Lee, K. S., Hwang, D. J., & Seo, J. J. (2003a). A study of responses between gifted middle school students and regular school students about open ended problems. *Communications of Mathematical Education* 17, 181-190.
- 이강섭, 황동주, 서종진 (2003b). 일반 창의성(도형)과 수학 창의성과의 관련 연구: TTCT Figural A와 MCPSAT A를 바탕으로. 수학교육 42(1), 1-10.
- Lee, K. S., Hwang, D. J., & Seo, J. J. (2003b). A study on the relationship between general creativity and mathematical creativity: based on TTCT Figural A and MCPSAT A. *The Mathematical Education* 42(1), 1-10.
- 이경화, 고진영, 박숙희 (2007). 영재의 지능, 창의성, 자아개념간의 관계 분석. 영재와 영재교육 6(1), 147-162.
- Lee, K. H., Koh, J. Y., & Park, S. H. (2007). Gender differences in creative thinking and creative personality among primary school students. *The Journal of the Korean Society for the Gifted and Talented*, 8(1), 147-162.
- 이규민(2003). 단위검사 개념의 적용-일반화가능도 이론을 중심으로. 교육평가연구 16(1), 53-70.
- Lee, G. M. (2003). An application of the testlet concept-A generalizability theory approach. *Journal of Educational Evaluation* 16(1), 53-70.
- 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 scores form holistic and analytic scoring methods in performance assessments of an elementary school science class. *The Journal of Child Education* 16(4), 169-184.
- 이기영 (2004). 평가 유형과 채점 방식에 따른 중·고등학교 과학 수행 평가의 일반화가능도에 관한 연구. 박사학위논문, 서울대학교.
- Lee, K. Y. (2004). *A study on generalizability of science performance assessment according to assessment type and scoring method in middle & high school*. (Unpublished doctoral dissertation, Seoul National University, Seoul, Korea).
- 이기영, 김찬중 (2005). 한국 지구과학 올림피아드 문항 분석을 통한 문항의 질 향상 방안. 한국지구과학회지 26(6), 511-523.
- Lee, K. Y., & Kim, C. J. (2005). Analysis of Korea earth science olympiad items for the enhancement of item quality. *Journal of the Korean Earth Science Society* 26(6), 511-523.
- 이기영, 안희수 (2005). 중등학교 과학 수행평가의 평가 유형과 채점 방식 및 신뢰도 분석. 한국교육학회지 25(2), 173-183.
- Lee, K. Y., & Ahn, H. S. (2005). Analyses of assessment types, scoring methods and reliability of science performance assessment in middle and high school. *Journal of Korean Association for Science Education* 25(2), 173-183.
- 이대현 (2014). 다양한 해결법이 있는 문제를 활용한 수학적 창의성 측정 방안 탐색. 학교수학 16(1), 1-17.
- Lee, D. H. (2014). A study on the measurement in mathematical creativity using multiple solution tasks. *School Mathematics* 16(1), 1-17.
- 이동희, 김관수 (2010). 수학적 창의성과 태도 및 작업에 미치는 등산학습법의 적용과 효과. 한국초등수학교육학회지 14(1), 23-41.
- Lee, D. H. & Kim, P. S. (2010) The Effect of Climbing Learning Method on Mathematical Creativity and Attitude toward Mathematical Creativity. *Journal of Elementary Mathematics Education in Korea* 14(1), 23-41.
- 이영식, 신상근 (2004). 다변량 일반화가능도 이론에 의한 말하기 시험의 타당도와 신뢰도에 관한 연구. 외국어교육 11(2), 249-265.
- Lee, Y. S., & Shin, S. K. (2004). An investigation into the dependability of ratings in a German speaking test using the multivariate generalizability theory. *Foreign Languages Education* 11(2), 249-265.
- 이옥수 (2007). 자연계열 통합교과형 논술 문항 분석. 석사학위논문, 서울대학교.
- Lee, O. S. (2007). *An item analysis of scientific essay for science and engineering courses*. (Unpublished master's thesis, Seoul National University, Seoul, Korea).
- 이종희, 김기연 (2007). 창의적 생산력 신장의 교육목표 이해를 위한 수학영재의 수학적 창의성 개념 탐색. 수학교육 46(4), 445-464.
- Lee, J. H., & Kim, K. Y. (2007). A study on the concept of

- mathematical creativity in the mathematically gifted aspect. *The Mathematical Education* 46(4), 445-464.
- 이지연 (2005). 국가대표볼링선수 선발방법의 신뢰도 추정: 일반화가능도 이론의 적용. 박사학위논문, 한국체육대학교.
- Lee, J. Y. (2005). *Estimating the reliability of the selection method of the national team bowlers: application of the generalizability theory*. (Unpublished doctor's dissertation, Korea National Sport University, Seoul, Korea).
- 이향 (2012). 말하기 수행 평가에서 발음 범주 채점의 최적화 방안 연구-일반화가능도 이론을 활용하여. 한국어 교육 23(2), 301-329.
- Lee, H. (2012). A study on the optimal conditions in rating of pronunciation scale for speaking performance assessment-using generalizability theory. *Journal of Korean Language Education* 23(2), 301-329.
- 이현숙 (2012). 혼합형 검사의 문항 유형별 가중치에 따른 신뢰도 및 다변량 일반화가능도분석. 교육평가연구 25(1), 95-116.
- Lee, H. S. (2011). Multivariate generalizability analyses for mixed-format tests with various compositions of MC and CR item weights. *Journal of Educational Evaluation* 21(1), 95-116.
- 정영환 (2009). 일반화가능도 이론을 적용한 태권도 심판 판정의 신뢰도 분석. 박사학위논문, 국민대학교.
- Jung, Y. H. (2009). *Analysis on reliability of referees judgment in Taekwondo competition applying generalizability theory*. (Unpublished master's thesis, Kookmin University, Seoul, Korea).
- 정유화 (2011). 보드게임 활동이 수학적 창의성과 수학적 태도에 미치는 영향. 석사학위논문, 서울교육대학교.
- Jung, (2011). *The effects of boardgame activity on mathematical creativity and attitude towards mathematics*. (Unpublished master's thesis, Seoul National University of Education, Seoul, Korea).
- 조석희, 황동주 (2007). 중학교 수학 영재 판별을 위한 수학 창의적 문제해결력 검사 개발. 영재교육연구 17(1), 1-26.
- Cho, S. H., & Hwang, D. J. (2007). Math creative problem solving ability test for identification of the mathematically gifted middle school students. *Journal of Gifted/Talented Education* 17(1), 1-26.
- 조재윤 (2009). 일반화가능도 이론을 이용한 쓰기 평가의 오차원 분석 및 신뢰도 추정 연구. 국어교육 128, 325-357.
- Cho, J. Y. (2009). A study on the analysis of error source in writing assessment and reliability estimation using generalizability theory. *The Education of Korean Language* 128, 325-357.
- 조정환 (1991). 일반화가능도이론에 의한 피하지방검사, 농구기능검사의 신뢰도 추정. 박사학위논문, 한국체육대학교.
- Cho, J. H. (1991). *Generalizability coefficient estimation of skinfold thickness test and basketball skill test*. (Unpublished doctoral dissertation, Korean National College of Physical Education, Seoul, Korea).
- 최연희 (2001). GENOVA and FACETS analysis of an English writing test: tasks, raters, and scales. 영어 교육 56(2), 125-142.
- Choi, Y. H. (2001). GENOVA and FACETS analysis of an English writing test: tasks, raters, and scales. *English Teaching* 56(2), 125-142.
- 최은선 (2010). 초등수학 영재교육을 위한 수학적 창의성 개념연구. 석사학위논문, 서울교육대학교.
- Choi, E. S. (2010). *A study on the concept of mathematical creativity for the education of mathematically gifted children*. (Unpublished master's thesis, Seoul National University of Education, Seoul, Korea).
- 하수현, 이광호 (2014). Leikin 의 수학적 창의성 측정 방법에 대한 고찰. 한국초등수학교육학회지 18(1), 83-103.
- Ha, S. H. & Lee, K. (2014). A study about the Leikin's method of measuring mathematical creativity. *Journal of Elementary Mathematics Education in Korea* 18(1), 83-103.
- 한혜승 (1996). 제6차 교육과정에 따른 중학교 수학교과서 분석: 1학년을 중심으로. 석사학위논문, 이화여자대학교.
- Han, H. S. (1996). *Analysis of math textbooks according to the 6th curriculum in middle school: Focus on 1st-grade textbooks*. (Unpublished master's thesis, Ewha Woman University, Seoul, Korea).
- 한혜정 · 박순경 · 이근호 · 이승미 (2012). 시·도 교육청 수준 교육과정 지침 실태 분석 및 개선 방안. 서울:

- 한국교육과정평가원.
- Han, H. J., Park, S. K., Lee, K. H., & Lee, S. M. (2012). *A study on the improvement of the MPOE curriculum organization & implementation guideline through analyzing its current situation*. Seoul: Korea Institute for Curriculum and Evaluation.
- 홍주연, 한인기 (2014). 수학 영역에서 창의적 산출물 의미 척도. *수학교육* 53(2), 291-312.
- Hong, J. Y. & Han, I. K. (2014). A study on creative products semantic scale in mathematics. *The Mathematical Education* 53(2), 291-312.
- 황우형, 이유나 (2009). 중등 영재학생과 일반학생의 완벽주의 성향과 수학교과에 대한 정의적 특성과의 관계. *수학교육 논문집* 23(1), 1-38.
- Hwang, W. H. & Lee, Y. N. (2009). The relationship between mathematically gifted students and regular students in perfectionism and the affective traits. *Communication of Mathematical Education* 23(1), 1-38.
- 황혜정, 김홍원, 박경미, 김수환, 김신영, 채선희 (1997). 창의력 신장을 돕는 중학교 수학과 학습 평가 방법 연구. 서울: 한국교육개발원.
- Hwang, H. J. & Kim, H. W., Park, K.M., Kim, S. H., Kim, S. Y., & Chae, S. H. (1999). *A study on the evaluation methods in the middle school mathematics to improve the mathematical creativeness*. Seoul: Korea Educational Development Institute.
- American College Testing (1989). *Preliminary technical manual for the Enhanced ACT Assessment*. Iowa City, IA: American College Testing.
- Balka, D. S. (1974). *The development of an instrument to measure creative ability in mathematics*. (Unpublished doctoral dissertation, University of Missouri-Columbia).
- Bauer, G. R. (1971). *A study of the effects of a creative classroom, creative problems, and mathematics education on the creative ability of prospective elementary teachers* (Unpublished doctoral dissertation, Stanford University).
- Becker, J. P., & Shimada, S. (1997). *The Open-Ended Approach: A New Proposal for Teaching Mathematics*. VA: National Council of Teachers of Mathematics.
- Brandau, L. I., & Dossey, J. A. (1979). *Processes involved in mathematical divergent problem-solving*. San Francisco: American Educational Research Association.
- Brennan, R. L. (2001a). *Generalizability theory*. New York: Springer.
- Brennan, R. L. (2001b). *Manual for mGENOVA version 2.1*. Iowa City, IA: Center for advanced studies in measurement and assessment, The University of Iowa.
- Brown, J. D. (2013). Score dependability and decision consistency. In A. J. Kunnan (Ed.), *The companion to language assessment*. Oxford, UK: Wiley-Blackwell.
- Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 9, 105 - 121.
- Burjan, V. (1991). Mathematical giftedness—Some questions to be answered. In *Education of the Gifted in Europe: Theoretical and Research Issues: report of the educational research workshop held in Nijmegen (The Netherlands)* (pp. 165-170).
- Colton, D. A. (1993). *A Multivariate generalizability analysis of the 1989 and 1990 AAP mathematics test forms with respect to the table of specifications*. ACT research report series.
- Cronbach, L. J., Gleser, G. C., & Nanda, H. Rajaratnam. N.(1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: Wiley.
- 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.
- 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.
- Ervynck, G. (1991). Mathematical creativity. *Advanced mathematical thinking* (pp. 42-53). New York: Springer.
- Evans, E. W. (1964). *Measuring the ability of students to respond in creative mathematical situations at the late elementary and early junior high school level*. (Unpublished doctoral dissertation, University of Michigan).
- Foster, J. (1970). An Exploratory attempt to assess creative ability in mathematics. *Primary Mathematics* 8(1), 2-8.
- Gao, X., Shavelson, R. J., & Baxter, G. P. (1994). Generalizability of large-scale performance assessments in science: Promises and problems. *Applied Measurement in Education* 7(4), 323-342.
- Godbout, P., & Schutz, R. W. (1983). Generalizability of ratings of motor performances with reference to various observational designs. *Research Quarterly for Exercise and Sport* 54(1), 20-27.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Haylock, D. W. (1987). A framework for assessing mathematical creativity in school children. *Educational Studies in Mathematics* 18(1), 59-74.
- Haylock, D. W. (1997). Recognizing mathematical creativity in schoolchildren. *Zentralblatt für Didaktik der Mathematik* 27(2), 68-74.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction* 26(4), 430-511.
- 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., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal* 105(1), 11-30.
- Ho, A. D., & Kane, T. J. (2013). The reliability of classroom observations by school personnel. research paper. MET Project. *Bill & Melinda Gates Foundation*.
- Hoeflinger, M. (1998). Developing mathematically promising students. *Roepers Review* 20(4), 244-247.
- Hollands, R. (1972). Educational technology: aims and objectives in teaching mathematics. *Mathematics in School* 1(6), 22-23.
- Imai, T. (2010). The influence of overcoming fixation in mathematics towards divergent thinking in open-ended mathematics problems on Japanese junior high school students. *International Journal of Mathematical Education in Science and Technology* 31(2), 187-193.
- Jarjoura, D., & Brennan, R. L. (1982). A variance components model for measurement procedures associated with a table of specifications. *Applied Psychological Measurement* 6(2), 161-171.
- Jarjoura, D., & Brennan, R. L. (1983). *Multivariate generalizability models for tests developed from tables of specifications*. New Directions for Testing & Measurement.
- Jensen, L. R. (1973). *The relationships among mathematical creativity, numerical aptitude and mathematical achievement*. (Unpublished doctoral dissertation, University of Texas at Austin).
- Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teachers: combining high-quality observations with student surveys and achievement gains. MET Project. *Bill & Melinda Gates Foundation*.
- Kim, H., Cho, S., & Ahn, D. (2004). Development of mathematical creative problem solving ability test

- for identification of the gifted in math. *Gifted Education International* 18(2), 164-174.
- Krutetskii, V. A. (1976). *The psychology of mathematical abilities in school children*. University of Chicago Press.
- Lee, G., Brennan, R. L., & Frisbie, D. A. (2000). Incorporating the Testlet Concept in Test Score Analyses. *Educational Measurement: Issues and Practice* 19(4), 9-15.
- 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.
- Lee, K. S., Hwang, D. J., & Seo, J. J. (2003). A development of the test for mathematical creative problem solving ability. *Research in Mathematical Education*, 7(3), 163-189.
- Leikin, R. (2009). Exploring mathematical creativity using multiple solution tasks. *Creativity in Mathematics and the Education of Gifted Students*, 129-145.
- Leikin, R., Koichu, B., & Berman, A. (2009). Mathematical giftedness as a quality of problem-solving acts. *Creativity in mathematics and the education of gifted students*, 115-127.
- Leikin, R., & Lev, M. (2007). Multiple solution tasks as a magnifying glass for observation of mathematical creativity. In *Proceedings of the 31st international conference for the psychology of mathematics education* (Vol. 3, pp. 161-168).
- 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.
- Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing* 15(2), 158-180.
- Mainville, W. E. (1972). *A Study of the Effect of Mathematics Activity Materials: Upon Certain Aspects of Creative Thinking Ability of Prospective Elementary School Teachers*. (Unpublished doctoral dissertation, Michigan State University).
- Mann, E. L. (2005). *Mathematical creativity and school mathematics: indicators of mathematical creativity in middle school students*. (Unpublished doctoral dissertation, University of Connecticut).
- Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale". *Educational Assessment* 13(4), 267-300.
- Maxwell, A. A. (1974). *An Exploratory Study of Secondary School Geometry Students: Problem Solving Related to Convergent-Divergent Productivity*. (Unpublished doctoral dissertation, University of Marine).
- Measures of Effective Teaching. (2009). *Measures of effective teaching*. Retrieved from www.metproject.org
- Messick, S. (1989). Validity. in R. L. Linn (Ed.), *Educational measurement* (3rd ed). New York: Macmillan.
- Morrow Jr, J. R., Fridye, T., & Monaghan, S. D. (1986). Generalizability of the AAHPERD health related skinfold test. *Research Quarterly for Exercise and Sport* 57(3), 187-195.
- Polya, G. (1973). *How to solve it*. Princeton: Princeton University Press.
- Powers, S., & Brennan, R. L. (2009). *Multivariate generalizability analyses of mixed-format exams*. In Annual Meeting of the National Council on Measurement in Education, San Diego, CA.
- 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., Lung, H., & Lewin, B. (1993). *On*

- Concept Maps as Potential "Authentic" Assessments in Science. Indirect Approaches to Knowledge Representation of High School Science.* National Center for Research on Evaluation Standards and Student Testing, Office of Educational Research and Improvement.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. London: Sage Publications.
- Sheffield, L. J. (1994). *The development of gifted and talented mathematics students and the National Council of Teachers of Mathematics Standards*. PA: Diane.
- Shriki, A. (2010). Working like real mathematicians: developing prospective teachers' awareness of mathematical creativity through generating new concepts. *Educational studies in mathematics* 73(2), 159-179.
- Shriki, A. (2013). A model for assessing the development of students' creativity in the context of problem posing. *Creative Education* 4(07), 430.
- Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *Zentralblatt für Didaktik der Mathematik* 29(3), 75-80.
- Sriraman, B. (2004). The characteristics of mathematical creativity. *The Mathematics Educator* 14(1), 19-34.
- Stamm, C. L., & Moore, J. E. (1980). Application of generalizability theory in estimating the reliability of a motor performance test. *Research quarterly for exercise and sport* 51(2), 382-388.
- Stratford, P. W., Norman, G. R., & McIntosh, J. M. (1989). Generalizability of grip strength measurements in patients with tennis elbow. *Physical therapy* 69(4), 276-281.
- Taylor, J. L. (1979). Development of the physical education observation instrument using generalizability study theory. *Research Quarterly. American Alliance for Health, Physical Education, Recreation and Dance* 50(3), 468-481.
- Torrance, E. P. (1966). *Torrance tests of creative thinking: norms technical manual* (Research Edition). Princeton, NJ: Personnel Press.
- Torrance, E. P. (1974). *Norms-technical manual: Torrance tests of creative thinking*. Lexington, MA: Ginn and Company.
- Tuli, M. R. (1980). *Mathematical creativity as related to aptitude for achievement in and attitude towards mathematics*. (Unpublished doctoral dissertation, Panjab University).
- Wan, L., & Brennan, R. L. (2005). *Reliability of scores for tests composed of testlets: A comparison of approaches in three measurement models*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada.
- Webb, N. M., Shavelson, R. J., & Maddahian, E. (1983). *Multivariate generalizability theory*. In I. J. Pyans, Jr.(Ed.), *Generalizability theory: Inference and practical applications* (pp.67-81). San Francisco, CA: Jossey-Bass.
- Webb, N. M., Schlackman, J., & Sugrue, B. (2000). The dependability and interchangeability of assessment methods in science. *Applied Measurement in Education* 13(3), 277-301.
- 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.

Analysis of weights depending on scoring domains of the mathematical creativity test

Sungyeun Kim

Incheon National University, 119 Academi-ro, Yeonsu-gu, Incheon, Korea

E-mail : syk@inu.ac.kr

This study analyzes the mathematical creativity test as an illustrative example with scoring domains of fluency, flexibility and originality in order to make suggestions for obtaining maximum reliability based on a composite score depending on combinations of each scoring domain weights. This is done by performing a multivariate generalizability analysis on the test scores, which were allowed to access publicly, of 30 mathematically gifted elementary school students, and therefore error variances, generalizability coefficients, and effective weights have been calculated. The main results were as follows. First, the optimal weights should adjust to .5, .4, and .1 based on the maximum generalizability coefficient even though the original weights in the mathematical creativity test were equal for each scoring domain with fluency, flexibility and originality. Second, the mathematical creativity test using the three scoring domains of fluency, flexibility, and originality showed higher reliability than using one scoring domain such as fluency. These results are limited to the mathematical creativity test used in this study. However, the methodology applied in this study can help determine the optimal weights depending on each scoring domain when the tests constructed in various researchers or educational fields are composed of multiple scoring domains.

* ZDM Classification : C43

* 2000 Mathematics Subject Classification : 97C40

* Key words : mathematical creativity, multivariate generalizability analysis, weight