

## 숫자 바로 따라 외우기와 거꾸로 따라 외우기의 차이에 관한 기준 연구\*

김 홍 근<sup>†</sup>

대구대학교 재활심리학과

박 태 진

전남대학교 심리학과

숫자외우기의 전통적인 채점 방법은 바로 따라 외우기(DF)와 거꾸로 따라 외우기(DB)의 점수를 합산하여 단일하게 채점하는 것이다. 그러나 DF와 DB의 개별적 채점 및 양자의 비교 필요성이 여러 연구자들에 의하여 제기되어 왔다. 이러한 견해를 수용하여 Wechsler 지능척도의 최신판인 WAIS-III는 요강 자체가 DF와 DB의 차이(DF-DB)에 대한 기준을 포함하고 있다. 그러나 한국에서는 DF-DB의 기준이 아직 제시된 바 없어서 양자의 차이에 대한 객관적 평가가 어렵다. 이에 본 연구에서는 정상인 201명의 자료에 기초하여 DF-DB에 관한 한국인 기준을 제시하였다. 연령, 학력, 지능 수준이 DF-DB에 미치는 효과를 검토한 결과 통계적으로 유의미하지 않았다. 그러므로 DF-DB의 기준은 연령, 학력, 지능 수준의 구분 없이 단일하게 작성하였다. 기준의 형식은 DF의 자릿수에서 DB의 자릿수를 뺀 결과를 백분위 점수화한 형식을 취하였다. 기준을 검토해보면 DF-DB가 3이하면 백분위 35.3%이상으로 '정상' 수준이며, 4면 백분위 16.4%로 '경계선' 수준이고, 5 이상은 백분위 6.5%이하로 '비정상' 수준이다. DF는 정보의 유지만을 요하지만 DB는 정보의 유지뿐 아니라 조작을 요한다. 이러한 점에서 DF는 단기기억을 측정하지만 DB는 단기기억과 더불어 작업기억(working memory)을 측정한다. 그러므로 DF-DB는 작업기억의 한 지표로 활용될 수 있다.

주요어 : 숫자외우기, 기준, 작업기억

\* 본 연구는 과학기술부의 2003-2004 뇌신경정보사업의 지원을 받아 수행되었다.

† 교신저자(Corresponding Author) : 김 홍 근 / 대구대학교 재활심리학과 / 대구시 남구 대명3동 2288, 705-714  
TEL : 053-650-8295 / FAX : 053-650-8259 / E-mail : hongkn@daegu.ac.kr

Wechsler 지능척도는 전세계적으로 가장 많이 사용되는 임상용 지능검사이다(Harrison, Kaufman, Hickman, & Kaufman, 1988). 한국에서도 Wechsler Adult Intelligence Scale-Revised(WAIS-R: Wechsler, 1981)를 번안한 K-Wechsler Adult Intelligence Scale(K-WAIS: 염태호, 박영숙, 오경자, 김정규, 이영호, 1992)이 임상용 지능검사로 가장 많이 사용되고 있다. 그러므로 K-WAIS의 임상적 활용도를 높이려는 작업은 매우 중요하다. 현재까지 이루어진 이러한 작업의 예로는 단축형 척도 개발(김중술, 이용승, 이민식, 1994), 언어성 지능(VIQ)과 동작성 지능(PIQ)의 차이 기준 제작(김홍근, 1999a), 기억지수(Memory Quotient)와의 비교 기준 제작(김홍근, 1999b), 관리 지능지수(Executive Intelligence Quotient)와의 비교 기준 제작(김홍근 2001a), 병전 지능 추정법 개발(김홍근, 2001b, 2001c), 측두엽간질(temporal lobe epilepsy)의 편재화(lateralization) 지표 개발(김홍근, 2002a) 및 3요인 해석을 위한 기준 제작(김홍근, 2002b) 등을 꼽을 수 있다. 그러나 해외의 풍부한 연구 성과(예, Kaplan, Fein, Morris, & Delis, 1991; Kaufman, 1990)와 비교한다면 K-WAIS의 활용도를 높이려는 작업이 더욱 요구된다. 본 연구는 K-WAIS의 임상적 활용도를 높이려는 시도의 하나로 특히 숫자외우기 소검사에 주목하였다. 숫자외우기가 여타의 Wechsler 소검사들과 다른 점의 하나는 그 자체가 두 개의 소검사, 즉 숫자 바로 따라 외우기(Digits Forward: 이하 DF)와 숫자 거꾸로 따라 외우기(Digits Backward: 이하 DB)로 구성된 점이다. 본 연구의 주목적은 DF와 DB의 차이(이하 DF-DB)에 관한 한국인 기준을 제시하는 것이다.

숫자외우기의 전통적인 채점 방법은 DF와 DB의 점수를 합산하여 단일하게 채점하는 것이다. 그러나 많은 연구자들이 이러한 채점 방법의 문

제점을 지적하고 DF와 DB의 개별적인 채점 및 양자의 비교 필요성을 제기한 바 있다(Banken, 1985; Black, 1986; Hayslip & Kennelly, 1982; Kaplan et al., 1991; Lezak, 1995; Ramsay & Reynolds, 1995; Rapport, Webster, & Dutra, 1994). 이러한 견해를 적극적으로 수용하여 Wechsler 지능척도의 최신 개정판인 WAIS-III(Wechsler, 1997)는 요강 자체가 DF-DB의 기준을 포함하고 있다. 그러나 한국에서는 DF-DB의 기준이 아직 발표된 바 없어서 양자의 차이에 대한 객관적 평가가 어렵다. 이에 본 연구에서는 한국인 정상인 201명의 숫자외우기 자료를 기초로 DF-DB의 기준 제작을 시도하였다. 기준 제작에 앞서 연령, 학력, 지능 수준이 DF-DB에 미치는 효과를 분석하였다. 이 분석의 목적은 DF-DB의 기준을 연령, 학력, 또는 지능 수준별로 작성해야 하는지 아니면 단일하게 작성해도 되는지를 결정하는 것이었다. 본 연구의 표집수는 기준 연구로서는 그리 크다고 할 수 없으며 특히 연령, 학력, 또는 지능 수준별 기준을 작성하기에는 절대적으로 부족하였다. 그러므로 연령, 학력, 지능 수준이 DF-DB에 미치는 효과가 극히 작아서 단일한 기준이 무방할 경우에만 실제적인 기준 제작이 가능할 것으로 예상하였다.

DF와 DB는 검사자가 불러준 일련의 숫자들을 피검자가 따라서 말하는 과제라는 점에서 매우 유사하며 다만 후자가 '역순'이라는 요소를 포함하는 점에서만 다르다. 이러한 점에서 DF와 DB는 일종의 통제과제-실험과제 짝으로 생각할 수 있다. 이러한 통제과제-실험과제 짝의 다른 예로는 Stroop Test(Stroop, 1935)에서 단순시행과 간섭시행, Trail Making Test(Reitan & Wolfson, 1993)에서 A시행과 B시행, Rey Complex Figure Test(Rey, 1941; 김홍근, 1999b)에서 모사시행과 회상시행 등을 들 수 있다. 이렇게 실험과제와 통제과제가

함께 실시되는 검사의 주요 특징은 두 과제의 수행을 비교하여 임상적 해석을 보다 구체화할 수 있는 점이다. 예를 들어 Stroop Test의 주요 목적은 '선택적 주의'를 측정하는 것이다. 그런데 만약 피검자가 Stroop Test의 간섭시행에서 저조한 수행을 보인다면 이는 선택적 주의의 문제로 해석할 수도 있지만, 일반 주의력의 문제, 일반 지능의 문제, 언어적 문제, 색채 지각의 문제 등 여러 가지 대안적 해석이 가능하다. 그러나 만약 피검자가 단순시행에 비하여 간섭시행에서만 특히 저조한 수행을 보이는 것이라면 선택적 주의의 문제라는 해석이 다른 대안적 해석들에 비하여 훨씬 더 설득력을 가진다. 그러므로 Stroop Test를 선택적 주의를 측정하는 검사로 활용하기 위해서는 단순시행과 간섭시행의 비교가 매우 중요하다.

숫자외우기에서 DF와 DB에 내재된 인지 과정을 살펴보면 DF는 정보의 유지만을 요하지만 DB는 정보의 유지뿐 아니라 조작을 요한다. 이러한 점에서 DF는 단기기억 과제이지만 DB는 단기기억과 더불어 작업기억(working memory)이 요구되는 과제이다. 작업기억이란 '일시적으로 유지되고 있는 정보에 대해 정신적 조작(mental operation)을 가할 수 있는 능력'을 의미한다 (Baddeley, 1995). 작업기억은 '암산'과 같은 대표적인 예는 물론이고 추론, 언어 이해, 공간적 정보처리, 계획 등의 여러 인지 과정에 폭 넓게 관여하므로(D'Esposito, 2001; Rains, 2002) 임상심리검사에서 평가할 필요성이 매우 크다. 숫자외우기를 작업기억을 평가하는 도구로 활용할 경우 (Conklin, Curtis, Katsanis, & Iacono, 2000; Foster, Black, Buck, & Bronskill, 1997; Kiefer, Apel, & Weisbrod, 2002) 다음과 같은 점의 고려가 필요하다. 즉 만약 피검자가 DB에서 저조한 수행을 보인다면 이는 작업기억의 문제로 해석할 수도 있

지만 단기기억의 문제, 일반 지능의 문제, 청각적 문제, 언어적 문제 등 여러 가지 대안적 해석이 가능하다. 그러나 만약 피검자가 DF에 비해서 DB에서만 특히 저조한 수행을 보이는 것이라면 작업기억의 문제라는 해석이 다른 대안적 해석들에 비하여 훨씬 더 설득력을 가진다. 그러므로 숫자외우기를 작업기억과제로 활용하려면 DF와 DB의 수행 비교가 매우 중요하다. 피검자의 DF와 DB 수행을 비교하려면 DF와 DB 각각에 대한 개별적 기준도 유용하지만 가장 결정적으로 필요한 것은 DF-DB의 기준이다.

## 방 법

### 피검자

피검자는 연령이 16-64세에 속하는 성인남녀 201명이었다. 피검자는 검사자의 지인이거나 이들이 소개한 사람들이었으며 거주지는 주로 대구 및 경북지역이었다. 면접 질문에서 신경과나 정신과적 주요 병력이 확인되는 사람들은 피검 대상에서 제외하였다. 표집 과정에서 성별 및 연령 분포는 가능한 균형이 맞도록 노력하였으며, 학력 분포는 K-WAIS 기준집단의 학력분포(염태호 등, 1992, p. 80)에 맞추고자 노력하였다. 실제로 표집된 피검자의 성별 분포는 남자가 89명, 여자가 112명이었다. 연령 분포는 16-19세가 34명, 20-24세가 41명, 25-34세가 43명, 35-44세가 23명, 45-54세가 34명, 55-64세가 26명이었다. 학력 분포는 0-6년이 21명, 7-9년이 33명, 10-12년이 97명, 13년 이상이 50명이었다. 표 1은 피검자의 연령별 학력 분포를 보여 준다. 표 1에서 볼 수 있듯이 연령이 높은 집단일수록 평균 학력이 상대적으로 낮았다. 이러한 연령별 학력 차이는 세

표 1. 연령 및 학력 수준에 따른 피검자 수

학력(년)	연 령 (세)					
	16-19	20-24	25-34	35-44	45-54	55-64
13이상	8	12	22	3	5	0
10-12	22	26	20	15	12	2
7-9	4	3	1	5	13	7
0-6	0	0	0	0	4	17
합	34	41	43	23	34	26
(%)	(16.9)	(20.4)	(21.4)	(11.4)	(16.9)	(12.9)

표 2. 연령 수준별 FIQ 평균, 표준편차, 최고점(Max)과 최저점(Min)

	연 령 (세)					
	16-19	20-24	25-34	35-44	45-54	55-64
<i>M</i>	104.7	106.5	109.8	112.7	111.0	97.7
<i>SD</i>	13.2	9.5	11.3	13.7	10.5	15.0
<i>Min</i>	64	84	82	80	85	70
<i>Max</i>	128	123	135	135	132	118

대간에 존재하는 실질적인 학력 차이를 반영하는 것으로 K-WAIS 규준집단에서도 이와 유사한 연령별 학력 차이가 존재한다(김홍근, 2001c). 피검자 전체의 K-WAIS 지능지수(FIQ) 평균은 107.2이고 표준편차는 12.7로 다소 편향된 표집을 이루었다. 표 2는 피검자의 연령별 FIQ 평균을 보여준다. 변량분석(ANOVA)을 사용하여 검증한 결과 피검자의 연령별 FIQ 평균은 유의미하게 달랐다,  $F(5, 195)=5.58, p < .001$ . Tukey's HSD test로 사후 검증한 결과( $\alpha = .05$ ) 16-19세에서 45-54세까지의 다섯 집단 간에는 FIQ 평균이 유의미하게 다르지 않았다. 그러나 55-64세 집단의 FIQ 평균은 20-24세, 25-34세, 35-44세, 45-54세 집단의 FIQ 평균에 비하여 각각 유의미하게 작았다.

#### 절차

모든 피검자들에게 K-WAIS, Rey-Kim기억검사(김홍근, 1999b), 관리지능검사(Executive Intelligence Test: 김홍근, 2001a)를 실시하였다. 본 연구에는 숫자외우기 소검사만 해당되므로 다른 검사들에 관한 기술은 생략한다. 검사자들은 K-WAIS의 실시 및 해석 방법을 교육받은 심리학 관련 학부 및 대학원생들이었다. 숫자외우기는 K-WAIS 요강에 따라 표준적인 방법으로 실시하였다. 요약하면 DF를 먼저 실시하였으며 연이어서 DB를 실시하였다. DF에서는 검사자가 일련의 숫자들을 불러주면 피검자가 이 숫자들을 곧바로 따라서 말하였다. DB에서는 검사자가 일련의 숫자들을 불러주면 피검자가 이 숫자들을 곧바로 역순

표 3. 바로 따라 외우기와 거꾸로 따라 외우기의 채점례

바로 따라 외우기			거꾸로 따라 외우기		
1.	5-8-2	(o)	1.	2-4	(o)
	6-9-4	(o)		5-8	(o)
2.	6-4-3-9	(o)	2.	6-2-9	(o)
	7-2-8-6	(o)		4-1-5	(o)
3.	4-2-7-3-1	(o)	3.	3-2-9-7	(o)
	7-5-8-3-6	(o)		4-9-6-8	(o)
4.	6-1-9-4-7-3	(o)	4.	1-5-2-8-6	(x)
	3-9-2-4-8-7	(o)		6-1-8-4-3	(x)
5.	5-9-1-7-4-2-8	(o)	5.	5-3-9-4-1-8	( )
	4-1-7-9-3-8-6	(o)		7-2-4-8-5-6	( )
6.	5-8-1-9-2-6-4-7	(x)	6.	8-1-2-9-3-6-5	( )
	3-8-2-9-5-1-7-4	(o)		4-7-3-9-1-2-8	( )
7.	2-7-5-8-6-2-5-8-4	(x)	7.	9-4-3-7-6-2-5-8	( )
	7-1-3-9-4-2-5-6-8	(x)		7-2-8-1-9-6-5-3	( )
점수		8	점수		4

으로 따라서 말하였다. 예를 들어 검사자가 “6-4-3-9”라고 불러주면 피검자의 과제는 “9-3-4-6”이라고 따라서 말하는 것이었다. DF와 DB 공히 검사자가 숫자들을 불러주는 속도는 1초에 1개 정도였다. DF는 자릿수가 3인 것(예, 5-8-2)에서 9인 것까지, DB는 자릿수가 2인 것에서 8인 것까지 단계적으로 실시하였다. 각 단계마다 2회의 시행을 실시하였으며 이 중 1회의 시행이라도 성공하면 다음 단계의 시행을 계속하였다. 그러나 2회의 시행을 모두 실패하는 경우에는 다음 단계의 시행들을 실시하지 않았다.

채점은 WAIS-III에서 DF-DB의 규준 제작에 사용한 방식을 원용하였다. 이 채점 방식은 DF-DB에 관한 다른 연구들(Kaplan et al., 1991; Lezak, 1995)에서도 공통적으로 사용되고 있어서 가장 표준적인 방식에 해당한다. 이 채점 방식에서 DF 점수는 피검자가 정확하게 바로 따라서 말할 수

있었던 숫자목록 중 가장 긴 것의 자릿수이다. DB점수는 피검자가 정확하게 거꾸로 따라서 말할 수 있었던 숫자목록 중 가장 긴 것의 자릿수이다. DF와 DB 공히 ‘가장 긴 숫자목록’은 2회의 시행을 모두 성공적으로 수행할 필요가 없고 1회의 시행만 성공적으로 수행하여도 해당된다. 표 3은 DF점수와 DB점수의 산출례를 보여준다. DF-DB점수는 DF점수에서 DB점수를 빼어서 산출한다. 표 3에 예시된 사례의 경우 DF점수가 8, DB점수가 4이므로, DF-DB점수는 4가 될 것이다. 피검자가 DF 또는 DB의 첫 단계에서 2회의 시행 모두를 실패하는 경우 채점이 애매할 수 있다. 본 연구에서는 이러한 경우 첫 단계의 숫자목록 보다 하나 적은 자릿수로 채점하였다. 그러므로 DF의 첫 단계를 실패하는 경우 DF점수는 2, DB의 첫 단계를 실패하는 경우 DB점수는 1로 채점하였다. 그러나 실제적으로 이러한 바닥효과

(floor effect)를 보인 피검자는 201명 중 1명에 불과하였다.

### 분석

규준 제작에 앞서 연령, 학력, 지능 수준이 DF-DB에 미치는 효과를 분석하였다. 이 분석의 목적은 DF-DB의 규준을 연령, 학력, 또는 지능 수준별로 작성해야 하는지 아니면 단일하게 작성해도 되는지를 결정하는 것이었다. 첫째로, 연령 변인이 DF-DB에 미치는 효과를 분석하였다. 이 분석에서는 연령 변인을 독립변인, DF-DB를 종속변인으로 ANOVA를 실시하였다. 연령 변인의 수준은 표 1에 제시된 것에 준하였다. 본 연구의 표집은 세대간의 실질적인 학력차이를 반영하여 연령이 높을수록 학력이 낮은 경향을 보였다(표 1 참고). 이러한 연령별 학력 차이를 통제된 상태에서 연령 변인의 효과를 검증하기 위하여 교육년수를 공변인으로 포함시킨 공변량분석(ANCOVA)도 실시하였다. 또한 본 연구의 표집은 연령별 FIQ 평균이 유의미하게 다른 편향성(bias)을 보였다(표 2 참고). 이러한 연령별 FIQ 차이를 통제된 상태에서 연령 변인의 효과를 검증하기 위하여 FIQ를 공변인으로 포함시킨

ANCOVA를 실시하였다. 둘째로, 학력 변인이 DF-DB에 미치는 효과를 분석하였다. 이 분석에서는 학력 변인을 독립변인, DF-DB를 종속변인으로 ANOVA를 실시하였다. 학력 변인의 수준은 표 1에 제시된 것에 준하였다. 본 연구의 표집은 세대간의 실질적인 학력차이를 반영하여 학력이 높을수록 연령이 낮은 경향을 보였다(표 1 참고). 이러한 학력별 연령 차이를 통제된 상태에서 학력 변인의 효과를 검증하기 위하여 연령을 공변인으로 포함시킨 ANCOVA를 실시하였다. 셋째로, 지능수준이 DF-DB에 미치는 효과를 분석하였다. 이 분석에서는 FIQ와 DF-DB간의 상관분석을 실시하였다.

## 결 과

### 연령의 효과

표 4는 연령 수준별 DF, DB, DF-DB의 평균과 표준편차를 보여준다. ANOVA에서 연령 변인이 DF에 미치는 효과는 통계적으로 매우 유의미하였다,  $F(5, 195) = 16.00, p < .001$ . 연령 변인이 DB에 미치는 효과도 매우 유의미하였다,  $F(5,$

표 4. 연령 수준별 DF, DB, DF-DB점수의 평균, 표준편차

연령(세)	n	DF		DB		DF-DB	
		M	SD	M	SD	M	SD
16-19	34	7.65	1.28	5.85	1.42	1.79	1.20
20-24	41	7.83	1.09	5.78	1.37	2.05	1.52
25-34	43	7.56	1.18	5.26	1.40	2.30	1.66
35-44	23	7.39	1.37	5.22	1.54	2.17	1.56
45-54	34	6.91	1.55	4.62	1.60	2.29	1.34
55-64	26	5.15	1.59	3.27	1.31	1.88	1.42

195) = 13.05,  $p < .001$ . 그러나 연령 변인이 DF-DB에 미치는 효과는 전혀 유의미하지 않았다,  $F(5, 195) < 1$ , *ns*. 표 4에서 볼 수 있듯이 연령 수준이 높을수록 DF와 DB는 점차 낮아지는 경향이 뚜렷하였다. 그러나 DF-DB는 연령 수준에 관계 없이 거의 일정하였다. 교육년수를 공변인으로 포함시킨 ANCOVA에서도 연령 변인이 DF, DB, DF-DB에 미치는 효과는 유사하였다. 즉 연령 변인이 DF에 미치는 효과와,  $F(5, 194) = 3.42, p < .01$ , DB에 미치는 효과는 유의미하였지만,  $F(5, 194) = 4.21, p < .01$ , DF-DB에 미치는 효과는 전혀 유의미하지 않았다,  $F(5, 194) < 1, ns$ . FIQ를 공변인으로 포함시킨 ANCOVA에서도 연령 변인이 DF, DB, DF-DB에 미치는 효과는 유사하였다. 즉 연령 변인이 DF에 미치는 효과와,  $F(5, 194) = 14.65, p < .001$ , DB에 미치는 효과는 유의미하였지만,  $F(5, 194) = 14.49, p < .001$ , DF-DB에 미치는 효과는 전혀 유의미하지 않았다,  $F(5, 194) < 1, ns$ .

#### 학력의 효과

표 5는 학력 수준별 DF, DB, DF-DB의 평균과 표준편차를 보여준다. ANOVA에서 학력 변인이 DF에 미치는 효과는 통계적으로 매우 유의미하였다,  $F(3, 197) = 34.64, p < .001$ . 학력 변인이

DB에 미치는 효과도 매우 유의미하였다,  $F(3, 197) = 30.19, p < .001$ . 그러나 학력 변인이 DF-DB에 미치는 효과는 전혀 유의미하지 않았다,  $F(3, 197) < 1, ns$ . 표 5에서 볼 수 있듯이 학력 수준이 높을수록 DF와 DB는 증가하였다. 그러나 DF-DB는 학력 수준에 관계 없이 거의 일정하였다. 연령을 공변인으로 포함시킨 ANCOVA에서도 학력 변인이 DF, DB, DF-DB에 미치는 효과는 유사하였다. 즉 학력 변인이 DF에 미치는 효과와,  $F(3, 196) = 13.43, p < .001$ , DB에 미치는 효과는 유의미하였지만,  $F(3, 196) = 10.44, p < .001$ , DF-DB에 미치는 효과는 전혀 유의미하지 않았다,  $F(3, 196) < 1, ns$ .

#### 지능의 효과

전체 피검자 집단에서 FIQ와 DF의 상관은 통계적으로 매우 유의미하였다,  $r(199) = .500, p < .001$ . FIQ와 DB의 상관도 매우 유의미하였다,  $r(199) = .512, p < .001$ . 그러나 FIQ와 DF-DB의 상관은 전혀 유의미하지 않았다,  $r(199) = -.043, ns$ . 표 6은 연령별로 FIQ와 DF의 상관, FIQ와 DB의 상관, FIQ와 DF-DB의 상관을 보여준다. 전체적으로 많은 수의 상관계수를 검증하므로  $p < .01$ 인 것만 유의미한 것으로 수용하였다. 표 6에서 볼 수 있듯이 FIQ와 DF의 상관은 6개 중 4개

표 5. 학력 수준별 DF, DB, DF-DB점수의 평균과 표준편차

학력(년)	n	DF		DB		DF-DB	
		M	SD	M	SD	M	SD
13이상	50	7.94	1.08	5.84	1.57	2.10	1.69
10-12	97	7.48	1.23	5.47	1.32	2.01	1.34
7-9	33	6.73	1.53	4.27	1.23	2.45	1.46
0-6	21	4.76	1.45	2.86	1.24	1.90	1.37

표 6. 연령 수준별 FIQ와 DF, FIQ와 DB, FIQ와 DF-DB의 상관계수

연령(세)	n	FIQ와 DF	FIQ와 DB	FIQ와 DF-DB
16-19	34	.669*	.423	.212
20-24	41	.240	.331	-.125
25-34	43	.052	.432*	-.329
35-44	23	.651*	.599*	-.017
45-54	34	.589*	.753*	-.219
55-64	26	.690*	.765*	.066

\* $p < .01$ .

가 유의미하였다. FIQ와 DB의 상관 역시 6개 중 4개가 유의미하였다. 그러나 FIQ와 DF-DB의 상관은 6개 중 어느 하나도 유의미하지 않았다.

#### DF-DB의 기준

앞의 분석들은 연령, 학력, 지능 변인들이 DF나 DB에는 뚜렷한 효과를 미치는 반면에 DF-DB에는 거의 효과가 없음을 보여준다. 그러므로 DF-DB의 기준은 연령, 학력, 지능 수준을 구분하지 않고 단일하게 작성하였다. 기준의 점수 단위는 백분위를 사용하였다. 이렇게 작성된 DF-DB의 백분위 기준을 표 7의 원편에 제시하였다. DF-DB 기준의 사용례를 든다면 표 3에 제시된 사례의 경우 DF-DB점수가 4이므로 백분위 16.4%에 해당한다. 임상에서 정상/비정상의 판별을 위한 통계적 기준으로는 다음을 잠정적으로 제시할 수 있다:  $x \geq 20\text{ile}$ 은 정상,  $10\text{ile} \leq x < 20\text{ile}$ 은 경계선,  $x < 10\text{ile}$ 은 비정상. 이 기준을 적용한다면 DF-DB점수가 3이하의 정상, 4는 경계선, 5 이상은 비정상에 해당한다. 그러므로 표 3에 제시된 사례의 DF-DB점수는 경계선 수준에 해당한다. 표 7의 중간에는 참고적으로 미국 WAIS-III의 해당 기준(Wechsler, 1997, p.

표 7. 본 표집, WAIS-III, K-WISC-III에서 제시된 DF-DB의 백분위 기준

DF-DB	본 연구 (N = 201)	WAIS-III (N = 2450)	K-WISC-III (N = 201)
7	0	0	0
6	1.5	.3	2.4
5	6.5	2.1	6.7
4	16.4	9.4	17.7
3	35.3	26.7	32.1
2	63.2	55.8	56.5
1	90.0	83.1	86.6
0	97.0	96.2	94.3
-1	99.5	99.1	98.6
-2	100.0	99.8	99.0
-3		100.0	99.5
-4			100.0

주. 본 연구의 표집 연령은 16~64세, WAIS-III의 표집 연령은 16~84세, K-WISC-III의 표집 연령은 16세임.

213)을 제시하였다. 이 기준은 16~84세의 통합 기준으로 2450명의 표집수에 근거한다. 표 7의 오른쪽에는 참고적으로 K-WISC-III의 해당 기준(곽금주, 박혜원, 김청택, 2001, p. 277)을 제시하



였다. 이 기준은 16세의 기준으로 201명의 표집수에 근거한다. 본 연구의 기준, WAIS-III의 기준, K-WISC-III의 기준을 비교하면 각각 문화 및 연령적 차이가 큼에도 불구하고 상당히 유사함을 볼 수 있다. 예를 들어 DF-DB점수가 3이면 본 연구, WAIS-III, K-WISC-III에서 각각 35.3, 26.7, 32.1의 백분위로 모두 '정상'에 해당한다. 반면에 DF-DB점수가 5이면 본 연구, WAIS-III, K-WISC-III에서 각각 6.5, 2.1, 6.7의 백분위로 모두 '비정상'에 해당한다. 이러한 세 기준간의 유사성은 연령이나 문화적 요인이 DF-DB의 분포에 미치는 효과가 근소한 수준임을 보여준다.

## 논 의

K-WAIS는 국내에서 가장 많이 사용되는 임상용 지능검사이다. 그러므로 K-WAIS 소검사의 하나인 숫자외우기도 많은 임상심리검사에서 실시되고 있다. 숫자외우기에서 DF와 DB의 개별적 채점 및 비교의 필요성은 많은 연구자들이 오래 전부터 제기해 온 문제이다. 이러한 주장을 수용하여 WAIS-III는 요강 자체가 DF-DB에 관한 기준을 제시하고 있다. 그러나 국내에서는 아직 DF-DB의 기준이 발표된 바 없어서 DF와 DB의 차이를 객관적으로 비교하기가 어려웠다. 이에 본 연구에서는 한국인 정상인 201명의 자료를 기초로 DF-DB의 기준 제작을 시도하였다. 기준 제작에 앞서 연령, 학력, 지능 변인이 DF-DB에 미치는 효과를 검증한 결과 극히 작았으며 통계적으로 유의미하지 않았다. 그러므로 DF-DB의 기준은 연령, 학력, 지능 수준의 구분 없이 단일하게 작성하였다. 기준의 형식은 WAIS-III의 형식을 적용하여서 DF의 자릿수에서 DB의 자릿수를 뺀 결과를 백분위 점수화하였다. 본 연구에서

제시한 기준에 따르면 DF-DB점수가 3이하이면 백분위 35.3%이상으로 '정상' 수준이며, 4면 백분위 16.4%로 '경계선' 수준이고, 5이상은 백분위 6.5%이하로 '비정상' 수준이다. 그러므로 예를 들어 DF점수가 7인 사례에서는 DB점수가 3이면 경계선수준에 해당한다. 그러나 이러한 진단 기준은 물론 절대적인 것이 아니므로 다른 검사 결과들을 고려하여 융통성 있게 적용되어야 할 것이다.

앞서 DF는 단기기억 과제이지만 DB는 단기기억과 더불어 작업기억이 요구되는 과제임을 언급하였다. 그러므로 DF-DB는 작업기억의 한 지표로 활용될 수 있다. DF-DB를 작업기억의 지표로 활용할 시 유의할 점의 하나는 피검자의 DF 수행 수준이 미치는 효과다. 본 연구의 기준에 따르면 DF-DB점수가 최소 4는 되어야지 '경계선'이나 '비정상'이라는 해석이 가능하다. 그런데 DB의 실질적인 최소 점수가 2임을 감안하면 4점의 차이가 나려면 DF점수가 최소 6은 되어야 한다. 그러므로 DF점수가 5이하인 사례들에서는 DF-DB가 '비정상'일 가능성이 없으며, DF점수가 6인 사례에서도 DF-DB의 민감도는 상당한 제한을 받는다. 결국 DF와 DB의 비교를 통한 작업기억의 정상/비정상 진단은 DF가 정상 수준인 피검자에서만 실질적인 타당성이 있다. 그런데 이러한 제한점은 '작업기억'이라는 개념 자체가 단기기억의 개념을 내포하는 것에 근본 원인이 있다고 할 수 있다. 앞서 언급하였듯이 작업기억이란 일시적으로 유지되고 있는 정보에 대해 정신적 조작을 가할 수 있는 능력을 지칭한다. 따라서 일시적으로 정보를 유지하는 능력, 즉 단기기억 자체가 비정상 수준인 피검자에서 작업기억의 정상/비정상성을 논하기란 극히 어려운 일이다. 이러한 관점에서 단기기억이 비정상 수준인 피검자에서 작업기억의 정상/비정상성을 측정하

기가 어려운 것은 '숫자외우기' 검사에만 국한된 것이 아니라 여타의 작업기억 검사에도 거의 공통적으로 해당되는 것으로 볼 수 있다.

임상에서는 검사 시간의 제약 때문에 '최소한의 검사로 최대한의 정보'를 얻을 필요성이 매우 크다. 그런데 작업기억에 특화된 검사들(예, Paced Auditory Serial-addition Task; Gronwall, 1977)은 보통 검사시간이 상당히 길다. 또한 대부분의 검사가 국내 기준이 없다. 그러므로 숫자외우기를 작업기억검사로 활용하는 것은 검사 시간의 절약이나 기준의 이용성 측면에서 상당한 장점을 가진다. 반면에 다음과 같은 제한점도 가진다. 첫째로 DB가 작업기억에 특화된 다른 검사들에 비하면 민감도가 낮다는 증거가 있다. 예를 들어 Stone, Gabrieli, Stebbins와 Sullivan(1998)은 정신분열 증환자군과 통제군을 듣기폭(listening), 계산폭(computation span), DB의 3가지 작업기억과제에서 비교하였다. 두 집단은 듣기폭과 계산폭에서는 매우 큰 차이를 보였지만 DB에서는 비교적 작은 차이를 보였다. DB의 민감도가 상대적으로 낮은 것은 작업기억에 대한 부하량(loadings)이 비교적 낮기 때문일 것이다. 둘째로 DB가 작업기억뿐 아니라 다른 정신적 과정도 반영할지 모른다. 예를 들어 DF에는 언어적 과정만 개입되지만 DB에는 언어적 과정과 시공간적 과정이 모두 개입된다는 주장이 제기된 바 있다(Banken, 1985; Black, 1986; Costa, 1975; Hoshi et al., 2000; Rapport et al., 1994; Robertson, 1990). 이러한 주장에 따르면 DB에는 불려준 숫자들을 심상화시킨 후 거꾸로 주사(scanning)하는 과정이 개입된다. 어쨌든 인지검사의 결과를 한 이론적 관점에서만 해석하는 것은 그에 따른 제한이 불가피한 만큼 DF-DB가 작업기억이 아닌 다른 정신 과정을 반영할 가능성도 항상 고려되어야 할 것이다.

본 연구에 따르면 연령이 DF-DB에 미치는 효

과는 전혀 유의미하지 않았다. 여러 선행 연구에서도 연령이 DF-DB에 미치는 효과가 유의미하지 않다는 결과가 얻어진 바 있다(Verhaeghen, Marcoen, & Goossens, 1993; Grégoire & Van der Linden, 1997에서 재인용). 연령이 DF-DB에 미치는 효과가 유의미하지 않은 것은 단기기억과 작업기억의 '노화 속도'가 유사한 수준임을 시사한다. 예를 들어 만약 단기기억 보다 작업기억의 노화 속도가 더 급격하다면 연령이 DF-DB에 미치는 효과가 유의미했을 것이다. 그러나 본 연구와 같은 횡단적(cross-sectional) 연구에서는 연령의 효과를 '노화의 효과'로 단순 해석할 수 없다. 예를 들어 본 연구에서 연령별로 교육 수준이 달랐으므로 연령의 효과는 노화의 효과이기도 했지만 학력의 효과이기도 했다. 본 연구의 분석에 따르면 연령이 DF-DB에 미치는 효과는 학력 차이를 공변인으로 통제했을 경우에도 유의미하지 않다. 그렇지만 이러한 통계적 통제를 실시한 경우에도 횡단적 연구에서의 연령 효과를 노화 효과로 해석하는 것에는 많은 어려움이 따른다. 그러므로 단기기억과 작업기억의 노화 속도에 차이가 있는지 없는지를 검증하는 것은 종단적(longitudinal) 연구에서 보다 합리적으로 검증될 수 있는 문제이다. 또한 종속변수의 경우도 다양한 종류의 단기기억/작업기억 검사가 포함되어야지 보다 타당한 결론이 도출될 수 있을 것이다. 그러므로 본 연구는 DF-DB의 기준 제시라는 본래의 실용적인 목적에 충실하고 연령이 DF-DB에 대한 미치는 효과에 대한 이론적인 해석은 더 이상 언급하지 않는다.))

1) 아울러 본 연구에서는 학력이나 지능 수준이 DF-DB에 미치는 효과도 유의하지 않았다. 이러한 결과의 이론적 해석 역시 여러 변인들이 혼입되어 있는 본 표집의 특성상 많은 어려움이 따르므로 더 이상 논의하지 않는다.

신경심리학적 연구들에 따르면 전두엽이 DB의 수행에는 매우 중요하지만 DF의 수행에는 상대적으로 덜 중요하다는 증거가 있다. 예를 들어 Canavan, Passingham, Marsden, Quinn, Wyke와 Polkey(1989)는 전두엽손상군과 통제군 간에 DB 수행에는 유의미한 차이가 있지만 DF 수행에는 차이가 없음을 보고하였다. Kiefer 등(2002)과 Stone 등(1998)은 정신분열증환자군과 통제군 간에 DB 수행에는 유의미한 차이가 있지만 DF 수행에는 차이가 없음을 보고하였다. 이 결과는 정신분열증이 전두엽이상 의심되는 임상집단이라는 점에서(Weinberger & Berman, 1998) 전두엽이 DB의 수행에는 중요하지만 DF의 수행에는 덜 중요하다는 가설을 지지한다. Hoshi 등(2000)은 광학적 단층촬영(optical tomography)을 이용한 연구에서 DF 수행 시에 비하여 DB 수행 시에 전두엽이 보다 활성화(activation)됨을 보고하였다. 이러한 연구 결과들은 DB에는 작업기억이 요구되지만 DF에는 요구되지 않는다는 가설을 간접적으로 지지한다. 왜냐하면 작업기억은 신경심리학적 관점에서는 전두엽-관리기능(frontal-executive function)의 한 요소로 볼 수 있기 때문이다(Baddeley & Della Sala, 1998; D'Esposito, 2001; Malloy & Richardson, 1994; Rains, 2002). 김홍근(2001a)은 전두엽-관리기능 측정의 이론적 모형을 제시하며 '전두엽-관리기능 의존도'라는 개념을 도입한 바 있다. 이 개념을 사용한다면 DF는 전두엽-관리기능 의존도가 낮은 과제의 한 예이고 DB는 전두엽-관리기능 의존도가 높은 과제의 한 예이다. 그러므로 DF와 DB의 차이는 신경심리학적 관점에서는 전두엽-관리기능의 한 지표로 활용될 수 있다.

마지막으로 본 연구의 주요 제한점을 논의하면 다음과 같다. 첫째로는 연령, 학력, 지능 수준 등 DF-DB에 잠재적으로 영향을 미칠 수 있는 변

인들이 균형적으로 층화표집(stratified sampling)되지 않은 점을 들 수 있다. 예를 들어 연령별 표본 수는 균일하지 않았으며 전체 FIQ 평균도 107.2로 다소 높아서 편향된 표집을 이루었다. 그렇지만 본 연구의 분석에 따르면 이러한 변인들이 DF-DB에 미치는 효과는 극히 작으므로 DF-DB기준에 심각한 편향을 끼치지 않았을 것으로 예상된다. 이는 본 연구, WAIS-III, 그리고 K-WISC-III의 기준이 유사한 것에서도 간접적으로 시사된다. 그렇지만 기준 연구에서는 각종 변인들이 균형적으로 표집되는 것이 가장 이상적인 만큼 추후 연구에서는 보다 정밀한 표집에 근거한 DF-DB의 기준 설정이 요구된다. 본 연구의 또 다른 주요 제한점은 DF와 DB의 개별적 기준을 제시하지 못한 것이다. 본 연구의 분석에 따르면 연령, 학력, 지능 변인 등이 DF나 DB에 미치는 효과는 통계적으로 매우 유의미하다. 그러므로 DF나 DB의 기준은 연령, 학력, 지능 수준별로 제작해야 하는데 본 연구의 표집수는 그러한 기준을 만들기에 절대적으로 부족하였다. 그러나 DF와 DB의 비교는 DF와 DB의 개별적인 기준만으로는 불충분하고 최종적으로는 DF-DB의 기준에 의거하게 된다. 이는 마치 VIQ와 PIQ를 비교할 시 VIQ와 PIQ에 대한 개별적인 기준으로 불충분하고 최종적으로는 VIQ와 PIQ의 차이에 대한 기준에 의거하는 것과 같다(김홍근, 1999a). 그러므로 DF와 DB에 대한 개별적 기준을 제시하지 못하고 DF-DB의 기준만 제시한 것은 본 연구의 제한점이긴 하지만 DF와 DB를 비교하려는 시도에서 결정적 장애가 되지는 않는다.

## 참고문헌

곽금주, 박혜원, 김청택 (2001). K-WISC-III 지

- 침서. 서울: 도서출판 특수교육.
- 김중술, 이용승, 이민식 (1994). K-WAIS의 단축형에 관한 연구. *정신의학*, 19, 121-126.
- 김홍근 (1999a). K-WAIS의 활용을 위한 세 가지 고찰. *한국심리학회지: 임상*, 18, 179-186.
- 김홍근 (1999b). Rey-Kim 기억검사: 해설서. 대구: 도서출판 신경심리.
- 김홍근 (2001a). Kims 전두엽-관리기능 신경심리 검사: 해설서. 대구: 도서출판 신경심리.
- 김홍근 (2001b). 병전 지능 추정의 허와 실. *한국심리학회지: 임상*, 20, 145-154.
- 김홍근 (2001c). 병전 지능 추정: 2001년 이후. *한국심리학회지: 임상*, 20, 155-164.
- 김홍근 (2002a). 간질의 임상신경심리학적 평가. 2002년 한국임상심리학회 춘계학술대회 발표집, pp. 89-117. 5월 25일. 전주: 전북대학교 산학협동관.
- 김홍근 (2002b). K-WAIS의 3요인 해석을 위한 기준 연구. *한국심리학회지: 임상*, 21, 631-645.
- 염태호, 박영숙, 오경자, 김정규, 이영호 (1992). K-WAIS 실시요강. 서울: 한국 가이던스.
- Baddeley, A. (1995). Working memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp.755-764). Cambridge, MA: MIT Press.
- Baddeley, A., & Della Sala, S. (1998). Working memory and executive control. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.), *The prefrontal cortex: executive and cognitive functions* (pp. 9-21). New York: Oxford University Press.
- Banken, J. A. (1985). Clinical utility of considering digits forward and digits backward as separate components of the Wechsler Adult Intelligence Scale-Revised. *Journal of Clinical Psychology*, 41, 686-691.
- Black, F. W. (1986). Digit repetition in brain-damaged adults: Clinical and theoretical implications. *Journal of Clinical Psychology*, 42, 770-782.
- Canavan, A. G. M., Passingham, R. E., Marsden, C. D., Quinn, N., Wyke, M., & Polkey, C. E. (1989). Sequencing ability in Parkinsonians, patients with frontal lobe lesions and patients who have undergone unilateral temporal lobectomies. *Neuropsychologia*, 27, 787-798.
- Conklin, H. M., Curtis, C. E., Katsanis, J., & Iacono, W. G. (2000). Verbal working memory impairment in schizophrenia patients and their first-degree relatives: Evidence from the digit span task. *American Journal of Psychiatry*, 157, 275-277.
- Costa, L. D. (1975). The relation of visuospatial dysfunction to digit span performance in patients with cerebral lesions. *Cortex*, 11, 31-36.
- D'Esposito, M. (2001). Functional neuroimaging of working memory. In R. Cabeza & A. Kingstone, *Handbook of functional neuroimaging of cognition* (pp.293-327). Cambridge, MA; MIT Press.
- Foster, J. K., Black, S. E., Buck, B. H., & Bronskill, M. J. (1997). Ageing and executive functions: A neuroimaging study. In P. Rabbit (Ed.), *Methodology of Frontal and executive function* (pp. 117-134). Hove, U.K.: Psychology Press.
- Grégoire, J., & Van der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, 4, 140-149.
- Gronwall, D. M. A. (1977). Paced Auditory Serial-Addition Task: A measure of recovery from

- concussion. *Perceptual and Motor Skills*, 44, 367-373.
- Harrison, P. L., Kaufman, A. S., Hickman, J. A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. *Journal of Psychoeducational Assessment*, 6, 188-198.
- Hayslip, B., & Kennelly, K. J. (1982). Short-term memory and crystallized-fluid intelligence in adulthood. *Research on Aging*, 4, 315-332.
- Hoshi, Y., Oda, I., Wada, Y., Ito, Y., Yamashita, Y., Oda, M., Ohta, K., Yamada, Y., & Tamura, M. (2000). Visuospatial imagery is a fruitful strategy for the digit span backward task: a study with near-infrared optical tomography. *Cognitive Brain Research*, 9, 339-342.
- Kaplan, E., Fein, D., Morris, R., & Delis, D. C. (1991). *WAIS-R as a Neuropsychological Instrument*. San Antonio, TX: The Psychological Corporation.
- Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston, MA: Allyn & Bacon.
- Kiefer, M., Apel, A., & Weisbrod, M. (2002). Arithmetic fact retrieval and working memory in schizophrenia. *Schizophrenia Research*, 53, 219-227.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
- Malloy, P. F., & Richardson, E. M. (1994). Assessment of frontal lobe functions. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 6, 399-410.
- Rains, G. D. (2002). *Principles of human neuropsychology*. New York: McGraw Hill.
- Ramsay, M. C., & Reynolds, C. R. (1995). Separate digits tests: A brief history, a literature review, and a reexamination of the factor structure of the test of memory and learning (TOMAL). *Neuropsychology Review*, 5, 151-171.
- Rappport, L. J., Webster, J. S., & Dutra, R. L. (1994). Digit span performance and unilateral neglect. *Neuropsychologia*, 32, 517-525.
- Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation* (2nd ed). Tucson, AZ: Neuropsychology Press.
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie*, 28, 286-340.
- Robertson, I. H. (1990). Digit span and visual neglect: A puzzling relationship. *Neuropsychologia*, 28, 217-222.
- Stone, M., Gabrieli, J. D. E., Stebbins, G. T., & Sullivan, E. V. (1998). Working and strategic memory deficits in schizophrenia. *Neuropsychology*, 12, 278-288.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journal of Gerontology: Psychological Sciences*, 48, 157-171.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997). *Administration and Scoring Manual for the Wechsler Adult Intelligence Scale-Third Edition*. San Antonio, TX: The Psychological Corporation.
- Weinberger, D. R., & Berman, K. F. (1998).

한국심리학회지 : 임상

Prefrontal function in schizophrenia: Confounds and controversies. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz, *The prefrontal cortex: Executive and cognitive functions*. New York: Oxford University Press.

원 고 접 수 일 : 2002. 6. 27

수정원고접수일 : 2002. 11. 15

게 재 결 정 일 : 2002. 12. 10

## Korean norm for the difference between digits forward and digits backward

Hongkeun Kim

Department of Rehabilitation Psychology  
Daegu University

Taejin Park

Department of Psychology  
Chonnam National University

The traditional method of scoring Digit Span Test is to sum the scores of digits forward (DF) and digits backward (DB). However, several researchers have articulated the need for scoring DF and DB separately and comparing the two. The WAIS-III accepts the view and includes a U.S. norm for the difference between DF and DB. However, no Korean norm for the difference between DF and DB has been published yet. Thus, the aim of the present study was to provide a Korean norm for the difference between DF and DB. A total of 201 Korean normal subjects were administered DF and DB. The effects of age, years of education, or FIQ on the difference between DF and DB were minimal and statistically nonsignificant. Thus, the norm for the difference between DF and DB was constructed regardless of age, years of education, or FIQ. The unit of measurement for the norm was the number of digits forward minus the number of digits backward. The difference of 3 or smaller corresponded to 35.3%ile and may be classified as normal; the difference of 4 corresponded to 16.4%ile and may be classified as borderline; and the difference of 5 or greater corresponded to 6.5%ile and may be classified as abnormal. The DF requires short-term maintenance of information, whereas the DB requires short-term manipulation as well as short-term maintenance of information. Thus, the DF is a short-term memory task, whereas the DB is both a short-term memory task and a working memory task. Viewed in this way, the difference between DF and DB could be used as an index of working memory capacity.

*Keywords* : digit span, norm, working memory