

학생발달 성취 관찰도구의 일반화가능성과 구인타당화 모형 검증

김 성 속
(인하대학교 교수)

《 요 약 》

이 연구는 학생발달 범주에서 기대되는 성취결과 영역을 측정하기 위한 관찰 프로파일의 일반화 가능성과 구인타당화 모형을 탐색하고자 하였다. 관찰 프로파일 또는 관찰도구는 아동의 개인-사회적 발달, 인지적 발달, 신체적 발달, 그리고 건강과 안전 영역에 이르기까지 총체적으로 파악하고자 개발된 도구이다. 학생 연령에 따라 구분할 수 있는 수준의 발달단계를 설정하고 11개의 구인과 55개의 하부구인을 추출하였다. 특히 구인 지도를 활용하여 발달단계에 따라 각 구인에서 개별 문항을 개발하는 과정을 전개하였다. 연구의 초점은 관찰도구 개발에서 제기되어 온 신뢰도 문제, 즉 결과의 일반화가능도에 대한 점검과 함께 이론적으로 구상한 기대결과요인을 각 구인이 어떻게 설명하는가를 확인하는 구인타당화 모형을 검증하였다. 일반화가능도 이론과 확인적 요인분석 방법(구조방정식 모형 중 공통요인분석의 일환)을 적용한 결과를 토대로 관찰도구 개선방안의 측정학적 청사진을 제시하였다. 관찰변수의 구성이 이론변수를 어떻게 설명하는가 다양한 모형을 제시하고 적합성 지수를 산출함으로써 이론 모형과 실제 변수의 변화를 가시화하였다.

주제어 : 기대성취결과, 관찰프로파일, 아동발달, 신뢰도, 구인타당화, 일반화가능도 이론, 확인적 요인분석

I . 들어가며

학생들의 행동특성을 파악하는 여러 가지 프로그램 중 최근 학생들의 발달상황을 연구하는 영역이 급격하게 증가하는 추세이다. 연구의 관심영역은 대부분 표준화 검사를 통하여 학생의 학업성취 발달상황을 탐색하는데 치중하여 왔다. 물론 학생 발달 상황 중 특히 인지영역을 반영하기 위해 표준화 학업성취검사 결과를 활용하는 것이 적절하지만 이를 통해 행동특성의 복합적 구성요인을 전반적으로 파악할 수는 없다. 따라서 아동발달 연구의 관점이

인지적 발달상황을 넘어서 다각적이며 총체적으로 접근하기 위한 일환으로 연구대상에 대한 관찰 연구가 확산되고 있다. 관찰방법은 실제 생활과 가장 근접한 상황에서 관찰 대상의 행동과 표현을 수집할 수 있는 참평가이다. 자기기록에 의한 검사응답이나 시험답안과 달리 관찰연구는 관찰자의 관찰에 의하여 연구대상의 특성을 파악하고 분석한다. 따라서 관찰자의 전문성과 객관성에 따라 질 높은 자료를 보장할 수 있는 반면 여러 명의 관찰자가 동원되는 경우 관찰자 내 일관성이나 관찰자간 합치도 수준이 주요 문제로 부각되어 왔다. 어떤 종류의 자료를 수집하던지 측정도구는 양질의 문항으로 구성되어 있을 때 목적에 충실할 수 있다. 좋은 측정도구는 측정하려는 것을 바르게 측정하고 있는가를 말하는 검사의 타당도를 충족시키는 동시에 정확하고 일관성있는 결과를 보장하고자 하는 검사의 신뢰도를 만족시켜야 한다. 관찰 자료는 연구대상 스스로가 응답하는 경우가 아니므로 관찰자가 관찰대상에서 보고자 하는 요인들을 관찰도구를 통해 얼마나 타당하게 구성하였는가가 매우 중요하다. 그리고 관찰결과가 어느정도 정확하고 신뢰로운가를 포함하여 관찰자가 서로 다른 이유로 관찰기록에 있어 오차가 발생하지 않도록 철저하게 관찰자 훈련이 선행되어야 한다.

미국의 경우 2001년 현 대통령 부시 행정부 내에서 'NCLB(No Child Left Behind)', 즉 '한 명의 낙오자도 없기'라는 법안을 채택함에 따라 각 주 마다 어떻게 학업성취 평가 결과를 활용함으로써 이 법안에 대응할 수 있는 적절한 대책을 구안하고 있다. 특히 1999년 연방정부에서 공포한 '공립학교 책무성 법안'(Public Schools Accountability Act)은 공립학교의 교육적 책무성 체계를 공식적으로 확립하기 위한 법안으로 공공교육을 받고 있는 모든 학생의 학업성취를 측정함과 동시에 전체적인 학력 향상에 그 목적을 두고 있다(<http://www.ed.gov>). 이와 더불어 최근 캘리포니아 주 교육부에서는 출생 후부터 14세까지 아동을 대상으로 폭넓은 영역의 발달 단계를 탐색하고 자료를 수집하고자 '성취기대결과 발달 프로파일' (DRDP : Desired Results Developmental Profile, 이하 DRDP와 혼용)을 개발하였다. DRDP는 개인-사회적 발달(social-personal), 인지적 발달(cognitive), 신체적(physical), 그리고 건강과 안전(health/safety) 네 영역에 대하여 도달해야할 성취기대결과를 관찰하여 발달 수준을 평가하는 것이다.

따라서 이 연구는 이론적 근거로 구현된 기대결과 영역을 학생들이 어떻게 수행하고 있는지 실제 학생을 접하는 상황에서 관찰 측정도구를 점검하고자 한다. 이미 여러 연구에서 지적하였듯이(Lanc, Liu, Ankermann, & Stone, 1996; Kim, 2000) 관찰도구 개발 시 직면하는 것은 도구에서 측정하고자 하는 구인이 얼마나 타당한가와 측정결과를 어느정도 일반화할 수 있는가를 검증하는 문제이다. 그러므로 이 연구의 목적은 사전조사를 통해 수집된 자료를 근거로 관찰도구의 양호도를 검증함으로써 관찰 프로파일(도구)의 구인 타당화 모형을 구축하고자 한다. 특히 구인타당도를 검증하기에 앞서 검사에서 측정하고자 하는 구인을 형성하는 이론적 구조모형을 설정하고 구조모형에 부합되는 개별구인의 적절성을 검토하였다. 타

당도와 신뢰도 수준을 보장하는 과정이 제시될 때 아동 발달수준을 파악하기 위한 관찰도구의 활용은 정당화될 수 있기 때문이다.

II. 이론적 배경

1. 성취 기대결과 (Desired Results)

Wilson(2002)은 최근 저서에서 측정도구를 개발하는 과정 중 최우선되어야 하는 것은 무엇을 측정하려고 하는가 즉, ‘구인’에 관한 연구라고 다시 한번 강조하였다. 측정과정에서 구인의 의미는 측정하고자 하는 대상의 기술, 태도, 성격, 지식, 행동 등에 관한 이론적 개념을 의미한다. 그리고 대부분 구인은 구인을 구성하는 하부요인이나 차원으로 세분화된다. 예를 들어 아동의 ‘사회성’이라는 구인은 어른과의 상호작용, 교우와의 관계, 자기 표현능력, 집단 내 역할 등과 같은 하부요인으로 구성되는 것이다.

캘리포니아 주 교육부에서 개발한 ‘성취기대결과 발달 프로파일’(DRDP : Desired Results Developmental Profile)을 구성하는 차원은 아동들이 연령별 발달단계에서 기대해야하는 성취수준은 무엇이며 어느정도 도달하는가를 나타내는 지표이다. 관찰 대상 구인을 설정하는 과정에서 아동발달의 기대되는 성취결과 영역에 대한 심층적인 문헌연구가 선행되었다. 초 중 학교를 다니는 학생에 대한 행동발달에 대한 최근 연구에 의하면(Carver, 1989; Cohen et al., 1983; Hadeed & Sylvia, 1999; Landerholm, Karr, & Mushi, 2000) 인지적 능력, 언어표현, 사회성, 감성발달, 신체발달 등 일반적 영역 뿐 아니라 자기 통제력, 다양성에 대한 지각, 안전에 대한 태도, 자아관의 발달 등 세부적 범주도 구분되어 제시하고 있다. 학교생활을 통해 정신적, 사회적, 신체적 발달이 어떻게 이루어지며 상호 영향을 주고있는가에 대한 이론적 근거에 초점을 맞추어 어떻게 학생들이 실제로 행동하는가 파악하려는 추세이다. 물론 사회성이나 감성, 인지능력, 그리고 언어발달 등에 대한 발달이론은 다양하게 전개되고 있지만 특히 관찰방법을 통해 관측 가능한 아동행동은 어떤 것이 있는지 중점적으로 이론적 구성틀을 개발하고자 한 것이다.

아동의 행동발달에 있어 관찰 가능한 영역에 대한 이론적 기본틀은 네가지 행동영역 즉 개인-사회적, 인지적, 신체적 그리고 안전성으로 구성되었다. 아동발달이론에 근거하고 충분한 논의를 거쳐 <표 1>에 정리한 바와 같이 55개의 하부구인을 규정하였다. 아동 연령에 따라 구분할 수 있는 수준의 발달단계를 설정하고 하부구인간의 관계에 따라 각 영역으로 분류한 결과 최종 11개의 구인을 추출하게 된 것이다(<http://www.cde.ca.gov>).

〈표 1〉 학생발달의 성취 기대 결과와 DRDP 구인 구성

성취기대결과	DRDP 구인	DRDP 하부 구인 (주제)
개인·사회적 영역	자아	자아개념 자아존중
	사회성 기술	어른과의 상호작용 관계 배려와 양보 어른에 대한 감정표현 친구에 대한 감정표현 어른과의 갈등해소/타협 친구와 갈등해소/타협 친구와 협동관계 후배에게 도움주기 우정 친구와 감정이입
	자기 통제력	규칙 충동 조절 만족 유보 집중력
	다양성	다양성 인식
	언어 능력	다양성 수용 이해력 언어 의미 대화 표현능력
인지적 영역	학습 흥미	이해 추구 질문 탐구심
	인지 능력	과제 지속성 기억력 인과관계 파악 문제해결 발명탐구심 검증 진술 비교와 대비능력 초 인지
	읽기쓰기 능력	어휘력 독해력 개인경험 표현 글 쓰기 작문 전략 작문 수정 편집
	수리 능력	수리감각과 산술 소수와 분수, 백분율 대수 측정과 기하
신체적 영역	신체적 기술	자료 정리와 구조화 근육발달과 조화 용구 사용시 근육발달과 조화 용구 사용시 집단 행동
건강과 안전영역	건강 습관과 안전성	기민성 건강식 운동 자기 관리 안전 규칙 안전에 대한 책임 활동 중 안전에 대한 책임 위기 대처 능력

2. 구인 지도 (construct map) 개발

심리 또는 교육연구에서 주로 사용되는 측정도구는 이미 지적인 바와 같이 일차적으로 구인과 하부구인을 설정하고 각 구인에 대응하는 문항을 개발함으로 시작된다. 물론 개발된 문항이 이론적으로 설정한 구인을 잘 반영하고 있는지를 검토하는 과정도 중요한 연구 영역이다. 여기에서 최근 Wilson(2002)과 동료(Kim, Brown, & King, 2004) 등은 구인 지도(construct map)라는 개념을 확대 적용하였다. 이전에 각종 심리 교육 측정도구를 개발할 때 선정된 구인과 개발된 문항들을 제시하였다면 그 과정 중에 구인 지도라는 개념을 적용하여 구체적으로 어떻게 각 구인에서 개별 문항으로 발전하게 되었는지 상세하게 가시화하는 것이다. 물론 이러한 검사개발 과정은 이미 이론적으로 정립되고 적용되어 왔지만, Wilson(2002)은 어떤 구인을 측정하기 위한 도구를 형성하는 과정을 4단계 구성모형으로 다른 각도에서 전개하였다. 그것은 첫째, 구인의 개념을 정의하는 구인모형(construct model), 둘째, 문항모형에 따라 문항을 개발하는 문항모형(item model), 셋째, 문항응답의 범주와 척도를 개발하는 점수모형(score model), 그리고 마지막으로 측정모형을 적용하여 자료를 분석하고 다시 피이드백을 제공하는 측정모형(measurement model)이다. 이 과정에서 구인 지도는 각 단계로 전환하는 근거로 사용되는 것이다.

기대결과를 나타내는 11개 각 구인에 대한 문항개발을 위하여 우선 구인 지도를 개발하였다. 각 구인지도는 연령에 따라 각 구인을 나타내는 아동발달의 구체적 행동양식을 적절하게 표현함으로 어떻게 관찰도구가 개발되고 있는지 요약하고 있다. 이미 구분된 11개 구인과 하부구인을 설정한 후에 아동발달 전문가와 문헌 검토를 통해 설정한 구인과 하부구인이 무엇을 의미하는가 각 개념을 정의하였다. 도구개발의 마지막 단계에 이르기까지 정상적인 아동의 발달과정에 따라 각 구인이 어떻게 발달해 가는지 지속적으로 수정보완하였다. 또한 각 구인과 하부구인을 관찰함에 있어 연령에 따라 차이가 생길 수 있으므로 그에 준한 채점 기준표를 개발하였다. 발달 단계에 따라 어떻게 행동이 변하는가, 관찰 가능한 행동으로 표현될 수 있는가, 나아가 전 단계와 어떻게 다르게 표현되고 관찰되는가 등을 구체화한 것이다.

이 과정을 통해 연령별 아동들의 발달상황을 나타내는 구체적 지표를 제시하고 어떻게 실제 행동으로 표현될 수 있는지 확인할 수 있다. 구인 지도는 관찰도구의 세부문항을 개발할 때 활용되며 관찰자가 사용할 채점표를 구성하는 기준이 된다. 각 구인과 하부구인, 그리고 각 연령의 발달단계를 정의함은 물론 아동행동을 관찰할 수 있는 단적인 기술어와 예문을 함께 제시하였다. 특히 특징을 강조한 한 두 가지 단어를 사용하여 각 발달단계를 구분하였다. 11개 구인 중 ‘자아’에 대한 구인 지도의 예를 <표 2>에 정리하였다.

〈표 2〉 구인 '자아 (self)'에 대한 구인지도

자아 SELF			
	하위구인	자아 개념(자기 묘사) Self-Concept	자기 존중(자기 평가) Self-Esteem
발달 단계	단계 설명	자신을 인지하는 정도 의미, 스스로에 대하여 묘사하는 정도를 관측함	자신에 대하여 가치기준에 근 거하여 평가하는 정도 사람이나 환경과의 상호작용 의 결과로 판단함
통합적 자아 (13세 이상)	자아를 완전히 심리적 구인으로 인식하고 자기의 역할을 인지한다. 자기 주장이 강해지고 자신을 과거 현재 미래의 연속선에서 파악한다. 자신에 대한 믿음이 더욱 실재화 된다.	자신의 상태를 과거와 미래 안에서 묘사한다. (예 : 나는 어렸을 때부터 항상 동물을 좋아했다. 나는 수의사가 되려고 한다)	자신에 대한 장점과 단점을 과거와 현재 그리고 미래에 비추어 평가한다. (예 : 나는 훌륭한 수의사가 될 것이다. 왜냐하면 나는 항상 동물들에 대하여 신경써왔고 어떻게 해야 하는지 알고 있기 때문이다)
인정적 자아 (11-14세 정도)	자아를 심리적 구성으로 인정하는 단계이며 자신의 내면적 감정, 사고, 믿음 등을 표현하고 '내가 무엇을 할 수 있는가' 보다 '나는 누구인가'에 대한 질문이 시작된다. 언어사용도 추상적이고 일반화된 표현을 많이 사용한다 (예 : '나는 내가 됐다고 여길때까지 공차기 연습을 한 나' 라는 표현대신 '나는 완벽주의자이다'라고 말한다)	자신에 대하여 자신의 감정, 생각, 흥미 등과 관련된 내면적 상태로 묘사한다. (예 : 나는 모든 일에 걱정이 많은 종류의 사람이다. 나는 동물의 권리를 믿는다.)	자신의 장단점에 대하여 일반화하여 진술한다. (예 : 나는 좋은 학생이다. 비록 운동은 잘 못하지만)
소속적 자아 (9-11세 정도)	자신을 거의 다른사람 특히 친구들과의 관계로 표현한다. 자아개념은 신체적으로부터 심리적으로 전환하고 정신세계와 동기 등을 중시여긴다. 특히 어디에 소속되어있는 것이 강한 동기로 작용된다.	자신에 대하여 친구 집단 안에서 차지하는 개인적 역할로 묘사한다. (예 : 나는 우리 축구팀 수비선수로 어떻게든 상대방이 골인하지 않도록 한다. 나는 컴퓨터 동아리이며 내 역할은 컴퓨터가 잘 작동하는지 지키는 것이다)	자신과 남과의 상호관계 안에서 자신을 판단한다. (예 : 컴퓨터 동아리 친구들은 내가 컴퓨터를 잘하고 항상 도움을 주는 편이므로 나를 좋아한다) 자신이 할 수 있는 자부심을 근거로 판단한다. (예 : 나는 무엇이 고장이든 친구 컴퓨터는 고칠 수 있다)

비교 자아 (7-9세 정도)	타인과의 비교로 자신을 설명한다. 자기존중에 대한 심리적 인식이 시작되고 인지적, 사회적, 신체적 그리고 일반적 자존을 의식하게 된다.	다른 사람과 비교함으로써 자신을 묘사한다. (예 : 나는 제인보다 공을 더 멀리 던진다, 나는 민이보다 불평하진 않는다)	다른 사람과 비교하여 자신을 평가한다. (예 : 나는 우리 야구팀에서 최고선수이다. 특히 유진보다는 훨씬 낫다) 자신을 인지적 사회적 그리고 신체적 특징으로 진술한다. (예 : 나는 학교 성적은 좋지만 낯선사람들 앞에서 무척 수줍어한다)
효과적 자아 (5-7세 정도)	개인 존재에 대한 인식이 시작된다. 자아에 대하여 기본적인 수행하는 행위나 신체적 특징에 대하여 표현한다.	자신을 수행하는 행위로 묘사한다. (예 : 나는 헛수영 하지 않고 물을 때릴 수 있다) 자신을 문화적, 지역적, 집단적 원인으로 표현한다. (예 : 나는 지금 1학년 아이예요. 나는 카톨릭 신자이다)	자기에 대하여 어떤 행위에 대하여 구체적으로 판단한다. (예 : 나는 더하기와 빼기를 잘한다, 나는 추구할 때 공을 잘 차는 편은 아니다.) 자신의 성취에 대하여 자랑한다. (예 : 와, 내가 이 나무에 오르는 거 봤어?)

이러한 구인지도는 문항제작의 길잡이로 사용된다. 각 발달단계에 대한 기술과 예문과 함께 하부구인을 위한 발달단계는 아동들이 특정 단계에서 나타나는 발달과정을 인목요연하게 제시하였다. 이러한 개요는 문항 구조와 설계를 결정함에 중요한 역할을 한다. 발달단계를 이해하는 것은 적도개발과 직접적으로 연결되며 각 단계에 대한 기술어와 예문은 실제로 관찰자가 점수를 부여하는 과정을 안내하는 것이다. <표 3>은 구인 ‘자아’의 하부구인인 ‘자아개념’과 ‘자기존중’ 중 ‘자아개념’을 어떻게 문항으로 나타내는지 요약하였다. <표 2>와 비교할 때 어떻게 구인지도에서 문항이 개발되는지 파악할 수 있다.

〈표 3〉 관찰 도구 개별문항 '자아개념'의 예

	효과적 자아	비교 자아	소속적 자아	인정적 자아	통합적 자아
	<input type="checkbox"/> 예 <input type="checkbox"/> 아니오	<input type="checkbox"/> 예 <input type="checkbox"/> 아니오	<input type="checkbox"/> 예 <input type="checkbox"/> 아니오	<input type="checkbox"/> 예 <input type="checkbox"/> 아니오	<input type="checkbox"/> 예 <input type="checkbox"/> 아니오
SC1. 자아 개념 (자기 묘사) 관찰기회 없음 <input type="checkbox"/>	자기 자신의 수행한 행위에 대하여 잘 표현한다. (예 : 나는 매우 잘 땔수 있다, 나는 쏘지 않고 밀가루와 기름을 섞을 수 있다)	자신을 항상 남과 비교하여 표현한다. (예 : 나는 진이보다 잘 땔 수 있다. 나와 조는 엄마의 도움 없이도 쿠키를 만들 수 있다)	자신에 대해 다른 사람과의 관계 속에서 또는 집단 내 역할로서 표현한다. (예 : 나는 육상 뒤에 속해있고 허물선수이다. 쿠키를 만들 때 내 역할은 밀가루를 짓는 것이다)	자신의 성격을 나타내거나 자신의 느낌, 믿음, 생각, 관심 등을 중심으로 표현한다. (예 : 나는 나의 달리기 실력을 향상시키기 위해 항상 걷는다, 나는 친구들이 내가 만든 쿠키를 맛있다고 하던 정말 좋다)	자신의 현재 모습을 과거와 미래의 진행 중의 모습으로 표현한다. (예 : 나는 항상 뛰고 이기는 걸 좋아했다, 결국 육상 선수가 되고 싶다. 난 항상 집에서 과자를 만들었고 다른 종류의 과자는 어떻게 만드는가 배우고 싶다)

예를 들어 자아개념의 첫 번째 단계에서 아동들이 실제로 나타내는 행위가 효과적으로 기술했는지 여부를 확인한다. 행위자체가 관찰가능한지 여부를 검토한 후 각 주제에 대한 체점 기준표가 함께 제작되는 것이다. 문항개발이 완성된 후에도 전문가와 패널에 의해 문항 선정, 구성과 연계 등 수정 보완되었다. 이런 과정을 통하여 최종적으로 총 55개의 지표와 각 지표에 대한 척도가 개발되었으며 앞서 제시한 11개 구인의 세부 문항 수와 척도의 범위를 <표 4>에 정리하였다. 11개의 구인은 자아개념(SC), 사회성(SIS), 자아 통제력(SR), 다양성(DI), 언어능력(LA), 학습흥미(IL), 인지능력(CD), 수리능력(MA), 읽기쓰기능력(LIT), 기술과 기능(MS), 그리고 안전과 건강(HH)이다.

〈표 4〉 DRDP 관찰 도구의 문항구조

기대 성취 결과(Desired Results)	구인	문항 수	척도 범위
개인적·사회적 기대결과 : 학생은 개인적으로나 사회적으로 유능하다(DR-1)	1-1 자아개념(SC)	2	0-5
	1-2 사회성(SIS)	10	0-3
	1-3 자기 통제력(SR)	4	0-3
	1-4 다양성(DI)	2	0-4
	1-5 언어능력(LA)	3	0-3
인지 학습 기대결과 : 학생은 효과적인 학습자이다(DR-2)	2-1 학습 흥미(IL)	4	0-6
	2-2 인지 능력(CI)	7	0-6
	2-3 수학 능력(MA)	6	0-4
	2-4 문장력(LIT)	6	0-7
신체 기능 기대결과 : 학생은 신체적으로나 기능적으로 유능하다(DR-3)	3-1 기술과 기능(MS)	4	0-5
건강과 안전 기대결과 : 학생은 안전하고 건강하다(DR-4)	4-1 안전과 건강(HH)	7	0-3

Ⅲ. 연구 방법

1. 표본과 자료

연구에 사용된 자료는 DRDP 도구를 본격적으로 적용하기에 앞서 시행된 예비조사 결과이다. 2003년 봄학기에 미국 캘리포니아 4개 지역에서 추출된 11개 학교를 대상으로 방과 후 프로그램에서 활동하는 학생들을 관찰하였다.

표본 대상은 다양한 교과재량과목과 특별활동으로 구성된 방과후 프로그램에 참여하고 있는 163명 학생이며 연령의 범위는 5세에서 14세 사이이다. 관찰자는 학교선생님, 지역교육청 교육전문가, 도구개발자 등 21명으로 구성되었으며 검사도구에 대한 사전이해와 관찰방법에 대한 관찰자 훈련을 시행하였고 비디오 관찰이 아닌 직접참여관찰 방법을 시행하였다. 단, 분석방법 특성에 따라 사용된 자료의 크기는 프로그램에 따라 차이 있음을 밝히고자 한다.

2. 분석 방법

가. 일반화가능도 이론의 적용

타당화 모형검증을 위하여 두 가지 분석방법을 적용하였다. 첫번째 분석은 관찰결과에 대한 일반화가능도, 즉 관찰점수의 일관성, 정확성, 합치성, 일반화가능성 등을 검증하는 절차이다. 도구의 신뢰도를 파악하기 위한 연구질문은 다음과 같다. 관찰도구를 통해 아동 발달 과정을 일관성있게 기록할 수 있는가? 관찰자들이 동일한 기준으로 기록하는가? 관찰자들이 일관성있게 기록하는가? 각 구인, 개별 문항, 그리고 관찰자들에 의해 학생들의 행동을 일반화할 수 있는가?

신뢰도에 대한 기본정보를 파악하기 위해 측정의 표준오차와 고전검사이론의 내적합치도 Cronbach α 계수를 산출하였고 일반화가능도 이론을 적용하여 오차 분산성분과 일반화가능도 계수를 추정하였다. 일반화가능도 이론(또는 G 이론으로 칭함)은 고전검사이론에서 설명하지 못한 측정오차의 중다 요인에 대한 분석이 가능하고 Cronbach와 그 동료(1972)의 이론 정립 이후 Brennan(1983)과 Shavelson(1991, 1993) 등에 의하여 실제 자료 분석에 활용할 수 있도록 영역을 넓혀왔다. 특히 GENOVA 프로그램(Crick & Brennan, 1983)이 개발된 이후 20여년 동안 지속적으로 연구범위가 다분화되었다. 최근 수행평가의 양호도 문제나 관찰자료의 일반화 문제, 그리고 중다 측정의 측정오차 문제가 부각되면서 G 이론은 오차요인을 분석하고 적정수준의 일반화가능도를 확보하기 위한 사전연구로 활발하게 적용되어 왔다(Baxter, Shavelson, Goldman, & Pinc, 1992; Colton, Gao, & Kolen, 1993; Lanc, Liu, Ankermann, & Stone, 1996; Shavelson et al., 1993; 김성숙, 2001). G 이론에서 관찰된 단일 행위는 관찰행동 전집을 대표하는 표집으로 보고 이러한 표집행위가 얼마나 일반화되어질 수 있는가 하는 일반화가능도를 추정하고자 한다. 특히 일반화연구(G study)는 각 피험자 점수 또는 관찰점수를 구성하는 허용가능한 관찰의 전집과 관계된 조건의 집합들의 분산성분 추정값을 산출하는 과정이다. 오차분산성분의 상대적 크기를 비교하여 각 오차원의 영향력을 분석함으로써 어떤 요인이 측정의 일반화과정을 저해하는지 설명할 수 있고, 결정연구(D study) 설계의 조건을 결정하는 근거로 활용한다. 결정연구는 이미 정의된 전집을 근거로 결정연구에서 각 오차의 상대적 영향력에 따라 일정 수준의 신뢰도를 보장하기 위한 측정조건을 제시한다. 이와 같이 일반화가능도 이론은 단순히 고전검사이론의 신뢰도 계수의 개선안이라기보다 전집에 대한 관찰점수의 일반화가능성을 이해하고자 전개된 측정이론의 한 영역이다(김성숙 · 김양분, 2001).

이 연구에서 일반화가능도 이론을 적용하기 위해 사용된 사전자료는 모든 구인의 모든 문

항을 사용한 3명의 관찰자에 의한 30명 학생 결과이다. 일반화연구 설계에서 측정대상은 물론 학생이며 오차요인은 관찰자와 문항이고, 각 학생들의 행동을 모든 문항에 대하여 세 명의 관찰자가 기록하였으므로 각 국면들이 서로 완전교차된 경우이다. 따라서 일반화 연구 설계는 $p \times r \times i$ 완전교차 모형이며 G 연구를 통해 각 오차요인이 상대적으로 어떤 영향을 주고 있는지 파악하고 D 연구를 통해 최적의 일반화가능도 계수를 산출하기 위한 각 국면의 조건(수)을 제시하였다. 측정의 일반화가능도 분석을 위해 GENOVA 프로그램(Crick & Brennan, 1983)을 사용하였다.

나. 확인적 요인분석의 적용

두번째 분석은 관찰도구가 복직에 충실하고 합당한 정도를 확인하는 측정의 타당도를 검증하였다. 특히 도구를 제작하는 과정에서 가장 중요하게 다루는 ‘구인’이 어느정도 이론적 구조에 부합되는가를 점검하고자 하였다. 이 연구에서 아동발달을 나타내는 네가지 영역에 대하여 이론적으로 추출한 11개 구인이 어떻게 설명하고 있는지에 대한 구인 타당화 모형을 설정하였다. 그리고 이론적 모형을 검증하기 위하여 구조방정식 모형의 한 부분으로 활용되는 확인적 요인분석(Confirmatory Factor Analysis)을 적용하여 구인간 관계를 가정적으로 설정한 각 모형이 어느정도 적절한가 설명하였다.

특히 탐색적 요인분석에서 고유분산을 사용하는 반면 확인적 요인분석은 공통요인을 나타내는 부분과 고유요인을 나타내는 부분을 분리하여 명시하고 있다. 또한 상관행렬을 사용하는 탐색적 요인분석방법과는 달리 확인적 요인분석은 공분산행렬을 주로 사용하기 때문에 비표준화된 측정변수를 사용하고 있다. 구조방정식 모형의 일반적 모양중 관찰변수를 사용하여 이론변수를 정의하는 즉, 이론변수가 어떤 관찰변수에 의해 구체적으로 정의 되는가를 나타내는 측정모형은 수학적으로 공통요인을 사용하고 관찰변수로 가설화된 이론변수를 검증하므로 확인적 요인분석으로 함께 부르는 것이다 (Kaplan, 2000; Kline, 1998; 이순목, 2000). 이미 여러 연구에서 소개되었던 (Steiger, 1990; Loehlin, 1987; 이기중, 2000) 구조방정식 모형은 요인분석과 경로분석의 강점을 조합하여 요인의 구조를 파악할 수 있는 포괄적이며 강력한 통계 방법이다. 구조방정식모형은 Joreskog(1973) 이후 관찰변수와 연속적 잠재변수 또는 이론변수들 간의 관계를 결정하는 분석도구로 광범위한 영역에서 사용되었다. 이 연구에서 관찰변수는 도구에서 사용된 각 문항점수이며 이론변수는 연구대상인 이론적 구인을 나타내고 있다. 모형의 모수 추정은 최대 우도 추정방법을 사용하였다. 이미 확인적 요인 분석과 구조방정식 모형을 적용하기 위한 다양한 분석 툴이 비교되어있고(Long, 1983; 서민희·성태제, 2001; 김재철, 2002; 김아영·김미진, 2004) 분석을 위해 Mplus 프로그램(Muthen & Muthen, 2001)을 활용하였다.

IV. 연구 결과

1. 일반화가능도 분석 결과

DRDP 도구의 신뢰도를 추정하기 위해 산출된 각 계수를 정리하면 <표 5>와 같다. 각 구인에 대한 측정의 표준오차와 내적 일치도를 나타내는 Cronbach α 계수는 ‘다양성’ 구인(.41)을 제외하고 적절한 수준이며(.62~.79) 일반화가능도 계수도 표본 수가 적은 경우 계산되지 않은 경우를 제외하고(수리능력), 수용할 만한 수준의 계수로 산출되었다(.60~.78).

<표 5> DRDP 구인의 표준오차, 신뢰도와 일반화가능도 기대 성취결과

기대 성취결과	구인	표준오차	Coefficient Alpha	G coefficient
개인 사회적 영역 (DR-1)	자아개념 (SC)	1.40	.78	.71
	사회성 (SIS)	3.18	.79	.75
	자기 통제력 (SR)	1.95	.76	.60
	다양성 (DIV)	2.17	.41	.74
	언어능력 (LA)	1.50	.75	.65
인지 학습 영역 (DR-2)	학습 흥미 (IL)	4.34	.62	.73
	인지 능력 (CI)	5.21	.79	.65
	수리 능력 (MA)	.76	*	
	읽기 쓰기 능력 (LIT)	1.01	*	.74
신체적 영역 (DR-3)	기술과 기능 (MS)	3.74	.63	.60
건강과 안전 영역 (DR-4)	안전과 건강 (HID)	.54	*	.62

* 절측값이 10% 이상인 경우 검사 신뢰도 값을 추정하지 않았다.

G 연구와 D 연구 결과에서 각 오차국면의 분산성분 뿐 아니라 각 오차간의 상호작용의 상대적 크기도 파악할 수 있다. 학생-관찰자 상호작용은 관찰자에 따라 학생관찰에 영향주는 정도를 의미하며 학생-문항 상호작용은 문항에 따라 학생이 다르게 행동하는 정도를 나타내고 관찰자-문항 상호작용은 문항에 따라 관찰자의 평정이 영향받는 정도를 파악할 수 있다. 지면 관계상 구인 ‘자아개념(SC)’과 ‘사회성(SIS)’에 대한 G 연구와 D 연구 결과를 <표 6>,<표

7>, <표 8>, 그리고 <표 9>에 정리하였다. G 연구 결과에서 나타나듯이 ‘자아개념’ 구인의 경우 관찰자-문항상호작용이 상대적으로 크게 산출되고 있다(11.6%). 반면 ‘사회성’ 구인의 학생-문항 상호작용은 상대적으로 크게 나왔지만(12.3%), 학생-관찰자 또는 관찰자-문항 상호작용의 상대적 크기는 의미있게 나타나지 않았다(각각 2.4%, 3.9%). 즉, ‘사회성’의 경우 학생을 관찰하는 과정에서 문항에 따라 영향을 주는 정도가 관찰자가 달라짐에 따라 영향을 주는 정도보다 상대적으로 크게 나타난다고 해석된다. 그러나 결과표가 제시되진 않았지만 구인 ‘다양성’을 보면 학생-관찰자 상호작용이 상대적으로 크게 나타났으며 관찰자에 따라 학생에 대한 평정에 차이가 보임을 알 수 있다. 즉 관찰자 효과를 분석함으로써 평정체계를 개선시켜야 한다는 이슈를 제기하고 있다. 뿐만 아니라 구인 ‘다양성’에 대한 신뢰도 개선과 측정결과와 오차에 대하여 재고의 필요성을 요구하고 있다.

한편, D 연구에서 일반화가능도 계수 .70 수준에 도달하기 위한 각 국면의 수를 조합한 결과, 구인 ‘자아개념’의 경우 관찰자 3명과 12개 문항, 관찰자 4명과 10개 문항 그리고 구인 ‘사회성’에 의 경우 관찰자 3명과 20개 문항, 또는 관찰자 2명과 40개 문항 등의 조합을 제시하고 있다 (<표 4>와 <표 6> 참조). 문항에서 발생하는 오차가 관찰자에 의한 오차보다 상대적으로 크게 추정되었으므로 문항의 수를 늘리는 것이 관찰자 수를 늘리는 것보다 효율적으로 신뢰도계수를 개선시킬 수 있다.

<표 6> ‘자아개념’ G 연구 결과 ($n_p=10$, $n_r=3$, $n_i=12$)효과

효과	자유도	EMS 분산성분	표준오차	분산성분 비율(%)
P	9	.0163300	.0099263	6.2
R	2	.0125140	.0118958	4.7
I	11	.1204265	.0526145	45.5
PR	18	.0117284	.0053603	4.4
PI	99	.0111392	.0049144	4.2
RI	22	.0306958	.0106610	11.6
PRI	198	.0617284	.0061728	23.3

〈표 7〉 ‘자아개념’ D 연구 결과

표본크기				분산성분						
SC	10	관찰자	문항	진 집	기대관찰	상대오차	절대오차	Mean	일반화	Phi 계수
				분산성분	분산성분	분산성분	분산성분		가능도 계수	
SC 1	10	1	5	0.01633	0.04263	0.02630	0.06904	0.04700	0.38305	0.19128
SC 2	10	1	10	0.01633	0.03535	0.01902	0.04664	0.03116	0.46201	0.25932
SC 3	10	1	12	0.01633	0.03413	0.01780	0.04291	0.02852	0.47845	0.27567
SC 4	10	2	15	0.01633	0.02499	0.00866	0.02397	0.01781	0.65335	0.40518
SC 5	10	2	5	0.01633	0.03059	0.01426	0.04768	0.03647	0.53375	0.25513
SC 6	10	2	10	0.01633	0.02639	0.01006	0.02990	0.02247	0.61869	0.35324
SC 7	10	3	12	0.01633	0.02288	0.00655	0.02161	0.01735	0.71365	0.43039
SC 8	10	3	15	0.01633	0.02235	0.00602	0.01891	0.01512	0.73052	0.46345
SC 9	10	3	5	0.01633	0.02658	0.01025	0.04056	0.03296	0.61431	0.28707
SC 10	10	4	10	0.01633	0.02192	0.00559	0.02153	0.01813	0.74501	0.43135
SC 11	10	4	12	0.01633	0.02148	0.00515	0.01895	0.01595	0.76037	0.46287
SC 12	10	4	15	0.01633	0.02103	0.00470	0.01637	0.01377	0.77638	0.49936

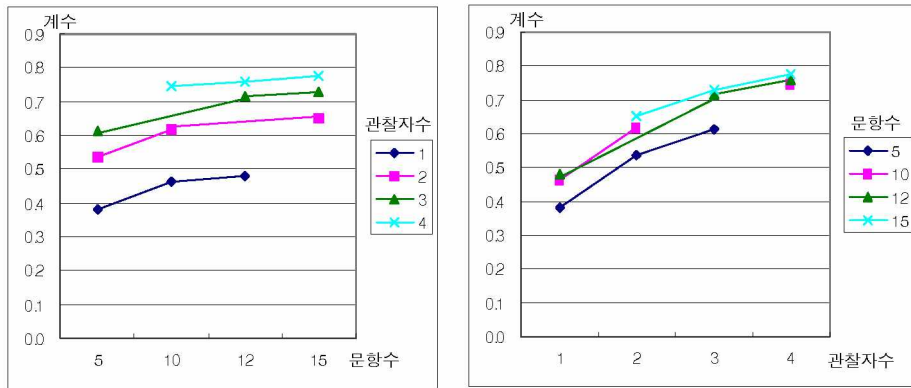
〈표 8〉 ‘사회성’ G 연구 결과 ($n_p=10$, $n_r=3$, $n_i=40$) 효과

효과	자유도	EMS 분산성분	표준오차	분산성분비율(%)
P	9	.0148623	.0078699	5.8
R	2	.0010518	.0015129	.4
I	39	.1075261	.0257949	42.4
PR	18	.0060209	.0025637	2.4
PI	351	.0313414	.0046843	12.3
RI	78	.0097816	.0028954	3.9
PRI	702	.0831458	.0044317	32.8

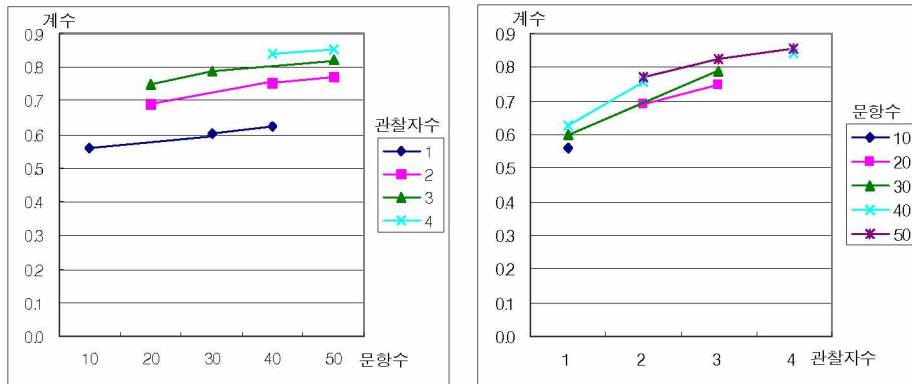
〈표 9〉 ‘사회성’ D 연구 결과

표본크기					분산성분						
학생	관찰자	문항			전집 분산성분	기대관찰 점수분산	상대오차 분산성분	절대오차 분산성분	Mean	일반화가 능도계수	Phi 계수
SIS 1	10	1	10		0.01486	0.02661	0.01175	0.01866	0.00958	0.55857	0.44332
SIS 2	10	1	30		0.01486	0.02470	0.00984	0.01480	0.00743	0.60173	0.50106
SIS 3	10	1	40		0.01486	0.02375	0.00888	0.01287	0.00636	0.62590	0.53597
SIS 4	10	2	50		0.01486	0.01933	0.00447	0.00724	0.00471	0.76883	0.67234
SIS 5	10	2	20		0.01486	0.02152	0.00666	0.01280	0.00830	0.69068	0.53722
SIS 6	10	2	40		0.01486	0.01970	0.00483	0.00817	0.00531	0.75460	0.64529
SIS 7	10	3	50		0.01486	0.01805	0.00319	0.00575	0.00437	0.82338	0.72089
SIS 8	10	3	20		0.01486	0.01982	0.00496	0.01085	0.00787	0.74978	0.57803
SIS 9	10	3	30		0.01486	0.01884	0.00398	0.00802	0.00593	0.78896	0.64954
SIS 10	10	4	40		0.01486	0.01767	0.00281	0.00582	0.00478	0.84107	0.71858
SIS 11	10	4	50		0.01486	0.01741	0.00255	0.00501	0.00420	0.85366	0.74788

문항수와 관찰자 수 변화에 따른 일반화가능도 계수의 변화를 그래프로 나타내면 [그림 1]과 [그림 2]와 같다.



[그림 1] ‘자아개념’ 문항 수와 관찰자 수 변화에 따른 일반화가능도 계수 변화



[그림 2] '사회성' 문항 수와 관찰자 수 변화에 따른 일반화가능도 계수 변화

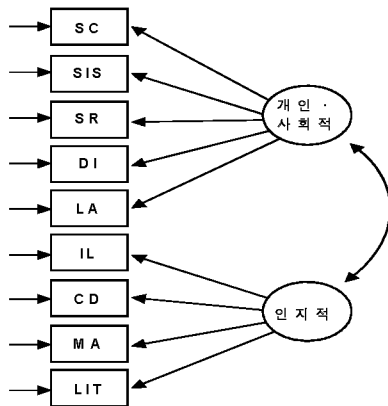
2. 구인 타당도 분석 결과

확인적 요인분석을 적용한 분석결과에서 관찰도구가 측정하려고 하는 '아동발달'의 이론적 구조와 각 문항간의 관계를 볼 수 있다. 특히 확인적 요인분석은 실제 관찰된 측정변수와 설명하려고 하는 이론변수간의 관계를 나타내고자 설정된 모형이 어느정도 타당한가를 검증하고 있다. 물론 이 연구에서 측정변수는 DRDP 관찰도구의 문항점수이며 이론변수는 이론적 배경에 근거한 아동발달에 대한 기대결과 구인을 의미한다. 설정된 모형이 얼마나 관계를 잘 설명하고 있는지 구조모형의 적합성 정도를 산출하였다. 이러한 적합도 검증은 관찰변수간의 독립성을 구체화하는 기본 모형과 비교할 수 있다.

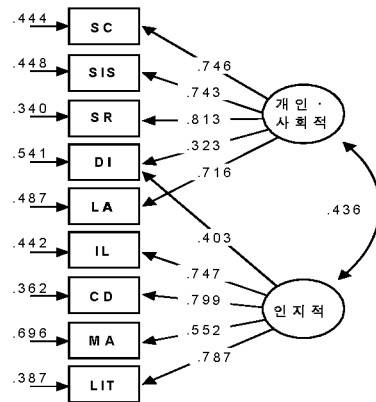
이론적으로 설정된 모형이 실제 관측에 의해 어느 정도 부합되는가에 대한 합치수준에 대하여 Hu와 Bentler(1995)는 비교 합치도(CFI), 표준합치도(NFI), 그리고 터커-루이스 지수(TLI) 등 3종류 지표를 비교하였다. 이 지표는 물론 모형이 어느정도 구인의 구조를 잘 설명하고 있는가를 나타내며 0에서 1의 구간으로 산출된다. 일반적으로 .95 이상이면 적합도가 매우 높다고 볼수 있으며 .90을 넘으면 수용할 만한 모형으로 지지한다고 해석한다. 또한 모형을 평가하는 다른 준거는 근접 오차에 대한 정보이다. 개략화 오차평균(RMSEA)은 얼마나 모형에 근접한가를 나타내는 값이며, Steiger(1990)과 Mulaik과 동료들의 연구에서(1989) RMSEA 적합도 수준을 실질적으로 .05에서 .08 수준이면 좋은 적합도이고 .08에서 .10 수준이면 수용할 만한 오차라고 제안하였다.

물론 확인적 요인분석의 결과물은 요인계수, 요인간 상관, 그리고 고유요인의 분산과 공분산에 대한 정보를 모두 포함하고 있다. 모형검증 과정은 초기 즉 가설모형으로 시작하여 전체적합도 수준과 R², 각 모수값의 유의도 등을 고려하여 모형을 수정해 가는 것이다. 이 예비조사 결과의 경우 자료의 크기가 만족할 만큼 크기 않았지만(163 사례수) 이론적 모형을

검증하기 위해 8가지 모형을 시도하였고 4가지 모형에 대하여 요약하였다. 구인간 관계를 설명하는 첫 번째 모형은 발달 이론에 의해 관찰도구를 구성하고 있는 기본모형이다. 즉 개인적-사회적 영역(DR1)은 자아개념, 사회성, 자기 통제, 다양성 그리고 언어능력 등 다섯가지 구인으로 설명하고 있다. 그리고 인지 영역(DR2)은 학습흥미, 인지능력, 수리력과 읽기 쓰기 능력 등 4개의 구인으로 설명한다고 설정하였다(〔그림 3〕 참조). 첫 번째 모형에 대한 분석 결과 ‘다양성’ 구인은 개인적-사회적 영역과 인지 영역 두 영역 모두와 관련지어 나타나는 경우 모형 적합도 지수가 향상되므로 모형에 대한 개선안을 제시하고 있다. 따라서 두 번째 개선된 모형은 ‘다양성’ 구인 모두 두 영역과의 관계를 경로로 제시하였다.(〔그림 4〕 참조). 모형 1의 경우 RMSEA는 .07인 반면 모형 2의 RMSEA는 .051로 전체 모형의 적합도 수준이 향상된 것을 알 수 있다. 각 모형의 적합도 지수는 <표 10>에 요약하였다.



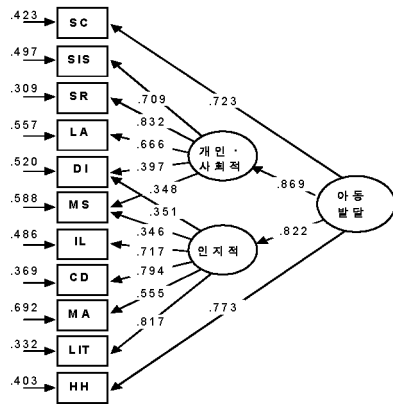
〔그림 3〕 개발 모형 1



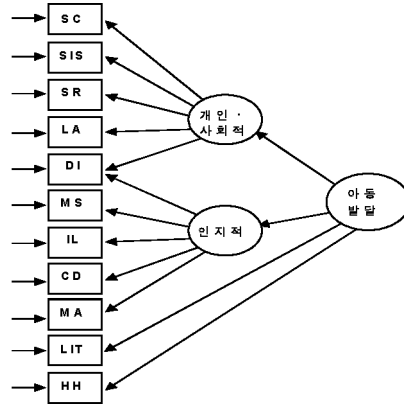
〔그림 4〕 개발 모형 2

다음 단계 모형 3은 신체적 영역을 나타내는 ‘기술과 기능’ 구인이나 안전성 영역을 나타내는 ‘건강과 안전’ 구인 모두를 포함하여 제시하였다. 이론변수를 설명하는 특정변수는 적어도 3개 이상은 넘어야 하므로 두 가지 구인을 포함하여 2단계 확인적 요인 분석을 시행하였다. 즉, 모형 3은 모형 2에 나타난 경로에서 한단계 더 발전하여 ‘기술과 기능’과 ‘건강과 안전’ 구인을 포함하여 관찰도구의 대상인 ‘아동발달’ 이론 변수를 설명하였다 (〔그림 5〕 참조). 모형 검증 결과 나타난 적합도 지수와 요인모수의 유의수준에 따라 개선된 모형 4는 ‘기술과 기능’ 구인은 두가지 영역 ‘개인적-사회적 영역’과 ‘인지적 영역’을 설명하고 ‘자아개념’과 ‘건강과 안전’ 구인은 직접 ‘아동발달’ 변수로 경로를 그린 모형이다. 이 경우 적합도 지수 RMSEA 값은 .111에서 .095 수준으로 개선되었다. 물론 모형 3과 모형 4 모두 ‘개인-사회적 영역’과 ‘인지적 영역’이 전체 아동발달 변수를 설명하는 것은 포함하고 있다 (〔그림 6〕 참조). 지면 관계상 개선된 모형 2와 모형 4의 경우만 각 구인과 영역간의 경로에 대한

추정모수값을 그림에 함께 제시하였다.



[그림 5] 개발 모형 3



[그림 6] 개발 모형 4

앞에서 언급한 바와 같이 이론변수 즉 개인·사회적 영역이나 인지적 영역 또는 그 다음 단계인 ‘아동발달’ 변수를 11개 구인들이 어떻게 구조화하여 설명하고 있는가에 대한 이론적 모형에 대한 검증은 모형의 전체 적합도 지수로 요약할 수 있다(<표 10> 참조). MPlus 수행 결과에 따라 산출된 적합도 지수는 카이제곱값, 비교합치도(CFI), 터커-루이스 지수(TLI), 알카이크 정보기준(AIC), 개략화 오차평균(RMSEA), 표준화 잔차평균(SRMR) 등이다. 4가지 모형에 대한 모형 적합도를 해석하면, CFI 값(.903~.983)과 TLI 값(.873~.976)은 모두 만족하나 잔차를 기준으로 산출한 RMSEA 값과 SRMR 값에 의하면 모형 3은 적합하지 않다고 볼 수 있다. 따라서 모형적합도 지수로만 준한다면 4가지 모형 중에서는 모형 2에 의한 구인 배열이 이론변수를 가장 잘 설명하고 있다고 볼 수 있다.

<표 10> 개발모형의 모형 적합도 지수 요약

	chi-square model	chi-square baseline	비교합치도 CFI	터커루이스 TLI	알카이크지수 AIC	개략화 오차평균 RMSEA	표준화 잔차평균 SRMR
모형 1	46.577 (p=.0079)	661.961 (p=.00)	.967	.954	8909.576	.070	.051
모형 2	35.546 (p=.0788)	661.961 (p=.00)	.983	.976	8900.544	.051	.040
모형 3	126.506 (p=.00)	923.892 (p=.00)	.903	.873	11268.010	.111	.120
모형 4	114.071 (p=.00)	923.892 (p=.00)	.916	.887	11257.570	.105	.072

V. 결론과 제언

연구의 수행 목적은 아동발달 범주에서 이론적으로 구상된 영역을 측정하기 위하여 개발된 관찰도구의 양호성과 모형의 적합성을 점검하는 것이었다. 이 관찰 프로파일은 그동안 아동 또는 학생발달 연구의 관심이 주로 표준화검사를 통한 학업성취 능력 발달에 치중한 것에 반하여 개인-사회적 발달, 인지적 발달, 신체적 발달, 그리고 안전 영역에 이르기까지 총체적으로 파악하고자 개발된 도구이다. 특히 이 연구의 초점은 관찰 도구 개발에서 제기되어 온 신뢰도 문제, 특히 관찰자간 일치도나 결과의 일반화 수준 등에 대한 점검과 함께 이론적 구인을 근거로 개발된 기대결과요인을 각 구인이 얼마나 잘 설명하고 있는지 구인타당화 모형을 검증하였다. 일반화가능도 이론과 확인적 요인분석 방법(구조방정식 모형 중 공통요인분석의 일환)을 적용한 결과를 근거로 예비조사의 도구를 어떻게 수정할 수 있는가 측정학적 청사진을 제시하고자 하였다.

각 구인에 대한 일반화가능도 이론의 적용 결과를 보면 각 구인의 일반화가능도 계수는 만족할 수준이었고 구인에 따라 관측에 있어 문항과 관찰자 각각 영향에 있어 상대적 차이를 보이고 있다. 이것은 관찰하고자 하는 구인의 특성에 의한 것도 있지만 각 구인을 구성하는 문항수가 다르므로 영향을 주는 것으로 해석할 수도 있다. 결과에서 보여주듯이 신뢰도 계수가 낮게 나타내고 있는 ‘다양성’ 구인에 대하여 개별 문항을 수정하여야 한다고 제안하고 있다. Lane과 동료들(1996) 연구에서 오차국면이 여러개의 경우인 관찰자료에 대한 일반화가능도 분석 결과를 근거로 볼 때 높은 수준의 일반화가능도는 연구자가 어떻게 구인을 이해하고 있는가에 따라 구인 중심의 신뢰도 분석이 수반되어야 한다고 설명하였다. 한편 D 연구 결과 관찰도구의 적정 수준의 일반화가능도 계수를 보장하기 위한 각 구인의 문항 수와 관찰자 수를 제한함으로써 본 조사를 실시할 때 유용가능한 조사수행 틀을 제공하였다. 물론 연구결과에서 관찰자 수를 증가하거나 문항 수를 늘릴때 전반적으로 어떻게 일반화가능도 계수가 향상되는가를 그림으로 나타내어 그 경향을 쉽게 파악할 수 있었지만 더욱 중요한 것은 관찰자 수나 문항 수에 대한 의사결정은 반드시 비용이나 시간과 관련된 효율성이 전제되어야 한다.

구인타당화 모형을 점검하기 위한 확인적 요인분석을 적용한 결과에서 개인적-사회적 영역은 자아개념, 사회성, 자기통제, 언어능력, 그리고 다양성으로 설명되고 인지적 영역은 학습흥미, 인지능력, 수리력, 그리고 읽기쓰기 능력과 다양성으로 설명되었다. 특히 구인 ‘다양성’은 인지적인 면과 개인사회적인 영역 모두 타당성있게 포함되고 있음을 구인 타당화 모

형에서 검증한 것이다. 이러한 모형은 어떤 구인의 개인적 차원을 확실하게 설명하는 배경이 되며 일반적인 구인보다 더 정확하게 이론변수를 예측하는 것이다. 물론 측정변수의 수가 충분하지 않아 ‘기술-기능’ 영역과 ‘안전-건강’ 영역은 모형에 포함시키지 않았지만 2단계 구조모형에서 추론할 수 있는 ‘기술-기능’ 구인은 ‘다양성’ 구인과 같이 어떤 특정 영역에 국한되는 것 보다 아동발달 영역에서 서로 복합적으로 반응하고 있다는 것을 간접적으로 제시하고 있다. 그리고 ‘자아개념’ 구인이나 ‘안전과 건강’ 구인은 직접 아동발달 개념을 설명하는 것이 다른 구인과 더불어 좋은 모형을 구현할 수도 있겠다. 어떤 측정변수의 구성이 이론변수를 잘 설명하는가 다양한 모형을 제시하고 그 적합성 지수를 산출함으로써 이론 모형과 실제 모수가 어떻게 개선되는지 가시화하였다. 특히 이 연구에서는 구인 지도를 사용하여 어떻게 각 구인으로부터 발달단계에 따라 개별 문항을 개발하는지 상세한 안내서를 전개하듯이 제시함으로써 도구개발 절차를 한단계 발전시켰다.

이 연구의 제한점과 후속 연구를 위한 제언을 다음 몇가지로 요약할 수 있다. 첫째, 연구에서 사용된 표본의 수가 충분하지 않다는 점이다. 특히 일반화가능도 이론을 적용하기 위해 필요한 설계의 조건, 즉 2명 이상의 관찰자(동일하지 않다 하더라도)가 동일한 학생을 관찰한 표본을 추출하기 위해 30명의 학생만을 사용하게 되었다. 그러나 예비조사의 경우 구인의 일반화가능도와 관찰상황의 각 조건 또는 수를 제안하는 것이 주요목적이므로 어떻게 분석하여 결과를 제시하는가 과정을 나열하는데 그 의의를 두고자 하였다.

둘째, DRDP 관찰 프로파일은 아동발달 단계에 대한 충분한 이론검토를 배경으로 개발되었음에도 불구하고 네가지 영역 11개 구인, 그리고 55개 하부구인의 수가 골고루 배열되어 있지 않았다. 이런 관찰도구 구성상의 문제로 구인타당화 검증과정에서 ‘개인-사회적’ 영역과 ‘인지적’ 영역 두 가지 영역으로 모형을 선정하였고 ‘기술-기능’ 영역과 ‘안전-건강’ 영역은 제외하고 하부 구인만 포함하였다. 그러므로 구인타당화 과정을 정당화하기 위하여 ‘아동발달’이라는 개념을 포함하여 2단계 구조 분석을 시행하게 된 것이다.

셋째, 예비조사 결과를 반영하여 본조사를 시행하던 그 결과는 단지 타당화 검증이나 신뢰도 파악 뿐 아니라 자료 성격에 따라 다층적 자료 분석을 위한 자료수집 방법과 분석을 적용할 것을 제안하였다. 각 지역과 학교의 특성을 살려 다층분석을 시행한다면 아동들에게 기대되는 성취결과에 대한 파악 뿐 아니라 각종 배경변인과의 관계를 다각적으로 연구할 수 있을 것이다(강상진, 1995).

결론적으로 이러한 관찰도구의 구인타당화 과정은 아동발달 단계를 파악하기 위한 관찰 프로파일을 개발하는 과정에서 이론에 근거한 구인 선정과 병행한 문항과 척도 개발의 질을 보장하는데 그 의의가 있다 하겠다. 즉 연구 결과는 아동발달 이론을 구축하는 관찰도구의 양호도 연구의 필요성을 지지함은 물론 다양한 평가 현장에서 실질적 함의를 뒷받침하는 것이다. 뿐만 아니라 측정의 타당화는 각종 조사활동 단계의 성격에 따라 적합한 타당화 논리

를 적용해야 한다는 배호순(1999)의 논점에 동의하며 도구나 과제의 개발 과정에서는 내용 타당화, 과정 타당화, 또는 구인 타당화를 적용하고 조사활동의 전개와 실시단계에서는 준거 타당화와 외적 타당화 논리를 적용하고, 결과를 해석하고 수정하는 단계에서는 준거 타당화와 외적 타당화 논리를 적용할 수 있음을 제안하고자 한다.

참 고 문 헌

<http://www.ed.gov>

<http://www.cde.ca.gov>

- 강상진(1995). 다층통계모형의 방법론적특성 및 활용방법. *교육평가연구*, 8(2), 63-94.
- 김성숙(2001). 채점의 변동요인 분석방법에 대한 고찰: 일반화가능도 이론과 다국면 라쉬모형의 적용과 제해석. *교육평가연구*, 14(1), 303-325.
- 김성숙·김양분(2001). *일반화가능도 이론*. 서울: 교육과학사.
- 김아영·김미진(2004). 교사효능감 척도 타당화. *교육심리연구*, 18(1), 37-58.
- 김재철(2002). *학생 배경변인과 수학에 대한 태도변화의 관계분석: 잠재변인 변화모형의 적용*. 박사학위논문, 서울대학교.
- 배호순(1997). 포트폴리오 평가의 질관리 방안의 모색. *교육평가연구*, 10(1), 75-104.
- 서민희·성태제(2001). 한국 고등학생에 대한 Holland 직업흥미이론의 문화적 타당성 평가. *교육평가연구*, 14(1), 237-262.
- 이기종(2000). *구조방정식 모형*. 서울: 교육과학사.
- 이순목(2000). *요인분석의 기초*. 서울: 교육과학사.
- Baxter, G.P., Shavelson, R.J., Goldman, S.R., & Pine, J. (1992). Evaluation of procedure-based scoring for hands-on science assessment. *Journal of Educational Measurement*, 29, 1-17.
- Bentler, P.M. (1990). Comparative fit indexes in structural model. *Psychological Bulletin*, 107, 238-246.
- Brennan, R. (1983, 1992). *Elements of Generalizability Theory* (2nd.). Iowa City, IA: The American College Testing Program.
- Carver, C. S. (1989). How should multifaceted personality constructs be tested? Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of Personality and Social Psychology*, 56, 577-585.
- Cohen, D., Stern, V., & Balaban, N. (1983). *Observing and recording the behavior of young children*, 3rd Ed. New York: Teachers Colleges Press.
- Colton, D. A., Gao, X., & Kolen, M. J. (1997). Assessing in reliability of performance level scores using bootstrapping. *Act Research Report Series*, 97(3), 41-55.
- Crick, J. E. & Brennan, R. L. (1983). *Manual for GENOVA: A generalized analysis of variance system*. The American College Testing Program, Technical Bulletin no. 43, Iowa City, Iowa.

- Cronbach, L. J., Glesser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The generalizability of behavioral measurement: The theory of generalizability for scores and profiles*. New York: John Wiley.
- Hadeed, J. & Syvia, K. (1999). Behavioral observations as predictors of children's social and cognitive progress in day care. *Early Child Development and Care*, 154, 13-30.
- Hu, L. & Bentler, P.M. (1995). Evaluating model fit. In R. H. Hoyle (ed.), *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.
- Joreskog, K.G. (1973). A general model for estimating a linear structural equation system, In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences*. Academic Press, 85-112.
- Kaplan, D. (2000). *Structural Equation Modeling: Foundations and extensions*. AQT 10, CA: Sage.
- Kim, S. (2000, April). *Investigating the generalizability of scores from different rating system in performance assessment*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans.
- Kim, S., Brown, R.S., & King, A. (2004, April). *Construct validation of an observational profile using generalizability theory and structural equation modeling*. Paper presented at the annual meeting of the AERA. San Diego.
- Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
- Landerholm, E. Karr, J., & Mushi, S. (2000). A Collaborative approach to family literacy evaluation strategies. *Early Child Development and Care*, 162, 65-79.
- Lane, S., Liu, M., Ankermann, R. D., & Stone, C. A. (1996). Generalizability and validity of a mathematics performance assessment. *Journal of Educational Measurement*, 33(1), 71-92.
- Loehlin, J.C. (1987). *Latent variable models: An introduction to factor, path, and structural analysis*. Hillsdale, NJ.: Lawrence Erlbaum Associate.
- Long, J.S. (1983). *Confirmatory Factor Analysis*. Beverly Hills, CA: Sage publications.
- Mulaik, S.A., James, L.R., Alstine, J.V., Benett, N. Lind, S., & Stilwell, C.D. (1989). An estimation of goodness of fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.
- Muthen, L. & Muthen, B (2001). *Mplus User's guide*, (2nd ed.). Los Angeles, CA: Muthen & Muthen.
- Shavelson, R., Baxter, P., & Gao, X. (1993). Sampling variability of performance assessment.

- Journal of Educational Measurement*, 30(3), 215-235.
- Shavelson, R. & Webb, N.M. (1991). *Generalizability theory: A Primer*. CA: Sage.
- Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Wilson, M. (2002). *Measurement: A constructive approach*. University of California, Berkeley.

• 논문접수 : 2004년 10월 13일 / 수정본 접수 : 2005년 5월 17일 / 게재 승인 : 2005년 6월 2일

ABSTRACT

An Investigation of Generalizability and Construct Validation of an Observational Developmental Profile

Sung-Sook Kim

(Professor, Inha University)

By providing such validity and reliability evidence for the use of an instrument, program providers can get a better picture of the entirety of a child's development progress. The purpose of this study, therefore, is to seek construct validity evidence from an initial pilot study for an observational instrument using generalizability theory (G-theory) and confirmatory factor analysis (CFA) to explore the reliability and validity of the instrument, respectively. As a result, evaluations of such programs will be better equipped to identify the influences these programs may have on the full range of participating children's developmental progression by utilizing instruments sensitive to and capable of measuring development. The instrument, the California Department of Education's Desired Results Developmental Profile (DRDP), is an observational profile used by child care providers in State sponsored facilities. Its purpose is to assess the developmental progress of children from birth to age 14 in four "desired results" areas of development : social-personal, cognitive, physical, and health/safety.

Looking at the results of generalizability analysis obtained using multiple observations, we found that high generalizability is achievable using a construct centered reliability approach where there is a high degree of fit between the substantive expectations generated from the researcher's understanding of the construct as realized by the measurement procedure and observation. The results analyzed by CFA suggest that the developmental domain of socio-emotional development can be explained by the indicators of self-concept, social interaction skills, self-regulation, language acquisition, and diversity. Similarly, the domain of school-age cognitive development can be explained by the constructs of interest in learning, cognitive competence, math skills, literacy skills, and diversity. Such a model provides excellent explanatory power in that use of individual dimensions of a construct can sometimes predict dependent variables more accurately than a more general construct.

Key Words : child's development, observation, desired result, reliability, construct validation, generalizability theory, confirmatory factor analysis.