

バウムテスト研究における評価

— Koch の 58 指標を用いた経験から評価の一致率と作業手順を考える —

佐 渡 忠 洋

心身マネジメント学科

On Rating in the Baum Test Research Consideration for Concordance Rates and Working Procedure through Experience with Koch's 58 Indexes

Tadahiro SADO

要 旨

本研究の目的は、Koch の 58 指標を使った自験例から、バウムテスト研究における評価の一致率および評価の作業手順の問題を考察することである。最初に、中学生が描いた 1,417 枚のバウム画の評価経験に基づいて、評定者間一致率の低い指標、評定者が定義を理解しにくい指標、定義のより具体化が望まれる指標が特定された。その後、評価の作業手順にはチェックリスト型評価と誘目的指標型評価の 2 通りがあると考えられた。前者は多数の指標から成る網によって全バウムを網羅的に評価し、バウムから意味（仮説）を汲み取ろうとする評価であり、後者はスポットライトを当てるように特定の指標で全バウムを貫いて評価し、仮説（意味）からバウムを捉えようとする評価である。両手順の比較により、数量化研究における評価ではチェックリスト型評価よりも誘目的指標型評価が実用であると示唆された。

キーワード：評価、指標、定義、作業手順

Abstract

This paper aims to consider the problems in Baum test research focusing on the concordance rates in rating and the working procedure for rating from our own rating experience using Koch's 58 indexes. First, based on the rating experience of 1,417 Baums (representations drawn in the Baum test) created by junior high school students, indexes that had low agreement between raters in the rating, were difficult to understand regarding their definitions for the raters, and were expected to contain a more concrete definition were identified. Then, the working procedure of actual rating was considered in two ways, Checklist-type Index Rating and Salient Index Rating, with the former being a working procedure that tries to understand any meaning from the Baums comprehensively through evaluating all Baums with a net of many indexes, and the latter being a working procedure that tries to grasp any meaning from the Baums by spotlighting through all Baums using specific indexes. By comparing the characteristics of the two procedures, it was suggested that Salient Index Rating was more practical than Checklist-type Index Rating for the rating in a quantification study.

Keywords : rating, index, definition, working procedure

1. はじめに

1.1 導入

バウムテストは日本で頻繁に活用されている心理（アセスメント）技法である。刊行されてきた専門書の数が増え、ロールシャッハ法に次いで多いことに、臨床家・研究者の本技法に対する関心の高さを、踏み込んでいえば、本技法の実用性を見ることができよう。

臨床実践においても研究においても、バウムテストで描かれた木（以下、バウム）を理解する手法に、指標を用いるものがある。特に日本における研究では、既存の指標（時に新規の指標）によってバウムを評価し、数量化した後で、集団間でそれらの頻度を比較する方法論が主流であった（佐渡，2011）。しかし、こうした数量化研究（岸本（2015）でいう「指標アプローチ」）には相応の諸課題が付随する。中でも、得られたバウムと採用した指標との合致の判断、つまり評価にまつわる問題は重要かつ複雑である。

多くの心理学研究には、分類・整理・スコアリングなどで得られたデータを処理するプロセスが入る。代表的な投映法であるロールシャッハ法では、以前よりこの点の議論がなされてきた（Baughman, 1958；Meyer et al., 2002 など）。たとえば、反応（産出プロセスとその結果）をどのようにスコア化するか、そのスコアがどれほど汎用性と信頼性を有するか、実施者の熟練度に拠らないスコアリング・システムとはどのようなものか、などについてである。分類やスコアの洗練を目的とした研究がなされる背景に、スコア化を重視する（より強調して言うところ、"スコアありき"の）姿勢を垣間見ることができかもしれない。バウムテスト研究がそうした志向へあまりに流れることを筆者は是としなが、指標という"視点"に一定の価値を認めるならば、やはり指標を用いた研究の精度はより高められるべきである。

1.2 先行研究と問題意識

バウムテストの評価に関わる問題について、次の2編の報告が参考となる。

佐渡（2010）は、日本で報告された数量化研究176編を検討し、評価者間信頼性の確保を試みた研究が23編（13.1%）のみであったことを指摘した。評価者間信頼性の確保がなされた数値のみが信頼できる、というわけでは決していないが、それでもこの指摘は、研究者らの評価作業に対する意識をある程度描出しているといえよう。つまり、先行研究の多くは、評価の慎重さを欠いているかもしれない、知見をできるだけ他者と共有しやすい形で報告をする努力を怠っているかもしれない、読み手が信用しやすい形で結果が書かれていない傾向にある、と。少なからず、先の13.1%という数値はそうした批判の根拠にはなり得る。

中島（2009）の報告は、Koch（1957/2010）の58指

標の判定基準を整理したものである。Kochの指標が正確に日本へ導入されたのは、おそらくこれが最初であろう。Koch自身、テキストの中で指標の判定基準を分かりやすく記してはいないので、中島の報告によりわれわれはKochの指標が意味するところ、そしてKochの指標観を理解する足場を得た。

ところで、指標を使った評価作業を実際に行うと分かるように、それは甚だ複雑で葛藤をともなう作業である。筆者の経験では、データの半分まで評価した後、何かに気づいてもう一度評価をやり直すことは少なくなかった。評価者間での討議で、はじめて自分が定義を誤解していることに気づくこともあった。指標の定義に困惑したり、1枚のバウムにある指標が該当するか否か頭を抱えたりすることは、1度や2度だけではなかった。

数編のバウムテスト研究に携わる中で、筆者は同様の経験をした研究者が多少なりともいることを知った。そして、バウムテスト研究の評価に関わる課題を明確化すること、その課題克服に資する研究の必要性を感じるようになった。

1.3 目的

以上より本研究では、筆者が行ったバウム評価経験それ自体を議論の対象としつつ、次の2点を検討する。第1に、自験例の評価者間一致率を検討し、Kochの58指標のうち、実際の評価と結果の理解とに注意を要する指標を明らかにする。第2に、評価作業の手順について見直しを試みる。なお、示すデータとその検討結果は、検証のためというよりも、議論を進める素材として用いることにしたい。

2. 評価経験の提示

2.1 検討データ

評価の対象となったバウムデータは、1967年と2011-13年に中学生が描いた計1,417枚のバウムである^{註1}。その全バウムをKoch（1957/2010）の58指標を使って、臨床心理士である評価者3名が個別に評価した。その際、各指標の評価基準は中島（2009）に従った。また、その評価作業で用いた用紙は、評価者の目線の動きを考慮して、図1に示すように形態部単位で順番を整理してある。本用紙を使用することで、評価者は1枚のバウムを見ながら、58指標に該当するか否かをチェックする形で評価作業を行った（この作業手順については後で詳しく議論する）。

その後、評価者3名の評価結果を照らし合わせ、2名以上が一致した場合を最終評価とし、各指標に合致したバウムの枚数とその頻度（以下、最終頻度）を求めた。そして、評価者3名すべての評価が一致した（3名中3名が指標と合致した、あるいは合致しなかったと判断した）ものと、3名全員の評価が一致しなかった（3名中

		バウムの ID				
No	指標	1	2	3	4	5
58	空間	上縁はみ出し				
33		幹下縁立				
5	幹①	一線幹				
6		二線幹				
42	幹②	幹上直				
45		幹の瘤や凹み				
17		暗く塗られた幹				
14		幹の中の葉や実				
7	枝①	一線枝				
8		一部一線枝				
9		二線枝				
43	枝②	枝先直				
38		管状枝				
3		直線枝				
1	枝③	全水平枝				
2		一部水平枝				
13		一部低在枝				
12		地面までの枝				
52		三次元(「目」を除く)				
18		暗く塗られた枝				
40		さまよって空間をうめる				
39		さまよった長すぎる枝				
10	分枝	全直交分枝				
11		一部直交分枝				
53		逆向きの分枝				
15	樹冠	樹冠のない幹、付属程度の短い枝のある幹				
16		日輪型や花型				
19		陰影手法の樹冠(枝なし)				
35		球形樹冠				
36		カール状樹冠				
37		もつれた線の樹冠				
20	実	実				
25		空中の実(球形樹冠)				
34	根	まっすぐな根本				
28		一線根				
29		二線根				
21	実葉花	葉				
22		花				
23		大き過ぎる実や葉				
24		黒塗りの実や葉				
26		落下中の、あるいは落下した実、葉、枝				
27		空間倒置				
41		樹冠における主題の変化				
47		ステレオタイプ				
48	その他	留め杭や支柱				
49		梯子				
50		格子で保護、針金				
54		付属品				
56	背景	ほのめかされるだけの風景				
55		多くの風景				
57		島や丘の形				
30	全体	モミ型幹				
31		半モミ型幹				
32		円錐幹				
4		十字型				
44		切断された枝、折れた枝、折れた幹				
46		積み重ね型				
51		変質型				

図1 使用した評価用紙(略)

1~2名のみが指標と合致したと判断した)ものを整理し、前者を「全一致」、後者を「不一致」として、バウムの枚数とその頻度を算出した(以下、この「不一致」の頻度は「不一致頻度」と記す)。

2.2 分析

導き出された値を使い、次の4点を検討する。

検討①:本データにおいて頻繁に認められる(出現頻度が高い)指標と、しばしば評価の評価が分かれる(評価者間の一致率が低い)指標との関連を理解するために、「最終頻度」と「不一致頻度」との間をピアソンの相関係数を用いて検討した。

検討②:評価作業での評価が分かれやすい指標を明らかにするために、「不一致頻度」が15%以上の指標を抽出した。この基準「15%以上」とは、「不一致頻度」が高い指標を特定するための便宜的なものである。

検討③:検討②で「不一致頻度」が高い指標を導き出し得たとしても、当の指標がどれほど好発するのかわかる点(「最終頻度」の高低)に影響を受けているだろう。そこで、「最終頻度」の影響を除いた「不一致頻度」の高低を理解するために、試みとして『「出現頻度」÷「不一致頻度」』の式で算出し、値が3.0以上の指標を抽出した。この「3.0以上」という基準も、特に理論的根拠はもたず、便宜をはかるためである。

検討④:全評価を終えた後、評価者3名に簡易アンケートを実施し、自身の評価経験をもとに58指標の中から「定義の理解が難しい指標」と「不注意で評価し忘れた可能性がある指標」にチェックを求めた。そして、2名以上がチェックしたものを評価者らの意見としてまとめた。

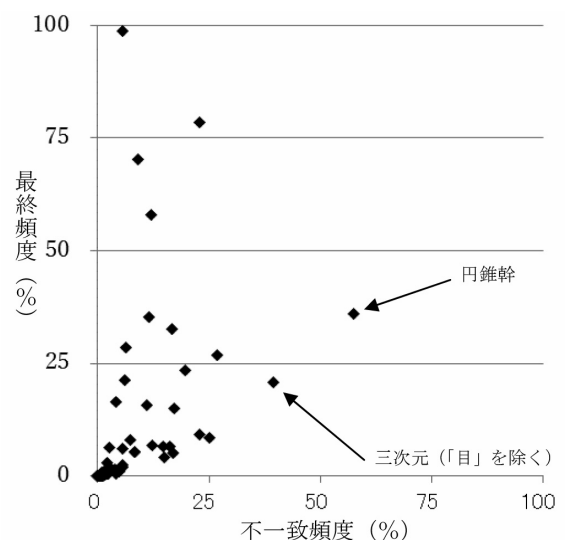


図2 検討①の散布図

表 1 評価結果の整理

No	指標	全一致 枚数 (%)	検討② 不一致 枚数 (%)	15% 以上	最終頻度 枚数 (%)	検討③ 計算式 値	3.0 以上	検討④ 定義 注意	No
1	全水平枝	1394 (98.4)	23 (1.6)		4 (0.3)	0.2			1
2	一部水平枝	1204 (85.0)	213 (15.0)	□	59 (4.2)	0.3			2
3	直線枝	1358 (95.8)	59 (4.2)		9 (0.6)	0.2		A	3
4	十字型	1400 (98.8)	17 (1.2)		2 (0.1)	0.1		A	4
5	一線幹	1413 (99.7)	4 (0.3)		2 (0.1)	0.5			5
6	二線幹	1338 (94.4)	79 (5.6)		1398 (98.7)	17.7	◇		6
7	一線枝	1386 (97.8)	31 (2.2)		42 (3.0)	1.4			7
8	一部一線枝	1137 (80.2)	280 (19.8)	□	334 (23.6)	1.2			8
9	二線枝	1094 (77.2)	323 (22.8)	□	1114 (78.6)	3.4	◇		9
10	全直交分枝	1415 (99.9)	2 (0.1)		0 (0.0)	0.0			10
11	一部直交分枝	1094 (77.2)	323 (22.8)	□	132 (9.3)	0.4		A	11
12	地面までの枝	1384 (97.7)	33 (2.3)		7 (0.5)	0.2		D	12
13	一部低在枝	1188 (83.8)	229 (16.2)	□	93 (6.6)	0.4		A	13
14	幹の中の葉や実	1409 (99.4)	8 (0.6)		2 (0.1)	0.3		A	14
15	樹冠のない幹、付属程度の短い枝のある幹	1406 (99.2)	11 (0.8)		0 (0.0)	0.0		D A	15
16	日輪型や花型	1417 (100.0)	0 (0.0)		0 (0.0)	—		D A	16
17	暗く塗られた幹	1313 (92.7)	104 (7.3)		112 (7.9)	1.1		D	17
18	暗く塗られた枝	1242 (87.6)	175 (12.4)		96 (6.8)	0.5		D	18
19	陰影手法の樹冠 (枝なし)	1401 (98.9)	16 (1.1)		0 (0.0)	0.0		A	19
20	実	1288 (90.9)	129 (9.1)		995 (70.2)	7.7	◇		20
21	葉	1247 (88.0)	170 (12.0)		820 (57.9)	4.8	◇	A	21
22	花	1402 (98.9)	15 (1.1)		9 (0.6)	0.6			22
23	大き過ぎる実や葉	1338 (94.4)	79 (5.6)		34 (2.4)	0.4		D	23
24	黒塗りの実や葉	1298 (91.6)	119 (8.4)		76 (5.4)	0.6			24
25	空中の実 (球形樹冠)	1360 (96.0)	57 (4.0)		235 (16.6)	4.1	◇		25
26	落下中の、あるいは落下した実、葉、枝	1380 (97.4)	37 (2.6)		90 (6.4)	2.4		A	26
27	空間倒置	1386 (97.8)	31 (2.2)		9 (0.6)	0.3		A	27
28	一線根	1338 (94.4)	79 (5.6)		27 (1.9)	0.3		D	28
29	二線根	1261 (89.0)	156 (11.0)		223 (15.7)	1.4		D	29
30	モミ型幹	1035 (73.0)	382 (27.0)	□	382 (27.0)	1.0		D	30
31	半モミ型幹	1061 (74.9)	356 (25.1)	□	121 (8.5)	0.3		D	31
32	円錐幹	601 (42.4)	816 (57.6)	□	512 (36.1)	0.6		D	32
33	幹下縁立	1255 (88.6)	162 (11.4)		501 (35.4)	3.1	◇		33
34	まっすぐな根本	1342 (94.7)	75 (5.3)		28 (2.0)	0.4		A	34
35	球形樹冠	1331 (93.9)	86 (6.1)		303 (21.4)	3.5	◇		35
36	カール状樹冠	1368 (96.5)	49 (3.5)		17 (1.2)	0.3		D	36
37	もつれた線の樹冠	1362 (96.1)	55 (3.9)		22 (1.6)	0.4		D	37
38	管状枝	1173 (82.8)	244 (17.2)	□	212 (15.0)	0.9			38
39	さまよった長すぎる枝	1348 (95.1)	69 (4.9)		15 (1.1)	0.2		D	39
40	さまよって空間をうめる	1396 (98.5)	21 (1.5)		4 (0.3)	0.2		D	40
41	樹冠における主題の変化	1384 (97.7)	33 (2.3)		22 (1.6)	0.7			41
42	幹上直	1396 (98.5)	21 (1.5)		9 (0.6)	0.4			42
43	枝先直	1298 (91.6)	119 (8.4)		76 (5.4)	0.6		D	43
44	切断された枝、折れた枝、折れた幹	1337 (94.4)	80 (5.6)		86 (6.1)	1.1		A	44
45	幹の瘤や凹み	1175 (82.9)	242 (17.1)	□	73 (5.2)	0.3		D A	45
46	積み重ね型	1417 (100.0)	0 (0.0)		0 (0.0)	—			46
47	ステレオタイプ	1360 (96.0)	57 (4.0)		14 (1.0)	0.2			47
48	留め杭や支柱	1405 (99.2)	12 (0.8)		12 (0.8)	1.0			48
49	梯子	1417 (100.0)	0 (0.0)		0 (0.0)	—			49
50	格子で保護、針金	1407 (99.3)	10 (0.7)		1 (0.1)	0.1			50
51	変質型	1402 (98.9)	15 (1.1)		1 (0.1)	0.1		D	51
52	三次元 (「目」を除く)	859 (60.6)	558 (39.4)	□	294 (20.7)	0.5		D A	52
53	逆向きの分枝	1207 (85.2)	210 (14.8)		92 (6.5)	0.4		A	53
54	付属品	1398 (98.7)	19 (1.3)		15 (1.1)	0.8			54
55	多くの風景	1377 (97.2)	40 (2.8)		24 (1.7)	0.6			55
56	ほのめかされるだけの風景	1179 (83.2)	238 (16.8)	□	463 (32.7)	1.9			56
57	島や丘の形	1399 (98.7)	18 (1.3)		7 (0.5)	0.4			57
58	上縁はみ出し	1326 (93.6)	91 (6.4)		406 (28.7)	4.5	◇		58

3. 経験から導き出されたもの

3.1 検討①について

検討①より、「最終頻度」と「不一致頻度」との間に弱い正の相関 ($\gamma = .40$, $\rho = .002$) が認められた (図 2 も参照)。

本結果は、出現頻度の高い指標は評定者間の評価が分かちやすい傾向を幾らか示している。ただし、散布図から分かるように、「円錐幹 [No.32]」と「三次元 (「目」は除く) [No.52]」の 2 指標がこの相関係数に影響を与えている可能性があるため、過度な解釈は避けねばならない。ただし、「円錐幹」と「三次元 (「目」は除く)」は評定作業においてとりわけ注意しなければならない指標である、ということではできよう。

3.2 検討②～④について

Koch の全 58 指標に関する結果を表 1 にまとめた。

検討②で「15%以上」であった指標は 12 個認められ、表には□で強調した。検討③で「3.0 以上」であった指標は 8 個認められ、表には◇を付した。検討④において評定者らに「定義の理解が難しい指標」と見なされた 19 個の指標には D を、「不注意で評定し忘れた可能性がある指標」と見なされた 16 個の指標には A を表に付けた。

検討②による 12 個の指標は、評定者の評価が分かちやすい指標と理解できる。ここで高い値が認められた指標は、評定者が定義を十分理解できなかった可能性 and/or 各指標の定義そのものに検討と具体化の余地が残されている可能性、が考えられる。

検討③による 8 つの指標は、出現頻度が高いために評定での「不一致頻度」の高さが然程目立たなかった指標と理解できる。つまり、これらの指標も評価が分かちやすい傾向をもつ指標であって、検討②と同じ解釈が可能である。

検討④より、今回の評定者らの内省報告を得ることができた。D と A が付された指標は、必ずしも検討②と③の値が高かったわけではない。しかし、この整理結果は評定者らがどの指標で困惑し、自らの評定に自信がなかったかなどを暗示していよう。

以上の検討で、評定の不一致という点で注意を要する指標、評定者の理解度を考慮せねばならない指標を抽出することができた。注意を要するからといって、数量化してきた結果が「意味をなさない」や「大いに疑わしい」とまではいえない。それは拡大解釈というものであろう。本結果は、研究者が評定において配慮を要する指標、論文内で提示された出現頻度を理解する際にも配慮を要する指標を、いくらか特定できたに過ぎない。したがって、今後のバウムテスト研究では、上の結果を考慮し、評定者に対して事前学習を十分設けるなどの工夫が必要となろう。なお、ここまで評定者の熟練度 (バウムテストをどれほど臨床実践で用いており、研究や評定を行う上で

必要な基礎知識を有しているかなど) には触れてこなかった。これは本稿の最後に述べることにしたい。

3.3 評定の作業手順の問題

記述してきた経験の後、筆者は今回用いた作業手順に問題があることに気づいた^{註 2}。

筆者が行った評定作業の具体的な手順は、次のようなものであった。評定者はまず Koch の 58 指標の定義を学び、評定に入る前に、各指標が意味するバウム形態をいくらか思い浮かべることができる状態になる。そして、評定作業で使用した用紙 (図 1) を使い、1 枚のバウムを見ながら、58 指標の中で合致するものすべてにチェックを入れていく。この作業を 1,400 枚超も繰り返していくというものである。この作業手順をここでは、「チェックリスト型評定」(Checklist-type Index Rating) と仮称しておく。

しかしそもそも、この作業手順に無理があったのではないか。58 指標を心にとどめながら 1 枚のバウムを評価するという作業が、如何に複雑かつ大変なことかは、改めて考えると容易に想像できる。1 枚のバウムが必ず 1 つの指標のみに該当するような評定であれば、大きな問題にはならない。けれども、Koch の 58 指標には微細な表現 (Koch はそれらも発達の検討では意味があると考えていたわけだが、彼の指標の設定には催眠実験からの仮説がある) を掬い取るようにも設定されているため、この作業手順ではいずれかの指標に「チェックのし忘れ」や「実感を伴わない評定」が、どうしても生じることになる (先の検討では、それがどの指標に生じるかを明らかにしたとも理解できる)。

以前、筆者はバウムテスト研究の方法論におけるスクリーニング法を批判的に論じたことがある (佐渡, 2010; 坂本ら, 2012)。つまり、多数の指標を、仮説を持たず、網羅的に検討する方法論 (スクリーニング法) では、本当に意味ある連関を議論できない可能性がある、と。今回採用したチェックリスト型評定という作業手順において、筆者はまさにそのスクリーニング法と類似の方法論的誤謬をおかしたことになるかもしれない。指標とは、先達らが創案したものであれ研究者が新たに創案したものであれ、それを用いる者の視点である。すなわち、研究で用いる指標とはその研究者の仮説的着目点である。しかし、研究者の仮説的着目点が本当に存在しているのであれば、その着目点を大事にしつつ、そこから全バウムを見通すことが、評定では重要になると考えられる。

実感を伴った指標からバウムを捉えるという立場に立つと、1 つの指標のみを念頭に全バウムを評価していく、そしてその評定法を、採用した指標の数だけ繰り返していくといった作業手順が考えられる。その場合、作業に使う評定用紙は図 1 とは変わり、用紙の左にバウム ID (描き手の ID) が記されることになるだろう。この作業手順を「誘目的指標型評定」(Salient Index Rating)

と、ここでは仮称しておく。

4. チェックリスト型評定と誘目的指標型評定

4.1 特徴の比較

要約すると、チェックリスト型評定とは、評定に採用したすべての指標を学んだ上で、それら全てを頭におきつつ、1枚のパウムを見て、用いる全指標に合致する可否かを網羅的にチェックしていく方法であり、こうした手続きを全パウムに渡って行っていく作業手順である。一方、誘目的指標型評定とは、採用した指標の内の1つをまず理解した上で、当の指標のみを頭におき、それに合致する可否かを全パウムに渡って判断していく方法であり、こうした手続きを採用した指標の数だけ繰り返していく作業手順である。

試みとして、両手順を視覚的に表すと図3のようになるだろう。チェックリスト型評定は漠とした視点（網羅的な指標、指標の網）によって全パウムを眺め・味わおうとする。そこには意味の"汲みとり"の姿勢がある。かたや、誘目的指標型評定は限られた着目点（具体的な視点・指標）でもって、スポットライトを当てるように、全パウムを理解しようとする。研究者の考え・仮説によって、全パウムを"貫いて"捉えようとする姿勢といえよう。

両者の特徴の差異を一層描出するために、4つの要素を設定して比較してみたい（表2）。

第1の「評定の抜け防止」とは、簡単に言うと、評定時のチェックし忘れを防ぐ要素である。誘目的指標型評定は1つの指標のみを念頭に評価を繰り返すため、「評価し忘れる」という単純ミスを防ぐことが期待できる。たとえば、多くのパウムで認められる「二線幹 [No.6]」などは、あまりに多くのパウムが該当するために、チェックリスト型評定だと作業中にチェックし忘れることが実際にはある（表1で「不一致頻度」が5.6%）。また、「三次元（「目」を除く）[No.52]」もチェックリスト型

評定では見逃されやすいだろう（表1で「不一致頻度」が39.4%）。"評価し忘れ"は、作業で生じるミスであるから、指標それ自体の特徴と理解すべきではない。ともあれ、誘目的指標型評定は人為的な評価の不一致を減らすという意味で、有用といえる。

第2の「採用した指標（仮説）の自覚」は、もっとも大切な要素であると筆者は考えている。指標を使って評定するとは、当然ながら、何故その指標を採用したのかに関する自覚をある程度持つことになる。その指標は、何がしかの心理学的意味と連関があると考え、その指標を採用することで研究者（絵の受け手）が"意味がありそうだと感じた表現を捉えることができると考えられており、その指標によって評価するに足る意義があるとの前提が存在するはずである。したがって、誘目的指標型評定によってある指標から全パウムを通して捉えることは、研究者の仮説の洗練のみならず、その指標を使って理解する事象への新たな気づきにもつながるかもしれない。

第3の「評定のやり直しのしやすさ」は、何かの指標に関して自らの誤解を認めた際、もう一度評定作業する時の効率の良さを表す。第1節で言及したように、作業の途中で評定をやり直すことは少なくない。誘目的指標型評定では、やり直すとしても特定の指標のみを行えばよく、それ故に実用的である。また、多くのパウムを評定していくと自らの指標の理解が揺らぎ始め、そのことを通して指標の理解が促進していく。誘目的指標型評定は焦点づけることを特徴とするため、指標の理解についての自らの誤りに気付きやすいと思われる。

第4の「作業にかかる時間」はそのまま、評定者が作業を終わるまでの時間を意味する。筆者の経験上、チェックリスト型評定よりも誘目的指標型評定の方が1.5倍ほど長くかかる。研究では、臨床心理士資格保有者（あるいはパウムテスト実践者）の作業時間を確保するという難点があり、仮にその評定者らに対して謝礼を払うので

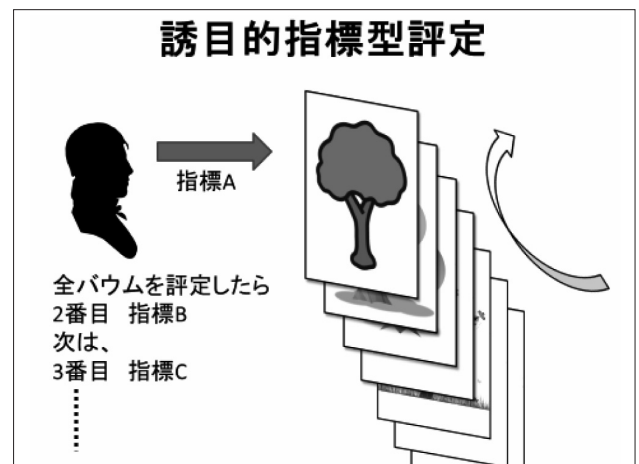
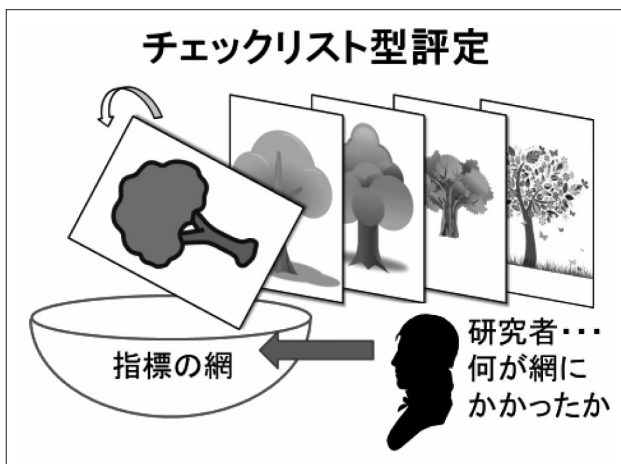


図3 チェックリスト型評定と誘目的指標型評定のイメージ図

表2 2つの評価手順の比較

チェックリスト 型評価		誘目的指標 型評価
1)	— 評価の抜け防止	+
2)	— 採用した指標（仮説）の自覚	+
3)	— 評価のやり直しのしやすさ	+
4)	— 作業にかかる時間	+

あれば、誘目的指標型評価では研究費をより必要とする（研究費の話は決して下賤なものではなく、研究遂行を考えていく上で極めて現実的な部分でもある）。誘目的指標型評価はここにデメリットを見ることができるが、第1～3のメリットと天秤にかければ、採用するに足る作業手順ではないだろうか。

以上より、バウムというそれ自体全体的な表現を、あえて別の形（たとえば数値）に置き換える評価という作業を行う場合、誘目的指標型評価には十分なメリットがあると分かる。また、そのメリットは、導きだされた結果を、研究者間でより共有しやすくすると考えられる。

4.2 再び自験例より

実際にチェックリスト型評価と誘目的指標型評価はどのように異なるのか、という疑問が生じる。そこで、誘目的指標型評価で行った筆者の評価経験も提示したい。この研究は子どものバウム表現の比較文化的検討である（Sado et al., 2015）。

評価の対象は、フランスの子ども（5～12歳）のバウム442枚と、ニューカレドニアの子ども（4～12歳）のバウム459枚の、計901枚である。

評価は、第2節の中学生のものとはほぼ同じであるが、誘目的指標型評価により評価したこと、評価者の内2名が入れ替わったことに相違がある。この2点の相違は評価（とその結果）において決定的なものだが、それ以上に提示・比較する価値を重視したい。ただし、本データの限界を意識しつつ、議論を進めていく。^{註3}

先のチェックリスト型評価とここでの誘目的指標型評価における「不一致」の数と頻度を表3に示した。さらに、両者の「不一致頻度」の差を見るために、「チェックリスト型評価－誘目的指標型評価」の式で算出し、値を表の右端に加えた。結果、幾つかの指標が誘目的指標型評価の導入により「不一致頻度」の低下をみせたが、逆に増加している指標もあった。その全般的特徴を捉えるために、全58指標の「不一致頻度」の平均値を求めると、チェックリスト型評価で8.4（標準偏差＝10.58）、誘目的指標型評価で6.3（標準偏差＝5.85）であり、誘目的指標型評価の平均値の方が若干の低い。しかし、両平均値をt検定により分析すると、有意差はなかった（ $t=1.32$, $df=114$, n.s.）。

全58指標の「不一致頻度」の平均値は両作業手順で

表3 両作業手順の評価経験の比較

No	指標	チェックリスト型評価 n = 1,417 枚数 (%)	誘目的指標型評価 n = 901 枚数 (%) %の差	
1	全水平枝	23 (1.6)	21 (2.3)	-0.7
2	一部水平枝	213 (15.0)	62 (6.9)	8.1
3	直線枝	59 (4.2)	103 (11.4)	-7.2
4	十字型	17 (1.2)	17 (1.9)	-0.7
5	一線幹	4 (0.3)	19 (2.1)	-1.8
6	二線幹	79 (5.6)	25 (2.8)	2.8
7	一線枝	31 (2.2)	98 (10.9)	-8.7
8	一部一線枝	280 (19.8)	88 (9.8)	10.0
9	二線枝	323 (22.8)	117 (13.0)	9.8
10	全直交分枝	2 (0.1)	10 (1.1)	-1.0
11	一部直交分枝	323 (22.8)	78 (8.7)	14.1
12	地面までの枝	33 (2.3)	24 (2.7)	-0.4
13	一部低在枝	229 (16.2)	46 (5.1)	11.1
14	幹の中の葉や実	8 (0.6)	14 (1.6)	-1.0
15	樹冠のない幹、付属程度の短い枝のある幹	11 (0.8)	137 (15.2)	-14.4
16	日輪型や花型	0 (0.0)	27 (3.0)	-3.0
17	暗く塗られた幹	104 (7.3)	26 (2.9)	4.4
18	暗く塗られた枝	175 (12.4)	20 (2.2)	10.2
19	陰影手法の樹冠（枝なし）	16 (1.1)	33 (3.7)	-2.6
20	実	129 (9.1)	77 (8.5)	0.6
21	葉	170 (12.0)	87 (9.7)	2.3
22	花	15 (1.1)	21 (2.3)	-1.2
23	大き過ぎる実や葉	79 (5.6)	7 (0.8)	4.8
24	黒塗りの実や葉	119 (8.4)	38 (4.2)	4.2
25	空中の実（球形樹冠）	57 (4.0)	39 (4.3)	-0.3
26	落下中の、あるいは落下した実、葉、枝	37 (2.6)	12 (1.3)	1.3
27	空間倒置	31 (2.2)	98 (10.9)	-8.7
28	一線根	79 (5.6)	43 (4.8)	0.8
29	二線根	156 (11.0)	35 (3.9)	7.1
30	モミ型幹	382 (27.0)	45 (5.0)	22.0
31	半モミ型幹	356 (25.1)	68 (7.5)	17.6
32	円錐幹	816 (57.6)	102 (11.3)	46.3
33	幹下縁立	162 (11.4)	80 (8.9)	2.5
34	まっすぐな根本	75 (5.3)	169 (18.8)	-13.5
35	球形樹冠	86 (6.1)	160 (17.8)	-11.7
36	カール状樹冠	49 (3.5)	66 (7.3)	-3.8
37	もつれた線の樹冠	55 (3.9)	20 (2.2)	1.7
38	管状枝	244 (17.2)	53 (5.9)	11.3
39	さまよった長すぎる枝	69 (4.9)	40 (4.4)	0.5
40	さまよって空間をうめる	21 (1.5)	31 (3.4)	-1.9
41	樹冠における主題の変化	33 (2.3)	16 (1.8)	0.5
42	幹上直	21 (1.5)	213 (23.6)	-22.1
43	枝先直	119 (8.4)	65 (7.2)	1.2
44	切断された枝、折れた枝、折れた幹	80 (5.6)	75 (8.3)	-2.7
45	幹の瘤や凹み	242 (17.1)	29 (3.2)	13.9
46	積み重ね型	0 (0.0)	202 (22.4)	-22.4
47	ステレオタイプ	57 (4.0)	42 (4.7)	-0.7
48	留め杭や支柱	12 (0.8)	1 (0.1)	0.7
49	梯子	0 (0.0)	0 (0.0)	0.0
50	格子で保護、針金	10 (0.7)	0 (0.0)	0.7
51	変質型	15 (1.1)	15 (1.7)	-0.6
52	三次元（「目」を除く）	558 (39.4)	6 (0.7)	38.7
53	逆向きの分枝	210 (14.8)	47 (5.2)	9.6
54	付属品	19 (1.3)	26 (2.9)	-1.6
55	多くの風景	40 (2.8)	56 (6.2)	-3.4
56	ほのめかされるだけの風景	238 (16.8)	207 (23.0)	-6.2
57	島や丘の形	18 (1.3)	11 (1.2)	0.1
58	上縁はみ出し	91 (6.4)	10 (1.1)	5.3

差は無かったが、これは一見、誘目的指標型評定の有用性を否定もするものであろう。しかし、誘目的指標型評定の「不一致頻度」で15%以上であった指標を見ると、「樹冠のない幹、付属程度の短い枝のある幹 [No.15]」、「まっすぐな根本 [No.34]」、「球形樹冠 [No.35]」、「幹上直 [No.42]」、「積み重ね型 [No.46]」、「ほのめかされるだけの風景 [No.56]」の6個のみで、チェックリスト型評定での12個よりも大幅に減少している。この点をもってして、誘目的指標型評定の有用性を見ることが許されるのではないだろうか。

もちろん、この作業手順結果の比較は、評定者の違いや検討対象データであるバウムの特徴の違いというものが影響を与えているはずである。それが、一方の「不一致頻度」の低下に、あるいは増加に、しかもある一部の指標にのみ現れた可能性もある。したがって、ここで提示した検討結果は誘目的指標型評定の有用性や実用性を実証的に示すものではなく、その利点を強調する立場から筆者の経験を記述したものである。

4.3 どのように活用するか

では、両作業手順をわれわれはどのように活用するのがよいのか。

チェックリスト型評定には意味の"汲みとり"の姿勢があると記した。したがって、この手順は本来的には仮説の生成のために用いるのが有益かもしれない。絵を味わい意味を探す方法論として、チェックリスト型評定のようなやり方は有効である。一方、誘目的指標型評定は"貫いて"捉えようとする姿勢と述べた。つまり、こちらの手順の特徴は、全体的把握を括弧に入れた上での部分的明確化と理解できるため、この手順は実際の評定作業において実用的になる。

このように、両作業手順にはまったく異なる方向性があり、異なる哲学がある。両者を混同すべきではない。作業手順のあり様と、どの作業手順を採用するかには、研究者のスタンスが滲み出る部分である。

チェックリスト型評定と誘目的指標型評定とが異なるものであるにも関わらず、どうして混同されて評定に用いられてきたのであろう^{註4}。スポットライト分析を提唱した論文でも記したように、先行研究では多くの指標を使って検討する方法論が大勢を占めていた（表4参照）。その背景に、指標選択時の仮説と自覚の薄さなどを考えることができる。それと同じスタンスが、チェックリスト型評定による評定に見ることができるだろう。チェックリスト型評定のようなバウムへのアプローチは決して否定されてはならないが、いざこの手順での評定となると、研究者の指標観が背景に押しやられ、数値化という目的が前景化するのだと考えられる。バウムから仮説を得たり、バウムを味わったりする時、Koch (1957/2010, p.20) が当初から強調していたように、バウムそのものを中心に置き、バウムがわれわれに語りかけてくるまで

表4 先行研究において検討に用いている指標の数
(坂本ら, 2012 より)

指標数	論文数 (%)
1～10	31 (18)
11～20	19 (11)
21～100	90 (51)
101 以上	28 (16)
不明	8 (5)
計	176

待つ必要がある。しかし、その語りかけてきたものを大切にしつつ、ある次元で別の形へと移し替えていくためには、細心の注意をもって選択された指標が中心に置かれるべきあり、その指標を抛り所としてバウムを捉える必要があるだろう。このように考えると、チェックリスト型評定と誘目的指標型評定とは、特に数量化研究の場合、受け取り手（研究者）とバウム（および描き手）との関わりのフェーズの差異と理解できるかもしれない。

5. おわりに

5.1 まとめ

本稿では、筆者の評定経験における評定者間の一致率から、定義の厳密化と再考が求められる指標、評定者らの評価が分かれやすい指標、評定に慎重になるべき指標、論文で提示された数値の理解に注意を要する指標を特定した。その後、チェックリスト型評定と誘目的指標型評定という評定の作業手順をとりあげ、両特徴を論じた。

評定の具体的な作業手順は、研究遂行に重要な部分である。しかし、これについて論じた報告を、筆者は寡聞にして知らない。少なからず、千編を超えようとしている日本のバウムテスト研究の中で、指標に関する議論は若干あるのみで（津田, 1980a&b；一谷・津田, 1982）、評定の作業手順を論じた研究を見つけることはできない。数量化研究に対して即座にアレルギー反応を示すのではなく、評定そのものの意味を問う研究や、評定の作業手順を考え直す研究は、もう少しなされても良いように思える。

5.2 残された課題

本研究はいくつかの限界を抱えている。今後議論が望まれる部分に焦点を当てて言及しておく。

第1に、評定者の性質を不問にしている点である。バウムテストを実践に使っており、指標なども熟知している評定者を、研究者がいつも十分に確保できるとは限らない。そもそも評定者の性質を研究間で同一にすることも不可能でもあし、臨床歴によって評定者の質を担保しようとするのは強引すぎる。したがって、筆者はこの部分の検証に限界を感じてはいるが、評定者の性質が評定

結果に及ぼす影響に関して、今後検討が求められるかもしれない。

第2は、評定の作業手順の中でも、評定者間信頼性の部分である。実際の評定には何人の評定者が加わるべきなのか。1名であれば、その研究者の仮説に最も符合した表現を評価できるかもしれないが、その結果は時に"評価の信頼性が弱い"と批判される。3名であれば、評定結果の信頼性を幾らか担保できるが、その研究者の仮説に符合した表現を評価するという点で開きが生じる(指標の定義の問題ともいえるが、「不一致頻度」がどうしても出てくる)。また、数名の評定結果を突き合わせた時、どこまでを最終評定とするかの統一見解も今のところない。3名中3名が一致した時のみを最終評定とするよりも、"民主的に多数決"とするのが現実的かもしれない。こうした評定の方法論やプロトコルについては、一応のマニュアルがあっても良いように思われる。

第3に、Kochの58指標以外を対象としていない点である。指標の種別で評定結果の特徴は異なるだろう。この部分も今後検討されるべきかもしれないが、究極的には、指標は研究と研究者の数だけ生まれうる。本研究ではバウムテストを体系化したKochの指標群にのみ焦点を当てたにすぎないが、参考となる数値はできる限り提示するよう努めることで、本稿自体が批判の対象となることも狙った。

最後に、本研究では評定について考えてきたが、その過程で指標について考え直す研究の必要性も感じた。その場合、われわれとバウムと描き手にとって指標とは何か、という問い立てがなされるかもしれない。そうした研究は、バウムテストという技法・領域の発展のためだけでなく、バウムテスト実践・研究に取り組む人々を支援するものにもなると思われる。機をみて取り組んでみたい。

註 釈

- 1) このデータは、佐渡・岸本・山中(2014)として報告している。
- 2) これは、ある研究会議での岸本寛史氏からの指摘に拠っている(岸本, 2014)。また、以後に記す評定の作業手順の命名も、岸本氏からの助言に負っている(岸本, 2016)。
- 3) 両作業手順の特徴を厳密に比較するには、同じバウムデータを使い、同一の評定者らが、カウンターバランスをとってチェックリスト型評定と誘目的指標型評定とで評定し、両結果で評価が分かれる程度を比較することが求められるだろう。そうした研究は確かに厳密さを持つが、それを行うことが本稿の目的ではない(また、そのためだけに評定を行う意味を筆者は感じなかった)。
- 4) 他の研究者との個人的対話から、本問題が筆者にのみ生じたものではなく、多くのバウムテスト研究、

描画法研究で生じていることは明らかである。ただし、それを証明する文献やデータは存在しない。

付 記

貴重な示唆と助言をくださった岸本寛史氏に感謝申し上げます。

本研究は、日本心理臨床学会第34回秋季大会でのポスター発表を整理・修正して、まとめ直したものである。なお、科研費JSPS:26780382の成果の一部でもある。

文 献

- Baughman, E. E. A New Method of Rorschach Responses, *Psychological Bulletin*, 55, 121-147, 1958.
- 一谷 彊・津田 浩一 「『バウム・テスト整理表』の作製とその具体的利用」『京都教育大学紀要 Ser.A』61巻、1-22頁、1982年 [一谷 彊・林 勝造・国 吉政一(編著)『バウムテストの基礎的研究』風間書房、431-477頁、1985年、所収]
- 岸本寛史「Personal Communication」於、高槻赤十字病院、2014年2月2日
- 岸本寛史『バウムテスト入門——臨床に活かす「木の絵」の読み方』誠信書房、2015年
- 岸本寛史「佐渡へのe-mail」2016年8月19日
- Koch, K. *Der Baumtest: der Baumzeichenversuch als psychodiagnostisches Hilfsmittel*. 3.Auflage. Bern: Hans Huber. 1957. 岸本寛史・中島ナオミ・宮崎忠男(訳)『バウムテスト 第3版——心理的見立ての補助手段としてのバウム画研究』誠信書房、2010年
- Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E. Jr., Fowler, J. C., Piers, C. C., and Resnick, J. An Examination of Interrater Reliability for Scoring the Rorschach Comprehensive System in Eight Data Sets. *Journal of Personality Assessment*, 78(2), 219-74, 2002.
- 中島ナオミ「コッホのドイツ語原著における58指標の判定基準」『関西福祉科学大学紀要』12号、71-90頁、2009年
- Sado, T., Yama, M., Kishimoto, N., Baratgin, J. & Jamet, F. Cross-cultural Study on Children's Representations in the Baum Test: A Statistical Perspective. The 5th International Conference of Expressive Psychotherapy. August 3-6, 2015.
- 佐渡忠洋「実施法と評定間信頼性からみたバウムテスト研究の精度——バウムテスト文献レビュー(第二報)」『岐阜大学カリキュラム開発研究』第28巻1号、21-32頁、2010年
- 佐渡忠洋「バウムテスト研究の可能性」In; 岸本寛史(編)『臨床バウム——治療の媒体としてのバウムテス

- ト』誠信書房、28-43 頁、2011 年
- 佐渡忠洋「日本におけるバウムテストの研究」『臨床心理学』第 10 巻 5 号、674-679 頁、2010 年
- 佐渡忠洋・岸本寛史・山中康裕「今昔の中学生のバウムテスト表現の検討——1960 年代と 2010 年代との発達指標を通して」『明治安田こころの健康財団 研究助成論文集』第 49 号、77-86 頁、2014 年
- 坂本佳織・佐渡忠洋・岸本寛史「バウムテスト研究におけるスポットライト分析」『心理臨床学研究』第 30 巻 1 号、41-50 頁、2012 年
- 津田浩一「「バウム・テスト整理表」の紹介（Ⅰ）」『心理測定ジャーナル』第 16 巻 9 号、17-22 頁、1980a 年
- 津田浩一「「バウム・テスト整理表」の紹介（Ⅱ）」『心理測定ジャーナル』第 16 巻 10 号、2-8 頁、1980b 年