

執筆者

上野 一彦 (大学入試センター)	第1章
繁樹 算男 (帝京大学)	第1章
石隈 利紀 (筑波大学)	第2章、第4章
立脇 洋介 (大学入試センター)	第2章
大六 一志 (筑波大学)	第2章、第4章
松田 修 (東京学芸大学)	第2章、第3章
山中 克夫 (筑波大学)	第2章、第4章

目 次

第1章 ウェクスラー知能検査の概要.....	1
1. 知能検査の歴史	
2. ウェクスラー知能検査の発展	
3. ウェクスラー知能検査と CHC 理論	
4. 認知構造の統計学的分析	
第2章 日本版 WAIS-IV の開発.....	9
1. 原版 WAIS-IV の概要	
2. 開発のための準備	
3. 下位検査及び合成得点の概要	
4. 日本版 WAIS-IV (研究版) の標準化	
第3章 発達障害のアセスメントにおけるウェクスラー検査の活用指針と 検査結果から見える発達障害者の臨床的特性.....	28
1. はじめに	
2. 発達障害者のアセスメントにウェクスラー検査を活用する際の基本指針	
3. ウェクスラー検査の活用事例	
4. ウェクスラー検査合成得点の活用指針	
5. まとめ	
第4章 日本版 WAIS-IV (研究版) の活用～認知能力の評価と発達障害者への活用.....	40
1. 10 検査による認知能力の評価	
2. 短縮版による認知能力の評価	
3. 認知能力の評価から支援へのヒントを探る：まとめにかえて	
文献.....	48

第1章 ウェクスラー知能検査の概要

1. 知能検査の歴史

20世紀に入って、人間の知的機能をより科学的に測定しようと、フランスのビネー(Binet,A.)らによって個別式の知能検査が開発されたことはよく知られている。ビネー式知能検査の誕生である。その最初の目的は、知的発達に大きな遅れのある児童を判別するための知能測定であった。精神年齢(MA)という、実際の年齢に対応する知的機能の発達年齢(精神年齢：MA)が当初用いられ、こうした目的による知能検査は、初等教育の普及とともに全世界に広がっていった。

やがて、アメリカに主要な舞台を移した知能検査は、ターマン(Terman,L.M)によるIQ(知能指数)の実用化、さらには大人にも適用されるようになりスタンフォード・ビネー知能検査として著しい発展を遂げた。やがて個別式検査から集団式検査の開発は、第一次世界大戦時の徴兵検査やその後の移民に対する政策のなかで、一定の知的能力を有する者とそうでない者との判別という目的でも使用された。この初期の「判別」という目的は、知能検査を「人間を仕分ける」ツールという極めて操作的な目的に限定されたために、多くの誤用と誤解を生む土壤ともなった。

また測度としてのIQ値は、MAと歴年齢(CA)による単純な比例IQから、正規分布仮説に基づく偏差IQへと変化し、心理統計学の進歩によって、より客観的な指標としての装いを整えていった。その利用にあたっても、IQで測定されるものは何かという心理学における根源的テーマとして常に議論の的となってきた。

実際には、心理学者の中では比較的一致性の高い概念である。たとえば、心理学者が100人集まれば、100通りの定義が出るとも言われるが、かつて661人の知能研究者への調査をした結果がある(Snyderman and Rothman,1987)。その結果は、95%が一致して、抽象的思考・推論、問題解決能力、知識の獲得能力を知能の一部としており、70%以上が一致して、記憶、環境への適応、心的スピードを知能を構成する要素としている。今日、「知能とは一般的な知的能力であり、とりわけ推論し、計画し、問題を解き、抽象的に考え、複雑な観念を理解し、敏速に学習し、経験から学習する能力を含んでいる。」というのが一般的定義である。ウェクスラー(Wechsler,D)は、「知能は目的を持って行動し、合理的に考え、効率的に環境と接する個人の総体的能力」と定義している。

知能検査が初期の目的から大きな変貌を遂げたのは、1960年代の「LD(learning disabilities)」という教育概念の登場である。この概念の最初の提唱者カーカ(Kirk,S.A.)は、単なる個人差から、一人の人間のなかでの認知能力の差異「個人内差」に着目した。これは心理学アセスメントにおける判定の主目的であった、集団のなかでの個人の相対的位置だけでなく、個人の認知構造からその知的活動や情報処理特性に目を向け、それらアセスメント情報を積極的に教育支援に活かすことにもつなげていく可能性を秘めていた。

これらの新しい変化のなかで、知能検査をめぐって神経心理学的アプローチと心理測定的なアプローチの研究が盛んになっていった。前者は、さまざまな認知能力検査の開発という果実をもたらしたし、後者は、LDの判断における知能と学力の差異を重視するディスクレパンシー・モデルによる操作的な判定法のという心理測定研究の隆盛をよんだ。特に後者は、米国の学校教育で学校心理士(School Psychologist)が心理アセスメントや個別教育計画(IEP)作成の中心となっ

て活躍する土壤を作っていました。

しかし、歴史的振り子現象というか、心理測定への偏重を批判し、もっと早期からの子どもの教育支援を重視すべきではないかとする動きが強くなっていました。それは知能検査や認知検査の安易な臨床的解釈への疑問にもつながり、知能検査類に対する次の改革を求める潮流ともなった。

21世紀に入ってから、米国ではRTI (Response to instruction : 子どもの指導に対する反応から次の指導介入をする) を重視する動向が一層強くなり、指導に役立つ情報をめぐっての議論がくりかえされた。それは過去の心理検査の安直な使用や解釈に対する批判という逆風ともなった。心理測定側から見れば、しっかりととした理論に裏付けされた心理アセスメントが強く求められたわけで、さまざまな心理検査類の開発と改訂が行われた。

ウェクスラー知能検査もそうした批判的動向のなかで、理論モデルの検証やプロセス分析を中心とする認知能力の理解と学習指導につながる統合版の開発等、これまでの検査の在り方の再検討を求められたといつても過言ではないだろう。今世紀に入って、ウェクスラー知能検査の改訂 (WISC-IVやWAIS-IV) が短期間のうちに次々と進行した背景にはこうした動向が強く影響している。

人間の知的機能についての状態の単なる判別から、人間のより広くより深い発達理解と支援のためのツールとして、知能検査や認知検査は 100 年を超える歴史のなかで発展を重ねてきた。本ガイドラインであつかうのは、最新版のウェクスラー成人用知能検査の日本版 WAIS-IV の基礎となる研究版とその短縮版である。多くの知的能力に関する心理学尺度が認知検査と称する風潮のなかで、あえて知能検査という伝統的名称を擁護する立場は、代表的な知能検査として多くの臨床実践からの圧倒的な実績と科学的な統計理論からの検証研究に立脚した確固とした自信がそこにあるからであろう。エビデンスを重視し、論理的な頑健さに裏付けされた科学と臨床の両面からしっかりと応答できる心理検査こそが今求められている。

2. ウェクスラー知能検査の発展

知能検査自体、100 年を優に超える歴史を持つが、そのなかでウェクスラー知能検査は 80 年もの間、世界的な使用と絶え間ない改善の積み上げによって今日の位置を築き上げてきた。医療、教育、福祉、司法・矯正などの諸領域において、また、幼児から児童、青年、成人、そして高齢者まで、人間のライフスパンでの知的な発達過程における認知構造を把握するための心理学ツールとして、今日、全世界で圧倒的な支持を得てきた理由がそこにある。

ウェクスラーは、さまざまな臨床現場で使用されている心理検査類のなかから識別に定評のある尺度を下位検査としてパッテリー化し知能検査を作った。彼自身が語る、最初の動機は、「臨床の目的に合った効率的で使いやすい道具を創ること」であった。

この検査は 1939 年のウェクスラー・ベルビュー検査 I、1946 年のウェクスラー・ベルビュー検査 II がその原型となっているが、当初は児童から成人までの、フルレンジ検査であった。その後、一躍、ウェクスラー検査を分析的な知能検査として全世界に普及させる礎となったのは 1949 年に刊行されたウェクスラー児童用知能検査 (WISC) と 1955 年のウェクスラー成人用知能検査 (WAIS) ではなかつたろうか。さらに 1967 年にはウェクスラー就学前・幼児用知能検査 (WPPSI) が加わり、3 系統の年齢群検査として改訂を重ね、現在に至っている。これら検査群は、米国におけるウェクスラー・ファミリーと総称されるが、その開発の系譜については、図 1-1 にタイム

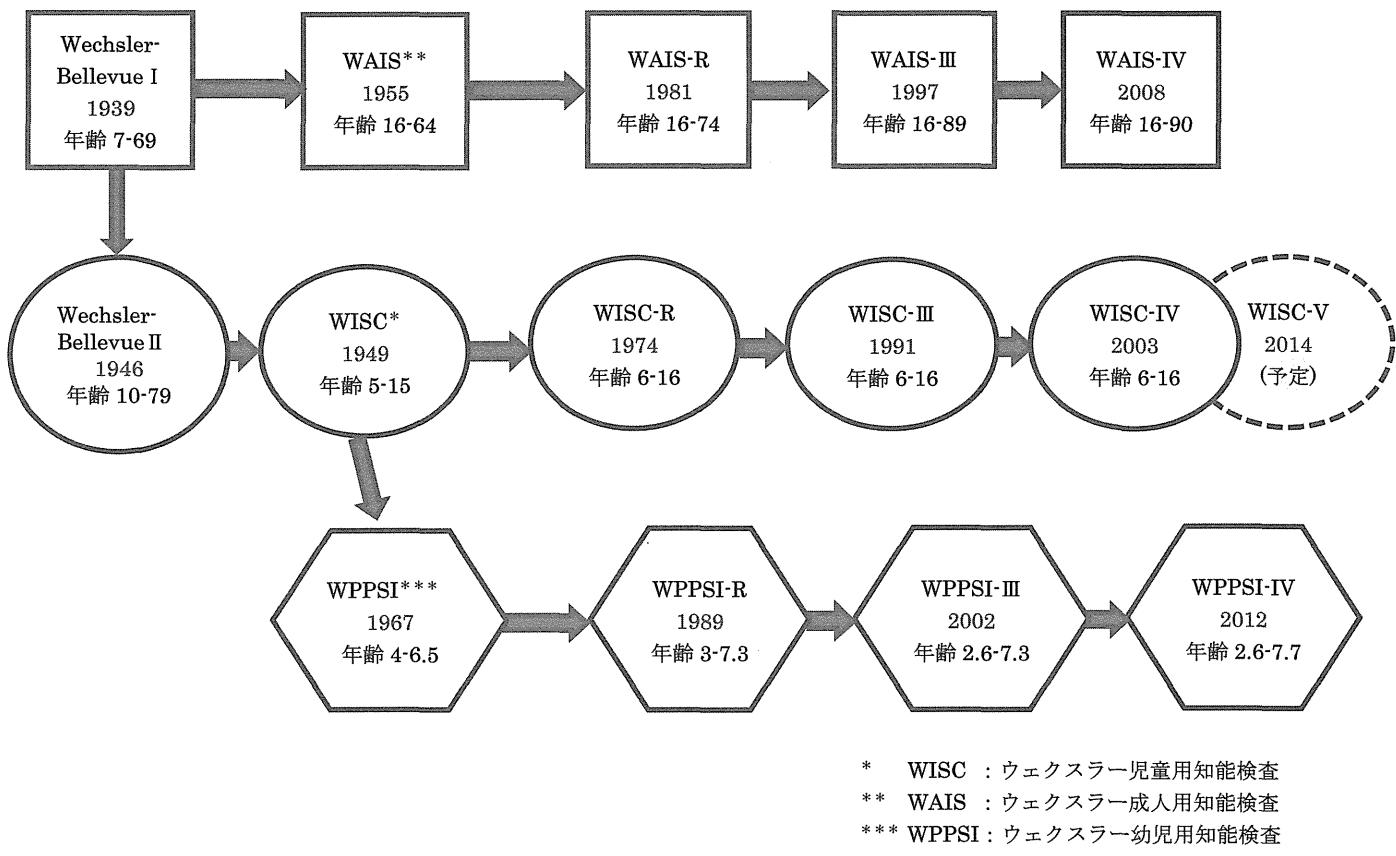


図 1-1 原版ウェクスラー検査の歴史

(上野他(2014)から作図)

ラインとして示される。WISC と WAIS は適用年齢は異なるが、これらウェクスラー・ファミリーにおける知的能力に関する基本概念の構造は共通しており、その意味で経年的変化としての連続性も保たれている。

1949 年に WISC が刊行されると、世界的な注目のなか、各国で翻案・標準化が進められた。わが国もその流れに加わり、1958 年に日本版 WISC を完成させていている。図 1-2 は、上段に日本版 WISC、下段に日本版 WAIS の今日までの刊行の歴史をタイムラインとして示している。

本ガイドラインは、最新のWAIS-IVの日本版WAIS-IVの作成を最終ゴールに置き、その中核となる研究版WAIS-IVの開発を試みている。前節でも触れたように、知能検査・認知検査等の心理尺度は、今世紀初頭に大きな質的転換が迫られ、次々と開発、改訂が繰り返されている。歴史的実績のあるウェクスラー知能検査もその例外ではない。

図1-2でいえば、その胎動は、合成得点という指標を実用化した1991年のWISC-IIIに始まり、1997年のWAIS-IIIで、ほぼ現在のコンセプトの原型は完成した。これらの改訂においては、時間的経過のなかでの尺度改訂として、一般的な改訂事項である、時代的变化に合わせた項目や下位検査の入れ替え、実施法や採点法の改善、新しい集団での標準化と換算尺度の作成、各種統計値の公表と判定基準の明確化などがなされている。

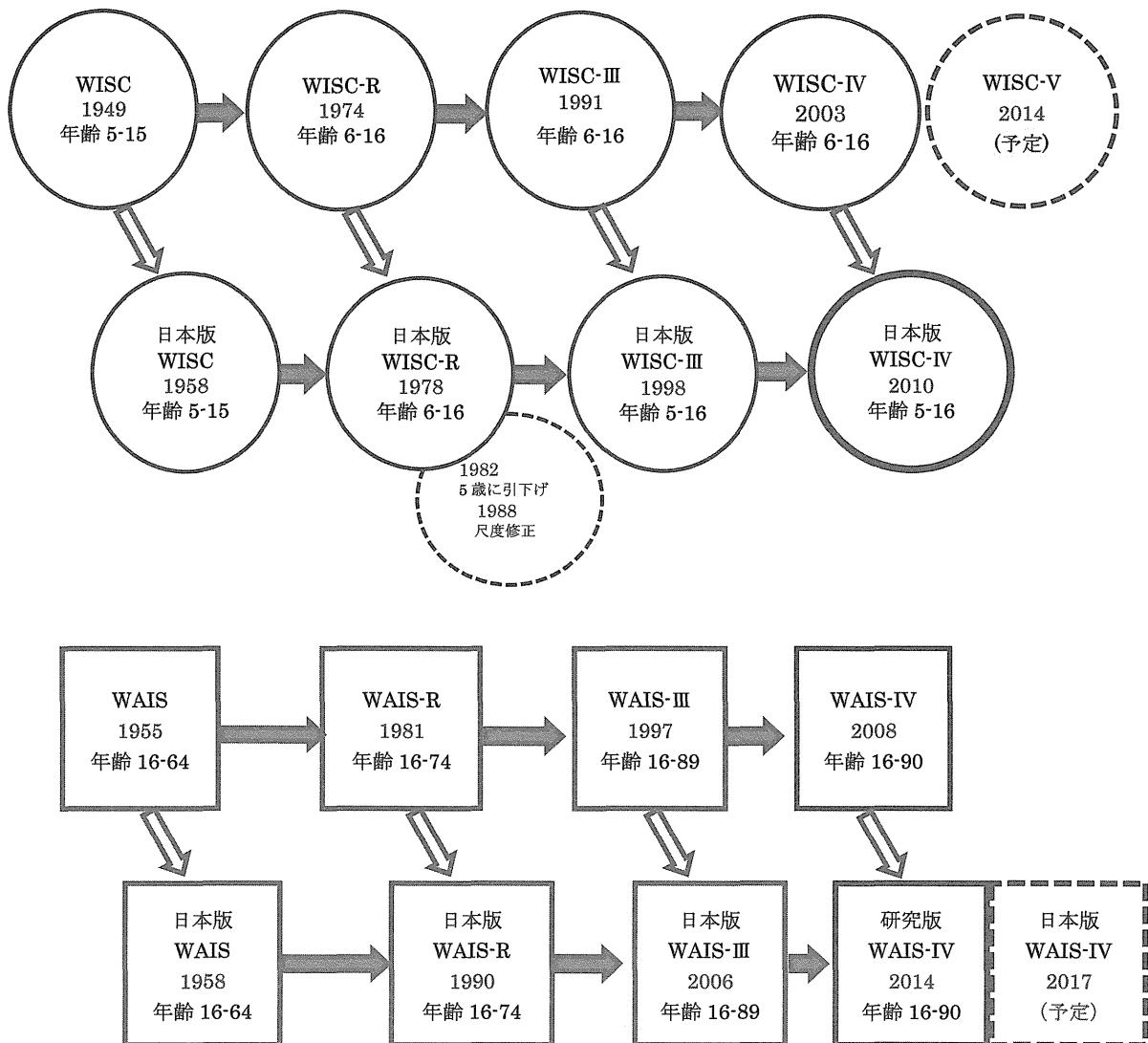


図 1-2 原版 WISC・WAIS と日本版 WISC・WAIS 刊行の歴史
(上野(2014)から作図)

しかし、今回の2003年のWISC-IV、2008年のWAIS-IVにみられる改訂は、ウェクスラー知能検査の歴史上、最も大きな改訂がなされたと思われる。その本格的改訂の起点となったWISC-IVの日本版作成チームが、今回の研究版WAIS-IVの開発チームの中核となっており、このガイドラインの作成にもあたっている。

3. ウェクスラー知能検査と CHC 理論

キャッテル (Cattell.R.B.) は知能検査の分野で因子分析の方法論を用い、多因子からなる知能構造は、流動性知能(Gf : fluid intelligence)と結晶性知能(Gc : crystallized intelligence)という 2 つの共通因子に大別されるとした。Gf は、文化の差異や言語の違いによる影響をほとんど受けない生得的な知能であり、新規の問題を解決するために自ら方法を作り出し、解決する知能とされる。Gc は、流動性知能を基盤とするが、経験の機会など環境因子、文化因子により強く影響される。過去の学習経験を高度に適用して得られた判断力や習慣(つまり経験の結果が結晶化されたもの)

に依存する知能である。

キャッテルの弟子ホーン (HornJ.L.) は Gf と Gc の 2 因子を強調、一般知的因子 g (以後、g 因子) は、さまざまな能力が 100 程度集まつたものであり、それらが Gf と Gc 二大別されることを示している。これらキャッテル、ホーンの研究を背景に、キャロル (Carroll,J.B.,1993) は知能構造に関する多くの研究を概観し、相關行列が計算できることなどいくつかの基準をクリアした 461 の研究から得られた知能検査データに対し、因子分析法を用いてメタ分析を実施した。それらの因子分析の結果を 3 つの層の因子構造として整理した。すなわち、まず、いろいろな認知能力テストの尺度から得られる因子を第 I 層として、系列推理、帰納、ピアジェ的推理、言語理解、語彙など限定因子 (narrow factor) をおいた。これらの上に、第 II 層として、流動性知能、結晶性知能、記憶・学習、視覚、聴覚、検索、認知速度、処理速度など、広範因子(broad factor)を置いた。さらに、最上位の第 III 層には、一般知能因子 (g 因子) を置いた。ところで、第 II 層におかれている流動性知能 (Gf) と結晶性知能 (Gc) とは、キャッテルによって提唱された構成概念である。キャッテルとホーンによる Gc-Gf モデルとキャロルのデータ分析による階層モデルが統合された、新しい知能理論が Cattell-Horn-Carroll の理論、略して CHC 理論と呼ばれる 3 つの階層構造(図 1-3 参照)をなすモデルとして展開されている。

この CHC 理論は、新しい知能理論の中でも、特にその妥当性の高さに注目が集まり、支持されている (三好・服部, 2010)。今日、多くの知能・認知検査は CHC 理論の多大な影響を受けているといっても過言ではない。WISC-IV も例外ではなく、今回の改訂にあたって、理論的な説明力を保証するためにも CHC 理論面からの検討は重要と考えられる。

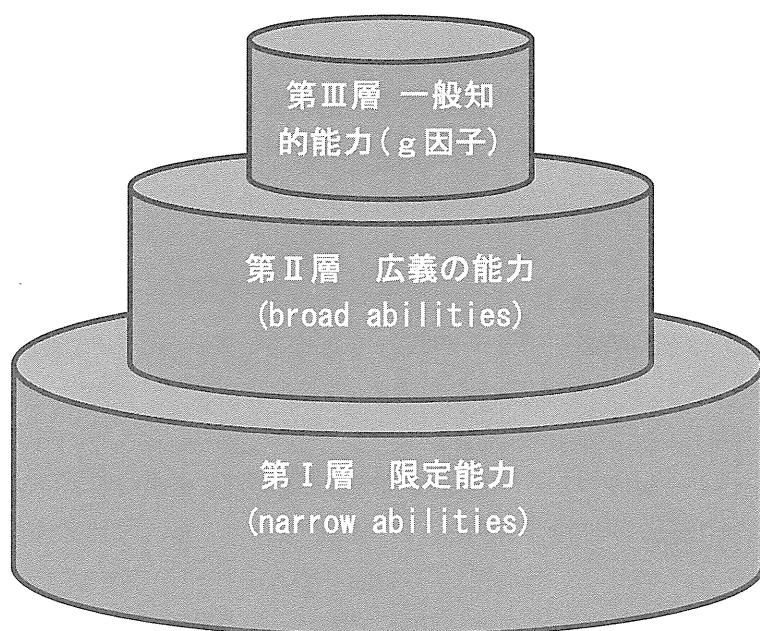


図 1-3 CHC 階層図

4. 認知構造の統計学的分析

(1) はじめに

新しいウェクスラー知能検査の因子構造にはその概念構造と共通性がある。本節では、最新の日本版 WISC-IV の標準化データに基づき、前節で紹介した CHC 理論からの日本版 WISC-IV の適合性を検証し、このガイドラインにおける日本版 WAIS -IVへの理論的な仮説検証とする。すなわち、この理論構造が日本で得られたデータに適合するかどうかは、同じ知能構造概念に立脚する、研究版 WAIS-IV、さらには日本版 WAIS-IVにおいても、CHC 理論が想定する構造が通文化的に妥当であるかどうかを確かめるための傍証となるであろう。そこで、CHC 理論のモデルと WISC-IV が仮定するモデルの因子構造を比較し、それぞれがどの程度の妥当性をもつかを調べる。

(2) 分析の方法

分析にあたっては、日本版 WISC-IV の標準化データを用いた（標準化に関しては、『日本版 WISC-IV 理論・解釈マニュアル』を参照）。標準化データは合計 1,293 名から得られている。その内訳は 5 歳から 16 歳までのそれぞれの年齢群で、111、102、112、116、127、113、123、101、103、93、93、99 名である。

上記の目的を達成するための基本的な分析方法は、検証的因子分析による。すなわち、WISC-IV の尺度間の相関係数行列についていくつかの構造モデルを作り、そのいずれが正しいかを統計的に検討する。これは統計学的にはモデル選択の問題である。統計的モデル選択の基準はいろいろあるが、情報量基準として AIC (Akaike's Information Criterion)、BIC (Bayesian Information Criterion)、データへの当てはまりの良さの基準として RMSEA (Root Mean Square Error of Approximation) を用いる。

AIC は、同じモデル分布から発生されるデータ全体に対する当てはまりの良さを近似するものであり、今手元にあるデータへの当てはまりをよくするだけの指標ではない。また、BIC は、ベイズ理論において、論理的に導かれるモデル比較の指標のベイズファクターの簡便な近似値としての意味を持っている。この 2 つの情報量基準は、2 つ以上のモデルを比較するための尺度であり、値が小さければ小さいほどよいモデルである。RMSEA は、データとモデルとの適合度指標の一種であり、データにモデルを適合させた残差の大きさを評価する。したがって、RMSEA もその値が小さいほどデータへの当てはまりが良好であることを示す。

(3) 結果

検証的因子分析を行い、WISC モデルと CHC モデルの適切さを比較した。WISC モデルおよび CHC モデルを、図 1-4 と図 1-5 に示す。図 1-4 は WISC モデル、図 1-5 は CHC モデルに基づいた因子負荷量の推定値も示している。ほぼすべての因子負荷量が十分高い値を示しており、この 2 つのモデルが妥当であることの 1 つの証左である。ただし、CHC モデルの一部にかなり小さい負荷量のものもあり、それらのパスを除去した場合の CHC モデルが図 1-6 である。

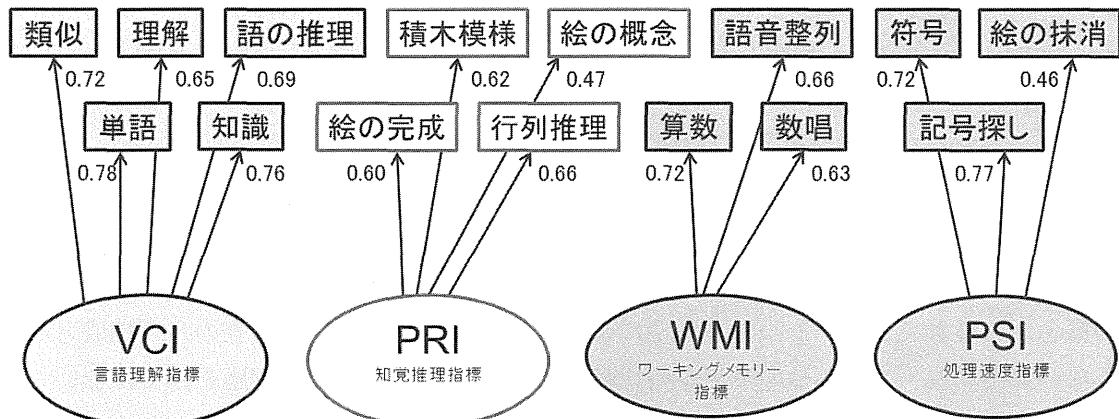


図 1-4 WISC モデルに基づく検証的因子分析

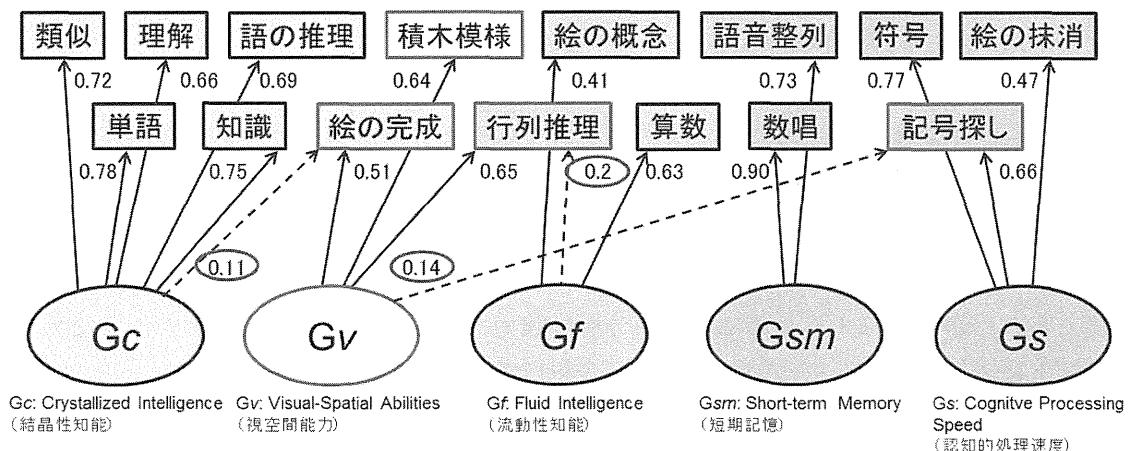


図 1-5 CHC モデルに基づく検証的因子分析

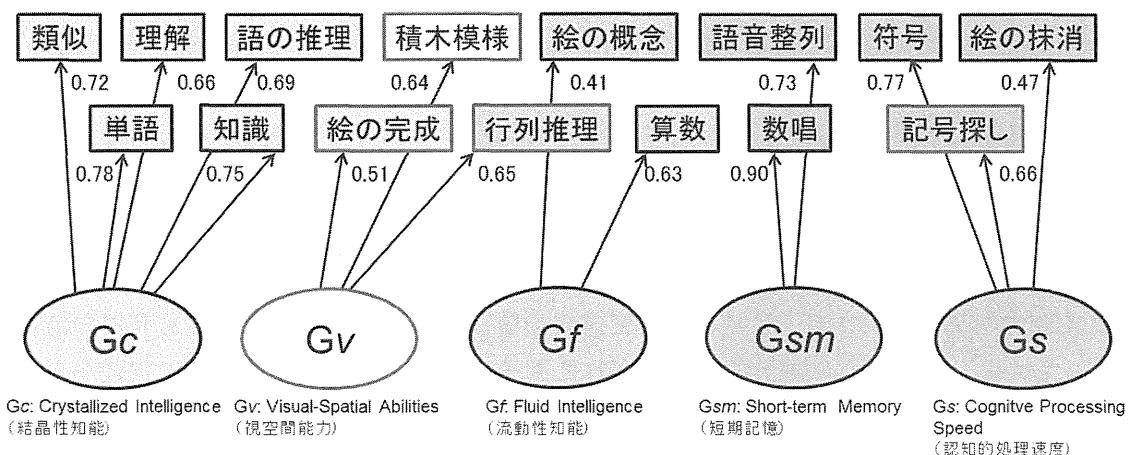


図 1-6 CHC モデルに基づく検証的因子分析（負荷量の高いもののみを表記）

この2つのモデル間の統計的モデル選択において、AIC、RMSEAの統計基準のいずれをとっても、CHCモデルの方が良好な値を示した。すなわち、WISCモデルのAIC = 364.1、RMSEA = .044、CHCモデルのAIC = 307.5、RMSEA = .038である。RMSEAは、0.1以下の値を示すことが望ましいとされることがあるが、その基準はWISCモデルとCHCモデルとともに満たしている。

ただし、これらの差が統計的に有意であるかどうかを知ることはできず、まして、実質的な差であるかどうかを判断することはできない。むしろ、あまり差がないと判断すべきかもしれない。また、知能検査において、1つの下位尺度が複数の群因子にかかわることは、Thurstone以来の単純構造の概念から外れており、2つの群因子に余分の相関性をもたらし、検査の尺度構成として望ましくない点があり、この2つのモデルの比較は単純な統計的モデルの問題のみではないことを指摘しておきたい。

(4) まとめ

知能の因子構造を、WISC-IVの尺度の基礎となっている構造と、CHC理論から示唆される因子構造とに分けて考察した。前者をWISCモデル、後者をCHCモデルと呼ぶ。全体のデータにおいてはCHCモデルの方が妥当であったが、大きな差ではなかった。

知能テストを含む認知能力のテストの設計・開発において、因子分析はどのような役割を果たすのかについての考察を加える。因子分析によって、テストが測定している構成概念を整理できることは確かである。独立した因子であることが検出されれば、それらは別の解釈を与え、別の実践的意味を持つはずである。また、因子間の相関も貴重なデータである。テストの測定目標として、多次元的な評価をするテストを目指すのか、1次元的なテストを目指すのか、それらの目標が適えられたかを吟味するために、因子分析は強力な技法である。また、どのようなテスト項目をテストに入れるべきかについても、因子分析は重要な情報を与える。

しかし、因子分析は万能ではない。因子分析によって抽出されたとされる因子は、いうまでもなく、因子分析に適用されたデータの範囲による。先に挙げた実証研究はWISCのデータであり、標準的な知能のテストが収集した情報の枠内での因子の整理である。実際、オーソドックスなテストによらない因子構造を提唱する理論もある。たとえば、ガードナーの多重理論があるが、この理論では、言語、音楽、論理数理、空間、身体運動、対人関係、自己内省という7つの因子を仮定している。そして、さらに、自然に対する知能、スピリチュアルな知能、存在論的知能、道徳的知能というい4つの知能因子を加えている(Hogan,2007による)。このような理論は、教育に携わる者に影響力があり、確かなデータとのつながりがあり、実践が重ねらなければならぬと思われるが、このように認知能力を広げることが知能に関する理論を深めることにつながるかどうかには疑問があり、知能とは別の次元の付加とみなすべきかもしれない。

最後にもっとも重要な点を強調しておく。因子分析は、先に述べたように、構成概念を整理するうえで強力なツールであり、テストの尺度構成や項目の選択という手続きにも有用な情報をもたらす。しかし、このように統計的に整理された因子や尺度に意味を与えるのは実践からもたらされる知見である(上野, 2012a, b)。因子分析は、統計的に得られた因子を解釈し、名称を与えることが普通に行われるが、この解釈が妥当であるかどうかは、実践によって検証されるべきである。解釈や因子の名称で終わるならば、それは中途半端なところでストップした研究であると言わざるを得ない。

第2章 日本版 WAIS-IV の開発

1. 原版 WAIS-IV の概要

第1章で紹介したように、アメリカでは2008年にウェクスラー成人用知能検査の最新版であるWAIS-IVが開発された。2003年のWISC-IVと2008年のWAIS-IVにみられる改訂は、ウェクスラー知能検査の歴史上、最も大きなものと思われる。今回の改訂ポイントをまとめれば次の6点となる。

①5つの合成得点の算出を基本とする。

10基本検査と補助検査による5つの合成得点（全検査IQ：FSIQ、言語理解指標：VCI、知覚推理指標：PRI、ワーキングメモリー指標：WMI、処理速度指標：PSI）の算出と、補助検査の代替ルールの明確化が図られた（図2-1を参照）。

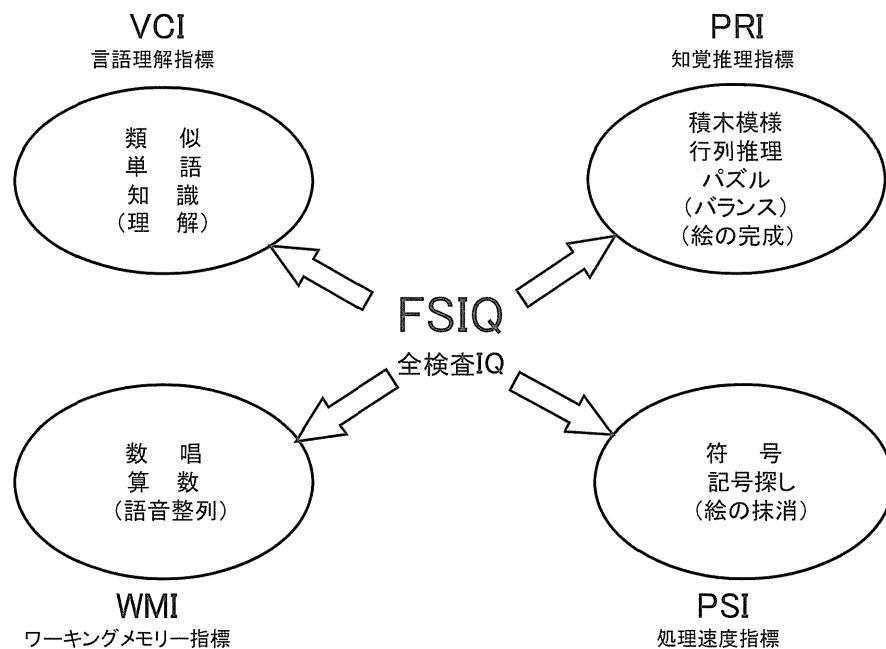


図2-1 WAIS-IVの理論的枠組み

②言語性IQ(VIQ)と動作性IQ(PIQ)の廃止

ウェクスラー知能検査の象徴的存在でもあった、VIQとPIQが廃止された。より純粋な言語理解力や知覚・非言語性推理力は、VCIやPRIによって測定可能されることとなった。

③指標得点の理論的洗練化

4つの指標得点の理論的な洗練化が図られた（推理能力と速度要因の分離など）。また、流動性能力（推理能力）やワーキングメモリーについての測定が強化された。

④実施所要時間の短縮と実施法の簡素化

これは10の基本検査による実施によって5つの合成得点が得られるが、実施に要する時間は従来よりも10～20%短縮された。また下位検査間で実施法の共通化が進められた。

⑤プロセス得点の採用

受検者の知的処理過程をさらに詳しく知るための得点として、プロセス得点を利用することができるようになった。

⑥標準出現率の採用

標準化サンプルにおいてどの程度の割合(累積パーセンテージ)で見られたかを表す数値である標準出現率(Base Rate：累積出現率ともいう)が採用された。

VIQとPIQの廃止、10基本検査と補助検査による5つの合成得点、さらには検査実施法の徹底的な精錬化などが大きな改訂とされる所以である。日本版WAIS-IVの研究版およびそのガイドラインは、これらWISC-IVの最新の改訂コンセプトと日本版WISC-IVの標準化データの解析を土台として作成された。

2. 開発のための準備

(1) 日本版の項目作成に関する主な方針

①項目の変更

- ・できるだけ原版 WAIS-IVの項目を採用する。
- ・変更が必要な場合は、明確な理由を挙げ、「過去の日本版項目（WAIS-III、WISC-IV）」または「新規作成項目」による代替を行う。

②図版の変更

- ・使えるものは原版のものを利用する。
- ・使えないものは、新規項目を日本で作成する。

③パイロットスタディ

日本版 WAIS-IV（研究版）作成の基本方針にそって、パイロットスタディを実施した。大学生中心に総計 557 名（例：理解 166 名、単語 50 名、類似 101 名）の学生を対象に実施した。その結果（項目の平均点・標準偏差、通過率）をもとに、研究版調査で使用する項目を決定した。項目は 1.5～2.0 倍を目安とした。

(2) 倫理的配慮

本研究において収集したデータは、研究担当者のもとに厳重に管理され、研究終了後は廃棄する。また、本研究で得た情報を、他の目的のために使用することはない。研究協力に一旦同意した場合でも、いつでも撤回できることとする。3 年間にわたる研究を通して、上記の点に注意しながら、各ワーキンググループがデータの収集を行った。

3. 下位検査及び合成得点の概要

(1) 下位検査の概要

①積木模様

「積木模様」は、知覚推理指標の基本検査で、制限時間内に手本と同じ模様を、積木を使って完成させる課題である。理論・解釈マニュアルによると、この検査は、主に、抽象的な視覚刺激を分析して統合する能力を評価している。その他、非言語性の概念形成と推理、視覚性知能、流動性知能、視知覚と視覚体制化、同時処理、視覚-運動協応、学習、および視覚刺激の中で全体を部分に分解する能力も反映していると考えられている。研究版の標準化調査では、練習問題と 14 問の本問題が実施された。

②類似

「類似」は、言語理解指標の基本検査で共通する概念を答えさせる問題である。理論・解釈マニュアルによると、言語概念形成や言語推理を評価することを目的としている。また、「類似」は、結晶性知能、抽象的思考、聴覚理解、記憶、類推的および分類的思考、重要な特徴とそうでない特徴の区別、および言語表現の程度も反映すると考えられている。研究版の標準化調査では、練習問題と 24 問の本問題が実施された。

③数唱

「数唱」は、ワーキングメモリーの指標得点を算出するための基本検査である。順唱と逆唱の 2 課題から成る WAIS-III、WISC-IV とは異なり、WAIS-IV では、口頭で提示された一連の数字を昇順に並べ替えてもらう「数整列」がさらに加えられた。この下位検査は題材が数字のみであることから特に文化差を考慮する必要もなく、研究版調査では原版 WAIS-IV の問題をそのまま用いた。

④行列推理

「行列推理」は、知覚推理指標の基本検査で、一部が空欄になった図版を提示し、その空欄に当てはまるものを選択肢の中から選ばせる課題である。理論・解釈マニュアルによると、この検査は、主に、流動性知能、視覚性知能、分類および空間能力、部分と全体の関係の理解、同時処理、知覚体制下を反映すると考えられている。研究版の標準化調査では、2 つの練習問題と 26 問の本問題が実施された。

⑤単語

「単語」は、言語理解指標の基本検査である。この検査には、絵の課題と語の課題が含まれる。絵の課題とは絵の名称呼称を求めるもので、原版 WAIS-IV の 3 問をそのまま翻訳して用いた。語の課題とは従来どおり語句の意味の説明を求めるもので、日本語版のパイロットスタディ用の項目 53 問を作成した。この語の課題について、大学生など 50 名を対象とした記述回答によるパイロットスタディを実施し、通過率を調べた。そして、(1) 通過率、(2) NTT データベースにおける親密度、(3) 品詞の分布、(4) 原版問題項目との文化的な等価性などを考慮し、研究版調査用項目 33 問を選定した。研究版調査後の調整を考慮し、最終版よりも 6 問多くした。パイロットスタディ

イの詳細は以下の通りである。

通過率と親密度の相関は.77 と高かったが、親密度が高いわりに通過率が低い（つまり正しい説明は難しい）項目や、親密度が低くても通過率が高い項目が存在した。そのような項目については、通過率を優先的に考慮して選定した。

選ばれた研究版調査用項目 33 間のうち、日本版 WAIS-III からの継承は 14 間 (42%) となった。原版 WAIS-IV では 78% を原版 WAIS-III から継承しており、それと比べると低い継承率となつたが、これは今回、原版との等価性を考慮したためである。

こうした原版との等価性を考慮する一方、日本のことばの文化的特徴については日本版 WAIS-III を継承するようにした。すなわち、四字熟語は日本語の特徴と考えられ、これは 6 間を選んだ（日本版 WAIS-III では 5 間）。また、主にひらがなで表される形容詞、形容動詞（大和言葉）も日本語の特徴と考え、日本版 WAIS-III の 2 間をそのまま継承した。カタカナで表される外来語については、日本版 WAIS-III の 3 間中 2 間を継承した。

以上の選定の結果、品詞の内訳は、名詞 10 (30%)、形容詞、形容動詞、副詞 12 (36%)、動詞 11 (33%、ただしサ変動詞の語幹のみのものを含む) であり、原版の 26%、41%、33% とほぼ同じ分布となつた。

⑥算数

「算数」はワーキングメモリー指標の基本検査である。制限時間内に暗算で回答させる。この下位検査については、原則、「原版通りの問題を使用すること」および「原版通りの実施方法および採点方法とすること」を確認した。ただし、「算数」の文章題に出てくる西欧人の名前は日本人には馴染みが少なく、記憶に負荷がかかることが考えられるため、日本人の一般的によくみられる名前に修正した。また、文化的・言語的に日本語の表現あまりみられない状況（パイを 99 個もっている）については、日本人になじみがある表現に修正した。これらの修正を行った上で、原版 WAIS-IV の問題項目 22 項目をそのまま採用することとした。

⑦記号探し

「記号探し」は処理速度指標の基本検査である。制限時間内に、記号グループをよく見てこれらの記号の中に刺激記号と同じ記号があるかないかを答える。この下位検査については、「原版通りの問題を使用すること」および「原版通りの実施方法および採点方法とすること」にした。教示については原版 WAIS-IV は重複も多く分量も多いので、日本の検査者が利用しやすいように微調整を行つた。

⑧パズル

「パズル」は知覚推理指標の基本検査であり、選択肢の中から、組み合わせると見本図版と同じになるもの 3 つを、制限時間内に選ぶ課題である。原版 WAIS-IV は 26 間で構成され、正答 1 間につき 1 点である。問題を検討した結果、言語や文化の差異を考慮する必要性は認められなかつたため、原版の 26 間をそのまま用いることとした。

⑨知識

「知識」は、重要な出来事、もの、場所、人に関する一般知識が問われるものであり、言語理解の指標得点を算出するための下位検査である。問題は口頭で提示され、同様に受検者も口頭で回答することが要求される。

原版 WAIS-IV は、全部で 26 間の問題からなるが、文化の違いを考慮した修正を加え、35 間からなるパイロットスタディ用の項目（開始問題とより易しいリバース実施問題を除く）を作成した。これらの問題は 169 名の大学生を対象に集団実施され、通過率や反応、採点のしやすさなどから分析し、研究版調査で用いる問題（開始問題とリバース実施問題と合わせて 32 間）を決定した。

⑩符号

「符号」は処理速度指標の基本検査である。見本を手がかりに、数字とペアになった記号を制限時間内に書き写させる。この下位検査については、「原版通りの問題を使用すること」および「原版通りの実施方法および採点方法とすること」とした。教示について、原版 WAIS-IV は重複も多く分量も多いので、日本の検査者が利用しやすいよう微調整を行った。

⑪理解

「理解」は、日常的な問題の解決や社会的ルールなどについての理解を問うものであり、「言語理解」の指標得点を算出するための下位検査である。問題は口頭で提示され、同様に受験者は口頭で回答することが要求される。米国版からの翻訳項目、日本版 WAIS-R および日本版 WAIS-III から継承した項目から 30 項目からなるパイロット調査を行い、その結果に基づき、日本の文化あるいは時代背景に合った項目といった観点から項目を検討し研究版調査では 26 項目を実施することとした。

⑫バランス

「バランス」は、知覚推理指標の補助検査であり、重りの一部が隠されている天秤を見て、その隠されている重りとして適切なものを選択肢の中から、制限時間内に選ぶ課題である。原版 WAIS-IV は 27 間で構成され、正答 1 間につき 1 点である。問題を検討した結果、言語や文化の差異を考慮する必要性は認められなかったため、原版の 27 間をそのまま用いることとした。

⑬絵の完成

「絵の完成」は、知覚推理の補助検査で、制限時間内に提示された視覚刺激の欠所を見つける課題である。理論・解釈マニュアルによると、この検査は、主に、視知覚、視覚体制化、集中力、対象の本質的な細部を見つける力を反映していると考えられている。絵の完成については、修正図版の原案の検討が終了した。

⑭語音整列

「語音整列」は、ワーキングメモリーの指標得点を算出するための代替検査である。原版 WAIS-IV では、口頭でランダムに提示された「数字」と「アルファベット」をそれぞれ順番に並べ替え

ることが要求される。日本の文化を考慮すると、アルファベットの並べ替えを「かな」の五十音順の並べ替えに変更する必要があるが、日本語では、「2」(数字)と「に」(かな)のように、数字と「かな」の両方に存在している音、まぎらわしい音(「9」と「く」)が存在しているため、こうした点を考慮しパイロットスタディ用の問題を作成した。また、日本の「かな」は行・段の二次元である点でアルファベットより並べ替えが難しいことが予想される。この点を検討するため、19名の大学生に対しパイロットスタディを行ったところ、予想通り原版ノルムに比べ長い系列の問題で極端に正答数が少ないことが明らかにされた。そのため、研究版調査で用いる項目には、難易度が中程度の問題を加えた。

⑯絵の抹消

「絵の抹消」は、処理速度の指標得点を算出するための代替検査である。この下位検査では、規則的に並んださまざまな図形の中から、特定の図形を走査してもらうものであり、制限時間内に正しく見つけられた数をもとに得点が算出される。

内容に関しては特に文化差は考えられないで、「数唱」と同様に、研究版調査ではそのまま原版 WAIS-IV の問題を用いる。

(2) 合成得点の概要

①FSIQ

FSIQ は、全般的な知能の水準を表す指標である。この指標は、3つの VCI 下位検査、3つの PRI 下位検査、2つの WMI 下位検査、2つの PSI 下位検査の得点を合成して算出された。研究版では、類似、単語、知識、積木模様、行列推理、パズル、数唱、算数、記号探し、符号が使用された。

②GAI

GAI は、VCI と PRI を合わせた指標であり、結晶性能力や流暢性推理能力、言語能力、視覚認知など、知能の核心部分の水準を表す指標である。FSIQ の算出に使用した下位検査の中から、WMI 下位検査と PSI 下位検査を除いた能力を表す指標である。3つの VCI 下位検査と 3つの PRI 下位検査の得点を合成して算出された。研究版では、類似、単語、知識、積木模様、行列推理、パズルが使用された。

③CPI

CPI は、WMI と PSI を合わせた指標であり、聴覚的ワーキングメモリーや筆記作業、数処理能力など、熟達して自動化することが期待される能力の水準を表す指標である。これらは知能の核心部分とは考えない研究者が少なからず存在する能力であるが、社会において知能を発揮するために不可欠な基礎力と考えられる。FSIQ の算出に使用した下位検査の中から、VCI 下位検査と PRI 下位検査を除いた能力を表す指標である。2つの WMI 下位検査と 2つの PSI 下位検査の得点を合成して算出された。研究版では、数唱、算数、記号探し、符号が使用された。

④VCI

VCI は、言語概念や言語による推理力・思考力、習得知識（結晶性能力）などを表す指標であり、3つの VCI 下位検査の得点を合成して算出された。研究版では、類似、単語、知識が使用された。

⑤PRI

PRI は、視覚認知、視覚-運動協応、非言語性の流動性推理能力などを表す指標であり、3つの PRI 下位検査を合成して算出された。研究版では、積木模様、行列推理、パズルが使用された。

⑥WMI

WMI は、聴覚的ワーキングメモリーや注意力、数処理能力などを表す指標であり、2つの WMI 下位検査を合成して算出された。研究版では数唱と算数が使用された。

⑦PSI

PSI は、単調な筆記作業を根気よく手際よく処理する能力を表す指標であり、2つの PSI 下位検査の得点を合成して算出された。研究版では、記号探しと符号が使用された。

4. 日本版 WAIS-IV (研究版) の標準化

(1) 調査の方法

①対象者

視覚、聴覚、運動などに著明な障害がない、16～34歳の健常成人321名を対象とした。年齢および性別の内訳は表2-1の通りである。

表2-1 対象者の年齢と性別

年齢群	男	女	合計
16-19	41	41	82
20-24	38	41	79
25-29	41	38	79
30-34	40	41	81
合計	160	161	321

対象者の居住地域については、特に統制しなかった。WAIS-III等の過去の知能検査において、知能検査の結果を左右するほどの文化の地域差は報告されていないためである。

対象者の最終学歴については、多様な学歴が含まれるよう配慮したが、大学在学および大学卒業者が68%を占めた。また、職業では学生が全体の47%を占め、特に16～19歳は全員学生となつた。このことから、各問題の通過率は高めになり、また、得点も全体的に高めになることが予想される。

②テスター

テスターは、業務で日常的に WAIS-III を使用している者54名であり、その多くは臨床心理士

の資格を有していた。

実施法の統一をはかるため、研究の趣旨の説明も含めて4時間の実施法研修会を開催した。研修会の講師は、日本版 WAIS-IV 〈研究版〉作成委員会委員5名が交代で勤めた。

③検査の実施

テスターは「日本版 WAIS-IV 〈研究版〉調査のための実施手引」に基づいて検査を実施した。WAIS-IV は15の下位検査で構成されるが、研究版調査では基本検査10個を実施することとした。その理由は、研究版の方が最終版よりも所要時間が長くなるため、受検者の負担を考慮したのである。研究版で所要時間が長くなるのは、①言語理解の下位検査で問題が多めに用意されていること、②最終版では多くの下位検査で問題3~6から開始であるのに対し、研究版ではすべて問題1から開始すること、③なるべく多くのデータを収集するために、下位検査を中止にする条件を緩和し、中止になりにくくしていること、等のためである。

実施した下位検査10個は、実施順に以下の通りである。「積木模様」「類似」「数唱」「行列推理」「単語」「算数」「記号探し」「パズル」「知識」「符号」

④倫理的配慮

検査実施にあたり、テスターは受検者に調査協力依頼状を提示して、研究の趣旨やプライバシー保護等の倫理的配慮を説明し、同意書への署名を求めた。なお、研究開始に先立ち、東京学芸大学研究倫理審査委員会による承認を受けた。

⑤調査期間

データの収集を2013年6月~11月に行い、分析を2013年11月~2014年2月に実施した。

(2) 調査の結果および考察

A. 下位検査と合成得点の基本統計量

①積木模様

粗点分布の基本統計量を表2-2に示した。

表2-2 積木模様における粗点の分布（問題項目数14、割増問題6、粗点最高66）

年齢群	16-19	20-24	25-29	30-34	全体
人数	82	79	79	81	321
平均値	53.41	54.46	55.52	55.32	54.67
中央値	57.00	57.00	56.00	57.00	57.00
標準偏差	9.44	8.94	6.93	7.04	8.17
最小値	24	24	34	34	24
最大値	65	66	65	65	66

年齢が上昇するにつれ平均が高くなり、また、標準偏差が小さくなる傾向が見て取れる。速い正答に対する割増点で天井効果の回避を試みたが、若干の天井効果が生じている可能性を考えら

れる。

②類似

粗点分布の基本統計量を表 2-3 に示した。

表 2-3 類似における粗点の分布（問題項目数 24、粗点最高 48）

年齢群	16-19	20-24	25-29	30-34	全体
人数	82	79	79	81	321
平均値	25.67	28.06	27.23	28.28	27.30
中央値	26.00	29.00	27.00	28.00	27.00
標準偏差	5.33	5.75	5.40	6.05	5.71
最小値	14	13	15	14	13
最大値	38	42	42	42	42

年齢が上昇するにつれ平均が高くなる傾向が見られるが、20~24 歳群の方が 25~29 歳群よりも平均が高くなっている。こうした傾向は類似だけでなく、数唱、行列推理、単語、算数など、多くの下位検査で見られたことから、サンプルの特性と考え、換算表の作成にあたっては、回帰曲線を推定することにより平滑化を行った。

平均が粗点最高とかなり離れており、また、平均の高い年齢群で標準偏差が小さくなる傾向は見られないことから、天井効果の可能性は低いと考えられる。

③数唱

粗点分布の基本統計量を表 2-4 に示した。

表 2-4 数唱における粗点の分布（問題項目数 16×3、粗点最高 48）

年齢群	16-19	20-24	25-29	30-34	全体
人数	82	79	79	81	321
平均値	29.49	31.08	29.80	30.59	30.23
中央値	29.50	31.00	29.00	29.00	30.00
標準偏差	4.29	5.06	5.21	5.44	5.03
最小値	21	21	21	20	20
最大値	41	42	44	46	46

3 つの年齢群で平均が 30 点付近になっているのに対し、20~24 歳群は平均 31.08 と若干高めになった。換算表の作成にあたっては、回帰曲線を推定することにより平滑化を行った。

④行列推理

粗点分布の基本統計量を表 2-5 に示した。

表 2-5 行列推理における粗点の分布（問題項目数 26、粗点最高 26）

年齢群	16-19	20-24	25-29	30-34	全体
人数	82	79	79	81	321
平均値	19.62	20.49	20.18	21.01	20.32
中央値	20.00	21.00	21.00	22.00	21.00
標準偏差	3.52	3.37	3.19	2.81	3.26
最小値	8	12	12	12	8
最大値	24	26	26	26	26

積木模様と同様、年齢が上昇するにつれ平均が高くなり、また、標準偏差が小さくなる傾向が見て取れる。天井効果が生じている可能性が考えられる。

なお、20~24 歳群の方が 25~29 歳群よりも若干平均が高くなっているが、類似や数唱と同様なので、サンプルの特性と考えられる。

⑤単語

粗点分布の基本統計量を表 2-6 に示した。

表 2-6 単語における粗点の分布（問題項目数絵 3+語 33、粗点最高 69）

年齢群	16-19	20-24	25-29	30-34	全体
人数	82	79	79	81	321
平均値	31.55	35.11	34.99	37.09	34.67
中央値	31.50	36.00	35.00	38.00	36.00
標準偏差	9.32	8.33	9.14	9.46	9.26
最小値	10	14	15	10	10
最大値	52	53	58	63	63

年齢が上昇するにつれ平均が高くなる傾向が見られるが、類似はじめ多くの下位検査がそうであるように、20~24 歳群の方が 25~29 歳群よりも若干平均が高くなっていることから、サンプルの特性と考えられる。

平均が粗点最高とかなり離れており、また、平均の高い年齢群で標準偏差が小さくなる傾向は見られないことから、天井効果の可能性は低いと考えられる。

⑥算数

粗点分布の基本統計量を表 2-7 に示した。

年齢が上昇するにつれ平均が高くなる傾向があるようにも見えるが、年齢群間で平均にほとんど差がないため、傾向ははつきりしない。特に類似はじめ多くの下位検査がそうであるように、20~24 歳群の方が 25~29 歳群よりも若干平均が高くなっていることから、いっそう傾向を不明瞭にしている。