

知能検査繰返しによる得点増加に関する研究

住 田 幸 次 郎

1. 問 題

知能検査法の研究は、近来、教育上の重要な問題として益々その意義が大きくとりあげられるようになってきた。「現代は評価の時代である」とまで言われるようになり、教育測定に対する関心が教育界においてとみに高まって来ているが、なかでも知能検査はその理論的發展において、また実用上の有用性において、諸教育測定の代表選手たる地位を保持しているといえよう。

しかし、このように言っても、現在の知能検査は、まだまだ知能の測定については未解決な基本的發展を多くかかえており、その有用性を十分に発揮するに至っていない。

その一つとして、いまからとりあげる再検査効果に関する組織的研究は、間接的には他の目的からなされた知能に関する研究結果から経験効果のあることがおぼろげにわかっているとしても、直接に諸条件の統制、比較をおこなって研究対象としたものがみられないのはその条件統制、組織的研究が困難なためである。知能検査を利用する際に、同一検査を同一の被験者に繰返し用いることは多くおこなわれる研究方法である。いまずぐに頭に浮んだものでも、reliability（信頼度）測定の一手段として再検査法を用いたり、知能の縦断研究に一定間隔で以てテストを施行したり、児童個人の知能を正しくとらえて学習指導に役立たせるために数年ごとに繰返す、といったことが頭にうかぶであろう。

ところで、知能検査の test-retest 研究は殆んど全部が reliability の測定に関連したものに限られている。

Kuder および Richardson¹⁾ が同一検査の折半による方法を軌道にのせ、諸種のそれに関連する式を發表する迄は、再検査法は最もしばしば信頼度測定に用いられた。これらについては Guilford²⁾ や Gulliksen³⁾ に要領よくまとめられている。

伸びの状態もしくはその要因について、知能検査の再検査を問題にとりあげた研究のすくない原因としては、非常に多くの状況要因の統制が困難であるし、被験者を test-retest の期間中一定に保ちにくいこと、偏差値方式をとらない粗点累加方式の検査では発達による自然能力増の推定が極めて困難であること等の理由に基づいているものと思われる。公開されているこの種の資料は、僅かに知能検査制作の際の資料として^{4) 5)}みられることがある程度である。ここでは紙面の都

1) Kuder, G.F. & Richardson, M. W. The theory of the estimation of test reliability, *Psychometrika*, 1937, 2. 151—160.

2) Guilford, J.P. *Psychometric Methods* (2nd. Ed.) 1954. McGraw-Hill P. 341—413.

3) Gulliksen, H. *Theory of Mental tests* 1950. John Wiley & Sons.

4) 小見山栄一編 教育標準検査ハンドブック 1959. 東洋館出版 P. 1—210.

5) 京大知能検査研究会編 京大知能検査報告(1)~(12) 1952~1959.

合上、個々の文献については詳述せず、必要に応じて以下文中で引用することにした。

知能検査は個人ひとりびとりの環境や経験の差により成績が左右されることのすくないような、誰もがよく知っている程度の知識と手段のみを用いて、どの個人にとっても新しい状況設定の下において課題を与え解決に至らせる。

それだから、理論上は再検査の伸びはあまり大きくないであろうと想像されうる。しかし実際には紙と鉛筆またはそれに近い道具のみを使用ししかも経験効果等を除去することは不可能に近い。

伸びの防止が不可能なもので避くべからざるものという証明がなされるならば、むしろ積極的に再検査値の補正や適当な *test-interval* を決めることが必要であろう。そのためには伸びの生ずる *process* を正しくとらえ、要因を分析し、テストの種類方法と伸びとの関連をできるだけ精細にとらえなくてはならぬ。この際の研究に慎重を期すためには同一検査の反復に限っていたのではだめで、たとえテストの名称やみかけの形式が表面上でどんなに異っていても、因子構成の類似や状況への馴れ、使用材料からくる相互の転移量などの後検査への作用が併せ知られることも必要であろう。

さて、もし知能検査の再検査結果が変容のはげしいものであるならば、従来数多くなされてきた知能の追跡研究や知能値変動に関する研究においては、この点についての何らかの考慮を払う必要があると思われる。そういった点を見無視して、もしくは資料の不足から何らかの処理をすることができずに従来の結果を解釈してきたことに非常な危惧を持つものである。この懸念を少しでもあきらかにしたい目的から、次のような知能の再検査方法に関する総合的な実験をなそうとするものである。

〔目的〕 知能検査における繰返し (*retest*) に際しておこる知能偏差値変化の状態を知、そのような変化をもたらす諸要因を分析する。

2. 実 験 方 法

(イ) 被験者 京都市内の3小学校児童男女。3～6学年児童全員を対象に同一期間(昭和33年6月)に一斉検査をおこなった。該当小学校の選択にあたっては、校区下における産業別人口動態調査の結果などを勘案して階層の偏りの影響が小さいようにした。⁶⁾ 再検査の場合には、これらの児童に学校差が生じないように各間隔条件を割当てた。

(ロ) 実験材料 京大NX8-12知能検査。同一検査を一定間隔後に反復実施する。⁷⁾

(ハ) 再実験の期間 原実験より後1ケ年にわたる期間。昭和34年6月迄。⁸⁾

(ニ) 実験実施者 原、再検査とも特に訓練し熟練した4名が繰返し実施。実験者による差を防

6) 京都市内 中京区、東山区、南区より各1校

7) 京大NX8-12の検査内容については Tab. 1 参照のこと

8) 実施期日の厳密な日時よりの変動はつぎの範囲に止めた。原テストに対して丁度4ヶ月後、8ヶ月後、12ヶ月後の日時より±3日以内、同一条件内において最も早い学校と遅い学校の差は1週間以内に実施した。

Tab. 1 京大NX8-12知能検査の内部構成

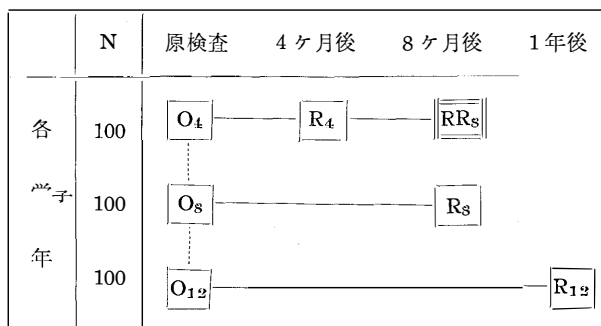
検査番号	問題内容	時 間	項目数	仮想因子	備 考
1	反 対 語	2 分	20	V	ある言葉と反対の意味の語をえらばせる
2	記 憶	2 分	12	(M)	有意義な文章(日常文)について質問する
3	同図形発見	1分30秒	18	S	同じ図形をさがす
4	異質発見	2 分	20	V(R)	ある語群のなかで、1つだけ他と異質的なものをのぞく
5	点 図 形	2分30秒	12	S	点をむすんで問題とおりの形をつくる
6	数 交 換	2分30秒	15	N	ある法則にしたがって数を交換し、正しい数を知る
7	語 い	1分30秒	30	V(W)	単語の空所を補充させる
8	数 計 算	2 分	20	N	一定の数になるようにする際に不要な数値を除かせる
9	単語マトリックス	3 分	20	V(R)	前後上下の関係からみて最も適切な単語をえらぶ

止するために計画的に4名を配置した。

(ホ) 実験計画 Fig. 1のように3種の検査間隔を違えた群を、各学年についてつくり、それぞれの群を知能偏差値平均および分散がほぼ等しいように抽出した。各小群とも男女各50、計100名である。説明上の便宜から

Fig. 1 実験計画図

原検査 O (original), 再検査 R (retest), 三度目の検査 RR (re-retest) の記号を用い、間隔条件(単位月)を数字で副え、 O_4 , R_{12} の如くする。故に O_8 , R_8 は被験者は同一で実施した時期に8ヶ月の差のある資料を言う。



Tab. 2 Original test における各群偏差値平均

	O_4	O_8	O_{12}	全 体
3 年	50.18	50.28	50.23	50.23
4 年	50.63	50.50	50.20	50.44
5 年	50.49	50.29	50.83	50.54
6 年	50.21	50.08		50.15
全学年	50.38	50.29	50.42	50.36

各N=100 (男50女50)

9) Tab. 2 参照のこと

(3) 各条件群における相関を、それぞれの知能偏差値を使って product-moment 法により算出する。普通、信頼度の測定には、この相関係数値を諸公式で修正したものを信頼度として用いるのであるが、後述のように信頼度をあてはめる方法により係数値およびその意味が変わるので、Pearsonの r を無修正のままで用いることにする。

以上3つの測度を中心に検定・分析を試みることにする。¹⁰⁾

結果の整理にはいる前に、Tab. 2に示された各群被験者の偏差値平均の等質性検定をおこなってみる。¹¹⁾6年生が条件不揃いのため、3～5年生のSS平均を分散分析で検定したところ、差が認められなかった(Tab. 3)ので、各群は原検査に関して等質と考えてよい。¹²⁾

結果の説明ならびに分析

Tab. 3 被験群等質性の検定

以上、資料にてあわせながら分析・検討を試みることにしたい。

さきに、original test時における各groupの知能偏差値が等質であることを確かめたので、再検査における伸びの状態を知るため各testの偏差値平均を全体および学年別について相互に検定した。偏差値の状態はFig. 2に示したとおりである。

変動因	df	S	M. S	F
条 件	2	101	50.5	
学 年	2	1483	741.5	1.230
誤 差	4	2412	603.0	
全変動	8	3996		

再検査による偏差値の伸びは O_{12} — R_{12} 間の比較をのぞき統計的に有意である。また、 R_{12} が最も原検査の状態に近いことがわかった。

この理由としては、小学3～6年といった年齢程度においては、1年間のCA発達に伴う知的作業量上昇がほぼ再検査上昇にひとしい程度のものであるからだろう。因みに粗点は R_{12} に於ても O_{12} より有意に伸びている。 R_4 と R_8 の偏差値はほぼ等しいと言えるが、 R_8 — RR_8 はいずれも原検査より8ヶ月後になされているが開きは大きい。 RR_8 には4ヶ月後の挿入検査(R_4)の効果があきらかであって、知能検査をペーパーテストで実施する場合の一つの限界をここにも示しているといえよう。知能検査は繰返し経験効果の比較的小さいものを下位検査問題として構成されているにしても、極めて短い期間での繰返しは知能偏差値または指数が過度に増加することが知られた。また、1年の間隔をあければ、この増加量はそれほど激しくないこともわかった。今回は研究の都合上、期間を原検査から1年後に限ったが、もっとこの傾向をはっきりと吟味するためには、更に2年後・3年後といった長い時間軸での研究が必要であり、これらの知見がまとめれば同一検査の反復を研究法とする実験、たとえば縦断・追跡研究における各時点での数値を単にそのまま比較してよいかどうかの回答となるであろう。

10) ここに分析した資料は、さきに日本心理学会第23回(北海道)においての梅本堯夫 奥野茂夫、広田実らとの共同研究 京大NX知能検査報告(12)における発表資料を含む。

11) 6年生が卒業のため、1年後の資料を欠く。

12) この際、個人SSでは等質であったが、各テストにおける変動は統制が困難のこともあり、視察するとあまり下位テスト間の差がないので、そのままみすごした。後の研究において、下位テスト同志、または下位テストについて検定する場合には、必要に応じて原テスト間の差も検定した。

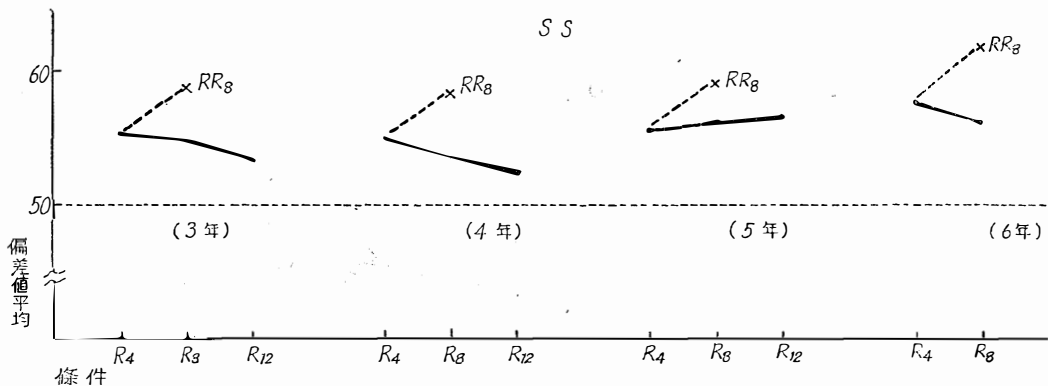
Tab. 4 結 果 一 覧 表

original testでの各グループは等質であろうか (SS)	知能の群間等質性の検定 (分散分析)	群間に差なし
test-retest 間隔と伸びとの傾向 (SS)	各条件間の各群検査偏差値平均の検定 (t検定)	(全体) (1) R_4-O_4 , R_8-O_8 に有意な伸び (2) $R_{12}-O_{12}$ は伸びが小さい (3) RR_8 の伸びは R_8 に比して大きい (4) R_{12} の分散は僅かではあるが他の再検査群よりも大である (学年別) 高学年の方がやや差の大きい傾向
偏差値に「伸び」が生じる要因となりそうな二、三の条件についての分析 (偏差値の伸びの差) 4年生の資料について	知能程度のちがう群間の伸び (分散分析) 男女別の伸び (分散分析) 記憶力量の差—記憶下位検査によるGP…両群とも知能SSは一定 (分散分析) 全体としてどの条件の伸びが大きい (U検定)	知能M群P群>G群 4ヶ月後の伸び>1年後の伸び 下位検査によって差あり 性差による差なし 記憶テスト力量は全体として差なし 記憶検査自体の伸び $P>G$ (下位検査) 単語マト・異質発見では $G>P$ 差の分布—差の分布は正規又は不正規 (O_4-RR_8) と他のものに差あり $O_{12}-R_{12}<O_4-R_4$, O_8-R_8
発達段階によって伸びに差はないか	条件別下位テスト別偏差値検定—original 同志 retest 同志 (t検査)	(原検査学年間)一部の例外を除き差なし (再検査学年間)条件下位テストの如何を問わず 大体において発達差上昇がみられる $6年 \geq 5年 > 4年, 3年$ (下位検査) 1. 点図形, 同図形, 反対語 (因子的に空間因子又はそれに近いもの) —伸びが大 2. 記憶— RR_8 において各学年差が顕か 3. 伸びの小さいもの 数検査 言語推理検査 単語完成 4. 伸びの逆転 (5年) 単語マトリクス
伸びの相違は下位検査によって生じ方がちがわないだろうか (全体および4年生) の資料について)	下位テスト別偏差値平均検定 (t検定)	(原検査下位テスト間) Test2 (記憶) が他に比して original な群で低かった*。他は概ねよい (再検査下位テスト間) 1. Test8 (数計算) の伸びが小さい 2. Test2 (記憶) の伸びが小さい* (RR_8 においてみられた特徴) 下位テストはつぎの二つに分けられた (イ) RR_8 において伸びの低い下位検査群 異質発見, 数交換, 単語完成, 数計算 (ロ) RR_8 において伸びの高い下位検査群 反対語, 同図形発見, 単語マト
test-retest 相関値間の変動比較 ($r \rightarrow z$ 変換)	test-retest間隔条件の相違と r test-retest学年段階の相違と r	(イ) SS R_4-O_4 の $r > R_{12}-O_{12}$ の r (ロ) RR_8-R_4 の rは他のR-Oよりも大 (ハ) 記憶下位検査は係数変動が大である (ニ) 数交換, 数計算では条件による変動が小 (ホ) 単語完成において $RR_8-O_4 > R_8-O_8$ $RR_4-O_4 > R_{12}-O_{12}$ (イ) O_8-R_8 においてやや高学年がたかい (ロ) $O_{12}-R_{12}$ では差がない (ハ) 反対語, 単語マト, 点図形で発達の差が大きい (ニ) 発達の差の少ないテストは数交換 同図形 単語マトである (ホ) 総体に安定している

統計的に有意なほどではないが、 R_{12} における分散は他条件の場合よりもやや大である。これは各学年ともにあらわれた傾向であるが、その1つの理由として R_{12} においてはP群における先検査効果が他の条件よりも小さくなるために、分散が幾分拡大すると考えられる。¹³⁾

今迄述べてきた全体的な傾向を学年別に考えてみてもほぼ同じことが言える。ただ、高学年児の方が低学年児集団よりも条件による伸びの

Fig. 2 偏差値平均の Test-Retest における上昇下降の比較



開きが大きい、これについては後に発達について触れる際に更に検討したいと思う。

さて、こんどは以上みてきたような再検査の伸びがいかなる要素により生じたと考えればよいのか、今迄に考えられたいくつかの要因の中から二三のものをとりあげて検証してみよう。各群の偏差値平均を比較するとき、原検査に差があったり、知能平均と別の基準により統制がなされたりした場合には、再テスト得点間の差が必ずしも妥当でない。そこで、4年生の資料を代表としてとりあげて、各個人の「伸び」、つまり原・再テスト間の偏差値差（raw score の差ではない）を測度に用いることにして、分析にはいる。

A. 知能の高さとの関係 知能検査成績のよい者は2度検査すると知能がもともと良いから益々高くなるだろうか。それとも、伸びは一樣にあがるもので知能の高低とは無関係なものだろうか。この問題をとらえるために知能の高いもの（G）、低いもの（P）、平均に近いもの（M）、各群20名男女半数ずつを抽出し¹⁴⁾、伸びを測度にして分散分析をおこなった。（Tab. 5）その結果、知能のG・M・P群別により伸びに差のあることがわかったので、さらに群間検定した結果知能G群は他群にくらべ伸びが小さいことがわかった。他にテスト期間や下位検査の相違が伸びに響

13) Fig. 2 参照。

14) 4年生の被験者。

Tab. 5 分散分析表 (知能群別)

変 動 因		df	S		F	備 考
主 効 果	A(テスト期間)	2	611.7	305.9	4.067	$P<.05$
	B(知 能)	2	1282.2	641.1	8.525	$P<.01$
	C(下位検査)	8	1359.3	169.9	2.259	$P<.05$
二次交互作用	A × B	4	572.1	143.0	1.901	
	A × C	16	1667.6	104.2	1.385	
	B × C	16	2594.7	162.2	2.156	$P<.01$
三次交互作用	A × B × C	32	3028.1	94.6	1.257	
誤 差 E		1539	115757.5	75.2		

くことがみられたが、これについては別にのべる。

一般に、知能の高い者は再検査でも増加が激しいのではないかと予想されたが、実験からはむしろM・P群の上昇が高い。この理由について考えてみると、(i) time-limit 形式による知能検査において悪い成績をとるもののなかには、課題解決の速度の遅さに起因しているものが多い。これが繰返しにより早くなる。(ii) 日常ほとんど使用しないような操作を要求する下位検査において、知能の低い群は高い群に較べて、まず「テストのやりかた」をのみこむことが困難である。再検査時には練習効果やCA 上昇に伴う MA 発達により困難さが減少する。

(ii) 知能の高い群は解決方法の理解や速度は既に初めのテストで発揮されているから、初めのテストで失敗した項目は自己の能力以上のものであることが多いから、再検査してもあまり伸びないのだろう。これらのことは、下位検査における性質と各群の伸びの割合を併せ考えれば納得できるのではなかろうか。

B. *sex difference* 知能検査による性差が殆んど不可避であることは Brown, M. H.¹⁵⁾ らの

Tab. 6 分散分析表 (男女差)

変 動 因		df	S		F	備 考
主 効 果	A 男 女	1	4760	4760.0	2.084	
	B 条 件	2	40145	20072.5	8.792	$P<.01$
	C 下位検査	8	109576	13697.0	6.000	$P<.01$
二 次 交 互	A × B	2	3157	1578.5	—	
	A × C	8	50463	6307.9	2.763	$P<.05$
	B × C	16	140109	8756.8	3.836	$P<.01$
誤 差		16	36524	2282.8		

15) Brown, M. H. & Bryan, G. E: Sex as a Variables in Intellignce test performance. *J. educ. Psychol.* 1957, 48, 273—278

主張にもみられる。従って、伸びにもそれが影響しないかとの仮定で、男女別に分散分析をおこなったが、男女間に差は認められなかった。(Tab. 6) 故に、性の要因は伸びを誘発する条件としてはそれほど大きくない。

C. *memory power* 再検査の伸びが原検査より上昇する場合、原検査の問題そのものか、または特殊な解決法を記憶しているからかもしれないとの推論は容易になりたつ。さすれば記憶力量のすぐれている者の方が伸びは高い筈であるから、これについて考察してみよう。原検査(第二検査 日常記憶)の記憶テスト成績の良いものと悪いものを個人 SS のほぼ等しい 2 名を対¹⁶⁾にしてえらびだし、伸びを分散分析したが、GP 群間に差はみられなかった。このことから、再検査伸びは個人の記憶力量の多少とはそれほど大きい関係がないことがわかり、再検査の伸びは問題そのものの記憶というよりは特定検査に対する set その他の条件変化が影響しているものと

Tab. 7 分散分析表 (記憶力量群別)

変 動 因		df	S		F	備 考
主 効 果	A(テスト期間)	2	169.6	84.8	1.184	
	B(記憶成績)	1	20.0	20.0	—	
	C(下位検査)	8	1130.9	141.3	1.973	$P < .05$
二次交互作用	A × B	2	34.4	17.2	—	
	A × C	16	2329.5	145.6	2.033	$P < .05$
	B × C	8	1659.7	207.4	2.896	$P < .01$
三次交互作用	A × B × C	16	1118.4	69.9	—	
誤 差 E		486	34838.0	71.6		

思われる。Tab. 7 の原表を下位検査別にながめてみると、日常記憶で G 群の伸びが殆んどなく、一方 P 群の伸びが非常に大きい。また単語マトリクスや異質発見は他の諸下位検査とは逆に G 群の伸びが高くなっている。これらの解釈を、ここに述べた資料のみから論ずるのは差し控えるが、記憶があまり問題にならないと考えられる言語推理検査に記憶力量の差があらわれている点は興味ぶかった。

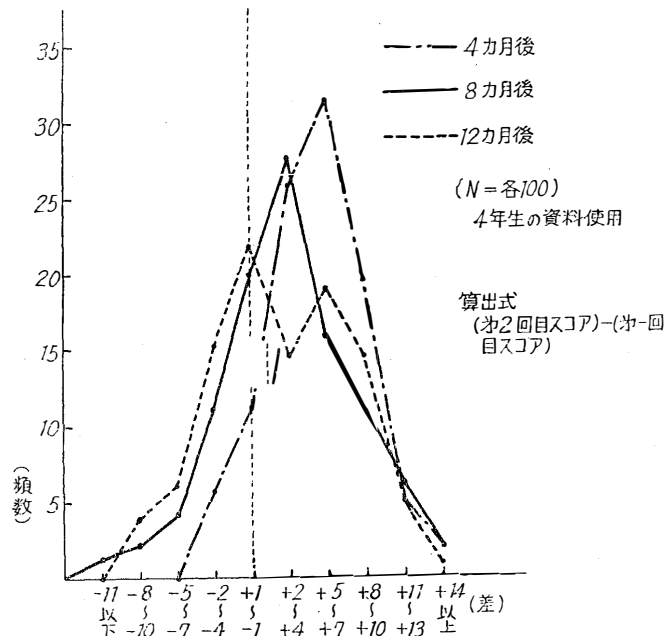
総体に、記憶力量と伸びとはあまり関係がないと思ってよいのではないか。

D. 間隔条件による伸びの差 さきに、間隔条件の異った場合における偏差値平均の差を比較したが、伸びの差を測度として考えた場合でも同様の傾向があるかどうかを知ろうとした。

伸びの差の分布は Fig. 3 のようになり、必ずしも正規分布するとは限らないので nonparametric な統計量を使い、伸びの差を単位に U-test したところ、RRs と他群との間に大きな差があり、また (O₁₂—R₁₂) の伸びの低さがめだっていて、概して傾向は平均値の場合に類似

16) 標準化資料における全検査SS—記憶検査SSの相関は $r = .623$ である。

Fig. 3 Original Test-Retest の SS 差分布



Tab. 8 伸び量の条件間における比較 U検定

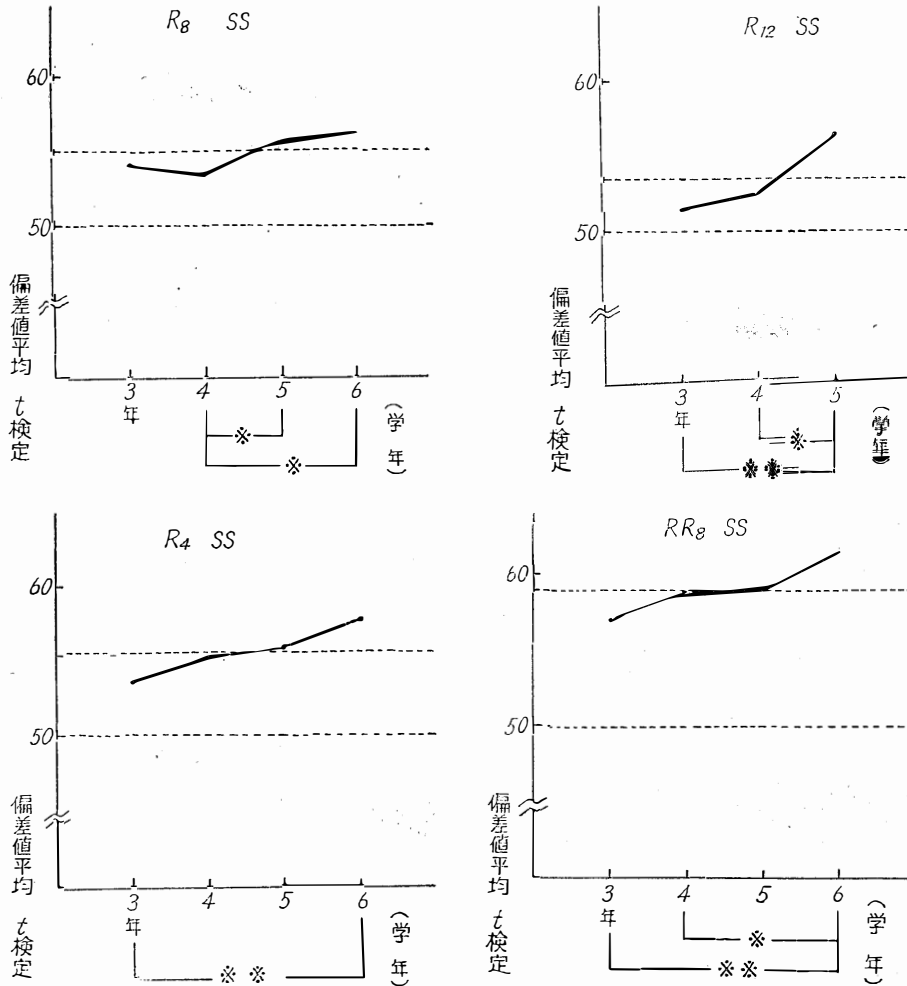
	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
3 年		※ ※		※	※ ※			※	※ ※	
4 年		※		※	※ ※			※ ※	※ ※	
5 年	※ ※	※			※ ※	※	※	※		
6 年	※				※ ※			※		
全 体		※ ※		※	※ ※			※ ※	※ ※	※

(注) A : ($R_4 - O_4$) の伸び量
 B : ($RR_8 - R_4$) 〃
 C : ($RR_8 - O_4$) 〃
 D : ($R_8 - O_8$) 〃
 E : ($R_{12} - O_{12}$) 〃

している。群間差が前の場合ほどはっきりしていないのは、列位検定(U)を用いたためである。

E. 発達 さきに各条件同志の検定比較をおこなった際に、高学年群の方が伸びが激しそうだという推測をしておいた。これを厳密に考えてみよう。個人偏差値平均は学年間に全く差が無いので、再検査成績相互の検定をおこなった。その様相は Fig. 4 に示したように発達差が顕著である。一般に 3・4 年と 5・6 年の 2 つの学年の group にわけることが出来、それぞれ他の group と差がみられる。一般的に言って、高学年は発達による自然増加量以上に伸びが低学年に比して大きい。また、3 年生と 4 年生の間には殆んど差はない。従って、低学年よりも高学年は非常には

Fig. 4



げしい上昇をもたらすので、同一検査を繰返して用いることにより知能測定をなすのは、検査の適用年齢範囲内では高学年となるほど好ましくないと言えよう。

高学年の伸びが、高まる原因について十分な説明をおこなうには資料が不足しているが、低学年においてはいくら繰返しても破れない困難度の壁があるわけだが、年齢が高まると各個人の能力量の増大が前経験などを有利に使って困難度の低下をもたらしていると言えよう。

ここで、さらに詳しく下位検査ごとにその関係をみたものを Tab. 9 に表示する。original test と re test の間において、発達差が間隔をどのように変えてみても生じるのは点図形・同図形発見・反対語といった空間因子および言語因子検査のうちで因子排列が最も空間因子に近い検査である。何故こういった空間因子の伸びが高まるか、これについては空間知覚発達とか問題の

Tab. 9 発達による偏差値上昇度

検査 条件	1 反 対	2 記 憶	3 同図形	4 異 質	5 点 図	6 数 交	7 単 語	8 数 計	9 単マツ
R ₄	※※3—⑥ ※※4—⑥ ※※(5—⑥)	※※3—⑥ ※4—⑥	※※3—⑥		※3—⑥			※※4—⑥	※④—5 ※5—⑥
RR ₈	※※3—⑥ ※※4—⑥ ※※(5—⑥)	※※3—④ ※※3—⑤ ※※3—⑥ ※※3—⑥ ※4—⑥ ※5—⑥	※※3—⑤ ※※3—⑥ ※4—⑥		※※3—⑤ ※3—⑥				※※④—5 ※※5—⑥
R ₈			※3—⑤ ※※3—⑥	※(3—⑥) ※※4—⑥				※※4—⑤	
R ₁₂	※※3—⑤ ※※4—⑤		※※3—⑤ ※※4—⑤	※※3—⑤ ※4—⑤	※※4—⑤	※3—⑤	※※3—⑤ ※※4—⑤		

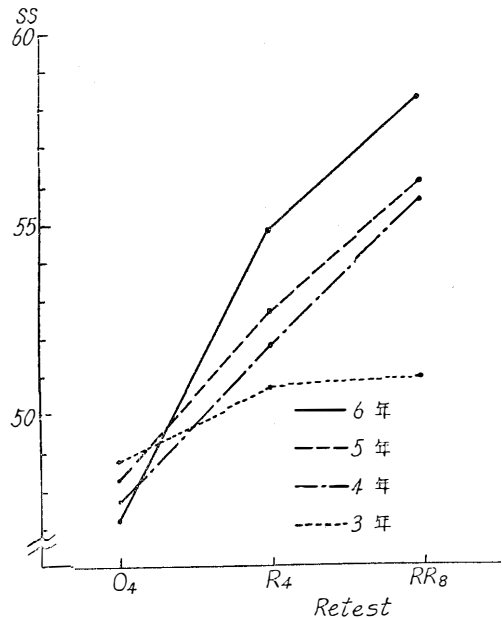
(注) 欄内数字は学年。円でかこんだ数字は、伸びが大である方の学年
 ※※… $P < .01$ ※… $P < .05$ カッコでかこんだものは original な方に差があるため厳密には意味をもたないが参考迄にあげたもの

解決に対する set, situation に関する諸種の研究が必要であるが、これら空間因子的な検査は高校生大学生の資料からの分析によれば知能の優れた者と劣っているものとを弁別するに非常に適した検査であることが、いろんな面から実証されているので、まだ知的発達の途次にあるこれら年齢段階においては、CA発達に伴うMA量の増加に発達勾配が加わっている、つまりMA差のようなものに伸びが左右されていると考えられよう。

日常記憶はRR₈（三度の繰返し）によって非常にきれいな発達差が生じている。練習効果が高学年ほど有効なことを示している（Fig. 5）。

伸びと発達との関係が殆んどみられないといつてよいのは数的検査で、数能力と共に関係把握や推理的な要素を含んでいると考えられるこういった検査では年齢に関係なく伸びはほぼ等しい。一方簡単な数を使つての推理力は学年が異っても学年相応にしか伸びないのは興味ある事実である。

Fig. 5 記憶下位検査の Retest 伸び量



単語完成のように語いと速度が中心になる (注) O₄においては各学年とも統計的に有意な差はない

ような検査では年令発達に伴う再テスト上昇が予期されていたのであるが、案外大きくはなかった。また、伸びの統計的に有意な逆転が単語マトリクスにおいてみられたが、この年令付近では近ごろ同一粗点における偏差値減少度がCA 1ケ年につき約3点であるが、5年生の再検査上昇比率がこれに追いつくのがやっとだったためである。これ以外の個所には逆転はみられていない。

下位検査種別による伸びの特徴発達要因について考えることをこの辺で留保して、つぎに下位検査の種々の特徴、形式の相違といったものからくる伸びの状態を観察してみよう。個人の知能が遺伝的要因による部分と環境的要因による部分とから成立っていると考えるのは現在においては常識であろう。ただし、その構成要素とか割合いとかについては、諸研究者によりまちまちであるが、一例として Burt, C が最近示した研究¹⁸⁾を挙げるならば、Tab. 11のような値となる。彼は mental test や scholastic test の成績を用いて相関係数を求め分析した結果、genetic な component は全体のほぼ77%を占め、environmental な要素が約17%、その他6%に大別されている。Burt 以外の研究者達の予測も大体これに近く、遺伝的素質が60~80%というところに落ちついているようである。

生理学的な考えを知能・脳の機能研究の領域に導入して縦横の解釈をおこなった Hebb¹⁹⁾ は、知能に生得的な知能と環境の影響に左右される知能、所謂A、B二種の知能を考えている。さらに²⁰⁾ 芋阪は集団知能検査における知能観にもこういった考え方を導入し、主として遺伝的知能を測定する下位検査と、環境的知能を測る検査とに分けられるのではないかといいことを提唱し、これより知能の構造や優秀児の知能特質を検討しようとしている。検査をどちらかに分類する方法をいくつか仮説的にとりあげているが、その一つにこの再検査における伸び量からの推測をあげている。

そこで、この検査において伸びの大きい下位検査と比較的にそうでない検査との状態を把握したいと思う。

群間の原検査における知能偏差値は個人偏差値で統制してある。そのため、各下位検査の偏差値は僅かではあるが変動がみられる。そこで、original test についての各検査間における偏差値の差を検定した。これは全体の資料について及び4年生の資料について実施した。両者は殆んど似たような傾向を示したが、全体の方がいくぶんテスト間変動が大なので、4年生の資料を用いることにする。再検査における下位検査間の伸びはテストにより若干の差があらわれた。

第一に気付くことは、Test 8 (数計算)の伸びが他検査に較べてあきらかに低い。この検査問題は方法の理解がそんなに困難なものではなく、また問題が覚えやすいものでも、解き方にコツ²¹⁾がいるものでもなく、迅速に的確な判断をなすことを要求される検査である。

18) Burt, C: The inheritance of mental ability, *Amer. Psychologist*, 1958, Jan. Vol. 13, No. 1 1—16

19) Hebb, D. O. The organization of behavior. 1949.

20) 芋阪良二他「京大SX知能検査報告(3)」発表内容 (1961. 10. 第3回日本教育心理学会総会 名大)

21) この検査においては、ある一定数 (たとえば20) を与えて、選択肢4つのうちから3つをえらんでその数を組立てさせる (たとえば8・6・9・5のうち6+9+5=20であるから8が不要となる)。不要な数8をさがさせる。

練習効果が小さいのは、上述のようなテストの性質からみて、テスト situation への馴れか
らくる効果が再検査においても変動することが小さいためと思われる。また、もう一つの数的検
査である数交換の伸びも他に比して小さいことから、数因子は test-statistic の影響が小さく、
生得的要素を比較的に測定しやすいものである可能性があろう。

他に、日常記憶の伸びも低い、original test の score 自体少々低いのでこの方法では何
ともいえない。別の分析が必要である。

その一例として、検査を3度繰返し (RRs) た場合の score を graph に示してみると、2
回 (R₄) や初回 (O₄) よりも伸びは更に大きくなっているし、上昇 score も他検査とあまり変
らない。

テスト2 (記憶) をさきにのべたような理由から一応 omit して、残りの下位検査を伸びの量
の状態から2つの group に分けることができる。伸びの大きい方の group は反対語・単語マ
トリクス、同図形発見、比較的に伸びの小さい方は数交換・数計算・単語完成・異質発見で、そ
れぞれの group 内同志では相互に差が小さい。

それぞれの group に属する各テスト共通の特徴を的確に言いあらわすことはむずかしいけれ
ども、大雑把に言って前者は問題数の比較的に多い (18~20)、各小問困難度がテストの前後を
通じほぼ一定であって、時間的要因が利きやすいものといえないであろうか。検査の繰返しは発
達による、また練習による test-set の smooth さ²²⁾ 上昇による速度増大を産みだすものと考
えられよう。

後者については数検査の伸びが小さいことを示している。数検査は他の推理系統諸検査のよう
に方法上そんなにむずかしいと考えられないのに伸びが小さいことから、比較的に数因子は検査
により生得的なものをとらえやすいと考えられよう。

信頼度係数と再検査法

いままでは再検査による偏差値上昇がテスト間隔や知能程度などの要因に影響される状態を特
定の知能検査を材料にとりあげて分析を試みてきた。本論文の研究目的はそこにある訳である
けれども、従来の諸研究者は test-retest reliability の研究に利用するのが殆ん
どである。そこで、一応相関 (test-retest) 分析にも目をむけてみることにしよう。

Kuder や Richardson が一回のテスト結果からの信頼度係数分析を試みる迄は、test-retest
法による信頼度係数は alternate-test の制作が困難なために専ら利用された。

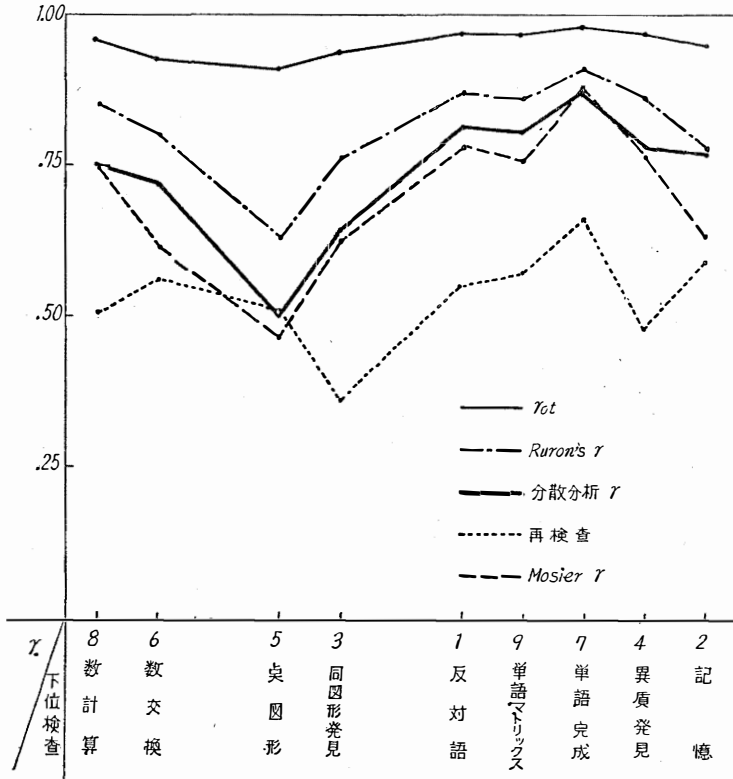
reliability 測定の根幹たる真分散をどのように推定するかについて各種の研究があいついでな
された。Guilford はこれを非常に手際よく整理してまとめている。²²⁾

Guilford はここで数個の例をあげて各方法の相対的な値の変動を示しているが、我々もこれ
を大ざっぱに追試する目的で、つぎのような整理をおこなった。

22) Guilford, J. P. Psychometric Methods. (2nd. Ed.) Chap.13,14

(被験者) さきの実験²³⁾に用いた資料のうち、発達による伸びの影響が小さかった4年生のO₁₂の被験者を取りあげた。さきの実験では他群との統制がなされているので、一応これを御破算にしてO₁₂に該当する全被験者より抽出しなおした男女各50, 計100名を用いた

Fig. 6 同一被験者の同一資料4年生(4ヶ月後男女各50)による reliability 係数



(方法) 各下位検査ごとに、主として粗点を用いて相関を求め集計した。相関の算出はすべて product-moment 法。求めた信頼度は r odd-even, Mosier の修正式 r , r total-odd, 分散分析, Rulon の公式, Flanagan の公式, Kuder-Richardson の公式[20][21], Tucker の修正K-R, などそれに r test-retest の各種である。

結果をながめてみると、以上の多くの方法はいくつかの傾向類似のパターンにわけられた。その主なものを図示したのが Fig. 6 で最も高い値を示したのは r odd-total (Mosierの修正前の値) であるのは当然かもしれないが、つぎに Rulon の r で代表されるいくつかの数式群(これには r odd-even, Flanagan の公式の r , などが属する)の値が高い。これはいずれも折半形式によって信頼度を測定した場合のもので、分散分析法による測定値に近い値の一群、つまり

23) 再テスト偏差値スコアの分析をさす。

24) K-R [20]や Tucker の修正 K-R がこれにあたるが、これに分散分析の r を加えた三者の係数は計算誤差の範囲内で一致する。

total なテストから信頼性を推定するものよりもどの下位検査においてもあきらかに高い。

r_{ot} に Mosier の修正を施した場合の値はほぼ全テストからの係数値に類似した。以上、全体傾向は Guilford の示した²⁵⁾例題と殆んど同じであった。

下位検査別にみると、Guilford が速度検査においては信頼度が若干高くなりやすいことをのべているが、正答率の多い検査がこの研究でも高くなっている。また、分散の小さい点図形や同図形発見の score が低い。この図よりテストの相違による変動よりも、あきら

かに方法による差異がつよいのは、分散に信頼度が左右されるためと思われる。

偏差値法により test-retest の相関値を無修正で算出したのは以下にのべる係数値の比較研究の参考に供するためのものである。この検査全体の $r_{test-retest}$ は .82~.85 であるが、下位検査別に算出すると低くなっている。

上述のように、test-retest の相関は同一 test の折半による係数より低いことがわかったし、方法の相違が下位検査の如何に不拘優先するから、係数の処理方法からの影響を除くために再検査相関をそのまま使うことにして、各条件ごとに r を求めたのが Tab. 13 である。

ここで、説明の都合上 O_4-R_4 の相関の値を (A) であらわすことにし、以下同様に²⁶⁾ B, C, D, E, とおけば、A-D の間に 5% で有意の差が出て、4ヶ月後再検査群の相関値がいくぶん高いことを示しているが、概して original-retest の係数値には間隔の差異による変動が極めて小さいことがわかった。相関の値は $A \geq E \geq D$ であるが統計的には有意となっていない。

下位検査相互の関係では、言語系統の検査

Tab. 10 算出に使用した資料の下位検査別結果

検査番号	問題内容	偏差値 平均	粗点平均	粗点分散	通過率
1	反対語	50.16	7.43	11.83	•378
2	記憶	48.09	4.77	7.44	•317
3	同図形発見	50.99	7.19	3.65	•400
4	異質発見	50.89	9.41	9.70	•470
5	点図形	48.93	3.55	2.21	•295
6	数交換	47.72	5.75	5.93	•383
7	語い	50.90	16.75	34.53	•570
8	数計算	50.76	8.94	6.82	•448
9	単語マト リックス	52.08	8.99	16.19	•428
※	S S	50.19			

Tab. 11

Analysis of Variance for Assessments
of Intelligence (By Burt. C.) ※

Source	Unadjusted test scores	Adjusted assessments
Genetic component :		
fixable	40.51	47.92
nonfixable	16.65	21.73
Assortative mating	19.90	17.91
Environment :		
systematic	10.60	1.43
random	5.91	5.77
Unreliability	6.43	5.24
Total	100.00	100.00

※ 18) 参 照

25) 各下位検査の正答率は Tab. 10 に記してある。

26) B: R_4-RR_8 C: O_4-RR_8 D: O_8-R_8 E: $O_{12}-R_{12}$.

が高く、数検査や空間検査が低い。これにはGulliksen²⁷⁾の言うような test の item 数などの影響と思われる。

Tab. 12 下位検査、条件別知能偏差値

学年	検査 条件	1	2	3	4	5	6	7	8	9
		反対語	記 憶	同図形発見	異質発見	点図形	数交換	語 い	数計算	単語マト リックス
3	R ₄	53.03	50.68	52.81	52.26	52.40	53.22	52.31	50.70	54.09
	RR ₈	55.31	50.90	55.52	53.56	59.11	53.40	53.31	54.67	56.25
	R ₈	53.55	51.59	52.27	52.79	52.74	51.91	51.97	52.81	54.33
	R ₁₂	51.69	50.89	50.32	50.90	51.12	50.47	50.99	51.26	52.71
4	R ₄	54.20	51.92	54.03	54.30	52.79	54.18	52.22	50.54	55.65
	RR ₈	57.31	55.61	57.58	54.66	58.54	54.86	53.88	52.61	58.37
	R ₈	52.33	48.76	53.89	51.94	53.94	52.59	52.45	50.83	54.29
	R ₁₂	51.91	49.76	51.51	52.49	51.84	50.92	51.05	52.05	52.73
5	R ₄	54.88	52.70	54.97	54.59	54.98	53.11	53.82	52.80	52.24
	RR ₈	56.61	56.03	59.10	55.94	58.56	53.77	55.12	53.77	53.67
	R ₈	54.40	49.88	55.06	53.69	55.12	51.98	54.32	53.80	56.58
	R ₁₂	56.02	51.57	55.21	55.50	56.31	53.35	54.85	52.76	53.68
6	R ₄	58.11	54.94	55.99	53.70	54.94	54.67	53.75	52.53	55.40
	RR ₈	60.79	58.25	61.06	55.35	60.40	55.83	55.69	53.63	58.10
	R ₈	54.69	49.54	55.66	55.76	56.23	53.24	54.48	52.09	55.61

また、発達年令段階と係数値との関係においても、偏差値平均の伸びの分析の際みられたようなはっきりした上昇勾配をみいだすことが出来なかった。これは全体にみても、下位検査別にながめても同様である。下位テスト種別では言語検査の相関信頼度が数検査・空間検査に比し低い結果があらわれている。

これらをまとめれば、再検査法による相関信頼度はテスト間隔、発達年令などの条件により統計的有意差を生じる程の変動はみられないといえる。故に、信頼度の変動はその算出の定義条件となる分散に依存するが、それはテストの長さ、内的整合性、テスト困難度などテスト構成上の要因の問題であって、それにはさきに検討して来たような test-retest 間における「伸び」の程度と直接には殆んど無関係であることがわかるであろう。さすれば、一般によくみられるように test-retest reliability を、そのテスト得点の変動しない度合、たしかさとして定義表現するのは、いささか不十分のそしりを免れぬといえよう。

27) Gulliksen, H. Theory of Mental tests. N. Y. John wiley, 1958.

京 都 大 学 教 育 学 部 紀 要 Ⅷ

Tab. 13 各 original, retest 偏差値の錯差横法による相関 (各cell N=100)

学年	検査 条件	1	2	3	4	5	6	7	8	9	SS
		反対語	記 憶	同図形 発 見	異質発見	点図形	数交換	語 い	数計算	言語マト リックス	
3	R ₄ ・O ₄	・503	・505	・471	・681	・511	・511	・766	・482	・715	・874
	R ₄ ・RR ₈	・675	・569	・453	・666	・612	・496	・690	・468	・747	・862
	RR ₈ ・O ₄	・524	・511	・423	・553	・474	・393	・713	・485	・715	・836
	R ₈ ・O ₈	・597	・589	・359	・701	・294	・441	・548	・625	・457	・790
	R ₁₂ ・O ₁₂	・503	・468	・444	・507	・454	・409	・581	・386	・518	・783
4	R ₄ ・O ₄	・731	・562	・454	・667	・647	・595	・703	・512	・463	・894
	R ₄ ・RR ₈	・754	・631	・631	・826	・636	・620	・713	・635	・664	・885
	RR ₈ ・O ₄	・690	・510	・479	・587	・661	・571	・767	・610	・425	・874
	R ₈ ・O ₈	・476	・635	・240	・591	・521	・311	・818	・526	・544	・769
	R ₁₂ ・O ₁₂	・657	・622	・270	・529	・474	・576	・685	・488	・590	・831
5	R ₄ ・O ₄	・633	・613	・401	・630	・371	・624	・807	・535	・602	・846
	R ₄ ・RR ₈	・717	・812	・548	・759	・673	・604	・867	・517	・687	・890
	RR ₈ ・O ₄	・559	・566	・431	・524	・317	・474	・742	・453	・565	・806
	R ₈ ・O ₈	・694	・545	・346	・460	・409	・511	・549	・455	・612	・742
	R ₁₂ ・O ₁₂	・679	・534	・305	・597	・424	・548	・729	・641	・607	・837
6	R ₄ ・O ₄	・714	・495	・354	・491	・672	・447	・705	・453	・500	・822
	R ₄ ・RR ₈	・770	・557	・401	・619	・766	・462	・665	・597	・775	・867
	RR ₈ ・O ₄	・699	・306	・376	・538	・621	・451	・741	・607	・469	・819
	R ₈ ・O ₈	・681	・485	・363	・651	・657	・553	・746	・627	・689	・875
全 学 年	R ₄ ・O ₄	・640	・510	・407	・610	・543	・556	・744	・490	・556	・849
	R ₄ ・RR ₈	・716	・627	・507	・725	・665	・559	・748	・533	・714	・878
	RR ₈ ・O ₄	・605	・428	・411	・540	・518	・464	・746	・523	・542	・816
	R ₈ ・O ₈	・649	・562	・317	・594	・491	・454	・657	・556	・564	・776
	R ₁₂ ・O ₁₂	・614	・536	・352	・523	・441	・494	・661	・517	・574	・803

reliability は、ある個人における test 得点が何回検査されたとしても、ほぼ同じ相対位置にある度合を示すものであり、繰返しの際に集団にかけられる影響因たとえばテスト間隔とか知能の高さからくる変動は係数に直接関与しない。極端な例を挙げるならば、再検査得点がいろいろの条件からの影響で原検査に比して10点ほど上昇することが知られているテストでも、相関は著しく高くなる可能性がある可能性は十分に存在する。

また、再テスト法により信頼度相関を求める際には、間隔を1年、8ヶ月、4ヶ月のいずれにとっても殆んど同じである。ただ、4ヶ月間隔の r がやや高めにでている。

単語完成のように項目数の多い下位検査では、テスト間隔による変化が僅かにみられる。

課 題

以上のように、再検査に関する総合的研究を報告したが、さらにこれらの知見を確認し推しすすめるために、次のような実験がおこなわれるべきと考える。

(1) 再テスト効果は2年後、3年後、さらに x 年後の間隔においてどうなるだろうか。これを検査の種類との関連においてとらえる必要があるだろう。

(2) 多くの下位テストについて、伸びの状態をしらべ類型化する。

(3) 年令段階によって再テストの伸びを喚起する要因は異ってくるか、異ってくるとすればどのようなか。

(4) 諸検査の再テストの程度を的確にあらわす公式を、たとえば係数といったようなもので、研究する。これらは、本研究の基礎にたち、なされるべきものと考えられる。

ま と め

再検査法により知能検査を実施した場合におこるテスト平均変動の状態とその要因を考察する実験をおこなった。

1. 再テスト間隔4ヶ月、8ヶ月、12ヶ月のいずれにも再テスト得点上昇がみられた。12ヶ月の伸びが最小である。

2. 伸びの生じる要因としてテスト間隔、知能の程度、年令発達、テストの種類などの強い影響をうける。性差や記憶力量差などは余り影響しない。

3. 数的な下位検査で測られた結果は、他テストよりも伸びが小である。これは数テストが環境要因からの影響をうけにくいということかもしれない。

4. 再テスト得点の上昇は問題そのものを記憶しているものでなく、テスト場面においての特定 situation を経験し、馴れるためと考えられる。

5. test-retest reliability は(2)でのべた諸種の要因にはあまり左右されない。また、偏差値の上昇度と係数値とは直接には関係が無い。

6. test-retest reliability は同一資料をテストの折半形式で求めた値よりも低い値をとる。

7. さらに2年後、5年後といった長期の、またいろいろの種類の下位検査について、伸びの研究がなされねばならない。