中国版K式発達検査の標準化に関する研究 第3報 —1歳～3歳児における検査の信頼性・妥当性について—

高 健¹

Standardization of Chinese Version Kyoto Developmental Test III:
A study on the reliability and stability
of the assessment in children aged 1～3 years

Jian Gao¹

To study the reliability and validity of Chinese Version Kyoto Development Test, we have analyzed the results of the test administered to 1096 Chinese children aged 1-3 years, including 558 boys and 538 girls, and the following conclusions were established.

1. Inter-observer Reliability: A random sample of 50 cases was selected. According to the results of the test of the two observers, it showed that the correlation coefficients between the two observers were 0.95-1.00, which reached significant level ($p < 0.001$).

2. Test-retest Reliability: The test-retest correlation coefficients of eight children with one week interval were 0.99-1.00, and the test-retest correlation coefficient between developmental quotients (DQ) ranged from 0.83-0.98. In addition, both of the results showed significant level ($p < 0.001$).

3. Split-half Reliability: The correlation coefficient and the reliability coefficient between the scores based on odd-even method were similar to the statistical mark of Chinese Version Kyoto Development Test.

4. Criterion-related Validity: 267 children were tested on Chinese Version Kyoto Development Test and Chinese Developmental Screening Test with one week interval. Except postural-motor domain, DQ for other domains and DQ for Chinese Version Kyoto Development Test and mental index of Chinese Developmental Screening Test all showed significant level ($p < 0.001$). Besides the above, Chinese Version Kyoto Development Test and Chinese Version S-M Social Living Ability Test were also carried on 302 subjects at the same time. The scores of each domain obtained from the two tests showed significant level ($p < 0.05$, $p < 0.001$). The results suggested that Chinese Version Kyoto Development Test seemed to be an available and effective developmental test.

Key words : Chinese Version Kyoto Developmental Test, inter-observer reliability, test-retest reliability, split-half reliability, criterion-related validity

中国版K式発達検査, 検者間信頼性, 再検査信頼性, 折半信頼性, 基準関連妥当性

¹神戸学院大学

¹Kobe Gakuin University

I 緒 言

心理検査の信頼性とは、理論上は、測定値の中にどれほど誤差が混入しているかを表すものである(下中, 1981)。実際の手続きでは、検査得点の一貫性または再現性を意味する(生澤ほか, 1998)。日本の新版K式発達検査の信頼性は折半信頼性によって検討され、統計的に信頼性の高い検査として異議をさしはさむ余地がない。なお、この検査を文化・社会的環境の異なる中国の幼児を対象として実施する場合には、信頼性を考慮する必要があると思われる。

筆者は第2報(高, 2004)では中国版「得点・発達年齢換算表」の作成と1歳～3歳児における検査結果の分析についての検討を行った。本報では第2報に続いて、中国版K式発達検査による中国被検児の検査結果に基づいて検査の信頼性と妥当性の検討を行ったので報告する。

II 方 法

1. 対象と検査方法

被検児の選定として第2報(高, 2004)の中国版K式発達検査第3葉標準化集団(1歳0ヵ月超～3歳0ヵ月: 1096名, 男児: 558名, 女児: 538名。以下: 標準化集団と称す)が対象となっている。被検児の年齢区分及び検査者、検査方法は第1報(高ほか, 2003)と同一である。なお、『2・3数復唱』の判定基準は「1/3」を「3/3」に変えた。『絵の名称II』の「電話」図版は中国版の「電話」図版を用いた。また、検査結果の算出は中国版「得点・発達年齢換算表」を用いた。

2. 分析方法

1) 検査の信頼性

(1) 検者間信頼性

中国版K式発達検査第3葉標準化実施中の1999年1月～2月に、標準化集団から50名の被検児をランダムに選び、検者間信頼性の検討を行った。実施に先立って、あらかじめ年齢別のサンプル数を決めておき(1: 0超～1: 3 10名, 1: 3超

～1: 6 10名, 1: 6超～2: 0 10名, 2: 0超～2: 6 10名, 2: 6超～3: 0 10名, 計50名の予定)、偶数番にきた者に対して、検者の1人が検査を行っている間に、もう1人の検者も同じ被検児を観察して検査項目の可否を同時に記録し、また、個別に得点算出を行うという方法を採用した。決めておいたサンプル数(50名)に達するまでこの手続きを繰り返した。統計分析には、検者間の各領域と全領域得点の一致率および項目の可否判断の一致率を求め、各領域(姿勢・運動以下: P-M, 認知・適応 以下: C-A, 言語・社会 以下: L-S)と全領域得点のPearson相関係数も算出した。

(2) 再検査信頼性

中国版K式発達検査第3葉標準化実施中の1999年12月～2000年2月に、標準化集団から80名の被検児をランダムに選び、再検査信頼性の検討を行った。

再検査信頼性について、被検者の年齢(発達段階)も大きく関係すると思われ(鎌原ほか, 1998)、低い年齢段階の子どもの場合、相対的に集中力が劣るので、検査結果は一過性であり、再現性の低い可能性がある。したがって、検査結果を歪ませる可能性のある条件を統制するために、本研究では、低い年齢段階の被検児数(1歳児)を多く取ることとした。1ヵ所の対象園において、あらかじめ年齢別のサンプル数を決めておき(1歳児45名, 2歳児25名, 3歳児15名, 計85名の予定)、火、水、木曜日の午前中に(月、金曜日に欠席者が多く、午後に早引き者がいるため)検査を実施した者に対して、同一の検者が再検査を行った。初回の検査と再検査との間隔は1週間である。検者が被検児における初回の検査時の行動を記憶している可能性を避けるために、再検査する際に、初回の検査記録を参照することはしなかった。なお、再検査の時、欠席した被検児が5名(1歳児: 2名, 3歳児: 3名)いたので、実施できた者は80名であった。統計分析には、検査一再検査による得点一致率を求め、各領域と全領域得点及び発達指数(以下: DQ)の安定性分析には対

応のあるt検定を用いた。

(3) 折半信頼性

標準化集団の被検児すべてを解析に用いた。日本の新版K式発達検査標準化資料（生澤ほか，1998）に基づいて検査項目を奇数番号の項目と偶数番号の項目に分けて，二つの項目群を作り，2群の各領域と全領域得点に信頼性分析を行い，各年齢区分及び全体（1：0超～3：0）でSpearman-Brownの補正を施し，Pearson相関係数とSpearman-Brown信頼性係数を算出した。

2) 基準関連妥当性

(1) 発達スクリーニング検査との相関

中国版K式発達検査第3葉標準化実施中の2000年12月～2001年3月に，標準化集団の対象園（23の園）から，園児数300名以上の1ヵ所〔1)の(1)と別の園〕を選び，全園児に対して，中国で標準化され広く用いられている発達スクリーニング検査CDST（Chinese Developmental Screening Test，略称：CDST）（鄭ほか，1997）を実施した。中国版K式発達検査とCDSTとの検査間隔は1週間であった。CDSTの検者は中国版K式発達検査と同一であった。なお，欠席と冬休みのため，検査不能だった者を除き，二つの検査がともに実施できた者は267名であった。統計分析には，中国版K式発達検査の各領域と全領域DQ及びCDSTのDQと知能指標（以下：MI）の間でPearson相関係数を算出した。

(2) 社会生活能力検査との相関

中国版K式発達検査第3葉標準化実施中の1999年1月～2月，1999年12月～2000年2月，2001年11月～12月に，中国版K式発達検査（1回目）が実施された被検児に対して，中国で標準化され広く用いられている中国版S-M社会生活能力検査（張ほか，1995）（以下：中国版S-M・SMS）を実施した。集団保育児において，中国版K式発達検査を実施した当日，被検児の名前を記入した中国版S-M・SMSの検査用紙を，被検児のことをよく知っている保育担当者に手渡し，1週間以内に記入するように依頼した（筆者が評価・記入方法をあらかじめ説明した）。家庭保育児におい

て，中国版K式発達検査を実施した後，検者は中国版S-M・SMSの項目について，項目毎に具体的に詳しく子どもの様子を保護者（主に母親）に聞きながら，評価した。分析対象となったのは，1歳から3歳までの計302名の被検児であった。統計分析には，中国版K式発達検査の各領域と全領域得点及び中国版S-M・SMSの全検査粗点の間でPearson相関係数を算出した。

3) 検査時間

検査実施の時刻には，午前（8時30分～11時30分）と午後（3時～5時）があり，検査する児の人数を各年齢区分に午前・午後ともに概ね半数ずつになるように調整した。また，検査所要時間を調べるために，被検児が最初に与えられた検査用具に手を出し始めた時から，検査の完了が明確になるまでの経過時間をストップウォッチで測定した。さらに，各年齢区分の被検児を午前実施群と午後実施群に分け，独立したサンプルのt検定によって2群間における検査の所用時間，各領域と全領域得点の平均値の差を比較した。

以上の統計処理はSPSS 10.0J for Windowsによって行った。

III 結 果

1. 検者間信頼性の検討

標準化集団からランダムに50名の被検児を抽出し，2名の検者によりそれぞれ独自に採点した。このようにして得られた各領域の得点一致率（表1-1）は86%～100%である。C-A領域で一番低く，L-S領域で一番高い。各領域内の検査項目をさらに詳細に検討してみると（表1-2），P-M領域の「両足跳び」，C-A領域の「折り紙」と「描画」においては，合否判断の一致率がそれぞれ順に94%，94%，92%であり，不一致となった項目はC-A領域で比較的に多かった。なお，検者間における各領域及び全領域得点の相関係数（表1-1）は，0.95～1.00の範囲で，いずれも有意な相関が得られた（ $p < 0.001$ ）。

2. 再検査信頼性の検討

1歳から3歳の年齢範囲にある80人の被検児を、同一の検者が1週間の間隔をおいて初回と同じ方法で再検査した。このようにして得られた得点一致率はP-M領域で98%, C-A領域で95%, L-S領域で96%, 全領域で86%であった。各領域及び全領域得点の1回目と2回目の相関係数(表2)は0.99~1.00の範囲で、いずれも有意であった($p < 0.001$)。また、各領域及び全領域DQの1回目と2回目の相関係数(表2)は、0.83~0.98の範囲、P-M DQで一番低く、C-A DQで一番高く、いずれも有意であった($p < 0.001$)。さらに、検査―再検査における得点とDQの平均値を比較した結果(表2)を見ると、1週間間隔のP-M得点、C-A得点、L-S得点、全領域得点の平均値の差は、それぞれ0.31, 0.81, 0.69, 1.81, P-M DQ, C-A DQ, L-S DQ, 全領域DQの平均値の差は、それぞれ0.60, 0.58, 0.58, 0.48であった。L-S DQを除いて、全体に1回目よりも2回目

のほうが上昇する傾向が見られたが、t検定の結果により、全領域の得点とDQにだけ有意差(それぞれ $p < 0.01$, $p < 0.05$)が見られた。各領域内の検査項目をさらに詳細に検討してみると、1回目で合格していなかったが、2回目で合格した項目は、P-M領域の「飛び降り」、C-A領域の「はめ板」と「入れ子」、L-S領域の「数の復唱」であった。

3. 折半信頼性の検討

原版と同様に、検査の内部一貫性をみる折半信頼性を検討した。表3に各年齢区分と全体(1:0超~3:0)における各領域と全領域の奇隅折半得点の相関係数および信頼性係数を示した。参照のため、その中には原版データ(生澤ほか, 1998)も合わせて示してある。

各領域と全領域の相関係数および信頼性係数の平均は全ての年齢区分を通して高いものであり、全般に満足すべき信頼性係数となっている。1歳6カ月から1歳9カ月の年齢区分において、相関係数も信頼性係数も他の年齢区分より幾分低い値であったが、原版と類似した傾向を示している。年齢区分を広く取った場合は(1:0超~3:0)年齢区分を狭く取った場合より高い値であり、これらの相関係数および信頼性係数は原版と類似した値を示している。

表 1-1 検者間の得点の一致率と相関 (N=50)

	得点一致率	相関係数
P-M score	94.0%	0.95 ***
C-A score	86.0%	0.99 ***
L-S score	100.0%	1.00 ***
total score	82.0%	0.99 ***

***: $p < 0.01$ (Pearson 相関係数)

表 1-2 各領域における検査項目の検者間の信頼性 (N=50)

P-M 領域	得点一致率	C-A 領域	得点一致率	L-S 領域	得点一致率
歩く 2. 3 歩	100.0%	積木課題 (注)	100.0%	数の記憶	100.0%
片手支持登る	100.0%	小鈴と瓶	100.0%	比較	100.0%
片手支持降りる	100.0%	自動車 (玩具)	100.0%	身体名称	100.0%
手すりで登降	100.0%	課題箱	100.0%	絵の名称	100.0%
両足跳び	94.0%	はめ板	100.0%	姓名	100.0%
飛び降り	100.0%	入れ子	100.0%	性の区別	100.0%
交互に足を出す	100.0%	折り紙	94.0%	――	――
――	――	図形の弁別	100.0%		
		四角構成	100.0%		
		隠しコップ	100.0%		
		描画	92.0%		

注：積木課題：積木の塔、積木の模倣。

表2 検査—再検査による得点及びDQの安定性 (N=80)

領域	1回目		2回目		相関係数	t 値
	平均	標準偏差	平均	標準偏差		
P-M score	75.75	9.55	76.06	9.47	0.99 ***	1.92
C-A score	178.31	41.53	179.13	41.89	1.00 ***	2.02
L-S score	65.19	23.27	65.88	24.33	0.99 ***	1.74
total score	319.25	72.13	321.06	73.62	1.00 ***	3.21 **
P-M DQ	98.87	9.40	99.47	8.63	0.83 ***	1.03
C-A DQ	104.62	12.49	105.20	12.04	0.98 ***	1.91
L-S DQ	108.98	9.71	108.41	9.17	0.96 ***	1.89
total DQ	104.13	6.39	104.61	6.68	0.95 ***	2.14 *

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$. t 値: 対応のあるサンプルの t 検定. 相関係数: Pearson 相関係数.

表3 奇偶折半得点の相関係数と信頼性係数

年齢区分	人数	Pearson 相関係数					Spearman-Brown 信頼性係数				
		P-M 領域	C-A 領域	L-S 領域	全領域	平均	P-M 領域	C-A 領域	L-S 領域	全領域	平均
中国版 K 式発達検査											
1：0 超～1：3	130	0.51	0.77	0.48	0.86	0.66	0.72	0.89	0.66	0.93	0.80
1：3 超～1：6	132	0.52	0.78	0.55	0.85	0.68	0.69	0.89	0.72	0.93	0.81
1：6 超～1：9	191	0.39	0.68	0.36	0.74	0.54	0.58	0.86	0.61	0.87	0.73
1：9 超～2：0	232	0.48	0.70	0.61	0.77	0.64	0.69	0.88	0.80	0.92	0.82
2：0 超～2：6	247	0.34	0.73	0.62	0.79	0.62	0.53	0.88	0.83	0.91	0.79
2：6 超～3：0	164	0.55	0.74	0.77	0.84	0.73	0.74	0.89	0.88	0.93	0.86
1：0 超～3：0	1096	0.85	0.96	0.94	0.98	0.93	0.92	0.98	0.97	0.99	0.97
日本新版 K 式発達検査 (注)											
1：0 超～1：3	79	0.81	0.63	0.44	0.82	0.67	0.89	0.78	0.61	0.90	0.80
1：3 超～1：6	70	0.76	0.69	0.58	0.80	0.71	0.87	0.81	0.73	0.89	0.82
1：6 超～1：9	50	0.55	0.78	0.59	0.80	0.68	0.71	0.88	0.74	0.89	0.80
1：9 超～2：0	46	0.72	0.66	0.63	0.74	0.69	0.84	0.80	0.77	0.85	0.81
2：0 超～2：6	60	0.44	0.72	0.83	0.84	0.71	0.61	0.84	0.91	0.91	0.82
2：6 超～3：0	57	0.50	0.70	0.78	0.82	0.70	0.67	0.83	0.88	0.90	0.82
1：0 超～3：0	362	0.89	0.96	0.93	0.98	0.94	0.94	0.98	0.96	0.99	0.97

注: 日本新版 K 式発達検査標準化集団 (1979 ~ 1981). すべて $p < 0.001$.

4. CDST との相関

1歳から3歳の267名の被検児に1週間において中国版K式発達検査とCDSTを行い、中国版K式発達検査の各領域および全領域DQとCDSTのDQ, MIとの相関を表4に示した。中国版K式発達検査のDQはP-M領域を除いて全ての年齢群でCDSTのDQ, MIとの間に高い相関が見られた。

5. 中国版S-M・SMSとの相関

1歳から3歳の306名の被検児に中国版K式発達検査と中国版S-M・SMSを同時に行い、中国版K式発達検査の各領域および全領域得点と中国版S-M・SMSの全検査粗点との相関を表5に示した。中国版K式発達検査の各領域および全領域得点は全ての年齢区分で中国版S-M・SMSの全検査粗点との間に高い相関が見られた。

表4 CDST との相関 (N=267)

年齢区分	人数	CDST	中国版 K 式発達検査			
			P-M DQ	C-A DQ	L-S DQ	total DQ
1 歳群	86	MI	0.11	0.45 ***	0.47 ***	0.52 ***
		DQ	0.20	0.50 ***	0.38 ***	0.51 ***
2 歳群	145	MI	0.08	0.48 ***	0.36 ***	0.49 ***
		DQ	0.14	0.49 ***	0.32 ***	0.54 ***
3 歳群	36	MI	0.08	0.46 ***	0.73 ***	0.63 ***
		DQ	0.13	0.57 ***	0.59 ***	0.62 ***

***: $p < 0.001$ (Pearson 相関係数)

表5 S-M・SMS との相関 (N=306)

年齢区分	人数	S-M・SMS	中国版 K 式発達検査			
			P-M score	C-A score	L-S score	total score
1 歳群	102		0.33 ***	0.43 ***	0.41 ***	0.43 ***
2 歳群	145	全検査粗点	0.23 **	0.42 ***	0.85 ***	0.73 ***
3 歳群	59		0.29 *	0.28 *	0.48 ***	0.61 ***

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$ (Pearson 相関係数).

表6 実施時刻による所要時間と得点の比較 (N=1096)

年齢	人数	実施時刻	所要時間	P-M score	C-A score	L-S score	total score
1:0 超~1:3	66	午前	22.82	***	58.17	105.82	198.67
	64	午後	26.31		57.67	106.89	198.88
1:3 超~1:6	66	午前	25.64	***	65.59	133.62	240.94
	66	午後	28.68		65.59	134.23	243.29
1:6 超~1:9	96	午前	37.53	*	71.24	161.76	285.70
	95	午後	36.15		70.74	161.58	284.21
1:9 超~2:0	116	午前	39.59		75.59	182.88	325.10
	116	午後	39.48		75.42	181.33	324.09
2:0 超~2:6	124	午前	42.29		80.85	205.90	373.98
	123	午後	41.86		80.26	205.38	372.93
2:6 超~3:0	82	午前	40.44		85.46	234.67	425.66
	82	午後	41.07		86.01	237.17	432.36

午前: 8:30 ~ 11:30, 午後: 15:00 ~ 18:00. ***: $p < 0.001$, *: $p < 0.05$ (独立したサンプルの t 検定)

6. 検査の実施時刻

午前と午後実施群間で t 検定を行った結果 (表 6), 全ての年齢区分において, 実施時刻による各領域と全領域得点および DQ の差異が見られなかった. なお, 1 歳 0 カ月以上から 1 歳 3 カ月, 1 歳 3 カ月以上から 1 歳 6 カ月の二つ年齢区分において, 午前実施よりも午後実施のほうは所要時間が

有意に長く, 時間差が顕著であった (いずれも $p < 0.001$). 1 歳 6 カ月以上から 1 歳 9 カ月の年齢群において, 午後実施よりも午前実施のほうは所要時間が有意に長かったが, 時間差が顕著ではなかった ($p < 0.05$). 他の年齢群において, 時間差は見られなかった.

IV 考 察

1. 中国版K式発達検査の信頼性について

信頼性を推定するには種々の方法があるが、中国版K式発達検査を作成する方針の一つは、日本の新版K式発達検査の標準化過程に可能な限り準拠することであるので、信頼性を検討する際に、原版と同様に折半信頼性によった。

中国版K式発達検査における各領域と全領域の折半信頼性については、狭く取った6つの年齢区分でも、広く取った1歳から3歳の全体でも全般に高い信頼性係数が示され、これらの数値は日本の新版K式発達検査と比較しても十分満足できる値と言える。このことから、中国版K式発達検査は同様な内容を測定していることが分り、中国版K式発達検査は内的一貫性の高い検査と考えられる。一方、信頼性を高める条件として、検査項目の適度の難易度（通過率50%）もあげられ（鎌原ほか，1998）、難易度が低い項目または高い項目からなるテストよりも、適度の難易度を持つ項目からなるテストは信頼性が高くなると思われる。本研究の第1報（高ほか，2003）で報告された、日中標準化集団における検査項目の生活年齢別通過率をみると、1歳6カ月から1歳9カ月の年齢区分において、該当年齢の検査項目（10項目）の通過率は、50%前後（40～56%）であった通過率の項目が日中とも少なかった（いずれも30%，3/10）。これはその年齢区分における相関係数および信頼性係数が、他の年齢区分よりやや低くなった原因であろう。

検者間信頼性の検討に際し、実施する前に予備的な訓練を行って一致度を確かめた（15人の被検児についての得点一致率は80.0%～93.3%であった）。検者の違いによる誤差を避けることに配慮した上、得られた検者間の得点の相関係数は0.95～1.00の範囲となった。ところが、これまでの先行研究では日本の新版K式発達検査の検者間信頼性に関する情報がないので、本検査に準拠すべき基準もない。検者間信頼性に関する先行研究を概観すると、CDSTでは検者間の得点の相関

係数が0.94（鄭ほか，1997）、日本版ミラー幼児発達検査スクリーニング検査では検者間の得点の相関係数が0.97～1.00の範囲（日本感覚統合障害研究会MAP標準化委員会，1989）、中国版ゲゼル発達検査では検者間の得点の相関係数が0.90～0.97の範囲（張ほか，1994）であった。これらの数値から本検査は検者間信頼性が概ね確保されたと判断できる。

合否判断の不一致となった項目について、不一致を説明する理由の一つに検査者の未熟性が考えられる。検査を適切に施行するのにもっとも重要なものは手引書である。検査項目の合否を正しく判断しているかどうかを確かめるために、手引書を注意深く勉強し、しばしば手引書を照合することが決定的に重要である。なお、検査場面において、被検児の反応は千差万別であり、必ずしも手引き書に書かれた正答例や誤答例通りの反応とは限らない。また、判定基準には許容範囲が明確に規定していない場合があつて、照らし合わせにくい面が見られた。「両足跳び」の課題はその一例である。実施手引書に規定された判定基準は「その場で両足で跳び上がれる」ことに過ぎない（生澤ほか，1998）。床から離れる高さ、被検児が跳び上がった地点に下りる必要の有無など許容範囲まで規定されていない。もう一例として「描画」課題において、各々の正答基準は明確に規定しているが、あげられている正誤実例が少ないと感じられた。判定基準または正答基準によく照らし合わせた後、その反応や図形を合格あるいは不合格のいずれにすべきか分らないので、厳しく判断しすぎる例が「両足跳び」、「折り紙」、「描画」で見られた。こうした理由から検査者の熟練の問題だけとも言え切れぬ、判定基準と正答基準の不備があると思われる。先行研究では、検査を身につけるために一番重要なのは、経験を積み、自分の体験の中から合格反応と不合格反応について実感を持つことが指摘されている（下中，1981）。しかし、初心者にとって最初から経験を持つのは不可能である。そのため、日本では新版K式発達検査が完成した8年目、初心者への事例集が出版

された。本研究では検査の実施方法や判定方法について具体例の補助教材も必要だと考えているが、これは今後検討されるべき課題の一つといえよう。

検査結果の安定性を保証するため、検査環境、実施項目、言語指示、制限時間、レポートなど十分配慮した上得られた検査—再検査の相関係数は、各領域と全領域得点が0.99～1.00の範囲、各領域と全領域DQが0.83～0.98の範囲であり、得点一致率は平均93.8%であった。再検査信頼性に関する先行研究を概観すると、CDSTは7日間～10日間の間隔で相関係数が0.90（鄭ほか、1997）、日本版ミラー幼児発達検査スクリーニング検査は1週間～4週間の間隔で相関係数が0.85～1.00（日本感覚統合障害研究会MAP標準化委員会、1989）、日本版デンバー式発達スクリーニング検査は1週間の間隔で項目の一致率は平均95.8%であった（上田、2000）。このことは中国版K式発達検査が高い安定性を持つ検査であることを証明している。得点の高さについては、比較的短い間隔で再検査を実施したことにより、被検児が検査項目に「熟知性」を生じ、再テストの得点を増加させたと解釈される。行為の記憶に関する「再認の2過程説」（Jacoby, 1991）はこの解釈に適用できると考えられる。「再認の2過程説」とは、「熟知性」と「意識的な回想」との2過程に基づいて、再認判断を行えるという考えである。それを参考すれば、「飛び降り」、「はめ板」、「入れ子」、「数の復唱」における再テストの得点の高さが「熟知性」による再認判断する過程を反映していると想定できよう。短い間隔で再検査を実施することにより、1回目テストの段階が学習の段階になってしまい、2回目テスト項目を見たときに、「見覚えがある」（「熟知性」）という感覚に基づいて「学習した」と判断することがあろう。そして、「熟知性」は、プライミング効果（先行刺激の処理によって、後続刺激の処理が促進されること）（太田ほか、2001）と同様に、1回目で処理した刺激を、2回目で再び処理する際には、その処理が流暢になったり、正確になったりされていると

考えられる。したがって、「飛び降り」、「はめ板」、「入れ子」、「数の復唱」は「熟知性」を生じやすいテスト項目であるといえよう。ただし、異なる間隔期間で「熟知性」の高さをはっきりさせるためには、さらにデータを集める必要があり、この点は今後の課題と考えている。

検査時間には二つの意味がある。一つは時刻を、もう一つは所要時間を意味する。検査の信頼性に対して、この両方とも重要である。日本の新版K式発達検査では幼児の検査に最適な時刻を指摘していないが、一般に朝の方（午前）が検査しやすいと思われる（日本感覚統合障害研究会MAP標準化委員会、1989；張ほか、1994）。本研究の結果により、時刻が所要時間に影響し、被検児の年齢が小さいほど、その影響が強いことが示唆された。所要時間が長くなると、被検児の検査への集中力が減少し、検査の成績にも影響を及ぼす可能性があると思われる。なお、時刻の検査成績への影響が認められなかったのは、2時間ほどの昼休（昼寝と間食時間を合わせる）により心身状態が調整されたことが考えられる。したがって、時刻に拘らず、疲労、空腹及び昼寝から目覚めた状態を避け、被検児のよい状態の時間帯に実施するのが重要である。

2. 中国版K式発達検査の妥当性について

妥当性とは検査がどの程度目的を達することができるかの程度を示すものである（下中、1981）。妥当性検証のためのいくつかの基準があるが、本研究では267名の被検児に1週間をおいて中国版K式発達検査とCDSTと二つの検査を行い（検査と基準変数）、基準関連妥当性を検討した。

CDSTは中国上海医科大学付属小児病院によって、中国の6カ所の都市部における0～6歳児の6569名を対象として標準化された発達スクリーニング検査である。1997年に公表され、現在、全国的に乳幼児の発達スクリーニング検査に広く利用されている。120個の検査項目は運動（30個）、社会適応（30個）、知的（言語と操作を含む60個）の3つの能力領域に大別され、結果は発達指数

(DQ) と知能指標 (Mental Index : MI) の二つの指標で示されている。運動能力は粗大運動、身体バランス及び協応動作を調べる検査項目から評価される。社会適応能力は対人関係及び衣服の着脱、食事、排泄などの身辺自立を調べる検査項目から評価される。知的能力は言語理解及び道具操作を調べる検査項目から評価される。検査結果について、DQは領域ごとの粗点合計と生活年齢から換算表によって求め、運動、社会適応、知的の3領域の総合的能力を表す。MIは知的領域の粗点と生活年齢から換算表によって求め、言語理解及び道具操作の知的能力を表す(鄭ほか, 1997)。本研究でこの検査を選んだのは被検児年齢への適用性であること、検査の臨床的実用性であることなどが主要な理由である。これまでの研究では、CDSTのDQ、MIと中国版ゲゼル発達検査のDQ ($r = 0.60, p < 0.01$; $r = 0.57, p < 0.01$) および中国版WPPSIの全検査IQ ($r = 0.53, p < 0.01$; $r = 0.68, p < 0.01$) との有意な相関が見られた(鄭ほか, 1997)。

本研究では、中国版K式発達検査のC-A DQ、L-S DQおよび全領域DQはCDSTのDQ、MIと高い相関が見られ、中国版K式発達検査のC-A DQ、L-S DQ、全領域DQとCDSTのDQとMIとの対応が高いことが示された。このことから中国版K式発達検査は子どもの発達全般の能力を反映すると共に、子どもの日常生活動作、身辺自立といった社会適応能力の基礎となる能力とも関連することが推測される。なお、P-M DQがCDSTのDQとMIとの相関が低かったことは、CDSTがスクリーニング形式の発達検査であるため結果の表示の粗さが主たる原因であると推測される。すなわち、P-M DQは単一の運動能力を表しているが、CDSTのDQは運動、社会適応、知的の3領域の総合的能力を表していることに関係している可能性があると考えられ、各検査にはそれぞれの特性があることを示唆している。このことは中国版K式発達検査作成の主要目的に合致するであろう。

中国版K式発達検査は子どもの日常生活動作、

身辺自立といった社会適応能力の基礎となる能力と関連するという仮説を明らかにするためには、発達検査以外の変数との関係をも検討する必要があると考えられたため、本研究では306名の被検児に中国版K式発達検査と中国版S-M・SMSと二つの検査を同時に行い、その相関を検討した。

S-M・SMSは知的能力とは独立した社会的適応能力を測定するのに都合がよく、岡野らはIQの低い子どもにおいては、知能検査に加えてS-M・SMSを併用することにより、知的側面以外の能力の発達を捉えることができると指摘している(岡野ほか, 1983)。鈴木らは脳性麻痺児にS-M・SMS、上肢・体幹下肢運動年齢検査(MAT)、知能検査(WISC-R)を行ない、それらの関連を検討し、社会生活能力は知能、体幹下肢の運動年齢と相関が見られず、上肢の運動年齢と相関が認められたと報告している(鈴木ほか, 1997)。

日本の新版K式発達検査は特定の知能理論に基づいて作成された知能検査ではないが、知的発達を測定する検査項目を広く集めたものであり、検査項目に社会適応能力と関与した項目も多く存在している。本研究では、中国版K式発達検査の各領域および全領域得点とS-M・SMSの全検査粗点との間に高い相関が見られ、中国版K式発達検査が子どもの日常生活動作、身辺自立といった社会適応能力の基礎となる能力と関連することを明らかにした。これは原版製作者の発達に関する考え方を反映したものと思われる。

V 結 論

中国版K式発達検査による中国被検児の検査結果に基づいて検査の信頼性と妥当性の検討を行った。その結果から以下の結論が得られた。

(1) 検者間信頼性の検討では、検者間の得点の相関係数が0.95～1.00であり、検者間信頼性が概ね確保されたと判断できる。

(2) 再検査信頼性の検討では、1週間を隔てた検査—再検査の相関係数が各領域と全領域得点で0.99～1.00、各領域と全領域DQで0.83～0.95と

なり、十分な信頼性が得られた。

(3) 折半信頼性の検討では、奇偶折半した項目間における得点の相関係数と信頼性係数は全ての年齢群と領域を通して高いものであり、原版と比較しても満足すべき信頼性係数と言える。

(4) 基準関連妥当性の検討では、1週間を隔てた中国版K式発達検査の各領域、全領域DQとCDSTのDQ、MIとの相関は、姿勢・運動領域を除くどの領域でも高い相関が見られた。中国版K式発達検査の特定の領域がCDSTと相関したことは、その特定の領域とCDSTのDQ、MIとの対応が高いことを意味し、中国版K式発達検査の妥当性を直接的に示している。また、中国版K式発達検査の全ての領域がCDSTと相関しなかったことは、中国版K式発達検査には検査の特異性があり、CDSTと類似していない部分があることを意味し、中国版K式発達検査の妥当性を間接的に示している。

(5) 発達検査以外の変数との関係の検討では、中国版K式発達検査と中国版S-M・SMSを同時に行った結果から、中国版K式発達検査の各領域及び全領域得点とS-M・SMSの全領域粗点との間に高い相関が見られ、中国版K式発達検査が子どもの日常生活動作、身辺自立といった社会適応能力と関連することを明らかにした。

文 献

- 生澤雅夫, 松下 裕, 中瀬 惇 (1998): 新版K式発達検査法—発達検査の考え方と使い方— (嶋津峯真監修), ナカニシヤ出版
- 高 健 (2004): 中国版K式発達検査の標準化に関する研究 第2報—中国版「得点・発達年齢換算表」の作成と1歳～3歳児の検査結果の分析, 民族衛生, 70(1), 31-45
- 高 健, 郷間英世 (2003): 中国版K式発達検査の標準化に関する研究 第1報—1歳～3歳児の検査項目の適切性と難易度の検討, 民族衛生, 69(4), 112-131
- 高 健, 郷間英世, 小沢寺敬子ほか (1999): 就学前児の社会生活能力について—日本と中国の比較を通しての検討—, 小児保健研究, 58(4), 459-464
- 高 健, 郷間英世, 宮田広善ほか (2001): 脳性麻痺幼児の社会生活能力に関する研究—日本と中国との比較を通しての検討—, 発達障害研究, 22(4), 55-61
- 鎌原雅彦, 宮下一博, 大野木裕明ほか (1998): 心理学マニュアル 質問紙法, 北大路書房
- 日本感覚統合障害研究会MAP標準化委員会 (1989): 日本版ミラー幼児発達検査スクリーニング検査検査マニュアル, HBJ
- Jacoby LL (1991): A process dissociation framework: Separating automatic from intentional use of memory. *Journal of Memory and Language*, 30, 513-541
- 岡野美年子, 永田陽子, 本多経子 (1983): 幼児相談における新版S-M社会生活能力検査の利用, 心理測定ジャーナル, 19(1), 9-16
- 太田信夫, 多鹿秀継 (2001): 記憶研究の最前線, 131-139, 北大路書房
- 下中邦彦 (1981): 新版心理学事典, 433-435, 平凡社
- 鈴木亮二, 吉田一成, 岩谷 力 (1997): 脳性麻痺児の社会生活能力について, 小児保健研究, 56(5), 621-626
- 鄭 慕時, 馮 玲英, 劉 湘雲 (1997): 0～6児童智能發育篩査測驗全国城市常模的制定, 中華兒科雜誌, 35(3), 117-120
- 張 致祥, 左 啓華, 雷 真武 (1995): 「嬰兒—初中學生社会生活能力量表」再標準化, 中国臨床心理学雜誌, 3, 12-15
- 張 秀玲, 李 寄平, 秦 明鏡 (1994): Gesell發展診斷量表3歳半～6歳北京修訂本的製定, 中国臨床心理学雜誌, 2(3), 148-150
- 上田礼子 (2000): 日本版デンバー式発達スクリーニング検査 増補版, 53-66, 医歯薬出版
- (受稿 2003. 5. 28; 受理 2003. 11. 27)