

한국 Bayley 영유아 발달검사수행의 안정성: 종단 연구*

The Stability of K-BSID-II Performance: A Longitudinal Study*

울산대학교 아동·가정복지전공
교수 박혜원
연구원 신민선

Dept. of Child & Family Welfare, Univ. of Ulsan,
Professor : Hyewon Park Choi
Graduate student : Minsun Shin

● 목 차 ●

- | | |
|------------|-------------|
| I. 서론 | IV. 논의 및 결론 |
| II. 연구 방법 | 참고문헌 |
| III. 연구 결과 | |

<Abstract>

This study followed the standardization study pool of K-BSID-II to examine the stability of K-BSID-II. One hundred and ninety five infants aged from 1-42 months tested twice with intervals of 5-16 month between tests. All participants had been full-term babies without any medical complications. There were no significant differences between the groups which were divided in terms of demographic variables such as mothers' education, infants' birth order and gender. The stability of K-BSID-II was significant and that of the mental scale was lower than that of the psycho-motor scale: There were high stabilities between raw scores in two sessions($r=.91$ for the mental scale and $r=.89$ for the psycho-motor scale) but there were significant but low correlations between mental index scores($r=.18$, $p<.05$) and psycho-motor index scores($r=.17$, $p<.05$). The stability was increased with initial test age and mothers' education. It was recommended for testers and researchers to use both raw scores and index scores for better interpretations in future research.

주제어(Key Words): 한국 베일리 영유아 발달검사(K-BSID-II), 수행 안정성(stability), 1-42개월 영유아(1-42 month old infants), 인지발달지수(MDI), 동작발달지수(PDI)

Corresponding Author: Hyewon Park Choi, Department of Child & Family Welfare, College of Living Science, University of Ulsan, San 29, Mugeo-dong, Nam-ku, Ulsan, 680-749, Korea Tel: 82-52-259-2367 Fax: 82-52-259-2888 Email: hyewonc@mail.ulsan.ac.kr

* 이 논문은 2003년도 울산대학교 교비연구비지원에 의하여 연구되었음.

1. 서론

1960년대 이후 본격적으로 시작된 영아인지에 대한 연구 결과 영유아기는 인간이 살아가는데 필요한 기본적인 능력과 기술을 발달시키는 중요한 시기로 밝혀지고 있다. 무엇보다도 영아는 선천적인 지각 능력과 기억 능력을 가지고 태어난다는 것이 적절한 연구 방법과 측정 도구를 통해 밝혀지고 있다(Snow & McGaha, 2003). 또한 의학의 발달로 영아기 사망률이 낮아지고 조산아의 생존률이 높아지면서 발달장애에 대한 관심이 고조되면서 정확한 영아기 발달 측정에 대한 관심 또한 매우 높다.

국내에서도 특수교육법의 개정으로 장애아를 조기진단하고자 하는 요구가 증가하고 있고, 영유아의 잠재 가능성을 제발하고자 하는 교육 흐름에 따라 영유아 발달에 대한 기초 자료가 시급히 요구되고 있다. 현재 영유아 발달을 진단하는 도구들이 몇몇 사용되고 있으나 주로 장애아 발달을 진단하는 검사이거나(강순규, 조윤경, 1994; 김승국, 김옥기, 1985) 간편용으로 개발되어(김숙현, 이경옥, 2001; 대한소아과학회, 2002), 우리나라 영유아의 발달을 보다 체계적으로 진단하고 연령별 발달을 기준지향적으로 평가하는 객관적인 발달검사가 필요한 실정이다.

영유아발달검사는 주로 미국에서 개발되어 왔는데 간편한 선별검사인 DDST, Brazelton 검사 등이 있고 세계적으로 가장 많이 사용되며 발달 지수까지 산출할 수 있는 것으로는 Bayley 영아발달검사(Bayley Infant Scale of Development-BSID)(Bayley, 1969)가 있다. 1930년대부터 개발(Bayley, 1932; 1933a; 1933b; 1935; 1936)이 시작된 BSID는 1969년 미국에서 표준화된 이래, 일본, 중국 등 동남아에서도 표준화되거나 변안된 검사가 오래 전부터 사용되고 있다(Aggarwal, Chaudhari, Bhave, Pandit, & Barve, 1998; Brooks-Gunn, & Weinraub, 1983). 그 후 또다시 20여년간의 오랜 연구를 통해 수정보완 과정을 거쳐 1993년에 2판인 Bayley Scale of Infant Development-II가 제작되었다(Bayley, 1993; Campbell, Siegel, Parr, & Ramey, 1986). 2판은 그 대상연령이 2-30개월용이던 1판에서 1-42개월용으로 확장되었다. BSID-II는 미국 전가정을 대표하는(성, 인종, 교육수준, 지역에 따라 층화표집) 1,700명의 아동을 대상으로 표준화되었으며 BSID, DDST, DAS, McCarthy Scale 등과의 공인 타당도가 증명되었다(Baird & Folsom, 1993; Black & Matula, 1999; Goldstein, Fogle, Wieber, & O'Shea, 1995; Johnson, Cook, & Kullman, 1992; King & Seegmiller, 1973; McChesne, 1995; Nellis & Gridley, 1994).

우리나라에서의 Bayley 영유아발달검사에 대한 연구를 살펴보면 1980년대이후 꾸준히 척도화를 위한 연구와 타당도에 대한 연구가 수행되었다(민동욱, 박혜원, 2003; 배운

희, 2001; 이일섭, 1992; 정문자, 1983; 정문자, 이은혜, 박경자, 1993; 정영미, 1992; 제경숙, 1984). 그러나 이러한 연구들은 Bayley 검사의 유용성에 대해 보고하고 있으나 한국 영유아에 대한 척도화를 이루지 못하였다. 이에 따라 2004년 조복희, 박혜원은 예비연구(박혜원, 조복희, 최호정, 2003)를 통해 개발된 한국 Bayley 영유아발달검사(Korean Bayley Scale of Infant Development-II, K-BSID-II)를 사용하여 정상분만한 1,700명의 영유아를 대상으로 표준화 연구를 실시하였다. 동 연구에서는 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 18, 21, 24, 27, 30, 36, 42개월의 전체 17개 연령집단의 남녀 각 50명을 대상으로 규준을 마련하는 자료를 수집하였다. 조복희, 박혜원(2004)은 미국 규준을 사용하여 한국 영유아의 수행을 분석하였는데 한국의 영유아의 수행은 인지발달 척도에서는 미국의 영유아와 유사하였으나 동작발달 척도에서는 미국의 영유아에 비해 유의하게 높게 나타남으로써 한국 규준마련의 필요성을 보고하였다.

그런데 Bayley 영유아 발달검사 등의 영아발달 검사가 보다 유용하기 활용되기 위해서는 검사 결과의 안정성에 대한 자료가 필요하다. 일반적으로 지능검사의 안정성, 특히 어린 아동의 지능검사의 안정성은 논쟁의 대상이 되어왔다(Bayley, 1970; Rubin, 1996; Sabourin, 1996). 영아기 지능의 종단 연구를 최초로 수행한 Bayley와 그녀의 동료들은 1929년에 착수하여 30년 이상 이루어진 인간의 성장 발달에 관한 종단적 연구(Bayley, 1970)에서 다음과 같은 결론을 내렸다. 첫째, 개인의 지능은 선천적으로 주어지는 것이 아니라, 환경 조건에 따라서 변화하는 인간 특성 중의 하나라는 것이다. 둘째, 여러 월령에 걸쳐서 IQ를 측정할수록 IQ가 높아질수록 IQ가 안정되어 간다는 것을 발견하였다. 또한 IQ의 변화는 출생 후 초기 수년에 걸쳐서 가장 크게 일어나는 것으로 보고되었다. 따라서 적어도 영아기에 있어서는 이후의 발달을 예측할 수 있는가와 관련된 안정성이 만족스럽지 못하였다. 셋째, Bayley(1965)의 연구 자료에 따르면 지능은 10대 후반이나 혹은 20대 초반에 절정을 이루는 것이 아니라 적어도 50세에 이르기까지 즉 전생애동안 계속 성장한다는 것이다.

이후의 안정성 연구에서도 주목할 만한 몇 가지 결론들이 얻어졌다(Jencks, 1972). 첫째 극단적으로 낮은 점수를 제외하고는, 영아기 검사에서의 수행은 성장후의 IQ를 예언하지 못한다. 즉 영아기 수행 점수와 그 이후의 수행 점수간의 상관은 "0"에 가깝다. 그러나 영아기 이후의 월령간 수행점수들은 서로 유의한 상관을 보인다. 일반적으로 IQ 점수는 비교되는 월령 간격이 짧을수록 높고 또한 처음 검사시의 월령이 높을수록 안정성이 높다. 따라서 안정성은 연령과 함께 증가한다고 볼 수 있다.

그런데 이렇듯 영아기에 측정된 지능수행이 이후의 발달과 유의한 상관을 보이지 않는 이유의 하나는 영아기에 측정하는 행동 또는 수행과 영아기 이후에 측정하는 행동 특성간의 차이때문일 수 있다. 즉 검사의 내용자체가 상당히 달라지기 때문에 검사 수행간 상관이 낮을 수 있다(Miller, 1987). 이는 실제로 영아기의 지적 능력은 이후의 지적 능력과 상관이 있음에도 불구하고 이를 측정하는 도구의 차이 때문에 안정성이 낮게 산출될 수 있음을 시사한다. 둘째는 실제로 영아기의 지능이 안정적이지 않기 때문일 수 있다. 1993년 표준화된 BSID-II는 1960년 이후 활발히 연구되어 온 영아기 재인지역, 신기성 선호성(Brody, 1993; McChesne, 1995) 등에 관한 문항을 첨가하여 안정성을 증가시키기 위해 노력하였다. 본 연구의 목적은 이러한 BSID-II를 변안하고 한국 영유아에 맞도록 수정한 K-BSID-II의 수행안정성을 살펴보고 이러한 안정성에 영향을 미치는 요인을 탐색하고자 한다. 이는 K-BSID-II의 유용성을 밝히는데 중요한 역할을 할 수 있을 것이다.

II. 연구 방법

1. 연구대상

본 중단연구의 대상아동은 한국 베일리 영유아발달검사 표준화과정에 참여한 울산 거주 아동 493명중 재검사안내에 협조한 195명이었으며 1차 검사 후 5개월이상 경과한 이후에 2차 검사를 실시하였다. 아동의 1, 2차 검사시 월령분포와 인구통계학적 정보는 <표 1, 2>와 같다.

<표 2>에서 보는 바와 같이 표준화 연구에서 영유아의 층화표집 계획시 영유아의 월령, 성별 그리고 어머니의 학력만 고려하였기 때문에 이 세 변인을 제외한 다른 변인은 결측치가 많았다. 중단연구에 참여한 아동 195명중 남아가 102명(52.3%), 여아가 93명(47.7%)이며 출생순위는 첫째가 57명(47.1%), 둘째 이상이 64명(52.9%)이었다. 검사아동의 어머니학력은 고졸이하가 130명(66.6%)으로 가장 많았고 아버지학력은 대졸이 35명(54.7%)으로 가장 많았으며 그 다음이 26명(40.6%)인 고졸이하로 나타났다.

2. 연구도구

K-BSID-II는 인지척도(Mental Scale), 동작척도(Motor Scale), 행동평정척도(BRS: Behavior Rating Scale)의 세 가지 척도로 구성되어 있으며 인지척도는 179문항, 동작척도는 114문항, 행동평정척도는 30문항으로 구성되어 있다. 행동평정 척도의 경우 월령집단에 따라 측정하는 하위 척도가 달라져 중단적 분석에 적절하지 않으므로, 본 중단분석에

<표 1> 1차 검사 및 2차 검사시 월령별 아동분포 N(%)

검사시 월령	1차 검사	2차 검사
6개월 미만	81(41.5)	0(0.0)
6개월~12개월	49(25.1)	58(29.7)
13개월~24개월	47(24.1)	80(41.1)
25개월~42개월	18(9.3)	57(29.2)
합 계	195(100.0)	195(100.0)

<표 2> 대상 아동의 일반적 특성

변 수	구 분	빈도(%)
아 동 성	남	102(52.3)
	여	93(47.7)
	계*	195(100.0)
출 생 순 위	첫째	57(47.1)
	둘째	55(45.5)
	셋째	9(7.4)
	계	121(100.0)
부 학 력	고졸이하	26(40.6)
	대졸	35(54.7)
	대학원이상	3(4.7)
	계	64(100.0)
모 학 력	고졸이하	130(66.6)
	대졸	64(32.8)
	대학원이상	1(0.6)
	계	195(100.0)

* missing으로 인하여 계에 차이가 있음.

서는 제외하였다. 인지 척도와 동작척도는 22개 연령집단별로 해당연령의 문항세트를 제시하도록 되어 있으며 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14-16, 17-19, 20-22, 23-25, 26-28, 29-31, 32-34, 35-37, 38-42개월용으로 나뉜다. 예비 연구결과 인지척도의 채점자간 일치도는 $r=.99$, 동작척도의 일치도는 $r=.99$ 였으며 표준화연구에서 보고된 것은 각각 $r=.91$ 및 $r=.99$ 에 해당하였다. 또한 본 도구의 지침서내 보고에 따르면(박혜원, 조복희, 미출간자료) K-BSID-II와 한국 웨슬러 유아지능검사(K-WPPSI)와의 상관은 $r=.65$ (인지척도), $r=.52$ (동작척도) 등으로 공준 타당도 등도 입증되었다. 그 외에 세부적인 도구에 대한 설명은 박혜원, 조복희, 최호정(2003) 및 조복희, 박혜원(2004)을 참조하라.

3. 연구절차 및 자료분석

중단연구를 위해 한국 베일리 영유아발달검사의 표준화에 참여하였던 아동중 울산에 거주하고 있는 아동 493명을 대상으로 우편을 통해 2차 검사에 대한 홍보를 실시하고 검사 희망 접수를 받았다. 2차 검사를 실시하기 위해서 참여한 검사자는 총 7명이며 아동학 석사과정중의 학생으로 훈련을 거쳐 검사에 대한 이론적 지식, 실시방법 및 채점방법에 있

어서 숙련된 검사자이다. 1차 검사 기간은 2002년 3월에서 2003년 2월사이였으며 2차 검사 기간은 2003년 3월에서 2003년 8월까지였다. 아직 K-BSID-II의 규준은 제작과정 중에 있어, 본 연구의 분석에 사용된 수행점수는 표준화연구와 마찬가지로 미국의 규준을 사용하였다. 수집된 자료는 SPSS/PC+ 10.0 프로그램을 활용하여 분석하였다.

III. 연구 결과

한국 Bayley 영유아발달검사의 수행안정성을 분석하기 전에 전체 아동을 대상으로 인구 통계학적 변인에 따라 수행이 달라지는지 살펴보기 위해 성, 출생순위 그리고 어머니의 학력에 따라 수행차이를 분석한 결과 이러한 변인에 따른 유의한 차이를 보이지 않았다.

〈표 3〉은 성별에 따른 K-BSID-II 척도별 수행점수이다. 〈표 3〉을 살펴보면 남아의 1차 인지지수는 98.69(SD=8.24), 여아의 1차 인지지수는 96.47(SD=8.98)로 남아가 높게 나타났지만 그 차이는 유의미하지 않았다. 동작지수 분석에서는 남아가 99.37(SD=12.73), 여아가 99.94(SD=10.65)로 여아가 높게 나타났지만 역시 유의미하지는 않았다. 2차 검사에서는 인지지수와 동작지수 모두에서 각각 여아가 96.89(SD=10.60), 103.92(SD=13.05)로 남아의 평균인 94.90(SD=11.07)과 102.36(SD=11.66)보다 높게 나타났지만 유의한 차이를 보이지 않았다.

출생순위에 따라서 1차 검사와 2차 검사시 K-BSID-II의 수행척도를 비교한 〈표 4〉를 보면 1차 검사 인지지수의 평균 값은 첫째 아이가 97.32(SD=8.95), 둘째 아이가 97.85(SD=7.68), 셋째 아이가 95.56(SD=9.22)으로 유의미한 차이

를 보이지 않았다. 동작지수에서는 첫째 아이가 100.51(SD=11.35)로 가장 높았으며 둘째 아이는 97.29(SD=10.44), 셋째 아이는 98.89(SD=11.24)로 나타났다. 2차 검사에서도 출생순위에 따른 일관적인 수행양상을 보이고 있지 않다.

〈표 5〉에서 어머니 학력이 12년이하(고졸이하)인 집단과 13년이상(대학중퇴이상)인 집단으로 구분하여 영유아의 1, 2차 검사 인지지수와 동작지수의 수행을 비교하였다. 1차 검사에서 인지지수는 어머니의 학력이 고졸이하인 집단이 평균 98.19(SD=8.76)이며 대중퇴이상의 집단의 평균은 96.51(SD=8.39)이었으나 유의한 차이를 나타내지 않았다. 또한 동작지수에서도 고졸이하인 집단의 평균이 99.54(SD=12.39), 대중퇴이상의 집단의 평균이 99.85(SD=10.47)로 유의미한 차이를 보이지 않았다. 2차 검사에서도 이와 유사하게 집단간 차이가 유의하지 않았다.

1. 1차 검사와 2차 검사 수행의 안정성

〈표 6〉은 2차 검사 월령별 1, 2차 검사 아동의 K-BSID-II 척도별 수행점수를 나타낸 표이다. 1차 검사의 인지지수는 97.63(SD=8.65)이며 동작지수는 99.64(SD=11.76)로 나타났다. 2차 검사의 경우도 인지지수는 95.85(SD=10.87), 동작지수 103.11(SD=12.33)로 인지지수보다 동작지수에서 높은 양상을 보이고 있다. 이는 조복희, 박혜원(2004)의 연구에서 밝힌 바와 같이 한국 영유아의 동작성 척도 수행이 미국의 영유아에 비해 높은 것과 일치하는 결과이다. 2차 검사에서 1차검사보다 인지지수의 경우 수행이 낮아지는 경향을 보이는 반면 동작지수는 높아져서 유의한 차이가 있었다.

1차 수행과 2차 수행간의 상관을 보면 〈표 7〉에 제시된 바와 같이 원점수간에 그리고 지수점수간에 모두 유의한 상관을 보였다. 특히 원점수간의 상관은 매우 높아 인지척도에서

〈표 3〉 성별에 따른 K-BSID-II 척도별 수행비교(평균(SD))

시기	척도	성별		t
		남(n=102)	여(n=93)	
1차	인지지수	98.69(8.24)	96.47(8.98)	1.795
	동작지수	99.37(12.73)	99.94(10.65)	-.333
2차	인지지수	94.90(11.07)	96.89(10.60)	-1.280
	동작지수	102.36(11.66)	103.92(13.05)	-.883

〈표 5〉 어머니의 학력에 따른 수행 비교(평균(SD))

시기	척도	모 학 력		t
		고졸이하(n=130)	대중퇴이상(n=65)	
1차	인지지수	98.19(8.76)	96.51(8.39)	1.649
	동작지수	99.54(12.39)	99.85(10.47)	.030
2차	인지지수	95.75(10.75)	96.06(11.17)	.036
	동작지수	103.12(12.55)	103.09(11.97)	.000

〈표 4〉 출생순위에 따른 K-BSID-II 수행 비교(평균(SD))

시 기	척 도	출 생 순 위			F
		첫째(n=57)	둘째(n=55)	셋째(n=9)	
1차	인지지수	97.32(8.95)	97.85(7.68)	95.56(9.22)	.299
	동작지수	100.51(11.35)	97.29(10.44)	98.89(11.24)	1.211
2차	인지지수	95.14(11.72)	97.60(9.53)	91.33(14.05)	1.577
	동작지수	102.96(11.88)	100.85(12.65)	102.67(15.26)	.414

〈표 6〉 2차 검사월령별 1, 2차 검사 아동의 K-BSID-II 척도별 수행점수

시 기	척 도	6~12개월(n=58)	13~24개월(n=81)	25~42개월(n=56)	전체(N=195)
		M(SD)	M(SD)	M(SD)	M(SD)
1차	인 지 지 수	97.10(7.62)	98.84(8.06)	96.43(10.26)	97.63(8.65)
	동 작 지 수	95.72(9.74)	98.99(14.37)	103.02(14.37)	99.64(11.76)
2차	인 지 지 수	98.02(9.65)	93.01(11.15)	97.91(10.90)	95.85(10.87)
	동 작 지 수	103.55(10.35)	100.17(12.41)	106.89(13.19)	103.11(12.33)

〈표 7〉 1차 검사와 2차 검사의 척도별 상관

	1차 검사 인지원점수	1차 검사 인지지수	1차 검사 동작원점수	1차 검사 동작지수	2차 검사 인지원점수	2차 검사 인지지수	2차 검사 동작원점수
1차 검사 인지지수	.04						
1차 검사 동작원점수	.99**	-.01					
1차 검사 동작지수	.30**	.44**	.35**				
2차 검사 인지원점수	.91**	-.04	.90**	.25**			
2차 검사 인지지수	.00	.18*	-.06	.06	.09		
2차 검사 동작원점수	.90**	-.06	.89**	.24**	.98**	.06	
2차 검사 동작지수	.15*	.08	.12	.17*	.15*	.47**	.28**

* $p < .05$, ** $p < .01$

$r = .91$, 동작척도에서는 $r = .89$ 의 상관을 보였다.

2. 한국 Bayley 영유아 발달검사 수행의 안정성에 미치는 요인분석

Miller(1987)은 일반적으로 지적 수행의 안정성에 영향을 미치는 시간적 요인으로 검사 연령, 검사간 간격을 언급하였다. 즉 안정성을 측정(비교)하는 기간이 가까울수록, 처음 검사시의 연령이 높을수록 상관이 높다. 본 연구에서는 따라서 첫 검사시의 연령과 1차, 2차 검사간의 간격에 따라 수행의 안정성이 어떻게 달라지는지 그리고 성별, 아동의 출생순위, 그리고 어머니 학력과 같은 인구통계학적인 변인에 따라 안정성이 달라지는지 살펴보았다.

1) 검사 연령에 따른 수행의 안정성

검사연령의 영향을 살펴보기 위해 1차 검사 월령의 평균을 중심으로 두 집단으로 구분하여 두 검사결과 간의 상관을 분석한 결과를 〈표 8, 9〉에 제시하였다. 선행연구들(Jencks, 1972)과 마찬가지로 전반적으로 1차 검사시 월령이 6개월이 하인 집단($n=99$)의 수행안정성보다 1차 검사시 월령이 6개월이상인 집단($n=96$)의 안정성이 높은 경향을 보였다. 인지

원점수의 경우 검사간 상관 즉 안정성이 월령이 낮은 집단은 $r = .41$, 높은 집단은 $.90$ 로 증가했으며 동작원점수의 경우도 $r = .43$ 에서 $r = .88$ 로 높아짐을 알 수 있다. 인지 지수의 경우는 월령이 낮은 집단은 유의한 상관이 없었으나 월령이 높은 집단에서는 $r = .22$ ($p < .05$)로 낮지만 유의한 차이를 보이고 있다. 반면 동작지수의 경우 월령이 낮은 집단은 $r = .20$ ($p < .05$)으로 유의한 상관을 보인데 반해 월령이 높은 집단에서는 $r = .15$ ($p < .13$)로 유의한 차이를 보이지 못했다. 〈표 7〉에서 살펴본 바와 같이 지수의 경우는 전반적으로 안정성이 유의하지만 낮기 때문에 동작성 지수의 경우 연령집단을 구분하여 집단별 아동수가 줄어 들게 됨에 따라 지수의 상관이 유의하지 않게 나타난 것으로 해석된다.

2) 검사 간격에 따른 수행의 안정성

1차 검사와 2차 검사 간격은 최소 5개월에서 최대 16개월이었다. 1차 검사와 2차 검사간 간격의 평균 기간은 8.8개월이었으며 검사간격에 따라 8개월 이하 집단($n=87$)과 9개월 이상 집단($n=108$)으로 나누어 보면 검사간격이 8개월 이하 집단이 87명(44.6%), 검사간격이 9개월 이상 집단은 108명(55.4%)으로 각 집단에서 1차, 2차 검사간 상관은 〈표 10,

〈표 8〉 1차 검사월령이 6개월이하 집단의 수행 안정성

	1차 검사 인지원점수	1차 검사 인지지수	1차 검사 동작원점수	1차 검사 동작지수	2차 검사 인지원점수	2차 검사 인지지수	2차 검사 동작원점수
1차 검사 인지지수	.18						
1차 검사 동작원점수	.98**	.11					
1차 검사 동작지수	.17	.52**	.23*				
2차 검사 인지원점수	.41**	.09	.38**	.03			
2차 검사 인지지수	.03	.08	.04	.16	.32**		
2차 검사 동작원점수	.45**	.06	.43**	.02	.86**	.13	
2차 검사 동작지수	-.04	.17	-.02	.20*	-.09	.32**	.22*

* $p < .05$, ** $p < .01$

〈표 9〉 1차 검사 월령이 6개월이상인 아동의 수행의 안정성

	1차 검사 인지원점수	1차 검사 인지지수	1차 검사 동작원점수	1차 검사 동작지수	2차 검사 인지원점수	2차 검사 인지지수	2차 검사 동작원점수
1차 검사 인지지수	.13						
1차 검사 동작원점수	.97**	.09					
1차 검사 동작지수	.20	.44**	.36**				
2차 검사 인지원점수	.90**	-.00	.86**	.11			
2차 검사 인지지수	.57**	.22*	.50**	.10	.66**		
2차 검사 동작원점수	.90**	-.06	.88**	.13	.96**	.58**	
2차 검사 동작지수	.45**	.01	.45**	.16	.44**	.61**	.59**

* $p < .05$, ** $p < .01$

11)과 같다. 검사간격이 8개월 이하인 집단에서 유의하지 않았던 인지지수간 상관인 검사간격이 큰 집단의 경우 유의($r=.03 \rightarrow r=.26$)하게 나타나는 등 수행 안정성이 검사간격이 큰 집단에서 오히려 더 높은 유형을 나타냄으로써 선행연구들과는 상반된 결과를 보였다.

3) 인구 통계학적 변인(성, 출생순위 및 어머니 학력)에 따른 수행 안정성분석

〈표 12〉와 〈표 13〉에 성별, 출생순위(첫째 vs. 둘째이상) 집단을 나누어 안정성을 분석한 결과 집단에 따른 일관성있는 수행 안정성의 차이는 없었다.

〈표 14〉에서 어머니의 학력에 따라 K-BSID-II의 수행 안정성을 비교해 보면 학력이 높은 집단의 상관이 더 높은 경향을 나타내 안정성이 큰 것으로 나타났다.

IV. 논의 및 결론

본 연구에서는 심리측정적인 측면에서 가장 정교하고 신뢰로운 영아발달검사로 알려진 1개월~42개월 영유아용 BSID-II를 한국형으로 개발한 한국 Bayley 영유아 발달검사(K-BSID-II) 수행의 안정성을 알아보고 안정성에 영향을

〈표 10〉 검사 간격이 8개월 이하 집단의 K-BSID II 척도별 상관관계

	1차 검사 인지원점수	1차 검사 인지지수	1차 검사 동작원점수	1차 검사 동작지수	2차 검사 인지원점수	2차 검사 인지지수	2차 검사 동작원점수
1차 검사 인지지수	.14						
1차 검사 동작원점수	.99**	.11					
1차 검사 동작지수	.34**	.13	.36**				
2차 검사 인지원점수	.91**	.08	.90**	.35**			
2차 검사 인지지수	-.12	.03	-.18	.05	.02		
2차 검사 동작원점수	.90**	.08	.90**	.40**	.98**	-.05	
2차 검사 동작지수	.17	.19	.15	.23*	.20	.35**	.30**

*p<.05, **p<.01

〈표 11〉 검사 간격이 9개월 이상 집단의 K-BSID II 척도별 상관관계

	1차 검사 인지원점수	1차 검사 인지지수	1차 검사 동작원점수	1차 검사 동작지수	2차 검사 인지원점수	2차 검사 인지지수	2차 검사 동작원점수
1차 검사 인지지수	.01						
1차 검사 동작원점수	.98**	-.04					
1차 검사 동작지수	.26**	.37**	.29**				
2차 검사 인지원점수	.92**	-.05	.91**	.21*			
2차 검사 인지지수	.13	.26**	.05	.24*	.20*		
2차 검사 동작원점수	.92**	-.09	.91**	.23*	.97**	.20*	
2차 검사 동작지수	.17	.00	.14	.25**	.22*	.55**	.38**

*p<.05, **p<.01

〈표 12〉 성별 K-BSID-II 척도의 상관비교

성 별		1차 검사 인지지수	2차 검사 인지지수	1차 검사 동작지수
남 아	2차 검사 인지지수	.17		
	1차 검사 동작지수	.31**	.17	
	2차 검사 동작지수	.08	.44**	.27**
여 아	2차 검사 인지지수	.21*		
	1차 검사 동작지수	.21*	.10	
	2차 검사 동작지수	.09	.50**	.17

*p<.05, **p<.01, ***p<.001

〈표 13〉 출생순위별 K-BSID-II 척도 수행 상관비교

출생 순위		1차 검사 인지지수	2차 검사 인지지수	1차 검사 동작지수
첫째 (n=57)	2차 검사 인지지수	.25		
	1차 검사 동작지수	.48**	.05	
	2차 검사 동작지수	.17	.57**	.23
둘째 이상 (n=64)	2차 검사 인지지수	.17		
	1차 검사 동작지수	.42**	-.07	
	2차 검사 동작지수	-.15	.45**	-.11

**p<.01

〈표 14〉 어머니 학력에 따른 수행 안정성

모학력		1차 검사 인지지수	2차 검사 인지지수	1차 검사 동작지수
고졸 이하 (n=130)	2차 검사 인지지수	.10		
	1차 검사 동작지수	.44**	.02	
	2차 검사 동작지수	-.02	.47**	.11
대중퇴 이상 (n=65)	2차 검사 인지지수	.34**		
	1차 검사 동작지수	.47**	.16	
	2차 검사 동작지수	.30*	.49**	.34**

*p<.05, **p<.01

미치는 요인을 탐색하였다. 2002년에서 2003년 사이에 실시된 한국 Bayley 영유아 발달검사 표준화와 타당도 연구에 참여한 울산지역에 거주하는 493명의 아동가운데 2차 검사를 신청한 195명을 대상으로 하여 중단연구를 실시한 것이다.

일찍이 Bayley 영유아검사의 초판인 BSID의 안정성을 연구한 개발자 Bayley(Bayley, 1965; 1970)는 비록 BSID의 결과는 아동의 발달 지연과 현재 수준을 설명하는데 유용하지만 성장후의 지능에 대한 예언력 즉 안정성은 낮은 것으로 보고하였다. 그러나 1960년 이후 활발한 영아기 연구를 통해 영아기 초기부터 안정적으로 인지능력을 측정할 수 있게 되었다. 예로 인지 발달 연구자들(Banks, & Ginsburg, 1985; Brody, 1993; Fantz & Fagan, 1975)은 영아가 복잡한 시각 자극을 선호하는 것은 신경학적인 성숙도를 반영하는 것이며 이러한 경향은 후의 수행을 예언할 수 있음을 보고하였다. 또한 사고, 추론, 단기 기억과 같은 정보처리 및 문제해결 능력이 어린 영아에서도 찾아볼 수 있음이 밝혀졌다. 그 외에도 Gelman & Gallistel(1978)은 수세기의 발달은 세부적인 하위단계로 나누어지는데 초기 단계는 영아기에 획득된다는 것을 보고하였다. 따라서 성장후 지적 수행과 밀접한 관계가 있는 측면을 측정하는 이러한 문항이 개발되어 BSID-II에 첨가되었다(Black, & Matula, 1999). 아직 미국의 경우도 BSID-II에 대한 중단적인 자료는 찾아보기 힘들지만 본 연구에서는 한국에서 2002년의 표준화연구의 표본을 중심으로 5개월에서 16개월의 간격으로 2회 측정된 본 검사의 영아기동안의 안정성을 분석할 수 있었다.

안정성분석에 앞서 인구통계학적 변인에 따라 수행을 분석했을 때 모의 학력, 아동의 출생순위 및 성별에 따라 일관성있는 수행의 차이는 나타나지 않았다. 5-16개월을 간격으로 실시된 2회 검사의 수행안정성은 〈표 7〉에 제시된 바와 같이 인지원점수간 상관은 인지척도 $r=.91(p<.01)$, 동작척도 $r=.89(p<.01)$ 에 이르는 매우 높은 정적 상관을 보였으나 지수간 상관은 유의하지만 낮은 상관(인지척도 $r=.18, p<.05$, 동작척도 $r=.17, p<.05$)을 보이고 있다. Bayley 검사와 같은 편차척도에서는 지수는 월령에 따라 증가하는 원점수를 월

〈표 15〉 검사간격과 1, 2차 검사 월령과의 상관관계

	1, 2차 검사간 간격	1차 검사월령
1차 월령	.184*	
2차 월령	.338**	.980**

*p<.05, **p<.01

령별 규준집단내에서 평균과 편차를 고려하여 환산된 표준 점수로 일반적으로 원점수에 비해 안정성이 낮다(McCall, 1983). 그럼에도 불구하고 본 연구에서 월령이 6개월 이하인 아이를 대상으로 5개월에서 16개월을 간격으로 검사를 실시했을 때 인지지수를 제외한 모든 수행에서 상관의 도출됨으로써 K-BSID-II의 안정성은 확보된 것으로 해석할 수 있다. 그러나 본 연구결과 원점수의 안정성은 매우 높아 영아기 수행의 안정성을 살펴보기 위해서는 원점수를 활용하는 것이 매우 유용함을 보여주고 있다.

안정성에 영향을 미치는 요인을 탐색하기 위해 검사시 월령, 검사 간격 그리고 인구학적 변인(성별, 아동의 출생순위 및 모학력)에 따라 집단을 분류하여 안정성을 살펴보았다. 일반적으로 안정성은 연령이 증가함에 따라 높아져 처음 검사 시기가 늦을 수록 그리고 두 검사간 간격이 길수록 높아진다(McCall, 1983; Miller, 1987). 〈표 8, 9〉를 비교해 볼 때 본 연구에서도 선행연구와 마찬가지로 처음 검사시 월령이 높았던 집단이 낮았던 집단에 비해 안정성이 높은 것을 보여주었다.

검사간격(8개월이하 vs. 8개월 이상)에 따른 안정성 분석 결과를 〈표 10, 11〉에서 비교해 볼 때 1차 검사와 2차 검사의 간격이 큰 집단에서 인지지수와 동작지수의 안정성이 더 큰 경향이 나타났다. 이는 일반적으로 기존의 선행연구에서 밝혀진 1, 2차 검사 간격이 짧을수록 안정성이 높다는 결과와 상반된다(McCall, 1983). 그러나 본 연구에서는 추후 분석(표 15 참조)에서 1, 2차 검사 간격이 긴 집단은 1차 검사와 2차 검사에서 모두 월령 또한 높은 경향이 있었으며 따라서 이 결과는 월령이 높을수록 지능의 안정성이 높아지는 일반적인 현상인 것으로 해석된다.

일반적으로 부모의 학력은 영아의 BSID수행에 영향을 미칠 수 있음이 보고되었다(Bayley, 1965). 본 연구에서 학력에 따른 수행의 차이는 유의하지 않았으나 모의 학력이 높을 수록 안정성은 높은 경향을 보였다. 이는 모의 학력이 높을 수록 영유아의 심리적, 물리적 환경도 안정적일 수 있을 것으로 사료되며 이것이 지적 수행의 안정성과 관련이 있을 것으로 해석된다. 그러나 출생순위와 성별에 따른 수행의 차이가 유의미 하지 않았던 것과 마찬가지로 수행의 안정성의 차이는 유의하게 나타나지 않았다. 성인이나 보다 나이든 아동을 대상으로 하는 선행연구(박혜원, 광금주, 박광배, 1995;

King & Seegmiller, 1973; Zajonc & Markus, 1975)에서 일반적으로 출생순위에 따라 지적 수행에 차이가 나타나고 있음을 보고하고 있는 점을 감안하여 지능의 안정성에서도 출생순위에 따른 차이가 있는지 살펴보았으나 본 연구 결과는 이러한 가정을 지지하지 않았다. 따라서 Bayley 검사의 수행안정성에 미치는 영향은 영아의 출생순위나 성차보다는 안정적인 환경임을 시사하였다.

본 중단연구에서는 중단 연구대상표집의 어려움 때문에 월령과 인구통계학적 변인에 고려한 표집을 하지 못하고 재검사에 응한 모든 아이들이 검사가능대상이었고 이로 인해 각 월령군의 인구통계학적 변인상 동질성 확보가 어려웠다. 후속연구에서는 검사 월령과 검사 간격에 따른 집단별 인구통계학적 변인을 통제하여 표집함으로써 안정성에 미치는 영향에 대한 보다 명확한 규명이 이루어 져야 한다.

■ 참고문헌

- 강순구, 조윤경(1994). 포테이지 아동발달 지침서. 서울: 특수교육.
- 김숙현, 이경옥(2001). 영아발달체크리스트 표준화연구. 2001. 한국유아교육학회 연차학술대회.
- 김승국, 김옥기(1985). 사회성숙도 검사 요강. 서울: 중앙적성출판사.
- 대한소아과학회(2002). 한국형 영유아발달검사. 서울: 대한소아과학회.
- 민동옥, 박혜원(2003). 조산아에 대한 베일리 영유아발달검사 실시 및 채점 방식에 관한 연구. 한국 아동학회 추계학술대회.
- 박혜원, 곽금주, 박광배(1995). 웨슬러 유아지능검사의 수행 분석: 형제관계를 중심으로. 아동학회지, 22(1), 57-69.
- 박혜원, 조복희(미출간원고). 한국 Bayley 영유아 발달검사: 해석지침서.
- 박혜원, 조복희, 최호정(2003). 한국 Bayley 영유아 발달검사 (K-BSID-II) 표준화연구: 예비연구. 한국심리학회지: 발달, 16(4), 12-134.
- 배윤희(2001). 베일리 영유아발달 검사-II의 한국 표준화를 위한 예비연구. 울산대학교 석사학위논문.
- 이일섭(1992). Bayley 발달검사를 이용한 고위험 저출생체중아의 발달평가. 연세대학교 석사학위논문.
- 정문자(1983). 유아발달을 위한 가정방문 프로그램 연구. 대한가정학회지, 21(4), 129-137.
- 정문자, 이은혜, 박경자(1993). 베일리 유아발달 척도의 한국 표준화를 위한 예비연구. 아동학회지, 14(1), 5-21.
- 정영미(1992). Bayley 발달검사를 이용한 유아의 발달평가. 연세대학교대학원 석사학위논문.
- 제경숙(1984). 베일리의 유아발달검사의 척도화를 위한 연구. 연세대학교 석사학위논문.
- 조복희, 박혜원(2004). 한국 Bayley 영유아발달검사(K-BSID-II) 표준화연구(1): 지역, 성별 및 모의 교육 수준에 따른 수행분석. 한국심리학회지: 발달, 17(1), 191-206.
- Aggarwal, P., Chaudhari, S., Bhave, S., Pandit, A., Barve, S. (1998). Clinical predictors of outcome in hypoxic ischaemic encephalopathy in term infants. *Annals of Tropical Paediatrics*, 18, 117-121.
- Baird, M., & Folsom, J. (1993). Concurrent validity of the Kent Scoring Adaptation of the Bayley Scales of Infant Development. *Diagnostic*, 18(2), 135-143.
- Banks, M., & Ginsburg, A. (1985). Infant visual preferences: A review and new theoretical treatment. In H. Reese (Ed.), *Advances in child development and behavior* (19, 207-246). Orlando, FL: Academic Press.
- Bayley, N. (1932). A study of the crying of infants during mental and physical tests. *Journal of Genetic Psychology*, 40, 306-329.
- Bayley, N. (1933a). Mental growth during the first three years. *Genetic Psychology Monographs*, 14, 1-92.
- Bayley, N. (1933b). The California First-Year Mental Scale. Berkeley: University of California Press.
- Bayley, N. (1935). The development of motor abilities during the first three years. *Monographs of the Society for Research in Child Development*.
- Bayley, N. (1936). The California Infant Scale of Motor Development. Berkeley: University of California Press.
- Bayley, N. (1965). Comparisons of mental and motor test scores for age, 1 - 15 months by sex, birth order, race, geographical location, and education of parents. *Child Development*, 36, 379-411.
- Bayley, N. (1969). Bayley scales of infant development. New York: Psychological Corporation.
- Bayley, N. (1970). Development of mental abilities. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed., Vol. 1, pp. 1163-1209). New York: Wiley.
- Bayley, N. (1993). Bayley scales of infant development-

- II. San Antonio: Psychological Corporation.
- Black, M. M., & Matula, K. (1999). *Essentials of Bayley Scales of Infant Development II Assessment*. New York: Wiley & Sons.
- Brody, N. (1993). *Intelligence* (2nd ed.). New York: Academic Press.
- Brooks-Gunn, J., & Weinraub, M. (1983). Origins of infant intelligence testing. Origins of intelligence: Infancy and early childhood (2nd ed, 25-66). In M. Lewis (Ed.) New York: Plenum Press.
- Campbell, S. K., Siegel, E., Parr, C. A., & Ramey, C. T. (1986). Evidence for the need to renorm the Bayley Scale of Infant Development based on the performance of population-based sample of 12-month-old infants. *Topics in Early Childhood Special Education*, 6, 83-96.
- Fantz, R., & Fagan, J. (1975). Visual attention to size and number of pattern details by term and preterm infants during the first six months. *Child Development*, 46, 3-18.
- Gelman, R., & Gallistel, D. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press
- Goldstein, D. J., Fogle, E. E., Wieber, J. L., & O'Shea, T. M. (1995). Comparison of the Bayley Scales of Infant Development-Second Edition and the Bayley Scales of Infant Development with premature infants. *Journal of Psychoeducational Assessment*, 13, 391-396.
- Jencks, C. (1972). *Inequality*. New York: Basic Book.
- Johnson, L. J., Cook, M. J., & Kullman, A. J. (1992). An examination of the concurrent validity of the Battelle Developmental Inventory as compared with the Vineland Adaptive Scales and the Bayley Scales of Infant Development. *Journal of Early Intervention*, 16(4), 353-359.
- King, W., & Seegmiller, B. (1973). Performance of 14- to 22- month-old black first-born infants on two tests of cognitive development: The Bayley Scales. *Developmental Psychology*, 8, 317-326.
- McCall, R. (1983). A conceptual approach to early mental development. In M. Lewis (Ed.), *Origins of intelligence* (2nd Ed.). New York: Plenum.
- McChesne, S. (1995). Concurrent validity of the Fagan test of infant intelligence with the Bayley Scales of Infant Development for infants from low socioeconomic and middle to upper socioeconomic status. University of South Carolina.
- Miller, S. (1987). *Developmental research methods*. New York: Prentice Hall.
- Nellis, L., & Gridley, B. E. (1994). Review of the Bayley Scale of Infant Development-Second Edition. *Journal of School Psychology*, 32, 201-209.
- Rubin, B. G. (1996). A Test-retest study of the mental scale of the Bayley Scales of infant development: Second-age children with significant developmental disabilities. Temple University.
- Sabourin, J. A. (1996). The stability of IQ in a clinic-referred child sample as measured by the Bayley Scales of Infant development, the Griffiths mental development scales and the Kaufman Assessment Battery for Children. Laurentian University of Sudbury (Canada).
- Snow, C., & McGaha, C. (3rd. Ed.) (2003). *Infant development*. New York: Prentice Hall.
- Zajonc, R., & Markus, G. (1975). Birth order and intellectual development. *Psychological Review*, 82, 74-88.

(2005년 10월 31일 접수, 2005년 12월 7일 채택)