

展 望

概念的妥当性の検証

—心理測定学的構成概念と認知心理学的構成概念の場合—

並 木 博

(元早稲田大学教育学部・慶應義塾大学文学部)

Construct Validation in the Constructs of Psychometrics and Cognitive Psychology

Hiroshi NAMIKI

(FORMERLY DEPARTMENT OF EDUCATION, WASEDA UNIVERSITY and DEPARTMENT OF LITERATURE, KEIO UNIVERSITY)

Since the publication of Cronbach and Meehl's (1955) article on construct validity in terms of a nomological network, this methodological approach has been widely acknowledged by researchers in psychology. When Cronbach (1989) reviewed the evolution of the concept 30 years later, he suggested a realistic strategy for improving construct validation. During that time, the general trend of psychology had drastically changed into an information-processing approach. Embretson (1983) proposed a new method of construct validation based on cognitive theorizing, calling it "construct representation". In the present article, following a discussion of the history of construct validation, the author reviews available methods for construct validation, including one proposed by Kyllonen (1996). A summary is presented of the present author's development of a neuropsychological diagnostic test for dementia based on working-memory theory and item-response theory. A new method of construct validation, called "a mutually compensatory method of construct validation", is proposed that would effect a synergy of the psychometric and cognitive approaches.

Key Words : construct validity, nomological network, item response theory, working memory, dementia

1955年に Cronbach and Meehl が法則定立的ネットワークによる概念的妥当性検証に関する論文を公にして以来、その方法論上の重要性が心理学研究者によって広く認められるようになった。30年後に Cronbach はこの考え方の展開を展望し、概念的妥当性検証の進歩のための現実的方略を提言した。その間、心理学全般の動向は、情報処理的アプローチへと大きく変貌を遂げた。Embretson は認知心理学的理論化に基づいて概念的妥当性検証の新しい方法を提唱し、それを“構成概念表現的アプローチ”と呼んだ。筆者は概念的妥当性検証の以上のような歴史の流れを展望した後、Kyllonen の最新の挑戦的研究を含めて、これまで用いられてきたその検証の方法を検討した。最後に、筆者らの行った作動記憶理論と項目反応理論に基づく認知症診断のための神経心理学的診断検査の開発の概要を報告し、また“概念的妥当性の相補的検証手続き”と名づけた方法に言及したが、これは心理測定学的、および認知心理学的アプローチの協同を企図したものであった。

キーワード：概念的妥当性 法則定立的ネットワーク 項目反応理論 作動記憶 認知症

1. はじめに

概念的妥当性は、1つの心理学的検査によって、その測定対象である構成概念がどの程度正確に捉えられているかを問うものであり、これまで提案された各種の妥当

性の中で、科学基礎論的に最も奥行き深い考え方であり、またより基礎的、より包括的、そしてより理論的な妥当性の概念である。また、基準関連の妥当性や内容的妥当性は、いずれも概念的妥当性に集約されるのであり、

その相異なる側面に過ぎないとも考えられる (Messick, 1995)。Cronbach and Meehl (1955) によって概念的妥当性に関する周到な議論が行われてすでに半世紀の歳月が流れたが、さらに Cronbach (1989) は “Construct Validation After Thirty Years” と題してその後の研究の進展を展望した。その間、心理学全般の流れは、認知的、情報処理的アプローチへと大きく変わり、心理検査の領域も例外ではなく、それまでのテスト理論に基づく検査得点の分析にとどまらず、検査課題に対して被検査者の反応が生じるまでのメカニズム、あるいはプロセスを認知心理学的理論によって解析することが盛んに行われるようになった。したがって、概念的妥当性検証の手続きも、それまでの主として因子分析法に基づく心理測定学的構成概念の場合とは異なる方法による可能性が生じてきた。

本稿においては、概念的妥当性をめぐる議論について、従来の心理測定学的構成概念に関するものと、新しく台頭してきた認知心理学的理論に基づくものとを考察した後、従来型の概念的妥当性検証の方法を網羅的に眺める。その中で特に、一般知能因子 g の概念的妥当性に関する研究成果をいくつか挙げる。また、新しい認知心理学的構成概念として、作動記憶を取り上げ、一般知能因子 g との関連性を実験によって探った研究を紹介する。最後に “概念的妥当性の相補的検証手続き” と筆者が名づけた新しい試みを、作動記憶理論と項目反応理論に基づく認知症診断のための神経心理学的検査の開発の経過の概要とともに報告したい。なお以下では、概念的妥当性 construct validity を CVy, 概念的妥当性検証 construct validation を CVn, また両者の区別の必要のないときにはただ CV とそれぞれ略記してスペースの節約を図ることをお許し願いたい。

II. 概念的妥当性をめぐる議論

Cronbach and Meehl (1955) の CVy に関する周到かつ精緻な議論には啓蒙的な意義があるので、その主要な部分、とりわけ法則定立的ネットワークに重点を置いて紹介したい。まず CVn に関して、彼らは次のように述べている。

CVn が問題になるのは、操作的に定義されていない属性、あるいは性質の測度としてあるテストを解釈すべきときである。研究者が直面する問題は、どのような構成概念がテスト成績の分散を説明するかである。CVy は何か新しい科学的アプローチを必要としているわけではない (p. 282)。

次に、テスト得点と基準との相関が高いことが CVy の裏づけにならないことを強調して次のように言う。

CVn は、ネットワークの中のある陳述が、観察デー

タ間の予測された関連性のとおりになって初めて容認される。ある観察データが “基準” と見なされるとして、基準自体の CVy が研究途上のものと考えられる (p. 300)。

次に、法則定立的ネットワーク (nomological network: 以下 NNW) がこの論文の中核になっているので、その基本的原理 6 カ条の要旨を述べておきたい。

- (i) 科学的に言うて、“何かが何であるかを明確にする” とは、それがその中で生じる法則を述べることを意味している。私たちは NNW として、理論構成の重なり合う法則体系といったものを提言したい。
- (ii) NNW の中の法則は以下のものを関連づける。①観察可能な属性あるいは量を相互に、②理論的構成概念を観察可能なものに、③相異なる理論的構成概念を相互に。これらの “法則” は統計学的、あるいは決定論的なものであろう。
- (iii) 構成概念が科学的に容認されるための必要条件は、少なくともその法則の中に観察可能な変数を含むものがあるような NNW の中で生じることである。また、NNW の中の法則性が観察可能な属性を相互に関連づけることから、ある出来事の生起を理論的に予測できることになる。さらに、構成概念は観察データに “還元される” のではなく、NNW の中の他の構成概念と結合して観察可能なデータを予測するに過ぎない。
- (iv) 理論的構成概念が “よりよくわかるようになる” とは、それが生じる NNW を精緻にする作業であり、またその構成要素の確かさを高めることである。少なくとも、1 つの構成概念の歴史の初期には NNW は限られており、構成概念の結びつきの数もわずかであろう。
- (v) 構成概念を加えたり、理論に関連性を加えたりして NNW を豊かにすることは、もし観察データで裏づけられる NNW を生じたり、あるいは同じ観察データを予測する NNW の数を減らしたりするならば、正当化される。もし観察データが NNW への当てはまりが悪いときには、他の構成概念を考えたり NNW の再組織を考えることで修正する。
- (vi) 質的にまったく異なった “操作” が、NNW の中の位置において同じ構成概念に結びついていれば、同じものを “測定” しているといえる。このように考えても疑義が生じないのは、2 つの操作が同じ NNW で定義された量の推定をしているという NNW 内の証明があれば、私たちはそれでよしとするからである (pp. 290-291)。

最後に、心理学的検査がいろいろな批判に耐えるため

の基本的な姿勢について、次のように述べている。

その答えは次のようになる。NNW が観察データと接点を持ち、推論の明示的かつ公共的なステップを示さない限り、CVy ありとは主張しえない (p. 291)。

Cronbach (1989) は、その後の CVy の進展を展望してまず次のように述べている。

CV の考え方は進化した。CV は妥当性検証の別のスタイルとして見られうるようになった——内容の分析や予測力の結果が妥当性を支持しえないときの最後の手段としての場合がほとんどだが。今日では CV は他のアプローチが依拠する基盤として見られている。

基準関連の妥当性検証においては、基準について不純物が混在していないか、何か欠如している成分がないかを探る必要がある。つまり基準の CVn が求められている。

内容的妥当性の検証は、テストがその“内容の明確化”と一致することを明らかにできれば十分である。しかし“内容の明確化”がよく吟味されているといえるならば、これは CV を具体的に主張することになる (p. 151)。

つまり、基準関連妥当性も内容的妥当性も、結局測定内容の概念を確認することに尽きると主張しているのである。次に、Cronbach は construction という用語を初めて用いたが、筆者はこれを「構成概念による解釈」と訳した。

テスト A は妥当性がある、テスト B は妥当性がないという言い方は非論理的である。特定の解釈こそが妥当性の裏づけである。“これまで得られた証拠は解釈と一致している”という言い方のほうが、“このテストは CVy がある”という言い方よりもはるかに意味がある。妥当性の検証は息の長い、果てしない過程である。テストの解釈がその根拠としてしっかりした理論を持っていることはまずない。ほとんどの場合、未完成の理論のスケッチに依拠している。テストの解釈においては、概念と含意をゆるい形でまとめたものを用いるのが典型的であるが、私はこれを“理論”と呼ぶよりもむしろ“構成概念による解釈 (construction)”と呼びたい (pp. 151-152)。

次に Cronbach は、テストへの反応に至るまでのプロセスに言及しており、時代の大きな流れを知ることができる。

テストの遂行行動の説明は、反応する人のプロセス、あるいはスタイルに依存している。したがって、あるテスト得点を 1 つの説明で済ますことは適切ではない。いくつもの説明がありうることを認めねばならないし、そうであればその人がなぜそのプロセスを用いるかを

説明しなければならない。

認知心理学が、攻撃的にスタイルとストラテジを探索するとすれば、テスト得点の解釈の適切なルールがあらゆる被験者に同じではない、あるいは同じ被験者についてすべての機会において同一ではないことが次第に明らかになるであろう。私たちは、情報処理による分析と古い研究デザインを統合することによって、実際に用いられているテストについてさらに明快な CV を期待しうるであろう (p. 155)。

Cronbach の継承者の一人である Snow は長年にわたった ATI 研究に一区切りをつけた後、この方針に従う研究に着手したが、その道半ばで亡くなっている。

Cronbach は最後の節で、CVn をコミュニティ過程として考えて、効率よくこれを進める現実的な方略を提案している。

たった 1 つの構成概念についても、確立した解釈が可能になるには、長い時間がかかる。知能を特徴づけようとした Binet の最初の努力から、現在の私たちの控えめな理解に至るまでに 1 世紀近い歳月が流れた。歩みは遅いではあるが、構成概念による解釈がより厳密になるにつれて、妥当性について行われる精緻で優れた調査研究によって、投入される余計な出費も十分その元が取れるようになる。……

科学、テクノロジー、そして政策いずれにおいても同じように、一つの理論は関係するコミュニティの広い支持を得て初めて成功を見るのであり、その過程は合理的であり、また社会的なものである。構成概念による解釈の容認は本来コミュニティ過程である。合理主義者の理想をこのように弱体化しても、これは決して Cronbach and Meehl (1955) の論文で推奨した研究の方針を撤回したわけではない (p. 164)。

テストの CVn を迅速に、しかも公平に行うには、テストの開発者とは別に、テスト評価の専門家が必要であると Cronbach は主張する。テスト開発者には、自分の仕事に対する思い込みがあるのが常であるので、評価の専門家はそのテストに対して中立的な立場をとりうる人であり、独自の調査を行って限られた時間内にその結果を報告できる人であってほしいと言う。評価の専門家が CVn の仕事を迅速に進めるために、問題点を整理して調査を進めるならば、テスト評価の専門家の仕事はコミュニティの代行人として CVn を引き受けることにほかならないと Cronbach は説いている。

Embretson (1983) は、心理学が認知心理学へと転じた変化を機能主義から構造主義へのパラダイムシフトであるとし、心理学における理論化も先行条件／結果の間の関連性の説明から、遂行行動をその礎になっているシス

テム、および下位システムで説明するように変化したとする。CVnの研究においても、このような変革に対応して、Embretson は従来型アプローチを法則定立的スパン (nomothetic span)、認知心理学的アプローチを構成概念表現 (construct representation) とそれぞれ呼んで区別することを提案した。また構成概念についても、Cronbach がこれを人について測定された属性であり、個人差の源となる理論的変数としたのとは異なり、構成概念表現ではやや広く、課題遂行の礎となっている理論的メカニズムの確認を目指している。

要するに構成概念表現の研究の目標は、課題を分解することにある。情報処理のパラダイムでは、構成概念表現とは、課題に対する反応が遂行行動に含まれるプロセス、方略、および貯蔵された知識のそれぞれにどの程度依存しているかを問題にすることを指している (p. 180)。

法則定立的スパンに関する研究は、典型的には心理学的、および教育的テストの CVy を支持するための一次的データとして提供される。法則定立的スパン研究は、個人差測定のためのテストとしての効用が評価される段階にあたる。あるテストで測られる構成概念に関する理論を支持するために、そのテストにおける個人差と、他のテスト、いくつかの集団への所属、社会的に重要性のある基準等との相関を求める。もしあるテストにおける個人差について、構成概念と相関を示すはずの変数との間に強い相関が得られるならば、NNW は広いスパンを持っていることになる (p. 191)。

Embretson はまた、以下のような構成概念表現の評価規準 4 カ条を提案した。

- (i) テストに対する個々人の遂行行動は、項目の刺激の特徴に関連づけられねばならない。
 - (ii) 課題に含まれる構成概念について、いくつか考えられる理論との比較に耐えられねばならない。CV のためには、競合する理論を数量的モデルとして操作化した上で、モデルの適合度を比較すべきである。
 - (iii) 特定の項目において、理論的構成概念の量化的方法が確立されていなければならない。つまり、ある項目を解くときに含まれる構成概念について、各項目のパラメータを必要とする。この条件を満たすことにより、特定の理論的特性を持つ項目のバンキングが可能になり、また項目困難度に理論的特性を持たせることができる。
 - (iv) この方法論が、当該構成概念上での個々人の測定を可能にしなければならない。つまり、構成概念の次元上で、個人差を表すパラメータが必要である。
- 以上の 4 カ条は、上述した同じく CV に関する Cron-

bach の NNW の 6 カ条とまったく異なることが興味深いのであり、ここでも心理学の潮流の変化を目の当たりにするのである。

Embretson は、構成概念表現的アプローチを大別して 3 種類挙げて、いずれも上の 4 つの規準に照らして評価した。

- (a) テスト項目の数学的モデリング。項目に対するメカニズムを探る方法として、①複雑さの要因による方法と、②サブタスクの方法によるコンポーネント・モデルとを挙げている。
- (b) 心理測定学的モデリング。項目反応理論であり、“人と項目との出会い”という表現をつけ加えている。構成概念は、“潜在特性”であり、項目間の共通因子にほかならない。Embretson は、Rasch model の linear logistic function を基本のモデルとして用いている。
- (c) 多重コンポーネントを持つ潜在特性モデリング。これは、上の(a)と(b)のアプローチを組み合わせたものであり、(a)に 2 通りのアプローチがあることから、この組み合わせにも 2 種類あり、linear logistic latent trait model (LLTM) と multicomponent latent trait model (MLTM) の 2 つである。前者は項目困難度として、操作の難しさと操作の数の一次式を指定する logistic function model である。後者は、ある項目の正答率を、一組のコンポーネントを指定した上で、各コンポーネントの通過率の積と仮定し、これを(b)の Rasch model と組み合わせたものである。

Embretson は以上の 3 種類のモデルにつき実測データへの適合度を調べ、さらに上述の 4 つの規準により評価を行った。また法則定立的スパンについては、当時先端的な技法であった共分散構造分析を用いて、CV を検討した。そして結語として次のように述べている。

構成概念のモデリングというアプローチ、特に構成概念表現という研究の成否は、課題に含まれる理論的メカニズムを定義する量的指標を開発しうるかかどうかという研究者とテスト開発者の力量にかかっている。ある場合には、特殊な理論的メカニズムを研究するための革新的方法が求められるかもしれない。テスト項目について課題を分解する研究は、特に活発な領域であり注目に値する。したがって、方法論上のさらなる進歩が期待される (p. 195)。

以上、従来の NNW による CVn は、次の節でも述べるように隔靴搔痒の感が否めないものであるが、認知心理学的モデリングを導入した構成概念表現のアプローチは CV へのより直截簡明な研究方法である。

III. 心理測定学的測度における概念的妥当性検証の方法

本節では、主として因子分析によった伝統的な構成概念の NNW による CVn の手法を網羅的に挙げたい (Cronbach & Meehl, 1955; Cronbach, 1971, 1990; Gregory, 1992; Klein, 1993; 並木, 1997; Thorndike, 1982)。

1. テストの均質性

テストを構成する全項目が単一の潜在特性を測るとき、この均質性は CVy の必要条件ではあるが、必要十分条件ではない。この性質は信頼性を保証しても妥当性を裏づけるわけではない。均質テスト作成の手順は、潜在特性に対して内容的妥当性の高い項目を選定した上で、部分-全体相関の値の低い項目を除いたり、あるいは主因子法の因子分析により得られた第二因子以下の固有値が 1 より小さい値となることを求める Kaiser (1960) の基準を満たせばよい。項目反応理論のいくつものモデルは、この均質性を前提条件としているが、実測データのモデルへの適合度が高いことは、均質性と同様、測定対象である潜在特性、あるいは構成概念に関する CVn にはならない。均質テストが捉えているものが何であるかについては、NNW における間接的な検証によるか、あるいは前節で述べたように、項目反応理論に情報処理モデルを組み合わせるか、あるいはまた、後述するように、筆者らの試みた項目反応理論と認知心理学的理論を突き合わせる手続きも一つの方法と考えられる。

2. 因子分析

新しく開発されたテストを、因子構造がよくわかっていないテスト項目の中に含めて因子分析を行うことにより、新しいテストの CVy を確かめることができる。一例として、Lansdell and Donnelly (1977; Gregory, 1992) の研究を挙げる。新しいテストの一つは Halstead-Reitan Category Test と呼ばれる神経心理学的検査であり、検査項目に対する反応だけでなく、強化の手続きによって学習させることにより認知能力を測ろうとするユニークなものである。もう一つの Finger-Tapping Test は単純で純粋な psycho-motor 的なテストである。彼らはこの 2 つのテストを WAIS の 11 の下位検査と一緒にして、合計 13 変数の相関行列につき因子分析を行った。なお、被験者 96 名はすべて何らかの精神疾患の患者であった。バリマックス回転後の因子行列につき、4 つの因子の解釈を行い、さらに新しい 2 つのテストの因子負荷量を検討した結果、これら 2 つのテストの測定内容について、既存のテストで測られるもの以外の新しいものは認めがたいという結論に達したという。この研究では否定的な結論になったとしても、因子分析による CVn のよい例である。

Cronbach and Meehl (1955) は、Guilford と Eysenck の言葉を引用しながら、因子分析で抽出された因子が構成概念としての機能を持っており、因子による説明が正確で安定しており、また因子の抽出は構成概念に関する仮説の検証に明示的に用いられうると説いている。

3. 収束的妥当性 (convergent validity) と弁別的妥当性 (discriminant validity) の確認

CVn に関して Campbell and Fiske (1959) は、これら 2 つの妥当性を考慮すべきであるとするきわめて重要な提言を行った。Cronbach (1989) は、彼らの考え方を“CV を探求するための調理法を提供するもの”と評価している。収束的妥当性は、共通の構成概念を測定するための検査得点間に高い相関が認められるとき、これらの得点が妥当性を持つことをいう。弁別的妥当性は、これらの共通の構成概念の検査得点と、これとは別の構成概念の検査得点との間に低い相関が認められるとき、これらの得点の妥当性が立証されることをいう。そして、このような妥当性を系統立てて検証する方法として、Multi-trait-Multimethod Matrix を用いて、2 つあるいはそれ以上の数の特性を 2 つあるいはそれ以上の方法によって測ることにより、これら 2 つの妥当性を同時に立証するという。

CVn の方法としてこのマトリックスを用いる手続きは非常に優れているが、完全にこの方法を施行することには大変な労力が伴う。例えば Thorndike (1982) は、一部に欠損値のある行列を挙げている。ほぼ完全にこの手続きに従った例として Wong, Day, Maxell, and Meara (1995) の社会的知能の研究があり、12 にも上る数の測度の妥当性を検討した。多くの研究では、ある特性の測度の妥当性の裏づけは、他の 2, 3 の特性の測度との相関を調べるだけで終わり、CVy に関する確定的な証拠が得られないのが通例である。彼らの方法は、測定方法の共通性によって本来の妥当性が水増しされる弊を排除すると同時に、測定方法の違いを超えて、ある特性あるいは構成概念を仮定する当否を明らかにすることにより、安易な心理検査の陥穽に警鐘を鳴らすものといえる。

4. 群間差

ある特性ないしは構成概念を測る検査得点により高低 2 群を設け、この構成概念と理論的に関連する作業成績を比較することによって、検査得点の CVy を立証する方法である。一例として、Campione, Brown, and Bryant (1985) は比較的容易な Raven Progressive Matrices タイプの 3 項目を学習課題とし、知能検査得点によって分けられた正常児群と軽度の精神遅滞児群の学習転移実験を行った。各項目を自力で解くようになるまでに必要な手助けの回数を学習の測度としたが、これは dynamic

assessment の手続きにあたる。3つの課題は、それぞれ回転、重ね合せ、引き算の規則によって正答が得られる。そして、一旦習得されたこれらの規則が両群について、2つの規則を組み合わせた転移課題にどの程度転移を示すかを比較した。必要な手助けの回数は、規則習得段階では両群の間に差はないが、規則の保持段階ですでに差が生じ、転移課題では遅滞児群の成績が著しく劣った。この結果は、両群を分けた一般知能の構成概念を情報処理の能力とメタ認知的な働きとするならば、学習転移を促進させたものはそのような構成概念よりなる知能の働きであると解釈することができる。

5. 理論的予測通りの実験処理効果の確認

例として、筆者らによる ATI パラダイムを用いた評価方法の比較研究を挙げる(鹿毛・並木, 1990)。

小学校6年生の算数の1つの単元の授業について、クラス単位で相対評価、到達度評価、および自己評価の3条件で評価する。約10回の授業は同一の学習プログラムを用いて行われた。一つの結果として、テスト不安傾向を適性の次元、授業で受ける強制感を結果の測度とした場合、3つの評価条件との間に顕著な交互作用が認められた。相対評価の回帰直線の傾きが一番大きく、不安傾向と強制感が強く相関している。また主効果としても、相対評価が高い強制感を生じている。一方、不安傾向の低い生徒については、3つの評価条件の間に差はない。

従来、評価方法の比較は主効果だけで論じられることが常であり、この研究は評価方法にも ATI 的視点の重要性を指摘している。さらに、実験結果が理論的に予測したとおりであり、また結果の解釈が容易であることは、教場実験を可及的に統制するためにプログラム学習を用いたこととともに、適性の測度、結果の測度、および評価条件の CVy がそれぞれ確保されていたためと考えられる。このように、ATI パラダイムは測度の CVy を NNW の中で確認する手続きという意味を持っている。Snow and Yalow (1982) の作成した一般知能とさまざまな教授処理とを関連づける ATI 仮説の一覧表によれば、例えば、発見的方法と情報処理の負担を軽減する教授処理を組み合わせると ATI 効果が生じることが予測される。この表は、一般知能の CVy を NNW の中で明らかにしている。

6. 構成概念の NNW における明確化

Jensen (1970; Eysenck, 1979) は IQ と学習の間に相関が生じる条件を以下の9カ条にまとめたが、これも上の Snow の表と同様に、IQ の CVy を NNW の中で明らかにするものである。

①学習が意図的に行われる性質を持ち、努力を要するものであるとき。②学習材料が階層的構造を持っている

とき。③有意味材料であるとき。④学習課題が、何らかの関連性がありながら、同一でない過去の学習からの転移を可能にする性質を備えているとき。⑤学習が洞察的な局面を持つとき。⑥学習材料が中程度の困難度と複雑さを持つとき。⑦全学習者に対して学習時間を一定にしたとき。⑧学習材料が年齢に関連しているとき。⑨新しいものを学習するとき、学習過程の初期の段階において。

以上、これらの学習状況は学校で生じる学習の性質を備えており、また IQ がそのような学習のための能力であるといえる。要するに IQ の CVy をいくつもの学習の条件という NNW の中で明らかにしている。

7. 多次元尺度法を用いて空間的に表現する方法

Marshalek ら (Marshalek, Lohman, & Snow, 1983; Jensen, 1998; Kyllonen, 1996; Snow, 1980) は、既成の34種類の下位検査を高校生に施行したデータを Guttman のノンメトリックな多次元尺度法 (Radex Model) を用いて分析し、知能の各因子の空間的位置関係によりその構成概念を明らかにしたが、これは CVn の新しい方法として多くの研究者の注目を集めた。

FIGURE 1 中の個々の点は下位検査であり、それらは同心円的なパターンを作っている。より一般的かつ複合的な検査内容を持つものが中心にあり(つまり他の検査との相互相関が平均的に高い)、周辺に近づくにつれて特殊性が高く、単純な内容の検査となり、これが radex と呼ばれる形である。また空間中の距離が近いことは、検査内容の類似性が高いことを示している。ここで、ちょうどパイを中心で切り分けたように、検査内容により言語的、数量的、および空間的の3領域に分かれることも興味深い。なお、図中の実線は従来の因子分析的研究で得られたクラスターを重ね合わせたものである。

彼らによれば、同じ下位検査間の相互相関に階層構造的因子分析を施行した結果は、この radex にほとんど完全に対応しているという。すなわち、同心円の中心にどれくらい近いかが g への因子負荷量に、特定のパイの中心線への近さは群因子への因子負荷量にそれぞれ対応している。

8. 心理測定学的、および認知心理学的構成概念を同一のシステムの中で分析する方法

Spearman が一般知能因子 g を確認して以来、以上挙げたいくつもの研究成果により、この因子の構成概念が次第に明らかにされてきたが、その因子の中でどのような情報処理が行われるかはまだ十分にわかっていない。

Kyllonen (1996) は、作動記憶という理論的構成概念によって、1世紀に近い歴史を持つ心理測定学的構成概念である g の認知心理学的性質を明らかにしようとした。彼は、Cognitive Abilities Measurement (CAM) という

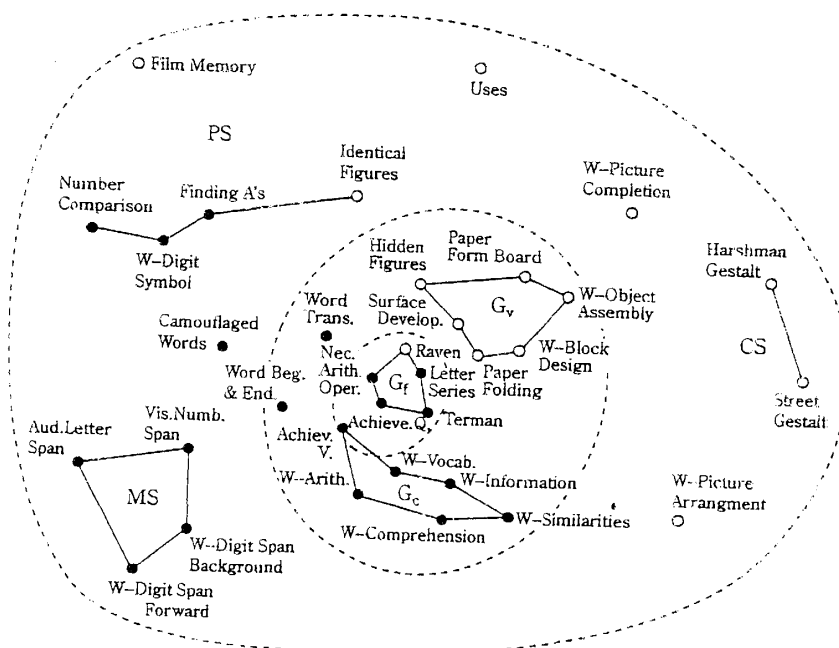


FIGURE 1 高校生 241 名に施行された 34 種類の検査間の相関を Radex model の多次元尺度法により分析して得られた図 (Snow, 1980; Marshalek, Lohman & Snow, 1983)

Gf: 流動性知能, Gc: 結晶性知能, Gv: 視覚的処理因子, MS: memory span, PS: perceptual speed, CS: closure speed, W: Wechsler Adult Intelligence Scale.

枠組みを考えて、4つの源泉要因 source factor (処理の速さ PS, 作動記憶 WM, 宣言的知識 DK, 手続き的知識 PK) と 3つの内容領域 (言語的 V, 数量的 Q, 空間的 S) を組み合わせて 12個 (4×3) のセルからなる分類体系を作った。例えば、作動記憶の要因で言語的内容 WMV というように、各セルに彼らが新しく作成したものも含めて、1～3種類の検査を準備した。2つのセルは適切な検査が得られず空欄であったが、計25種類からなる検査バッテリーを350名の被験者に施行した。このデータについて、25の検査に単一の因子 g のみを仮定するモデル、各セルに各々独立した因子を仮定するモデル、4つの源泉要因とこれと直交する3つの内容領域を組み合わせて7因子とするモデル (これは3.で述べた Multitrait-Multimethod の方法に相当する)、単一因子 g といくつかの下位因子を仮定する階層構造モデル等、9種類のモデルについて分析を行い、各モデルのデータへの当てはまりを検討した。

これらの分析から得られた一つの結論として、3つの作動記憶因子が一般 (推理) 因子と非常に高い相関 (約 $r = .95$) を示しており、作動記憶容量の個人差が一般知能の個人差を説明することが明らかであるとしている。また彼は得られた知能の因子といくつかの学習課題との関連性を分析して、同様に作動記憶容量の決定的な重要性を確認している。

Kyllonen はこの論文の要約の中で次のように述べている。

瞠目すべき発見は、作動記憶容量が認知的活動にお

いて中心的な要因であることが首尾一貫して明らかになったことである。逐次研究を進めるにつれて、多種の操作化に基づき、多様な基準を用いて、作動記憶容量が他の認知能力テストの成績との間、また短期的および長期的学習との間で、他のどの認知的因子と比べても、より深く関連していることが確認されたのである。作動記憶容量の因子が中心的役割を担っているという発見により、作動記憶容量は正しく Spearman's g に違いがないという結論に達したのである (p. 73)。

さらに彼は、作動記憶容量という理論的構成概念の利点を4つ挙げている。①作動記憶容量はその確認のために因子分析によらないために、統計的 artifact という批判を受けなくて済む。②作動記憶という考え方は、個人差の因子にレッテルを貼ったものではなく、記憶研究の中で実験結果を説明するために理論的に展開してきた構成概念である。③作動記憶容量は相対的尺度ではなく、絶対性を備えている。④作動記憶は心理測定学と認知心理学の2領域の協同の可能性を拓ける。

Kyllonen の意欲的な研究は、心理測定学と認知心理学の代表的な構成概念間の関連性を明らかにすることにより、両者の CVn を同時に行ったものであり、次節の筆者らの研究もその方法は異なるとしても、その意図は彼の主張と軌を一にしている。

IV. 概念的妥当性の相補的検証手続き

アルツハイマー型認知症 (DAT) や脳血管性認知症 (VD)

に見られる多様な症状が、心理学の対象領域に深く関わることから、これらの疾患の心理学的研究が盛んに行われてきた (Bayles, Boone, Tomoeda, Slauson, & Kaszniak, 1989; Costa, Whitefield, & Stewart, 1989; 並木, 1993, 1996)。筆者らは医学部神経内科の医師との協同のプロジェクトとして、新しい認知症診断の神経心理学的検査の開発を目標とする研究に着手し、この度ほぼその作業を終えた (並木・篠原・山本・米倉, 2002; 米倉, 2003; 米倉・篠原・山本・並木, 2005)。わが国でこれまで広く用いられてきた診断検査は、記憶機能のみに偏るもの、また外国の検査の翻訳ないしは翻案、模倣によるものが多く、理論的背景を欠き、特に日本語の特質 (Namiki, 1999) を考慮したものもない。このような問題点をふまえて筆者らは以下を特徴とする新しい検査を開発した (Sato, Namiki, Ando, & Hatano, 2004)。

(1) 作動記憶理論に基づいて検査項目の作成と分析を行った。近年、加齢に伴う認知機能の低下に作動記憶が深く関連しており、特に DAT においてその容量が著しく減少することが確認されている (Baddeley, 1999; Morris, 1994; 並木, 1999; 米倉, 2001)。Baddeley ら (Baddeley, Logie, Bressi, Della Sala, & Spinnler, 1986; Baddeley, Bressi, Della Sala, Logie, & Spinnler, 1991) は、単純な課題の場合には DAT 患者群と健常高齢者群の間に作業成績の差はほとんどないが、二重課題とすることにより作動記憶に大きな負荷をかける場合には、患者群の成績が著しく低下することを見出した。この結果は、被験者の作動記憶容量と課題の作動記憶に対するデマンドとの間の統計学的交互作用によるものと考えられる (並木, 1997)。作動記憶理論は、検査項目をサブルーティンに分解することによりそのデマンド数を求めることを可能にするので、この交互作用を反映しうる適度のデマンド数、つまり困難度を持つ検査項目を作成して、認知症患者と健常者、さらに DAT 患者と VD 患者とを識別する検査を構成することができた。

(2) 項目困難度と項目識別力の 2 つのパラメータを持つロジスティック関数モデルによる項目反応理論 (Embretson & Hershberger, 1999; Embretson & Reise, 2000; Mislevy & Bock, 1982) を適用することにより、通常計算される検査合計点よりすぐれた尺度値が得られただけでなく、検査得点からより多くの情報を得ることができた。まず項目について得られるいくつかの指標を用いて診断の目的にあわせた簡易型検査を構成して臨床的効用をさらに高めることができた (米倉他, 2005)。またこの理論は項目単位で分析することから、判別分析に適したデータが得られる。

本研究の結果の一つとして、下位検査項目を 51 個に細

分化して、臨床的に確定した 2 つの病型 DAT と VD との一致度を判別分析により計算したところ、87.6% という高い値が得られ、この検査が臨床場面におけるスクリーニングに十分役立つと考えられる (米倉・篠原・並木, 2003)。なお、この結果は、検査得点の NNW による CVn に相当する。

(3) 検査得点の CVn を客観的に、また定量的に行う方法を確立することができた。項目反応理論は、単一因子のみを測定する均質テストを前提条件としており、本検査得点の主因子法の因子分析により、第一因子の固有値が突出して大きく、均質性に関する Kaiser (1960) の規準を十分満たしていることが確認された。

一方、Case の作動記憶理論 (Case, 1980, 1985, 1992; Case & Edelstein, 1993; 並木, 1982) によれば、ある課題の遂行の可否について、被験者の作動記憶容量が課題の最大デマンドと等しいか、あるいはより大きいときのみ遂行が可能となる。これは決定論的項目特性曲線をモデルとする項目反応理論に相当し、また Guttman の尺度分析と同じ考え方である。したがって、課題項目の困難度はその最大デマンド数で決定される。筆者らは各項目を下位項目に分けて、その最大デマンドを Case の方法によって求めた。

TABLE 1 の値は 13 個の下位検査項目について得られたものであり、これが理論的困難度である。TABLE 1 に各下位検査項目の実測正答率と項目反応理論により計算された推定困難度を示したが、理論的困難度とこれらの間の相関が非常に高いことから (36 個に細分化した場合、 $r = .93, .95$)、作動記憶による理論的解析が全体として正しいと考えられ、またこの検査が捉えている測定対象が作動記憶容量であるという CVy を立証している。一方、検査を構成する 5 種類の項目から得られる 8 つの検査項目を第一因子負荷量の大きさと並べると、その順位と各項目の最大作動記憶デマンドの順位とが完全に一致し、この結果も検査得点の CVy を裏づけている。他方各項目の項目特性曲線への当てはまりがよく、さらに Guttman の尺度化モデルへの適合度を調べるために Green (1956) の再現性指数を計算したが、この値も十分高い。以上の事実から、本検査が捉えている単一次元が作動記憶容量であるという確証が得られた。上述した Embretson のアプローチは、項目特性曲線のパラメータに心理学的意味を持たせるものであるが、筆者はこれとは異なり、項目反応理論と認知心理学的理論によりそれぞれ独立に項目困難度を求め、両者の一致度を検討することを試みた。

以上、本検査の開発は臨床的効用の高さとともに、心理学的検査の CVn の新しい手法を確立したことで一応の成果が得られたと筆者は考えている。なお、候補とし

TABLE 1 各下位検査項目の最大作動記憶デマンド数, 実測正答率, 推定困難度, および理論的困難度(最大デマンド数)と実測正答率, 推定困難度との順位相関係数(米倉・篠原・山本・並木, 2005)

13個の下位検査項目		最大デマンド数	13個の下位検査項目	
			実測正答率(%)	推定困難度
1	事物の名称	5	67.9	-0.58
2	階層構造的分類 1	8	55.6	-0.06
3	階層構造的分類 2, 3, 4	7	57.7	-0.12
4	おはじきの計数 2 桁	4	96.9	-2.16
5	おはじきの計数 3 桁	5	84.2	-1.24
6	おはじきの数の再現 2 桁	6	73.5	0.70
7	おはじきの数の再現 3 桁	7	70.4	-0.53
8	漢字の読み	4	84.7	-1.23
9	漢字の仲間はずれ	8	58.7	-0.14
10	迷路 1	7	51.0	0.09
11	迷路 2	9	39.3	0.44
12	迷路 3	10	31.6	0.66
13	4 桁の順列	6	82.1	-1.15
理論的困難度との順位相関係数			0.90	0.91

た12種類の項目を用いる2つの予備実験の後で、最終的に5種類の項目を選定したが(TABLE 1 参照), うち数唱(Linn, Wolf, Bachman, Knoefel, Cobb, Belanger, Kaplan, & D'Agostino, 1995)を除く4種類の項目は筆者らのオリジナルのものであった(Kawada, Naito, Namiki, Yamamoto, & Yasuoka, 1979; 並木・内藤・安岡, 1976; 並木他, 2002)。

本実験では200名の認知症患者に個別に検査を施行し, そのデータの分析を終えるまでに12年の歳月を要したことを付記したい(米倉・深井・仲村・並木, 1997; 米倉・鎗田・仲村・並木, 1999; 米倉・野澤・仲村・並木, 2000; 米倉他, 2003; 米倉他, 2005)。

V. おわりに

心理学的検査が, 心理学の研究道具としてだけでなく, 教育, 医学, 企業等の各界において実用になる道具として, いろいろな批判に耐えて受け入れられるためには, 検査得点のCVが必須の条件である。2005年は, Binetの検査が完成を見てちょうど100年, またCronbach and Meehlの論文が公にされて50年にあたり, 心理学にとって記念すべき節目を迎えたのである。筆者はここで原点に立ち戻ってCVを再考し, さらに今後の展開の見通しを立てることを企図した。心理測定学と認知心理学の相補的な関係(Sternberg, 1980), 統合(Cronbach, 1989), ないしは協同(Kyllonen, 1996)が次の50年に向けてさらなる進展を遂げて, 心理学的測定の理論と技術の改善をもたらすであろう。

引用文献

- Baddeley, A. D. 1999 *Essentials of human memory*. Hove, England : Psychology Press. Pp. 251-272.
- Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R., & Spinnler, H. 1991 The decline of working memory in Alzheimer's disease : A longitudinal study. *Brain*, **114**, 2521-2542.
- Baddeley, A. D., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. 1986 Dementia and working memory. *Quarterly Journal of Experimental Psychology*, **38A**, 603-618.
- Bayles, K. A., Boone, D. R., Tomoeda, C. K., Slauson, T., & Kaszniak, A. W. 1989 Differentiating Alzheimer's patients from the normal elderly and stroke patients with aphasia. *Journal of Speech Hearing Disorders*, **54**, 74-87.
- Campbell, D. T., & Fiske, D. W. 1959 Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**, 81-105.
- Campione, J. L., Brown, A. L., & Bryant, N. R. 1985 Individual differences in learning and memory. In R. Sternberg (Ed.), *Human abilities : An information-processing approach*. NY : Freeman. Pp. 103-126.
- Case, R. 1980 The underlying mechanism of intellectual development. In J. R. Kirby & J. B. Biggs (Eds.), *Cognition, development, and instruction*. NY : Academic Press. Pp. 5-37.
- Case, R. 1985 *Intellectual development : Birth to adulthood*. Orlando, FL : Academic Press.
- Case, R. 1992 Neo-Piagetian theories of child devel-

- opment. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development*. Cambridge, UK : Cambridge University Press. Pp. 161-196.
- Case, R., & Edelstein, W. 1993 Introduction : Structural approaches to individual differences. In R. Case & W. Edelstein (Volume Eds.), K. Deanna (Series Ed.), *Contributions to human development*. NY : Karger. Pp. 1-10.
- Costa, P. T., Jr., Whitefield, J. R., & Stewart, D. (Eds.) 1989 *Alzheimer's disease : Abstracts of the psychological and behavioral literature*. Washington, D. C. : American Psychological Association.
- Cronbach, L. J. 1971 Test validation. In R. L. Thorndike (Ed.), *Educational measurement*. (2nd ed.). Washington, D. C. : American Council on Education. Pp. 443-507.
- Cronbach, L. J. 1989 Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy*. Urbana, IL : University of Illinois Press. Pp. 147-171.
- Cronbach, L. J. 1990 *Essentials of psychological testing*. (5th ed.). NY : Harper & Row.
- Cronbach, L. J., & Meehl, P. E. 1955 Construct validity in psychological tests. *Psychological Bulletin*, **52**, 281-302.
- Embretson (Whitely), S. 1983 Construct validity : Construct representation versus nomothetic span. *Psychological Bulletin*, **93**, 179-197.
- Embretson, S. E., & Hershberger, S. L. (Eds.) 1999 *The new rule of measurement : What every psychologist and educator should know*. Mahwah, NJ : Lawrence Erlbaum Associates. Pp. 1-35.
- Embretson, S. E., & Reise, S. P. 2000 *Item response theory for psychologists*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Eysenck, H. J. 1979 *The structure and measurement of intelligence*. Berlin : Springer Verlag. Pp. 199-201.
- Green, B. F. 1956 A method of scalogram analysis using summary statistics. *Psychometrika*, **21**, 79-88.
- Gregory, R. J. 1992 *Psychological testing : History, principles, and applications*. Needham Heights, MA : Allyn & Bacon.
- Jensen, A. R. 1970 Hierarchical theories of mental ability. In W. B. Dockrell (Ed.), *On intelligence*. London : Methuen.
- Jensen, A. R. 1998 *The g factor : The science of mental ability*. Westport, CT : Praeger Publishers.
- 鹿毛雅治・並木 博 1990 児童の内発的動機づけと学習に及ぼす評価構造の効果 教育心理学研究, **38**(1), 36-45.
- Kaiser, H. F. 1960 Application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20**, 141-151.
- Kawada (Fujitani), T., Naito, T., Namiki, H., Yamamoto, J., & Yasuoka, R. 1979 Experimental verification of mental-space theory and its problem at issue. *Studies in Sociology, Psychology, and Education* (Keio University), **19**, 91-97.
- Klein, P. 1993 *The handbook of psychological testing*. NY : Routledge.
- Kyllonen, P. C. 1996 Is working memory capacity Spearman's g ? In I. Dennis & P. Tapsfield (Eds.), *Human abilities : Their nature and measurement*. Mahwah, NJ : Lawrence Erlbaum Associates. Pp. 49-75.
- Lansdell, H., & Donnelly, E. E. 1977 Factor analysis of the Wechsler Adult Intelligence Scale subtests and the Halstead-Reitan Category and Tapping Tests. *Journal of Consulting and Clinical Psychology*, **45**, 412-416.
- Linn, R. T., Wolf, P. A., Bachman, D. L., Knoefel, J. E., Cobb, J. L., Belanger, A. J., Kaplan, E. F., & D'Agostino, R. B. 1995 The "preclinical phase" of probable Alzheimer's disease : A 13-year prospective study of the Framingham Cohort. *Archives of Neurology*, **52**, 485-490.
- Marshalek, B., Lohman, D. F., & Snow, R. E. 1983 The complexity continuum in the radix and hierarchical models of intelligence. *Intelligence*, **8**, 107-127.
- Messick, S. 1995 Validity of psychological assessment. *American Psychologist*, **50**, 741-749.
- Mislevy, R. J., & Bock, R. D. 1982 *Maximum likelihood item analysis and test scoring with logistic models*. Mooresville, IN : Scientific Software.
- Morris, R.G. 1994 Working memory in Alzheimer-type dementia. *Neuropsychology*, **8**, 544-554.
- 並木 博 1982 操作と作業記憶能力 波多野完治 (監), 天岩静子 (編), ピアジェ双書第5巻 : ピアジェ派心理学の発展II 国土社. Pp. 171-198.
- 並木 博 1993 痴呆の定性的・定量的診断はどこまで

- 可能か—心理学の立場から— 臨床成人病, **23**(4), 61-64.
- 並木 博 1996 知能の生涯発達 総合リハビリテーション (医学書院), **24**(8), 739-744.
- 並木 博 1997 個性と教育環境の交互作用: 教育心理学の課題 培風館. Pp. 113-128.
- Namiki, H. 1999 Aging and cognitive functions in the case of kanji processing : Development of a neuro-psychological test for senile dementia based on the theory of working memory. *Proceedings of German-Japanese Association of Social Sciences*. Tokyo: Waseda University Press. Pp. 339-344.
- 並木 博 1999 加齢に伴う知能の変化 東清和(編), エイジングの心理学 早稲田大学出版部. Pp. 31-64.
- 並木 博・内藤俊史・安岡龍太 1976 ピアジェと測定: 数量的思考と道徳的判断の場合 慶應義塾大学大学院社会学研究科紀要, **16**, 11-19.
- 並木 博・篠原幸人・山本正博・米倉康江 2002 作動記憶理論と項目反応理論に基づく痴呆患者の神経心理学的検査 (T-K-W 式検査): 検査法作成の経過 精神神経学雑誌, **104**(8), 690-709.
- Sato, T., Namiki, H., Ando, J., & Hatano, G. 2004 Japanese conception of and research on human intelligence. In R. J. Sternberg (Ed.), *International handbook of intelligence*. Cambridge, UK : Cambridge University Press. Pp. 302-321.
- Snow, R. E. 1980 Aptitude processes. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), *Aptitude, learning and instruction : vol.1. Cognitive process analyses of aptitude*. Hillsdale, NJ : Lawrence Erlbaum Associates. Pp. 27-63.
- Snow, R. E., & Yalow, E. 1982 Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge, UK : Cambridge University Press. Pp. 493-585.
- Sternberg, R. J. 1980 Factor theories of intelligence are all right almost. *Educational Researcher*, **9**, 6-13, 18.
- Thorndike, R. L. 1982 *Applied psychometrics*. Boston, MA : Houghton Mifflin. Pp. 186-193.
- Wong, C. T., Day, J. D., Maxell, S. E., & Meara, M. N. 1995 A multitrait-multimethod study of academic and social intelligence in college students. *Journal of Educational Psychology*, **87**, 117-133.
- 米倉康江 2001 作動記憶理論による学習困難の分析: 学業不振と老年痴呆の場合 早稲田大学大学院教育学研究科紀要(別冊), **8**, 85-96.
- 米倉康江 2003 痴呆診断のための神経心理学的検査の開発—認知発達心理学の精神神経学への応用— 早稲田大学教育学研究科博士学位請求論文.
- 米倉康江・深井英里子・仲村禎夫・並木 博 1997 作動記憶理論に基づく老年痴呆の神経心理学的診断検査の作成と妥当性の検証 日本教育心理学会第39回総会発表論文集, 363.
- 米倉康江・鎗田智子・仲村禎夫・並木 博 1999 老年痴呆の神経心理学的診断テストの理論的裏付け 日本教育心理学会第41回総会発表論文集, 467.
- 米倉康江・野澤雄樹・仲村禎夫・並木 博 2000 老年痴呆の神経心理学的診断検査得点の分析—作動記憶理論と項目反応理論との関連性— 日本教育心理学会第42回総会発表論文集, 715.
- 米倉康江・篠原幸人・並木 博 2003 作動記憶理論と項目反応理論に基づく痴呆の為の神経心理学的検査 (T-K-W 検査) の開発 日本心理学会第67回発表論文集, 181.
- 米倉康江・篠原幸人・山本正博・並木 博 2005 作動記憶理論に基づく痴呆診断の神経心理学的検査 (TKW 式検査) の開発: 項目反応理論による重症度と病型診断のための用途別簡易検査の作成 高次脳機能研究, **25**(1), 8-16.