

초등학생용 ACCENT 지능검사의 구인타당도 연구

김 동 일

서울대학교 교육학과 교수

여 승 수*

한국교육개발원 연구위원

홍 성 두

광주여자대학교 초등특수교육과 교수

《 요 약 》

이 연구의 주된 목표는 새로 개발된 ACCENT(Achievement-Cognitive ability ENdorsement Tests) 지능검사의 구인타당도를 살펴보는 것이었다. 특별히 이 검사는 WISC(Wechsler Intelligence Scale for Children) 검사에서 이론적으로 가설된 3요인 모형을 기반으로 개발되었기 때문에, 3요인 모형이 2요인 모형보다 더 우수한지에 대한 검증이 필요했다. 본 연구의 참여한 학생 수는 1,596명이며 초등학교 1학년부터 6학년까지 동등한 비율로 표집되었다. 두 모형을 통계적으로 검증하기 위해서 확인적 요인 분석이 사용되었다. 2요인 모형과 3요인 모형의 비교에서 3요인 모형의 적합도가 더 우수한 것으로 나타났다. 임상적인 측면에서 본 연구는 주의집중에 어려움이 있을 수 있는 학생들을 추가적으로 판별하는데 있어서 3요인 모형이 더욱 유용할 수 있음을 보여주고 있다. 성별에 따른 측정 동등성 측면에서 남녀간의 차이가 존재하지 않았다. 끝으로 교육학적 함의와 후속 연구에 대한 제안을 제시하였다.

주제어 : 지능검사, 구인타당도, 측정 불변성, 주의집중 장애

I. 서 론

역사적으로 지능검사는 학교생활의 성공과 학업성취의 중요한 예언변인으로 여겨져서(Joanna, 2003), 학교나 교육관련 기관에서는 지능검사를 지속적으로 사용하고 있고 이러한 이유로 인해 사람들은 지능검사란 단어에 익숙해 있다. 그러나 이들에게 지능이 무엇인지 물어 본다면 그들이 생각하는 지능의 정의는 매우 다양하여 일치된 함의를 얻는 것은 쉬운 일이 아닐 것이다. 이렇게 일반인 혹은 학자들 간에

* 교신저자(yssdreams@gmail.com)

지능의 정의를 쉽게 합의할 수 없다는 측면은 현재까지도 지능에 대한 지속적인 논의가 진행되고 있는 중요한 원인 중 한가지 일 것이다(하대현, 1997).

이런 지능에 대한 학자들 간의 다양한 정의는 지능검사의 개발과 직접적인 관련성을 갖게 되고, 결국 특수교육현장에서 다양한 이론적 배경을 가진 여러 지능검사들이 활용되고, 지능검사의 종류에 따라 전혀 다른 결과를 얻게 되는 원인이 된다. 심지어 동일한 검사에 있어서도 구인을 어떻게 볼 것인지에 따라서 전혀 다른 구조를 가진 검사로 변화될 수도 있다.

예컨대, Wechsler(1974)가 개발한 WISC(Wechsler Intelligence Scale for Children) 검사에서는 지능을 개인에게 둘러싸인 세계와 직면하고 그 세계를 이해하는 총합적인 능력으로 정의하였다. 그는 지능을 측정하기 위해 영역별 하위검사를 실시하고, 그러한 하위검사들의 총합을 바탕으로 일반적인 지적 능력(general intellectual ability)을 지능으로 정의하는 이론적 틀을 가지고 있다(곽금주 외, 2001). 그런데, 이 WISC 검사의 경우도 2요인 모형(언어성 요인과 동작성 요인)을 지지하는 연구(Gass, Demsky, & Martine, 1998, Stewart & Moley, 1983)들과 2요인 모형의 언어성 요인에서 '숫자'와 '산수'의 소검사를 따로 분리하여 주의집중 요인을 추가한 3요인 모형이 더 적합한 모형이라고 주장하는 연구(Kaufman, 1975; Hodges, 1988; Reynolds & Ford, 1994)들이 현재까지 지속적인 논쟁의 여지로 남아있다.

이는 지능검사의 결과를 특수교육 대상자를 선정하고 진단하며 배치할 때 준거자료로 활용하고(Joanna, 2003), 특수교육요구학생과 부모의 교육 권리를 보장해주는 근거자료로 이용하고 있는(Carroll, 1978) 특수교육현장에서는 중요하게 고려해야 할 사항임에 틀림없다. 따라서 지능검사를 개발하는 연구자들은 개별문항에 대한 분석과 신뢰도 및 타당도와 관련된 심리측정에 관한 연구 결과와 함께 검사를 구성하는 요인의 구조가 적합한지에 대한 평가도 함께 이루어져야 할 것이다.

본 연구는 검사의 구조에 대한 중요성을 고려하면서, 김동일 등(2004)에 의해서 새롭게 개발된 ACCENT(Achievement-Cognitive ability ENdorsement Tests) 지능검사의 구인타당도 특히 요인 모형을 평가하는 것이 주된 연구 목표이다. ACCENT 지능검사는 새롭게 개발된 검사 도구이며 이 검사가 특수교육현장에서 활용될 수 있기 위해서는 반드시 신뢰도와 타당도에 관한 정보가 제시되어야 한다. 이 중 신뢰도와 준거타당도에 관한 정보는 신현수(2004)에 의해서 이미 보고되었다. 신현수의 연구에 따르면 ACCENT 검사 점수는 WISC-III 전체 점수와 K-ABC(Kaufman Assessment Battery For Children)점수간에 높은 상관을 보였다. 그러한 연구 결과를 바탕으로 ACCENT 지능검사는 적합한 준거타당도를 확보하고 있는 것으로 판단될 수 있을 것이다. 적합한 준거 타당도와 함께 새로 개발된 연구는 구인타당도에 관한 경험적인 증거를 제공해야한다.

따라서, 본 연구의 첫 번째 연구 주제는 새롭게 개발된 ACCENT 지능검사의 요인의 수를 결정하는 구인타당도 연구이다. 이러한 구인타당도 연구와 함께 본 연구의 두 번째 연구 문제는 임상적으로도 구인타당도에서 결정된 모형이 특수교육이 필요한 아동을 위해 의미 있게 해석될 수 있는지를 조사하는 것이다. 구인타당도에서 나온 결과는 단순히 통계의 수치에 의한 결과만으로는 만족될 수 없으며, 실제로 교육현장에서 검사를 사용하는 교사들에게 아이들을 평가하고 교수방법을 계획하는데 있어서 의미 있는 정보를 제공해야한다.

끝으로 구인타당도에서 증명된 검사의 요인들이 성별 간에 동등하게 영향을 미치고 있는지를 밝히는 측정 불변성(measurement invariance)을 검증하는 것이었다. 실제로 지능연구자들은 오랜 기간 동안 지능검사가 집단에 따라 동일한 결과를 제공하는지에 대한 의문을 갖고 있었으며 이러한 문제의 해답을 찾기 위해서 측정 불변성에 관심을 가져왔다(배미란, 1998). 만약 성별 간에 유의한 차이가 발생한다면 검사의 결과해석에 있어서 추가적인 주의가 필요하게 될 것이다. 따라서 본 연구에서 지능검사가 성별 간에 측정 불변성이 유지되고 있는지에 대한 검증을 실시하였다.

1. 연구 문제

- 1) 새롭게 개발된 ACCENT 지능검사의 구인타당도는 2요인 모형과 3요인 모형 중 어떠한 모형이 더 적합한가?
- 2) 구인타당도 분석에서 확정된 검사의 모형은 특수교육 요구 아동을 선별하는 데 있어서 의미 있게 해석되고 활용될 수 있는가?
- 3) 구인타당도 연구에서 확정된 검사의 모형은 성별간의 측정 불변성이 유지되고 있는가?

II. 연구 방법

1. 연구 참여자

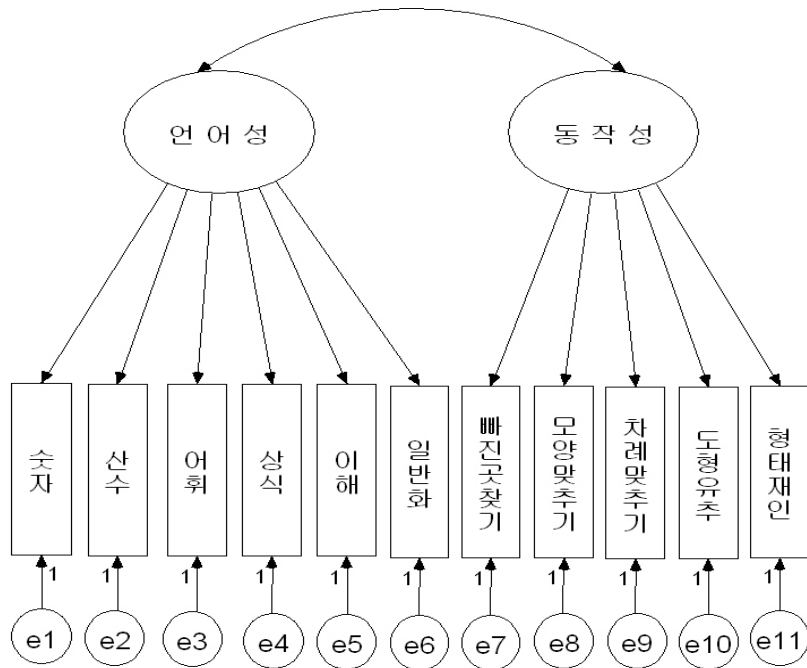
본 연구의 대상은 ACCENT 지능검사의 규준을 만들기 위해서 2003년 5월에 전국 11개의 초등학교에서 표집된 학생들의 자료를 사용하였다. 참여한 학교들은 경기, 인천 충청도, 전라도 지역을 중심으로 선별되었다. 전체 학생 수는 1,726명이며, 초등학교 1학년부터 6학년까지 동등한 비율로 표집 되었다. 그러나 본 검사에서는

listwise 방법을 사용하여 결측값이 있는 학생들은 제외되었으며, 결국 1,596명의 학생이 최종적으로 분석에서 사용되었다. 성별 비율은 남학생이 52%, 여학생이 48%이다. 지역별 학교분포는 대도시가 52.2%, 농어촌이 28.6%, 중소도시가 18.9%를 차지하였다.

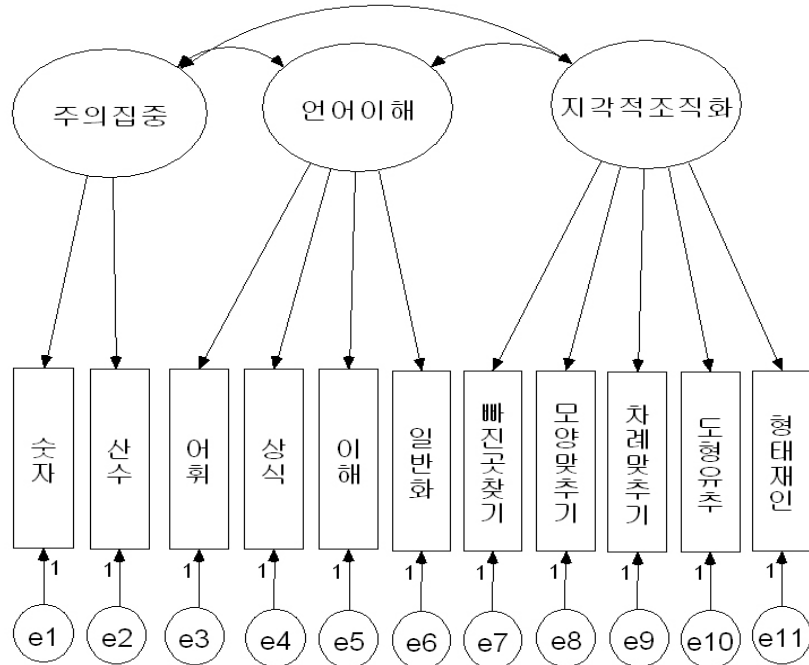
2. 연구 절차 및 연구 방법

1) 모형설정

ACCENT 지능검사의 요인 구조를 확인하기 위해서 2개의 비교 모형을 설정하였다. 먼저 첫 번째 모형은 2요인 모형이며 두 번째 모형은 본 검사의 이론적 배경이 되는 3요인 모형이다. 2요인 모형은 <그림 1>이고 3요인 모형은 <그림 2>에 제시하였다.



<그림 1> 2요인 모형



<그림 2> 3요인 모형

2요인 모형은 ACCENT 지능검사의 11개 소검사를 크게 언어성 요인과 동작성 요인으로 나눈 것이다. 즉 2요인 모형의 요인 구조는 WISC-R과 WISC-III의 기본 모형이다. 3요인 모형은 11개의 소검사를 크게 3요인으로 나눈 것으로 2요인과 비교해보면, 언어성 요인에서만 주의집중 요인과 언어이해 요인으로 분리된 모형이다. 바로 이 3요인 모형이 본 검사가 가정하고 있는 모형이다.

2) 자료 점검

먼저 각각의 소검사들의 일변량 정규성을 검정하였다. 각 변수의 일변량 정규성 검정을 위해서 왜도(skewness)와 첨도(Kurtosis)를 산출하였다. 왜도와 첨도는 일반적으로 +1과 -1 사이의 값을 가질 수 있으며 0에 가까울수록 정규분포로 근접하고 있다고 해석할 수 있다(Mertler & Vannatta, 2000).

왜도와 첨도에 대한 정보와 함께, 정규분포에 대한 기본 가정을 검사하기 위해서 Kolmogorov-Smirnov을 사용했다. 결론적으로 Kolmogorov-Smirnov 검정결과 영가설을 기각하였다. 다시 말해서, 11개의 하위 요인은 정규분포를 따르지 않는 것으로 나타났다. 다변량 정규성 검정은 Scatterplot Matrix를 사용하여 점검하였

다. 점검결과 다변량 정규분포가 아닌 것으로 나타났다. 자료의 이상점을 분석하기 위해서 SPSS 10.0의 BOXPLOT을 사용한 결과 특이한 이상점은 발견하지 못했다.

마지막으로 자료의 결측치에 대한 분석을 하였다. 결측치는 무작위 결측치(missing at random)와 체계적 결측치(systematic missing)로 구분될 수 있는데(배병렬, 2002) 본 자료에서 체계적 결측치로 판단할 수 있는 뚜렷한 근거가 없었다. 따라서 무작위 결측치로 판단되어 결측치를 처리하는 방법은 listwise 방법을 사용하였다. 채택된 자료는 각 변수간 공분산 행렬을 사용하여 분석하였다. 소검사의 공분산 행렬은 <표 1>과 같다.

<표 1> 소검사간의 공분산 행렬

	숫자	산수	어휘	상식	일반화	이해	모양 맞추기	차레 맞추기	형태 재인	빠진 곳 찾기	도형 유추
숫자	1.92										
산수	1.39	3.79									
어휘	1.48	2.82	5.91								
상식	.78	1.39	1.89	2.80							
일반화	1.12	2.28	3.18	1.63	5.85						
이해	.39	.74	1.11	.60	1.09	.98					
모양 맞추기	.35	.71	.80	.45	.77	.28	.86				
차레 맞추기	.32	.58	.74	.40	.62	.22	.22	.56			
형태재인	.53	.98	1.24	.76	1.16	.37	.39	.32	1.68		
빠진 곳 찾기	.66	1.12	1.45	.71	1.22	.40	.39	.38	.68	1.73	
도형유추	1.13	1.92	2.41	1.14	2.00	.79	.80	.67	1.00	1.15	4.06

<표 2>는 학년별에 대한 ACCENT 검사의 하위 요인의 기술 통계치를 보고하고 있다.

〈표 2〉 ACCENT 지능검사의 하위 요인에 대한 평균 및 표준편차

학 년 (n)	숫자		산수		어휘		상식		일반화		이해		모양 맞추기		차례 맞추기		형태 제인		빠진 곳 찾기		도형 유추	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1학년 (198)	2.06	1.69	1.57	1.22	3.36	1.48	2.17	1.27	2.87	1.75	.78	.85	1.64	.88	.47	.65	2.05	1.21	1.27	1.15	2.23	1.65
2학년 (246)	2.17	1.49	2.17	1.22	4.11	1.61	4.11	1.61	2.50	1.45	3.30	1.80	1.98	.83	1.02	.73	2.39	1.26	2.23	1.22	3.60	1.77
3학년 (280)	2.86	1.01	3.11	1.29	5.09	1.93	3.27	1.47	3.67	1.71	1.12	.76	2.23	.85	1.24	.73	2.99	1.11	2.76	1.11	2.76	1.11
4학년 (288)	3.39	1.01	4.07	1.56	6.15	2.07	3.50	1.56	4.17	1.97	1.47	.96	2.50	.80	1.44	.64	3.20	1.13	3.13	1.06	4.94	1.64
5학년 (280)	3.56	1.09	4.71	1.63	7.11	1.96	4.27	1.52	5.35	2.38	1.70	1.00	2.85	.83	1.60	.56	3.82	1.09	3.23	1.09	5.55	1.51
6학년 (304)	3.91	1.01	5.39	1.62	7.84	2.00	4.22	1.66	6.35	2.61	2.00	.87	2.83	.81	1.65	.57	3.79	1.05	3.50	1.09	5.94	1.47
전 체 (1,596)	3.08	1.38	3.67	1.95	5.80	2.43	3.41	1.69	4.41	2.41	1.37	.99	2.39	.93	1.29	.74	3.12	1.30	2.79	1.32	4.61	2.02

3) 확인적 요인 분석을 사용한 모형의 적합도 검증

본 연구에서 제시된 모형의 적합도를 구하기 위해서 확인적 요인 분석을 사용하였다. 확인적 요인 분석은 일반적인 요인 분석(factor analysis)의 특별한 유형으로서, 주된 사용의 목적은 검사개발단계에서 잠재된 요인의 수를 확정하기 위해서 사용된다(Kline, 2005). 확인적 요인 분석을 하기 위해서 사용된 프로그램은 AMOS 4.0이었다. 모형의 모수치를 추정하는 방법은 가중최소자승추정방법(Weighted Least Square estimation)이 사용되었는데, 그 이유는 다변량 정규성 가설을 만족하지 않았기 때문이다(손원숙, 2002).

모형의 적합도는 2가지로 구분된다. 내재모형(nested model)을 비교할 때 사용되는 적합도 지수로 χ^2 , AGFI(Adjusted GFI), RMSR(Root Mean Square Residual), CFI(Comparitive Fit Index)가 사용되었으며, 비내재모형(non-nested model)을 비교하기 위해서 RMSEA(Root Mean Square Error of Approximation), AIC(Akaike Information Criterion), CAIC(Consistent AIC)값을 이용하여 검증하였다.

4) 성별에 따른 집단의 측정불변성(measurement invariance) 검증

집단 간 측정 불변성 검증은 다집단 확인적 요인 분석을 사용하여 실시하였다. 설현수(2003)에 의하면 이 방법은 4가지 단계로 구분되어 진다. 첫 번째 단계는 구조 동등성(configural invariance) 검증이다. 이 단계는 동일한 측정변수가 동일한 요인에 부하가 되고 있는지를 검증하는 것이다. 구조 동등성이 확인되면 측정 동등성을 검증할 수 있다. 이 단계는 집단간 요인 계수가 동일한지를 검증하는 것이다. 이 방법은 아무런 제한을 두지 않은 기저모형의 χ^2 과 측정동일성의 χ^2 의 차이를 산출하여 동일성이 유지되는지를 판단할 수 있다. 세 번째로는 요인 분산/공분산 동일성(factor variance/covariance invariance) 검증이고 마지막 단계는 오차변수의 분산 동일성(error variance invariance) 검증이다.

III. 연구 결과

1. 통계적 준거에 의한 모형간 비교

ACCENT 지능검사의 요인 구조를 WISC 검사의 2요인 모형과 3요인 모형으로 구성하였고, 확인적 요인 분석을 통해 모형별 적합도를 산출하였다.

<표 3> 2요인 모형과 3요인 모형의 적합도 비교

학 년(n)	2요인 모형						3요인 모형					
	χ^2	적합도 지수					χ^2	적합도 지수				
		df	χ^2/df	AGFI	RMR	CFI		df	χ^2/df	AGFI	RMR	CFI
1학년(198)	105.57*	43	2.45	0.85	0.16	0.90	48.13	41	1.17	0.93	0.07	0.95
2학년(246)	52.28	43	1.21	0.94	0.08	0.96	48.18	41	1.17	0.94	0.07	0.97
3학년(280)	60.02*	43	1.39	0.94	0.07	0.91	55.35	41	1.35	0.94	0.07	0.93
4학년(288)	48.34	43	1.12	0.95	0.06	0.98	45.72	41	1.11	0.95	0.06	0.98
5학년(280)	80.93*	43	1.88	0.92	0.08	0.90	72.11*	41	1.75	0.92	0.07	0.92
6학년(304)	80.56*	43	1.87	0.93	0.09	0.92	73.72*	41	1.79	0.93	0.09	0.94
전체(1,596)	164.09*	43	3.81	0.97	0.06	0.97	106.70*	41	2.60	0.98	0.04	0.98

*p<.05

<표 3>의 결과를 살펴보면, 2요인 모형과 3요인 모형의 적합도는 전반적으로 양호한 것으로 나타났다. 그러나 두 모형 간 비교하여 판단해 보면 3요인 모형이 전반적으로 2요인 모형보다 양호한 적합도 지수를 나타냈다. 특히 1학년 모형에서는 3요인 모형이 2요인 모형보다 더욱 월등히 높게 나타났다.

모형 간 적합도를 비교하기 위해 <표 4>에 제시된 적합도 지수를 추가적으로 산출하였다.

<표 4> 요인 모형간 적합도 비교표
(전체학년 n=1,596)

모형	x^2	적합도			
		x^2/df	AIC	CAIC	RMSEA
2요인 모형	164.09*	3.81	210.09	356.72	0.04
3요인 모형	106.70*	2.60	156.70	316.08	0.03
비교결과	◇	◇	◇	◇	◇

*p<.05,

◆ 2요인 모형의 적합도 지수가 더 적합한 경우,

◇ 3요인 모형의 적합도 지수가 더 적합한 경우

2요인 모형과 비교했을 때, 3요인 모형은 <표 4>에서 제공된 적합도 지수에서도 더 뛰어난 결과를 산출했다. <표 3>과 <표 4>에서 제공된 결과를 종합해 보면, 3요인 모형이 더 뛰어난 적합도 지수를 제공하고 있다고 판단할 수 있다.

<표 5> 2요인 모형과 3요인 모형의 표준화 계수

학 년 하위검사	2요인 모형							3요인 모형							
	1 학년	2 학년	3 학년	4 학년	5 학년	6 학년	전 체	1 학년	2 학년	3 학년	4 학년	5 학년	6 학년	전 체	
숫자	0.67	0.42	0.20	0.42	0.45	0.43	0.58	0.99	0.48	0.33	0.47	0.54	0.51	0.62	
산수	0.62	0.47	0.30	0.50	0.54	0.57	0.76	0.46	0.55	0.47	0.56	0.67	0.69	0.82	
어휘	0.20	0.51	0.56	0.67	0.64	0.64	0.78	0.34	0.55	0.55	0.68	0.65	0.65	0.79	
상식	0.14	0.40	0.43	0.45	0.42	0.51	0.58	0.03	0.41	0.44	0.46	0.42	0.52	0.58	
이해	0.17	0.24	0.38	0.45	0.52	0.52	0.57	0.56	0.26	0.41	0.47	0.53	0.54	0.59	
일반화	0.13	0.46	0.48	0.52	0.67	0.68	0.67	0.63	0.50	0.52	0.53	0.69	0.69	0.68	
빠진 곳 찾기	0.34	0.46	0.48	0.36	0.31	0.59	0.63	0.38	0.46	0.49	0.36	0.31	0.60	0.64	
모양	0.31	0.18	0.38	0.36	0.32	0.51	0.55	0.31	0.18	0.38	0.37	0.32	0.51	0.56	
차례 맞추기	0.39	0.56	0.39	0.25	0.41	0.38	0.59	0.36	0.59	0.40	0.26	0.42	0.39	0.60	
도형유추	0.60	0.60	0.41	0.66	0.46	0.42	0.71	0.57	0.61	0.41	0.66	0.46	0.41	0.71	
형태재인	0.37	0.48	0.39	0.27	0.30	0.34	0.57	0.42	0.48	0.40	0.27	0.30	0.39	0.58	
요인 간 상관								a	0.03	0.73	0.51	0.83	0.75	0.80	0.87
								b	0.31	0.67	0.57	0.69	0.74	0.63	0.84
								c	0.62	0.66	0.60	0.68	0.91	0.78	0.87

a=주의집중과 언어이해의 상관,
b=주의집중과 지각적 조직화의 상관,
c=언어이해와 지각적 조직화의 상관

두 모형간 표준화 계수는 <표 5>에 제시하였다. 두 모형의 표준화계수는 저학년에서 고학년으로 올라갈수록 전반적으로 증가하고 있다. 주목할 사항은 3요인 모형에서 ‘숫자’와 ‘산수’ 영역의 소검사 점수가 모형간 상대적으로 큰 차이를 나타내고 있다는 것이다. 이러한 결과는 ‘숫자’와 ‘산수’ 영역을 따로 분리한 3요인 모형이 경험적 자료에 더 적합한 모형이라고 판단할 수 있다.

2. 임상적 판별 준거에 의한 모형간 비교

1) 표준점수 80점을 기준으로 사용한 방법

특수교육 요구 학생을 선별하기 위해서 다음과 같은 방법을 사용하였다. 먼저 언어성 요인의 점수 합을 표준점수로 변환한 후에 경계선급(at risk) 이상(80점 이상)의 점수를 받은 학생들을 선별하였다. 80점 이상을 받은 학생들의 언어성 요인에 대한 기술 통계치는 다음과 같다.

<표 6> 언어성 요인의 기술 통계치

학 년(n)	원점수		표준점수	
	M	SD	M	SD
1(240)	12.53	4.04	102.81	12.83
2(246)	15.55	4.42	102.51	13.44
3(265)	19.69	4.22	102.51	13.14
4(268)	23.59	5.18	102.32	13.30
5(263)	27.52	5.66	102.59	12.50
6(280)	30.89	5.51	102.83	12.01
총 합 (1,562)	21.93	8.06	102.60	12.85

<표 7> 주의집중 요인의 기술 통계치

학 년(n)	원점수		표준점수	
	M	SD	M	SD
1(17)	0.00	0.00	77.73	0.00
2(16)	0.69	0.48	74.97	3.31
3(13)	2.69	0.85	72.47	7.25
4(13)	3.69	0.63	72.76	4.56
5(11)	4.45	0.82	74.99	5.39
6(13)	5.92	0.27	77.43	1.86
총 합 (83)	4.58	2.19	75.19	2.65

선별된 학생 중 언어성 요인에서 주의집중 요인인 ‘산수’와 ‘숫자’ 영역을 분리하여 표준점수로 환산하여 점수를 산출하였다. 언어성 요인에 대한 원점수와 표준점수는 <표 6>에 나타나 있다. 선별된 학생 중 주의집중 요인의 표준점수가 80 이하의 점수를 받은 학생의 기술 통계치는 <표 7>과 같다. 경계선급 이상의 1,562명의 학생 중에서 주의집중에 문제가 있을 것이라고 예상되는 83명의 학생이 선별되었다. 선별된 학생들은 학년별로 11명 이상이였다. 이러한 선별은 2요인 모형에서는 확인될 수 없는 사항으로 3요인으로 지능검사가 개발되어야만 교사나 연구자들은 교육적으로 더 의미 있는 정보를 지능검사로 얻을 수 있음을 보여주고 있다.

2) 요인 간 점수차의 백분율을 사용한 방법

요인 간 점수차의 백분율을 사용하는 방법은 K-WAIS의 3요인(주의집중 요인)을 탐색하기 위해서 사용한 김홍근(2002)의 연구 방법으로서 본 검사 도구에서도

동일하게 사용하였다. 앞서서 사용한 통계방법과 가장 큰 차이점은 80점을 기준으로 아이들을 선정한 것이 아니며 요인들간 점수 차이에 의해서 아동을 선별하게 된다. 요인 모형간 점수차를 구하는 공식은 다음과 같다.

2요인 모형은 점수차이 = | 언어성 요인 - 동작성 요인 | 의 공식을 이용하였고, 3요인 모형에서는 점수차이 = max { | 언어이해(VC) - 주의집중(FD) | , | 언어이해(VC) - 지각적조직화(PO) | , | 주의집중(FD) - 지각적조직화(PO) | } 를 사용하였다.

먼저 2요인 모형에서 요인 간 점수차의 백분율을 구하면 <표 8>과 같다.

<표 8> 2요인 간 점수차의 기준

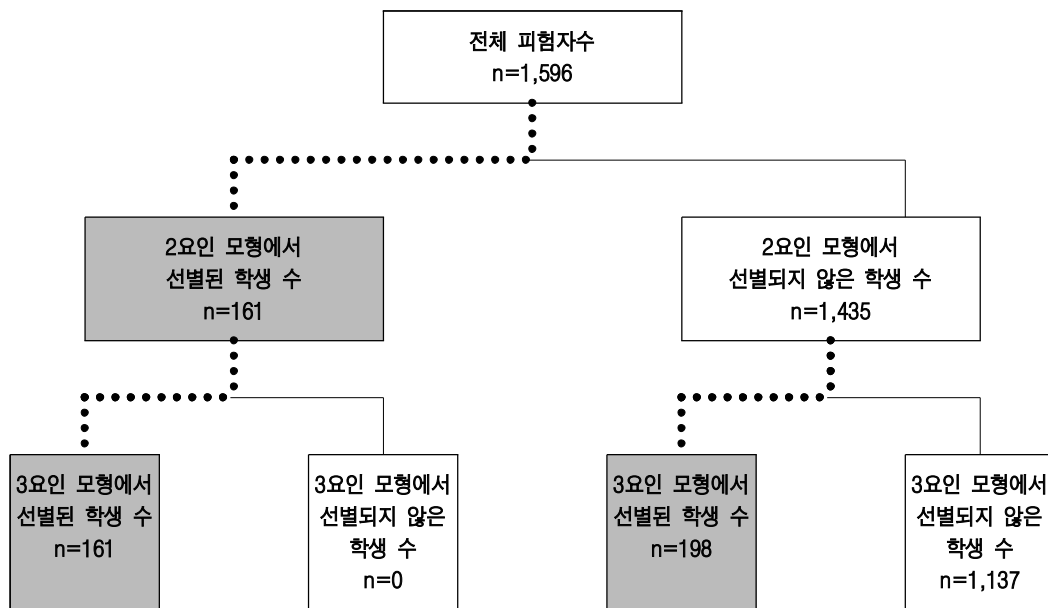
2요인 간 차이 (절대값)	백분율	2요인간 차이 (절대값)	백분율
≥58	0.0%	26	13.7%
53	0.2%	25	15.1%
51	0.4%	24	16.3%
50	0.4%	23	18.5%
49	0.5%	22	19.8%
48	0.5%	21	21.3%
47	0.7%	20	22.9%
46	0.7%	19	25.1%
45	0.8%	18	28.1%
44	1.1%	17	29.6%
43	1.2%	16	34.6%
42	1.5%	15	39.1%
41	1.7%	14	43.3%
40	2.1%	13	46.3%
39	2.3%	12	49.3%
38	2.4%	11	52.9%
37	3.0%	10	58.0%
36	3.4%	9	60.8%
35	4.5%	8	64.7%
34	5.2%	7	71.0%
33	5.2%	6	76.5%
32	6.3%	5	81.1%
31	7.0%	4	83.9%
30	8.7%	3	88.4%
29	9.1%	2	94.1%
28	10.6%	1	97.0%
27	11.6%	0	100.0%

<표 9> 3요인 간 점수차의 기준

3요인 간의 차이 (절대값)	백분율	3요인간의 차이 (절대값)	백분율
≥58	0%	27	10.5%
54	0.1%	26	12.0%
53	0.2%	25	14.4%
52	0.2%	24	16.2%
51	0.2%	23	18.2%
50	0.3%	22	20.2%
49	0.4%	21	22.1%
48	0.4%	20	23.6%
47	0.5%	19	26.8%
46	0.6%	18	28.8%
45	0.7%	17	31.7%
44	0.9%	16	33.6%
43	1.0%	15	37.3%
42	1.2%	14	41.5%
41	1.4%	13	44.2%
40	1.7%	12	47.6%
39	2.0%	11	51.1%
38	2.2%	10	55.5%
37	2.4%	9	59.8%
36	2.8%	8	63.8%
35	3.4%	7	68.1%
34	3.9%	6	74.3%
33	4.4%	5	79.3%
32	5.3%	4	83.6%
31	5.8%	3	87.0%
30	7.1%	2	94.0%
29	7.5%	1	97.5%
28	9.1%	0	100.0%

임상적으로 큰 의미가 있는 10% 미만의 백분위 기준(김홍근, 2002)에서 분석해보면, 10% 미만의 점수는 29점 이상이였으며 전체 피험자 중 161명의 학생이 포함되었다. 동일한 방법으로 3요인 간 점수차를 <표 9>에서 제시하였다. 요인 간 차이 점수에서 임상적으로 의미 있는 10% 미만의 점수는 28점 이상이었고, 총 359명의 학생들이 포함되었다. 2요인 모형으로 사용했을 때 보다 198명의 학생들이 추가적으로 선별되었다.

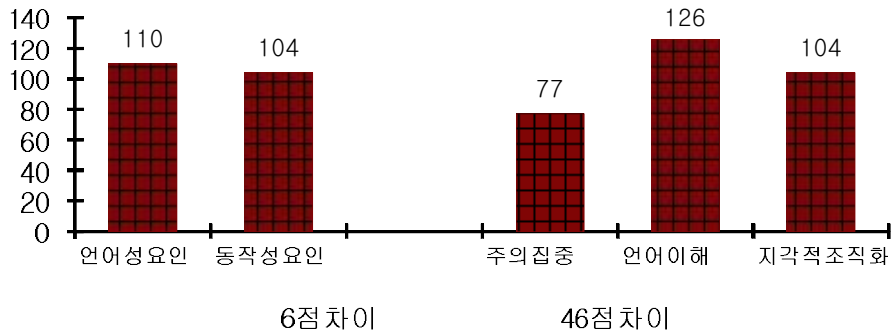
2요인 모형과 3요인 모형의 전체적인 비교는 <그림 3>과 같다. 2요인 모형간의 10% 미만의 준거에 의해서 선별된 학생은 161명이였으며 2요인 모형에서 선별된 161명은 모두 3요인 모형에서도 선별되었다. 반면, 2요인 모형에서 선별되지 않은 1,435명중에서 3요인 모형에서 선별된 학생은 198명이었고, 나머지 1,137명은 3요인 모형에서도 선별되지 않았다.



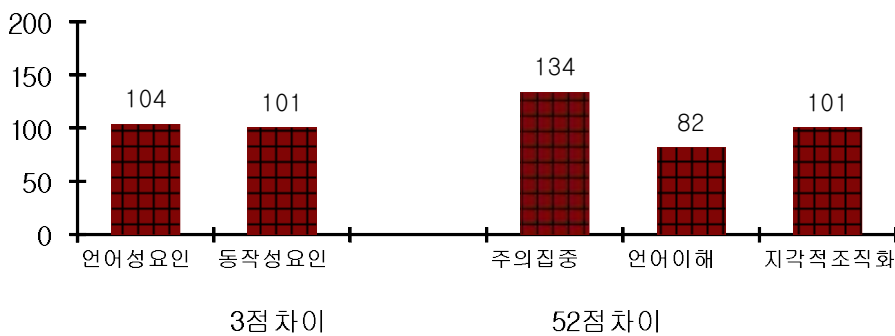
※점선은 유의한 차이가 있는 경로를 의미함

<그림 3> 2요인과 3요인 간 분석 비교

<그림 4>는 3요인 모형이 더 유용한 개별학생의 검사 프로파일의 예이다. 이 학생은 2요인 모형에서는 언어성 요인이 110점, 동작성 요인이 104점으로 정상아동으로 판정될 수 있다. 그러나 3요인 모형으로 이 학생의 점수를 해석하면 이 학생은 높은 언어이해 점수 때문에 낮은 주의집중영역의 점수가 2요인 모형에서는 정상으로 판정될 수 있다. 따라서 2요인 모형에서 미리 조기 발견할 수 없었던 주의집중에 어려움이 있을 수 있는 학생들이 3요인 모형에서는 쉽게 선별할 수 있음을 보여주고 있다.



<그림 4> 2요인 모형보다 3요인 모형이 유용한 특정 사례 1



<그림 5> 2요인 모형보다 3요인 모형이 유용한 특정 사례 2

<그림 5>는 <그림 4>와는 정반대의 상황이다. 이 학생의 경우에도 마찬가지로 언어성 요인과 동작성 요인에서는 평균점수를 얻었다. 그러나 3요인으로 분리하여 프로파일을 해석해보면, 높은 주의집중점수 때문에 낮은 언어이해점수를 보완해주고 있다. 따라서 이 학생은 언어이해의 점수는 매우 낮지만 뛰어난 주의집중점수 때문에 정상적인 범주의 언어성 요인의 점수를 획득하였다. 결론적으로 3요인 모형의 사용은 교사들에게 지나치게 높은 주의력을 가지고 있지만 낮은 언어이해 능력을 가지고 있는 학생을 선별할 수 있는 정보를 제공하게 된다.

3. 성별에 따른 측정불변성(measurement invariance) 검증

마지막으로, 3요인 모형이 남학생과 여학생 간 측정 불변성을 지니고 있는지 확인하기 위해서 다집단(multi-sample) 확인적 요인 분석을 실시하였다. 이전에 설명된 것처럼, 4가지 단계의 동등성 검증이 실시되었다.

<표 10> 성별 측정불변성 검증결과

모형	χ^2	적합도				χ^2	
		df	AGFI	RMR	RMSEA	$\Delta\chi^2$	Δdf
모형 1	149.86 ***	82	0.97	0.05	0.02		
모형 2	163.14 ***	90	0.97	0.07	0.02	13.28	8
모형 3	166.23***	93	0.97	0.09	0.02	3.09	3
모형 4	183.85***	104	0.97	0.10	0.02	17.62	11

***p<.001

모형 1은 아무런 제약을 가하지 않은 모형, 모형 2는 요인 계수에 의한 요인 동등성 제약이 가해진 모형, 모형 3은 공분산 동일성의 제약이 가해진 모형, 모형 4는 오차 분산의 제약이 가해진 모형

$\Delta\chi^2$ 는 이전 모형과 본 모형의 χ^2 값의 차이

Δdf 는 측정 동등성 자유도에서 구조 동등성 자유도를 뺀 값

<표 10>에 따르면, 아무런 제약을 가하지 않은 모형 1에서 χ^2 는 149.86으로 집단간 모형에 차이가 없다는 가설을 기각시켰다. 그러나 χ^2 값은 사례수에 영향을 받는 문제점이 발생하기 때문에 다른 적합도의 값을 함께 고려해야한다. AGFI는 0.97이고, RMR는 0.05, RMSEA는 0.02로 전반적으로 적합도가 양호하였다. 따라서 성별에 따른 집단간 모형검증에서 구조 동등성이 성립되었다고 할 수 있다.

구조 동등성이 성립이 되었기 때문에 다음 단계로는 요인 계수의 동일화 제약을 부여한 측정동등성을 검증하였다. 측정 동등성은 요인 계수에 대한 동등화 제약이 설정된 모형이 제약이 없는 모형에 비해서 얼마나 x^2 이 좋지 않은지를 기준으로 평가한다(설현수, 2002). 제약을 가하지 않은 모형 1의 x^2 의 값은 149.86이고 자유도는 82이다. 이 값을 기저모형으로 설정하고 모형 2의 x^2 값(163.14)과의 차이를 구하면 유의수준 .05에서 통계적으로 유의하지 않다고 나타났다. 즉 ‘두 집단 간 요인 계수가 동일하다’라는 영가설을 기각하지 못했기 때문에 집단간 측정동등성이 성립된다고 판단할 수 있다.

세 번째 단계로 공분산 동등성(factor variance/covariance invariance)을 검증하였다. <표 10>에서 나온 결과에 따르면, 남녀차이에 따른 모형의 공분산은 동일한 것으로 나타났다. 3요인 모형의 공분산은 두 집단간에 통계적으로 유의한 차이가 나타나지 않았다. 마지막으로 오차변수의 분산 동일성(error variance invariance)이 검증되었는데, 이 단계에서도 마찬가지로 남녀 차이는 나타나지 않았다.

IV. 결론 및 논의

1. 연구 요약 및 의의

본 연구는 새롭게 개발된 ACCENT 지능검사의 구인타당도를 구하는 것을 목적으로 하고 있으며 WISC 검사에서 연구되고 있는 구인타당도의 관점을 바탕으로 분석되었다. 연구 결과를 살펴보면 다음과 같다.

첫째, 2요인 모형과 3요인 모형에 대한 적합도 차이를 비교하였다. 분석결과 2요인 모형과 3요인 모형의 적합도 모두 양호한 결과를 산출하였다. 전통적으로 인정받고 있는 2요인 모형이 본 연구에서 획득된 자료와 심하게 부합된다고 판단할 수는 없지만, 전반적으로 2요인 보다는 3요인 모형의 적합도가 조금 더 우수했다.

둘째, 모형들간의 표준화 계수를 비교하였다. 비교 결과 2요인 모형의 표준화 계수보다는 3요인 모형의 표준화 계수가 전반적으로 더 높은 수치를 나타냈다. 특히 소검사인 ‘산수’와 ‘수학’ 영역의 표준화 계수에 가장 큰 차이를 나타냈다. 즉 2요인 모형에서의 두 소검사(산수와 수학)의 표준화 계수는 3요인 모형에서 더 높은 수치를 나타냈다. 이러한 분석 결과로 판단 할 수 있는 것은 주의집중 요인에 관한 추가적인 정보를 얻기 위해서는 2요인 모형보다는 3요인 모형으로 검사의 요인이 구성되어야 함을 보여주고 있다.

셋째, 두 모형을 임상적인 결과의 해석 및 예측 정보 측면에서 비교하였다. 3요인 모형을 설정하여 주의집중에 문제가 있을 것이라고 예측되는 아동을 추출하였다. 대상이 되는 학생은 2요인 모형에서 표준점수로 환산했을 때 언어성 요인에서 80점 이상의 점수를 획득했지만 3요인 모형으로 분류했을 때 주위집중 요인에서 80 이하의 점수를 획득한 학생이다. 이와 같은 준거로 분석한 결과 비록 2요인 모형에서 선별되지 않았지만 3요인 모형에서 선별된 아동들은 각 학년별로 10명 이상이었고 전체 1,562명중 83명(6%)이었다. 또한 요인 간 점수차의 백분율을 이용한 방법을 사용하여 특수교육이 필요한 아동을 선별하였다. 분석 결과 2요인 모형에서는 요인 점수간 차이가 나타나지 않은 학생들이 3요인 모형에서는 198명이 추가적으로 선별되었다. 이러한 임상 결과는 3요인 모형의 적합도가 양호 할 뿐만 아니라 임상적으로도 3요인 모형이 더 의미 있는 모형임을 보여주고 있다.

또한 3요인 모형의 타당성을 밝히기 위해서 성별에 따른 측정 불변성을 검증하였다. 검증결과 3요인 모형은 성별에 따른 요인 구조 및 계수, 공분산, 및 측정오차의 차이가 없는 것으로 나타났다. 이러한 결과로 판단해 보면 3요인 모형을 사용할 때에는 성별에 따른 점수의 습득 및 해석에서 차이가 나타나지 않는다고 판단할 수 있다.

이와 같은 연구 결과가 특수교육 현장이나 지능검사 도구를 사용하는 교육전문가들에게 주는 시사점은 다음과 같다.

첫째, 본 연구 결과의 의의는 새롭게 개발된 ACCENT 지능검사의 구인 타당도에 관한 정보를 제공하게 된다. 검사가 개발되고 매뉴얼을 제작하게 되면 본 연구에서 분석한 모형의 적합도 값이 구인타당도의 자료로 사용될 것이다.

둘째, 모형의 적합도 비교 결과 특히 1학년에서 모형간 적합도 지수에서 가장 큰 차이가 나타났다. 2요인 모형에서 1학년의 적합도는 전반적으로 적합한 수치를 보이지 않았다. 그러나 동일한 1학년 자료를 3요인 모형으로 분석한 결과를 살펴보면 전반적으로 2요인 모형보다 적합도가 크게 상승한 것으로 나타났다. 이러한 결과가 의미하는 것은 1학년과 같이 난이도가 문제가 될 수 있는 검사 상황에서는 2요인 모형보다는 3요인 모형이 더 적합한 것으로 나타났다. 따라서 학습부진아동이나 학습장애 아동을 선별하기 위한 검사 모형으로는 3요인 모형이 더 타당할 것이라는 시사점을 제공해주고 있다.

셋째, 3요인 모형은 임상적으로도 중요한 의미가 있음을 밝혀냈다. 미국에서는 이미 3요인 모형에 대한 많은 연구들이 진행되고 있으며, 3요인 모형에 대한 임상적인 결과를 발표하고 있다. 그러나 우리나라에서는 3요인 모형에 대한 관심이 부족하며 임상적인 결과를 제공하는 연구가 부족한 상황이다(김홍근, 2002). 따라서 본 연구는 한국에서 주의집중요인에 대한 임상적인 연구 결과를 제공하고 있으며, 실제로 3요인 모형의 교육적 유용성을 보여주고 있다. 특히 이러한 결과는 검사 도구를 사

용하는 특수교사나 검사전문가들에게 중요한 관점을 제공한다. 특수교육에서 아동을 조기선별하게 된다면 교육적 효과가 뛰어나며 교육비용의 절감 효과를 얻을 수 있기 때문에 3요인 모형은 특수교육현장에서 중요한 정보를 제공할 수 있는 틀이 될 것이다.

넷째, 본 연구의 결과는 차후에 개발되는 지능검사에 중요한 정보를 제공할 수 있다. 한국에서 개발된 지능검사들을 보면 대부분이 2요인 모형으로 구성되어 있다. 2요인 모형으로 검사가 제작되면 규준도 2요인 모형에 근거하여 제시된다. 따라서 3요인 모형으로 프로파일을 해석하고 싶어도 할 수 없는 상황이다. 따라서 본 연구의 결과를 바탕으로 이후에 개발되는 지능검사들은 2요인 모형뿐만 아니라 3요인 모형도 고려해야 함을 제언하고 있다.

2. 연구의 제한점 및 제언

비록 본 연구는 특수교육분야에서 중요한 정보를 제공할 수 있지만 다음과 같은 연구의 제한점과 함께 해석되고 현장에서 적용되어야 한다.

첫째, 본 연구에 3요인 모형으로 제작된 지능검사의 사용은 특수교육 요구 아동을 선별하기 위한 더 많은 정보를 교사들에게 제공할 수 있음을 보여주고 있다. 그러나 실제로 특수교육이 필요한 학생으로 진단 받은 학생들의 점수를 비교하지는 못했다. 그러므로 특수교육이 요구되는 아동들과 일반아동 집단을 함께 표집 하여 3요인 모형의 프로파일이 두 집단간 유의한 차이를 보이는지에 대한 추가적인 연구가 필요할 것이다.

둘째, 검사의 타당도에 대한 정보는 검사 자체에 대한 불변의 특성이 아니며, 다양한 교육환경과 연구 대상에 따라 다양한 타당도 결과가 도출될 수 있음을 인식하는 것은 매우 중요하다. 이러한 타당도의 관점은 Schmidt 와 Hunter(1977)가 제안한 타당도 일반화 가능성도(Validity Generalization)이론으로 설명될 수 있다. 즉 그들의 이론을 부연해서 설명하면, 메타분석(meta analysis)적 방법을 사용하여 ACCENT 검사의 타당도 검사를 오랜 시간동안 수집하여 종합된 결론을 내릴 필요성이 있음을 보여주고 있다. 따라서 본 연구와 유사한 연구들이 차후에 지속적으로 진행되어야 하며, 충분한 타당도 연구들이 축적되었을 때 타당도 일반화 가능성도 연구 방법을 사용하여 검사의 타당도에 관한 종합적 평가가 이루어질 수 있을 것이다.

참고문헌

- 곽금주, 박혜원, 김청택 (2001). **WISC-III 지침서**. 서울: 도서출판 특수교육.
- 김동일, 신중호, 여상인, 이대식, 이재희 (2004). **ACCENT 개발 최종보고서**. 서울대학교 교육연구소.
- 김홍근 (2002). K-WAIS의 3요인 해석을 위한 기준 연구. **한국심리학회: 임상**, 21(3), 631-645.
- 배미란 (1998). 인지과정 측정을 위한 Pass모형의 타당성 연구. 박사학위 논문, 연세대학교.
- 배병렬 (2002). **구조방정식 모델: 이해와 활용**. 대전: 대경 출판사.
- 설현수 (2003). 구조방정식 모형에서 χ^2 검증의 문제점 및 대안탐색. **교육평가연구**, 16(1), 105-123.
- 손원숙 (2002). 성별에 따른 “수학에 대한 태도”의 차원성 연구: 다집단 확인적 요인 분석의 적용. **교육심리연구**, 16(3), 235-253.
- 신현수 (2004). ACCENT 지능검사의 타당화 연구. 석사학위 논문, 서울대학교.
- 하대현 (1997). 우리나라 아동의 지능 개념: 원형적 접근. **교육심리연구**, 11(3), 245-270.
- Carroll, J. B. (1978). *On the theory-practice interface in the measurement of intellectual abilities*. In P. Suppes(Ed.), *Impact of research on education: Some case studies*. Washington, D.C.: National Academy of Education.
- Gass, C. S., Demsky, Y. L., & Martin, P. C. (1998). Factor analysis of the WISC-R(spanish version) at 11 age levels between 6-sup-1/-sub-2 and 16-sup-1/sub-2 years. *Journal of Clinical Psychology*, 54, 109-113.
- Hodges, K. K. (1988). Factor structure of the WISC-R for a psychiatric sample. *Journal of Consulting & Clinical Psychology*, 50, 141-142.
- Joanna, M. W. (2003). *The predictive validity of WISC-III factor-based index scores for diverse sociocultural groups*. Temple University.
- Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 61/2 and 161/2 years. *Journal of Consulting and Clinical Psychology*, 43, 135-147.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Mertler, C. A., & Vannatta R. A. (2000). *Advanced and multivariate statistical methods: practical application and interpretation* (4th ed). Pyrczak Pub, Glendale, CA.
- Reynolds, C. R., & Ford, L. (1994). Comparative three-factor solution of the WISC-III and WISC-R at 11 age levels between 6-1/2 and 16-1/2 years. *Archives of Clinical Neuropsychology*, 9, 553-570.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Stewart, K., & Moely, B. (1983). The WISC-R third factor: What does it mean? *Journal of Consulting & Clinical Psychology*, 51, 928-929.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.

Construct Validity of the Intelligence Test in
ACievement–Cognitive ability ENdorsement Tests(ACCENT)

Kim, Dong-II

Seoul National University

Yeo, Seung-Soo

Korean Educational Development Institute

Hong, Sung-Doo

Kwangju Women's University

<Abstract>

The main purpose of this study was to examine the construct validity of ACCENT(AChievement–Cognitive ability ENdorsement Tests). In the current study, data (n=1,596) were equally selected from first thorough sixth graders. Based on the intelligence theory on the Wechsler Intelligence Scale for Children (WISC), it was necessary to investigate whether this test is fitted to two or three factors. As a result, the three – factor model provided most acceptable fit for all grade levels. In addition, the three – factor model was superior to the two – factor model in terms of clinical interpretation. Finally, this study investigated the measurement invariance (MI) between male and female. Results showed that the measurement structure of ACCENT was invariance between the two groups.

Key Words

: Intelligence Test, Construct Validity, Measurement invariance

논문 접수: 2009. 10. 20 심사 시작: 2009. 11. 13 게재 확정: 2009. 12. 18