

# 第2章

## 妥当性と信頼性

第1節 妥当性 42

第2節 信頼性 50

性格検査に限らず、知能検査や適性検査を含めて、謂うところの心理テストが科学的に使い得るものかどうか、これを数値をもって示すのが「妥当性」と「信頼性」という概念です。妥当性とは、テストが測定しようとしている目標、例えば学力とか知能とか性格特性をどれほど正しく測定できるのか、信頼性とは、テストの結果(得点)がテスト時の様々な変動要因（例えば被検者の健康状態、気候や気温など）を超えて、どれほど被検者の真の得点に近かつ安定的に迫り得るのか、それぞれ統計的な数値をもってその程度を示す概念です。真つ当な性格検査なら、必ずマニュアルその他でこの2つの概念と数値を示しているはずで、もしこれらの概念と数値に触れていないとするなら、それは古いやまやかし物同然のテストということになります。

2つの概念は、一般には信頼性、妥当性の順に述べられています。それは、広い意味では妥当性も信頼性の下位概念の一つと考えられるからですが、本書では、カラー・ピラミッド・テストにより近くなってもらうため、まずは何を測定することができるのか、「妥当性」の方から説明していくことにいたします。

## 第1節 妥当性

妥当性と言えば、心理検査の各書は、例えば構成概念妥当性、併存的妥当性、因子的妥当性といった、なにやら難しい用語を手がかりに説明するのが普通です。しかしここでは、一般の読者にはどこか分かりにくく馴染みにくいこのような用語や概念を避けて、カラー・ピラミッド・テストが一体どのような性格次元を測定できるのか、もっとシンプルに説明して参りたいと思います。

ところで、色彩と感情、あるいは色彩とパーソナリティとの関係については、もう十分すぎるほど分かっているのではないか、今さら妥当性云々ということもないのではないか、と思う人がいるかも知れません。たしかに、色彩に関わる数多の先駆的な研究が、たくさんの興味深い知見を残しています。そのほんの一端は、序章で触れました。しかし残念なことに、それらの知見をそのまま借用してカラー・ピラミッド・テストで選択された色彩の解釈を行うことはできません。何故

なら、それぞれの調査や研究はそれぞれ固有の材料と方法で行われ、例えば、呼称上は同じ赤でも、それらはみな異なる色と形、ときには単に言語刺激だけで与えられていることさえあり、それ故、それぞれのところで得られた結果は、それぞれの実験上の具体的な色彩刺激と結び付けられて、初めて解釈可能な事柄になるのです。カラー・ピラミッド・テストもそれとまったく同じ理由で、カラー・ピラミッド・テスト独自の調査や実験によって色彩とパーソナリティ次元との関連を明かし、それに基づいて初めて確かな診断指標とするということになります。

とは言え、色彩と感情、あるいは色彩とパーソナリティとの関わりについては、これまでの研究結果を通観してみれば、各色彩の意味合いが最大公約数的に纏まってくることが分かります。そして、これらの多くは、同じように色彩を用いるカラー・ピラミッド・テストでもほとんど同様に通用し、性格診断上の重要な指標となっています。もちろん、そうした各色彩の意味合いも、すべて、カラー・ピラミッド・テストにおける調査・実験の結果、一つずつひとつずつ、実証していったものであることは、言うまでもありません。

例えば、赤色について、また青・緑について、それを実証する簡明かつ興味深い資料を見てみましょう。

赤・緑・青について言われている諸特性（例えば、情熱的な、繊細な、沈着な・・・）の中から、単なるその折（一過性）の気分や感情状態を示すようなレベル（例えば、昂揚した、傷ついた、落ち込んだ・・・）を除いてパーソナリティ特性を示す次元に着目すると、赤は刺激に対する反応性とか、外向的、行動的な活動性、それに対して、緑・青は刺激に対する統制性とか内向的、思考的な情動性が取り上げられます。この、赤そして緑・青がそれぞれ投影すると思われるパーソナリティ次元で、一般的に、日本人と米国人とで比較してみたらどうでしょうか。近年では、内弁慶で国際試合では弱いなどと言われていた日本のスポーツ選手達もむしろ対外試合の方が燃えてくるなど、日本人全般で島国根性など吹き飛ばしてきた感もなきにしもあらずですが、それでもなお、大勢的には、やはり彼方アメリカ人は自由闊達で、より行動的、より外向的なのに対し、此方日本人では控え目で、より思考的、より内向的な傾向があることは否めません。もし、赤そして

緑・青が、ここで指摘するようなパーソナリティ次元と関係しているなら、この両国におけるカラー・ピラミッド・テストの結果にそれが反映してくるはずで  
す。即ち、米国人では、より行動的、より外向的傾向を投影する赤をより多く好  
むのに対して、日本人では、より思考的、より内向的な傾向を投影する緑・青を  
より多く好む、といった傾向が見られるのではないのでしょうか。

そこで、シャイエとハイス(1964)が行った米国の標準データと私たちが行った  
我が国の標準データにおける同一年齢(6歳～18歳)での色彩の選好順位を比較し  
てみましょう。結果は、次の通りです。

### 男子

米国：①赤②青③緑④黄⑤橙⑥黒⑦紫⑧茶⑨白⑩灰

日本：①緑②青③赤④黄⑤橙⑥紫⑦白⑧黒⑨茶⑩灰

### 女子

米国：①赤②青③緑④黄⑤紫⑥橙⑦茶⑧白⑨黒⑩灰

日本：①青②緑③赤④黄⑤橙⑥紫⑦白⑧茶⑨黒⑩灰

物の美事に、米国人はより行動的で外向的、日本人はより思考的で内向的とい  
う上の仮説が実証されています。米国では男女ともに最も好まれる第1位の色は、  
緑・青を押さえて赤ですが、日本では男子が緑と青、女子が青と緑をもって1・  
2位を占め、米国では男女ともに1位の赤を3位に押さえています。(なお、こ  
の選好順位は、米国では美P、本邦では好Pという若干ニュアンスの異なる条件  
下のものですが、好嫌条件をカラー・ピラミッド・テストの標準教示にするまで  
に数多く行ってきた私たちの美醜刺激での研究でも全く同様の結果が得られてい  
ます。)

日米両国の標準データを比較して得られるこのシンプルな仮説一検証の結果  
は、色彩刺激を用いるカラー・ピラミッド・テストが十分に妥当性の高い性格検  
査たり得るのではないかと、まずはその思いを抱かせてくれますが、さて、カラ  
ー・ピラミッド・テストの妥当性の研究は、用いられる各種色彩とその色彩チッ  
プによって構成されるピラミッド図形(模様)がどのような感情やパーソナリテ  
ィ次元を測定し得るのか、という基本的なところから始まります。そして、さらに

進んで、第5章および第6章で詳述される色彩症候群、経過形式、形態変動性といったカラー・ピラミッド・テストのさまざまな臨床カテゴリーの妥当性を検討していくこととなります。これまで、カラー・ピラミッド・テスト研究の諸先達が、健常者を対象に、あるいは健常者における検査結果をベースに神経症、統合失調症、躁鬱病、非行・犯罪、知的障害など多様な臨床群の反応を比較・検討し、各臨床カテゴリーの妥当性を検証していますが、私たちも同様にそうした研究をつみ重ね、色彩と構造の意味を明かして参りました。

ここで、その一端を紹介しますと、例えば、一般にパーソナリティ発達が未熟で不安が高く衝動的で行動も不安定になりやすい非行少年群と非行のない健常少年群の反応を比較・対照し、以下のような結果を得ました。まず色彩反応に関して申しますと、非行少年群は健常少年群よりも、赤や紫を好み、茶を好んだり好まなかったりしてアンビバレンスな感情を抱き、また、好ピラミッドで最も多く選択される赤・緑・青の選択の合計(標準症候群)が非行少年群の方がより高く、さらに、各3回ずつ反応する好嫌それぞれのシリーズを通じて1回のピラミッドにしか用いられない色彩の数(経過形式のMaS)が非行少年群の方により多く見られる等々…。もちろん、単純に推論するわけではありませんが、これらの結果から一応、次のような仮説が成り立ちます。赤は衝動性、紫は不安、茶は自我意識や耐性などを投影、標準症候群は未成熟やステレオタイプの指標、MaSは行動の不安定性や情動不安定の徴表…。次いで形態反応に関しては、非行少年群では明らかに構造の形質が低く、個別的に言えば、構造性の低いじゅうたん型や多色成層型に偏り、非対称力動構造などの構造性の高い形質がほとんどありません。これらの結果は、カラー・ピラミッド・テストの形態反応が、パーソナリティの分化と成熟に関わる臨床カテゴリーたり得ることを十分に示しています。

妥当性に関するたくさんの貴重な研究結果があります。それらの恩恵に浴しながら、第5章と第6章ではカラー・ピラミッド・テストの解釈編を展開致しますが、その解釈の基礎たるべき研究、とりわけ病理学的臨床群を用いた諸研究を一覧表にした坂野雄二氏の労作を以下にご紹介したいと思います。

表 2-1 正常者と病理学的臨床群との比較 (坂野雄二, 1983)

研究者(発表年)	被験者	統制群	主 な 結 果
Spreen, O. (1951)	肺結核患者 神経症	あり	<ul style="list-style-type: none"> <li>・結核患者, 神経症患者は, 健常者に比べ, 刺激症候群の値が低い。</li> <li>・神経症患者では健常者に比べて無彩色(白・灰・黒)の選択が多く, 黄・橙の選択が少ない。</li> <li>・神経症患者では, 構造型が少なく, マント型が多い。</li> </ul>
Wewetzer, K. H. (1951)	統合失調症 躁鬱病 てんかん	あり	<ul style="list-style-type: none"> <li>・統制群に比べ実験群3群では, 緑・紫・白・灰の選択が多く赤・黄・橙の選択が少ない。また, 構造型が少なく, じゅうたん型が多く, 色彩に対する反応が優勢である。</li> <li>・統合失調症では, 特に白・紫3の選択が目立ち, 「分裂じゅうたん」を構成した者が多い。色彩に対する反応が優勢であり, 白が他の色彩に関係なく無秩序に出現する。</li> </ul>
Bregelmann, J. C. (1953a)	神経症 鬱病 統合失調症	あり	<ul style="list-style-type: none"> <li>・健常者に比べ, 神経症患者では茶色, 鬱病・統合失調症では紫の選択が多い。</li> <li>・選択行動のパターンでは, 健常者と患者群に顕著な差はない。</li> <li>・経過形式では, 健常者と患者群に顕著な差は見られない。</li> <li>・反応時間は, 健常者に比べ, 患者群で有意に長い。</li> </ul>
Bregelmann, J. C. (1953b)	神経症 鬱病 統合失調症	あり	<ul style="list-style-type: none"> <li>・ピラミッドの形質について, 健常者と患者群では, 統計的に有意な差異は認められない。患者群に特有なピラミッド構成方法は見られない。</li> </ul>
Frohoff, W. (1953)	統合失調症	なし	<ul style="list-style-type: none"> <li>・紫・灰・白の選択が多く, 黄・橙の選択が少ない。</li> <li>・構造よりも色彩に対する反応が優勢である。</li> <li>・緊張性統合失調症患者では, 特に構造型のピラミッドが少ない。</li> </ul>
秋谷たつ子 (1955)	知的障害児	あり	<ul style="list-style-type: none"> <li>・知的障害児は, 情緒面で精神年齢に等しい統制群と近似の結果を示す。</li> <li>・知的障害児に紫・緑の選択が多く, 安定性の欠如が示されている。</li> </ul>
Sacher, H. (1955)	統合失調症	あり	<ul style="list-style-type: none"> <li>・破瓜型患者では単一色ピラミッドが多い。この傾向は分裂病質の者には見られない。</li> </ul>
秋谷たつ子 (1956)	統合失調症	あり	<ul style="list-style-type: none"> <li>・色彩選択に関し, 統合失調症は黒・白の選択が多い。</li> <li>・統合失調症にじゅうたん型のピラミッドが多い。</li> </ul>

Ziolko, H. V. (1956)	神経症	なし	・試行を追うにしたがって、黒の選択が増加する。同時に色彩の変化が減少し、単一色ピラミッドが増加する。
O' Reilly, P. O., Holzinger, R. & Blewett, D. (1957)	統合失調症 躁鬱病 強迫神経症 パラノイア	あり	・病歴年数による差異は見られない。 ・統合失調症の患者は、そうでない患者あるいは健常者に比べ、白の選択頻度が高く、ピラミッドの一部が欠損した破壊型のピラミッド構成が有意に多い。
Seyfried, H. (1957)	非行少年	あり	・非行少年群は統制群に比べ紫の選択が有意に多い。 ・形質も非行少年群では構造化の進んでいない者が多く見られる。 ・一般的に色の変化が乏しい。
徳山孝之 (1958)	非行少年	なし	・緑・茶の選択が多い。 ・Y-G性格検査との比較の結果、内向性と青・緑・白が正の相関を、黄・橙・灰が負の相関を示した。 ・Y-G性格検査とCPTの一致度は年少者の方が年長者よりも高く、性格異常者が健常者よりも高い。
O' Reilly, P. O. & Blewett, D. (1959)	統合失調症 その他の 精神病	あり	・精神病患者群では、健常者群よりも紫の選択が多いが、統合失調症と健常者の間には差は認められない。 ・患者群は健常者群より橙の選択が多い。 ・統合失調症では、白の選択が多い。 ・標準症候群の得点は、群間で差が認められない。
秋谷たつ子 (1960)	統合失調症	あり	・統合失調症者は、美・醜両ピラミッド間の変化に乏しい。 ・統合失調症者の色彩選択頻度には個人差が大きい。 ・経過形式は健常者と統合失調症者で逆のパターンを示す。
秋山博之 丸岡 瞭 (1960)	非行少年	なし	・窃盗犯では緑3・青4の選択が多く、緑4・黒の選択が少ない。 ・暴力犯では赤2・黒の選択が多く、青4の選択が少ない。 ・性犯罪者では、緑4・黒・青4の選択が多く、橙の選択が少ない。
丸岡 瞭 秋山博之 (1960)	非行少年	なし	・偏執性の者は、美ピラミッドで緑3・青4、醜ピラミッドで赤2・橙2の選択が多い。 ・発揚性の者は、美ピラミッドで橙・青3、醜ピラミッドで茶・紫の選択が多い。 ・ヒステリー性の者は、美ピラミッドで赤4・茶、醜ピラミッドで白・赤2・青3の選択が多い。

秋山博之 丸岡 瞭 (1961)	非行少年 交通事故者 統合失調症	あり	<ul style="list-style-type: none"> <li>・非行少年群では、秩序傾向を伴うじゅうたん模様が多い。また、交通事故者群では成層構造が多く見られる。</li> <li>・精神病群では純粹じゅうたん、多色成層が多く見られる。</li> <li>・ピラミッド構成の複雑性・安定性はパーソナリティの安定性と関連あり。</li> </ul>
秋山博之 (1963)	非行少年	なし	<ul style="list-style-type: none"> <li>・ピラミッドの構造化の程度とロールシャッハ・テストのM反応の間には正の相関が認められる。</li> </ul>
Houben, A. M. J. (1963)	神経症	あり	<ul style="list-style-type: none"> <li>・神経症者は健常者に比べて標準症候群に属する色彩の選択が少なく、無彩色あるいは紫の選択が多い。</li> <li>・醜ピラミッド・シリーズに神経症者の特徴が顕著に現れる。</li> <li>・形質の「未熟性」が神経症者には見られる。</li> </ul>
丸岡 瞭 (1963)	麻薬常用者 一般受刑者 非行少年	あり	<ul style="list-style-type: none"> <li>・麻薬常用者では赤・黄1・青2の選択が多い。</li> <li>・標準症候群の得点は低く、逆に刺激症候群の得点が高い。</li> <li>・不均衡じゅうたん模様が多く見られる。</li> </ul>
川久保芳彦 (1964)	神経症	あり	<ul style="list-style-type: none"> <li>・神経症群は健常者に比べ、緑・黄・紫の選択が多く、青・黒が少ない。この違いは3回目のピラミッドにおいて顕著である。</li> <li>・形質では、神経症群に成層型が少ない。</li> <li>・不安神経症では緑・赤が多く、茶・黒が少ないが、神経衰弱ではその逆である。</li> <li>・不安神経症では神経衰弱、象徴形成状態群よりも構造型が多く、じゅうたん型はその逆である。</li> </ul>
Schaie, K. W. & Heiss, R. (1964)	知的障害	あり	<ul style="list-style-type: none"> <li>・知的障害を示す男子では、健常者に比べ、茶が多く青が少ない。</li> <li>・醜ピラミッドで赤・橙が多く黒が少ない。</li> <li>・女子では美ピラミッドで橙・紫が多く、醜ピラミッドでは青・白が多く黒が少ない。</li> <li>・被験者をMAで統制すると、上記のうち男子美ピラミッドの青・茶、女子美ピラミッドの橙・紫は差が見られなくなる。</li> </ul>
山下栄子 長山泰久 (1971)	統合失調症	なし	<ul style="list-style-type: none"> <li>・反応時間が短かく、カラー・チップの交換がほとんど見られない。</li> <li>・色彩・形質ともに、3回の試行間で変化に乏しい。</li> <li>・美・醜両ピラミッド間の差異に乏しい。</li> <li>・反応に対する固執性が強く観察される。</li> </ul>

<p>中川大輪 (1972)</p>	<p>非行少年 (窃盗・強盗・強姦・殺人)</p>	<p>あり</p>	<ul style="list-style-type: none"> <li>・赤1・赤2の選択が非行群に多く、窃盗群で特に顕著である。一方、橙・黄は非行群で少ない。</li> <li>・経過形式では、窃盗・強盗群で健常者群に比べてCSの得点が高い。</li> <li>・統制群に比べ、殺人群では刺激症候群の得点が高い。強盗群では無彩色症候群の得点が高い。</li> <li>・非行群でじゅうたん型が多く、構造型が少ない。単一色成層型も非行群に多い。</li> <li>・非行4群間に顕著な差はないが、統制群との識別は可能である。</li> </ul>
<p>中川大輪 (1975a)</p>	<p>統合失調症</p>	<p>あり</p>	<ul style="list-style-type: none"> <li>・統合失調症は健常者に比べ反応時間が有意に短い。</li> <li>・統合失調症は健常者に比べ、茶の選択が多くて青が少ない。また、標準症候群の色彩が少ない。</li> <li>・形質では、統合失調症にじゅうたん模様が多く、構造型が少ない。また、美・醜両ピラミッド間で変化が少なく、固執傾向が見られる。</li> </ul>
<p>中川大輪 (1976)</p>	<p>統合失調症</p>	<p>あり</p>	<ul style="list-style-type: none"> <li>・反応時間は統合失調症の方が短い。</li> <li>・無彩色症候群の得点は統合失調症の方が低い。また紫の選択が多い。</li> <li>・統合失調症の反応において固執傾向が顕著である。</li> <li>・統合失調症の形質に構造型が少なく、分裂じゅうたんが多い。</li> </ul>
<p>宗内敦 鎌原恵子 宗内美映子 (1979)</p>	<p>非行少年</p>	<p>あり</p>	<ul style="list-style-type: none"> <li>・色彩選択に関し、非行少年は統制群に比べて、美ピラミッドで緑1・紫2が多く灰が少ない。醜ピラミッドでは黄2・緑1・茶1が多く、黄1・緑3が少ない。</li> <li>・非行群は統制群に比べ、美ピラミッドで標準症候群・内向性症候群の得点が高く、無彩色症候群の得点が高い。醜ピラミッドでは、標準症候群・内向性症候群の得点が高い。</li> <li>・CPTの尺度間で幾つかの内部相関が見いだされた。</li> </ul>
<p>宗内美映子 宗内 敦 鎌原恵子 (1980)</p>	<p>非行少年</p>	<p>あり</p>	<ul style="list-style-type: none"> <li>・宗内ら (1979) に加え、色彩選択に関し、非行群は統制群に比べ、美ピラミッドで赤4・紫3・茶2の選択が多く、醜ピラミッドで赤4・緑4が少ない。</li> <li>・経過形式では非行群でMaSの得点が高い。</li> <li>・各色彩別に解釈が行われている。</li> </ul>
<p>宗内 敦 宗内美映子 鎌原恵子 (1981)</p>	<p>非行少年</p>	<p>あり</p>	<ul style="list-style-type: none"> <li>・形質では、非行少年に単一色成層・多色成層が多く、構造じゅうたん・対称構造が少ない。</li> <li>・形態水準は非行群が統制群よりも低い。</li> <li>・ピラミッド・シリーズ内での変動性は非行群の方が低い。</li> </ul>

## 第2節 信頼性

心理テストの科学性を証明するもう一つの柱が、信頼性です。あるテストが、いえ、カラー・ピラミッド・テストが妥当性のあるテストであるとしても、そのテスト結果が不安定で、同じ被検者に再度テストを行ったときにも同じような結果が得られるという保証がなければ、困ります。テストスコアが不安定になるようなテストは、たまたま行ったそのときの結果をもって被検者の本来的な姿を推測することは到底できない、つまりは、信頼できないテスト（非科学的なテスト）なのです。

この信頼性に関しても、妥当性と同じように、手続きが複雑でなじみにくい迫り方がいくつかあります。例えばテストを折半し、あるいは同質の並行テストを作成して、それぞれの結果の一致度を見る方法で、前者は「折半法による信頼係数」、後者は「並行テストによる信頼係数」と呼ばれます。また、テストの内部一貫性というものを測定したり、分散分析を用いて求める信頼性もあります。しかし、カラー・ピラミッド・テストについては、その構造から見て並行テストを作ったりテストを折半したりすることは難しく、複雑な計算式を用いて算出する他の方法もあまり実際的とは言えません。このような事情から、カラー・ピラミッド・テストの信頼性については、専ら、同一の被検者にテスト期間を置いて再度同じ検査を行う再検査法によって行われてきました。本節では、これまでに行われてきた諸研究に加え、私たちが行った研究の結果を述べて、カラー・ピラミッド・テストの信頼性を見てみたいと思います。

再検査法によって得られる相関係数は、2回のテスト結果の近似度、いわばテスト結果の安定性に関連するものですから、「安定性係数」と呼ばれます。この安定性係数の値が統計的に有意な水準にあれば、そのテスト結果は偶然に支配されるような不安定なものではなく、被検者の真の値を推定するのに十分信頼しうる安定的なものということになり、例えばカラー・ピラミッド・テストにおいては、色彩の選択数や経過形式や形態水準などのスコアが、被検者の安定的かつ本

来的なものであること、即ち、その結果によってパーソナリティ診断を行ってよいものであることを示しています。(但し、前提条件として、各臨床尺度の妥当性が確かめられていなければ意味がありません。)

再検査法を用いた信頼性の検討は、大別して2つの方法によって行われてきました。第1の方法は、健常な被検者を用いて一定期間は自分でテストを2回実施し、2回のテスト結果の相関関係を求めて信頼性(安定性)を検討するものです。ここでは、「健常な被検者は再検査までにパーソナリティの諸構造にあまり変動がなく、比較的安定している」という仮説を前提にしています。従って、テスト間の相関が高いほど信頼性(安定性)が高いということになります。再検査法を用いて信頼性を検討する方法としては最も一般的に行われるやり方で、私たちが行った研究もまた、この方法によっています。

第2の方法は、「健常群」と精神的に不安定な「病理群」に各2回のテストを行い、2回のテスト結果の変動の程度を比較するものです。ここでは、次の仮説を前提にしています。「病理群は健常群に較べて再検査までの間に情動性やパーソナリティ上の変化が生じる可能性が大きい」。パーソナリティ・テストでは、このことを利用して角度を変えて信頼性の検討を行うことができます。即ち、病理群が健常群よりも2回のテスト間の“関係が不安定”あるいは“差が大”であることを示して逆説的にテストの信頼性を示すのです。

第1の方法には、ジョラス(Jolas, G. 1953)が行って以来、たくさんの研究があります。表2-2は色相・色彩・色彩症候群・経過形式など臨床カテゴリー別に相関係数のレンジと平均値を一覧表にしたものですが、これらのスコアは、一般に質問紙性格検査によって得られるスコアに近く、わが国における代表的な性格検査であるY-G性格検査の0.56~0.82(平均0.69)、M-G性格検査の0.49~0.72(0.63)と比べ、ほぼ同じレベルです。因みに、投影法性格検査の安定性係数は質問紙性格検査に比べてかなり低く、例えばカラー・ピラミッド・テストと同様に色彩刺激を用いる色彩象徴テストでは0.23~0.67(松岡1964)、ロールシャッハ・テストの色彩反応に関する1つの研究では0.20~0.47(久保松1963)ですから、これに鑑みますと、カラー・ピラミッド・テストの信頼性は、質問紙

表 2-2 先行研究における再検査信頼性係数

研究者(年)	種類	被検者	方法	期間	主要色彩(平均)	色相(平均)	色彩症候群(平均)	経過形式(平均)
Jolas(1953)	24色版	50人(男)	好P	4週間	.47~.81 (.61)	.33~.72 (.56)		.57~.83 (.72)
Karl(1956)	"	43人(男)	中間値	3~5 "	.66~.86 (.79)			.51~.79 (.68)
Schröder(1960)	"	50人(男)	"	2 "	.39~.71 (.60)			
Brehmer(1960)	"	45人(男)	好P	5 "	-.04~.71 (.41)		.22~.64 (.43)	.46~.58 (.51)
	"	60人(女)	"	5 "	.27~.84 (.55)		.50~.73 (.61)	.19~.58 (.45)
Reinert(1958)	10色版	100人(男)	好P	4 "	.63~.78 (.72)	左に同じ		.48~.66 (.59)
			嫌P		.62~.77 (.72)			.61~.69 (.64)
Michel(1959)	14色版	45人(男)	好P	3~5 "		.54~.83 (.68)		.72~.76 (.75)
			嫌P			.60~.79 (.71)		.58~.75 (.67)
Waniek(1974)	"	31人(男)	SD法	3 "		.55~.76 (.66)		

(被検者はすべて18歳以上)

検査法のそれに比肩し得るほどの高い水準にあると言えるでしょう。

第2の方法には、代表的なものとしてホーベン(Houben, A. M. J. 1962)とワニーク(Waniek, W. 1974)の2つがあります。ホーベンは、成人健常者(43名)、サナトリウムで保養中の患者(50名)、心理治療中の心身症患者(93名)に14色版カラー・ピラミッド・テストを実施し、それぞれ3~6週間後、3週間後、1~4週間後に再検査を行い、各群における色彩選択頻度の差の大きさを比較しました。情動やパーソナリティの不安定性は、心身症群>保養群>健常群の順序が仮定できるので、差の大小についての仮説もまたこれとまったく同じ順序になりますが、結果は仮説通りでした。その後、ワニークも同じ論法で健常群、神経症群、統合失調症群を用いて3週間後に再検査を行い、色彩刺激に対する反応の差の大きさについて、仮説通りに、統合失調症群>神経症群>健常群の結果を得ています。但し、ワニークのものは、カラー・ピラミッド・テストそのものを実施したのではなく、SD法によって各色彩ごとの印象価を求め、その再検査結果に基づく再検査安定性係数です。

このように、従来の研究によれば、カラー・ピラミッド・テストはかなり高い水準の信頼性を有しているように思われます。が、少し慎重に見直してみると、どの研究も、再検査までの期間が1ヶ月前後と少しばかり短めであり、その点が気になります。これは、性格検査の信頼性を検討する再検査期間としては短すぎ、

その結果によってカラー・ピラミッド・テストの信頼性を結論づけるにはかなり問題があるように思われるからです。

再検査法を用いて信頼性を推定し、あるいは検討しようとするとき、最も問題になるのは再検査までの期間です。一般に、テスト間の相関は、再検査期間が短すぎると記憶効果や学習効果が入りこんで高くなり、逆に長すぎると発達の要因や環境の変化によるパーソナリティの変容が生じて低くなります。こうした影響を排除する最も妥当な期間を求めるのは難しいことですが、多くの性格検査はその標準化を行うに際して、3～6ヶ月を再検査期間としています。因みに、比較のために先述したY-G性格検査とM-G性格検査の再検査期間はどちらも記憶効果や学習効果を除き得る3～6ヶ月となっています。そういう意味では、表2-2に挙げたカラー・ピラミッド・テストの信頼性研究の結果をこれと比較することは、かなり無理のある、いえ、間違ったやり方であったと言えるかと思いません。

本テストにおける私たちの臨床的な経験では、期間1ヶ月前後の再検査では被検者のほとんどがまだ前回の反応内容を記憶していますが、数ヶ月を過ぎてからの再検査では、そうした影響がほぼ消滅していることが分かっています。また、カラー・ピラミッド・テストは折々のムードに影響されて反応しやすいと言われる色彩刺激を用いることから、その反応の中には被検者の一時的あるいは短期間のムードや感情状態が反映されて必ずしもパーソナリティの安定した情動構造を投影しないのではないか、例えば季節の移動によって色彩刺激への反応は微妙に変わってくる、といった疑問や批判があります。これはロールシャッハ・テストにおいても色彩反応が最も信頼性が低いという報告(例えば久保松 1963)を援用するなどしながらもよく言われるところです。

これらの点を勘案しますと、カラー・ピラミッド・テストの信頼性に関しては、先行の諸研究とは異なり、一定期間を経過した後にもなお一定水準の安定性を持つか否かを検討することが、とりわけ重要だと考えられます。そこで私たちは、どの被検者も再検査期間を3ヶ月から4ヶ月として、カラー・ピラミッド・テストの信頼性の検討を行うことに致しました。被検者は、大学生男女73名(男30名、

表 2-3 再検査信頼性係数(男:30名  
女:43名)(宗内 敦, 1983a)

好ピラミッド		色彩形式	嫌ピラミッド	
男	女		男	女
.53**	.20	11	.31	.39**
.76***	.31*	12	.25	.51***
-.01	.18	13	.12	-.03
.54**	-.04	14	.06	.31
.43**	.18	Red(1)	.10	.38**
.25	-.01	21	.27	.03
.41*	-.01	22	.67***	.43**
.46**	-.09	Orange(2)	.70***	.45**
.13	.06	31	.48**	.12
.43**	-.05	32	.41*	.40**
.55***	-.08	Yellow(3)	.46**	.40**
.56***	.45**	41	.39*	.55***
.34*	.60***	42	.49**	.29*
.15	.08	43	.63***	.32*
-.09	.12	44	.21	.26*
.28	.54***	Green(4)	.38*	.32*
.73***	.12	51	.67***	-.09
.45**	.16	52	.01	.67***
.48**	.36**	53	.06	.34*
.23	.37**	54	.30	.28*
.75***	.36**	Blue(5)	.27	.29*
-.02	.32*	61	.78***	-.03
.20	.32*	62	.66***	.20
.03	.13	63	.32*	.23
-.03	.22	Purple(6)	.46**	.14
.18	.18	71	.30	.43**
-.03	.46**	72	.47**	.52***
.04	.42**	Brown(7)	.52**	.61***
.33*	.46**	White(8)	.11	.05
-.09	.16	Gray(9)	.22	.38**
.52**	.41**	Black(10)	.58***	.54***
男	女	色彩症候群	男	女
.47**	.18	Nsyn	.52**	.33*
.61***	.27*	Ssyn	.66***	.54***
.66***	.43**	Isyn	.43**	.36**
.53**	.37**	Asyn	.51**	.56***
男	女	経過形式	男	女
.64***	.34*	CS	.54**	.45**
.32*	.32*	MiS	.40*	.07
-.12	.43**	MaS	.01	.27*
.42**	.54***	AS	.78***	.59***
男	女	形態水準 形態変動性	男	女
.37*	.32*	FL	.58***	.39**
.10	.03	FV	.27	.25*
Σ FL:		(男).74***	(女).30*	
Σ FV:		(男).21	(女).29*	

有意性水準 \*5% \*\*1% \*\*\*0.1%  
(小数点3位以下切り捨て)

女43名)。実施時期は1回目が6月，2回目が季節感の異なる9月～10月です。

結果は，表2-3の通りです。私たちはこの表に細かな分析を加え、『教育心理学研究』（宗内<sup>(敬)</sup>1983a）に発表しました。分析の詳細はそれに譲って，結論的に申しますと，これらの結果は，3～4ヶ月経てもなお，カラー・ピラミッド・テストの再検査結果は一定水準の安定性を保ち，性格検査としての基本的要件（信頼性）を備えていることを示しています。今少し説明を加えるために，表2-2の先行研究の結果と比較しやすい資料を作ってみますと，表2-4のようになります。

表 2-4 再検査信頼性（安定性）係数

	10色彩	24色相	色彩症候群	経過形式
好P男	-.09～.75(.32*)	-.09～.76(.29)	.47～.66(.57**)	-.12～.64(.33*)
好P女	-.09～.54(.28*)	-.05～.60(.22)	.18～.43(.31*)	.32～.54(.41**)
嫌P男	.10～.70(.38*)	.01～.78(.37*)	.43～.66(.53**)	.01～.78(.43**)
嫌P女	.05～.61(.36*)	.09～.67(.30*)	.33～.56(.45**)	.07～.59(.35*)

有意性水準 \*5% \*\*1% \*\*\*0.1%: ( )内は平均値

これから見ると，諸データのスコアは先行研究のそれに比べて多少見劣りのする結果になっています。しかしそれは，再検査期間が先行研究の再検査期間に比べて3～4倍に延長されていることからむしろ当然の結果だと言えますし，また，好ピラミッドにおける全24色相別の反応を除き，あとは男女ともすべて統計的に有意な水準にあることを勘案しますと，この研究の結果をもって，カラー・ピラミッド・テストの信頼性は大いに確認されたと言ってよいでしょう。

そしてこのカラー・ピラミッド・テストの信頼性の研究は，付随的に妥当性の側面に関しても興味深い情報を残してくれました。こうした点もない交ぜにしながら，私たちは以下のような要約と結論を残しました。本書の読者の中には，今後カラー・ピラミッド・テストの研究に従事してくださる方もいらっしゃるのではないかと期待し，参考のためにも，その折の文章を一部訂正の上ここに記して，本節を締めたいと思います。

- ① 概括的に言えば、信頼性(安定性)は反応カテゴリーごとに水準が異なる。  
即ち、色彩症候群 $\equiv$ 形態水準 $>$ 経過形式 $>$ 色相・色彩である。なお、形態変動性については、今後の研究に待つところが多い。
- ② 色彩症候群は極めて安定性に富んだ臨床カテゴリーである。その安定度は、質問紙性格検査のそれに近い。
- ③ 経過形式の中では、CSとASの安定性が高く、色彩症候群と同水準である。
- ④ 色相・色彩の安定性係数のスコアは、個々には高いものがあるものの、全体としては必ずしも高くはない。が、一応の水準には達しているように思われる。
- ⑤ 好Pにおける高刺激色の選択に関して、顕著な性差があった。即ち、男子の安定性係数は総じて高く、女子のそれは総じて低い。
- ⑥ 好Pと嫌Pの間では、全般的に言えば、安定性は嫌P $>$ 好Pである。しかし、これを確言するには、もう一段の実験と検討が必要である。

カラー・ピラミッド・テストのように客観的・数量的に信頼性を検討できる投影法性格検査は、ほかにはなかなか見あたりません。前章(表1-2)で見たように、投影法テストは数多くありますが、テスト構成上の問題として、その多くは、客観的な数値をもって信頼性を示すことが困難です。そのため、それらのテストは科学的に使用し得るか否かに大きな疑問が生じます。比べて本検査の信頼性は、概ね質問紙性格検査のそれに近いほどの、かなりの水準にあり、それは、カラー・ピラミッド・テストが十分に臨床的な使用に耐え得る性格検査であることを示しています。