

MMPIの回答に及ぼす態度の影響と妥当性尺度*

防衛大学校

岩 脇 三 良**

I 序

目録法形式の検査法では、言語刺激に対する被検者の内省が基礎となつているところから、回答が歪みやすいという批判がある。被検者がどのように動機づけられているかによつて、ごまかしがきき、検査得点変動するというのである。しかし、実際に、どの程度まで、故意に得点を歪めることができるものであろうか。また、その歪みは臨床尺度の得点を妥当性のないものにしてしまうのであろうか。本研究では、被検者に一定の態度をとらせた場合、MMP Iの臨床尺度得点がどのような変動を示すかを検討することをもつて、第1の目的とした。

MMP Iでは、質問紙法のもつ弱点に留意し、回答の歪みを検出し、さらに、その歪みのある程度修正できるように工夫されている。Pichot(14)は回答の信頼度を高めるのに必要な4つの原理をあげているが、これはだいたい、MMP Iを中心に考えられたもので、MMP Iのすぐれた着想がうかがえる。しかし、L尺度には適用の限界があり、F尺度、K尺度も、Comrey(2, 3)が因子分析法によつて示しているように、多義的な内容をもっている。また、筆者(12)がすでに示唆したように、L、F、Kの3尺度だけを利用するのでは不十分のように思われる。さらにMMP Iでは自己防衛的態度を修正するためにK尺度が利用されているが、はたして、これで十分なものであろうか。検査を受ける態度が妥当性尺度あるいは指標にどのように反映するかを検討するのが、本研究の第2の目的である。

II 対象と方法

検査対象は防衛大学校学生で、教養課程として心理学概論を聴講しているもの227名である。ただし、検査は2回にわけて行なわれたため、2回とも検査を受けた214名

の資料を検討することにした。被検者は3群にわかれ、これに対しMMP I東大改訂版(集団用)が昭和36年4月中旬と、2週間のインターバルをおいて、同月下旬に実施された。

第1回目の検査では、全員、普通の教示に従つて、回答した。検査結果は採点済みしだい、各人に返すことを約束し、自分の性格を知るためには、正直に、まじめに回答しなければならないこと、読みとばしのないよう注意すること、つとめて、ハイ・イエのどちらかへ印をつけることが特に強調された。

第2回目の検査では、F群(71名)には、できるだけ好ましい印象を与えるように、U群(69名)にはできるだけ好ましくない印象を与えるように、N群(74名)には、第1回と同じ条件で、回答させた。この場合の教示を具体的に示すと、次のようになる。

F群に対しては「ある学生から、このような性格検査であれば、自分を良い人物にみせようとするれば、望ましい得点が得られるのではないかという質問をうけた。はたして、そうなるものか調べてみたらよいと思う。たとえば就職試験を受けにいったところ、そこでは、このMMP Iの結果から、すぐれた人物だけを採用することがわかった。その職場にはぜひとも入りたい。このような状況に直面したと想定して、試験官にできるだけ良い印象を与えるよう回答していつてください。」

U群に対する教示。「ある学生から、このような性格検査であれば、自分の思うままの人間像をつくりあげることができるのではないかと質問をうけた。はたして、そうなるものか調べてみたらよいと思う。非常につらくいやな勤務に任命された。しかし、幸いなことに、精神的に異常であることがわかれば、療養を命ぜられ、2度とその勤務にはつけないことがわかった。このような状況に直面したと想定して、検査者に好ましくない人間像をもっていると印象づけるように回答していつてください。」

N群に対しては「先日行なつた性格検査の結果をまとめ、諸君の一般的傾向を示し、各人に、採点結果を返そ

* Fakability and validity indicator in the MMPI.

** by Saburo Iwawaki (Defense Academy)

うと思つたが、なかに、どうも、いい加減に答えているものがいるようで、一般的傾向を発表することができない。こういう検査には、正答、誤答というものではなく、正直に、ありのままに答えることが、自分の性格像を知るためにも大切なことである。2度と、このように再検査しないように。前回、まじめに答えたものには気の毒であるが、もう一度、読みとばしのないように注意して回答してください。」

いずれの場合も、検査に際しては、特別時間制限を設けず、全問、回答したものから順次退場させた。

III 結果および考察

普通の教示のもとに実施された第1回目のMMP Iの各尺度の粗点平均値（K修正を加えてある）をグループ別に表示すると、Table 1 のようになる。各尺度とも、グループ間には統計的に有意の差はなかつた。214 名全員同一母集団に属する標本と考えられる。

つぎに、前記のF、U、N群に対し2週間後、それぞれ異なる教示のもとに実施されたMMP Iの平均得点（K修正）のプロフィールを Fig 1 にしめす。

妥当性尺度については、後で取り上げることにして、

Table 1 普通の教示によるMMP I 平均得点
(K修正)

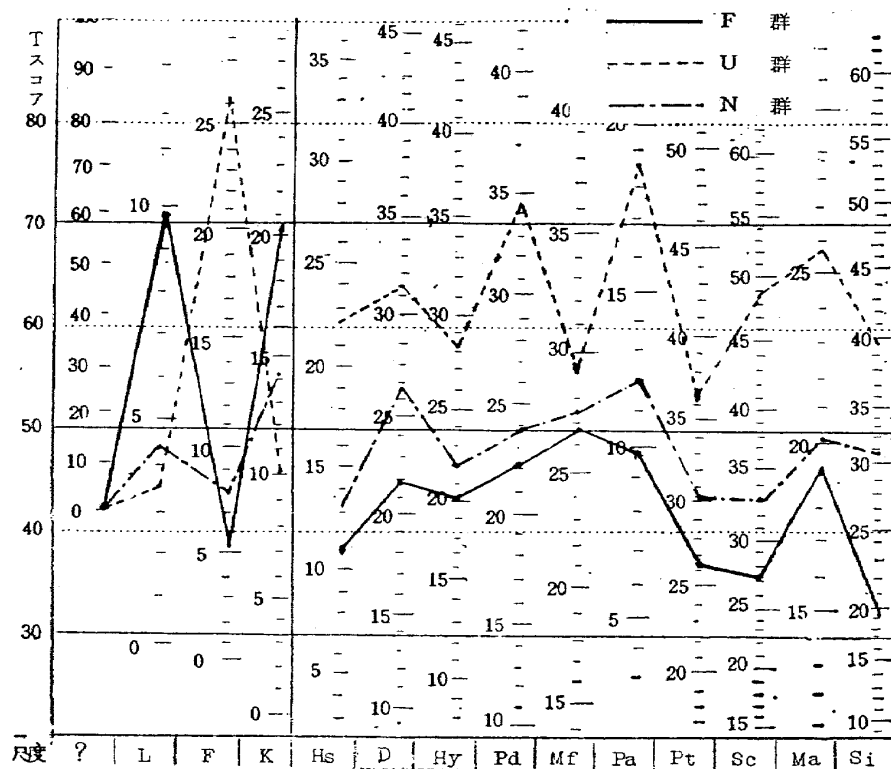
群別 人数	F 群 71	U 群 69	N 群 74
尺度			
?	5.34	6.02	5.79
L	4.32	4.41	4.36
F	8.79	8.26	8.08
K	14.50	13.96	14.14
Hs	14.38	13.57	13.89
D	27.67	26.73	26.33
Hy	21.63	22.44	23.54
Pd	23.64	22.84	23.32
Mf	28.46	28.33	28.38
Pa	11.88	11.75	11.93
Pt	31.63	30.40	29.62
Sc	33.67	32.45	33.00
Ma	20.61	19.64	19.98
Si	31.60	32.26	32.40

ここでは臨床尺度についてのみ検討してみよう。Fig 1 において、正直に答えることを強調したN群を対照群と

して、他の2群の得点平均を比較してみると、F群は全般的にN群よりも低く、U群は全般的にN群よりも高い。Si尺度を除き、U群とN群の差の方がF群とN群の差より大きい。Tスコアにして、10点以上の差があるのはF群ではSi尺度だけであるのに反し、U群ではMf尺度を除き、だいたいTスコア10点ないし、それ以上の差を示している。Fig.1 のプロフィールはK修正された結果であるが、U群のような態度で回答した場合には、K修正は有効に働いていないように思われる。

しかし、対照群と考えたN群が第2回目の教示により、実際にはF群に近い態度で回答しているかもしれないと思われる。そこで普通の教示により行なわれた第1回目のときの平均に対し、3群おのおのが第2回目のMMPI粗点にどれだけ有意な差を生じたかを比較検討してみた (Table 2)。すなわ

Fig. 1 態度の差によるMMP I の比較



ちK修正の効果を知るために、K修正された尺度については、第1回目と第2回目における3群のK修正前の粗点平均を示し、両回の差を比較し、Table 3 に示した。

Table 2, および Table 3 によると、N群では各尺度に有意の差はみられなかつた。ところが、第2回目のとき、一定の態度をもつて回答したF群とU群においてはK修正されていない場合、Mf 尺度を除き、統計的に有意な得点差をみせている。ここで注目されるのはMf 尺度が受検態度に影響されないことである。Mf 尺度は興味は男性的であるか、女性的であるかを測定し、同性愛

Table 2 第1回と第2回の間の態度の差による平均得点差 (K修正)

	F 群		U 群		N 群	
	差	t	差	t	差	t
Hs	-3.28	**	8.57	**	-0.48	0.25
D	-5.80	**	4.81	**	0.52	0.45
Hy	-1.63	*	4.73	**	-0.92	0.53
Pa	-1.30	1.08	11.67	**	0.37	0.50
Mf	0.98	1.69	1.00	1.23	-0.56	0.89
Pa	-2.14	**	7.01	**	0.07	0.18
Pt	-5.13	**	6.47	**	0.51	0.75
Sc	-6.58	**	16.20	**	-0.04	0.11
Ma	-1.31	1.52	6.17	**	0.27	0.52
Si	-12.51	**	7.62	**	-1.56	0.53

- (注) 1) **は1%レベルで有意 (対応ある差の検定による)
 2) *は5%レベルで有意 (//)
 3) 一値は第1回目の時より第2回目の方が平均値が低くなつたことを示す

的症狀を診断するために作成されたものであるが、各項目はアメリカ文化を中心に考えてつくられたものであり原版をそのまま翻訳してMf 尺度として、日本で利用するには疑問の点がある。矢田部—ギルフォード性格検査ではMf 尺度ははぶかれているが、わが国で性格検査を標準化する場合、この尺度を臨床尺度として採用すべきかどうか、また採用するにしても、どのように利用し、どのように解釈するかについては十分検討を要するものと考えられる。われわれの研究で、Mf 尺度だけが有意の差がなかつたことは、この尺度だけごまかしがきかないように巧く作文されているのか、それとも臨床尺度として、十分な弁別力をもっていないことを示すものかはつきした結論を出すわけにはいかないが、いずれはせよMf 尺度の扱いについてある問題を提供することになるのではあるまいか。少なくとも、わが国においてはMMP I のMf 尺度について再吟味する必要があるように思われる。

Mf 尺度を除き、K修正がないと各尺度とも、態度によつて有意の差が生ずることは、他のパーソナリティ・インベントリに比して、作為的な回答が困難とされているMMP I においても、ごまかしがきくことを物語るものであろう。K修正は態度の役割りを除去し、各臨床尺度の弁別力を増すために実施されているのであるが、各尺度によつて、態度の影響を受ける度合に差があるので、K尺度により修正値が異なつている。われわれの資料から直ちに、各尺度の修正値をいくりに訂正すべきであるというわけにはいかないが、少なくともSi 尺度にはK修正の必要があるように思われる。また、K修正を加えても、態度により有意差のなかつたのはPdとMaの両尺度だけ、しかも、よい印象を与えようとしたF群においてみられただけである。F群では、K修正した場合、得点差が少なくなつているから、自己防衛的態度を表わす

Table 3 態度の差による平均得点 (K修正なし) の比較

	F 群				U 群				N 群			
	第1回	第2回	差	t	第1回	第2回	差	t	第1回	第2回	差	t
Hs	7.08	1.08	-6.00	**	6.89	17.27	10.38	**	6.83	6.11	-0.72	0.78
Pd	17.95	14.33	-3.62	**	17.26	30.42	13.16	**	17.69	18.37	0.68	0.94
Pt	17.64	6.46	-11.18	**	16.44	26.83	10.39	**	15.78	15.93	0.15	0.50
Sc	19.03	7.78	-11.25	**	18.68	38.61	19.93	**	18.85	18.76	-0.09	0.45
Ma	17.71	15.22	-2.49	*	17.20	23.81	6.61	**	17.18	18.41	1.23	1.08

- (注) 1) **は1%レベルで有意 (対応ある差の検定)
 2) *は5%レベルで有意 (//)
 3) 一値は第1回目より第2回目の方が平均値が低かつたことを示す

K尺度による修正の効果はある程度認めなければならないが、また完全なものとはいえない。一方、悪い印象をあたえようとした群では、K修正により、幾分差はせばまっているが、F群に比し、はるかに大きい得点差を示している。隠蔽的な意味をもったK尺度による修正ではU群のような態度で回答されると、K修正の効果は不完全のようである。しかもPd、Pa両尺度ではTスコアにして70点以上の平均得点になり、正常者でも悪い得点を得ようとするれば、異常と診断される可能性がある。われわれの資料を防大生以外に一般化するわけにはいかないかもしれないが、要するに、ごまかしがきくのである。

N群の資料に基づき、再検査法によつて、信頼度係数を算出した結果 (Table 4) によると、まじめに答えれば、MMP Iは相当信頼度の高い検査であるといえる。それが好ましい方向、あるいはその逆の方向へ歪めようと企てると、その方向へ、だいたい、有意の差をもつて変質できるのである。少なくとも、東大改訂版MMP Iでは、たとえK修正をしても、それが実現可能であるといえよう。すると、この回答の歪み程度を検出でき

Table 4 MMP Iの信頼度係数 (K修正)

	Hs	D	Hy	Pa	Mf	Pa	Pt	Sc	Ma	Si	平均
r	.83	.82	.80	.84	.82	.88	.82	.89	.86	.78	.83

る手段が要求されてくる。そのために、MMP Iでは、いくつかの妥当性尺度あるいは指標を利用できるようになっている。回答する際の態度がそれらの尺度なり、指標なりにどのような影響を及ぼし、どれだけの検出力があるかを次に検討してみたいと思う。

？点

？点数については、前論文⁽¹²⁾でも述べたように、回答を歪めようとする態度を表わすかどうかは疑問であるが、本研究では次のような結果を得た。第1回目のときには、？が皆無というものの214名中、104名であり、最大？数は46であつたのに反し、第2回目の実施では最大？数は4で、214名中、206名までが？なしという状態であつた。他の尺度では差のなかつたN群においても、第1回のとき、？数がゼロでなかつたものが約半数であつたのに、第2回目では3名に減っている。他の2群でも同様の傾向を示したところから、？点数は回答を歪めようとする態度の反映したものであると考えることはむしろかくなる。むしろ、教示によつて、？数が多くなることは防げるものと思われる。しかし、無理矢理「ハイ」「イエエ」のどちらかへ回答することは、かえつて、真実の回答から遠ざけることになりはしまいかという問題が生ずるであろう。われわれの資料では、N群において、他

の尺度で有意の差がなかつたから、少なくとも、？数46が最大であるような程度では大して影響はないものと考えられるが、はつきりした結論はのべられない。ただ、いえることは、回答を歪めようと積極的な態度をとると歪める方向のいかんにかかわらず、？数は少なくなるということである。

他の妥当性尺度

MMP Iでは？点の他にL、F、K尺度が回答の信頼度を検出するために、作成されているが、次に、この尺度について検討してみよう。Fig. 1によると、L尺度ではF群が、F尺度ではU群が、K尺度ではF群が、著しく高い平均得点を示している。臨床尺度の例になつて、この3尺度が3群でどのような変動をみせたか、Table 5に表示する。Table 5の差は第1回目と第2回目の平均粗点である。

Table 5 態度の差による妥当性尺度の比較

	F 群		U 群		N 群	
	差	t	差	t	差	t
L	5.55	** 11.33	-0.86	0.83	-0.03	0.46
F	-3.28	* 2.46	18.15	** 9.55	-0.12	0.28
K	5.54	** 6.44	-3.92	* 2.53	-1.02	0.76

(注) 1) **は1%レベルで有意 (対応ある差の検定)

2) *は5%レベルで有意 (//)

3) 一値は第1回目の時より第2回目の方が平均値が低いことを示す

これによると、N群では妥当性尺度はいずれも有意差がないことがわかる。L、F、Kの順序にF群とU群を比較してみよう。

L尺度

F群では有意に点数が高くなつていのに反し、U群では有意の差はみられなかつた。すると、U尺度はよく見せようとする場合には普通の態度で回答した時より有意に得点が高くなることはわかるが、悪い印象を与えようとするときには、ほとんど変動がないといえるようである。Tスコアが70以上になるL尺度は全回答が疑問視される。第1回目のMMP Iの結果では214名中、4名(1.9%)がTスコア70以上になつた。ところが、第2回の結果では、次のようになつた (Table 6)。U群とN群ではLがTスコア70以上になることは皆無であつたのに反し、よい印象を与えようとしたF群では、60.3%がTスコア70以上になつている。しかし、39.7%はTスコア70以下であり、N群のL>70がゼロであることを加味し

て、19.85%はこの分類では誤りをおかすことになる。

Table 6 L尺度の比較

L	F 群	U 群	N 群
≤70	28	69	74
>70	43	0	0

Delay, Pichot et Perse(4)によると、この分類による誤りは13.75%となつていますが、これは男女を合計しての結果である。かれらは算出していないが、男性の結果だけをとりあげて再計算してみると、18.50%となり、われわれの資料に接近する。以上から、L尺度はよくみせようとする態度をある程度、検出できる尺度のひとつであるといえよう。L尺度はうわべを飾ろうとすると、高くなるとされているが、検査に慣れた被検者には有効でないといわれている。われわれの被検者は少なくとも第1回目の検査を経験しており、L尺度に警戒した被検者がいたのかもしれない。

F尺度

Tabel 5によると、U群においてF得点が著しく高まるのがわかる。F尺度は、元来、普通ごくまれにしか回答されない問題からできており、まれにしか回答されない問題に回答することは好ましくない自己記述を表わすのであるから、悪い印象をあたえようとするときには当然得点は高まるであろう。このF尺度は、そのほかに不注意な回答、質問の誤解などによつても高まるといわれており(11)、U群の得点が高いのはそういった要素が含まれているかもしれない。しかし、われわれの被検者はすでに普通の教示でMMP Iに回答しており、その場合は第2回目のF得点に比し、はるかに低くかつたのであるし、N群では有意差なく、F群では逆に5%水準で低くなつてるところから、U群の第2回目の回答で得点が著しく高くなつてゐるのは、その態度によるものであると考えたい。また、F群でF得点が低くなつてゐる事実はF得点が低いからといつて、ただちに、回答が信用できるものであると判定するわけにはいかないようである。F尺度の項目を因子分析した Comrey(3)がF尺度は臨床尺度と大差ないと結論しているように、病的症状を示す被検者ではF得点が高まるといわれており(13)、F得点が高いからといつて、必ずしも回答が信頼できないとはいえない。以上のように多くの原因からF得点は高まるものであり、F得点の高い被検者の回答の解釈には慎重でなければならない。当研究からは悪い印象をあたえようとする、F得点は著しく高まり、よくみせようとする、幾分低くなる事実、すなわち、F得点は態度によつても変動するものであるということが示された

Delay たち(4)はFのTスコアが臨床尺度のTスコアの最高のものよりもさらに高いときには、その回答は信用できないという。Fig. 1によると、U群のプロフィールはFが最高になつてゐる。当研究では、正常者群だけを対象とし、精神障害者群が含まれていないので、病状があるためにFが高くなつてゐる被検者のプロフィールと比較できる資料が手もとにないのは残念である。Delay たち(4)によると、まじめに回答した患者ではFのTスコアは最高にならないというが、その資料は提出されていない。標準化の違うと思われる山田(16)の資料によると、平均化されたプロフィールではFは臨床尺度より高くなつてゐない。もし、正直に回答した患者では一般に山田の示すようなプロフィールになるとすれば、FのTスコアで臨床尺度より高くなつてゐる事例は質問の理解不足、いい加減な回答あるいは悪い印象をあたえようとする態度を表わすものであるといえよう。患者群の資料がなく片手落ちであるが、われわれの資料から、態度により、このFのTスコアの高さがどのように異なつてくるかを示す。筆者(12)は前の論文でF-Kプラス群でFのTスコアがプロフィールで最高になつたものは、

Table 7 F尺度のTスコアがプロフィール中最高である事例数

	F 群	U 群	N 群
被 検 者 数	71	69	74
事 例 数	0	58	0
%	0	84.1	0

%は各群の被検者数を母数とした

43.5%であることを示したが、回答を好ましくない方向に歪めようとした場合にはその倍になるようである。前回報告した資料は普通の教示で実施された適性検査の一環としてのMMP Iの結果を分析したものであり、またF-K値の臨界点に問題があるためこのような差を生じたものと考えられる。Delay たち(4)のいうように、F得点だけよりはプロフィールでFのTスコアが最高になるか否かをみる方が回答の妥当性を云々する場合には有効のように思われる。ただ、患者群のプロフィールがないため、症状によるFの高さと比較できない点が残された問題となろう。

K尺度

Fig. 1およびTable 5によると、F群ではK点が著しく高まり、U群では低くなることがわかる。しかし、K修正が必ずしも臨床尺度の得点を十分に修正していないことは前述のとおりである。Tスコア70以上および30以下の事例数をTable 8に示す。F群では60.6%がTス

Table 8 K尺度スコアによる比較

Tスコア	F 群	U 群	N 群
≥70	43	2	2
≤30	0	0	0

コア70より高くなっている。一方、K点のTスコア30以下のものは各群において皆無である。K得点は態度のいかにかわらず、極端に低くはならないようである。U群でK点のスコアが70以上となつていものが2名いるが、2名とも、F得点はさらに高くなっている。

K尺度については、多くの解釈がなされており、K修正の有用性についても賛否両様の資料が提出されておりK尺度自身について総括的な研究が望まれる。最近ではComrey(2)が因子分析により、K尺度から8つの因子を抽出し、修正因子として、K尺度の使用は望ましくないというし、Smith(15)は自己防衛的なものは洞察力が低いものであるから、正常者の自己防衛性を測定するものさしとして、K尺度を用いるべきではないという。しかし、一方において、Hanley(10)は社会的望ましさによつて応答する構え(set)を測定するのには、K尺度の妥当性が高いというし、Fricke(6)はそのような構えはすてにより適応を示していることになると結論している。われわれの資料では決定的なことはいえないが、K修正だけでは不十分、特にU群のような態度で回答されたMMP Iには満足な修正はできないことが示唆できよう。Fordyce(8)はK尺度を中心に社会的望ましき(S-D)尺度を構成し、検査をうける態度は社会的に望ましい項目に回答するレディネスの有無に依存するといっているが、K尺度については、各項目につき、内容的に項目分析をして再検討する必要がある。

Hs尺度

微妙な内容をもつ質問よりははつきりとした質問の方が回答が歪みやすい。Hanley(10)は防衛度を示す項目の得点とはつきりした内容の質問得点とは負の相関を、微妙なものには正の相関があることを示した。この意味では、Hsに関する質問の内容は非常にはつきりとしており、Hs尺度が好ましい方向へ歪めようとする態度を検出する一指標になりうると考えられる。その可能性は、すでにDelayたち(4)および筆者(12)により示唆されているが、当研究のF群で、どのようになるであろうか。Table 9に、Hs粗点(K修正なし)の分布をF群とN群

Table 9 Hs粗点の比較

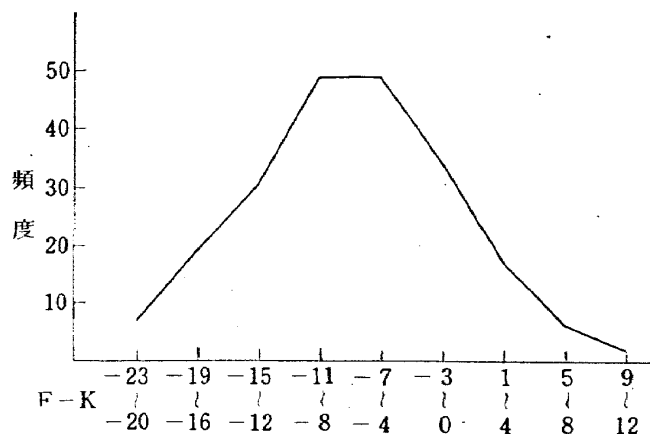
Hs	0	1	2	3	4以上	計
群別						
F群	22	26	11	5	7	71
N群	0	1	2	5	66	74

で比較した結果を示してみる。N群ではs H尺度で、粗点が3点以下になることは10.8%で非常に少ないことがわかる。ところが、よい印象を与えようとする(F群)3点以下の出現率が90.1%にはねあがる。Hs粗点3点を態度検出の臨界点とすると、F群で9.9%N群で10.8%の誤りを生ずることになる。つまり全体として、 $(9.9\% + 10.8\%) / 2 = 10.4\%$ の誤りを生じうる。H尺度は結構すぐれた妥当性尺度、少なくとも、良い印象をあたえようとする被検者を見出すのには、L尺度におとらず使用できる尺度であるといえる。しかし、なかにはほんとうに健康で、身体的悩みのない人がいるかもしれない。Hs粗点だけで判断するのは危険であろう。同時に、よくみせようとする態度が働いていても、現実身体的欠陥があつたり、たまたま健康を害していたりすると、Hs粗点が3点以下にとどまることはすくないであろう。しかし、誠実な回答をしたN群で、粗点3点以下が約10%であり、よくみせようとしたF群で4点以上が約10%であることから判断して、Hs粗点が極端にすくない被検者は、いちおう、よい印象を与えようとして回答をしたと疑つてよいようである。

F-K指標

好ましい方向にしる、好ましくない方向にしる回答を歪めようとする態度を検出するのにF-K差は有効であるといわれている(4, 5, 9)。筆者(12)は適性検査時におけるF-K指標について、若干の分析を試みてみたが、ここに再びとりあげ、F-Kが態度によつてどのように変動するか検討してみたいと思う。まず、特別の教示なくMMP Iに回答したときのF-K指標の得点をFig. 2に示してみよう。この得点分布はいわゆる正規分布をなしているといえる。

Fig. 2 F-K差の分布



第1回目に全員にMMP Iを実施したときのF-K指標の平均得点は-7.5、標準偏差は6.62であつた。この数値は前回報告した適性検査用として実施したMMP I

のF-K差の平均値にはほぼ等しい ($M = -7.8$ $SD = 6.53$)⁽¹²⁾

N群においては、F-K差につき、第1回目と第2回目の平均値に有意の差がみられなかつたので、第2回目に実施したKMP IのF-K指標のFおよびU群の平均値をN群のそれと比較すると、Table 10 のようになる。

Table 10 F-K差の比較

	N	M	SD	差	t
F群	71	-15.66	4.745	8.05	8.656**
N群	74	-7.61	6.319		
U群	69	+18.57	13.193		

(注) **は1%レベルで有意

良い印象を与えようとしたり、その逆の態度をもつて回答すると、F-K値は統計的に有意に差を生ずるのである。すなわち、好ましい方向へ歪曲しようとする、F-K指標はマイナス値で大きくなり、好ましくない方向へ歪曲しようとする、プラス値で大きくなるといえる。

F-K指標について、わが国で標準化されたものがないので、第1回に実施した全員の平均値をもとにして、F-K指標の検出力を推定してみる。T=30およびT=70をその限界点とすれば、F-K指標はそれぞれ-21および+6になる。この2つを態度検出用の限界点として第2回のF-K値の分類をしてみると、F群のような良い印象を与えた態度には殆ど役立たないが、悪い印象を与えようとする態度にはすぐれた力をもつことがわかる (Table 11-a 参照)。限界点の置き方が問題視されるの

Table 11-a F-K指標による分類

F-K	N 群	F 群	
> -21	73(98.6%)	61(85.9%)	$\phi = -.19$
≤ -21	1(1.4%)	10(14.1%)	
F-K	N 群	U 群	
< 6	69(93.2%)	10(14.5%)	$\phi = -.72$
≥ 6	5(6.8%)	59(85.5%)	

注 () 内は各群の全員を母数とした%である

で、最良の分類に必要な得点を決定するために、F-K指標による限界点を変え、N群とF群、およびN群とU群で、点相関係数 ϕ を算出した。F群とN群では-14、U群とN群では+7を限界点とすると、最大の ϕ 係数がえられた。前者では $\phi = -.52$ 、後者では $\phi = -.81$ となり、F-K指標による検出力は悪い印象を与えようとする態度に対してよりすぐれているようである。+7はGough(9)が最良の限界点と結論した数値に一致する。

-14と+7を臨界点として分類してみると、Table 11-b のようになる。

Table 11-b F-K指標による分類

F-K	N 群	F 群	
> -14	51(68.9%)	13(18.3%)	$\phi = -.52$
≤ -14	23(31.1%)	58(81.7%)	
F-K	N 群	U 群	
< +7	72(97.3%)	12(17.4%)	$\phi = -.81$
≥ +7	2(2.7%)	57(82.6%)	

-14を好ましくみせようとする態度検出の限界点とすると、まじめに答えた回答の68.9%を、ごまかし回答の81.7%を正しく分類する。逆にいえば、 $(18.3\% + 31.1\%)/2 = 24.7\%$ の分類誤差があることになる。一方仮病あるいは悪印象を与えようとする態度検出の限界点+7によると、平均10.1%の誤差を生ずることになる。Gough(9)のいうようにF-K指標は好ましくない方向への歪みを見破るのに、特にすぐれているといえる。しかしF-K指標の高い被検験者が回答中に示すうそ発見者の記録から、この指標の実用的価値を認めようとしたり(1)、就職応募者のF-K指標がその扶養家族数と高い相関(.61)のあることを示す(5)研究などによつてF-K指標は好ましい方向への虚偽回答検出にも有力であることが支持されており、さらに検討を加えねばならない。わが国におけるMMP Iの標準化に際し、この指標の標準化も必要であり、さらに正常人においても、患者においても、有力な検出力のあることが証明されねばならないであろう。

L+K-Hs指標

F-K指標の分析結果からもわかるようにMMP Iの妥当尺度を単独に用いるよりは、これらの尺度を組み合わせの方が効力があるようである。第1回目に実施したときのMMP Iの得点を全員214名について、K、L、Hsの3尺度の相関係数を算出してみると、L対Kは.51、L対Hsは-.33、K対Hsは-.48となる。それぞれ好ましい方向への歪曲を検出するのに役立つ尺度であろうがあまり高い相関を示さないことは、おのおの異つた側面にふれているためではあるまいか。

Hsは好ましい方向に歪めようとする著るしく粗点が低くなる。つまり、HsはKとLに対しては負の相関があり、KとLは正の相関である。したがってL+K-Hsを態度検出の指標として、F、U、Nの3群の第2回目の回答において平均値を比較してみた。Table 12からもわかるように、F群ではL+K-Hs値はN群に比

べて著しく大きい。またこの値はF-K指標とは逆に、U群ではマイナス値になる。

Table 12 L+K-Hs指標の比較

	N	M	SD	差	t
F群	71	29.91	5.079	19.00 13.23	20.772 7.691
N群	74	10.91	5.228		
U群	69	-3.32	9.280		

(注) **は1%レベルで有意

N群とF群、およびN群とU群で、最良の臨界点を求めるため、最高の点相関係数を算出したところ、前者ではL+K-Hs得点が+21で $\phi = -.88$ 、後者では+5で $\phi = -.55$ という最大係数がえられた。これによつて、おのおの2群を分類すると、Table 13 のようになる。

Table 13 L+K-Hs指標による分数

L+K-Hs	N 群	F 群
<+21	72(97.3%)	7(9.9%)
≥+21	2(2.7%)	64(90.1%)
L+K-Hs	N 群	U 群
>+5	67(90.5%)	25(36.2%)
≤+5	7(9.5%)	44(63.8%)

L+K-Hs 指標によると、よい印象をあたえようとした被検者の約90%を検出でき、まじめに答えた被検者の97%を正しく分類できることになる。すなわち、L+K-Hs得点 21 を臨界点とすると、 $(9.9\% + 2.7\%) / 2 = 6.3\%$ の分類誤差しかない。この数値は好い印象を与えようとする態度を見破るのに非常にすぐれた指標となりうることを物語るものであろう。L, K, Hs にはそれぞれ、妥当性尺度としての意味以外のものが含まれており、単独で用いるよりは、このように組み合わせた方がすぐれた検出力をもちうるものと思われる。L, K, Hs 指標は元来、よくみせようとする回答の歪みを見破るための尺度を組み合わせたものであるから、悪い印象を与えようとする仮病にはそれほどの効力があるとは期待できないが、Table 12 からわかるように、その場合でもこの得点は有意に小さくなるようである。しかし、その検出力は5を臨界点とした場合、誤り22.9%となり、好ましい方向への歪みをみつけるほど、強力ではないことがわかる。

われわれの資料は患者が含まれておらず、病状のためHs 尺度が高い被験者においても、この指標で有効な力をもつかどうかは断言できない。しかし、正常者における好ましい方向への歪曲回答には、少なくとも、非常に

有効であると思われる。L+K-Hs 指標については、他に資料を見出すことができず比較検討するわけにはいかない、今後の吟味を必要とするであろう。

プロフィールの型

MMP I ではプロフィールの型によつて、病状の臨床診断ができるようになってきているが、検査をうける際の態度の影響はプロフィールにどのようにあらわれるであろうか。Fig. 1 をみると、F群で顕著な型を示している。悪い印象を与えようとする仮病のはこぎりの型のようなプロフィールを描くようである。すなわち、D, Pd, Pa, Maとくに後の3尺度で著しくピークをなしている。他方 Delay たち(4)の資料ではむしろScが最高のピークをなすのはこぎり型が示されている。この相違をただちに民族の差、文化の差とするのは危険であろう。どちらも日仏両国の代表的な標本より得られた資料であるかどうか保証の限りでないのであるから。

ところで、Fig. 1 のプロフィールは全被検者の平均値から描かれたものであるから、のこぎり型のプロフィールを示す者がはたしてどの程度いるかをしらべる必要があろう。U群のうちから、類似のプロフィールは23名(33.3%)が見出された。F群にはなかつたが、N群では2名(2.7%)であつた。N群の2名は最高ピークがTスコア50以下であつたが、のこぎり型をしていたものである。他方U群の他の46名のプロフィールは必ずしもPd, Pa, Maでピークをなさず、のこぎり型のプロフィールではなかつた。プロフィールはその記入票の基礎となる尺度の標準により異なってくるはずで、米国の標準化を利用したのでは、違つたプロフィールが得られよう。Fig. 1 のプロフィールは東大改訂版によつたものである。のこぎり型はU群のような特殊な条件におかれて回答した被検者の33%にすぎないのであるから、型によつて回答の虚偽を見破るにはかなりの熟練が必要とされる。むしろ、数値的にはつきり示しうるF-K指標を利用した方が無難であろう。

IV 結 語

回答のごまかしということはパーソナリティ検査を臨床診断用に利用する場合、常に問題となる。本研究においてはMMP I 東大改訂版(集団式)によつて、検査をうけるときの態度が回答を有意に歪曲できることが証明された。MMP I ではこの事実留意して、修正値が加えられるようになってきているが、われわれの資料によると自己をよく見せようとする被検者の回答はある程度修正されうるが、仮病を示す被検者の回答には十分効果をあげることができない。これはMMP I 形式の質問票を利

用するかぎり避けられない限界であろう。比較的安易に作成され、実施も採点も特殊な名人芸を必要としないので、性格検査用の新しいペーパーテストが次々あらわれてくるが、検査利用者は結果の解釈にあたっては、回答のごまかし、その他による歪みをよく知っておかねばなるまい。

MMP I では、回答のごまかしという問題を処理する一助として、妥当性尺度が診断尺度の他に構成されている。？、L、F、Kがそれであるが、最近の研究では社会的望ましさを測定するS-D尺度(8、10) 回答のバイアスを知るB尺度(7)という別の尺度を提唱する人がある。しかし、それでなくても質問数が多く、採点に時間がかかり、面倒がられているMMP I に、態度検出のため新尺度を加えることは考えものである。本研究では回答の信頼度を検出する既成の4妥当性尺度を検討し、さらに、好ましい方向に歪めようとする態度を見破るため、Hs 尺度粗点を、悪い印象を与えようとする態度に対しプロフィールの利用を行なってみたが、Gough(9)のいうようにこれらを単独で利用するより、組み合わせた方が効果がある。またF-K指標に関しては、Goughの結果を支持できた。F-K指標は仮病回答に有効である。良い印象を与えようとする態度にはL+K-Hs 指標だけで、非常にすぐれた弁別力をもつことが証明された。

本研究ではいわゆる正常者群だけによる実験的な資料であり、異常者群でも、同様の検出力をもつかどうか、特にL+K-Hs 指標のような新指標については一まつの不安がある。今後の研究を必要とする問題である。また、項目別に分析し、歪みやすい質問とそうでないものとの検討してみる必要もあろう。

質問票形式の検査では回答のごまかしが可能であるとすれば、その検出が必要である。検出のためには妥当性尺度を付加しなければならない。しかし、その尺度は申しわけ的につけたり、ただひとつだけでは不十分であるといえよう。

この研究にあたって、たびたび御助言くださった
東京大学肥田野直助教授に深く感謝します。

文 献

- (1) Calvin, A. D. and Hanley, C.: An investigation of dissimulation on the MMPI by means of the lie detector. *J. appl. Psychol.*, 1957, 41, 312—316.
- (2) Comrey, A. L.: A factor analysis of items on the K scale of the MMPI. *Educ. psychol. Me-*

- asmt.*, 1958, 18, 633—639.
- (3) Comrey, A. L.: A factor analysis of items on the F scale of the MMPI. *Educ. psychol. Measmt.*, 1958, 18, 621—632.
- (4) Delay, J., Pichot, P. et Perse, J.: La détection de la simulation à l'Inventaire Multiphasique de Personnalité du Minnesota (M. M. P. I.). *Rev. Psychol. appl.*, 1960, 10, 249—262.
- (5) Drasgow, I. and Barnette, W. L. Jr.: F-K in a motivated group. *J. consult. Psychol.*, 1957, 21, 399—401.
- (6) Fricke, B. G.: Subtle and obvious test items and response set. *J. consult. Psychol.*, 1957, 21, 250—252.
- (7) Fricke, B. G.: A response bias(B) scale for the MMPI. *J. counsel. Psychol.*, 1957, 4, 149—153.
- (8) Fordyce, W. E.: Social desirability in the MMPI. *J. consult. Psychol.*, 1956, 20, 171—175.
- (9) Gough, H. G.: The F minus K dissimulation index for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1950, 14, 408—413.
- (10) Hanley, C.: Social desirability and responses to item from three MMPI scales: D, Sc, and K. *J. appl. Psychol.*, 1956, 40, 324—328.
- (11) 肥田野直: テスト法による性格診断——主としてM. M. P. I. を中心として——. 輿論科学協会研究紀要 No. 18, 1957.
- (12) 岩脇三良: MMP I の回答妥当性について. 教育心理学研究, 1961, 9, 101—105.
- (13) 児玉 省: 性格診断法(Ⅱ). 「心理学講座」第7巻, 中山書店, 1953.
- (14) Pichot, P.: *Les testes mentaux*. Paris, P. U. F., 1954.
- (15) Smith, E. E.: Defensiveness, insight, and the K Scale. *J. consult. Psychol.*, 1959, 23, 275—277.
- (16) 山田 旭: MMP I 及びロールシャッハ・テストによる精神障害者の類型診断——精神分裂病及び神経病について——. 教育心理学研究, 1960, 8, 92—101.

(1961年9月5日原稿受付)

ABSTRACTS*

A Q-TECHNIQUE STUDY OF THE TYPES OF PRIMARY SCHOOL TEACHERS : I

—The Type Factors of Educational Attitude—

by

Takeshi Takeuchi

Shirō Yabuki

&

University of Chiba

Osaka Gakugei University

1. The purpose of this study was to find factors in the educational attitudes of primary school teachers. In this investigation W. Stephenson's Q-technique was used.

2. A collection of nearly 450 statements was first made of opinions and attitudes that 130 primary school teachers had to four hypothetical educational situations. From the collection, 100 unstructured samples of statements were selected and Q-sorted by 25 primary school teachers.

3. Those 25 Q-arrays were correlated and factor-

analyzed. Three type factors were extracted and each of them was twice orthogonally rotated. The rotated factors were interpreted by the method of computation of the factor scores of each statement and designated under the names of (a) compulsion-freedom, (b) generosity-severity, (c) harmony-estrangement.

4. With relation to the interpretation of factors, several problems about the methodology of Q-technique were considered.

FAKABILITY AND VALIDITY INDICATOR IN THE MMPI

by

Saburo Iwawaki

Defense Academy

The Japanese edition of MMPI was administered twice to a group of 214 Japanese cadets. The first administration was under the usual instructional set. Two weeks later, the Ss were divided into three subgroups consisting of approximately an equal number of Ss which differed according to test-taking instructional set. Two experimental groups (favorable set group and unfavorable set group) and one

control group (neutral set group) were utilized.

The introduction of favorable or unfavorable set produced significant changes in the MMPI scores of testees. The K correction has not been found sufficiently valuable to indicate the faking bad profiles.

In order to test the efficiency of validity indicators, it has been planned to make a comparison

* We owe editing of English abstracts in this issue to Dr. Ray Simpson of the University of Illinois.

of the validating scores obtained from the three different set groups. The results reported in the present study indicated that the three keys (L, F, K), considered singly, are able to identify faking sets to some extent and that Hs raw score, utilized singly, was more successful in discriminating the faking good from the normal scores, but their maximum efficacy is realized in combination.

The results of this study confirmed Gough's results that the F-K index has a superior value to detect a normal person who has tried to fake. In the present data, the F-K cutting score for the faking bad profiles was plus seven. This cutting score

would correctly classify 97 per cent of the normal cases and 83 per cent of the faking bad cases.

A new index — L raw score plus K raw score minus Hs (Hypochondriasis scale) raw score — was presented to detect the faking good set. The highest phi coefficient was given by a cutting score of plus 21. This cutting score would correctly classify 97 per cent of the normal cases and 90 per cent of faking good cases.

The F-K index and the L+K-Hs index may be considered as efficient measures to detect malinger and test dissimulation in the MMPI.

COMMUNICATION IN EMOTIONAL FRUSTRATION

by

Hiroko Kameyama

Tokyo Woman's Christian College

What do we do in case of frustration which gives us on emotion like hostility or shame? Some of us may blame others and some may be embarrassed. The present study is concerned in a restricted way with the role of communication in reducing emotional tensions, and is related to the type of behavior toward frustration.

The major hypothesis to be tested was as follows: In a situation where emotional frustration and hostile attitude is instigated, opportunity for communication directed toward the instigator (when compared with the lack of this opportunity) will lead to relatively less residual hostility toward the instigator.

EXPERIMENT

Subjects: The subjects were 60 female students of the college, chosen as the high aggressive type of frustration and the less aggressive type, by rating of the Rosenzweig P-F Study and the questionnaire made for this research. **Procedure:** The subjects who were scheduled to appear individually at the experimental room were introduced to a participant of the experiment. The experimenter, in her preliminary instructions, said the purpose of this study was to

discover how students could learn another's personalities by direct communication. This served as the pretext for the experiment. Before proceeding with the main business of communication to one another, both the subject and the participant were each instructed to write an initial a "personality sketch" consisting of her first impressions of the other. This much accomplished, subjects began the communication process, on which the restriction that the content of communication must refer to self was imposed. The participant intended to characterize her as arrogant and opinionated, instigating hostility in the subjects. Her role was standardized and well rehearsed in advance. 30 of the subjects (No-Com group) were not permitted to communicate after the instigation. The remaining 30 subjects (Com group) were permitted additional communication after the instigation. At the conclusion of this period, each of them wrote a final personality sketch of her partner.

RESULTS

In treating the data, content analyses were done separately for their pre-experimental and the post-experimental "personality sketches" of the parti-