

答案返却の方法が学習成果に及ぼす影響の研究*

横浜国立大学

橋 本 重 治**

I 問題と目的

効果的な学習場面の構造の機能的一部として評価を考えるというのが、今日の教育評価での支配的傾向であるが(1)p.262—266, そのように考えれば、答案の採点や返却のような、テスト結果の処理の方法が重要問題としてとりあげられざるを得ない(2)p.87,(3)。本研究は、そのうちの答案返却の方法だけを問題とするのである。答案の返し方といつても、そこにはいろいろの方法が区別されうるわけであるが、実験を単純にするために、ここでは仮りに次のI, II, III, IVの4種の返却法をとることとし、その条件下においての影響を検した。

I——テストするだけで、後に全然答案を返却しないし、また口頭でもテストの結果について一言も言及しない。

II——正答には○, 誤答には×印をつけて採点した答案を単に返却するだけ。結果については一言も講評しないし、また生徒各自に検討するための時間も教師監督の下では与えない。

III——方法IIの場合と同じ要領で採点した答案を生徒に返却し、引つづいて数分間(場合によつて異なるが3分から5分位)の余裕を与えて、生徒各自で自分の答案のでき不できを自己検討するよう命ずる。教師の講評や矯正は全く加えない。

IV——上と同様に採点した答案を生徒に返却し、更に教師が一問一問毎に講評を加え、更に生徒多数の共通誤謬については簡単に矯正しておく。しかしそのために要する時間は、上記方法IIIの場合の自己検討のために与える時間と略々同じ長さとする。

この研究の目的とするところは、第一には、これら4つの返却法がその学習内容に関する学習効果に及ぼす影響にはどんな差異があるであろうか、どの返却法が最も

有利でどれが不利であろうか、ということを実験的に明らかにすることにある。第二の目的は、これら4つの返却法のアチーブメントに及ぼす効果は、生徒の能力の高低によつて相違するところがあるかどうかについても、判断の資料を求むることにある。第三には、——これは副次的目的ではあるが——、学習の評価に関する仮説の正否を検証するためのデータをも併せて求めようとの、理論的興味についての目的も附加された。

学習の評価としてテストを試みる場合、そのテストはどんな意味において生徒の学習に望ましい影響を与えるのであろうか。その根拠として理論的に仮定されうることとは次のようなことがらである、

(1) 生徒自身の立場からいつて、テストが学習指導において動機づけ効果(motivating effects)をもつことはC. C. Ross や Julian. C. Stanley の強調するところである(4)p.303—327。テストが、(a)それを受ける事前において、生徒により成績をとり或は逆に悪い成績をとることを回避しようとの一種の構え(attitude)をとらせることによつて、生徒の学習を動機づけ、(b)それを受けた後において、どこはでき、どこはできなかったかの、自分の学習の成功失敗を自己確認させることによつて、成功した内容についてはその学習を強化し、失敗した内容についてはその再学習を動機づけ*、(c)以上a, b二つの経験をしばしば反覆することによつて、平常の学習においての注意深かさや、不断に自己反省を加えて確実な学習をつづけるなどの好ましい一般的学習態度を形成するであろう**。

(2) 第二にテストする者、即ち教師の立場から、テストが、(a)教師をして学級全体並に個々の生徒のでき・不できを明らかにし、それに基づいて直接的に助長又は矯正の手段をとることを可能にし、(b)あるいは爾後の

* Effects of Various Methods of Returning Test-papers upon Achievement of Pupils.

** by Hashimoto, Juji (Yokokama National University)

* これにも、他律的に教師から結果を知らされる場合と、自発的に自己評価する場合との区別があろう(s)。

** しかし、テストがこのようにいつも生徒を積極的に動機づけるとは限らないことはいうまでもない。逆に消極的に動機づけることもあり得よう。

指導計画や指導の方法を改善・修正させて間接的に生徒の学習効果を高める*。

さて、これらの仮説が正しいとすれば、前述の4つの答案返却法の影響はどうなるであろうか。他はともかくとして、少なくとも上記仮説(1)の(b), , 即ち生徒の自己確認の効果については、返却法Ⅰは、答案を返却しないからあまり多くを期待できない(テストを受けた以上多少の自己確認は生じたはず)。これに対しⅡ, Ⅲ, Ⅳ, 殊にⅢとⅣは著しく有利であることが予想される。更に、上記仮説(2)の(a), 即ち教師の直接矯正の効果は、方法Ⅳでは期待できるが、他の方法では期待できない。こうして、他の条件がうまく統制されているならば、これら4つの返却法中、Ⅳが最も望ましい影響を学習成果に与え、Ⅲがこれにつぎ、逆にⅠが最も効果をあげ得ないはずである。かくして、テスト効果に関する仮説の検証ということも本研究のついでに目的とされた。

Ⅱ 実験手続

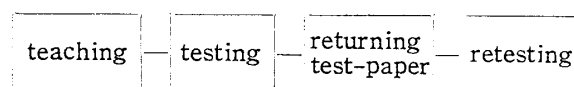
(1) 答案返却法が学習成果に及ぼす影響というのは、教室においての具体的な操作としては、或種の返却法が平常屢々繰り返されていることの総体的影響であると考えられるので、予備実験としてそのような実験計画を立てて実施してみたが、結果の反省の結果、その手続では所定の実験条件が攪乱され、あいまいにされることが知れたので**、長時間にわたる学習で連続的に何回もテストと答案返却を反覆する方法を止めて、単に1回のテストとそれに基づく1回だけの答案返却の影響をみるよう

* こういうことは、要するに、教育測定が教師に及ぼす motivating effects (動機づけ効果) であると C. C. Ross 等は述べている(4)p.305。

** 予備実験として、15時間単位の教材をとり、3時間毎にその時間の終りに10—15分の小テストを行い、これによつて前記4種の答案返却を試みた。従つて全部で5回の小テストとその答案返却を反覆したことになる。こういう手続で国語と算数の2科目につき2つの予備実験を試みたが、この場合、同一学級の児童を4分して4つの等質群をつくり、前述の4つの返却法を適用したので次のような欠点を生じた。すなわち、15時間の授業が連続した一単元をなしていたので、第Ⅳの返却法を適用する実験群の生徒の答案についての誤謬分析とその矯正を試みた教師の経験は、後の授業の際、しらずしらずのうちに他の返却法を適用する3つの被検者群にも影響して、所定の各実験群毎の実験条件が厳守されがたかつた。第2に、この実験が20日間もの長期に亘つて、5回ものテストと群毎に異なる答案返却を繰返したので、同じ学級の4つの被検者群の児童達の間互に競争・警戒の意識が発生し、それが所定の実験条件を乱した。

な手続をとるようにし、Fig 1 に示す実験プロセスに従つた。すなわち、一定時間を要する単元又は教材の指導後、その内容に関してテストし、その翌日答案返却し、その後2日の間隔をおいて3日目に再テストした。

Fig 1 実験のプロセス



(2) 4つの返却法を適用するところの4つの等質群の学習程度に影響を及ぼすと考えられる、教師の指導・生徒の学習・テスト条件などをできる限り齊一に保つことが要請される。そのためには別々の学級で4つの等質群をつくつてはまづい。そこで、1学級児童をa, b, c, dの4つの等質群にわけ、それぞれⅠ, Ⅱ, Ⅲ, Ⅳの返却法を適用することとした。更に、このような学級を数ヶ学級用意し——これらの学級はその教材・指導要領・テストなどすべての条件を同一にするようにした——、その各学級のa, b, c, d群をそれぞれ縦に合計して結果の信頼度を高めることをはかつた。この関係を表示すれば Table 1 のようになる。

Table 1 4つの等質群と返却法との関係

返却法 学級	Ⅰ	Ⅱ	Ⅲ	Ⅳ
1	a ₁	b ₁	c ₁	d ₁
2	a ₂	b ₂	c ₂	d ₂
3	a ₃	b ₃	c ₃	d ₃
全 体	Σ a	Σ b	Σ c	Σ d

(3) 4つの等質群は、Fig 1 に示した実験のプロセスにおいての最初のテストの成績によつてこれを作成した*。

(4) 答案返却の方法の詳細は、最初のテストの翌日の放課時、a群はそのまま帰宅させ(返却法Ⅰ)、b群には返却して直ちに帰宅を命じ(方法Ⅱ)、c群にはその次に返却して自己検討させた後直ちに帰宅させ(返却法Ⅲ)、最後にd群だけを残して返却法Ⅳを適用した。

以上のような実験手続により、学習内容と被検者をかえて次に述べるような5つの実験を試みた。

* 予備実験では、学力標準テストの成績なども等質群形に利用してみたが、必ずしも合理的ではなかつた。また、単に知能や学力のみでなく、生徒の性格・興味・家庭環境などの非知的(non-intellectual)な方面までも齊一にしなければならないことが反省されたが、こうなると等質群の形成ということにはなはだ困難なことである。

III 実験結果

第1実験

最初の実験は、昭和29年4月、鎌倉市第1小学校6年生の3ヶ学級で行われた。学習内容は算数科の割合に関する教材である。その実験過程は既述の通りであつて、3つの学級とも同歩調でその教材の4時間の取扱いに引つづいて直ちにテストし、その結果に基づいて4つの等質群を各学級毎に作り、その翌日、それぞれ既述のⅠ、Ⅱ、Ⅲ、Ⅳの返却法で答案を返却し、その後2日間を隔てて3日目に再テストを施行した。テストと再テストの間では算数の授業を行わないこととした。テストと再テストの問題は、その原理や困難度はおよそ同等にして、単に数量の大きさ・形式・提出順序を変えて、生徒に新しい問題であるとの印象をもたせることに努めた。

各学級並に全体について、a, b, c, d各群（従つてⅠ、Ⅱ、Ⅲ、Ⅳ各返却法）の被検者数N, 両テストでの平均得点M, 再テストの成績から始めのテストの成績を減じた差 diff. を示せば、Table 2 のようである。これは両テストとも100点満点で採点されたものである。

Table 2 第1実験結果（小6，算数）

学級	返却法	Ⅰ		Ⅱ		Ⅲ		Ⅳ	
	実験群	a		b		c		d	
	人数・得点	N	M	N	M	N	M	N	M
1	test	12	70.8	12	71.3	12	71.3	12	71.3
	retest	12	69.2	12	70.4	12	73.3	12	80.0
	diff		-1.6		-0.9		2.0		8.7
2	test	12	52.9	12	53.3	12	53.3	12	53.3
	retest	12	57.1	12	55.8	12	57.9	12	60.0
	diff		4.2		2.5		4.6		6.7
3	test	12	72.0	12	72.0	12	71.6	12	72.0
	retest	12	72.6	12	72.9	12	72.9	12	79.5
	diff		.6		.9		1.3		7.5
全体	test	36	65.3	36	65.6	36	65.4	36	65.6
	retest	36	66.3	36	66.4	36	68.1	36	73.2
	diff		1.0		.8		2.7		7.6

この結果を解釈するために2-3の統計的操作を加えてみよう。Table 2 に基づいて、各学級それぞれa, b, c, dの4群（従つてⅠ、Ⅱ、Ⅲ、Ⅳの4つの返却方法）の両テストにおける得点差(diff), 並にその方法別と学級別の得点差の合計を Table 3 の如くに表示し、この表にもとづいて要因分析をすれば Table 4 が得られる。

Table 3 両テストの得点差（第1実験）

返却法 学級	Ⅰ	Ⅱ	Ⅲ	Ⅳ	計
1	-1.6	-0.9	2.0	8.7	8.2
2	4.2	2.5	4.6	6.7	18.0
3	.6	.9	1.3	7.5	10.3
計	3.2	2.5	7.9	22.9	36.5

Table 4 要因分析（第1実験）

要 因	変 動	自 由 度	不偏分散	分散比	危険率
返却法による変動	90.08	4-1	30.03	10.18	1%
学級による変動	13.31	3-1	6.66	2.26	25%
誤差変動	17.70	(4-1)(3-1)	$2.95 = \epsilon^2$		
全 変 動	121.09	12-1			

$$* F_4^3(0.01)=9.739 \quad F_3^3(0.25)=1.762$$

Table 4 に示されるように、第1実験の結果の、4つの返却法による変動は1%の危険率で棄却できる。即ちこの4つの方法間のどこかには有意の差があることになる。そこで、そのどこに差があるかを危険率1%で検定すると、 $F_4^3(0.01)=13.74$ であるから

$$\frac{D^2}{2 \cdot 3 \epsilon^2} \geq 13.74$$

の公式によると、 $|D| \geq 15.59$ ならば1%で有意の差を認めてよい。すると

$$|I - II| = |3.2 - 2.5| = 0.7$$

$$|I - III| = |3.2 - 7.9| = 4.7$$

$$|I - IV| = |3.2 - 22.9| = 19.7^*$$

$$|II - III| = |2.5 - 7.9| = 5.4$$

$$|II - IV| = |2.5 - 22.9| = 20.4^*$$

$$|III - IV| = |7.9 - 22.9| = 15.0$$

であるから、* 印のところ即ちⅠとⅣ、ⅡとⅣの間には1%の有意差があることになる。更にⅢとⅣの差も大略有意差と認めてよいから、要するに返却法Ⅰ、Ⅱ、ⅢとⅣとの間に大体1%で有意差があると結論される。即ち答案を返却して簡単に矯正を加える方法が、他の3つの返却法よりも明らかに効果的であるといえる。これに対し、ⅠとⅡとⅢの間の差は有意ではない。

第2実験

この実験は、横浜市南吉田小学校4年生4ヶ学級の児童を用いて行つた。学習内容は漢字の書取りである。その実験手続は第1実験と同じであるが、ただ第1実験の場合のように実験に際して新しく教材を学習させ、それ

についてテストすることなく、既習教材をそのまま利用して、直ちにそのテストから出発した。いうまでもなくテストの問題は学級共通で、1字又は数個の漢字よりなる熟語20問とし、その同じものを、それが含まれている短文と、その提出順序を変えただけで再テストに使用し

Table 5 第2実験結果(小4書取)

学級	返却法	I		II		III		IV	
	実験群	a		b		c		d	
	人数・得点	N	M	N	M	N	M	N	M
1	test	13	15.8	13	15.7	13	15.8	12	16.2
	retest	13	15.3	13	16.2	13	16.8	12	17.9
	diff		-.5		.5		1.0		1.7
2	test	13	14.0	13	13.5	12	13.8	12	13.9
	retest	13	15.7	13	14.5	12	16.9	12	17.9
	diff		1.7		1.0		3.1		4.0
3	test	12	14.5	12	14.1	11	13.9	12	13.8
	retest	12	15.7	12	14.8	11	16.1	12	16.3
	diff		1.2		.7		2.2		2.5
4	test	12	15.7	12	16.2	12	16.3	12	16.2
	retest	12	16.0	12	17.9	12	18.2	12	18.3
	diff		.3		1.7		1.9		2.1
全体	test	50	15.0	50	14.9	48	15.0	48	15.0
	retest	50	15.7	50	15.9	48	17.0	48	17.6
	diff		.7		.9		2.0		2.6

Table 6 両テストの得点差(第2実験)

返却法	I	II	III	IV	計
1	-.5	.5	1.0	1.7	2.7
2	1.7	1.0	3.1	4.0	9.8
3	1.2	.7	2.2	2.5	6.6
4	.3	1.7	1.9	2.1	6.0
計	2.7	3.9	8.2	10.3	25.1

Fig 7 要因分析(第2実験)

要因	変動	自由度	不偏分散	分散比	危険率
返却法による変動	9.35	4-1	3.12	14.85	1%
学級による変動	7.02	4-1	2.34	12.00	1%
誤差変動	1.85	(4-1)(4-1)	0.21 = s^2		
全変動	18.22	16-1			

た。何れも20点満点による採点である。今回もテスト後、再テストする迄の2日間には漢字の指導を行うことを禁止した。その結果を表示すれば Table 5 の通りである。

この結果につき、上記第1実験の場合と同様の統計的処理を施してみよう。この表から、両テストの得点差だけを各返却法(従つて各実験群)について抜き出して示した Table 6 に基づいて、要因分析を試みれば Table 7 のようになる。

Table 7 についてみるに、返却法の相違による変動は1%の危険率で棄却される。即ちこの4つの返却法間には有意の差があることを示している。そこでどこに有意差が存するかを検定してみよう。 $F_{19}(0.01)=10.56$ であるから、 $\frac{D^2}{2 \cdot 4s^2} \geq 10.56$ とおくと、 $|D| \geq 4.11$ であれば1%水準で有意差を認めてよいことになる。すると、

$$|I-II|=1.4 \quad |I-III|=5.5^* \quad |I-IV|=7.6^* \\ |II-III|=4.1^* \quad |II-IV|=6.2^* \quad |III-IV|=2.1$$

となるから、これをまとめて結局、I・II; III・IVの間には1%の危険率で有意差があるといえる。即ち答案を返却して自己検討させる方法(III)、及び返却して矯正を加える方法(IV)は、他の無返却(I)及び単に返却するだけ(II)よりも、生徒の学習成果に対して有利に影響した。方法IとIIの間並にIIIとIVの間には有意の差は認めることができない。

第3実験

第3実験は茅ヶ崎市西浜小学校5年生の3学級により、昭和29年11月行つた。算数における名数と分数の問題につき、3学級歩調を合わせて5日間毎日1時間づつ学習させ、その結果について実験したものである。その実験手続や結果の処理の方法は、これまで述べた2つの実験とほとんど同じである。その結果を Table 8 に示す。採点は両テスト共に20点満点である。

本書の結果もこれまでと同様の統計的方法で吟味される。即ち、まず Table 9 のような得点差の表を作成し、これから Table 10 のような要因分析を試みる。返却法による変動は25%の危険率でなら棄却されうを示している。即ちこの4つの方法間には25%の危険率でなら有意の差を認めてよい。

そこで、この4つの方法間のどこに差があるかを吟味してみよう。 $F_{16}(0.25)=1.621$ であるから、 $\frac{D^2}{2 \cdot 3s^2} \geq 1.621$ の公式で、 $|D| \geq 2.82$ ならば25%水準の有意差であるが、

$$|I-II|=1.1 \quad |I-III|=3.6^* \quad |I-IV|=2.6 \\ |II-III|=4.7^* \quad |II-IV|=3.7^* \quad |III-IV|=1.0$$

Table 8 第3実験結果(小5, 算数)

学級	返却法	I		II		III		IV	
	実験群	a		b		c		d	
	人数・得点	N	M	N	M	N	M	N	M
1	test	10	10.4	10	10.2	10	10.4	11	9.5
	retest	10	9.5	10	8.7	10	10.0	11	9.2
	diff		-.9		-1.5		-.4		-.3
2	test	12	8.8	12	8.8	11	9.6	11	9.5
	retest	12	8.4	12	9.5	11	12.2	11	9.9
	diff		-.4		.7		2.6		.4
3	test	11	8.5	11	7.9	11	8.0	10	8.8
	retest	11	9.4	11	7.2	11	9.0	10	10.9
	diff		.9		-.7		1.0		2.1
全体	test	33	9.2	33	8.9	32	9.3	32	9.3
	retest	33	9.1	33	8.5	32	10.4	32	10.0
	diff		-.1		-.4		1.1		+.7

Table 9 両テストの得点差(第3実験)

返却法	I	II	III	IV	計
学級					
1	-.9	-1.5	-.4	-.3	-3.1
2	-.4	.7	2.6	.4	3.3
3	.9	-.7	1.0	2.1	3.3
計	-.4	-1.5	3.2	2.2	3.5

Table 10 要因分析(第3実験)

要因	変動	自由度	不偏分散	分散比	危険率
返却法による変動	4.81	4-1	1.60	1.95	25%
学級による変動	6.83	3-1	3.41	4.15	10%
誤差変動	4.93	(4-1)(3-1)	0.82 = ϵ^2		
全変動	16.57	12-1			

となり、この中 I と IV の差は僅かに不足するがこれをも認めて整理すると、I・II; III・IV の間に25%の危険率で有意差を認めることができる。その他の I と II, III と IV の間には有意の差は見出せなかった。従つてここでもまた、第2実験と同様、答案を返却して自己検討させる方法(III)と、返却して更に矯正する方法(IV)とは、他の2種の方法よりは効果的であるとの結論に到達する。

第4実験

ここでは小学4年3ケ学級の生徒を用い、教材は算数で、2位数×2位数、1位数を除数とする割算の計算問

題並に事実問題によつた。実験は昭和30年1月中に、茅ヶ崎市西浜小学校で行われた。実験手続はこれまでと全く同じである。テスト並に再テスト共に20点満点で採点し、その各群(方法)の人数・平均得点並に差は Table 11 に示す。

Table 11 第4実験結果(小4, 算数)

学級	返却法	I		II		III		IV	
	実験群	a		b		c		d	
	人数・得点	N	M	N	M	N	M	N	M
1	test	10	12.9	11	12.9	11	12.4	10	12.6
	retest	10	12.4	11	12.4	11	12.5	10	12.7
	diff		-.5		-.5		.1		.1
2	test	10	16.3	10	15.7	10	15.7	10	15.8
	retest	10	14.9	10	14.6	10	14.9	10	14.8
	diff		-1.4		-1.1		-.8		-1.0
3	test	11	12.5	11	12.3	11	11.7	10	12.3
	retest	11	12.3	11	11.4	11	12.3	10	13.3
	diff		-.2		-.9		.6		1.0
全体	test	31	13.8	32	13.6	32	13.2	30	13.6
	retest	31	13.2	32	12.7	32	13.2	30	13.6
	diff		-.6		-.9		0		0

返却法 I, II, III 及び IV のおのおのにおいて、テスト・再テストの得点差はそれぞれ -.6, -.9, 0, 0 であるから、絶対値の上では方法 I, II よりも III と IV の方が望まない影響を及ぼしたかのように見える。然し、これまでの実験に適用したと同様の統計的処理を施してみると、返却法による変動は 50 % でも棄却できない。即ちこの4つの方法間の差は有意差ではないことを示している。

第5実験

この実験は、上記第4実験と同時に同一の小学校で、6年3ケ学級を対象に実施した。漢字の書取り力につき、既習教材中から20語を選び、20点満点で採点した。他の実験手続はこれまでの場合と全く同様である。a, b, c, d の各実験群——従つて I, II, III, IV の各返却法における N・M, 並に, diff は Table 12 に示される。

そこで、ここでもまた、これまでの場合と同じ統計的吟味を加えてみなければならない。そのために、先ず Table 13 のような得点差表をつくり、これに基づいて要因分析を試みると Table 14 のようになる。

本表でみるように、この実験結果では、4つの返却法

Table 12 第5実験結果(小6, 書取り)

学級	返却法	I		II		III		IV	
	実験群	a		b		c		d	
	人数・得点	N	M	N	M	N	M	N	M
1	test	11	7.9	11	7.6	11	8.1	10	8.0
	retest	11	9.5	11	10.5	11	13.9	10	14.7
	diff		1.6		2.9		5.8		6.7
2	test	11	9.9	11	9.8	11	10.0	11	10.2
	retest	11	12.0	11	12.4	11	12.5	11	13.7
	diff		2.1		2.6		2.5		3.5
3	test	11	6.6	11	6.5	10	5.9	10	6.4
	retest	11	9.5	11	9.8	10	9.4	10	12.0
	diff		2.9		3.3		3.5		5.6
全体	test	33	8.2	33	8.0	32	8.1	31	8.3
	retest	33	10.4	33	10.9	32	12.0	31	13.5
	diff		2.2		2.9		3.9		5.2

Table 13 両テストの得点差(第5実験)

返却法	I	II	III	IV	計
学級					
1	1.6	2.9	5.8	6.7	17.0
2	2.1	2.6	2.5	3.5	10.7
3	2.9	3.3	3.5	5.6	15.3
計	6.6	8.8	11.8	15.8	43.0

Table 14 要因分析(I) (第5実験)

要因	変動	自由度	不偏分散	分散比	危険率
返却法による変動	15.88	4-1	5.29	4.64	10%
学級による変動	5.31	3-1	2.65	2.32	25%
誤差変動	6.81	(4-1)(3-1)	1.14 = ϵ^2		
全変動	28.00	12-1			

による変動は10%の危険水準で棄却することができる。そこでそのどこに差があるかを、一先ず危険率10%で検定してみると、 $F'_{0.10}(0.10)=3.78$ であるから、

$\frac{D^2}{2 \cdot 3\epsilon^2} \geq 3.78$ とおく。すると $|D| \geq 5.08$ であれば10%で有意差を認めてよいことになる。ところが

$$|I-II|=2.2 \quad |I-III|=5.2* \quad |I-IV|=9.2*$$

$$|II-III|=3.0 \quad |II-IV|=7.0* \quad |III-IV|=4.0$$

であるから、*印を附した場合のみ、即ち4種の返却法中、I; II; IV 並に I; IIIの間には10%の危険率で有意の差があることになる。IとII, IIとIII, IIIとIVの間

には10%の有意差は認め得ないが、ただ $|III-IV|=4.0$ であつて、10%の有意水準 $|D|=5.08$ に僅かに及ばないだけであるから、方法IIIとIVの比較については更にその差の精度を変量分析で検定してみた結果、Table 15に示すように25%の危険率でならば有意差であることが見出された*。

Table 15 要因分析(第5実験)

要因	変動	自由度	不偏分散	分散比	危険率
S III-IV	2.67	2-1	2.67	2.34	25%
S I-II	0.80	2-1	0.80	0.701	50%
S E	6.81	6	1.14		

実験のまとめ

以上5つの実験の結果をまとめて概括しよう。いま、各実験において、I, II, III, IVの4つの返却法間のどこに有意の差があつたか、その所在を一括してみると(括弧内は危険率)

第1実験—I, II, III; IV (1%)

第2実験—I, II; III, IV (1%)

第3実験—I, II; III, IV (25%)

第4実験—差の絶対値ではI, IIよりIII, IVが有利であつたが、然しそれは有意の差ではない。

第5実験— $\begin{cases} I, II; IV (10\%) \\ I; III (10\%) \\ III; IV (25\%) \end{cases}$

これを、一そう意味をくみとりやすくするために、分解して整理するとTable 16のようになる。表中の不等号は返却法の効果の大小を意味するものである。

すなわち、返却法IVが、I, II, IIIのすべてに勝っている場合が1回(危険率1%, 第1実験に於て)、方法I IIの両者にだけ勝っている場合が3回(第2, 第3, 第5実験においてそれぞれ1%, 25%, 10%の危険率で)、

* 以上5つの実験の外に、実はいま一つの実験を試みた。それは漢字の読みについての実験であつたが、問題が易しくて各群の最初のテストでの平均得点が、20点満点で既に18点内外に達し、頭打ちの状況であつた。これでは、各種の答案返却法の影響を受けて、再テストにおける各群の成績が変化する(この場合は各群とも上昇的に変化した)が)余地がほとんど残されていないことになり、この研究の目的からは不適切な実験ということで、資料から除外したことを附記しておきたい。ただ、この経験から、この種の実験計画について一つの教訓を学びとることができた。それは最初のテストの成績があまり良すぎないようにしなければならないということである。

Table 16 実験結果の整理

差の所在	危険率	実験番号
IV > I, II, III	0.01	1
IV > I, II	0.01	2
IV > I, II	0.25	3
IV > I, II	0.10	5
IV > III	0.25	5
III > I, II	0.01	2
III > I, II	0.25	3
III > I	0.10	5
I, II, III, IV	有意差なし	4

更に方法Ⅲだけに勝る場合が1回（危険率25%，第5実験）である。したがって全5回の実験で、返却法ⅣがⅠ，Ⅱに勝る場合が4回，Ⅲに勝る場合が2回である。その逆の，方法Ⅰ，Ⅱ，ⅢがⅣに勝る場合は1回もなかった。

次に，返却法Ⅲについていえば，それがⅠ，Ⅱに勝る場合が2回，Ⅰだけに勝る場合が1回であり，逆に方法Ⅰ，ⅡがⅢに勝る場合は1回もない。

更に5回の実験中1回だけは4種の方法間に統計上有意の差はみられなかったが，然し Table 11 で示したように両テストでの得点差の大小からいえば，方法Ⅲ，ⅣがⅠ，Ⅱに勝っていたのであるから，この結果は何等上述の4方法間の優劣関係と矛盾したり，それに逆らったりするものではない。

したがって，われわれは，この論文の初めに規定した4つの答案返却法が学習成果に及ぼす効果について，次のような推論を導き出すことができよう。

- (1) 返却法ⅢとⅣは，ⅠとⅡに比較して効果的である。この結論は最も確実性が高い。
- (2) 返却法ⅢとⅣについては，Ⅳがすぐれているとの相当の証拠がある。即ち答案を返却して自己検討させるよりは，返却して教師が講評し矯正する方がより効果的だということである。然しこの結論は十分確かではない。
- (3) 方法ⅠとⅡについてはその効果において差がない。すなわち答案を採点して単に返却するだけで何等の配慮も加えないのでは，答案を全く返却しないのと変りはない。

以上のような結論となる。更にこの結論は，本研究の目的として冒頭に掲げた処の，評価の基礎理論に関する仮説に対しても若干の証拠を提供している。すなわちそこで仮定したように，方法Ⅲ，ⅣがⅠ，Ⅱより効果的であるという上述の結論は，「評価は，自己の学習の成敗

を自己確認させて，成功した学習についてはそれを強化し，失敗した部分についてはその再学習を動機づける」との根本仮定を是認しているかのようであり，また方法Ⅳが最も優れているのは，この自己確認の効果の上に教師の直接矯正の効果が附加されているからという仮説を支持しているものと言えよう。然しそれはこれだけの証拠では十分の証明とはならない。

Ⅳ 生徒の能力との関係

1 問題と整理法

テストに関する方策が，生徒の能力の高低によつて異なるということを示す証拠がこれまでいくつかあげられている。例えば，S. L. Pressey は毎週テスト（Weekly test）の効果は能力の低い生徒に於て大きいということを見出し(4)p.312, J. E. Kirpatrick も頻繁にテストすることが知能の劣つた児童に対して最も有効であることを見出した(4)p.310。I. Paniasigui は，テストの結果を知らせることは能力の劣つた生徒よりも優秀な生徒に対してよりよき影響を与えることを見出した(4)p.315—316。

ここでは次の2つの問題を明らかにしてみたい。

問題Ⅰ テスト並に答案返却の諸方法の影響は，能力の高い生徒群（優群）と低い生徒群（劣群）との間で相違があるが。

問題Ⅱ このような優群と劣群における差異は，4つの答案返却法によつて異なるところがあるであろうか。換言すれば，4つの返却法中特に何れかの返却法が，優群あるいは劣群を著しく有利にしているということがあるか。

そのために，先ず，返却法Ⅰ，Ⅱ，Ⅲ，Ⅳのそれぞれを適用して実験した被検者について，その第1回のテストの成績によつて優・劣2群にわけ，その各群の再テストでの成績と第1回目のテストの成績との差（これを進歩量と名づけておく）を算出した*。

すると，返却法Ⅰ（ P_1 の符号で示す）について優群（ $P_{1.1}$ ）劣群（ $P_{1.2}$ ），返却法Ⅱ（ P_2 ）の優群（ $P_{2.1}$ ）劣群（ $P_{2.2}$ ），返却法Ⅲ（ P_3 ）の優群（ $P_{3.1}$ ）劣群（ $P_{3.2}$ ），返却法Ⅳ（ P_4 ）の優群（ $P_{4.1}$ ）劣群（ $P_{4.2}$ ）の8つの群についての進歩量が算出できる。そこで，次にこれら8つの進歩量の相互間の差に着目し，その有意性の検定を試みた。その結果を資料として上述の2つの問題に答えようとするのである。

* したがって，ここでいう能力とはこれらのテストにおいて発揮した生徒の学力のことであつて，知能テストによる知能ではない。

2 結 果

上述の視点から既に述べた5つの実験結果を再整理してみよう。先ず第1実験の結果は Table 17 のようである*。これにもとづいて要因分析を試みた結果、この各群の進歩量間の変動は25%の危険率で棄却できること、即ち $P_{1.1}$, $P_{1.2}$, $P_{2.1}$ …… $P_{4.2}$ の間には25%で有意の差があることが知られた**。そこで次に、この8つの群間のどこに有意の差があるかを検定してみなければならない。

Table 17 各群の進歩量 (第1実験)

返却法	I		II		III		IV	
群	$P_{1.1}$	$P_{1.2}$	$P_{2.1}$	$P_{2.2}$	$P_{3.1}$	$P_{3.2}$	$P_{4.1}$	$P_{4.2}$
人 数	18	18	18	18	18	18	18	18
進歩量の計	-50	80	-75	105	85	20	95	180
その平均	-2.8	4.4	-4.2	5.8	4.7	1.1	5.3	10.0

そこで各群の進歩量の差を Table 17 の進歩量の合計より、例えば、 $P_{1.2} - P_{1.1} = 80 - (-50) = 130$, $P_{2.1} - P_{1.1} = -75 - (-50) = -25$ ……というように計算して、一覧表に Table 18 のようにあらわしてみる。すると、この場合は $|D| \geq 103.50$ ならば 25 %の有意差を認め

Table 18 各群の進歩量の差*** (第1実験)

	$P_{1.1}$	$P_{1.2}$	$P_{2.1}$	$P_{2.2}$	$P_{3.1}$	$P_{3.2}$	$P_{4.1}$	$P_{4.2}$
$P_{1.1}$								
$P_{1.2}$	+130*							
$P_{2.1}$	-25	-155*						
$P_{2.2}$	+155*	+25	+180*					
$P_{3.1}$	+135*	+5	+160*	-20				
$P_{3.2}$	+70	-60	+95	-85	-65			
$P_{4.1}$	+145*	+15	+170*	-10	+10	+75		
$P_{4.2}$	+230*	+100	+255*	+75	+95	+160*	+85	

* この表は、要因分析をするためには、一そう詳細に1人1人の生徒の進歩量まで示しておく必要があるが、スペースの関係でそれを省略した。以下Table 19, 21, 20, 24についても同様である。

** これもスペースの関係で、結論のみを述べて、具体的な要因分析表を掲載することを割愛した。

*** 表中の正負の記号は左側欄に記された群を基準にみたものである。以下4つの同種の表においても同じ。

てよいから*、表中*印を附したものは全部25%の有意差ということになる。

同様に、第2実験の能力群別結果を示すと Table 19, 20 のようになる。

Table 19 各群の進歩量 (第2実験)

返却法	I		II		III		IV	
群	$P_{1.1}$	$P_{1.2}$	$P_{2.1}$	$P_{2.2}$	$P_{3.1}$	$P_{3.2}$	$P_{4.1}$	$P_{4.2}$
人 数	19	19	18	18	18	18	18	18
進歩量の計	9	21	8	19	21	57	16	82
その平均	0.5	1.1	0.5	1.1	1.1	3.2	0.9	4.6

Table 20 各群の進歩量の差 (第2実験)

	$P_{1.1}$	$P_{1.2}$	$P_{2.1}$	$P_{2.2}$	$P_{3.1}$	$P_{3.2}$	$P_{4.1}$	$P_{4.2}$
$P_{1.1}$								
$P_{1.2}$	+12*							
$P_{2.1}$	-1	-13*						
$P_{2.2}$	+10*	-2	+11*					
$P_{3.1}$	+12*	0	+13*	+2				
$P_{3.2}$	+48*	+36*	+49*	+38*	+36*			
$P_{4.1}$	+7	-5	+8	-3	-5	-41*		
$P_{4.2}$	+73*	+61*	+74*	+63*	+61*	+25*	+66*	

この場合は、 $|D| \geq 10.056$ なら1%の危険率で有意の差となるから(詳細は省略)、Table 20 中、*印のある差はどれも1%の有意差である。

同様に、第3実験については、各群の進歩量は Table 21 のようなり、これについて要因分析を行ってみたが、各群間の差は50%の危険率でしか棄却できなかった。即ちこの場合の各群の進歩量の間には有意の差は存在しない。

Table 21 各群の進歩量 (第3実験)

返却法	I		II		III		IV	
群	$P_{1.1}$	$P_{1.2}$	$P_{2.1}$	$P_{2.2}$	$P_{3.1}$	$P_{3.2}$	$P_{4.1}$	$P_{4.2}$
人 数	17	16	16	17	16	16	15	16
進歩量の計	-3.5	0	-3	-5.5	17	19.5	4.5	17
その平均	-0.2	0	-0.2	-0.3	1.1	1.2	0.3	1.1

更に同様に第4実験の結果をまとめると Table

* これも結果のみでその具体的な計算過程を掲げること省略する。

22, 23 のようになる。

Table 22 各群の進歩量 (第4実験)

返却法	I		II		III		IV	
群	P _{1.1}	P _{1.2}	P _{2.1}	P _{2.2}	P _{3.1}	P _{3.2}	P _{4.1}	P _{4.2}
人数	15	16	16	16	15	17	15	15
進歩量の計	-2.5	-9	-30	1	-16	16	-10	11
その平均	-1.6	-6	-1.9	0.1	-1.1	0.9	-7	0.7

Table 23 各群の進歩量の差 (第4実験)

	P _{1.1}	P _{1.2}	P _{2.1}	P _{2.2}	P _{3.1}	P _{3.2}	P _{4.1}	P _{4.2}
P _{1.1}								
P _{1.2}	+16							
P _{2.1}	-5	-21						
P _{2.2}	+26	+10	+31*					
P _{3.1}	+9	-7	+14	-17				
P _{3.2}	+41*	+25	+45*	+15	+32*			
P _{4.1}	+15	-1	+20	-11	+6	-26		
P _{4.2}	+36*	+20	+41*	+10	+27	-5	+21	

この場合は、 $|D| \geq 30.48$ なら5%の有意差を認めうるから (詳細は略), Table 23 中 *印のものだけが5%の危険率で有意の差といえることができる。

最後に、第5実験についてもこれまでと同様の方法によつて、各群の進歩量の差の有意性を吟味してみよう。

この場合は、 $|D| \geq 41.76$ であれば1%の有意差であり、また $|D| \geq 31.576$ であれば5%の危険率で有意差であるから (詳細は略), Table 25 中 *印のある差だけが1%又は5%の危険水準で有意の差といえる。

3 問題Iの検討

本節の冒頭に第1の問題として提案したところの「テスト並に答案返却の方法が学習成果に及ぼす影響は、能力の高い生徒群と劣った群との間に相違があるか？」に答えるために、これまで5つの実験について述べた整理結果から拾つて、4つの返却法毎に優群と劣群の進歩量の大小を比較し、且つその進歩量の差の有意性を Table 26 のように一括表示してみよう。

本表をみるに、すべてで $4 \times 5 = 20$ 回の優・劣群の比較において、優群の進歩量が劣群のそれを凌駕している場合は僅かに3回 (第1実験の P_{3.1}, 第3実験の P_{2.1} 及び第5実験の P_{2.1} であつて、何れも有意の差ではない) だけであつて、残る 17 回は劣群の方が進歩量が

Table 24 各群の進歩量 (第5実験)

返却法	I		II		III		IV	
群	P _{1.1}	P _{1.2}	P _{2.1}	P _{2.2}	P _{3.1}	P _{3.2}	P _{4.1}	P _{4.2}
人数	15	15	16	17	16	16	15	16
進歩量の計	35	37	49	48	63	64	64	98
その平均	2.3	2.5	3.1	2.8	3.9	4.0	4.3	6.1

Table 25 各群の進歩量の差 (第5実験)

	P _{1.1}	P _{1.2}	P _{2.1}	P _{2.2}	P _{3.1}	P _{3.2}	P _{4.1}	P _{4.2}
P _{1.1}								
P _{1.2}	+2							
P _{2.1}	+14	+12						
P _{2.2}	+13	+11	-1					
P _{3.1}	+28	+26	+14	+15				
P _{3.2}	+29	+27	+15	+16	+1			
P _{4.1}	+29	+27	+15	+16	+1	0		
P _{4.2}	+63*	+61*	+49*	+50*	+35*	+34*	+34*	

Table 26 能力別による進歩量の比較

実験	優群 — 劣群	差の有意性
第1実験	P _{1.1} < P _{1.2}	25%で有意差
	P _{2.1} < P _{2.2}	〃
	P _{3.1} > P _{3.2}	有意差ではない
	P _{4.1} < P _{4.2}	〃
第2実験	P _{1.1} < P _{1.2}	1%で有意差
	P _{2.1} < P _{2.2}	〃
	P _{3.1} < P _{3.2}	〃
	P _{4.1} < P _{4.2}	〃
第3実験	P _{1.1} < P _{1.2}	有意差ではない
	P _{2.1} > P _{2.2}	〃
	P _{3.1} < P _{3.2}	〃
	P _{4.1} < P _{4.2}	〃
第4実験	P _{1.1} < P _{1.2}	有意差ではない
	P _{2.1} < P _{2.2}	5%で有意差
	P _{3.1} < P _{3.2}	〃
	P _{4.1} < P _{4.2}	有意差ではない
第5実験	P _{1.1} < P _{1.2}	有意差ではない
	P _{2.1} > P _{2.2}	〃
	P _{3.1} < P _{3.2}	〃
	P _{4.1} < P _{4.2}	5%で有意差

きい。そしてこの 17 回中、4 回は 1% で、3 回は 5% で、2 回は 25% でそれぞれ有意差であつて、8 回だけは有意の差ではない。

このことは、一般的にいつて、テスト並に答案返却は、能力の高い生徒に対してよりも能力の低い生徒（できない生徒）に対してが、より大きな積極的効果を及ぼす傾向があることを示すものと解釈してよからう。この結果は、その実験条件は異なるけれども、前に引用した S. L. Pressey 並に J. E. Kirpatrick の実験結果と軌を同じくしている。

4 問題Ⅱの検討

一般的にいつて、テスト並に答案返却は、優群に対してよりも劣群に対してがより多くの効果を与えたのであるが、この傾向は答案返却法の種類によつて異なるところがあるであろうか。換言すれば、どの返却法が最もよく劣群の学習を有利にするであろうか。この問題を明らかにするために、前述の整理結果から、各実験毎に、どの群（方法）が最初のテストより再テストへと最も大きな進歩を示したか、その進歩量の大きいものから順に 3 位まで拾つて Table 27 に一括して示そう。

Table 27 進歩量の順位

実験	順位	群	差の有意性
第1実験	1	P _{4.2}	他の3群に対して25%で有意の差
	2	P _{2.2}	他の2群に対し25%で有意の差
	3	P _{4.1}	他の2群に対し25%で有意の差
第2実験	1	P _{4.2}	他の7群全部に対して1%で有意の差
	2	P _{3.2}	他の6群に対して1%で有意の差
	3	P _{2.2} P _{1.2} P _{3.1}	共に他の2群に対し1%で有意の差
第3実験	1	P _{3.2}	有意の差ではない
	2	P _{4.2} P _{3.1}	共に有意の差ではない
第4実験	1	P _{3.2}	他の3群に対し5%で有意の差
	2	P _{4.2}	他の2群に対し5%で有意の差
	3	P _{2.2}	他の1群に対し5%で有意の差
第5実験	1	P _{4.2}	他の4群には1%, 3群には5%で有意差
	2	P _{4.1}	有意の差ではない
	3	P _{3.2}	有意の差ではない

この表からみるように、比較的一貫して最も進歩量の顕著なものは、P_{4.2}、即ち返却法Ⅳを用いた場合の能力劣等群である。5つの実験で3回は第1位、2回は第2

位である。そして5つの実験で他の群と比較できる全機会 $7 \times 5 = 35$ 回中、11回は1%で有意差、5回は5%で有意差、3回は25%で有意の差である。

次に進歩量の顕著なものはP_{3.2}、即ち返却法Ⅲによる劣群であり、5つの実験中2回は第1位、2位の場合3位の場合がそれぞれ1回である。そして可能な比較 $7 \times 5 = 35$ 回中、1%で有意の差である場合が6回、5%で有意の場合が3回だけである。この他に、P_{2.2}、P_{4.1}、P_{2.1}なども比較的顕著な進歩をとげた場合もあるが、然し上の二つの群には遙かに及ばない。

以上の吟味から、返却法Ⅳは能力の劣つた生徒の学習に対して特に大きな効果を及ぼしているということ、逆にいえば、劣群の生徒に対して最も効果的な答案返却法はⅣであるということができよう。返却法Ⅲも亦、Ⅳについて、劣群の学習に有効な影響を与える方法である。

V 要 約

この研究は、答案返却の方法として、Ⅰ—無返却、Ⅱ—単に返却するのみ、Ⅲ—返却して自己検討、Ⅳ—返却して簡単な講評と矯正を加える、の4つを仮定し、それが学習成果に及ぼす効果並に他の2、3の関聯問題を追求することにあつた。その結果を要約すれば

- (1) 返却法Ⅲ、ⅣはⅠ、Ⅱに比してたしかに効果的だとの結論を得た。
- (2) 返却法ⅢとⅣとでは、Ⅳがすぐれた方法であると考えべき相当な証拠があつた。
- (3) 方法ⅠとⅡについては、何等有意の差を見出すことはできなかつた。
- (4) 以上の結論は、学習評価の価値に関する著者の基本仮定のいくつかに関係し、それを支持しているようである。
- (5) テスト並に答案返却は、一般的にみて、能力の高い生徒に対してよりも、能力の低い生徒に対してがより大きな積極的効果を及ぼした。
- (6) 全体を通して最も顕著な進歩を示したのは、返却法Ⅳを能力劣等群に適用した場合であつた。このことは、能力の劣つた生徒への返却法としてはⅣが最も効果的であることを意味するであろう。
- (7) 返却法Ⅲも亦、Ⅳについて劣等群の生徒の学習に対して効果的であつた。

この研究の実験計画並に結果の統計的処理に関しては、横浜国立大学助教授金井達蔵氏の絶大な御協力を得た。深甚な感謝の意をささげたい。

文 献

- (1) James L. Mursell, Successful Teaching, Its Psychological Principles, 1946
- (2) William J. Micheels and M. Ray Karnes, Measuring Educational Achievement, 1950
- (3) E. D. Curtis and G. G. Woods, "A Study of the Relative Teaching Values of Four Common Practice in Correcting Examination Papers", School Review, 37; 615—623. October, 1929 (C.

C. Ross, Measurement in Today's schools, 1949 による).

- (4) C. C. Ross and Julian C. Stanley, Measurement in Today's School, 1954
- (5) J. M. Mason and G. W. Angell, "An Experiment in Evaluation in Biological Science", The Journal of Educational Psychology, Vol. 44 No. 5 (May, 1953), 296—304

—1955.9.7 受稿—