

한국형 유아지능검사(WPPSI-R)의 표준화 : 예비연구

박혜원
울산대학교

곽금주
오산전문대

박광배
충북대학교

3세에서 7세사이의 유아용 웨슬러 유아지능검사의 제작을 위하여 전국적으로 3개 지역 (서울, 청주, 울산)에서 3세부터 6개월간격으로 총 9개 연령집단의 남녀 301명을 대상으로 수정, 번역된 웨슬러 유아 지능검사를 실시하였다. 검사실시에 앞서 원 지침서를 번역하였고 한국적 문화배경을 고려하여 상식, 이해, 어휘 및 공통성 소검사등의 언어성 검사의 내용과 동작성 검사의 몇가지 그림항목들을 수정하였으며 문항 변별력이 낮은 문항들을 대체할 새로운 문항들을 첨가하여 예비연구용 검사문 제작하였다. 우선 30명의 아동의 자료를 2명의 채점자가 독립적으로 채점한 결과 채점자간 신뢰도가 이해 .94, 어휘 .97, 공통성 .88, 미로 .80, 도형 .94로 나타나 높은 신뢰도를 나타냈다. Spearman-Brown 반분 신뢰도계수에 의한 11개 소검사들의 신뢰도는 원검사인 미국의 WPPSI-R과 유사하여 모양맞추기(.56)을 제외한 모든 검사가 .80이상의 높은 수준을 보였다. 속도검사인 동물짜짓기 소검사에 대해서는 반분신뢰도점중은 적절치 않으므로 계산치 않았다. 또다른 검사 일관성 지표로서 측정의 표준오차도 계산하였다. 소검사들의 요인 구조 분석결과 원검사와 유사한 언어성 요인(6가지 언어성 검사)과 동작성 요인(5가지 동작성 검사)이 밝혀졌으나 미국의 자료와는 달리 유일한 속도검사이고 동작성 검사인 동물짜짓기 소검사요인이 제 3의 요인으로 밝혀졌다. 이 세 요인은 모든 연령집단에서 전체 변량의 60% 이상을 설명하였다. 전체 지적 수행상에 남녀 차이는 없었으나 대부분의 소검사에서 여아가 남아보다 우세한 경향을 보였다. 특히 도형 및 동물짜짓기 소검사에서는 여아의 수행이 유의하게 높았다. 본 예비연구에서 한국형 웨슬러 유아지능검사의 도구제작을 위한 작업이 이루어졌으며 후속 표준화연구를 통하여 규준표의 작성과 기타의 타당도연구가 이루어져야 한다.

지능에 대한 정의는 최초의 지능검사인 Binet의 검사가 1905년 발표된 이래 끊임없는 논란이 되어 왔다. 그러나 지능을 자기와 상황에 대한 이해능력과 환경에 대한 적응력, 유연성, 또는 훈련가능성으로 보는 점에서는 그 의견을 같이 한다(Wechsler, 1967). 이러한 점에서 한 인간의 지능은 그에 대한 이해의 중요한 지표로서 교육적, 임상적, 그리고 연구조사 등에

* 이 논문은 1993년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

서 매우 큰 의미를 갖게 된다.

아동용 지능 검사의 발전과정은 위에 언급한 Binet검사를 선두로 하여 Gesell의 발달력 검사, Bayley검사, Denver Developmental Screening Test(DDST), Wechsler Intelligence Scale for Children(WISC)등이 개발되어 있으며 우리나라에서도 WISC와 WISC-R(개정판)이 표준화되어 있다. 즉 한국교육개발원에서 1960년대에 처음 WISC를 한국표준화한 이래 1991년도 5세부터 15세 아동용 KEDI-WISC를 개발하였다.

그러나 그 이하의 연령에 해당하는 아동이나 영아기에 대한 지능검사는 현재 표준화되어 있지 못하다. Bayley지능검사의 경우 한국형 표준화 작업을 1980년 초에 착수하였으나 아직 그 결실을 보지 못하고 있다. 이 검사는 2개월부터 30개월된 영아에게 실시할 수 있는 개인 검사인데 가장 큰 문제점으로는 이 검사의 결과가 후의 지능을 예언하는데 있어 그 예언력이 매우 낮다는 점으로 들어 나고 있다 (Bornstein & Sigman, 1986; Brooks-Gunn & Weinlaub, 1983, Lewis & Sullivan, 1985).

예를 들어 Lewis와 Sullivan(1985) 영아기의 지능검사결과와 취학전 연령의 아동의 지능간에는 상관이 거의 없다는 것을 지적하고 있다. 이 점에 대해 Bayley 검사의 타당도에 문제가 있는 것으로 해석하는 연구자들이 있다. 그러나 이렇게 상관이 낮은 이유중의 하나는 영아기에 측정되는 기술과 그 이후에 측정되는 행동의 속성과 범위가 다르기 때문이다. 즉 영아와 아동은 행동 범위의 폭이 다르기 때문에 두 집단에게서 지적인 능력을 측정하기 위해 사용하는 과제는 다를 수 밖에 없다. 그러나 기억, 언어와 수학적 개념에 대한 이해, 유사성과 상이성에 대한 지각, 결과에 대한 기대, 새로운 문제의 해결 그리고 공간에서의 형태에 대한

시각적 표상능력과 같은 포괄적인 지적 능력은 영아기부터 계속해서 관찰할 수 있는 지적 능력이다.

따라서 1967년 Wechsler는 기존의 아동용 지능검사인 WISC의 대상 연령인 6세보다 낮은 연령의 아동을 대상으로 하는 WPPSI (Wechsler Primary and Preschool Scale of Intelligence)를 개발하였다. 그후 1990년에는 이를 개정한 WPPSI-R을 내놓았다. WPPSI-R은 웨슬러 아동지능검사 (개정판:WISC-R, 1974)의 대부분의 하위검사를 연장적용함으로써 전세계적으로 보편화되어 사용되는 웨슬러 지능검사의 시리즈의 하나로서 3세이상의 어느 연령집단과도 거의 동일한 능력을 측정하는 것을 보증받을 수 있는 장점을 지녔다. 즉 웨슬러 지능검사는 이제 3세이후의 모든 연령에 있어서 광범위한 지적 능력을 비교적 일관성있게 측정할 수 있어 명실공히 인간의 지능을 평생동안 추적하거나 서로 다른 연령의 집단을 비교할 수 있게 해 주고 있다.

우리나라에서도 1991년도에 한국임상심리학회에서 성인용 Wechsler 지능검사를 재표준화하였고 비슷한 시기에 한국교육개발원에서도 아동용 Wechsler지능검사(KEDI-WISC)를 재표준화하여 내놓았다. 따라서 빠른 시일내에 본 연구의 유아지능검사가 개발된다면 거의 동일한 시기에 제작된 웨슬러지능검사 시리즈를 통하여 3세부터 성인까지의 지능을 비교할 수 있게 되는 이점을 지닐 수 있게 된다.

오늘날 유아의 조기교육이 보편화되면서 과거에 비해 아동의 장애나 지적 능력에 대한 진단요구 또한 매우 높다. 어린 시기의 지능검사는 영재성 또는 정신 지체등의 특성을 조기에 발견하여 조치함으로써 가장 적절한 교육적 조치를 마련할 수 있으며 문제를 예방하거나 아동의 잠재적 능력을 고양하는데 기여할 수 있

다는 점에서 매우 의미있는 도구가 된다.

한국형 웨슬러 유아지능검사(Korean WPPSI-R)개발을 위한 전체 연구는 1차년도에는 예비연구를 그리고 2 3차년도에는 표준화 연구를 수행하게 되는데, 본 연구는 예비연구에 해당하는 것으로 검사도구의 제작을 목적으로 한다. 따라서 연구방법은 문항번안, 수정, 보완을 중심으로 각 하부척도의 신뢰도를 확인하고 극대화하는 것과 평정자의 일치도를 높이기 위한 일반화가능도의 검증에 초점을 둔다.

방 법

연구대상

본 예비연구에서는 전국을 서울 경인지역, 충청중부지역, 그리고 그 이남의 남부지역으로

구분한 후 서울, 청주, 울산읍 기점으로 세 지역에 거주하는 만 3세에서 7년 6개월까지의 아동을 대상으로 하였다. 아동의 연령은 미국의 본 검사 제작방식과 마찬가지로 분류하였는데 정확한 연령구분은 검사실시일기준 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 5개월 15일까지의 9개 연령집단의 아동 301명이 참여하였다. 이들의외에 7명의 아동이 검사에 참여하였다가 끝나치지 못하여 분석에서 제외되었다. 증도에 탈락한 아동은 주로 만 3세 정도의 어린 아동이었다. 본 연구에 참여한 아동의 지역별 분포와 부모의 교육수준 그리고 직업분포는 다음과 같다.

표1. 연령별 지역분포(빈도)

	3.0세-4.0세	4.5세-5.5세	6.0세-7.0세	전체
서울	20	26	37	83
청주	36	36	36	108
울산	37	39	34	110
전체	93	101	107	301

표2. 연령별 아버지의 직업분포(빈도)

	3.0세-4.0세	4.5세-5.5세	6.0세-7.0세	전체
무직	1	0	2	3
전문, 관리	42	28	31	101
판매, 행정	16	21	31	68
서비스(피고용)	8	9	17	34
농업, 어업, 임업	0	0	0	0
기술(생산)	15	19	18	52
기술(직공/노동)	0	3	4	7
전체	82	80	103	265

주: 각 빈도는 전체 응답자 301명 중 아버지 직업에 대한 무응답을 제외한 나머жим.

표3. 연령별 아버지의 교육수준(평균 교육 연한)

	3.0세-4.0세	4.5세-5.5세	6.0세-7.0세	전체
평균	14.30	14.36	14.53	
표준편차	2.02	2.07	2.50	
N	82	84	103	

주 : 각 빈도는 전체 응답자 301명 중 아버지 교육에 대한 무응답을 제외한 나머жим.

도구

WPPSI-R은 WISC-R과의 비교를 용이하게 하기 위해 5가지 동작성검사와 6가지 언어성 검사의 11개 소검사로 이루어진 WPPSI와 달리 동작성검사에 모양맞추기(Object Assembly)를 첨가하여 12가지 소검사로 구성되어 있다. 그외에 검사중에 관찰한 행동이나 일상행동에 대한 부모의 보고에 따라 행동 관찰기록을 참고자료로 수집한다. 각 소검사의 종류를 표 4에 제시하였다. 각 소검사 옆에 붙은 번호는 일반적으로 검사를 실시하는 순서를 표시한다. 이 12개의 하위검사중에 문장과 동물짜짓기는 선택사항으로 지정되어 있다.

WPPSI-R에서는 다른 웨슬러 지능검사척도의 모양맞추기 검사와 동일한 그림을 사용하지만 그림을 원색으로 인쇄하였다.

예) 문항 5 : 얼굴



2. 상식

상식 소검사는 주변에서 일어나는 사건이나 물건들에 대한 지식을 알아본다. 언어표현이 부족한 어린 아동을 위하여 구두보고대신 그림을 선택하는 문항을 새로이 몇 개 첨가하였다.

표4. WPPSI-R의 소검사

동작성	언어성
1. 모양맞추기(Object Assembly)	2. 상식(Information)
3. 도형(Geometric Design)	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 짜짓기(Animal Pegs)	*12. 문장(Sentences)

주 : *표는 선택

각 소검사의 내용을 살펴보면 다음과 같다.

1. 모양맞추기

모양맞추기 소검사에서는 표준화된 형태로 나열된 여러 개의 퍼즐조각을 제시하고 제한된 시간안에 모양이 되도록 짜맞추게 한다.

나머지 문항들은 WISC-R의 상식 검사에 해당하는 것들로, 구두로 대답하게 되어 있다.

예) 구두문항 7 : 코

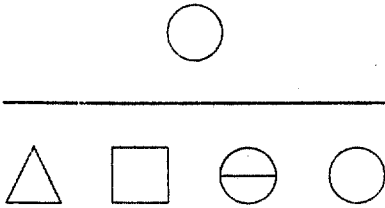
“코는 어디 있습니까?”

코를 만져보세요”

3. 도형

도형 소검사는 두 가지 유형의 과제로 구성 되어 있다. WPPSI-R에만 있는 첫 유형은 시각적 재인 과제이다. 여기서는 아동에게 단순한 도형을 보여주면서 보기로 제시된 4개의 도형 중에서 이것과 똑같은 도형을 지적하게 한다. 나머지 그리기 과제에서는 인쇄되어 있는 모형을 보고 따라서 그리게 한다.

예) 문항 1 :



4. 이해

이해 소검사는 어떤 일에서 행동의 원인과 그 결과에 대한 아동의 사고능력을 측정한다.

예) 문항 1 :

“뜨거운 난로를 만지면 어떻게 될까요?”

5. 토막짜기

토막짜기 소검사는 아동에게 제한된 시간 안에 두가지 색으로 구성된 토막들에 의해 만들어진 무늬를 보고 똑같이 재구성하게 한다.

예) 문항 1 :



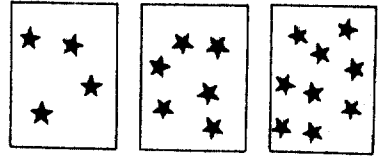
6. 산수

산수 소검사는 기본적인 수와 양의 개념에 대한 이해를 측정한다. 산수검사는 그림문항에

서 시작하여 단순한 셈하기 그리고 마지막에는 보다 어려운 응용문제로 진행된다.

예) 문항 5 :

“어디에 별이 가장 많습니까?”



7. 미로

미로 소검사는 아동에게 제한시간안에 미로를 통과해 출구를 찾아내도록 하는 지필검사이다. 맨 앞에는 어린 아동에게 적합한 간단한 미로문제들을 제시하였다.

예) 미로 3A



8. 어휘

어휘 소검사는 두 부분으로 나뉘어 진다. 첫째 부분은 그림속의 물체의 이름을 묻는 쉬운 그림문항이다. 둘째 부분은 말로 제시되는 단어들의 뜻을 구두로 대답하도록 하는 구두문항이다.

예) 문항 4 :

“칼은 무엇입니까?”

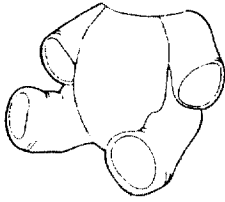
9. 빠진 곳 찾기

이 소검사에서는 일상적인 물체나 사건에 대한 그림에서 무엇이 빠졌는지를 찾아내게 한다.

예) 문항 1 :

“이 그림에서 어느 부분이 빠져있습니까?”

까?”

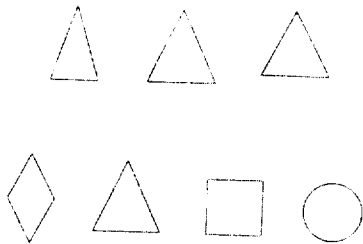


10. 공통성

공통성 소검사는 세가지 유형의 검사를 통하여 아동이 지닌 공통성에 대한 이해를 측정한다. 첫째 유형은 그림문항으로 보기의 윗줄 그림들과 공통적인 특징을 가지고 있는 것을 아래 줄의 그림중에서 골라야 한다. 그림문항에서는 말대신 손가락으로 그림을 지적하기만 하면 된다. 두번째 유형인 문장완성문항에서는 문장의 빠진 부분에 단어를 넣어 공통성이나 은유의 관계가 이루어지도록 문장을 완성해야 한다. 마지막 구두문항유형에서는 구두로 제시된 두 단어의 공통점을 말로 설명해야 한다.

예) 그림문항 1 :

“이것들과 비슷한 것은 어떤 것입니까?”



구두문항 14 :

“손가락과 포크는 어떤 점이 비슷합니까?”

11. 동물짜짓기

동물짜짓기 소검사는 보충검사로 여기서는 일련의 동물그림 밑에 있는 구멍에 맞는 색의 나무토막을 끼워야 한다. 이때 정확성과 속도를 모두 고려해서 채점한다.

12. 문장

문장 소검사도 상황에 따라 사용하는 보충검사이다. 잘 알려진 숫자 (digit span)검사의 형식을 이용해 만들었는데, 숫자대신에 단어를 사용하였다. 검사자가 문장을 큰소리로 읽어주면 아동은 이를 그대로 따라해야 한다.

예) 문항 7 :

“영희는 인형두개와 밤색 곰인형 한개를 가지고 있다”

절차

본 연구를 수행하기에 앞서 WPPSI-R의 제작자인 Psychological Cooperation과 본 검사의 지침서의 번역과 표준화, 그리고 배급을 위해 계약을 체결하였다. 그 후 지침서의 번역과 동시에 미국의 WPPSI-R을 한국의 문화적 배경을 고려하여 검사도구를 수정, 제작하였다. 특히 동작성 검사와는 달리 언어성 검사는 문화적 차이에 따른 반응의 특이성에 영향받을 것으로 예상되었다. 또한 미국판 WPPSI-R의 언어성 검사문항을 한국어로 번역하는 과정에서 문항의 속성이 변질될 가능성도 있다. 뿐만 아니라 미국판 WPPSI-R의 언어성 문항들중 본 연구의 3 세부터 7 세까지의 피험자 아동들이 100% 정답을 하거나 100% 오답을 하는 문항들이 발견될 수 있다. 이러한 문항들은 지능이 우수한 아동과 열등한 아동을 변별해내지 못하므로 결과적으로 검사의 신뢰도 및 타당도를 저하시키는 역할을 한다. 따라서 미국판 WPPSI-R의 언어성 검사문항중 본 연구의 자료에서 변별력 (문항과 총점간의 상관)이 낮게 나타난 것들을 새로운 문항으로 교환하여야 한다. 이를 위하여 미국판 WPPSI-R의 문항에 다 기존의 유사한 지능검사와 학업준비도 검사에서 발췌한 문항과 및 본 연구자들이 제작한

문항들을 자료수집전에 첨가하였고, 이 첨가된 문항들에 대한 아동들의 반응을 함께 수집하였다. 각 소점사별 첨가된 문항은 다음과 같다.

2. 상식

- 28. 무릎은 어디에 있습니까 ?
- 29. 한쪽 발에는 발가락이 몇 개 있습니까 ?
- 30. 코끼리는 다리가 몇 개입니까 ?
- 31. 1년은 몇 개월(달)입니까 ?
- 32. 빵은 무엇으로 만듭니까 ?
- 33. 눈이 오는 계절은 언제입니까 ?
- 34. 새끼 닭은 무엇이라고 부릅니까 ?
- 35. 여름 다음 계절은 무엇입니까 ?
- 36. 편지 봉투위에는 무엇을 써야 합니까 ?
- 37. 시계에는 바늘이 몇 개입니까 ?

4. 이해

- 16. 병원은 왜 필요한 것일까요 ?
- 17. 불이 나면 어떻게 해야 하나요 ?
- 18. 옷은 왜 입습니까 ?
- 19. 왜 청소를 해야 합니까 ?
- 20. 해가 나면 눈사람은 어떻게 될까요 ?
- 21. 이는 왜 닦아야 할까요 ?
- 22. 친구의 공을 가지고 놀다가 잃어버렸다

- 면 어떻게 해야 할까요 ?
- 23. 자동차는 왜 필요할까요 ?
- 24. 공부는 왜 해야 합니까 ?
- 25. 물은 끓으면 어떻게 될까요 ?

8.어휘

- 26. 육교가 무엇입니까 ?
- 27. 예방은 무슨 뜻입니까 ?
- 28. 의사는 무엇입니까 ?
- 29. 간호란 무슨 뜻입니까 ?
- 30. 태풍은 무엇입니까 ?
- 31. 소원이란 무엇입니까 ?
- 32. 결혼이란 무엇입니까 ?
- 33. 구조란 무엇입니까 ?
- 34. 경쟁하다는 무슨 뜻입니까 ?
- 35. 질투하다는 무슨 뜻입니까 ?
- 36. 교통순경은 무엇입니까 ?
- 37. 호기심이란 무엇입니까 ?
- 38. 협력이란 무엇입니까 ?

10.공통성

- 21. 양초 - 전등
- 22. 바지 - 치마
- 23. 피아노 - 바이올린

표5. 미국판 WPPSI-R에서 교체된 문항

소검사	미국판 문항	교체 문항
상식	7. 코	29. 발가락
	8. 귀	30. 코끼리 다리
이해	13. 집-아프다	24. 공부
	16. 영웅	25. 육교
어휘	17. 관람자	26. 태풍
	22. 도박	36. 교통순경
공통성	20. 행복-슬픔	23. 피아노-바이올린

- 24. 사과 - 배
- 25. 바퀴 - 공
- 26. 서랍 - 상자
- 27. 목걸이 - 반지
- 28. 강 - 바다
- 29. 야구 - 축구
- 30. 줄넘기 - 달리기

이상의 첨가문항중 원래의 문항과 난이도(오답비율)가 유사한 것중 Latent Class Analysis를 통하여 변별력이 가장 높은 것을 선택하여 최종적으로 본 연구의 분석에 포함하였다. 교체된 문항은 상식 소검사에서 2개, 이해 소검사에서 1개, 어휘 소검사에서 3개, 공통성 소검사에서 1개였다. 따라서 총 7개의 기존문항을 새로운 문항으로 교체하였는데 이를 표 5에 제시하였다.

조사자 훈련과 실시:

조사자는 총 14명으로 이들의 훈련은 각 지역별 연구책임자(서울 경인지역: 박금주; 충청중부지역: 박 광배; 영호남 남부지역: 박 혜원)와 책임조교 1명의 지도로 이루어 졌다. 모든 실시지침은 지침서에 실었고 채점상의 논란이 있을 수 있는 문제에서는 가급적 모든 응답을 기록하게 한 후 추후에 회의를 거쳐 기준을 마련하였다.

분석 (통계처리) 방법:

본 연구에서는 SAS를 이용하여 채점자간 신뢰도를 계산하였고 문항변안, 수정, 보완을 통해 구성한 각 하부척도(소검사)의 신뢰도를 확인하기 위하여 반분신뢰도를 산출하였으며 '측정의 표준오차 (standard error of measurement)'에 근거하여 내적일관도를 검증하였다. 그 외에 구성개념타당도를 위한 요인분석과 성

별 및 지역에 의한 변산을 파악하기 위한 분석을 수행하였다. 그런데 이미 표준화가 완료된 미국판 WPPSI-R의 경우에는 '환산점수 (scaled score)'와 지능지수(IQ)에 대하여 이러한 통계적 속성들이 파악되었지만, 본 연구는 한국판 WPPSI-R의 개발을 위한 예비연구이기 때문에 모집단을 편파없이 반영하면서 표준화가 가능할 정도로 규모가 큰 표본을 구하지 못하여 환산점수 및 지능지수의 산출이 불가능하였다. 따라서 여기에서 보고되는 모든 통계절차는 원점수(raw score)에 의해서만 이루어 졌다. 현재 수행중인 표준화 연구가 완료되면 여기에서 보고되는 통계치들이 환산점수와 지능지수에 의해서 재산출되고 비교될 수 있을 것이다.

결 과

본 최종결과분석에는 많은 무응답반응이나 코딩의 문제가 있는 7명의 아동의 자료를 제외하고 301명의 아동의 자료를 이용하였다. 각 분석별 결과는 다음과 같다.

1. 채점자간 신뢰도

WPPSI-R은 채점자 신뢰도(interscorer reliability)를 검증하는 것이 필수적이다. 따라서 2명의 채점자가 30명의 아동의 5가지 소검사결과를 독립적으로 평정한 후 일반화가능도를 검증하였다. 이들 소검사는 채점자의 판단에 따라 차이를 보일수 있는 것들로 미국판 검사의 경우에도 채점자간 신뢰도분석이 이루어진 이해, 어휘, 공통성, 미로 도형 소검사이다. Pearson 상관을 통한 분석결과 이해 .94, 어휘 .97, 공통성 .88, 미로 .80, 도형 .94로 높은 수준을 유지하였다.

2. 검사의 신뢰도

각 소검사의 신뢰도는 미국판 WPPSI-R과 비교를 위하여 원저자들이 사용한 반분신뢰도지수를 통해 검증하였다. 미국판 WPPSI-R의 신뢰도는 3 세부터 7 세까지의 아동을 연령에 따라 6 개월 단위로 9 개의 연령집단으로 나누어 각 연령집단에서 산출되었다. 그러나 본 연구에서는 피험자 아동의 수가 적어서 9 개의 연령집단으로 나누는 경우 각 집단의 사례수가 현저히 적어지므로 1 년 6 개월 단위로 3 개의 연령집단으로 분류한 후 각 집단 내에서의 반분신뢰도를 산출하였다. 이 신뢰도는 표 6에 제시하였다. 표 6에 제시된 반분신뢰도계수는 다음과 같은 Spearman-Brown 공식에 의해 수정된 것이다. $\rho_{rr'}$ 는 반분검사들간의 상관계수이고, ρ_{rr} 는 최종적으로 수정된 반분신뢰도계수이다.

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

또한 또한 표 6에는 측정오차의 정도를 나타내는 '측정의 표준오차(standard error of measurement)'도 제시하였다. 이 측정의 표준오차는 다음과 같은 공식에 의하여 산출되었다. σ_e 는 소검사점수의 표준편차이다.

$$\sigma_e = \sigma_w \sqrt{1 - \rho_{rr}}$$

표 6의 신뢰도 계수는 전체적으로 미국판 WPPSI-R의 그것과 매우 유사한 양상을 보이고 있다. 가장 낮은 신뢰도계수는 0.56 이고 가장 높은 것이 0.89이다. 미국판의 경우 신뢰도계수의 범위는 0.54-0.90 이다.

미국판(.63)과 마찬가지로 모양맞추기 소검사의 신뢰도(.64)가 다른 소검사들의 신뢰도보다 낮은데, 그 한 이유는 모양맞추기 소검사는 문항수가 6 개 밖에 지나지 않아서 다른 소검사들에 비하여 적기 때문이다. 또한 미국판의 경우도 모양맞추기 신뢰도가 4.5 세부터 5.5 세 연령집단에서 가장 낮은데(.56), 본 연구의 자료에서도 유사한 양상(.56)이 발견되고 있다. 언어성 검사들의 평균신뢰도는 본 연구에서 제작한 한국판과 원래의 미국판이 거의 동일한 것으로 나타났다. 그러나 동작성 검사들의 경우에는 한국판의 평균 신뢰도가 미국판의 평균 신뢰도보다 약간 높다.

표 6에 제시된 측정의 표준오차는 산수소검사에서 가장 작고 도형소검사에서 가장 큰 수치를 보이고 있다. 그러나 세 연령집단간에는 매우 유사한 결과를 보여주고 있다.

표6. 소검사의 연령별 신뢰도 계수와 측정의 표준오차

소검사	연령집단						평균 신뢰도
	3.0세-4.0세(N=93)		4.5세-5.5세(N=101)		6.0세-7.0세(N=107)		
	신뢰도	표준오차	신뢰도	표준오차	신뢰도	표준오차	
모양맞추기	.70	1.82	.56	2.56	.68	2.64	.64
도형	.90	3.72	.81	4.47	.67	3.54	.81
토막짜기	.88	2.12	.81	2.44	.83	2.63	.84
미로	.86	1.67	.78	1.68	.80	1.55	.82
빠진곳 찾기	.87	1.62	.71	1.63	.78	1.35	.80
동물 짝짓기	-	-	-	-	-	-	-
상식	.83	1.71	.84	1.78	.83	1.81	.84
이해	.85	1.98	.77	2.74	.85	2.32	.83
산수	.87	1.02	.83	1.37	.78	1.20	.82
어휘	.80	2.01	.84	2.12	.85	2.88	.83
공통성	.79	1.42	.81	1.98	.86	2.21	.82
문장	.87	2.14	.81	2.74	.88	2.31	.86

표 7. 연령집단별 소검사의 요인구조

연령집단	소검사	요인 1	요인 2	요인 3	Communality
3.0세 - 4.0 세 (N=93)	모양맞추기	0.22820	0.73950	0.36478	0.7320
	도형	0.19575	0.82389	-0.22562	0.7680
	토막짜기	0.14097	0.75951	0.01142	0.5969
	미로	0.30377	0.71562	0.18525	0.6387
	빠진곳 찾기	0.23794	0.73649	-0.07824	0.6051
	동물 짝짓기	0.04121	-0.01243	0.92441	0.8746
	상식	0.74129	0.41445	0.24073	0.7792
	이해	0.69497	0.16936	0.27215	0.5857
	산수	0.75170	0.39244	-0.03340	0.7202
	어휘	0.71705	0.26741	-0.15978	0.6112
	공통성	0.71249	0.12834	0.01429	0.5243
	문장	0.77699	0.09929	0.21980	0.6619
	Eigenvalue	3.50216	3.30949	1.28624	
	설명변량의 비율	29.2%	27.6%	10.7%	
4.5 세 - 5.5 세 (N=101)	모양맞추기	0.16035	0.53524	0.42269	0.4909
	도형	0.25559	0.70726	-0.18418	0.5995
	토막짜기	0.26947	0.70740	0.14474	0.5940
	미로	0.07543	0.70444	-0.07958	0.5083
	빠진곳 찾기	0.21045	0.62320	0.32313	0.5371
	동물 짝짓기	0.05991	-0.00105	0.83073	0.6937
	상식	0.80078	0.33943	0.03342	0.7576
	이해	0.78141	0.12905	0.22118	0.6762
	산수	0.65805	0.38666	-0.16074	0.6084
	어휘	0.84926	0.07190	0.16512	0.7537
	공통성	0.69901	0.17779	0.30331	0.6122
	문장	0.68418	0.21590	-0.21932	0.5628
	Eigenvalue	3.58005	2.53651	1.27764	
	설명변량의 비율	29.8%	21.1%	10.6%	
6.0 세 - 7.0 세 (N=107)	모양맞추기	0.16268	0.59057	0.45887	0.5858
	도형	0.43618	0.66516	-0.15745	0.6574
	토막짜기	0.41270	0.72627	0.01464	0.7023
	미로	0.18399	0.78224	-0.07637	0.6515
	빠진곳 찾기	0.25470	0.77618	-0.12010	0.6817
	동물 짝짓기	0.10158	-0.14198	0.85603	0.7726
	상식	0.82010	0.35266	0.07601	0.8027
	이해	0.78977	0.18986	0.11571	0.6731
	산수	0.61359	0.42377	-0.20442	0.5978
	어휘	0.79714	0.12565	0.14989	0.6736
	공통성	0.72550	0.37733	0.07277	0.6740
	문장	0.70796	0.24231	-0.00028	0.5599
	Eigenvalue	3.83185	3.12385	1.07733	
	설명변량의 비율	31.9%	26.0%	8.9%	

12개의 소검사들중 동물 짝짓기는 속도검사(speed test)이고, 따라서 반분신뢰도가 적용되지 않는 검사이므로 여기에서는 그것의 신뢰도 지수와 측정오차가 산출되지 않았다. 표 6에는 각 소검사의 평균신뢰도계수 또한 제시하였는데 이것은 각 소검사의 연령별 신뢰도계수 세 개를 Fisher 의 z 통계치로 변환하고, 그들의 가중평균을 구한 후, 그것을 다시 신뢰도 계수로 역변환한 것이다.

3. 요인구조

WPPSI-R의 소검사들은 이론적으로 동작성 검사들과 언어성검사들로 구분된다. 이러한 이론적인 구분이 실제로 존재하는 지를 검증하는 것은 '구성타당도 (construct validity)'의 검증이 된다. 따라서 12 개의 소검사점수에 대하여 각 연령집단별로 요인분석을 실시하였다. 요인 추출방식은 주축요인방식 (principal axis factoring)을 이용하였고, 요인의 수를 정하는 기준은 1 보다 큰 eigenvalue 를 이용하였다. 요인부하는 varimax 회전후 산출된 것이고, 각 요인이 설명하는 변량 (eigenvalue) 의 비율은

전체 변량에 대한 비율이다. 결과를 표 7에 제시하였다.

표 7의 요인구조를 보면 모든 연령집단에서 동작성 요인과 언어성 요인이 명확하게 구별된다. 모든 연령집단에서 제 1 요인이 언어성 요인이고, 제 2 요인이 동작성 요인이다. 그런데 세번째로 '동물 짝짓기'요인이 추출되었는데, 이 요인이 다른 동작성 검사들과 구별되는 이유는 이것이 유일하게 '속도검사 (speed test)' 이기 때문인 것으로 사료된다. 세개의 요인은 모든 연령집단에서 전체 변량의 60% 이상을 설명한다.

4. 성별 및 지역차이

성별과 지역에 따른 지능의 차이를 알아보기 위하여 각 소검사 점수를 종속변인으로 하고, 성별과 지역을 각각 독립변인으로 하는 단순회귀분석을 하였다. 성별과 지역이 서로 독립적이므로 성별과 지역을 모두 독립변인으로 하는 중다회귀분석을 무의미하여 수행하지 않았다. 지역은 서울, 청주, 울산의 세군데를 두개의 무의미 변인(dummy variable)으로 코딩하여 그

표8. 소검사 원점수의 성별차이

소검사	여자			남자			F
	N	평균	SD	N	평균	SD	
모양맞추기	153	20.75	5.43	148	20.80	6.09	ns
도형	153	46.73	14.18	148	42.26	16.09	6.52**
토막짜기	153	23.58	8.51	148	23.21	10.03	ns
미로	153	16.75	5.65	148	15.97	5.45	ns
빠진곳 찾기	153	19.24	4.50	148	18.21	4.66	ns
동물 짝짓기	153	16.95	7.51	148	16.55	7.27	3.78*
상식	153	14.95	6.36	148	14.16	6.90	ns
이해	153	13.67	7.80	148	13.83	8.23	ns
산수	153	15.69	5.00	148	15.20	5.32	ns
어휘	153	14.65	7.00	148	14.80	8.58	ns
공통성	153	10.88	6.12	148	10.04	6.34	ns
문장	153	17.74	7.88	148	16.21	8.20	ns

주: *p<.05 **p<.01

효과를 검증하였다. 표 8과 표 9에 그 결과를 제시하였다.

표 8에 의하면 지능에 대한 성별의 효과는 없다. 언어성 검사와 동작성 검사 모두에서 여아가 남아보다 약간 우세한 경향이 있으나 '도형' 소검사를 제외하고는 통계적으로 유의하지 않다. 지역에 따른 차이를 보면 모든 소검사에 걸쳐서 울산지역이 서울과 청주보다 낮은 점수를 보이고 있다. 반면에 서울과 청주 사이에서는 큰 차이가 발견되지 않았다. 울산지역의 아동들이 낮은 점수를 보이는 이유중의 하나는 표 1에 제시된 바와 같이 이 지역에서는 원점수상의 수행이 낮은 3세에서 5세반까지의 어린 연령집단이 다른 지역에 비하여 많이 참여하였기 때문이다. 본 연구에서는 아동의 지능검사의 수행을 지능(IQ)이라는 환산점수로 계산한 것이 아니라 원 점수를 사용하였기 때문에 어린 아동일수록 지적 수행이 낮게 나오는 경향을 보였다.

논 의

지능검사가 아동의 발달에 대한 매우 중요한 정보를 제공해 준다는 것은 주지의 사실이다. 아동을 대상으로 하는 교육학이나 심리학등의 많은 연구에서 지능검사가 사용되는데 지적인 능력외에 정서 상태나 심리장애의 진단에도 지능검사의 유용성은 널리 알려져 있다. 그러나 매우 어린 아동의 지능을 측정하는 데에는 실시상의 어려움뿐 아니라 측정된 지능의 안정성이 낮다는 문제점이 제기되고 있다. 본 연구는 1939년 제작된 웨슬러 성인용 지능검사로부터 응용되어 3세의 어린 아동부터 7세까지의 아동의 지능을 측정하도록 개발되어 있는 웨슬러 유아지능검사를 표준화하기 위한 도구를 개발하려는 목적을 갖고 있다. 이를 위하여 전국적으로 3개 지역을 중심으로 301명의 3세에서 7세 사이의 아동을 대상으로 번역, 수정된 WPPSI-R을 실시하였다.

표9. 소검사 원점수의 지역차이

소검사	서울			청주			울산			F
	N	평균	SD	N	평균	SD	N	평균	SD	
모양맞추기	83	22.25	5.85	108	21.00	5.64	110	19.43	5.52	6.04**
도형	83	44.06	16.72	108	44.03	15.95	110	45.37	13.49	ns
토막짜기	83	25.71	8.56	108	22.26	9.35	110	22.75	9.48	3.71**
미로	83	17.21	5.16	108	16.26	5.88	110	15.68	5.46	ns
빠진곳 찾기	83	19.13	4.47	108	18.25	4.67	110	19.90	4.61	ns
동물 짝짓기	83	18.85	3.45	108	19.99	0.09	110	12.02	10.21	47.67**
상식	83	15.30	6.68	108	15.47	6.28	110	13.02	6.74	4.30**
이해	83	16.66	7.39	108	14.98	7.80	110	10.35	7.40	18.73*
산수	83	16.18	5.28	108	15.25	4.99	110	15.09	5.19	ns
어휘	83	16.62	9.00	108	15.47	7.58	110	12.55	6.47	7.57**
공통성	83	11.97	7.27	108	11.53	5.70	110	8.27	5.24	11.61**
문장	83	17.43	7.30	108	18.42	8.02	110	15.24	8.38	4.55**

주: *p<.05 **p<.01

가장 먼저 살펴본 것은 본 지능검사의 소검사중 까다로운 채점절차가 요구되는 5개 소검사에 대해 두명의 서로 다른 지역에 근무한 채점자가 30명의 아동의 자료를 반복 채점하였다. 검사원점수를 이용하여 Person 상관계수를 계산한 결과 .80 - .97사이의 높은 신뢰도를 나타내었다. 그러나 소검사간의 신뢰도 계수상에 미국의 경우(.88 - .96)보다는 큰 변이를 보이고 있다. 미국판의 경우는 이해 .96, 어휘 .94, 공통성 .96, 미로 .94, 도형 .88로 공통성과 미로의 경우 본 연구에 사용된 검사의 신뢰도가 낮고 도형의 경우는 한국의 자료가 더 우수하였다. 그러나 이 문제의 해석은 본 연구에서는 151 - 188명의 자료를 사용한 미국의 분석과는 달리 30명이라는 적은 수의 아동을 대상으로 산출하였기 때문에 후속 표준화 연구에서 더 명확히 논의될 수 있을 것이다.

각 소검사의 내용을 분석하기 위하여 Spearman-Brown 반분신뢰도를 살펴본 결과 각 소검사의 신뢰도 계수는 전체적으로 미국판 WPPSI-R의 것과 매우 유사하였다. 그런데 측정의 표준오차는 미국판의 경우보다 높은 수치를 보이고 있다(Wechsler, 1990). 그러나 이것이 한국판이 미국판보다 측정오차를 더 많이 유발한다는 것을 의미하는 것은 아니다. 왜냐하면 신뢰도지수에서 한국판이 미국판보다 결코 더 열등하지 않을 뿐만 아니라 경우에 따라서는 더 우수하기 때문이다. 미국판의 측정오차는 위에 제시된 공식을 위하여 환산점수의 표준편차를 이용하였고, 본 연구에서는 원점수의 표준편차를 이용하여 산출하였다. 환산점수의 범위(range)는 원점수의 범위보다 현저하게 작다. 예를 들어 토막짜기의 원점수가 42인 경우 이것을 환산점수로 환산하면 19가 된다. 따라서 원점수의 표준편차가 환산점수의 표준편차보다 클 수 밖에 없고, 이것이 한국판에서

산출된 측정의 표준오차를 미국판의 그것보다 커지게 한 요인이다. 후속연구에서 한국판의 표준화가 이루어지면 환산점수에 의한 측정의 표준오차가 산출될 것이다.

가장 중요한 분석의 하나인 요인 분석의 결과 미국의 자료는 동물 짝짓기 요인이 다른 동작성 검사들과 구별되지 않았는데, 본 연구에서는 언어성 요인과 동작성 요인외에 동물짝짓기 소검사요인이 밝혀졌는데 이러한 요인구조는 미국판의 그것보다 더 합리적이다. 왜냐하면 속도검사는 역량검사(power test)와 척도속성이 매우 달라야 마땅하기 때문이다(Allen & Yen, 1979).

그외에 지능에 대한 전반적인 성별의 효과는 없는 것으로 나타났다. 언어성 검사와 동작성 검사 모두에서 여아가 남아보다 약간 우세한 경향이 있으나 '도형' 소검사를 제외하고는 통계적으로 유의미하지 않다. 성별의 차이가 나타나지 않은 이유의 하나는 나이가 어린 취학 전 아동들이기 때문인 것으로 생각된다. 도형 소검사에서 남여차이 또한 후속 표준화 연구에서 재검토되면 보다 명확한 해석이 가능해질 것이다.

지역별 분석에서는 울산지역의 아동들이 낮은 점수를 보였는데 그 이유중의 하나는 이 지역에서는 앞에서도 언급되었듯이 원점수상의 수행이 낮은 3세에서 5세반까지의 어린 연령 집단이 다른 지역에 비하여 많이 참여하였기 때문이다. 이점외에 다른 지역적인 차이가 있는지는 각 아동의 점수를 환산 점수로 바꾸어 분석하고, 각 지역의 특성이 반영될 수 있을 정도의 보다 많은 아동의 자료가 표집되면 분명히 밝혀질 것으로 생각된다. 후속 연구에서는 확정된 검사에 대한 타당도 검증과 표준화에 초점을 두고자 한다.

참 고 문 헌

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Ca.: Brooks-Cole.
- Bornstein, M. H., & Sigman, M.D. (1986). Continuity in mental development from infancy. *Child Development*, 50, 251-274.
- 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.(1990). *Manual for the Wechsler Preschool and Primary scale of intelligence - Revised*. New York: The Psychological Co.

The standardization of WPPSI-R: A pilot study

Hyewon Park Choi

Univ. of Ulsan

Keum-Joo Kwak

Osan College

Kwang-Bae Park

Choongbuk Univ.

As a pilot study for the development of Korean version of WPPSI-R, 301 preschool and primary school children were tested with a revised WPPSI-R in three locations (Seoul, Chung-joo, and Ulsan). The pilot version of the test was prepared based on the translation of the original material(manual and the test items) as well as the addition of new items for subtests of Information, Comprehension, Vocabulary, and Similarities. Interscorer reliability coefficients measured by two independent scorers with 30 children's data were satisfactory: .94 on Comprehension, .97 on Vocabulary, .88 on Similarities, .80 on Mazes and .94 on Geometric Design. Spearman-Brown split-half reliability coefficients for 11 subtests were found very similar to those for the original WPPSI-R. The reliability estimate for 'Animal Pegs' was not obtained since the split half technique is inappropriate for estimating the reliability of speeded tests. All coefficients were above .80 except the Object Assembly(.56). Another index of the consistency of scores, the standard error of measurement was also obtained. Factor analyses revealed three significant factors, the performance factor, the verbal factor as well as the 'animal pegs' factor which explained over 60 % of total variance in all age groups. Girls performed slightly higher than the boys in all subtests especially in Geometric Design and Animal Pegs. This study developed test material and the standardization study for the IQ norms and other validity studies will be followed.