

評定尺度はリカートが発明したわけではない —リカート(1932)より前の世界—

椎名乾平 (早稲田大学)

キーワード：評定尺度，歴史，心理測定

現在，評定尺度と言えば事実上 Likert の方法をさす。Likert (1932, A technique...)以前の評定尺度の有様について調べた。以下，文献記述は不完全だがすべて検索可能なのを確認してある。そもそも Rating scale を「評定尺度」と訳したのは増田幸一(1927, 心理学研究, p. 129)のようである。Likert 法には理論的根拠があるが(椎名, 教心予稿集, 2016), それ以前にも同様のデータ収集法が強い根拠なしで多用されていたのがわかった。尚，現在の使用法と直接繋がらない古い例は McReynolds & Ludwig (1987, On the history of rating scales...)に詳しい。現在と同様の評定尺度は，19 世紀末頃から，以下の 5 領域，A：遺伝・知能，B：教育測定，C：人事査定・職業適性，D：性格，E：態度，でかなり大々的に使用されていた。Garrett & Schneck (1933, Psychological Tests...)では，評定尺度が 4 分類されている。すなわち，R1) Man-to-Man Rating Scale ある特性について 5 人の具体的人物を列挙しその人物に関係付けて評定する。現在は全く使われていないと思われる。下記 C：参照 R2) Graphic Rating Scale (下記は Bills, 1923 より)

QUALITIES	REPORT				
I. APPEARANCE. Consider neatness of person and dress.	Appropriate	Neat	Ordinary	Possible	Slowly only

目盛は等間隔である必要はなく，また線上のどこをマークしても良い。

R3) Numerical Rating Scale パーセントを数値で答える方法。廃れて来ていると言う。

R4) Descriptive or Adjective Rating Scale 現在の Likert 法とほぼ同じもの。すなわち

The method here is to use common descriptive terms, or broad categories, such as "excellent," "good," "average," "poor," in rating an individual. Thus a teacher's appearance may be rated by

分野別に使用例を挙げれば以下のようなだろう。

A：遺伝・知能 Galton (1869, *Hereditary genius*, p.34)の才能による人間等級化は有名だが (R3 に近い)，実データを取ったわけではないので，評定尺度と呼べるものの最初の使用例は Statistics of mental imagery (1880, *Mind*)であろう(R4)。心像の明瞭性について 9 段階尺度で聞いている。Pearson (1903, On the Inheritance of the Mental..., p.210)では下図のような 5 段階評定尺度が使われている (R4。ここでは姉妹の健康状態の比較が目的である)。その後の Pearson の心理学的研究でも似たようなデー

I. PHYSIQUE:

	Very Strong.	Strong.	Normally Healthy.	Rather Delicate.	Very Delicate.
ELDER SISTER ...					
YOUNGER SISTER ...					

タが扱われている。

B：教育測定 周知のとおり 20 世紀初頭に教育測定運動が Thorndike を主導者にして興隆したが，ここでも評定尺度が用いられた。Bradshaw (1930, American Council on Education Rating Scale: its reliability...)に総説がある。また，評定尺度作成・使用についての留意点 39 項目がまとめられている。この論文では教師の人事評価に用いられた Miner (1917, The evaluation of a method for finely graduated) の評定尺度 (次図) が重要なものとされている。

	Lowest 5th	Fourth 5th	Middle 5th, Average	Second 5th	Highest 5th
Common sense					
Energy					
Intuitive					
Leadership					

さらに古く Boyce (1915, Methods for measuring teachers' efficiency)では，教師の勤務評定用紙(Score Card)の中で評定尺度が広く使用されたことが述べられており(R4)，各用紙で用いられたカテゴリー段階数についての集計結果まである。ちなみに 2, 3, 4, 5, 6, 7 段階尺度がそれぞれ 2, 11, 24, 12, 4, 1 回使用されたそうである。

C：人事査定・職業適性 第一次世界大戦で Yerkes が主導して知能テストが作成・実施されたのは良く知られているが，これと平行して Scott, W.D.の指導の下に軍隊用適性(職能)検査が作成された。米国陸軍の評価としては後者の方が高かったようである。具体的には Army Rating Scale (これが R1)と，Graphic Rating Scale が開発され利用された。戦後，Scott は Scott Company を設立して，これらの手法を民間企業に広めた。また遠藤公嗣(1999)によれば，大戦後 1924 年より ↓ R3 Paterson(1923, The Scott Company graphic rating...)。

QUALITIES	REPORT				
I. Ability to Learn: Consider the ease with which the employee is able to learn new methods and to follow directions given him.	Very Superior	Learns With Ease	Ordinary	Slow To Learn	Dull
II. Quantity of Work: Consider the amount of work accomplished and the promptness.	Unusually	Satisfactory	Limited	Unsatisfactory	

米国連邦政府は公務員の勤務評定法として Graphic Rating Scale を使用し始めた (ただし 1935 年に廃止)。お上のお墨付きがついたこともあり評定尺度(特に R2)は様々な産業領域で使われるようになったと考えられ，現在日本の「人事考課」にも繋がるという。

D：性格 Pearson にも性格特性の研究といえるものがあるが，本格的性格心理学での評定尺度の嚆矢は Webb (1915, Character and intelligence)であろう。非常に丁寧に用心深く評定尺度を使用しているのが現在でも参考になる。使われているのは 7 段階尺度であるが符号付数値

(2) In the columns under each subject's name place one of the marks
+ 3, + 2, + 1, 0, - 1, - 2, - 3
for each of the qualities specified. To avoid errors, please put the

を直接回答させているようである。また結果がなるべく正規分布になるように回答に分布制限を定めている。

E：態度 他の研究分野よりやや遅れて，1920 年代から態度研究が盛んになる。教育測定運動の尺度にも態度系のものがある(佐藤隆之，2018，市民を育てる学校)。

結論 20 世紀初頭，評定尺度は広く深く使用されていた。当時，実験計画法や統計的検定はまだ普及しておらず，ここで挙げたすべての研究が心理特性の測定あるいはその相関を念頭においたものである。わが国でも米国の動向に 4, 5 年程度の遅れで同じように研究され実際に使用されている。主導者は上野陽一，増田幸一である。