

# 触法者におけるリスクアセスメントのための知的発達評価技法 の研究および開発

## —新田中 B 集団式知能検査の信頼性・妥当性の検討—

研究代表者

名古屋大学医学部附属病院、親と子どもの心療科

宇野洋太

### 1. 研究要旨

目的：本研究の目的は、短時間で実施が可能である集団式知能検査法のひとつである新田中 B 式知能検査の信頼性および併存妥当性を検証し、標準化したり、臨床的実用性を検討することである。

方法：少年鑑別所に入所した 81 例（平均年齢±標準偏差 15.2±0.7 歳）に対して新田中 B 式知能検査および Wechsler Intelligence Scale for Children-Third Edition (WISC-III) を実施した。信頼性はクロンバックの  $\alpha$  係数を用い評価した。併存妥当性は新田中 B 式知能検査の、WISC-III に対する級内相関係数を用い評価した。また WISC-III での全検査 IQ が 70 未満の知的能力障害をスクリーニングするため Receiver operating characteristic (ROC) 分析も行った。さらに知的能力障害を診断するための層別尤度比 (Stratum-specific likelihood ratios (SSLR)) も算出した。

結果：新田中 B 式知能検査の IQ (BIQ) のクロンバックの  $\alpha$  は 0.86 で、WISC-III の全検査 IQ に対する級内相関係数は 0.83 であった。ROC 分析の Area under the curve は 0.89 (95%信頼区間 0.85 - 0.96) であった。さらに BIQ 65 以下の層の SSLR は 13.8 (95%信頼区間 3.9 - 48.9)、BIQ 76 以上の層の SSLR は 0.1 (95%信頼区間 0.03 - 0.4) で

あり、知的能力障害を診断もしくは除外することができた。

結語：本研究の結果より、新田中 B 式知能検査が高い信頼性と WISC-III に対する併存妥当性を有することが示された。さらに、知的能力障害をスクリーニングする場合において、検査後確率も算出することができた。新田中 B 式知能検査は簡便で、様々な状況において実施することができる。このことは、現在知能検査の実施が困難な場面においても、今後知的発達の評価ができるようになる可能を示す。

### 2. 背景

知的発達に遅れなどの障害があると、学習面はもとより、身辺面や就業面を含めた自立機能、対人関係機能、コミュニケーションやコーピングスキルなど生活のあらゆる側面に影響を与える。したがって知的能力障害を有するものを、早期に発見・診断し、そのものに対して、早期から適切に養育・介入を開始することが重要である。しかし知的発達の遅れに気付かれず、適切にケアを受けられない場合には、社会への不適応がおこり、自尊心の低下や行動上あるいは情緒的な問題をきたすことが知られている (Benson et al., 1997,

Goodman et al., 1995, Lynam et al., 1993, Stattin and Klackenber-Larsson, 1993, Moffitt, 1993, Frick et al., 1991, Kaslow et al., 1990)。実際、犯罪を惹起したもののうち、知的に障害があるものが多く含まれているとの報告もある(Herrington, 2009, Vanny et al., 2009, Burke et al., 2012, Hayes et al., 2007)。ヘリントンらの報告によれば、イギリスの18歳から21歳の受刑者を対象に調査した結果、平均IQは84で、10.4%程度のものにIQ70未満の知的能力障害を認め、9.8%のものがIQ70 - 74、13.6%のものがIQ75 - 79であったと報告している(Herrington, 2009)。またヘイズらの報告によれば、イギリスのある刑務所において、21歳から70歳の受刑者を対象に調査した結果、平均IQは87.1で、7.1%程度のものにIQ70未満の知的能力障害を認め、23.6%のものがIQ70 - 79であったと報告している(Hayes et al., 2007)。

本邦においては、法務省矯正統計年報2011によれば、犯罪受刑者のうち、IQ70以下のものの割合は23%程度で、境界域も含めると45%以上のものが知的配慮を要するということがわかった。もっともそれらが全て知的能力障害かは検討の余地があるが、他の要因を差し引いても知的能力障害の割合が極めて高いであろうことは予想が付く。またそれにもかかわらず、逮捕以前から知的能力障害の診断を受けていたものは1%にも満たないということも特筆すべき点である。つまり、そのほとんどが非行や犯罪により保護、逮捕されてから初めて背後にある知的発達の問題に気付かされている。これらから元々知的発達の問題があるにもかかわらず、早い段階から発見・診断されていないために、適切な養育や配慮、サービスを受ける機会を失ってきてしまい、その結果、社会や家庭内で不適応をきたし、さらには非行・犯罪化しているケースが数多く存在しているということが示

唆される。

したがって子育てや教育等の場面において、個人の知的発達を含めた認知特性を把握し、それに合わせた教育や福祉あるいは医療上のサービスを提供することが重要である。また社会安全という観点からも、知的発達に遅れがあるものに適切なサービスを提供することは重要で、不幸にして犯罪を惹起したものに対しても知的発達の側面に配慮した取り組みが、再犯抑止のために求められている(Breckon et al., 2013, Mannynsalo et al., 2009, Lindsay et al., 2013)。そのためには個人の知的発達をそれぞれの場面において、適宜、適切に把握することが必要である。

知的発達を示す指数としては通常知能指数(IQ)が用いられる。個人の知的発達年齢(精神年齢)を、その方の生活年齢との比で示す(精神年齢÷生活年齢×100)方法と、同年齢集団内での位置で示す偏差知能指数とがある。昨今一般に使用されている検査のうち、新田中ビネーVの一部を除いて、多くは偏差知能指数を用いて算出している。

標準偏差を15とした偏差知能指数の場合、1標準偏差内である85 - 115の間には約68%の人が、また2標準偏差内である70 - 130の間には約95%の人が存在することになる。2標準偏差より大きく平均から乖離している場合に異常となり、適応機能と合わせて検討し、知的能力障害を診断することとなる。

またIQを測定するための検査法を一般に知能検査といい、測定する内容からA式、B式、あるいはAB混合式と分類することがある。A式は文章題などを用い、主に言語性の能力を測定する。一方B式では図形や数字などを主に用い、非言語的な面を測定する。よってB式知能検査は文化的、言語的な影響は受けにくい特徴があると考えられている。

知能検査として、世界的にも最も広く用いられ

ているもののひとつにウェクスラー (Wechsler) 式知能検査がある (Wechsler, 1991)。これは本邦でも標準化されており、最も一般的な検査のひとつである。本検査は標準偏差を 15 とし、IQ が概ね 40 を下限、160 を上限として算出される。対象とする年齢に応じて 3 種類に分かれる。3 歳 10 ヶ月から 7 歳 1 ヶ月を対象とする Wechsler Preschool and Primary Scale of Intelligence (WPPSI)、6 歳から 16 歳 11 ヶ月を対象とする Wechsler Intelligence Scale for Children (WISC)、16 歳以上を対象とする Wechsler Adult Intelligence Scale (WAIS) である。いずれも所定の検査器具を用い、個別に実施し、全般的な知能のほか、言語理解、知覚統合、作動記憶、処理速度などといった群指数や、知識や理解、数唱等といった様々な下位領域の能力を評価することができる。実施には 1~2 時間程度の時間を要し、検査者は同検査に精通していることが必要である。一部の低位検査項目のみで全体の知的発達水準を推測できる短縮版もある (Donders, 1997, Donders, 2001) もの、個別で実施する点、所定の検査器具を用いる必要がある点、実施者には検査への習熟が必要な点等にかわりはない。そのため、例えば矯正施設などのような、一度に多数のものが対象となり、一方で知的能力障害やメンタルヘルスケアの専門家が必ずしも多くない場面において、全例に個別式知能検査を実施するのは現実的ではない。したがって、簡便に実施できる知能検査もしくはスクリーニング検査が求められる。

一方で、ひとりの検査者が複数の被験者に対して同時に実施できる集団式知能検査法がある。著者の知りうるところでは、英語圏の国々においてはいくつかの集団式検査が標準化され、使用されている (Lohman and Hagen, 2002, Ahmann, 1985, Kuhlmann and Anderson, 1953, Raven, 2000, Raven et al., 1998, Shipley, 2009) が、本邦を

含む非英語圏においては標準化されているものはほとんどない。その数少ないもののひとつが新田中 B 式知能検査である (Tanaka et al., 2003)。本検査は多数の被験者に対して同時に実施でき、筆記用具と検査用紙のみで、特別な検査器具などを必要としない。また本検査は言語的なやり取りが簡易な教示のみで、複雑な言語的やり取りを必要としない。したがって、様々な教育背景、言語環境、言語レベルのものに対して実施できると考えられている。実施時間も 40 - 45 分程度と短時間であり、全体の知的発達を評価する上で非常に簡便である。本検査はスタンフォード・ビネー法をもとに 1936 年田中寛一によって開発され、その後約十年ごとの度重なる改定や再標準化が行われた。最近では 2001~2002 年に再標準化がなされ、因子構造の確認 (表 1.) と、高い折半法信頼性 ( $r = 0.89 - 0.96$ ) およびテスト-再テスト信頼性 ( $r = 0.73 - 0.79$ ) が確認された。また国語、数学、理科、社会、英語の各教科の学力試験の成績と新田中 B 式知能検査のスコアとの相関は、国語とは相関係数  $r = 0.42 - 0.56$ 、算数とは  $r = 0.50 - 0.66$ 、理科とは  $r = 0.49 - 0.67$ 、社会とは  $r = 0.43 - 0.52$ 、英語とは  $r = 0.52 - 0.61$  であり、またこれらの総合学力とは  $r = 0.69 - 0.78$  と強い相関を認め、高い妥当性が示された。(Tanaka et al., 2003)。

しかし学校成績は知的発達のみならず、教育的背景など様々な環境的要因の影響も受けることが予測され、知的発達を評価するための外部基準とは言い難い。新田中 B 式知能検査と他の知能検査との関連を調査した研究は著者の知り得るところ存在しない。

### 3. 目的

本研究では、集団式知能検査法のひとつである新田中 B 式知能検査と、国際的に最も広く使用さ

れている知能検査法のひとつである Wechsler Intelligence Scale for Children – Third Edition (WISC-III) との併存妥当性を検討した。さらに知的能力障害をスクリーニングするための臨床的有用性を検討した。本研究によって新田中 B 式知能検査が標準化されたなら、これまで知能検査の実施が困難であった場面においても、知能検査の実施が可能となる。そのことで適宜、的確に個人の知的発達を把握し、その者の目標を設定したり、それを達成するための方略を考えたりすることができるようになることに貢献出来ると思われる。

## 4. 方法

### A. 新田中 B 式知能検査

新田中 B 式知能検査では図形や絵画が用いられ、文章題は用いられていない(Tanaka et al., 2003, Matsuura et al., 2009)。回答の方法も図形や記号等を記載するため、読み書きなどの学習の差や言語的・文化的影響を受けにくいなどの特徴がある。具体的には、「迷路」、「立方体の計算」、「図形と数字の置き換え」、「異同弁別」、「数系列の完成」、「図形の抹消」、「図形の完成」といった7つの下位項目からなり、これらから『図形の認知速度』、『記号情報の思考(一部、言語情報の認知)』、『空間関係の思考』といった3つの知能因子を測定できる(表1.)。6~8歳を対象にしたもの、8~10歳を対象にしたもの、10~12歳を対象にしたもの、12~14歳を対象にしたもの、14歳~成人を対象にしたものと、対象年齢に応じ、5つに分かれている。今回はそのうち14歳以上を対象にしたものを用いた。

### B. 対象者

2009年1月1日から2010年12月31日までの間、ある少年鑑別所に入所したもののうち、新田

中 B 式知能検査及び WISC-III の両検査を受検した全 81 例である。検査はランダムに実施されており、両検査は互いに独立して行われた。検査の実施方法に従い、新田中 B 式知能検査は集団で、WISC-III は個別で実施された。調査期間内に一度は退所したものの、再度入所したものが 8 名いた。ただ、いずれも 2 回目の入所では検査は実施されておらず、同一被験者が重複して対象としてエントリーされていることはない。

対象者の平均年齢は 15.2±0.7 歳で、性別は男性 77 例、女性 4 例であった。知的発達は WISC-III の全検査 IQ (Full IQ; FIQ) において、平均 76.5 (標準偏差 15.0、幅 51~127) であった。全 81 例中 58 例 (71.6%) は FIQ < 85 で、さらに FIQ < 70 のものは 26 例 (32.1%) であった(図 1.)。しかしいずれのケースも、テスト以前に知的能力障害の存在に気付かれておらず、知的能力障害に対するサービスを受けていたものはひとりもいなかった。

また注意欠如・多動性障害と診断されたケースが 5 例、特定不能の広汎性発達障害、素行障害、身体化障害と診断されたケースがそれぞれ 1 例みられた。重複して診断名を持つケースはいなかった。診断は精神障害の診断と統計マニュアル第 4 版テキスト改訂版を用い、臨床経験豊富な精神科医師によって行われた。

なお少年鑑別所とは原則県庁所在地を中心に、各都道府県に 1 か所ずつ(ただし北海道は 4 か所、東京都は 2 か所、福岡県は支所含め 2 か所) 設置されている。非行や触法行為を行った都道府県内あるいは、非行・触法行為を行ったものが居住している都道府県内の鑑別所に、主として 12 歳から 20 歳未満までのものが、審判が行われるまでの間の最長で 8 週間、本人の評価と今後の教育方針を検討することを目的に収容される公的な機関で、法務省矯正局によって管理・運営されている。

入所した各個人は、その能力や養・教育ニーズ等を評価するため、入所後概ね3日以内に矯正局が定めたテストバッテリーによって評価される。そのテストバッテリーに新田中B式知能検査やパーソナリティ検査等が含まれ、いずれも集団で実施できるものである。検査は3～15名程度で同時に行われる。各鑑別所はそれに加え、独自の判断で臨的に必要だと思われる検査を個別に実施することができる。当該鑑別所において新田中B式知能検査はじめ、共通のテストバッテリーに含まれる検査のいくつかが十分に標準化されていないことなどもあり、独自の判断でウェクスラー式知能検査などを追加の検査として行うことがある。ウェクスラー式知能検査などは集団式検査後、審判までの間に、個別に、心理士によって実施される。

表 1. 新田中 B 式知能検査における知能因子および下位検査項目の構成

知能因子	下位検査項目
図形の認知速度	図形と数字の置き換え 異同弁別 図形の末梢 (迷路)
記号情報の思考	立方体の計算 数列の完成
空間関係の思考	図形の完成

### C. 統計学的解析

#### 1) 検出力の算出

本研究の有する検出力を求めるため、パワー分析を行った。新田中 B 式知能検査における IQ (BIQ) ならびに FIQ の標準偏差を 15 と見積もり、両者の差の 95%信頼区間が±5 以内となることを

確認するための検出力を算出した。

#### 2) 内的整合性

新田中 B 式知能検査において、折半法による信頼性と、テスト再テスト信頼性といった外的信頼性は既に確認されている。これらは繰り返し検査を実施した場合に同一の結果が得られるかの程度を示す指標であり、検査の再現性を意味する。しかし一方で新田中 B 式知能検査の検査項目である下位項目相互間の均質性は確認されていない。したがって、新田中 B 式知能検査の内的整合性を確認するために、同検査の迷路、立方体の計算、図形と数字の置き換え、異同弁別、数系列の完成、図形の抹消、図形の完成といった 7 つの下位項目についてクロンバックの  $\alpha$  係数を求めた。クロンバックの  $\alpha$  係数は可能であれば 0.8 以上が望ましいとされている。

#### 3) FIQ に対する BIQ の精度

BIQ と FIQ の分布を示し、両者の差を求めた。また BIQ の FIQ に対する偏りおよび精度を検討するために、平均パーセント誤差 (Mean percentage error; MPE) および二乗平均平方根誤差 (Root mean squared error; RMSE) を算出した (Sheiner and Beal, 1981)。

#### 4) 併存妥当性および臨床的有用性

学力検査の成績に対する新田中 B 式知能検査の妥当性は特に総合科目で相関係数  $r = 0.69 - 0.78$  と強い相関を認めた。ただ学力検査成績は知的発達以外の様々な要因が関与することは明らかであり、かつ併存妥当性を証明するためには、ゴールドスタンダードとの妥当性を検討する必要がある。本邦で既に標準化され、偏差知能指数を側的にしており、また個別式知能検査のうちもっとも汎用されているものはウェクスラー式知能検査で

ある。したがって、併存妥当性を証明するためには外的基準としてウェクスラー式知能検査を用いるのが適当と考える。WISC-III 検査に対する新田中B式知能検査の併存妥当性を検討するために、両検査における IQ スコアに関して、級内相関係数 (One-way Analysis of Variance Intraclass correlation coefficient (ANOVA ICC)) を求めた。

次に、FIQ 70 未満のものを新田中B式知能検査によってスクリーニングするためのパフォーマンスを求めるために、受信者動作特性 (Receiver Operating Characteristic (ROC)) 分析を行った。さらに層別尤度比 (Stratum-specific likelihood ratio; SSLR) を用い、BIQ 各層において、FIQ < 70 を同定するための尤度比および Posttest probability (検査後確率) を算出した (Furukawa et al., 1997, Furukawa et al., 2003)。

SSLR とは、ある疾患を有する者のうち、当該検査で陽性となる者の割合を、その疾患を有しない者のうち、同検査結果が陰性となる者の割合で割ったオッズである。結果として得られる値を、一定の値幅で層別化し、層ごとに SSLR や検査後確率を算出する。各層における SSLR は、

$$SSLR = (n1g / N1) / (n0g / N0)$$

で計算される。n1g は g 番目の層における、疾患を有するものの人数で、N1 は全体のうちの疾患を有するものの人数である。n0g は g 番目の層における、疾患を有さないものの人数で、N0 は全体のうちの疾患を有さないものの人数である。検査後確率と SSLR、検査前オッズ、検査後オッズとは次のような関係となっている；

$$\text{検査前オッズ} \times \text{SSLR} = \text{検査後オッズ}$$

また

検査後確率

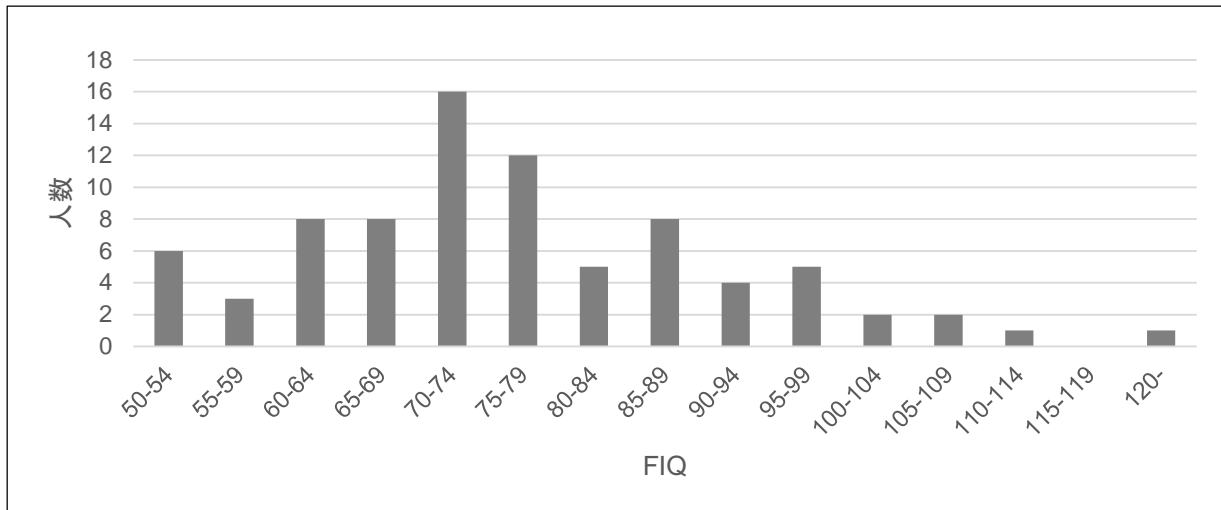
$$= (\text{検査後オッズ}) / (1 + \text{検査後オッズ})$$

で、算出される (Yang et al., 2004)。したがって、SSLR = 1 の場合、その検査の判別精度は偶然の確率に等しいことを示し、SSLR が 1 よりも大きいほど疾患の可能性が高くなり、1 よりも小さいほど疾患の可能性が低くなる。SSLR が 10 よりも大きかったり、反対に 0.1 よりも小さかったりする場合は、その疾患を確定もしくは除外することができる。従来、検査の妥当性の検討は、感度・特異度によるシングルカットオフポイントアプローチが行われてきた。この場合、カットオフポイントを満たさないものは、連続量で結果の得られるものにおいてもその程度に関係なく均質に扱われてしまうという欠点があった。つまりギリギリでカットオフポイントを満たさなかったものも、大幅にカットオフポイントを下回ったものも同等の重み付けでネガティブと捉えられてしまう。ところが実際にはこれらは偽陰性である確立などは違うはずである。したがって連続量で結果の得られるものにおいては、SSLR を用いることで、結果の値を層別化し、それぞれの層ごとに疾患の確率を算出できるようになる。

#### D. 倫理面への配慮

本研究は日本矯正医学会および名古屋大学医学部附属病院の生命倫理委員会の承認を受け、それに則り実施された。すべての対象者はすでに鑑別所を退所しており、追って同意を得ることが不可能であった。またさらに本研究に用いたデータはすべて臨床上、診断や介入を検討するために必要で得たデータであり、それを二次的、かつ回顧的に用いたものである。したがってインフォームドコンセントは得ていない。データは個人が特定されないように連結不可能匿名化し、解析を実施した。

図1. 全被験者におけるFIQの分布



Notes. 縦軸は人数で、横軸はFIQスコア. FIQ, WISC-III における全検査 (Full) IQ

## 5. 結果

### A. 検出力の算出と内的整合性

本研究のサンプルサイズ 81 例では検出力は 0.85 であった。またBIQの下位7項目におけるクロンバックの $\alpha$ 係数は 0.86 と、高い内的整合性が認められた。

### B. FIQに対するBIQの精度

図2~4はWISC-IIIの各IQスコアとBIQスコアの分布を示している。FIQは平均 $76.5 \pm 15.0$ であるのに対し、言語性IQ (Verbal IQ; VIQ)、および動作性IQ (Performance IQ; PIQ)の平均はそれぞれ $79.0 \pm 14.2$ 、 $78.8 \pm 15.7$ であった。またBIQは平均 $78.5 \pm 16.9$ で、FIQとの平均値の差は2.0 (標準誤差1.0, 95%信頼区間-4.1 - 0.1)であった。また平均パーセント誤差 $\pm$ 標準誤差は $0.03 \pm 0.01$ 、二乗平均平方根誤差は0.13であり、BIQのFIQに対する偏りはわずかであるとともに、高い予測精度を有することが確認された。

BIQのFIQに対するANOVA ICCは0.83と非常に高かった。またBIQはWISC-IIIにおけるVIQ、およびPIQの双方に対しても高いICC(それぞれICC = 0.72, 0.81)を認めた。このことはつまり、VIQもしくはPIQのいずれか一方のみを反映しているのではなく、言語性と動作性の双方およびIQ全体を反映していると考えられる(表2.)。

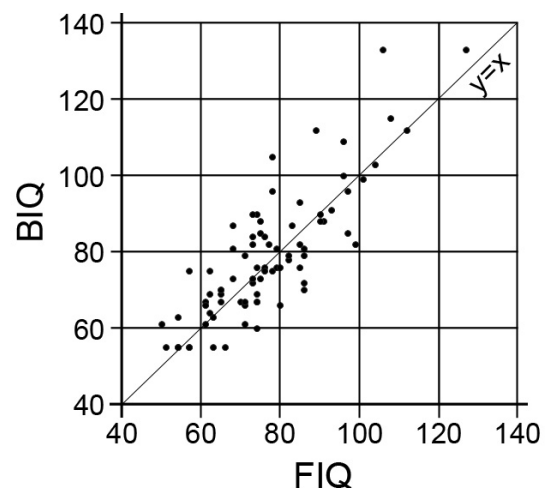


図2. BIQとFIQスコアの分布 (Uno et al., 2014)

Notes. 直線は $y=x$ の対角線. BIQ, 新田中B式知能検査におけるIQ; FIQ, WISC-III における全検査 (Full) IQ.

### 3) 基準関連妥当性と臨床的有用性

#### a. 級内相関係数 (ANOVA ICC)

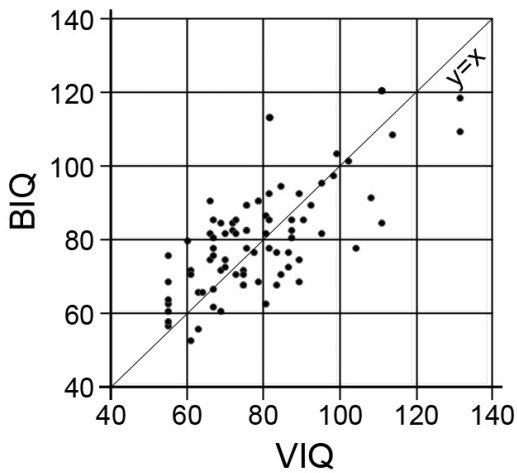


図3. BIQ と VIQ スコアの分布  
Notes. 直線は  $y=x$  の対角線. BIQ, 新田中 B 式知能検査における IQ; VIQ, WISC-III における言語性検査 (Verbal) IQ.

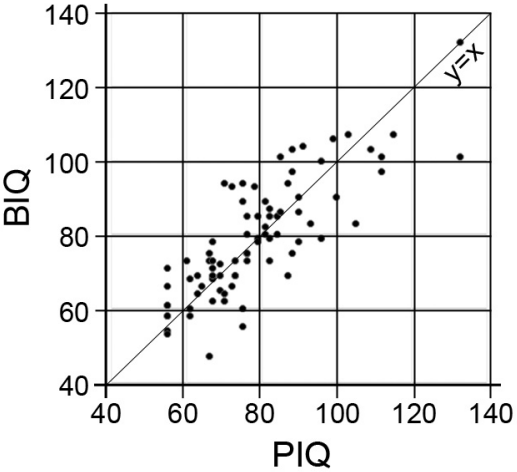


図4. BIQ と PIQ スコアの分布  
Notes. 直線は  $y=x$  の対角線. BIQ, 新田中 B 式知能検査における IQ; PIQ, WISC-III における動作性検査 (Performance) IQ.

b. ROC 分析および SSLRs

ROC 分析の結果、Area under the curve (AUC, 曲線下面積) は 0.89 (95% CI: 0.85 - 0.96) であった。AUC は ROC 曲線の下部の面積のことで、これが大きいことはすなわち、ROC 曲線が左上角に近いことを示す。つまり ROC 曲線の適合がよいことを意味する。よって本研究においては AUC が

大きいと、BIQ は高い識別能力を持って FIQ < 70 をスクリーニングすることができるということが示された (図5.)。

次に BIQ のスコアを 5 - 10 ポイントごとに層別化し、SSLR および検査後確率を算出した。BIQ 51 - 60 の層および BIQ 61 - 65 の層においては SSLR が 10 以上 (各層の検査後確率は 89, 83%) であり、FIQ < 70 をルールインすることができた。他方、BIQ 76 - 85 の層および BIQ 86 以上の層においては SSLR が 0.1 程度 (両層の検査後確率は 5%) であり、FIQ < 70 をルールアウトすることができた。

BIQ 66 - 70 の層および 71 - 75 の層においては SSLR がそれぞれ 2.4, 1.1 で、検査後確率は 33% および 53% であった (表3.)。BIQ 65 以下の群全体での SSLR は 13.8 (95% CI: 3.9 - 48.9, 検査後確率 87%) で FIQ < 70 のルールインが、BIQ 76 以上の群全体の SSLR は 0.11 (95% CI: 0.03 - 0.4, 検査後確率 5%) で FIQ < 70 のルールアウトが可能であった (表4.)。

表2. WISC-III の各 IQ 値とそれに対する BIQ の ICC

	平均 (標準偏差)	各 IQ に対する BIQ の ICC
<b>FIQ</b>	76.5 (15.0)	0.83
<b>VIQ</b>	79.0 (14.2)	0.72
<b>PIQ</b>	78.8 (15.7)	0.81
<b>BIQ</b>	78.5 (16.9)	-

Notes. ICC, 級内相関係数; BIQ, 新田中 B 式知能検査における IQ; FIQ, WISC-III における全検査 (Full) IQ; VIQ, WISC-III における言語性 (Verbal) IQ; PIQ, WISC-III における動作性 (Performance) IQ.



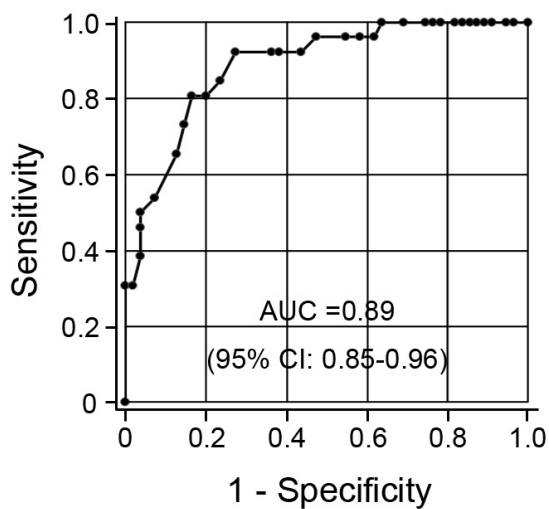


図 5. FIQ < 70 を識別するための BIQ の ROC 曲線とその AUC. (Uno et al., 2014)

Notes. FIQ, WISC-III における全検査 (Full) IQ; ROC 曲線, 受信者動作特性 (Receiver Operating Characteristic) 曲線; AUC, 曲線下面積 (area under the curve); 95% CI, 95% 信頼区間 (confidence interval); Sensitivity, 感度; Specificity, 特異度.

## 6. 考察

新田中 B 式知能検査は、集団式知能検査であり、実施時間が短く、特殊な検査器具も必要としないため、大勢に対して簡便に実施できる。本検査は、高い折半法信頼性 ( $r = 0.89 - 0.96$ )、テスト-再テスト信頼性 ( $r = 0.73 - 0.79$ )、また学習成績との高い妥当性 ( $r = 0.69 - 0.78$ ) を有することが示されていた。しかし、標準化された知能検査として本検査を用いるにはゴールドスタンダードな外的基準となり得る個別式知能検査との併存妥当性などの情報が不十分であった。したがって、本研究は新田中 B 式知能検査の、WISC-III に対する併存妥当性を検討したものである。さらに SSLR を用いて、新田中 B 式知能検査の臨床的な有用性も検討した。本研究の結果、14 歳以上を対象とした新田中 B 式知能検査は、高い内的整合性 ( $\alpha = 0.86$ ) および WISC-III の FIQ に対する高い併存

妥当性 (ANOVA ICC = 0.83) を有していることがわかった。したがって、WISC-III 等の個別式知能検査を実施することが困難である場合においても、簡便に実施できる集団式知能検査法である新田中 B 式知能検査を実施することで、知的発達を評価できることが示された。

表 3. FIQ < 70 に対する BIQ の SSLR と検査後確率

BIQ	人数		SSLR (95% CI)	検査後確率
	70 ≤ FIQ	FIQ < 70		
51-60	1	8	16.9 (3.2-90.2)	0.89
61-65	1	5	10.6 (1.9-60.6)	0.83
66-70	7	8	2.4 (1.0-5.8)	0.53
71-75	6	3	1.1 (0.3-3.6)	0.33
76-85	19	1	0.1 (0.02-0.6)	0.05
86 ≤	21	1	0.02 (0.02-0.5)	0.05
全体	55	26		

Notes. SSLR, 層別尤度比 (Stratum-specific likelihood ratios); 95% CI, 95% 信頼区間 (confidence interval); BIQ, 新田中 B 式知能検査における IQ; FIQ, WISC-III における全検査 (Full) IQ.

また FIQ < 70 をスクリーニングするに際して、ROC 分析の結果 AUC = 0.89 と高い識別能力を有していた。さらに新田中 B 式知能検査による IQ が 65 以下の層においては SSLR が 10 以上 (検査後確率 87%) で、知的能力障害ありと診断することが

できた。一方で、BIQ 76 以上の層では SSLR が 0.1 程度（検査後確率 5%）となり、知的能力障害なしと除外することができた。つまりこれらの層の間である BIQ が 66 - 75 のもので知的能力障害の有無の鑑別が必要となり、これらのものを対象に精密な知能検査を実施することで効率的に知的能力障害を診断することができることが判明した。これらから新田中 B 式知能検査が、知的能力障害をスクリーニングするのに高い識別能力を有した、簡便で、有効な検査法であるといえる。

表 4. FIQ < 70 に対する BIQ の SSLR と検査後確率

BIQ	人数		SSLR (95% CI)	検査後確率
	70 ≤ FIQ	FIQ < 70		
51 - 65	2	13	13.8 (3.9 - 48.9)	0.87
66 - 70	7	8	2.4 (1.0 - 5.8)	0.53
71 - 75	6	3	1.1 (0.3 - 3.6)	0.33
76 ≤	40	2	0.11 (0.03 - 0.4)	0.05
全体	55	26		

Notes. SSLR, 層別尤度比 (Stratum-specific likelihood ratios); 95% CI, 95% 信頼区間 (confidence interval); BIQ, 新田中 B 式知能検査における IQ; FIQ, WISC-III における全検査 (Full) IQ.

本検査は言語的な能力に直接関連する課題が少なく、図形や数字などを主に用いた B 式の知能検査である。したがって非言語的な側面のみを反映し、言語的な側面の問題を反映しないのではとの危惧もあるかもしれない。もっとも PIQ と ANOVA

ICC = 0.81 と強く相関を示すものの、VIQ との級内相関係数も ANOVA ICC = 0.72 と高く、動作性と言語性のいずれの能力も強く反映する結果となった。さらには FIQ との級内相関は係数 ANOVA ICC = 0.83 と最も高く、総合的な結果を最もよく反映していた。したがって言語性検査課題に対する能力での問題がある場合においても、新田中 B 式知能検査がスクリーニングとして十分有効であると考えられる。

本検査は、集団で、また 40 - 45 分の短時間で実施でき、検査用紙と筆記用具程度の検査器具しか必要としない。さらに検査者も特別なトレーニングを受ける必要がない。したがって、様々な状況で実施でき、例えば教育現場や司法施設のような、いままで WISC-III 等の実施が困難であった場面においても知的発達を評価できる可能性を広げるものと考えられる。さらに本検査は言語的なやり取りが簡易な教示のみで、複雑なやり取りを必要としない。したがって、様々な言語環境、言語レベルのものに対して実施できる可能性もある。さらに本研究により、シングルカットオフポイントを示すことではなく、SSLR を算出したことで、得られた結果に対する予測される検査後確率を求めることができた。このことは臨床的実用性を考える上で重要で、意義深いことと考える。

WISC-III が一般人口において標準化される際、標準偏差が 15.0 に設定されていることに比して、本研究の対象者の FIQ の標準偏差も 15.0 であった。つまりサンプル全体のバラツキは標準的であると考えられる。また知能の範囲も FIQ 51 - 127 と、一般的な範囲また知的能力障害の鑑別を要する範囲の概ねの層をカバーできている。さらに検査がランダムに行われたことで精査バイアスおよび鑑別所における範囲バイアスは小さいものと考えられる。しかし平均 FIQ が 76.5 ± 15.0 と低く、FIQ が 70 未満のものも全体の 32.1% と多い。これは、矯

正施設の入所者の平均 IQ は一般人口での IQ と比べ低いことが知られており (Douma et al., 2007, Anckarsater et al., 2007, Rayner et al., 2005, Harris and Bucens, 2003, Borum, 2003, Kroll et al., 2002, Lindsay, 2002, Hayes et al., 2007, Herrington, 2009)、そのことが影響していると考えられる。一方で FIQ が 50 以下となるような中重度の知的能力障害のあるものはみられなかった。したがって母集団を一般人口とみなすには偏りがあるものの、知的能力障害かどうかの境界に近い IQ 域のものが多く、本研究を進めるにあたりむしろそのことが強みとなった。今後は対象者を一般人口にさらにひろげて検討することが新田中 B 式知能検査を一般化して使用する際には必要となるかもしれない。さらには、鑑別所に収容されるものの性比から、本研究における対象者は男性が 95.1% と多くなった。したがって性別の影響も含め検討することも必要かもしれない。ただし、日本やアメリカ人で実施された研究において、VIQ, PIQ そして FIQ において、性差は無視できるとの報告もあり (Liu and Lynn, 2011)、性差を考慮する必要があるかは議論・検討の余地がある。

また知的能力障害以外でも様々な精神障害が認知機能の偏りを有していることが知られている。したがって、サンプルに精神障害を持つものが多く含まれる場合、サンプルバイアスを生じうる。本研究において、知的能力障害を除いた、注意欠如・多動性障害などの精神障害を持っているものの有病率は一般人口におけるそれらの障害の有病率と比べ、明らかに高いということにはなかった。したがって知的能力障害以外の精神障害が本研究結果に与える影響はあまり大きくないと考える。一方で、ほとんどの症例が非行・触法行為を惹起したり、それらの行為に関与しており、衝動性の問題や認知的な偏り等がみられたかもしれない。

非行や触法行為という点では一般人口と比較するとサンプルバイアスが存在する可能性は否定できない。

本研究により、集団式知能検査法である新田中 B 式知能検査は十分な信頼性および併存妥当性を持っていることが示された。また知的能力障害をスクリーニングする際の臨床的実用性も示された。今後はさらに対象を広げ、一般人口でのランダムサンプルを基に検討するなど、より幅広い場面でも妥当性を検証することが、一般化して使用していく上で求められると考える。さらには、本検査は簡単な教示のみで実施でき、言語的影響を受けにくいことから、今回調査した日本語圏のみならず、多文化的また多言語的な状況においても実施できる可能性を有し、今後検討できるかもしれない。知的発達が適切に評価され、それに基づいて適切に養育されたり、必要なサービスが提供され、知的能力障害のあるものやそのものを含めた周囲の人たちの生活の質が改善することを期待するとともに、本研究がそのことに貢献できることを望む。

## 7. 研究発表

### A. 論文発表

**Uno, Y., Mizukami, H., Ando, M., Yukihiro, R., Iwasaki, Y. & Ozaki, N. (2014). Reliability and validity of the new Tanaka B Intelligence Scale scores: a group intelligence test. *PLoS ONE*, 9, e100262.**

### B. 学会発表

なし

## 8. 利益相反

なし

## 9. 引用文献

- Ahmann, J. S. (1985). Otis-Lennon School Ability Test. *Measurement and Evaluation in Counseling and Development*, 17, 226-229.
- Anckarsater, H., Nilsson, T., Stahlberg, O., Gustafson, M., Saury, J. M., Rastam, M. & Gillberg, C. (2007). Prevalences and configurations of mental disorders among institutionalized adolescents. *Developmental Neuropsychology*, 10, 57-65.
- Benson, S., Bernstein, G. A., Bernet, W., Bukstein, O., Kinlan, J., Leonard, H., McClellan, J., Perry, M., Rogeness, G. & Shaw, J. (1997). Practice parameters for the assessment and treatment of children and adolescents with conduct disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 122S-139S.
- Borum, R. (2003). Managing At-Risk Juvenile Offenders in the Community Putting Evidence-Based Principles Into Practice. *Journal of Contemporary Criminal Justice*, 19, 114-137.
- Breckon, S. E., Smith, I. C. & Daiches, A. (2013). What makes offenders with an intellectual disability ready to engage with psychological therapy? A qualitative study. *Research in Developmental Disabilities*, 34, 1408-1416.
- Burke, M. M., Griggs, M., Dykens, E. M. & Hodapp, R. M. (2012). Defendants with intellectual disabilities and mental health diagnoses: faring in a mental health court. *Journal of Intellectual Disability Research*, 56, 305-316.
- Donders, J. (1997). A short form of the WISC-III for clinical use. *Psychological Assessment*, 9, 15.
- Donders, J. (2001). Using a short form of the WISC-III: sinful or smart? *Child Neuropsychology*, 7, 99-103.
- Douma, J. C., Dekker, M. C., De Ruiter, K. P., Tick, N. T. & Koot, H. M. (2007). Antisocial and delinquent behaviors in youths with mild or borderline disabilities. *American Journal on Mental Retardation*, 112, 207-220.
- Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A., Hart, E. L. & Tannenbaum, L. E. (1991). Academic underachievement and the disruptive behavior disorders. *Journal of Consulting and Clinical Psychology*, 59, 289.
- Furukawa, T., Hirai, T., Kitamura, T. & Takahashi, K. (1997). Application of the Center for Epidemiologic Studies Depression Scale among first-visit psychiatric patients: a new approach to improve its performance. *Journal of affective disorders*, 46, 1-13.
- Furukawa, T. A., Kessler, R. C., Slade, T. & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological medicine*, 33, 357-362.
- Goodman, R., Simonoff, E. & Stevenson, J. (1995). The impact of child IQ, parent IQ and sibling IQ on child behavioural deviance scores. *Journal of Child Psychology and Psychiatry*, 36, 409-425.
- Harris, K. R. & Bucens, I. K. (2003). Prevalence of fetal alcohol syndrome in the Top End of the Northern Territory. *Journal of paediatrics and child health*, 39, 528-533.
- Hayes, S., Shackell, P., Mottram, P. & Lancaster, R. (2007). The prevalence of intellectual disability in a major UK prison. *British Journal of Learning Disabilities*, 35, 162-167.
- Herrington, V. (2009). Assessing the prevalence of intellectual disability among young male prisoners. *Journal of Intellectual Disability Research*, 53, 397-410.
- Kaslow, F. W., Lipsitt, P. D., Buka, S. L. & Lipsitt, L. P. (1990). Family law issues in family therapy practice: Early intelligence scores and subsequent delinquency: A Prospective study. *The American Journal of Family Therapy*, 18, 197-208.
- Kroll, L., Rothwell, J., Bradley, D., Shah, P., Bailey, S. & Harrington, R. C. (2002). Mental health needs of boys in secure care for serious or persistent offending: a prospective, longitudinal study. *The Lancet*, 359, 1975-1979.
- Kuhlmann, F. & Anderson, R. G. (1953). Kuhlmann-Anderson

- intelligence tests. *Journal of Consulting Psychology*, 17, 320.
- Lindsay, W. R. (2002). Integration of recent reviews on offenders with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 15, 111-119.
- Lindsay, W. R., Holland, A. J., Carson, D., Taylor, J. L., O'Brien, G., Steptoe, L. & Wheeler, J. (2013). Responsivity to criminogenic need in forensic intellectual disability services. *Journal of Intellectual Disability Research*, 57, 172-181.
- Liu, J. & Lynn, R. (2011). Factor structure and sex differences on the Wechsler Preschool and Primary Scale of Intelligence in China, Japan and United States. *Pers Individ Dif*, 50, 1222-1226.
- Lohman, D. F. & Hagen, E. P. (2002). *CogAT, Form 6: Research Handbook*: Riverside Pub.
- Lynam, D., Moffitt, T. E. & Stouthamer-Loeber, M. (1993). Explaining the relation between IQ and delinquency: Class, race, test motivation, school failure, or self-control? *Journal of Abnormal Psychology*, 102, 187.
- Mannynsalo, L., Putkonen, H., Lindberg, N. & Kotilainen, I. (2009). Forensic psychiatric perspective on criminality associated with intellectual disability: a nationwide register-based study. *Journal of Intellectual Disability Research*, 53, 279-288.
- Matsuura, N., Hashimoto, T. & Toichi, M. (2009). A structural model of causal influence between aggression and psychological traits: survey of female correctional facility in Japan. *Children and Youth Services Review*, 31, 577-583.
- Moffitt, T. E. (1993). The neuropsychology of conduct disorder. *Development and psychopathology*, 5, 135-135.
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive psychology*, 41, 1-48.
- Raven, J., Raven, J. C. & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*, Oxford, UK: Oxford Psychologists Press.
- Rayner, J., Kelly, T. & Graham, F. (2005). Mental health, personality and cognitive problems in persistent adolescent offenders require long-term solutions: a pilot study. *Journal of Forensic Psychiatry & Psychology*, 16, 248-262.
- Sheiner, L. B. & Beal, S. L. (1981). Some suggestions for measuring predictive performance. *Journal of pharmacokinetics and biopharmaceutics*, 9, 503-512.
- Shipley, W. C. (2009). *Shipley-2: Manual*: Western Psychological Services.
- Stattin, H. & Klackenber-Larsson, I. (1993). Early language and intelligence development and their relationship to future criminal behavior. *Journal of Abnormal Psychology*, 102, 369.
- Tanaka, K., Okamoto, K. & Tanaka, H. (2003). *The New Tanaka B Intelligence Scale*, Tokyo: Kaneko shobo.
- Uno, Y., Mizukami, H., Ando, M., Yukihiro, R., Iwasaki, Y. & Ozaki, N. (2014). Reliability and validity of the new Tanaka B Intelligence Scale scores: a group intelligence test. *PLoS ONE*, 9, e100262.
- Vanny, K. A., Levy, M. H., Greenberg, D. M. & Hayes, S. C. (2009). Mental illness and intellectual disability in Magistrates Courts in New South Wales, Australia. *Journal of Intellectual Disability Research*, 53, 289-297.
- Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children*: Psychological Corporation San Antonio, TX.
- Yang, H. J., Soong, W. T., Kuo, P. H., Chang, H. L. & Chen, W. J. (2004). Using the CES-D in a two-phase survey for depressive disorders among nonreferred adolescents in Taipei: a stratum-specific likelihood ratio analysis. *Journal of affective disorders*, 82, 419-430.