

수직 척도점수를 이용한 학년 간 비교의 타당성

- 초등학생용 ICT 리터러시 검사 -

민 경 석(세종대학교 조교수)*

《 요 약 》

시간의 흐름(연령 혹은 학년)에 따른 학생의 성장을 측정하는 것은 사회과학의 중요한 관심 영역으로 다루어졌으며, 측정이론에서 학생의 변화/성장을 측정하는 방법으로 주로 수직 척도가 활용되어져 왔다. 이 논문의 목적은 교육/심리 분야에서 다양하게 활용되는 수직 척도점수를 이용한 학년 간 비교의 타당성을 검증하는 데 있다. 구체적으로 2008년 초등학생용 ICT 리터러시 검사에서 개발된 수직 척도점수의 의미를 분석하기 위하여 단계 간, 학년 간 측정모형의 동등성을 다집단 구조방정식모형을 이용하여 검증하였다. 측정모형의 비교 결과, 단계 간 검사의 일차원성이라는 요인 구조 동등성은 만족되었으나, 잠재요인이 각 관찰점수(능력영역)에 영향을 미치는 정도(요인부하량)는 검사 단계에 따라 다르게 나타났다. 이는 수직 척도점수가 제공하는 성장 정보가 단계, 학년에 따라 차별적이며, 결국 측정모형의 동등성 정도에 따라 수직 척도점수의 활용 수준이 제한되어야 함을 의미한다. 또한 학생의 성장/변화를 보다 구체적이고 일관되게 확인하기 위해서는 일반적으로 활용되는 단일 총점 보다는 현실의 다차원성을 반영하는 다양한 하위영역 점수를 활용할 필요가 있음이 제안되었다.

주제어 : 수직 척도, 초등학생용 ICT 리터러시 검사, 다집단 구조방정식모형

I . 서론

사회과학 분야에서 개인 혹은 집단의 변화와 성장에 대한 측정과 예측은 지속적으로 중요한 연구주제로 다루어져 왔다(Collins, & Horn, 1991; Collins, & Sayer, 2001; Gottman, 1995; Lissitz, 2006). 특히 교육학 영역에서 시간 흐름(연령 혹은 학년)에 따른 인지적, 정서적, 신체적 성장은 적절한 교육적 개입(보충교육, 학년 진급 등)의 결정을 위한 근거로 활용될 수 있

* 제1저자 및 교신저자, minkyungseok@sejong.ac.kr

으며, 집단 수준의 변화는 학교/교사의 책무성 혹은 교육과정의 적절성과 관련지어 논의되어 왔다(Ballou, Sanders, & Wright, 2004; McCaffrey et al., 2004).

학생의 변화와 성장을 판단할 수 있는 근거로서 주요하게 수직 척도점수(vertically scaled scores)가 활용되어져 왔다(성태제, 2000; Kolen, & Brennan, 2004; Mislevy, 1992). 수직 척도점수는 동일한 내용/구인을 측정함에도 불구하고, 피험자 집단의 능력수준에 따라 난이도에 차이가 있는 검사의 점수를 연계하는 측정이론적 방법으로, 동등화(equating)와 동일한 통계적 절차를 활용하여 수준별 검사의 비교를 가능하게 한다. 대표적인 수직 척도점수로는 전통적인 학년/연령점수(Mehrens, & Lehmann, 1991), Thurstone 척도(Thurstone, 1925, 1928)와 1980년대 이후 주로 활용된 문항반응이론 방법(item response theory, Lord, 1980) 등이 있다.

학년 간 점수 비교를 위한 수직 척도점수에 대한 주요 논쟁점은, 먼저 방법론적 측면에서 측정하고자 하는 구인(construct)의 실체에 부합하는 척도가 무엇인가에 집중되어져 왔다. 예를 들어, 성취도 검사에서 전통적인 Thurstone 방법은 학년이 높아짐에 따라 학생 점수의 분산이 커지는 경향을 보이는 반면, 문항반응이론 방법은 분산이 일정하거나 작아지는 경향을 보인다. 이때 학생의 내재적 성취도 수준은 실제 관찰되지 않기 때문에 어느 척도화 방법이 적절한 것인가는 연구자의 방법론적 편의성 혹은 상식 수준의 판단에 의하여 결정되는 제한점을 보인다. 두 번째, 내용적 측면에서 수직적으로 연계된 점수를 학년 간에 비교하는 경우, 고학년과 저학년 사이의 점수 차이가 의미하는 바가 무엇인가이다. 일반적으로 단일 총점을 보고하는 검사는 명시적 혹은 암묵적으로 측정하고자 하는 구인의 일차원성(unidimensionality)을 가정한 것이라 할 수 있다. 이 경우 학년 간 차이가 많은 경우(예, 초등학교 3학년과 6학년의 비교) 일관된 의미에서 점수비교가 가능한 것인가에 대한 질문이다. 예를 들어, 수학과목 성취도 검사에서 초등학교 저학년의 점수 차이는 낮은 수준의 ‘수학적 연산’에서 주로 나타날 것이며, 고학년의 점수 차이는 보다 추상적인 ‘수학적 이해’ 혹은 ‘문제 해결력’에서 나타날 수 있을 것이다. 이때 수직적으로 연계된 점수를 활용한 학년 간 비교에서, 저학년에서의 10점 차이는 고학년의 10점 차이와 수치상 동일한 점수 간격을 나타내지만 내용적으로 서로 다른 의미를 갖는다(Schulz, & Nicewander, 1997; Yen, 1986; Yen, & Burket, 1997).

이러한 학년 간 비교를 위해 수직적으로 연계된 점수에 대한 방법론적, 내용적 측면의 논쟁에도 불구하고 국외 검사(예, ACT, SAT 등)의 경우 다양한 형태의 수직 척도점수가 개발/보고되어온 반면(ACT, 2007; College Board, 2008), 우리나라에서는 학년 간 수직적 점수연계가 전국수준 표준화 검사에서 실제 활용되기 보다는 학술적 연구수준의 논의에 제한되어왔다(김성훈, 2000; 민경석, 서순식, 2009; 부재율, 2000, 2005; 부재율, 서동엽, 2000).

이 논문의 목적은 2008년 개발된 초등학생용 ICT 리터러시(information & communication technology literacy) 검사의 수직 척도점수(민경석, 서순식, 2009; 서순식 외, 2008)를 통한 학년

간 점수비교의 타당성을 검증하기 위하여 단계 간, 학년 간 측정모형 동등성을 비교하는 것에 있다. 측정모형 동등성은 검사 점수 연계의 전제조건으로, 동등성의 정도에 따라 수직 척도점수의 활용 및 해석 수준이 제한된다. 구체적으로 피험자 수준을 고려하여 수준별로 구성된 검사가 측정하는 구인의 집단간 동일성과 차별성을 다집단 구조방정식모형(multi-group structural equation models)을 이용하여 분석하였으며, 이에 따른 학년 간 수직 척도점수의 차이와 그 의미를 논의하였다.

Ⅱ. 초등학생용 ICT 리터러시 검사

1. 검사구성 및 척도점수

2008년 한국교육학술정보원에서 개발한 초등학생용 ICT 리터러시 검사(서순식 외, 2008)는 피험자 수준(학년)을 고려하여 3개의 검사지로 구성되었다. 구체적으로 1단계 검사는 1-2학년, 2단계 검사는 3-4학년, 3단계 검사는 5-6학년을 대상으로 한다. 3개 단계 검사의 구성은 ‘초/중등학교 정보통신 교육 운영지침(교육인적자원부, 2005, 2006)’에 기반하였으며, 각 단계 검사의 내용 및 능력영역별 구성은 <표 1>과 같다.¹⁾

<표 1>에 제시된 바와 같이 초등학생용 ICT 리터러시 검사는 각 단계별로 3개 내용영역(학생들이 학습할 내용)과 6개 능력영역(ICT를 활용하여 수행하는 활동)의 교차표에 따라 구성되었다. 국가수준 교육운영지침에 근거하여 1단계 검사(1-2학년용)에서 3개 내용영역 중 ‘정보처리’ 영역의 문항이 출제되지 않았으며, 또한 저학년이라는 피험자 수준을 고려하여 상대적으로 적은 문항(20문항)으로 검사가 구성되었다. 이러한 검사의 내용 구성과 문항 수의 차별성을 고려하여, 수직 척도점수의 학년 간 비교를 위한 측정모형 동등성 분석에서 초등학생용 ICT 리터러시 검사 중 2단계(3-4학년)와 3단계(5-6학년)만을 대상으로 하였다.

1) 초등학생용 ICT 리터러시 검사의 구성과 점수 체제에 대한 보다 상세한 내용은 서순식 외(2008)를 참조할 것.

〈표 1〉 초등학생용 ICT 리터러시 검사의 구성

영역	하위 요소	1단계 문항수	2단계 문항수	3단계 문항수
내용영역	컴퓨터와 네트워크	8	18	15
	정보처리	-	6	9
	정보 사회와 윤리	12	8	8
능력영역	문제 인식	3	2	2
	정보 탐색	4	7	3
	정보 분석과 평가	2	4	4
	정보 조직과 창출	3	7	10
	정보 활용과 관리	6	7	9
	정보 소통	2	5	4
	합 계	20	32	32

초등학생용 ICT 리터러시 검사 결과의 보고를 위하여 각 단계별로 척도점수와 성취수준이 설정되었다. 척도점수는 2단계와 3단계에서 각각 250, 350점으로 설정되었으며, 두 단계 모두에서 표준편차는 10점으로 고정되었다. 여기서 척도점수는 원점수에 대한 아크사인 전환(arcsine transformation) 절차(반재천, 2006; Kolen & Brennan, 2004)를 따른 결과이며, 점수 해석의 편의를 위하여 평균과 표준편차가 임의로 결정되었다. 또한 각 단계별 척도점수를 활용하여 ‘우수수준’과 ‘기본수준’ 등의 두 가지 성취수준이 설정되었다.

2. 수직 척도점수

초등학생용 ICT 리터러시 검사에서 수준별 2단계와 3단계 검사의 척도점수는 각 단계 내에서 해석되는 자체적 완결구조를 갖는다. 즉, 각 단계의 척도점수는 임의의 평균과 표준편차를 갖는 점수로서 단계 간 혹은 학년 간 비교를 통하여 학생의 ICT 리터러시 성장을 진단하지 못한다. 이러한 점수 해석의 제한점을 해결하고, 초등학교 정보통신교육이 단계적이고 연속적인 국가수준 교육과정에 기반한 점에 주목하여 민경석과 서순식(2009)은 단계 간 점수 비교를 위하여 3단계 검사를 2단계 척도(평균 250점, 표준편차 10점)로 연계한 수직 척도점수를 개발하였다. 두 연구자는 공통 피험자(common examinees)를 활용한 다양한 척도 전환방법 중 평균방법(mean method), 선형방법(linear method)과 동백분위 방법(equal percentile method)의 결과를 비교하여 선형방법이 가장 안정적인 것으로 보고하였으며, 이들의 절차에 따른 전체 피험자에 대한 수직 척도점수의 기술통계치는 〈표 2〉와 같다.

〈표 2〉 단계별, 학년별 수직 척도점수

검사 수준	학년	사례수	평균	표준편차	학년	사례수	평균	표준편차
2단계	3-4	1670	250.01	9.96	3	806	248.48	9.90
					4	864	251.44	9.82
3단계	5-6	1773	258.63	8.63	5	897	256.49	8.50
					6	876	260.81	8.22

〈표 2〉에 제시된 바와 같이 수직 척도점수를 통한 2단계에서 3단계로의 변화는 8.62점(=258.63-250.01)이며, 보다 세부적으로, 학년 간 평균 점수의 변화는 약 3-7점에 해당하는 것으로 나타났다. 그러나 수직 척도점수를 통한 단계 간 혹은 학년 간 비교에서 중요한 점은 평균차이로 표현된 변화가 의미하는 바가 무엇인가이다. 만약 단계별 혹은 학년 간 검사가 동일한 구인을 측정하지 않는다면, 즉 학년별로 학생들이 활용하는 ICT 지식과 기술이 다르다면, 통계적 방법으로 설정된 수직 척도점수의 의미는 명확히 제시될 수 없을 것이다.

Ⅲ. 연구방법

1. 연구대상 및 원점수 분포

2008년 개발된 초등학생용 ICT 리터러시 검사는 전국 초등학교의 분포 지역과 학교 규모를 고려한 유층무선표집(stratified random sampling)을 통한 31개 초등학교에서 실시되었으며, 2단계와 3단계 검사의 원점수(각 32점이 최고점)에 대한 기술통계치는 〈표 3〉과 같다.

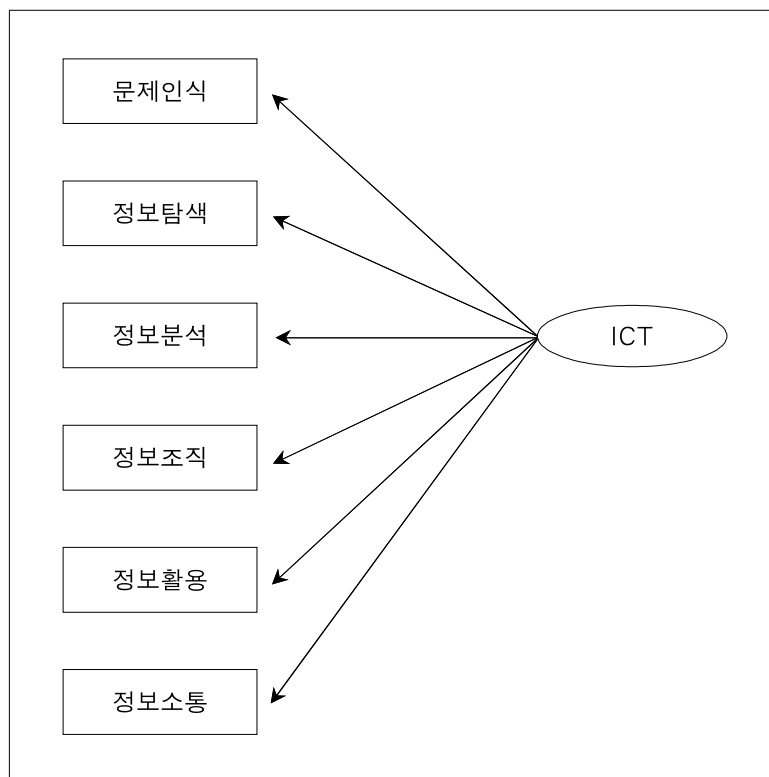
〈표 3〉 단계별, 학년별 원점수 기술통계치

검사 수준	학년	사례수	평균	표준편차	학년	사례수	평균	표준편차
2단계	3-4	1670	15.94	7.00	3	806	14.81	6.91
					4	864	17.00	6.93
3단계	5-6	1773	17.68	6.07	5	897	16.17	6.01
					6	876	19.23	5.74

〈표 3〉은 각 단계 검사의 원점수 분포를 제시한 것으로, 단계 간, 학년 간 비교를 위해서는 〈표 2〉에 제시된 수직 척도점수를 활용할 수 있다.

2. 측정모형의 동등성

수직 척도점수의 올바른 해석과 활용을 위하여 전제되어야 할 것은 연계된 검사가 측정하는 구인의 동등성이다. 초등학생용 ICT 리터러시 검사는 학년 간 내용적 수준 차이에도 불구하고 공통적으로 ‘문제 인식’, ‘정보 탐색’, ‘정보 분석과 평가’, ‘정보 조직과 창출’, ‘정보 활용과 관리’, ‘정보 소통’ 등 6개 능력영역으로 구성되며(〈표 1〉 참조)²⁾ 이를 종합하여 전체 점수(ICT 리터러시)를 산출한다. 이를 개념적 측정모형인 구조방정식 모형으로 표현하면 [그림 1]과 같다.



[그림 1] ICT 리터러시 검사의 개념적 측정모형

[그림 1]에서 단일 구인이 가정되었으며, ‘ICT 리터러시’라는 잠재구인이 6가지 능력영역에 영향을 주는 것으로 나타난다. 즉, ICT 리터러시 검사는 6개 능력영역을 구분하여 문항이 구성되며, 최종적으로 단일 총점이 보고되기 때문에 하나의 잠재구인을 측정하는 일차원 모

2) 측정모형에서 ICT 활용능력에 주안점을 두어 내용영역이 아닌 능력영역을 하위 요인으로 설정하였음.

형이 가정된 것이라 할 수 있다.

[그림 1]에 제시된 측정모형을 수리적으로 표현하면 식 (1)과 같다(민정석, 2006; Bollen, 1989).

$$X_i = \lambda_i \xi + \delta_i \quad (1)$$

여기서 X_i 는 i 번째 능력영역 관찰점수(6개 영역)를 나타내며, 각 영역별 관찰점수의 분산은 잠재구인(ξ , ICT 리터러시)의 분산(ϕ)과 오차항(δ_i)의 분산(θ_{δ_i})으로 구성된다. 요인부하량(factor loadings), λ_i 는 표준화계수(standardized coefficients)로 표현되는 경우, 각 능력영역 관찰점수와 잠재구인의 상관계수를 의미한다. 또한, 식 (1)의 측정모형은 오차항과 잠재구인이 서로 독립적이며($\rho_{\xi\delta} = 0$), 오차항과 잠재구인의 기대값이 각각 $0(E(\delta) = 0, E(\xi) = 0)$ 임을 통계적으로 가정한다.

식 (1)로 표현된 측정모형의 집단간 동등성은 상관행렬³⁾, 요인구조(factor structure), 요인부하량 비교 등 세 가지 단계로 검증될 수 있다(Joreskog & Sorbom, 1989, 9장 참조). 먼저 상관행렬의 동등성은 관찰변수의 상호 관계가 일관적임을 의미하며, 요인구조의 동등성은 관찰변수와 잠재변수의 관계를 나타내는 구조행렬(factor structure matrix)의 일관성(여기서는 일차원성)을 의미한다. 또한 요인부하량의 동등성은 요인구조의 동등성에 기반하여 각 관찰변수에 대한 요인부하량의 추정치가 일정한가를 검증한다. 상관행렬의 동등성을 검증하기 위한 구조방정식모형에서는 관찰변수 개수와 동일한 수의 잠재변수를 설정하기 때문에 단일 잠재변수를 가정한 요인구조, 요인 부하량 비교 모형과 통계적으로 위계적 관계를 갖지 않는다. 그러나 집단간 상관행렬의 비교에서 만약 동등성이 기각된다면, 이는 요인구조 혹은 요인부하량의 차별성에서 원인을 찾을 수 있을 것이다. 이러한 측면에서 세 모형은 내용적으로 위계적 특성을 가지며, 상관계수 동등성의 검정은 실제 관찰된 측정변수의 상호관계를 직접적으로 비교한다는 장점을 보인다.

또한 초등학생용 ICT 리터러시 검사는 일종의 성취도 검사이며 피험자의 성장에 따라 단계 간, 학년 간 능력수준이 다르다는 점에 근거하여, 측정모형의 다집단 비교를 위하여 위 세 가지 모형 외에 절편 동등성 및 평균 동등성 등에 대한 검정(박현정, 2008 참조)은 이 논문에서 고려되지 않았다.

수직 척도점수의 학년 간 비교를 통하여 학생의 성장과 변화를 측정하기 위해서는 동일한 측정모형이 모든 검사 수준 혹은 학년 간에 일정하게 유지됨을 전제로 한다. 구체적인 측정

3) 수준별 검사의 능력영역별 문항수가 다르기 때문에 공분산행렬이 아닌 상관행렬을 통하여 관찰변수의 상호관계 동등성을 검정하였음.

모형의 집단 간 동등성을 확인하기 위하여 LISREL(version 8.54, Joreskog & Sorborn, 2003)의 다집단 구조방정식모형(Joreskog & Sorborn, 1989)을 이용하였다.⁴⁾

IV. 연구결과

초등학교용 ICT 리터러시 검사의 수직 척도점수를 통한 학년 간 비교의 타당성을 확인하기 위하여, 먼저 2단계 검사와 3단계 검사의 측정모형 동등성을 검정하였으며, 다음으로 각 단계 내 두 개 학년에 대한 비교를 실시하였다.

1. 단계 간 측정모형의 동등성

초등학교용 ICT 리터러시 검사에서는 3-4학년과 5-6학년을 대상으로 각각 2단계 검사와 3단계 검사를 실시하였다. 이에 따른 단계별 6개 능력영역의 관찰점수 상관행렬은 <표 4>와 같다.⁵⁾

<표 4> 2단계, 3단계 검사의 상관행렬

능력영역	1	2	3	4	5	6
1. 문제 인식	–	0.24**	0.37**	0.39**	0.43**	0.28**
2. 정보 탐색	0.40**	–	0.32**	0.39**	0.40**	0.27**
3. 정보 분석과 평가	0.31**	0.48**	–	0.51**	0.53**	0.35**
4. 정보 조직과 창출	0.37**	0.56**	0.51**	–	0.57**	0.39**
5. 정보 활용과 관리	0.44**	0.59**	0.52**	0.63**	–	0.44**
6. 정보 소통	0.39**	0.55**	0.47**	0.57**	0.59**	–

* $p<.05$, ** $p<.01$

– 대각선 아래는 2단계 검사, 위는 3단계 검사의 상관행렬을 나타냄.

<표 4>의 상관행렬과 표준편차를 다집단 구조방정식모형의 입력자료로 활용하였으며, 단계 간 상관행렬, 요인구조, 요인부하량의 동등성 검정 결과는 <표 5>와 같다.

4) 단계 간 검사의 측정모형 동등성 검정을 위한 세 가지 모형의 LISREL 명령문이 부록에 제시됨.

5) 이후 제시되는 상관행렬은 원점수(정답한 문항 수)를 이용하여 계산된 것임.

〈표 5〉 단계 간 측정모형 동등성 검정

검정모형	χ^2	df	RMSEA	CFI	NNFI	IFI	GIF
상관행렬	221.55**	21	0.08	0.98	0.97	0.98	0.98 0.98
요인구조	28.36	18	0.02	1.00	1.00	1.00	1.00 1.00
요인부하량	646.23**	23	0.13	0.94	0.92	0.94	0.93 0.95

* $p < .05$, ** $p < .01$

〈표 5〉에서 다집단 구조방정식모형의 전체적인 적합성을 평가하기 위하여 전통적 검정 통계치인 χ^2 와 함께 RMSEA(root mean square error of approximation), CFI(comparative fit index), NNFI(nonnormed fit index), IFI(incremental fit index) 등의 적합도 지수를, 개별 집단의 모형 적합도 지수로 GIF(goodness of fit)를 제시하였다. GIF 열의 각 칸에서는 2단계 검사와 3단계 검사 순으로 지수를 표시하였다.

측정모형 동등성 검정 결과를 요약하면, 2단계와 3단계의 상관행렬과 요인부하량의 동등성은 기각되었으며(두 모형의 χ^2 와 RMSEA가 상대적으로 큰 값으로 나타남.), 요인구조 동등성은 기각되지 않았다. 즉, 2단계 검사와 3단계 검사에서 6개 능력영역의 상호관계는 일관성을 보이지 않는 반면, [그림 1]에서 제시된 단일 구인을 가정한 일차원 측정모형은 두 검사에서 일관되게 적용된다. 그러나 일차원 측정모형에서 요인부하량은 두개 단계 검사에서 일정하지 않은 형태로 나타난다. 결국, 2단계와 3단계 검사는 단일 구인을 측정하는 일차원적 요인구조를 가정에도 불구하고, 잠재구인이 각 하위점수에 미치는 영향력은 차별적이며, 이는 관찰점수 상관행렬의 단계 간 차이로 나타난다.

통계적으로 기각되지 않은 요인구조 동등성 모형에서 추정된 각 단계별 요인부하량과 각 관찰변수에 대한 잠재구인의 설명력 추정치는 〈표 6〉과 같다. 〈표 6〉에서 요인부하량은 두 단계 검사의 관찰변수와 잠재변수 모두에 대한 공통의 척도를 가정한 표준화 계수(common metric completely standardized solution)로 표시되었으며, 설명력은 각 집단 내에서 표준화(within group completely standardized solution)된 잠재요인에 의하여 설명된 관찰변수의 분산비율로 정의된다.⁶⁾ 여기서 완전 표준화(completely standardized solution)란 관찰점수와 잠재변수가 모두 표준화 된 것을 의미한다(LISREL 명령문의 OU줄에 SC 명령어 활용, 부록 참조)⁷⁾.

6) 공통척도 표준화와 집단내 표준화 절차의 수리적 차이로 〈표 6〉에서 요인부하량의 제곱이 설명력 추정치와 일치하지 않는다.

7) 추가적으로 잠재변수만을 표준화(OU 명령어 SS)한 요인부하량의 패턴(상대적 크기) 또한 〈표 6〉과 동일하게 나타남.

요인구조의 동등성을 가정한 모형을 통하여 2단계 검사에서 ‘ICT 리터러시’라는 단일 잠재요인은 주요하게 ‘정보 탐색’과 ‘정보 소통’에 많은 영향을 주는 반면, 3단계 검사에서는 ‘정보 조직과 창출’과 ‘정보 활용과 관리’가 상대적으로 많은 영향을 받는다. 즉, 2단계와 3단계 검사는 포괄적으로 ‘ICT 리터러시’를, 세부적으로 6개 능력영역을 측정함에도 불구하고 검사 단계(학년)에 따라 최종 검사 점수에 대한 하위요인의 영향 정도에 있어 차별성을 나타낸다고 할 수 있다. 또한 요인부하량의 수준별 차이는 관찰점수의 상호관련성을 나타내는 상관행렬의 차이로 이어진다.

〈표 6〉 단계 간 요인부하량과 설명력 추정치

능력 영역	2단계 검사		3단계 검사	
	요인부하량($\hat{\lambda}_i$)	설명력	요인부하량($\hat{\lambda}_i$)	설명력
문제 인식	0.53	0.27	0.53	0.29
정보 탐색	0.96	0.54	0.32	0.25
정보 분석과 평가	0.66	0.42	0.66	0.46
정보 조직과 창출	0.70	0.59	0.79	0.54
정보 활용과 관리	0.80	0.66	0.80	0.62
정보 소통	0.88	0.54	0.42	0.29

두 단계별 검사에서 요인부하량의 차이는 두 가지 방향에서 해석될 수 있다. 첫 번째는 ICT 리터러시의 능력영역에 있어 3·4학년과 5·6학년이 실제적인 능력의 차이를 나타내기 때문이라 할 수 있다. 즉, 저학년인 3·4학년 학생은 기본적인 ‘정보 탐색’과 ‘정보 소통’에서 주로 변별된다면, 고학년인 5·6학년은 ‘정보 조직과 창출’과 ‘정보 활용과 관리’에서 차이를 나타낸다고 볼 수 있다. 즉, 학년 집단에 따라 민감하게 반응(변별)하는 하위 능력영역이 다를 수 있다는 것이다. 두 번째는 검사지 구성의 차이에서 원인을 찾을 수 있다. 구체적으로, 2단계 검사에서 상대적으로 높은 부하량을 보인 ‘정보 탐색’과 ‘정보 소통’을 측정하는 문항이 3단계보다 많으며, 3단계 검사는 ‘정보 조직과 창출’과 ‘정보 활용과 관리’를 상대적으로 많은 문항으로 측정한다(〈표 1〉 참조). 즉, 많은 문항으로 구성된 하위 영역은 상대적으로 큰 분산을 나타내며, 이러한 특성은 전체 점수와 높은 상관으로 이어진다.

결국, 하나의 잠재요인을 측정하는 일차원성이라는 요인구조의 동등성에도 불구하고 각 단계별 요인부하량의 차별성은 통계적으로 설정된 수직 척도점수의 해석에 중요한 의미를 갖는다. 〈표 2〉에 제시된 바와 같이 2단계와 3단계 검사에서 나타난 수직 척도점수에서 약 9점 정도 차이가 나타난다. 이는 3·4학년에서 5·6학년으로 학년이 높아질수록 포괄적으로 ‘ICT

리터러시'가 약 9점 성장했음을 의미하며, 보다 세부적으로는 고학년으로 올라갈수록 '정보 탐색'과 '정보 소통'으로 부터 '정보 조직과 창출'과 '정보 활용과 관리' 능력에 대한 변화/성장이 일어난 것으로 해석할 수 있을 것이다.

2. 학년 간 측정모형의 동등성

초등학교용 ICT 리터러시 검사의 단계별 검사는 두 개 학년을 대상으로 하며, 각 학년의 능력영역 관찰점수 상관행렬은 각각 <표 7>, <표 8>과 같다.

<표 7> 3학년과 4학년의 상관행렬

능력영역	1	2	3	4	5	6
1. 문제 인식	-	0.38**	0.32**	0.39**	0.47**	0.40**
2. 정보 탐색	0.41**	-	0.45**	0.54**	0.58**	0.56**
3. 정보 분석과 평가	0.28**	0.51**	-	0.52**	0.51**	0.47**
4. 정보 조직과 창출	0.34**	0.56**	0.49**	-	0.64**	0.57**
5. 정보 활용과 관리	0.40**	0.59**	0.54**	0.59**	-	0.58**
6. 정보 소통	0.38**	0.53**	0.46**	0.55**	0.59**	-

* p<.05, ** p<.01

- 대각선 아래는 3학년, 위는 4학년의 상관행렬을 나타냄.

<표 8> 5학년과 6학년의 상관행렬

능력영역	1	2	3	4	5	6
1. 문제 인식	-	0.15**	0.37**	0.34**	0.42**	0.26**
2. 정보 탐색	0.29**	-	0.25**	0.34**	0.34**	0.23**
3. 정보 분석과 평가	0.33**	0.33**	-	0.49**	0.51**	0.32**
4. 정보 조직과 창출	0.41**	0.41**	0.50**	-	0.53**	0.38**
5. 정보 활용과 관리	0.41**	0.41**	0.50**	0.57**	-	0.45**
6. 정보 소통	0.28**	0.28**	0.34**	0.36**	0.40**	-

* p<.05, ** p<.01

- 대각선 아래는 5학년, 위는 6학년의 상관행렬을 나타냄.

학년 간 측정모형 동등성을 검정하기 위하여, <표 7>과 <표 8>의 상관행렬과 관찰점수 표준편차를 입력자료로 활용하였으며, 다집단 구조방정식모형의 적용 결과는 <표 9>와 같다.

<표 9>에서 각 칸의 첫 번째 숫자는 3-4학년의 측정모형 동등성 검정 결과를, 두 번째 숫자는 5-6학년의 비교 결과를 나타내며, GIF 열에는 각 학년에 대한 적합도 지수가 제시되었다. 학년 간 측정모형 동등성 검정결과를 요약하면, 상관행렬, 요인구조, 요인 부하량 모형에 대한 모든 χ^2 통계치가 통계적으로 유의하게 나타났다. 그러나 χ^2 통계치가 사례수에 민감

하여 표본수가 커질수록 과도하게 모형 동등성을 기각하는 경향이 있다는 점이 고려될 필요가 있다. 또한 표본의 크기와 모형의 간명성을 고려한 적합도 지수는 모두 적절한 수준으로 나타났다.⁸⁾ 특히, 단계 간 측정모형 동등성 검정 결과인 <표 5>와 비교하여, <표 9>에서 상관행렬과 요인부하량 동등성 모형의 모든 적합도 지수가 매우 높아졌다.

〈표 9〉 학년 간 측정모형 동등성

검정모형	χ^2	df	RMSEA	CFI	NNFI	IFI	GIF
상관행렬	47.56**	21	0.04	1.00	0.99	1.00	0.99
	43.30**	21	0.04	0.99	0.99	0.99	0.99
요인구조	33.03*	18	0.03	1.00	1.00	1.00	0.99
	30.57*	18	0.03	1.00	0.99	1.00	0.99
요인부하량	47.75*	23	0.04	1.00	0.99	1.00	0.99
	40.51*	23	0.03	1.00	0.99	1.00	0.99

* $p<.05$, ** $p<.01$

결국, 각 단계 검사는 두 개 학년을 대상으로 동일한 내용을 측정한다고 할 수 있다. 이러한 단계 간 비교와 단계 내 학년 간 비교의 차별성은 먼저 단계 내 두 개 학년에는 정확히 동일한 검사가 시행되었다는 검사 체계적 특성과 학년 차이가 상대적으로 적은 대상을 비교했다는 피험자 수준 요인의 두 가지 측면이 동시에 작용한 것이라 할 수 있다. 특히, 일차원 측정모형(요인구조와 요인부하량)의 일관성을 보이는 각 단계 내 두 개 학년은 교육현장에서 실제 활용되는 관찰점수의 상관행렬 또한 동등성을 유지하는 것으로 나타났다.

V. 결론 및 논의

한국교육학술정보원의 초등학생용 ICT 리터러시 검사는 국가수준 교육과정에 근거하고 학년별 피험자 수준을 고려하여 3개 단계의 검사로 2008년에 개발되었으며, 초등학생의 성장

8) Bollen(1989)과 Kline(1998)은 경험적으로 $RMSEA<0.05$, $CFI>0.9$, $NNFI>0.9$, $IFI>0.9$ 일 때 연구모형이 적합한 것으로 제안함.

에 따른 ICT 리터러시의 변화를 측정하기 위하여 단계별 검사를 연계한 수직 척도점수가 제안되었다. 동일한 능력을 측정하는 검사를 활용하여 피험자의 변화와 성장을 진단하는 것은 교육 분야의 중요한 과제로 미국의 경우 우리나라 대학수학능력시험에 해당하는 ACT, SAT와 하위학년 검사(예, EXPLORE, PLAN; PSAT/NMSQT)를 연계하여 학생, 고등학교, 대학교에 다양한 정보를 제공하여 왔다(ACT, 2007; College Board, 2008). 이러한 수직 척도점수가 제공하는 성장과 변화에 대한 정보의 중요성에도 불구하고, 수직 척도점수에서 학년 간 차이가 의미하는 바가 무엇이며, 수직 척도점수의 방법론적 내용적 타당성과 활용 한계에 대한 논의는 지속되어왔다(Kolen & Brennan, 2004; Reckase, 2004; Yen, 1986).

수직 척도점수의 학년 간 비교를 통하여 학생의 성장과 변화를 측정하기 위해서는 동일한 측정모형이 모든 검사 수준 혹은 학년 간에 일정하게 유지됨을 전제로 한다. 이 논문에서는 초등학생용 ICT 리터러시 검사에서 개발된 수직 척도점수를 이용한 학년 간 비교의 타당성을 확인하기 위해서 다집단 구조방정식모형을 활용하여 2단계와 3단계 검사의 측정모형 동등성(상관계수, 일차원성의 요인구조, 요인부하량)을 검정하였다. 또한 각 단계 내에 두 개 학년이 포함된 점을 고려하여, 동일한 검사가 실시된 두 개 학년에 대하여 단계 간 비교와 같은 방법으로 측정모형 동등성 검정을 실시하였다.

먼저 단계 간 비교에서 상관계수 동등성과 요인부하량 동등성은 기각되었으며, 잠재구인의 일차원성에 대한 요인구조 동등성 모형이 가정될 수 있었다. 즉, 2단계 검사와 3단계 검사는 포괄적으로 ‘ICT 리터러시’를 측정한다는 점에서 일관성을 보이는 반면, 하위학년인 검사인 2단계는 주로 ‘정보 탐색’과 ‘정보 소통’에 높은 요인부하량이 나타나며, 3단계 검사는 ‘정보 조직과 창출’과 ‘정보 활용과 관리’ 영역을 상대적으로 민감하게 변별하는 특성을 보인다. 또한 각 단계 내 3학년과 4학년의 비교, 5학년과 6학년의 비교 모두에서 측정모형 동등성 수준이 단계 간 비교 결과 보다 양호하게 나타났다. 즉, 학년별 요인부하량은 단계별 요인부하량과 비교하여 매우 유사하며, 이는 각 단계 내의 두 개 학년은 학년 격차가 많지 않고, 정확히 동일한 검사가 실시되었다는 점에 근거한다. 결국, 일차원성을 가정한 측정모형에서 요인부하량의 동등성 여부는 실제 교육현장에서 활용되는 관찰점수의 상관행렬 동등성의 전제조건으로 작용한다.

이상의 결과는 단일 총점이라는 일차원성을 가정하는 단순한 측정모형에 있어서도 피험자의 수준(학년)에 따라 검사 하위영역에 대한 반응이 차별적으로 나타나고 있음을 의미한다. 또한 난이도 차이를 보이는 검사를 통계적으로 연계한 수직 척도점수의 의미가 집단에 따라 동일하지 않을 수 있음을 의미한다. 예를 들어, ICT 리터러시 검사의 수직 척도점수에서 3학년과 4학년의 점수 차이가 5점이고 5학년과 6학년의 차이가 5점으로 동일한 간격일지라도 3, 4학년의 점수 차이는 주요하게 ‘정보 탐색’과 ‘정보 소통’에서 나타나며, 5, 6년의 차이는 ‘정보 조직과 창출’, ‘정보 활용과 관리’ 등에서 나타날 가능성이 높다. 곧 수직 척도점수 상에

서 같은 크기의 점수 차이일지라도 이것이 의미하는 바는 비교 대상 학년에 따라 달라질 수 있는 것이다. 이러한 차이 점수의 차별적 의미는 동일한 검사를 서로 다른 학년에 실시하는 경우에는 학년 간 능력 특성의 패턴에 따라 주요하게 결정될 것이며, 수준별 검사라는 서로 다른 검사를 학년별로 실시하는 경우에는 검사 특성(하위요인별 문항수, 문항 난이도 등)과 피험자 특성이 복합적으로 작용하여 차이 점수가 의미하는 바를 더욱 복잡하게 한다.

근래, 학부모, 교사, 학교행정가, 교육정책 결정자 등 다양한 교육 관계집단에서 교사/학교의 책무성에 관심을 보이며, 방법론적으로 다층모형과 구조방정식 모형과 같은 보다 엄밀한 분석방법이 발전함에 따라 추이분석, 성장점수모형, 부가가치모형 등 다양한 교육자료 분석이 이루어지고 있다(Ballou, Sanders, & Wright, 2004; McCaffrey et al., 2004 참조). 그러나 중요한 점은 이러한 통계모형이 비교하는 점수가 측정한 바가 무엇인지에 대한 명확한 확인이 전제될 때 이들 점수의 변화 혹은 차이의 의미가 밝혀질 것이며, 이에 대한 교육적, 정책적 개입의 방향성 또한 도출될 수 있을 것이다. 특히 학생들의 성장은 다양한 영역을 포괄하기에 보다 정확한 성장모형을 설정하기 위해서는 단일 차원성 보다는 현실의 다차원성을 고려할 필요가 있을 것이다(Reckase, 2004). 결국, 학년 간에 연계된 수직 척도점수일지라도 저학년의 점수차이와 고학년의 점수차이에 영향을 주는(혹은 민감하게 반응하는) 하위영역이 서로 다른 경우, 하나의 총점으로 비교되는 것이 아니라, 다양한 하위영역 점수로 학생의 성장을 진단하여 보다 상세한 정보를 제공하고 해당 영역에 대한 교육적 개입을 집중한다면 학생 성장을 보다 극대화할 수 있을 것이다.

참 고 문 헌

- 교육인적자원부(2005). **초 중등학교 정보통신기술 교육 운영지침**. 서울: 저자.
- 교육인적자원부(2007). **초 중등학교 정보통신기술 교육 운영지침 해설서(안)**. 서울: 저자.
- 김성훈(2000). 능력 변화과정 추정을 위한 시험자료의 동등화 방법 연구. **교육평가연구**, 13(1), 101-125.
- 민경석(2006). PISA 2003 정의적 특성의 측정모형 동등성. **한국교육**, 33(3), 31-49.
- 민경석, 서순식(2009). 초등학생용 ICT 리터러시 검사의 수직 척도화. **교육정보미디어연구**, 15(3), 1-19.
- 박현정(2008). 학습동기, 자아개념, 학업성취간 관계의 집단간 동등성 분석: PISA 2006 중심으로. **교육평가연구**, 21(3), 43-67.
- 반재천(2006). 국가수준 학업성취도 평가의 척도개발. **교육평가연구**, 19(2), 113-135.
- 부재율(2000). 수직적 동등화 방법들의 양호성 실증 비교 - 학년 간 자료를 이용하여. **교육평가연구**, 13(1), 81-100.
- 부재율(2005). 가교문항 수직적 동등화를 이용한 국어, 영어, 수학 학업성취도의 학년 간 비교 연구. **교육평가연구**, 18(1), 127-152.
- 부재율, 서동엽(2000). 수직적 동등화 결과 비교 분석 연구 - 단계형 수준별 교과(수학)에서. **교육과정평가연구**, 3(1), 141-160.
- 서순식, 민경석, 황경연, 장윤정, 김혜숙(2008). **초등학생용 ICT 리터러시 검사도구 타당화 연구**. 서울: 한국교육학술정보원.
- 성태제(2000). 다른 교과의 점수를 연결하는 통계적 조절 방법의 타당성. **교육평가연구**, 13(1), 63-99.
- ACT (2007). *PLAN Technical Manual*. Iowa City, IA: Author.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37-66.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.
- College Board (2008). *Official Educator Guide to the PSAT/NMSQT 2008/2009*. New York, NY: Author.
- Collins, L. M., & Horn, J. L. (Eds.) (1991). *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions*. Washington, D.C.: American Psychological Association.
- Collins, L. M., & Sayer, A. G. (Eds.) (2001). *New Methods for the Analysis of Change*. Washington, D.C.: American Psychological Association.

- Gottman, J. M. (Ed.) (1995). *The Analysis of Change*. Mahwah, NJ: Lawrence Erlbaum.
- Joreskog, K., & Sorborn, D. (1989). *LISREL 7 User's Reference Guide*. Chicago, IL: Scientific Software International, Inc.
- Joreskog, K., & Sorborn, D. (2003). *LISREL 8.54*: Computer program for structural equation models. Scientific Software International, Inc.
- Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York, NY: Gilford.
- Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices (2ed Ed.)*. New York, NY: Springer.
- Lissitz, R. (Ed.) (2006). *Longitudinal and Value Added Models of Student Performance*. Maple Grove, MN: JAM Press.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
- McCaffrey, D. F., Lockwoold, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-102.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology (4th Ed.)*. New York, NY: Holt, Rinehart and Winston.
- Mislevy, R. J. (1992). *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects*. Princeton, NJ: ETS.
- Reckase, M. D. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics*, 29(1), 117-120.
- Schulz, E. M., & Nicewander, W. A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement*, 34(4), 315-331.
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451.
- Thurstone, L. L. (1928). The absolute zero in intelligence measurement. *The Psychological Review*, 36, 175-197.
- Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23(4), 299-325.
- Yen, W. M., & Burket, G. R. (1997). Comparison of item response theory and Thurstone methods of vertical scaling. *Journal of Educational Measurement*, 34(4), 293-313.

• 논문 접수 : 2009년 12월 29일 / 수정본 접수 : 2010년 2월 8일 / 게재 승인 : 2010년 2월 22일

[부록 1]

LISREL 명령문

1. 상관행렬 동등성	2. 요인구조 동등성	3. 요인부하량 동등성
GRADE 3 AND 4 DA NG=2 NI=6 NO=1670 LA f1 f2 f3 f4 f5 f6 KM FI=GRADE3-4.COR SD FI=GRADE3-4.COR MO NX=6 NK=6 LX=DI TD=ZE VA 0.73 LX 1 VA 1.70 LX 2 VA 1.19 LX 3 VA 1.88 LX 4 VA 1.92 LX 5 VA 1.54 LX 6 LK F1 F2 F3 F4 F5 F6 PD OU AL SC GRADE 5 AND 6 DA NI=6 NO=1773 LA f1 f2 f3 f4 f5 f6 KM FI=GRADE5-6.COR SD FI=GRADE5-6.COR MO NX=6 NK=1 LX=PS LK ICT PD OU AL SC	GRADE 3 AND 4 DA NG=2 NI=6 NO=1670 LA f1 f2 f3 f4 f5 f6 KM FI=GRADE3-4.COR SD FI=GRADE3-4.COR MO NX=6 NK=1 FR LX 2 LX 3 LX 4 LX 5 LX 6 VA 1 LX 1 LK ICT PD OU AL SC GRADE 5 AND 6 DA NI=6 NO=1773 LA f1 f2 f3 f4 f5 f6 KM FI=GRADE5-6.COR SD FI=GRADE5-6.COR MO NX=6 NK=1 LX=PS LK ICT PD OU AL SC	GRADE 3 AND 4 DA NG=2 NI=6 NO=1670 LA f1 f2 f3 f4 f5 f6 KM FI=GRADE3-4.COR SD FI=GRADE3-4.COR MO NX=6 NK=1 FR LX 2 LX 3 LX 4 LX 5 LX 6 VA 1 LX 1 LK ICT PD OU AL SC GRADE 5 AND 6 DA NI=6 NO=1773 LA f1 f2 f3 f4 f5 f6 KM FI=GRADE5-6.COR SD FI=GRADE5-6.COR MO NX=6 NK=1 LX=IN LK ICT PD OU AL SC

ABSTRACT

Evaluation of Vertical Scaling When Comparing Various Grades

- ICT Literacy Tests for Elementary Students -

Kyung-Seok Min(Assistant Professor, Sejong University)

Measuring students' growth and change has been an important issue in educational research and usually vertical scaling is a way to assess students' growth. The purpose of this study was to evaluate validity of multiple grades comparison using vertical scales of ICT Literacy Tests for elementary students developed in 2008. Multi-group structural equation models were used in order to assess equivalence of measurement models across two test forms and four grades. As a result, equivalent factor structure was confirmed but factor loadings varied across test forms. It meant that the meaning of vertically scaled scores might not be consistent when students in different grades were compared. So, it would be better to generate multiple sub-scores reflecting multidimensionality of ability areas rather than to use a single total score, even when vertical scales were used to evaluate students' growth. Suggestions for further studies on vertical scaling were also provided.

Key words : vertical scaling, ICT Literacy Tests, multi-group structural equation models.