

한국 웨슬러 지능검사(KWIS) 단축형의 정확성 및 타당도에 관한 연구

- WARD7형과 Doppelt형의 비교 -

임영란† · 이우경 · 이원혜 · 박종원

용인정신병원

웨슬러 지능 검사(Wechsler, 1981)의 단축형은 지능 검사를 전체로 실시하는데 소요되는 시간을 절약 시켜 주기 때문에 주의력이나 동기 저하 등의 문제를 보이는 정신분열병 환자들에게 적용할 경우 그 유용성이 크다고 볼 수 있다. 지금까지 여러 유형의 단축형 지능검사가 개발되었지만 단축형 추정치를 통해 전체 지능을 예측할 때 그 정확성이나 타당도가 문제가 되곤 하였다. 따라서 본 연구의 목적은 정신분열병 환자를 대상으로 두 단축형 지능 검사를 통한 지능 추정의 정확도와 타당도를 비교 분석해 보기 위한 것이다. 연구를 위해 105명의 정신분열병 환자들의 한국 웨슬러 지능검사(KWIS)를 7개의 소검사(상식, 산수, 공통성, 숫자, 바퀴쓰기, 빠진곳찾기, 토막 짜기)로 구성된 WARD7 단축형과 4개의 소검사(어휘, 산수, 토막 짜기, 차례 맞추기)로 구성되어 있는 Doppelt 단축형으로 전체 지능을 각각 추정해 본 뒤 그 정확성을 비교 검토해보았다. 연구 결과, 두 단축형 지능과 전체 지능간에 매우 높은 상관을 보이고 있어 단축형 지능검사를 통한 전반적인 지능 추정이 상당히 정확하고 신뢰롭다는 것이 밝혀졌다. WARD7 단축형과 Doppelt 방식을 비교한 결과, WARD7 단축형의 추정 오차 평균은 ($M=2.43$) Doppelt방식($M=4.31$)에 비해 유의하게 작았다. 또한 WARD7 단축형에서 86.7%가 전체 지능과 ± 5 오차 범위에 있었던 반면 Doppelt 방식에서는 65.7%가 전체지능과 ± 5 오차범위에 있었고 추정 오차크기가 6점보다 큰 경우는 WARD7단축형, Doppelt방식에서 각각 2사례(1.9%), 20사례(19%)에 해당되었다. 더욱이 WARD7 단축형은 표준화된 지능 범주에 대한 분류오류율이 9.5%로 Doppelt형에(35%) 비해 더 낮게 나타났다. 따라서 본 연구 결과, WARD7 단축형 지능 검사가 Doppelt형에 비해 정신분열병 환자들의 지적 기능을 더 정확하고 타당하게 예측해주는 것으로 나타났다. 이러한 연구 결과에 대한 의미와 본 연구의 제한점 및 향후 연구과제에 대해 고찰하였다.

† 교신저자(Corresponding Author) : 임영란 / 용인정신병원 경기도 용인시 구성면 상하리 4번지 / Fax : 031-288-0184 / E-mail : limpsy@hotmail.com

성인용 웨슬러 지능검사(Wechsler Adult Intelligence-Revised: WAIS-R, 1981)는 성인 지능을 측정하는데 있어 지금까지 가장 신뢰로운 도구로 평가받고 있다. 그러나 전체 소검사를 실시하는데 드는 시간적, 인적 효율성이 종종 문제가 되곤 한다. 특히 정신과적 환자들을 위해 재할 계획을 수립하고 장애 정도를 파악할 때, 전반적인 지적 기능을 간편하게 추정하는 것은 매우 중요하다. 만성 정신 질환자들의 경우 주의 집중력의 저하, 피로감, 동기 부족으로 인해 2 시간 가량 소요되는 전체 소검사를 한 번에 모두 실시하기가 어려울 때가 많다(Ryan & Rosenberg, 1984). 따라서, 이러한 피검자들을 위해서 전반적인 지적 기능을 정확하게 추정할 수 있는 단축형 지능 검사 도구가 개발될 필요성이 대두되었다.

단축형 지능검사가 처음 사용되기 시작한 것은 이미 오래 전부터이다. Terman과 Merrill(1937)은 제한된 시간으로 인해 전체 소검사를 실시하기 어려울 경우 Stanford-Binet의 단축형을 사용할 것을 제안하였고 이에 따라 연령수준을 고려한 네 가지 과제가 사용되었다.

이후 Doppelt(1956)의 주도하에 많은 연구자들이 가장 정확한 단축형을 확인하려고 노력해왔다. 웨슬러 지능검사의 단축형들은 대부분 전체 11개의 소검사들 중 몇 개의 소검사들을 채택하여 간편하게 실시하도록 구성되어 있다. 신뢰도(Cyr & Brooker, 1984)와 같은 기준들이 강조되기도 하였지만 단축형 지능 검사를 구성하는 하위 소검사를 결정할 때 가장 중요한 기준은 단축형의 타당도 계수(coefficients)의 크기이다.

한편 요인분석을 통하여 WAIS의 구성 요인을 밝히고 이러한 결과를 바탕으로 구성 타당도를 검토하려는 시도가 있어왔다. 지능 검사가 어떤 구성 요소로 되어 있는가라는 문제는 이론적 관점뿐만 아니라, 지능 검사의 프로파일을 해석하는 임상가의 입장에서도 매우 중요한 문제이다. 연구자들은 지능 검사의 요인 구조를 탐색해 보는 것이 소검사 점수를 의미있게 해석하는 근거가 될뿐만 아니라 단축형 지능검사를 개발하는데 기초 자료가 된다고 보고 있다. WAIS-R의 요인구조는 정상 집단과 비정상 집단의 표집, 연

령, 성, 인종 집단에 따라 크게 두개 또는 세개 요인으로 밝혀져 왔다(Atkinson & Cyr, 1984; Beck, Horwitz, Seidenberg, Parker, & Frank, 1985; Gutkin, Reynolds, & Galvin, 1984; Parker, 1983; Ryan, Rosenberg, & DeWolfe, 1984; Silverstein, 1982). 두개 요인은 언어성 소검사들과 동작성 소검사들로 밝혀지고 있고 세개 요인일 경우 언어적 이해 능력(verbal comprehension), 지각적 조직화 능력(perceptual organization), 주의 지속능력(freedom from distractibility)으로 밝혀지고 있다. 언어적 이해 요인에는 상식, 어휘, 이해, 공통성 문제가 포함되며, 지각적 조직화 요인에는 빠진곳찾기, 차레맞추기, 토막짜기, 모양맞추기가 해당되고, 주의 지속 요인에는 산수, 숫자, 바꿔쓰기 등이 포함된다. 이밖에 요인분석을 통해 하위 소검사들과 지능의 일반 요인(g factor)의 관련성 그리고 하위 소검사 각각이 측정하는 독특한 심리 측정적 특징(s factor)을 알 수 있다.

요인분석에 관한 국내의 연구들을 살펴보면, 정상 집단에서는 3개 요인(박상규, 홍창희, 김재환, 1992)과 4개 요인(이창우, 1964)을 얻었으며 정신분열병 환자 집단에서는 3개 요인(박상규 등 1992; 채영숙, 김현정, 오상우, 1992)과 5개 요인(한덕웅, 1975)을 얻었다. 또한 정신병 집단과 신경증 집단(최미례와 김중술, 1990), 비정신분열병인 정신과 환자 집단(박상규 등, 1992)에서는 모두가 3개 요인으로 분석되었다. 최근 정신분열병 환자를 대상으로 한 오상우의 연구(1997)에서는 1요인을 얻었다.

이와같은 지능 검사의 요인 구조는 전체 지능 검사가 측정하고 있는 구성 개념을 가장 잘 반영하는 단축형 소검사를 개발하려고 할 때, 타당도에 대한 이론적 근거를 제공해준다. 그러나 단축형 IQ와 전체 IQ간의 타당성을 가늠하는 실제적 준거들은 Resnick와 Entin(1971)이 제시한 바 있다. 이들은 첫째, 단축형 IQ와 전체 IQ간의 상관성이 높아야 하고, 둘째, 단축형 IQ의 평균과 전체 IQ의 평균간에는 유의한 차이가 없어야 하며, 셋째, 피험자들은 각 지능 수준으로 범주화하는데 있어 단축형과 원래 검사간에 큰 차이가 없어야 한다고 주장하였다. 이러한 준거에 따라

정신과 환자들에게 단축형이 얼마나 타당하게 적용될 수 있는지에 대해 많은 연구들이 이루어졌다.

Silverstein(1982, 1884, 1985)은 두개 또는 네개 하위 소검사들로 이루어진 단축형 검사들을 개발하였다. 이중 어휘, 토막짜기 두가지 소검사로 이루어진 단축형은 심리측정적인 효용가치가 있으며, 어휘, 토막짜기, 산수, 차례맞추기 네가지 하위 소검사로 구성되어있는 단축형은 통계적인 지표들이 훌륭하다는 평가를 받고 있기는 하나 피검자의 능력 수준을 잘못 분류하는 비율이 높아 임상 장면에서 사용하는데에는 한계가 있다. Thompson(1987)은 Silverstein의 단축형 검사를 사용하여 연구한 결과, 이러한 단축형들이 전체 지능을 과잉추정한다는 것과 시간이 너무 많이 걸린다는 제한점이 있다고 보고하였다. Silverstein의 단축형에 대한 좋은 대안으로 Reynolds, Willson, Clark(1983)는 상식, 산수, 빠진곳 찾기, 토막짜기 네가지 소검사로 이루어진 단축형 지능 검사를 개발하였으나 그 타당도는 널리 입증되지 않았다. 뒤이어 Kaufman, Ishikuma, Kaurman-Packer(1991)는 두 개 소검사(상식, 빠진곳찾기), 세 개 소검사(상식, 빠진곳찾기, 바뀔쓰기) 네 개 소검사(공통성, 산수, 빠진곳찾기, 바뀔쓰기)로 이루어진 심리측정적으로 타당한 단축형 지능 검사들을 개발하였다. 이밖에 Feingold(1982)는 상식, 어휘 소검사로 이루어진 단축형을 개발하였으나 언어성 소검사 위주로 되어 있어 그다지 널리 사용되지 않았다. Satz와 Mogel(1962) 역시 일찍이 숫자, 바뀔쓰기로 이루어진 단축형을 개발하였으나 균형을 적용하는데 문제가 있어 임상적인 유용성이 떨어진다는 지적을 받았다. 또한 언어성 소검사들만으로 구성되어 있는 Slosson 지능 검사(Slosson, 1982)와 자기 보고식의 어휘, 공통성 검사로 이루어져 있는 Shipley Hartford(1940) 지능 검사 등 여러 가지 단축형 지능 검사들이 개발되었지만 이중에서 Silverstein 방식(1982), Kaufman 방식(1972), Reynolds 방식(1983) 등이 임상적으로 가장 폭넓게 사용되어왔다.

국내에서는 Doppelt형 지능 검사가 김중술, 김영환(1974)에 의해 처음 소개되었고 최근 김중술, 이용승, 이민식(1993)은 K-WAIS의 원자료를 분석하여 두가

지 소검사를 이용한 IQ 추정치와 네가지 소검사를 이용한 IQ 추정치를 제시하면서 대략적인 IQ를 알고자 할 때에는 이러한 단축형 검사를 이용한 추정이 매우 타당하다고 보고하였다.

이중 우리나라에서 현재까지 가장 널리 사용되어온 Doppelt 단축형 검사는 어휘, 산수, 토막짜기, 차례맞추기 등 4가지 소검사로 이루어져 있는데, 지능 검사의 3개 요인 구조에 따르면 언어적 이해(어휘), 지각적 조직화(토막짜기, 차례맞추기), 주의력(산수)을 측정한다고 볼 수 있다. 또한 전체 WAIS-R 실시 시간의 약 47% 정도(30.4-43.2분)가 소요된다. 하지만 임상적인 경험을 통해 볼 때, 이러한 방법으로 추정된 IQ가 실제 IQ와 유의한 차이를 보이는 경우가 많다. 또한 최근 연구가 활발한 신경심리검사들과의 비교를 위해서 단축형을 실시하더라도 단순히 대략적인 전체 지능만을 추정하는 것보다 정보 가치가 높은 소검사들의 수행을 비교할 필요성이 대두되게 되었다.

비교적 최근 개발된 Ward(1990)의 단축형 지능 검사는 WAIS-R의 소검사들 중 상식, 숫자, 산수, 공통성, 빠진곳찾기, 토막 짜기, 바뀔쓰기의 7개 소검사로 구성되어 있다. 요인 구조로는 언어적 이해(상식, 공통성), 지각적 조직화(토막짜기, 빠진곳찾기), 주의력(산수, 숫자, 바뀔쓰기)을 측정하고 있다. 더욱이 지능의 하위 소검사들중 가장 높은 특수성(specificities)을 지닌 숫자, 바뀔쓰기 소검사가 포함되어 있어 주로 지능의 일반요인(g factor) 부하량이 높은 소검사들로 구성되어 있는 Doppelt형에 비해 일반 요인, 특수 요인(s factor)을 골고루 갖추고 있다는 점에서 임상적으로 의미있는 정보를 제공해준다고 볼 수 있다. 또한 연구 결과에 따르면 WARD의 단축형은 검사 소요 시간을 약 45-50% 단축시켰을 뿐만 아니라(Ryan et al., 1984; Ward, Sleby, & Clark, 1987), 언어성 지능, 동작성 지능, 전체 지능(Axelrod, Woodard, Schretlen, & Benedict, 1996)을 각각 추정해주는 신뢰로운 예측치를 제공해주고 있다. 정신과 환자들의 지능평가시 가장 많이 쓰이는 단축형 지능 검사들의 정확성을 비교한 최근의 연구 결과(Allen, Huegel, Gurklis, Kelley, Barry, & Kammen, 1997), 진단적으

로 이질적인 정신과적 표집을 비롯하여 정신분열병 집단 혹은 정신분열형 정동장애 집단에서 Ward의 단축형 소검사가 WAIS-R의 가장 신뢰롭고 타당한 단축형 소검사로 밝혀졌다(Benedict, Schretlen, & Bobholz, 1992; Abraham, Axelrod, & Paolo, 1997).

위의 연구 결과를 바탕으로, 본 연구의 목적은 만성 정신분열병 환자들을 대상으로 하여 WARD7 단축형 지능검사로 IQ를 추정하는 것이 얼마나 정확한지를 평가하기 위한 것이다. 또한 WARD7 단축형에서 나온 추정치와 현재까지 임상장면에서 가장 널리 쓰이고 있는 Doppelt(1956) 단축형 추정치를 그 정확성 및 타당도면에서 비교검토하고자 한다.

연구방법

연구 대상

본 연구의 피험자는 1998년 3월부터 1999년 12월까지 용인정신병원에 입원중인 환자들 중, DSM-IV (American Psychiatric Association, 1994)에 따라 정

신분열병 진단에 해당되는 환자들을 대상으로 하였다.

전체 인원은 105명으로 남자 63명, 여자 42명이었고 연령분포는 18세에서 53살까지로 평균 연령은 35.90세, 표준편차는 8.60세였다. 첫 발병 연령은 26.02세 였으며, 표준 편차는 8.01세 였다. 전체 지능 지수 분포는 66에서 133까지 였으며 평균 지능은 95.90이었고 표준편차는 13.08이었다.

연구절차

KWIS의 표준화된 실시 지침에 따라 각 환자들에게 개별적으로 검사를 시행하였다. 검사자는 대학원 심리학과에서 임상 심리학 전공으로 석사학위를 받고 용인정신병원에서 임상심리전문가의 지도 감독하에 수련을 받고 있는 사람들이었다. WARD7형은 언어성 지능(VIQ), 동작성 지능(PIQ), 전체 지능(FIQ)을 모두 추정하는 것이 가능하나 Doppelt 방식은 전체 지능(FIQ)만 추정하게 되어 있다.

표 1에는 두 단축형을 구성하는 소검사 항목, 계산 공식, 평균 소요 시간 및 추정가능한 지능이 제시되

표 1. 단축형 소검사의 구성

단축형	소검사 구성	계산 공식	소요시간(분)	추정치
WARD7 (Ward, 1990)	상식, 공통성, 산수, 숫자, 바퀴 쓰기, 빠진 곳찾기, 토막짜기	언어성 점수: 2(상식+공통성)+산수+숫자	36.8 - 45.4	VIQ
		동작성 점수: 2(빠진곳찾기+토막짜기) +바퀴쓰기		PIQ
Doppelt 방식 (Doppelt, 1956)	산수, 어휘, 토막 짜기, 차례맞추기	전체점수: 언어성 점수+동작성 점수	30.4- 43.2	FIQ
		15-19 세: 2.7(산수+어휘)+1.6(차례+토막)+18		
		20-24 세: 2.55(산수+어휘)+1.74(차례+토막)+21		
		25-34 세: 2.38(산수+어휘)+2.57(차례+토막)+9.5		
		35-55 세: 3.84(산수+어휘)+1.49(차례+토막)-3		

어 있다.

자료분석방법

자료 처리 절차를 구체적으로 기술하면 다음과 같다. 우선 환자 105명의 KWIS 자료를 가지고 전체 KWIS와 WARD 단축형 IQ로 환산한 뒤 언어성 지능, 동작성 지능, 전체 지능 각각에 대해 상관 계수를 구하고 평균치 차이에 대해 paired t-test를 하였다. Doppelt형에 대해서는 전체 지능만 추정된 뒤 마찬가지로 상관계수와 평균치 차이 검증을 하였다. 이어서 연령 수준에 따른 전체 KWIS와 WARD7형, Doppelt 방식으로 추정된 지능의 상관 관계를 비교 분석해 보았다. 또한 전체 KWIS의 FIQ와 두 단축형의 절대 오차 분포를 비교한 뒤 절대 오차에 대해 t검증을 하였다. 마지막으로 Wechsler의 표준 분류 방식에 따라 전체 KWIS의 FIQ와 두 단축형의 FIQ추정치를 분류한 뒤 분류 오류 비율을 빈도분석하였다.

모든 KWIS 자료는 SPSS PC+ Version 8.0을 이용하여 통계 처리를 하였다.

결 과

전체 KWIS와 WARD7형의 FIQ, VIQ, PIQ의 평균 및 표준 편차 그리고 두검사간의 상관 계수가 표 2에 제시되어 있으며 전체 KWIS와 Doppelt형의 FIQ 평균 및 표준편차 그리고 상관계수는 표 3에 제시되어 있다. 표 2에서 알 수 있듯이 전체 KWIS의 언어

성 지능과 WARD7형의 VIQ와의 상관은 .96, 전체 동작성 지능과 WARD7형의 PIQ와의 상관은 .96, FIQ와의 상관은 .97 이었다. 반면 전체 KWIS와 Doppelt 방식에서의 FIQ와의 상관은 .95였다. 전체 KWIS와 단축형 지능검사간의 평균치 차이에 대한 paired t-test 결과, WARD7형의 PIQ와 전체 KWIS의 PIQ에서는 유의한 차이가 발견되지 않았다($t=1.34, p>.15$). 하지만, 전체 KWIS와 WARD7형의 VIQ($t=-6.57, p<.01$), FIQ($t=-3.32, p<.01$) 및 전체 KWIS와, Doppelt 방식에서의 FIQ($t=6.18, p<.001$)는 유의한 차이를 보였다.

연령 수준에 따른 전체 KWIS와 WARD7의 VIQ, PIQ, FIQ 및 Doppelt 방식에서의 FIQ의 평균, 표준 편차 및 상관 계수를 표 4와 표 5에 제시해 놓았다. 연령 집단별 전체 KWIS와 WARD7형의 VIQ, PIQ, FIQ의 상관은 .90-.99로 높게 나타났다. 또한, paired t-test 결과 45 세 이상~54 세 이하를 제외한 전 연령 범위에서 전체 KWIS의 VIQ와 WARD7형의 VIQ에서 유의한 차이가 발견되었으며, 35세~44세 집단에서는 FIQ와도 유의한 차이가 발견되었다($t=-2.98, p<.01$). 전체 KWIS와 Doppelt 방식에서의 연령 집단별 상관은 .89-.96이었으며, 이 경우에는 전 연령 범위에 걸쳐 FIQ 간의 유의한 차이를 보였다.

표 6에는 전체 KWIS의 FIQ와 두 단축형의 절대 오차 분포가 제시되어 있다. WARD7의 경우에는 86.7%가 전체 KWIS의 FIQ와 ±5점 이내의 차이를 보였고, 98.1%는 ±6 점 이내에 분포하였으며, 최대 오차는 8점 이었다. Doppelt 방식의 경우에는 65.7%가 전체 KWIS와 ±5 점 이내의 차이를 보였고 98.1%

표 2. KWIS와 WARD7의 언어성 지능, 동작성 지능, 전체 지능에 대한 평균, 표준 편차 및 타당도 계수

Scale	KWIS		WARD7		평균 차이	r
	Mean	SD	Mean	SD		
VIQ	99.84	12.43	101.97	11.85	2.13***	.96***
PIQ	90.92	14.01	90.35	14.92	.57	.96***
FIQ	95.90	13.08	96.89	12.87	.98**	.97***

** $p<.01$. *** $p<.001$.

표 3. KWIS와 Doppelt 방식에서의 전체 지능에 대한 평균, 표준 편차 및 타당도 계수

IQ	KWIS		Doppelt		평균 차이	r
	M	SD	M	SD		
FIQ	95.90	13.08	93.06	14.53	2.85***	.95***

** $p < .01$. *** $p < .001$.

표 4. 연령에 따른 KWIS와 WARD7의 언어성 지능, 동작성 지능, 전체 지능에 대한 평균, 표준 편차 및 타당도 계수

Age	n	Scale	KWIS		WARD7		평균 차이	r
			M	SD	M	SD		
24세 이하	9	VIQ	97.00	9.47	101.57	10.41	4.56*	.90**
		PIQ	91.78	12.90	89.89	13.00	1.89	.94***
		FIQ	94.44	10.96	96.56	8.59	2.11	.93***
25세~34세	37	VIQ	98.62	12.07	101.27	11.68	2.65***	.97***
		PIQ	90.30	15.71	89.76	17.25	.54	.95***
		FIQ	95.32	14.01	96.38	13.89	1.05	.97***
35세~44세	39	VIQ	98.46	12.48	100.44	12.14	1.97**	.97***
		PIQ	88.21	13.34	88.13	14.04	.08	.96***
		FIQ	93.62	12.54	94.92	12.54	1.31**	.98***
45세~54세	20	VIQ	106.05	13.02	106.45	12.57	.04	.98***
		PIQ	97.00	11.19	96.00	12.06	1.00	.95***
		FIQ	102.10	12.15	101.80	11.97	.30	.99**

* $p < .05$. ** $p < .01$. *** $p < .001$.

표 5. 연령에 따른 KWIS와 Doppelt 방식에서의 언어성 지능, 동작성 지능, 전체 지능에 대한 평균, 표준 편차 및 타당도 계수

Age	n	Scale	KWIS		Doppelt		평균 차이	r
			M	SD	M	SD		
24세 이하	9	FIQ	94.44	10.96	89.00	11.08	5.44*	.89**
25세~34세	37	FIQ	95.32	14.01	92.92	15.08	2.41**	.95***
35세~44세	39	FIQ	93.62	12.54	90.54	14.01	3.08***	.94***
45세~54세	20	FIQ	102.10	12.15	100.05	14.44	2.05*	.96***

** $p < .01$. *** $p < .001$.

가 ± 10 점 이내에 분포하였으며, 최대 오차는 19 점이었다.

또한, 전체 KWIS의 FIQ와 두 단축형의 FIQ의 절

대 오차에 대해 t 검증 결과를 표 7에 제시해 놓았다. 이 경우, WARD7에서 추정된 FIQ와 전체 KWIS의 FIQ간의 오차평균은 2.43이었던 반면 Doppelt에

표 6. 두 단축형을 통한 전체 지능의 추정시 절대 오차값의 빈도 분포

Absolute Error	WARD7 FIQ			Doppelt FIQ		
	Frequency	Percent	Cumulative Percent	Frequency	Percent	Cumulative Percent
0	23	21.9	21.9	9	8.6	8.6
1	17	16.2	38.1	15	14.3	22.9
2	20	19.0	57.1	10	9.5	32.4
3	18	17.1	74.3	14	13.3	45.7
4	8	7.6	81.9	13	12.4	58.1
5	5	4.8	86.7	8	7.6	65.7
6	12	11.4	98.1	16	15.2	81.0
7	1	1.0	99.0	7	6.7	87.6
8	1	1.0	100.0	3	2.9	90.5
9				3	2.9	93.3
10				5	4.8	98.1
19				2	1.9	100.1
Total	105	100.0		105	100.0	

서 추정된 FIQ와 전체 KWIS의 FIQ의 오차 평균은 4.31로써, 두 오차 평균이 $p < .001$ 수준에서 유의한 차이를 보였다.

다음은, Wechsler가 제시한 표준 분류방식에 따라 전체 KWIS의 FIQ와 두 단축형의 FIQ추정치를 지능 수준으로 분류하였다. 이 표준 분류 방식은 다음과 같다: 128이상=최우수 수준, 121- 127=우수 수준, 112-120=보통상 수준, 89-111=보통 수준, 79-88=보통하 수준, 70-78=경계선 수준, 69이하=정신지체수준(전용신, 서봉연, 이창우, 1963). 표 8에서는 단축형으

표 7. 단축형의 FIQ와 원래 KWIS의 FIQ간의 절대 오차값에 대한 t 검증

Absolute Error	M	SD	Mean Difference
KWIS - WARD7	2.43	2.05	1.89***
KWIS - Doppelt	4.31	3.41	

*** $p < .001$.

로 지능 수준을 추정했을 때의 분류 오류 비율을 제시해 놓았다. WARD7형에서는 10명이(9.5%), Doppelt형에서는 36명(34%)이 전체 소검사 지능 수준과 다르게 분류되어 WARD7형이 Doppelt형에 비해 분류오류율이 현저히 낮은 것으로 나타났다.

표 8. 전체 KWIS와 각 단축형의 일치 수준 및 분류 오류 비율

단축형	KWIS와의 일치 사례	KWIS와의 불일치 사례	전체 사례수
WARD7	95(90.5%)	10(9.5%)	105
Doppelt	69(66.0%)	36(34.0%)	105

표 9. 각 단축형의 분류 오류 비율

단축형	1 수준상단	1 수준하단	2 수준상단	전체사례수
WARD7	8(7.6%)	2(1.9%)		10(9.5%)
Doppelt	2(1.9%)	32(30.4%)	2(1.9%)	36(34.2%)

이를 좀더 구체적으로 살펴보면, WARD7형에서는 전체 KWIS보다 1수준 높게 분류된 경우가 8명(7.6%), 1수준 낮게 분류된 경우가 2명(1.9%)이었다. Doppelt 방식에서는 1수준 높게 분류된 경우가 2명(1.9%), 1수준 낮게 분류된 피험자가 32명(30.4%)이었으며 2수준 상단에 분류된 피험자도 2명(1.9%)에 해당되어 WARD7 형에 비해 Doppelt 방식이 분류 오류의 범위가 넓은 것으로 나타났다.

논 의

본 연구에서는 만성 정신분열병 환자들의 전반적인 지능 추정을 위한 단축형 지능 검사의 정확성과 타당도를 검토하고자 하였다. 연구 결과 두 단축형 추정치와 전체지능간에 매우 높은 상관을 보이고 있어 단축형 지능검사를 통한 전반적인 지능 추정이 상당히 정확하고 신뢰롭다는 것이 밝혀졌다. 그 중에서도 Ward의 단축형 소검사는 다음과 같은 점에서 더 정확한 것으로 나타났다. 우선 Ward의 7개 소검사를 통해 추정된 지능은 Doppelt방식에 비해 전체 KWIS의 지능과 상관이 더 높았을 뿐 아니라 오차 크기는 대부분의 경우 ± 5 점 이내로 전반적인 지능을 정확하고 효과적으로 추정해주는 것으로 나타났다(86.7% vs 65.7%). 뿐만 아니라 WARD7 단축형이 Doppelt 방식에 비해 오차 범위도 훨씬 작았으며(8점 vs 19점), 표준화된 지능 수준에 따른 분류에 있어서도 잘못 분류되는 확률이 훨씬 낮게 나타나(9.5% vs 34.0%) 정확도면에서 더 우수하다고 볼 수 있겠다. 또한 연령 집단별로 구분해 분석해 보았을 때에도 전반적으로 WARD7 단축형이 Doppelt 방식에 비해 전체 지능과의 상관이 더 높았으며, 평균 차이도 더 작게 나타났다. 더욱이 Doppelt방식의 경우에는 전 연령 범위에 걸쳐 전체 지능과 유의한 차이가 발견되었다. 즉, Doppelt 방식이 WARD7 단축형에 비해 전반적으로 정확도가 일관성 있게 떨어지는 것을 알 수 있다. 이러한 결과는 만성 정신과적 환자(Benedict et al., 1992; Abraham et al., 1997; Ryan, 1997)와 정신분

열병 진단을 받은 환자를 대상으로한 Allen 등(1997)의 연구 결과와 일치하는 것이다. 다만, 평균 차이 검증에서 WARD7형의 동작성 지능 추정치는 전체 KWIS의 동작성 지능과 유의한 차이가 나지 않았던 것에 비해 언어성 지능, 전체 지능에서는 WARD7 단축형과 전체 KWIS가 유의한 차이를 보였다. 이는 7개의 소검사로 구성되어 있는 WARD 단축형의 PIQ가 VIQ, FIQ에 비해 원래의 지능을 보다 정확하게 예측해주는 것을 의미한다. 이와 같은 결과는 Ryan, Weilage, 그리고 Spaulding(1999)의 연구결과와 상반된 것으로 이에 대한 가능한 설명은 다음과 같다. 우선 정신분열병 환자들에게서 흔히 취약성 지표(vulnerability marker)인 산수, 숫자와 같은 주의력과 관련된 소검사가 WARD7 단축형 언어성 지능 산출 공식에 포함되어 있다는 사실이다. 대부분의 정신분열병 환자들이 숫자, 산수와 같은 소검사의 수행이 저조하여 소검사들간에 편차가 큰(scattered) 프로파일을 이루는 경향이 있기 때문에 이들 소검사가 포함되는 언어성 지능의 경우 큰 편차가 반영되는 경향이 나타난다. 반면 동작성 소검사들은 대부분의 급성 정신분열병 환자들뿐만 아니라 관해형 정신분열병 환자들에서도 소검사들에 걸쳐 일관성 있게 수행이 저하되어 나타나기 때문에 상대적으로 소검사들간에 편차가 작아 단축형과 원래 동작성 지능간의 불일치 정도가 낮게 나왔을 가능성이 있다. 이에 대해서는 향후 여러 임상 집단에 걸쳐 그 타당성을 검토할 필요가 있겠다.

본 연구의 의의는 WARD7 단축형이 기존에 임상 장면에서 많이 사용되어왔던 Doppelt방식에 비해 보다 정확하고 신뢰롭게 전체 지능을 예측해줄 수 있다는 것을 밝힌 데 있다. 특히 동기나 의욕 저하 등으로 인해 전체 지능 검사를 모두 실시하기가 어려운 만성 정신분열병 환자들의 전반적인 지능을 추정해주는 데 Ward의 단축형 소검사가 매우 유용함이 입증되었다.

이상의 결과를 바탕으로 Doppelt 방식에 비해 WARD7 단축형의 장점을 구체적으로 살펴보면 다음과 같다. 첫째, Doppelt 방식은 FIQ만 추정이 가능한

것에 비해 WARD7 방식의 경우에는 FIQ에 대한 추정치뿐만 아니라 언어성 지능과 동작성 지능에 대한 추정치도 제공해준다는 이점이 있다. 이는 WARD7 단축형이 단순히 FIQ만 추정하는 도구로 그치는 것이 아니라 신경인지 기능 평가나 재활에 필요한 기능 평가시 매우 유용한 소검사 자료를 살릴 수 있다는 것을 의미한다. 특히 만성 정신 분열병 환자들의 기능 평가시 필요한 주의력, 현실 검증력, 심신 운동 속도 등과 같은 정보를 줄 수 있다는 점에서 그 활용 가치가 크다고 볼 수 있다. 또한 WARD7 단축형은 Doppelt형에 비해 지능의 일반적 요인 부하량이 높은 하위 검사들이 많이 포함되어 있어 심리 측정적인 면에서 더욱 타당하다고 볼 수 있다. 한편 WARD7 단축형은 Doppelt 방식에서 소요되는 시간보다 5-20%, 대략 2분-6분 정도 더 소요되는 것으로 알려져 있다 (Allen et al., 1997). 이렇듯, WARD7 단축형은 Doppelt 방식과 비교해볼 때, 실시하는데 소요되는 시간 차이가 별로 나지 않으면서도 정확도가 더 높고 더 많은 정보를 제공해 줄 수 있다는 점에서 그 효용성이 더 크다고 할 수 있다.

이상과 같이 WARD7 단축형이 전체 지능을 예측해주는 데 보다 정확하고 신뢰로운 것으로 밝혀졌지만 그럼에도 불구하고 일반적으로 단축형을 통해 전체 지능을 예측할 때에 몇 가지 유의할 점이 있다. 본 연구 결과와 선행연구를 살펴보면 모든 단축형 지능검사들은 표준적으로 사용되는 지능 범위를 기준으로 할 때, 잘못 분류되는 비율이 일정 수준 나타나고 있기 때문에 단축형 소검사를 사용하여 지능 범위를 결정할 때에는 특별히 주의가 요구된다고 볼 수 있다. 또한 선행 연구 결과, 특별히 임상적인 해석을 필요로 하는 언어성 지능과 동작성 지능간의 불일치(VIQ-PIQ)는 단축형과 원래 지능검사가 차이를 보이는 경우가 있으므로 이에 대한 해석 역시 조심스럽게 적용할 필요가 있을 것이다.

한편 연령 집단별로 두 단축형을 통한 지능 추정을 살펴보면 연령에 따라 전체 소검사 지능과의 상관계수에 다소 차이가 있음이 발견되었다. 예를 들어, WARD 7 단축형에서 연령이 24세 이하인 경우 다른

연령에 비해 전체 지능과의 상관계수가 낮았던 반면, 45세 이상에서는 단축형 추정치와 전체 지능의 상관계수가 가장 높게 나타났다. 그러므로 단축형을 통해 전체 지능을 추정할 때 연령에 따른 차이를 고려해야 할 것이다. 앞으로 이러한 차이가 연령 변화에 따른 표집의 차이가 반영된 것인지 아니면 단순히 피검자수의 차이에 기인하는 것인지 후속 연구를 통해 밝혀질 필요가 있겠다.

한편 본 연구에서는 지금까지 자료가 축적되어온 KWIS로 측정된 점수를 사용하여 필요한 분석을 해보았으나 K-WAIS로 측정된 몇몇 사례를 가지고 같은 방법으로 분석 해본 결과, KWIS를 사용했을 때 보다 단축형에 의한 전체 소검사 지능 추정이 더 정확하게 나타났다. 따라서 향후 K-WAIS를 가지고 이러한 단축형 지능검사의 정확성과 타당도를 재검증할 필요가 있겠다.

참고문헌

- 김은정, 김중술 (1993). K-WAIS의 요인분석: 정신과 환자 집단 자료를 중심으로. 한국심리학회지: 임상, 12(2), 52-61.
- 김중술, 김영환 (1974). KWIS의 단축형에 관한 연구 - Doppelt형. 장병림 교수 회갑기념 논총: 1-11.
- 김중술, 이용승, 이민식 (1993). K-WAIS의 단축형에 관한 연구. 정신의학, 19, 121-126.
- 박상규, 홍창희, 김재환 (1992). KWIS의 요인구조에 관한 연구. 정신건강연구, 11, 215-224.
- 염태호, 박영숙, 오경자, 김정규, 이영호 (1992). K-WAIS 실시요강. 서울: 한국가이던스
- 오상우 (1997). K-WAIS의 요인구조: 정신분열증 환자 집단 자료를 중심으로. 한국심리학회지: 임상, 16(1), 163-170.
- 이창우 (1964). KWIS의 인자 분석적 연구. 성균관대학교 대학원 석사학위논문
- 전용신, 서봉연, 오창우 (1963). KWIS 실시요강. 서울: 중앙교육연구소

- 채영숙, 김현정, 오상우 (1992). KWIS 분산도 분석에 의한 인지기능의 평가-정신분열증과 조증 환자를 중심으로. 한국심리학회 '92 연차 학술대회 논문집, 621-630.
- 최미례, 김중술 (1990). 정신병 환자의 KWIS 분산도 분석-신경증 환자와의 비교. 한국심리학회: 임상, 9(1), 192-205.
- 한덕웅 (1975). 정신질환자의 지능 측정술 위한 연구: KWIS의 진단적 사용의 선결문제. 성균관대학교 논문집, 20, 255-268.
- Abraham, E., Axelrod, B. N., & Paolo, A. M. (1997). Comparison of WAIS-R selected subtest short forms in a clinical population. *Assessment*, 4, 409-417.
- Allen, D. N., Huegel, S. G., Gurklis, J. A., Kelley, M. E., Barry, E. J., & van Kammen, D. P. (1997). Utility of WAIS-R short forms in schizophrenia. *Schizophrenia Research*, 26, 163-172.
- American Psychiatric Association. (1994). *DSMIV: Diagnostic and Statistical Manual of Mental Disorders*(4th ed). Washington, D.C
- Atkinson, L., & Cyr, J. J. (1984). Factor analysis of the WAIS-R: Psychiatric and standardization samples. *Journal of Consulting & Clinical Psychology*, 53, 402-405.
- Axelrod, B. N., Woodard, J. L., Schretlen, D., & Benedict, R. H. B. (1996). Corrected estimates of WAIS-R short form reliability and standard error of measurement. *Psychological Assessment*, 8, 222-223.
- Beck, N. C., Horwitz, E., Seidenberg, M., Parker, J., & Frank, R. (1985). WAIS-R factor structure in psychiatrics and general medical patient. *Journal of Consulting & Clinical Psychology*, 50, 652-660.
- Benedict, R. H. B., Schretlen, D., & Bobholz, J. H. (1992). Concurrent validity of three WAIS-R short form reliability and standard error of measurement. *Psychological Assessment*, 8, 222-223.
- Cyr, J. J., & Brooker, B. H. (1984). Use of appropriate formulas for selecting WAIS-R short forms. *Journal of Consulting and Clinical Psychology*, 52, 903-905.
- Doppelt J. E. (1956). Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. *Journal of Consulting Psychology* 20, 63-66.
- Feingold, A. (1982). The validity of the Information and Vocabulary subtests of the WAIS. *Journal of Clinical Psychology*, 38, 169-174.
- Gutkin, T. B., Reynolds, C. R., & Galvin, G. A. (1984). Factor analysis of the Wechsler Adult Intelligence Scale-Revised(WAIS-R): An examination of the standardization sample. *Journal of School Psychology*, 22, 83-93.
- Kaufman, A. S. (1972). A short form of the Wechsler Preschool and Primary Scale of Intelligence. *Journal of Consulting and Clinical Psychology*, 39, 361-369.
- Kaufman, A. S., Ishikuma, T. A., & Kaufman-Packer, J. L. (1991). Amazingly short forms of the WAIS-R. *Journal of Psychoeducational Assessment*, 9, 4-15.
- Parker, K. (1983). Factor analysis of the WAIS-R at nine age levels between 16 and 74 years. *Journal of Consulting Psychology*, 61, 302-308.
- Resnick, R. J., & Entin, A. D. (1971). Is an abbreviated form of the WISC valid for Afro-Americans? *Journal of Consulting and Clinical Psychology*, 36, 97-99.
- Reynold, C. R., Willson, V. L., Clark, P. L. (1983). A four-subtest short form of the WAIS-R for clinical screening. *Clinical Neuropsychology*. 5, 111-116.

- Ryan, J. J., Rosenberg, S. J. (1984). Administration time estimates for WAIS-R subtests and short forms in a clinical sample. *Journal of Psychoeducation Assessment*, 2, 125-129.
- Ryan, J. J., Rosenberg, S. J., & DeWolfe, A. S. (1984). Generalization of the WAIS-R factor astructure with a vocational rehabilitation sample. *Journal of Consulting Clinical Psychology*, 52, 311-312.
- Ryan J. J. (1997). Two types of tables for use with the seven subtest short form of the WAIS-R. *Journal of Psychoeducation Assessment*, 15, 308-313.
- Ryan J. J., Weilage, M. E., Spaulding, W. D. (1999). Accuracy of the seven subtest WAIS-R short form in chronic schizophrenia. *Schizophrenia Research* 39, 79-83.
- Satz, P., & Mogel, S. (1962). Abbreviation of the WAIS for clinical use. *Journal of Clinical Psychology*, 18, 77-79.
- Shipley, W. C. (1940). A self-administering scle for measuring intellectual impairment and deterioration. *Journal of Psychology*, 9, 371-377.
- Silverstein, A. B. (1982). Two-and four subtest short forms of the Wechsler Adult Intelligence Scale-Revised. *Journal of Consulting and Clinical Psychology*, 50, 415-418.
- Silverstein, A. B. (1984). Estimating Full Scale IQs from short forms of Wechsler's scales: Linear scaling versus linear regression. *Journal of Consulting and Clinical Psychology*, 52, 919.
- Silverstein, A. B. (1985). Two-and four subtest short forms of WAIS-R: a closer look at validity and reliability. *Journal of Clinical Psychology*, 41, 95-97.
- Slosson, R. L. (1982). Slosson Intelligence Test(2nd Eds.). East Aurora, NewYork: Slosson Educational Publications.
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence*. Boston: Houghton Mifflin.
- Thompson, A. P. (1987). Methodological issues in the clinical evaluation of two-and four-subtest short formsof the WAIS-R. *Journal of Clinical Psychology*, 43, 142-144.
- Ward, L. C., Selby, B., Clark, B. L. (1987). Subtest administration times and short forms of the WAIS-R. *Journal of Clinical Psychology*, 43, 276-278.
- Ward, L. C. (1990). Prediction of verbal, performance, and full scale IQs from the seven subtests of the WAIS-R. *Journal of Clinical Psychology*, 46, 436-400.
- Wechsler, D. (1981). *WAIS-R: Manual: Weschsler Adult Intelligence Scale-Revised*. The Psychological Corporation, San Antonio, TX.

원고접수일 2000. 1. 20

수정원고접수일 2000. 4. 12

게재결정일 2000. 4. 21 ■

The Study on the Accuracy and Validity of Korean Wechsler Intelligence Scale short forms: A Comparison of the WARD7 subtest vs Doppelt subtest

Young-Ran Lim · Woo-Kyeong Lee · Won-hye Lee · Jong-won Park

Yongin Mental Hospital

The short forms of the Wechsler Adult Intelligence Scale-Revised(Wechsler, 1981) have particular utility for individuals with schizophrenia because of their ability to provide estimated IQ scores with substantial time saving and, as a result, are less taxing for these patients who often show impaired attention and deficient motivation. Although valid short forms of the WAIS-R exist, they often require validation study. The purpose of this study is to compare the accuracy and validity of two short forms. The Korean-Wechsler Intelligence Scale(KWIS, 1963) was given to 105 schizophrenic patients, and IQs were estimated from Ward's seven subtests(Information, Arithmetic, Similarities, Digit Span, Digit Symbol, Picture Completion, Block Design) and Doppelt subtests (Vocabulary, Arithmetic), (Block Design, Picture Arrangement) that require about half the administration time of the full test. Results indicated that 86.7% of the estimated Full Scale IQs were ± 5 points of their actual scores. Estimates correlated well with Verbal, Performance, and Full Scale IQs, and mean estimated IQs differed by less than 1 IQ point from actual means. Error sizes in predicting Full Scale IQs were small($M=2.43$) and exceeded 6 IQ points in only 2 cases(1.9%). In comparison, error sizes for the Doppelt(1956) abbreviation($M=4.31$) were significantly larger and were greater than 6 IQ points in 20 cases(19%). Furthermore, Ward's seven subtest short form had lower misclassification rate(9.5%) than Doppelt's subtest(35%). This abbreviated Wechsler Scale may be used with schizophrenic patients when only general estimates of intellectual functioning are required. However, all short forms had substantial misclassification rates, indicating that caution is warranted when using these forms to classify individuals according to standard levels of intellectual functioning. Finally, several limitations and suggestions of the present study were discussed.