韓國心理學會誌 : 發達
*Korean Journal of Psychology : Developmental*
1996. Vol. 9, No. 1, 60-70.

# 한국형 유아지능검사(WPPSI)의 표준화: 표준화연구(1)*

박혜원 울산대학교
곽금주 오산전문대
박광배 충북대학교

전국적으로 1,236명의 유아(3세-7세)를 대상으로 한국형 웩슬러 유아지능검사를 실시하여 표준화된 규준을 작성하고자 하였다. 만 3세부터 6개월 간격으로 총 9개 연령집단으로 나누어 WPPSI-R (Wechsler Preschool and Primary Scale of Intelligence-Revised)을 예비연구를 거쳐 제작한 한국형 검사를 실시한 결과 미국 규준에 의한 한국 유아의 지능은 평균은 104.6 (SD=14.7)이며 45-160의 범위에서 분포하였고 정상분포를 이루었다. 채점시 검사자의 해석이 많이 관여되는 5개 소검사에 대한 채점자간 신뢰도는 도형 소검사 .92, 미로 소검사 .95, 이해 소검사 .97, 어휘소검사 .94, 공통성 소검사 .96으로 매우 높았다. 154명의 아동을 대상으로 3주에서 7주간의 간격을 두고 재검사한 결과 검사-재검사 신뢰도는 동작성 IQ는 .85, 언어성 IQ는 .84이며 전체 IQ는 .89였다. 전체아동의 수행을 12개 소검사별로 요인구조를 분석한 결과 언어성 요인과 동작성 요인의 두 요인이 추출되었다. 각 연령집단별 분석에서도 마찬가지로 언어성 요인이 제 1요인으로, 동작성요인은 제 2요인으로 추출되었다. 공준타당도검증을 위하여 기존의 지능검사인 KEDI-WISC및 고대-BINET검사와의 상관을 분석한 결과 KEDI-WISC와는 .49(N=48, p<.001), 고대-BINET검사와는 .77(N=89, p<.0001)의 상관을 보였다. 이상에서 본 검사의 타당도와 신뢰도를 검증하였고 이를 토대로 한국형 규준을 제작하여 출판하는 작업을 남겨두고 있다.

오늘날 유아의 조기교육이 보편화 되면서 과거에 비해 아동의 장애나 지적 능력에 대한 진단요구 또한 매우 높다. 어린 시기의 지능검사는 영재성 또는 정신 지체등의 특성을 조기에 발견하여 가장 적절한 교육적 조치를 취함으로써 문제를 예방하거나 아동의 잠재적 능력을 고양하는데 기여할 수 있다는 점에서 매우 의미있는 도구이다.

아동용 지능 검사의 발전과정을 보면 1905년 프랑스의 Binet검사를 선두로 하여 Gesell의 발달력 검사, Bayley검사, Denver Developmental Screening Test (DDST), Wechsler Intelligence Scale for Children (WISC), 그리고 본 유아지능검사의 모체인 Wechsler Preschool and Primary Scale of Intelligence (WPPSI) 및 그 개정판 WPPSI-R 그리고 McCarthy 검사, Kaufman (K-ABC) 검사 등이 개발되어 있다.

우리나라에서는 1960년대에 고대-비네검사와 WISC가 표준화되었고 1991년도에는 한국교육개발원에서 WISC-R을 확장, 번안하여 5세부터 15세 아동용 KEDI-WISC를 개발하였다. 그런데 이들 검사 중 오래된 것은 시대적으로 타당성이 낮은 문항들을 포함하고 있고 최근 검사의 경우에도 그 절차와

* 이 논문은 1994년도 한국 학술진흥재단의 공모과제 연구비에 의해서 연구되었음

채점등에서 문제점이 지적되고 있다.

더구나 그 이하의 연령에 해당하는 유아나 영아에 대한 지능검사는 현재 표준화되어 있지 못하다. 예를 들어 Bayley지능검사의 경우 한국형 표준화 작업을 1980년 초에 착수하였으나 아직 실용화되고 있지 않다.

우리나라에서의 웩슬러 지능검사의 최근 현황을 살펴보면 1991년도에 한국임상심리학회에서 성인용 Wechsler 지능검사를 재표준화하였고 같은 시기에 한국교육개발원에서도 아동용 Wechsler 지능검사를 재표준화하여 내놓았다. 따라서 빠른 시기에 본 연구의 유아지능검사가 개발된다면 거의 같은 내용의 지능검사에 의해 같은 시대의 3세부터 성인을 대상으로 개발된 검사를 실시하여 지능을 비교할 수 있게 되는 이점을 지닐 수 있게 된다. 인간의 지능을 평생 추적하거나 서로 다른 연령 집단의 지적 능력을 비교하기 위해서는 안전성있는 지능검사의 개발이 필수적이다. 본 웩슬러 유아 지능검사는 성인용으로 부터, 아동용, 유아용으로 확장됨으로 해서 안정성이 높은 장점이 있다.

한국형 웩슬러 유아지능검사 (Korean WPPSI) 개발을 위해서는 1차년도에 예비연구를 통하여 검사도구를 수정, 보완하였고 2차년도에는 본 표준화 연구를 수행하였다. 여기에서는 완성된 한국 규준과 지능검사를 출판하기에 앞서 검사의 신뢰도와 타당도 분석결과및 미국 자료와 비교, 분석한 결과를 보고하고자 한다.

## 방 법

### 연구대상

본 연구에서는 전국을 서울및 경인지역, 충청중부지역, 그리고 그 이남의 남부지역으로 구분하여 서울, 청주, 울산을 기점으로 세 지역근처에 거주하는 만 3세에서 7세 3개월까지의 아동을 대상으로 하였다. 검사실시는 1995년 3월부터 7월에 걸쳐 이루어졌다. 아동의 연령은 미국의 본 검사 제작방식과 마찬가지로 분류하였는데, 정확한 연령구분은 검사실시일 기준으로 2세 11개월 16일부터 3세 5개월 15일까지는 만 3세 집단으로, 만 3세 5개월 16일부터 만 3세 11개월 15일된 아동은 3세 반 집단등으로 구분하여 본 연구에는 만 2세 11개월 16일부터 7세 3개월 15일까지의 9개 연령집단의 아동 1236명이 참여하였다. 이들외에 7명의 아동이 검사에 참여하였다가 끝마치지 못하여 분석에서 제외되었다. 중도에 탈락한 아동은 주로 만 3세에 가까운 어린 아동이었다. 본 연구에 참여한 아동의 연령 분포와 모의 교육수준 및 부의 직업분포는 표 1, 2 와 같다.

### 도구

본 한국형 웩슬러 유아지능검사는 1989년에 개정된 웩슬러 유아지능검사 (Wechsler Preschool and Primary scale of Intelligence-Revised)를 수정, 번안한 것으로 기존의 다른 웩슬러 지능검사와 마찬가지로 6가지 동작성검사와 6가지 언어성 검사의 12개 소검사로 이루어져 있다. 소검사외에도 기록용지에는 검사중에 관찰한 행동이나 일상행동에 대한 부모의 보고내용을 적을 수 있는 행동관찰 기록면

표 1.연령별 지역분포(빈도)

| 연령(세) / 지역 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 | 5.5 | 6.0 | 6.5 | 7.0 | 전체 |
|---|---|---|---|---|---|---|---|---|---|---|
| 서울,경인 | 43 | 49 | 47 | 51 | 46 | 54 | 51 | 55 | 30 | 426 |
| 충북 | 20 | 42 | 36 | 47 | 50 | 53 | 48 | 35 | 46 | 377 |
| 경남 | 47 | 48 | 47 | 52 | 53 | 50 | 53 | 48 | 26 | 424 |
| 전체 | 110 | 139 | 130 | 150 | 149 | 157 | 152 | 138 | 102 | 1227 |

표 2. 지역별 모의 교육수준 및 아버지의 직업분포(빈도)

| 모의교육 / 부의 직업 | 서울,경인 | | 충북 | | 경남 | | 전체 | |
|---|---|---|---|---|---|---|---|---|
| | 고졸이하 | 전문대 이상 | 고졸이하 | 전문대 이상 | 고졸이하 | 전문대 이상 | 고졸이하 | 전문대 이상 |
| 무 직 | 0 | 1 | 2 | 2 | 2 | 0 | 4 | 3 |
| 전문, 관리, 행정 | 41 | 160 | 56 | 59 | 177 | 119 | 274 | 338 |
| 판매 (자영업) | 82 | 107 | 98 | 35 | 58 | 16 | 238 | 158 |
| 서비스 (피고용) | 5 | 21 | 55 | 8 | 10 | 7 | 70 | 36 |
| 농업, 어업, 임업 | 0 | 0 | 26 | 3 | 0 | 1 | 26 | 4 |
| 기술 (생산) | 4 | 0 | 9 | 2 | 27 | 2 | 40 | 4 |
| 기술(직공/노동) | 1 | 0 | 6 | 0 | 5 | 1 | 12 | 1 |
| 전 체 | 133 | 289 | 252 | 109 | 279 | 146 | 664 | 544 |

주: 각 빈도는 전체 응답자중 무응답을 제외한 나머지임

이 포함되어 있다. 표 3에는 각 소검사의 종류를 제시하였다. 각 소검사 옆에 붙은 번호는 일반적으로 검사를 실시하는 순서를 나타낸다. 이 12개의 하위 검사중에 문장과 동물짝짓기는 선택검사로서 특별히 상세한 분석이 필요한 아동이나 검사중 문제로 인하여 다른 검사를 실시하지 못한 경우에 대체용으로 사용된다.

본 연구에서는 표준화과정이므로 모든 아동에게 12가지 소검사를 모두 실시하였다. 그러나 각 아동의 지능을 산출하기 위해서는 10개의 본 검사의 수행만을 사용하였다.

표 3. WPPSI-R의 소검사

| 동작성 | 언어성 |
|---|---|
| 1. 모양맞추기 (Object Assembly) | 2. 상식 (Information) |
| 3. 도형 (Geometric Disign) | 4. 이해 (Comprehension) |
| 5. 토막짜기 (Block Design) | 6. 산수 (Arithmetic) |
| 7. 미로 (Mazes) | 8. 어휘 (Vocabulary) |
| 9. 빠진곳 찾기 (Picture Completion) | 10. 공통성 (Similarities) |
| *11. 동물 짝짓기 (Animal Pegs) | *12. 문장 (Sentences) |

주: *는 선택적으로 실시하는 검사

**절차**

미국판 WPPSI-R의 언어성 검사문항중 예비 연구결과(박혜원, 곽금주, 박광배, 1994)에서 변별력(문항과 총점간의 상관)이 낮게 나타난 것들을 새로운 문항으로 교체하였다. 또한 문화적인 차이를 보이는 그림자료들은 우리 실정에 맞는 것(예: 가스오븐대신 가스레인지)으로 대체하였고 화폐의 단위등은 우리 것으로 바꾸었다(예:페니 대신 원).

그러나 가능한 원 검사의 내용을 충실하게 유지시키려는 입장에서 소검사별로 미국판 WPPSI-R과 완전히 다른 문제는 표 4와 같이 총 7문항밖에 되지 않는다.

**조사자 훈련과 실시:**

조사자는 총 28명으로 이들의 훈련은 각 지역별로 연구책임자(서울 경인지역:곽금주; 충청중부지역:박광배; 경남지역:박혜원)와 책임조교 1명의 지도로 이루어 졌다. 모든 실시지침은 지침서(박혜원, 곽금주, 박광배, 1995)에 실려 있는 내용을 따랐다.

표 4. 미국판 WPPSI-R에서 교체된 문항

| 소검사 | 미국판 문항 | 교체 문항 |
|---|---|---|
| 상식 | 7. 코 | 발가락 |
| | 8. 귀 | 코끼리 다리 |
| 이해 | 13. 집-아프다 | 공부 |
| 어휘 | 16. 영웅 | 육교 |
| | 17. 관람자 | 태풍 |
| | 22. 도박 | 교통순경 |
| 공통성 | 20. 행복-슬픔 | 피아노-바이올린 |

**분석 (통계처리) 방법:**

본 연구에서는 모든 분석을 SAS를 이용하였다. 대상 아동의 인구통계학적인 특성에 대해서는 기술통계를 산출하였고 채점자간 신뢰도, 검사-재검사 신뢰도및 공준 타당도 검증을 위해서는 Pearson 상관을 산출하였다. 구성개념 타당도를 위해서는 주축요인 방식에 의한 요인분석을 실시하였다. 미국 WPPSI-R의 경우에는 미국 규준에 의한 '환산점수(scaled score)' 와 지능지수(IQ) 에 대하여 이러한 통계적 속성들이 파악되었지만, 한국형 WPPSI의 규준은 아직 개발중에 있기 때문에 본 보고서에는 미국 규준에 의해 환산점수와 지능을 산출하였다.

## 결 과

본 결과분석에는 1236명의 아동중 각 분석에서 무응답반응이나 코딩의 문제가 있는 아동을 제외한 자료를 제시하였다. 먼저 전체 아동의 지능의 분포를 표 5에 제시하였다. 미국 규준에 의한 아동의 평균은 104.6(SD=14.7)이며 45-160의 범위에서 정상분포를 이루었다.

### 1. 채점자간 신뢰도

웩슬러 유아지능검사의 경우 더 나이든 개인을 대상으로 하는 검사보다 채점자간 신뢰도(inter - scorer reliability)를 검증하는 것이 필수적이다. 왜

표5. 전체아동의 IQ 분포

| IQ분포 | 빈도 | (%) | 누적빈도 | (누적 %) |
|---|---|---|---|---|
| 130 - | 61 | (4.9) | 61 | (4.9) |
| 120 - 129 | 115 | (9.3) | 176 | (14.2) |
| 110 - 119 | 269 | (21.8) | 445 | (36.0) |
| 90 - 109 | 521 | (50.2) | 1066 | (86.2) |
| 80 - 89 | 125 | (10.1) | 1191 | (96.4) |
| 70 - 79 | 30 | (2.4) | 1221 | (98.8) |
| - 69 | 15 | (1.2) | 1236 | (100.0) |

냐하면 어린 유아에 있어서는 언어적 표현력이나 이해력이 부족하여 검사자와 채점자의 영향을 많이 받기 때문이다. 따라서 서로 다른 채점자가 136명 아동의 5가지 소검사 결과를 독립적으로 평정한 후 그 상관을 분석하여 일반화 가능도를 검증하였다. 이들 소검사는 채점자의 판단에 따라 차이를 보일 수 있는 것들로 미국판 검사의 경우에도 채점자간 신뢰도 분석이 이루어진 이해, 어휘, 공통성, 미로 및 도형 소검사이다. Pearson 상관을 통한 분석결과 이해 .97, 어휘 .94, 공통성 .96, 미로 .95, 도형 .92로 모두 .90이상의 높은 수준을 유지하였다.

### 2. 검사-재검사 신뢰도

검사-재검사 신뢰도분석을 위해서는 154명의 아동을 대상으로 3주에서 7주 간격을 두고 (평균; 4주) 검사를 반복 실시하였다. 미국판 WPPSI-R에서는 연령집단을 나누어 분석하였으나 본 연구에서는 전체 아동에 대한 분석 자료를 표 6에 제시하였다.

소검사별로 살펴볼 때 .50에서 .82사이에 분포하였고 가장 낮은 것은 동물짝짓기(r=.50)와 모양맞추기(r=.52)로 첫번째 검사에 비해 두번째 검사에서 많은 수행증가를 나타낸 것을 알 수 있었다.

### 3. 검사의 반분신뢰도

미국판과의 비교를 위하여 원 저자들이 사용한 방식에 따라 검사의 반분신뢰도를 통해 검증하였다.

표 6.검사-재검사의 IQ 안정성

| | 첫번째 검사 | | 두번째 검사 | | 검사-재검사 상관 |
|---|---|---|---|---|---|
| | 평균 | 표준편차 | 평균 | 표준편차 | |
| 모양맞추기 | 9.4 | 2.7 | 10.7 | 3.1 | .52 |
| 도 형 | 13.2 | 2.8 | 13.2 | 2.8 | .61 |
| 토막 짜기 | 12.3 | 2.6 | 13.4 | 2.7 | .72 |
| 미 로 | 11.6 | 3.4 | 11.9 | 3.9 | .71 |
| 빠진곳찾기 | 12.4 | 3.2 | 12.9 | 3.1 | .82 |
| 동물짝짓기 | 10.6 | 2.9 | 11.8 | 2.7 | .50 |
| 상 식 | 7.5 | 3.5 | 8.1 | 3.7 | .82 |
| 이 해 | 8.3 | 2.9 | 8.9 | 3.0 | .66 |
| 산 수 | 11.9 | 3.1 | 11.9 | 3.2 | .68 |
| 어 휘 | 8.6 | 3.0 | 9.5 | 3.0 | .63 |
| 공 통 성 | 7.7 | 2.9 | 8.3 | 2.6 | .69 |
| 문 장 | 8.5 | 2.7 | 8.7 | 2.7 | .69 |
| 동작성 IQ | 114.3 | 17.5 | 119.8 | 18.0 | .85 |
| 언어성 IQ | 94.0 | 14.9 | 97.3 | 14.7 | .84 |
| 전체 IQ | 104.2 | 16.2 | 108.8 | 16.8 | .89 |

미국판 WPPSI-R의 경우는 3세부터 7 세까지의 아동을 연령에 따라 6 개월 단위로 9 개의 연령집단으로 나누어 각 연령집단에서 반분신뢰도를 산출하였으나 본 연구에서는 사례수가 비교적 적은 관계로 1년 6 개월 단위로 3 개의 연령집단으로 분류한 후 각 집단내에서의 반분신뢰도를 산출하였다. 이 신뢰도는 표 7에 제시하였다. 표 7에 제시된 반분신뢰도계수는 다음과 같은 Spearman-Brown 공식에 의해 수정한 것이다. $\rho_{yy'}$ 는 반분검사들간의 상관계수이고, $\rho_{xx'}$ 는 최종적으로 수정된 반분신뢰도계수이다.

$$\rho_{xx'} = \frac{2\rho_{yy'}}{(1+\rho_{yy'})}$$

또한 표7에는 측정오차의 정도를 나타내는 '측정의 표준오차(standard error of measurement) '도 제시하였다. 이 측정의 표준오차는 다음과 같은 공식에 의하여 산출되었다. $\sigma_x$ 는 소검사 점수의 표준편차이다.

$$\sigma_e = \sigma_x\sqrt{1-\rho_{xx'}}$$

표 7에서 평균신뢰도의 경우 가장 낮은것은 0.72 이고 가장 높은 것이 0.95이다. 미국판의 경우 평균신뢰도계수의 범위는 0.63-0.86 이다. 미국판과 마찬가지로 동작성 요인의 소검사들의 신뢰도가 다른 소검사들의 신뢰도보다 낮은데, 그 한 이유는 모양맞추기 소검사등 이들 동작성 소검사들의 경우 문항수가 다른 소검사들에 비하여 적기 때문이다.

표 7에 제시된 측정의 표준오차는 전반적으로 미로 소검사에서 가장 작고 이해 소검사에서 가장 큰 수치를 보이고 있으며 또한 세 연령집단간에는 매우 유사한 유형을 보여주고 있다.

12개의 소검사들중 동물짝짓기는 속도검사 (speed test)이고, 따라서 반분신뢰도가 적용되지 않는 검사이므로 여기에서는 그것의 신뢰도지수와 측정오차

표 7. 소검사의 연령별 신뢰도 계수와 측정의 표준오차

| | 연령집단 | | | | | | |
|---|---|---|---|---|---|---|---|
| | 3.0세-4.0세(N=379) | | 4.5세-5.5세(N=456) | | 6.0세-7.0세(N=392) | | 평균( N=1227) |
| 소검사 | 신뢰도 | 표준오차 | 신뢰도 | 표준오차 | 신뢰도 | 표준오차 | 신뢰도 |
| 모양 맞추기 | .64 | 2.90 | .51 | 3.16 | .37 | 3.08 | .72 |
| 도 형 | .86 | 4.06 | .68 | 3.98 | .51 | 2.67 | .85 |
| 토막 짜기 | .74 | 2.65 | .79 | 2.69 | .73 | 2.78 | .88 |
| 미 로 | .81 | 1.79 | .78 | 1.79 | .73 | 1.62 | .89 |
| 빠진곳 찾기 | .90 | 1.76 | .81 | 1.70 | .57 | 1.71 | .86 |
| 동물 짝짓기 | — | — | — | — | — | — | — |
| 상 식 | .67 | 2.95 | .81 | 2.36 | .77 | 1.91 | .90 |
| 이 해 | .71 | 3.18 | .70 | 3.54 | .65 | 3.01 | .83 |
| 산 수 | .82 | 1.29 | .88 | 1.31 | .78 | 1.31 | .95 |
| 어 휘 | .77 | 2.23 | .76 | 2.96 | .77 | 3.80 | .83 |
| 공 통 성 | .79 | 1.63 | .84 | 1.86 | .87 | 2.12 | .92 |
| 문 장 | .80 | 1.64 | .88 | 1.88 | .82 | 2.63 | .91 |

표 8. 전체 아동의 소검사 요인구조

| 소 검 사 | 요인 1 | 요인 2 |
|---|---|---|
| 모양 맞추기 | 0.15285 | 0.55379 |
| 도 형 | 0.17103 | 0.60281 |
| 토막 짜기 | 0.28987 | 0.64281 |
| 미 로 | 0.13931 | 0.55447 |
| 빠진곳 찾기 | 0.28454 | 0.53783 |
| 동물 짝짓기 | 0.27612 | 0.41640 |
| 상 식 | 0.73923 | 0.19184 |
| 이 해 | 0.67115 | 0.22369 |
| 산 수 | 0.54783 | 0.34439 |
| 어 휘 | 0.69429 | 0.13774 |
| 공 통 성 | 0.54155 | 0.29436 |
| 문 장 | 0.47802 | 0.27240 |

가 산출되지 않았다. 표 7에는 각 소검사별 전체아동의 평균신뢰도계수 또한 제시하였는데 이것은 각 소검사의 연령별 신뢰도계수 세개를 Fisher의 z 통계치로 변환하고, 그들의 가중평균을 구한 후, 그것을 다시 신뢰도 계수로 역변환한 것이다.

4. 요인구조

WPPSI-R의 소검사들은 이론적으로 동작성 검사들과 언어성검사들로 구분된다. 이러한 이론적인 구분이 존재하는 지를 검증하는 것은 '구성타당도(construct validity)'의 분석이 된다. 따라서 12개의 소검사점수에 대하여 전체아동 및 연령집단별로 요인분석을 실시하였다. 이때 주축요인분석은 Varimax 회전 방식을 이용하였다.

표 8에 제시된 바와 같이 요인 구조를 보면 동작성 요인과 언어성 요인이 명확하게 구별된다. 연령집단별 분석에서도 모두 같은 경향을 보여 제 1 요인이 언어성 요인이고, 제 2 요인이 동작성 요인인데 이는 미국의 원자료의 결과와도 일치한다. 예비연구에서는 이 두 요인외에도 세번째로 '동물 짝짓기'요인이 추출되었는데, 이 요인이 다른 동작성 검사들과 구별되는 이유는 이것이 유일하게 '속도검사 (speed test)' 이기 때문인 것으로 사료되었다. 그런데 보다 많은 수의 아동을 대상으로 한 본 연구에서 이 속도검사 요인은 동작성 지능요인에 포

함된다는 것을 알려주고 있다.

### 5. 소검사간 상관분석

본 연구에서는 각 검사간의 상관유형이 Wechsler 지능이론에 부합하는지 살펴보기 위해 각 소검사의 환산점수간의 상관을 분석하였다.

표 9에 제시된 바와 같이 1235명의 수행을 바탕으로 12개 소검사와 언어성 지능, 동작성 지능 및 전체 수행간의 상관을 분석한 결과 모든 계수들이 .001수준에서 유의미한 상관이 있었다.

동작성 소검사간의 상관을 살펴보면 .22에서 .45의 상관을 보였는데 도형검사와 토막짜기검사간의 상관이 가장 높았다. 또한 보충검사로 사용되는 동물짝짓기 검사와 다른 검사간의 상관이 가장 낮은 것을 알 수 있었다. 언어성 소검사간의 상관을 살펴보면 .35에서 .55사이에 분포하여 동작성 소검사간의 상관보다 다소 높은 것을 알 수 있다. 동작성 지능과 언어성 지능간에는 .49의 상관을 나타내고 있으며 언어성 지능과 전체지능간에는 .87, 동작성 지능과 전체 지능간에는 .84의 상관을 보이고 있다. 동작성 소검사와 언어성 소검사간의 상관은 . 17에서 .38사이의 약소한 상관만을 보여주고 있는데 단지 공간지각능력을 측정하는 토막짜기(동작성 검사)와 산수(언어성)검사간에는 .38의 비교적 높은 상관을 나타내고 있다.

### 6. 공준 타당도 분석

공준 타당도 분석을 위하여 기존의 지능검사인 KEDI-WISC와 고대 BINET 검사를 각각 본 검사의 전후에 실시하였다. 검사간격은 1주일에서 4주 사이로 각 검사의 순서는 counter-balancing하였다. KEDI-WISC의 경우 48명의 아동에게 검사를 실시하였는데 전체 IQ의 경우 .49 (N=48, p<.001)의 상관을 보였다. 고대 - BINET의 경우 89명의 아동을 대상으로 검사를 실시하였는데 .77 (N=89, p<.0001))의 높은 상관을 보였다.

표 9. 소검사간(환산점수) 상관관계*(N=1235)

| 소검사 | OA | GD | BD | MA | PC | AP | IN | CO | AR | VO | SI | SE | Perf | Verb | TOT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 도형 맞추기(OA) | 1.00 | | | | | | | | | | | | | | |
| 도 형(GD) | .33 | 1.00 | | | | | | | | | | | | | |
| 토막 짜기(BD) | .45 | .43 | 1.00 | | | | | | | | | | | | |
| 미 로(MA) | .32 | .41 | .37 | 1.00 | | | | | | | | | | | |
| 빠진곳 찾기(PC) | .34 | .37 | .41 | .36 | 1.00 | | | | | | | | | | |
| 동물 짝짓기(AP) | .29 | .29 | .36 | .22 | .33 | 1.00 | | | | | | | | | |
| 상 식(IN) | .23 | .22 | .35 | .21 | .30 | .29 | 1.00 | | | | | | | | |
| 이 해(CO) | .22 | .27 | .35 | .21 | .32 | .24 | .53 | 1.00 | | | | | | | |
| 산 수(AR) | .24 | .31 | .38 | .32 | .30 | .31 | .54 | .43 | 1.00 | | | | | | |
| 어 휘(VO) | .17 | .20 | .29 | .18 | .28 | .26 | .55 | .51 | .40 | 1.00 | | | | | |
| 공 통 성(SI) | .23 | .29 | .35 | .20 | .31 | .29 | .41 | .44 | .37 | .43 | 1.00 | | | | |
| 문 장(SE) | .25 | .23 | .27 | .23 | .28 | .23 | .39 | .38 | .36 | .35 | .40 | 1.00 | | | |
| 동작성 IQ(Perf) | .64 | .67 | .69 | .69 | .65 | .41 | .33 | .32 | .41 | .25 | .36 | .36 | 1.00 | | |
| 언어성 IQ(Verb) | .29 | .33 | .44 | .31 | .39 | .37 | .80 | .72 | .70 | .72 | .66 | .51 | .49 | 1.00 | |
| 전 체 IQ(TOT) | .53 | .56 | .64 | .57 | .59 | .42 | .67 | .62 | .64 | .58 | .60 | .51 | .84 | .87 | 1.00 |

*모든 계수들은 .001수준에서 유의미한 상관이 있다

## 논 의

지능검사의 표준화작업은 잘 계획된 예비연구들을 통하여 검사자료를 개발하고 규준작성을 위해 전국의 해당 아동을 대상으로 조사하고 분석하는 다단계 연구로 이루어 진다. 본 연구에서는 예비조사에서 개발된 검사도구를 사용하여 표준화 연구를 실시한 후 규준표를 작성하기에 앞서 자료분석의 일부를 살펴보았다.

본 연구에서 대상아동은 전국적으로 해당 아동을 대상으로 층화표집을 실시하는 것이 이상적이나 서울, 경인, 충청 북도, 경남지역에 거주하는 아동으로 국한하였다. 따라서 본 지능검사의 규준이 잘 적용되는 집단의 특성을 분석할 예정이며 교육수준이나 부모의 직업과 같은 사회 경제적인 특성에 따른 지능의 차이를 분석하고 만일 차이가 있으면 이를 고려하여 최종 규준표의 작성시에는 교정할 계획이다.

유아지능검사의 경우 더 나이든 개인을 대상으로 하는 검사보다 채점자간 신뢰도 (interscorer reliability) 의 검증이 더 중요하다. 특히 우리나라의 어린 유아에 있어서는 검사상황에서 언어표현이 부족하여 검사자와 채점자의 영향을 많이 받을 것으로 예상되었다. 두 채점자가 독립적으로 136명의 아동의 5가지 소검사결과를 평정한 후 그 상관을 분석하여 채점자간 신뢰도를 검증하였다. 언어표현과 채점자에 따라 달리 채점될 수 있는 이해, 어휘, 공통성, 미로및 도형 소검사의 수행을 Pearson 상관을 통해 분석한 결과 .92에서 .97 사이로 매우 높은 일반화 가능도를 나타냈다.

선행연구(박 혜원 등, 1994)나 Wechsler (1989)의 보고에서 비교적 낮은 신뢰도를 보인 공통성과 미로 소검사(각각 .88, .80)에서도 본 연구에서는 매우 높은 일치도를 보였다.

검사-재검사 신뢰도분석결과 미국 연구에 비해 비교적 낮은 안정도를 보인 소검사는 동물짝짓기 (r=.50)와 모양맞추기(r=.52)로 첫번째 검사에 비해 두번째 검사에서 많은 수행증가를 나타낸 것을 알 수 있었다. 이 두검사 모두 동작성 검사로 한국 아동의 경우 동작성 수행에 대한 전이 효과가 큰 것이 아닌가 추정할 수 있으나 이 문제에 대해서는 추후 연구가 요망된다.

검사의 타당도분석을 위해서는 반분 신뢰도, 요인분석, 소검사간 상관분석및 공준 타당도연구가 실시되었다. 먼저 반분신뢰도는 미국판과의 비교를 위하여 원저자들이 사용한 반분신뢰도계수를 통해 검증하였다.

각 연령별 신뢰도 계수는 미국판에 비해 (.54-.90) 본 연구에서는 약간 큰 범위로 0.37- 0.90 사이에 분포하였다. 특히 6세이상의 아동집단에서 동작성 검사의 신뢰도가 낮았으며 모양맞추기의 신뢰도는 매우 낮았다. 이 문제는 추후 연구를 통해 개선해야 할 문제점으로 지적되었다. 그러나 전반적인 양상은 비슷하여 미국판과 마찬가지로 모양맞추기 소검사 등 동작성 검사의 신뢰도가 언어성 소검사들의 신뢰도보다 낮았다.

동작성 소검사들은 비교적 문항수가 적었으며 특히 가장 낮은 신뢰도를 보인 모양맞추기 소검사는 6문항 밖에 되지 않았다. 표7에 제시된 측정의 표준오차는 미로소검사에서 가장 작고 이해소검사에서 가장 큰 수치를 보여주고 있다. 표준오차의 분포는 세 연령집단간에는 매우 유사한 결과를 보여주고 있다.

본 연구에서 요인 분석의 결과 미국의 자료와 마찬가지로 언어성 요인과 동작성 요인이 추출되었다. 예비연구(박 혜원등, 1994)에서는 이 두요인외에 속도검사요인이 따로 추출되었으나 보다 많은 아동을 대상으로 하는 본 연구에서는 웩슬러의 지능이론에 부합하는 결과를 보여주고 있다.

소검사간의 상관을 통해 본 검사의 요인구조와 이론적 구성개념(Construct)을 분석한 결과에서도 두 요인간 소검사들의 상관이 각 요인 검사내의 상관보다 낮아서 언어성 요인과 동작성 요인을 잘 구

표 10. 순서조정 문항

| 문항 번호 | 최저 점수 | 최고 점수 | 평균 | 표준편차 | 바꾼 순서 |
|---|---|---|---|---|---|
| 모양맞추기 4 | 0 | 6.00 | 3.22 | 2.09 | 5 |
| 모양맞추기 5 | 0 | 8.00 | 5.37 | 2.23 | 4 |
| 모 양 6 | 0 | 3.00 | 2.44 | 0.99 | 6 |
| 모 양 7 | 0 | 3.00 | 0.38 | 0.82 | 11 |
| 모 양 8 | 0 | 4.00 | 0.72 | 1.20 | 8 |
| 모 양 9 | 0 | 4.00 | 0.62 | 1.15 | 10 |
| 모 양 10 | 0 | 4.00 | 1.25 | 1.60 | 7 |
| 모 양 11 | 0 | 5.00 | 0.84 | 1.35 | 9 |
| 모 양 12 | 0 | 5.00 | 0.23 | 0.80 | 12 |

분해 주고 있었다.

또한 비록 서로 다른 요인간 검사긴 하지만 공간능력(토막짜기:동작성)검사와 산수능력(산수:언어성)검사간에 비교적 높은 상관관계도 확인되었다.

공준타당도 분석에서는 모두 공준타당도를 입증해 주고는 있으나 웩슬러 지능검사 시리즈의 하나로서 본 검사와 보다 그 내용이 유사한 KEDI-WISC보다 고대-BINET 검사와의 상관이 높게 산출되었다. 조사 대상의 연령이나 검사간격등에 대한 상세한 분석을 통하여 그 원인을 분석해 볼 필요가 있다.

그 외에도 각 문항에 대한 수행 수준을 분석한 결과 모양맞추기의 4번보다 5번 문항의 수행수준이 유의하게 높았고 문장 검사의 7번에서 12번사이에서도 난이도의 점증적인 증가경향을 나타내지 못하였다. 따라서 이 문항들의 순서는 최종 표준화 검사에서는 조정할 예정이다. 이 문항들을 표 10에 제시하였다.

지능검사가 아동의 이해에 중요한 기본틀이 된다는 것은 주지의 사실이다. 특히 본 검사의 모체인 웩슬러 유아지능검사는 지능의 다양한 측면을 인정하면서도 지능이 각 영역에서 개별적인 별개의 능력으로 존재하는 것이 아니라 다차원적인 속성을 가진 총체적인 것으로 파악하고 있다(Wechsler, 1967). 따라서 12가지 소검사를 통하여 아동의 수행을 측정한 후 이를 합산하여 전체 지능을 산출하는데 이 소검사들은 그 특성에 따라 동작성 지능과 언어성 지능으로 대별되는 것으로 보고 있다. 그런데 이 두가지 종류의 지능의 수행양상은 단순히 지능지수뿐 아니라 적응상태나 정신상태에 대한 정보를 제공할 수 있다고 보고된다. 앞으로 본 지능검사의 자료를 바탕으로 우리 실정에 따른 프로화일 해석등 질적 해석방법이 개발되어야 할 것이다.

매우 어린 아동의 지능을 측정하는 데에는 실시상의 어려움뿐 아니라 측정된 지능의 안정성이 낮다는 문제점이 제기되고 있다(Brooks-Gunn, & Weinlaub, 1983; Lewis, & Sullivan, 1985). 본 연구는 1939년 제작된 웩슬러 성인용 지능검사로부터 응용되어 3세의 어린 아동부터 7세까지의 아동의 지능을 측정하도록 확장된 웩슬러 유아지능검사를 표준화하는 데 목적이 있다. 따라서 본 검사는 다른 어떤 지능검사보다도 연령변화에 따라서도 안정성을 보일 것으로 기대되므로 앞으로 종단연구를 통한 안정성 검증이 필요하다. 본 지능검사가 원 웩슬러 지능검사의 취지와 명성에 맞는 것이 되기 위해서는 앞으로 많은 분석과 수정 작업및 연구가 수행되어야 할 것이다.

## 참 고 문 헌


박혜원, 곽금주, 박광배(1994). 한국형 유아지능검사(WPPSI-R)의 표준화: 예비연구. 한국심리학회: 발달. 7(2), 38-52

박혜원, 곽금주, 박광배(1995). WPPSI-R지침서. 서울: 도서출판 특수교육

Brooks-Gunn, J. & Weinlaub, M. (1983). Origins of infant intelligence. In M. Lewis(Ed.), *Origins of intelligence: Infancy and early childhood*(pp. 25-66). New York: Plenum.

Lewis, M. & Sullivan, M. W. (1985). Infant intelligence and its assessment. In B.B. Wolman(Ed.), *Handbook of intelligence: Theories, measurements and applications*(pp. 505-599). New York: Wiley.

Wechsler, D.(1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence.* New York: The Psychological Co.

Wechsler, D.(1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence - Revised.* New York: The Psychological Co.

韓國心理學會誌 : 發達
*Korean Journal of Psychology : Developmental*
1996. Vol. 9, No. 1, 60-70.

# The development of Korean version of WPPSI: The standardization study(1)

Hyewon Park Choi
U. of Ulsan

Keum-Joo Kwak
Osan College

Kwang-Bae Park
Choongbuk Univ.

Korean version of WPPSI-R(Wechsler Preschool and Primary Scale of Intelligence-Revised) was adminstered to 1236 3- to 7-year-olds in three locations nationwide. According to the U.S. norm, Korean children's IQ were normally distributed between 45-160 with a mean of 104.6(SD=14.7). Interscorer reliability coefficients obtained with 136 children were high: .97 on Comprehension, .94 on Vocabulary, .96 on Similarities, .95 on Mazes and .92 on Geometric Design. For the test-retest reliability, 154 children were tested twice with intervals of 3-7 wks. The reliability coefficients were .85 for Performance IQ, .84 for Verbal IQ and .89 for the Total IQ. Factor analyses revealed two factors of 'verbal' and 'performance' IQ as predicted by Wechsler's two factor theory of intelligence. The pattern of correlations among 12 subtests was also consistant with the two factor theory: Correlations between subtests within a factor were higher than those between subtests across factors. Correlations among 'Performance' subtests were lower (.22 to .45) than those among 'Verbal' subtests(.35 to .55). Optional subtests of 'Animal Pegs' and 'Sentence' revealed the lowest correlations with other subtests. The concurrent validity was approved by analyses of correlations between WPPSI-R and KEDI-WISC as well as Kodae-BINET. Based on these validity and reliability data, Korean WPPSI will be published.