HOKKAIDO UNIVERSITY

Title	精神疾患における日本の臨床現場で実施可能な社会認知機能検査の検討
Author(s)	秋山, 久
Description	配架番号：2863
Degree Grantor	北海道大学
Degree Name	博士(医学)
Dissertation Number	甲第16054号
Issue Date	2024-06-28
DOI	https://doi.org/10.14943/doctoral.k16054
Doc URL	https://hdl.handle.net/2115/92756
Type	doctoral thesis
File Information	AKIYAMA_Hisashi.pdf, この論文の全文

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文

精神疾患における日本の臨床現場で実施可能な社会認知機能検査の検討

（A study of social cognitive measurements for psychiatric disorders in Japanese clinical practice）

2024 年 6月

北海道大学

秋山　久

学位論文

精神疾患における日本の臨床現場で実施可能な社会認知機能検査の検討

（A study of social cognitive measurements for psychiatric disorders in Japanese clinical practice）

2024 年 6月

北海道大学

秋山　久

目次

発表論文目録および学会発表目録・・・・・・・・・・・・・・・・・・・・ 1 頁

要旨・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 2 頁

略語表・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 5 頁

緒言・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 6 頁

研究方法・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 9 頁

研究結果・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 16 頁

考察・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 25 頁

総括および結論・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 29 頁

謝辞・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 31 頁

利益相反・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 31 頁

引用文献・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 32 頁

発表論文目録および学会発表目録

発表論文目録および学会発表目録

本研究の一部は以下の論文に投稿中である。

The Evaluation Study for Social Cognition Measures in Japan: Psychometric properties and recommendations

Hisashi Akiyama, Ryo Okubo, Atsuhito Toyomaki, Akane Miyazaki, Sachiko Hattori, Mariko Nohara, Yohei Sasaki, Ryotaro Kubota, Hiroki Okano, Kanami Takahashi, Yumi Hasegawa, Izumi Wada, Takashi Uchino, Kazuyoshi Takeda, Satoru Ikezawa, Takahiro Nemoto, Yoichi M Ito, Naoki Hashimoto

Asian Journal of Psychiatry, in submission

本研究の一部は以下の学会に発表した。

1.　秋山　久

精神疾患における日本の臨床現場で実施可能な社会認知機能検査の検討

第 118 回日本精神神経学会学術総会、 2022 年 6 月 16 日、福岡

2.　秋山　久

日本の臨床現場で実施可能な社会認知機能検査の信頼性・妥当性の検討

BPCNPNPPP4 学会合同年会（第 44 回日本生物学的精神医学会年会・第 32 回日本臨床精神神経薬理学会年会・第 52 回日本神経精神薬理学会年会・第 6 回日本精神薬学会総会・学術総会）、2022 年 11 月 5 日、東京

要旨

【背景と目的】

社会認知機能は「他者の意図や感情を理解する人間としての能力を含む、対人関係の基礎となる精神活動」と定義され、その障害は統合失調症をはじめとする精神疾患をもつ患者の対人関係や就学・就労などの社会参加を阻む大きな要因の一つとされている。そのためさまざまな側面から社会認知機能障害を評価する検査が開発されたが、検査の乱立や各検査の信頼性・妥当性の検討が十分なされていないことが課題となっていた。このような背景から 2012～2017 年にかけて米国において、社会認知機能検査に対する計量心理学的研究（Social Cognition Psychometric Evaluation：SCOPE 研究）が行われ、108 個の候補検査から最終的に 6 個の検査が精神疾患の社会認知機能検査として推奨された。しかし SCOPE 研究で推奨された各検査について、英語以外の言語や米国とは異なる文化圏における信頼性、妥当性の検証はまだ十分ではない。2020 年に Lim らがシンガポールで行った信頼性妥当性研究では、必ずしも SCOPE 研究の結果が再現されず、英語を公用語としている人々であっても文化圏が異なると、信頼性妥当性の評価の結果が異なる可能性が示唆された。今後、社会認知機能障害に対する治療法の開発を進めていく上でも、基盤となる症状評価に対する文化差の影響の評価はきわめて重要である。また Uchino らの報告によれば、統合失調症の患者は社会認知に大きな困難を抱え、それが社会認知機能障害と強く関連していると自覚しているにも関わらず、社会認知に関する検査や治療を受けている者はほとんどいなかった。つまり社会認知機能障害はアンメット・ニーズであり、日本の臨床場面で実施可能な社会認知機能検査が明らかとなることで、精神障害を持つ患者の社会認知機能を定量的に評価し、治療につながることが期待される。また今後の国際治験で統一的な評価が可能となることで、新規の治療法開発につながることも期待される。以上の背景から、今回我々は日本の臨床現場で実施可能な社会認知機能検査の検討を行うこととした。

【対象と方法】

本研究は国立精神・神経医療研究センター倫理委員会および北海道大学病院自主臨床研究審査委員会の承認を得て、2021 年 3 月～2023 年 3 月にかけて実施された。本研究の前段階として、国内のエキスパートパネルによる 2 回の投票・会議を実施し、研究で実施する 9 個の社会認知機能検査（the Bell Lysaker Emotion Recognition Task（BLERT）、Facial Emotion Selection Test （FEST）、Hinting task（Hinting）、Metaphor and sarcasm scenario test（MSST）、Ambiguous Intentions and Hostility Questionnaire（AIHQ）、Intentionality Bias Task（IBT）、Social Attribution Task-Multiple Choice（SAT-MC）、SAT-MCII、Biological motion task（BM））を選定し

た。研究実施施設は北海道大学病院、国立精神・神経医療研究センター、東邦大学医療センター大森病院、国立病院機構帯広病院の 4 施設であり、目標症例数は統合失調症をもつ患者 140 例、健常者 70 例とした。患者群の選択基準は、本研究に参加する施設に通院中の患者であり、1）評価時の主診断が DSM-V の統合失調症に該当する患者、2）過去 2 ヶ月にわたって入院がなく、過去 6 週間にわたって向精神薬の種類に変更がなく、過去 2 週間にわたって向精神薬の用量の変更がない者、3）同意取得時において年齢が 20 歳以上 59 歳以下の者、4）本研究の目的・内容を理解し、研究への参加同意を文書によって取得できる者、とした。また健常者の選択基準は、1）同意取得時において、20 歳以上 59 歳以下の者、2）同意取得時において、精神疾患の診断を受けていない者、3）本研究の目的・内容を理解し、研究への参加同意を文書によって取得できる者、とした。各研究参加者に研究内容を文書を用いて説明し、文書による参加同意を得た上で、参加基準を満たすか確認した後、1 週間以内に初回の評価および検査を実施した。社会認知機能検査は初回検査後、約 1 ヶ月後に再検査を実施した。主要評価項目は階層的重回帰分析による社会機能の増分妥当性とし、副次評価項目は各社会認知機能検査の信頼性・妥当性および社会機能との関連とした。

【結果】

解析対象者は患者群が 121 名、健常対照者群が 70 名だった。すべての検査が 15 分以内に施行可能であった。また各検査における被験者の主観的な満足度評価に大きな問題はなかった。再検査信頼性については、患者群における IBT と BM 以外のすべての検査で級内相関係数が 0.5 を超えた。学習効果については、SAT-MCII で Cohen’s d=0.83 の高い学習効果を認めた。天井効果および床効果については、MSST で 1 割程度の患者に床効果、5 割以上の患者に天井効果を認め、SAT-MC で 1 割程度の患者に天井効果を認めた。患者群と健常対照者群との比較では、ほとんどの検査において患者群で有意に成績が低かった一方で、AIHQ においては Cohen’s d=-0.42 とやや小さかったほか、IBT と BM においては有意差を認めなかった。社会機能との関連では、Hinting、MSST、SAT-MC、SAT-MCII、BM が日常生活技能の評価である University of California, San Diego Performance-based Skill Assessment- Brief（UPSA-B）に対して有意な関連を認め、FEST が実世界の社会機能を評価する尺度である Specific Levels of Functioning Scale（SLOF）に対して有意な関連を認めた。階層的重回帰分析では UPSA-B に対して Hinting の増分妥当性が認められた。

【考察】

神経認知機能の影響を考慮した上で、Hinting の社会機能に対する増分妥当性が認められた。また BLERT、FEST、Hinting、MSST、AIHQ、SAT-MC、SAT-MCIIは再検査信頼性が確認された。SLOF に対して FEST の、UPSA-B に対して Hinting、MSST、SAT-MC、SAT-MCII、BM の収束的妥当性が確認された。先行研究から一貫

してHintingの良好な計量心理学的特性が確認された一方で、BLERTとIBTの計量心理学的特性は先行研究に比し、劣る結果であった。BLERTは文化的差異によるコミュニケーション方法の違いが、IBTは言語の読字時間という要素が結果に影響した可能性がある。

【結論】

今回の研究により、各検査の日本語版の信頼性と妥当性が評価され、今後の国際臨床試験で使用できる統一的な評価方法が提供された。このことは、統合失調症をもつ患者の社会認知機能障害に対する治療法の開発の基盤となる成果である。今後、各検査の特性に留意しながら日本の臨床現場で活用されることで、社会認知機能障害に関するさらなる知見の蓄積が望まれる。

略語表

本文中及び表中で使用した略語は以下の通りである。

AIHQ	Ambiguous Intentions and Hostility Questionnaire
BACS-J	Brief Assessment of Cognition for Schizophrenia, Japanese Version
BLERT	Bell Lysaker Emotion Recognition Task
BM	Biological motion task
CP	chlorpromazine
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, fifth Edition
ER-40	Penn Emotion Recognition Task
FEST	Facial Emotion Selection Test
Hinting	Hinting Task
IBT	Intentionality Bias Task
ICC	intraclass correlation coefficients
IQ-PANSS	Informant Questionnaire for the PANSS
JART-25	Japanese adult reading test-25
M.I.N.I.	Mini-International Neuropsychiatric Interview （精神疾患簡易構造化面接法）
MSST	Metaphor and sarcasm scenario test
PANSS	Positive and Negative Syndrome Scale （陽性陰性症状評価尺度）
SAT-MC	Social Attribution Task-Multiple Choice
SAT-MCII	Social Attribution Task-Multiple Choice II
SCI-PANSS	Structured Clinical Interview for the PANSS
SCOPE	Social Cognition Psychometric Evaluation
SLOF	Specific Levels of Functioning Scale （特定機能レベル評価尺度）
UPSA-B	University of California, San Diego（UCSD）Performance-based Skill Assessment-Brief （日常生活技能簡易評価尺度）

1. 緒言

1.1　統合失調症と認知機能障害

統合失調症は、幻覚や妄想などの陽性症状、情動表出の減少や意欲欠如などの陰性症状、まとまりのない発語や行動などの解体症状を特徴とする疾患であり、しばしば機能障害をもたらす（American Psychiatric Association, 2013）。統合失調症の有病率は 1%と比較的低いが (McGrath *et al*, 2008)、世界的に見ても障害要因の上位 15 位の一つである (G. B. D. Disease Injury, 2017)。統合失調症をもつ患者の 80%以上は完治せず、良好な転帰をたどった患者であっても、その診断は人生を変える結果をもたらす (Jaaskelainen *et al*, 2013; Jauhar *et al*, 2022)。発症早期は幻覚や妄想などの陽性症状に特徴づけられるが、その前駆期から認知機能障害を生じており、このことが統合失調症をもつ患者の機能的転帰により影響していると考えられている。しかし、神経認知機能で説明できる機能的転帰の変動は一般的に小さく、20～40%であるとする研究が大半であることが示されており (Green *et al*, 2000)、このことは機能的転帰の分散の大部分がまだ説明されていないことを示唆している (Kurtz *et al*, 2001; Medalia & Choi, 2009; Wykes *et al*, 2011)。その中において社会認知機能は、神経認知機能よりも社会機能に強く関連すると考えられており (Pinkham & Penn, 2006)、近年ますます注目されている。

1.2　社会認知機能について

社会認知機能とは、「他者の意図や感情を理解する人間としての能力を含む、対人関係の基礎となる精神活動」と定義され (Green *et al*, 2008)、その障害は精神病症状の発症前から出現し、寛解後も残存すると考えられている (Addington *et al*, 2008; Pinkham *et al*, 2007)。社会認知機能障害は精神症状より強く、神経認知機能障害より直接的に社会的転帰に関与するとされ (Brekke *et al*, 2005; Nuechterlein *et al*, 2011)、患者の社会参加能力や社会的関係の構築・維持能力に直接影響し、生活の質を著しく低下させる。統合失調症をもつ患者は社会認知に大きな困難を抱えており、それが社会機能と強く関連していると自覚しているにも関わらず、社会認知に関する検査や治療を受けている者はほとんどいない (Uchino *et al*, 2022)。このことは、統合失調症をもつ患者における社会認知機能障害は、アンメット・ニーズであることを示唆している。

統合失調症においては以下の 4 領域の社会認知が障害されると考えられている。1 つ目は情動認知であり、他者の感情を知覚および自らの感情を表出する能力のことであり、障害されると表情に現れた相手の感情を識別することが困難になる。2 つ目は心の理論で、他者の意図、意向を推し量る能力であり、障害されると他者との意思疎通を妨げ、自分自身の行動をセルフモニタリングすることが困難になる。3 つ目は原因帰属バイアスで、特定の出来事の因果関係を理解する能力である。障害されると、

特に不快な出来事を体験したときに、自己帰属、他者帰属、外部帰属の 3 つの帰属様式のうち、自己ではなく、外部や他者に原因を帰属する傾向が強くなる。4 つ目の社会知覚は、文脈や身振り、仕草などを手がかりから社会的状況を理解する能力であり、障害により身振り・仕草などを手がかりに対人関係を理解することが困難になる。

1.3　社会認知機能検査について

これまで社会認知機能の様々な側面を評価するために多くの測定法が開発されてきたが、異なる理論や解釈に基づく多種多様な測定法が存在し、信頼性や妥当性が十分検証されていない検査が多く (Pinkham *et al*, 2014)、社会認知機能障害を測定する検査が十分に確立しているとはいえない。こうした背景から 2012～2017 年にかけて米国において、社会認知機能検査に対する計量心理学的研究（Social Cognition Psychometric Evaluation：SCOPE 研究）が行われた (Pinkham *et al*, 2018; Pinkham *et al*, 2016)。SCOPE 研究は、統合失調症における社会認知機能を前述の 4 つの中核領域（情動認知、心の理論、原因帰属バイアス、社会知覚）に分けて評価するというコンセンサスをもとに、既存の社会認知機能検査の計量心理学的指標を系統的に評価することを目的として行われた。その結果、統合失調症をもつ患者群と健常対照者群を対象とした 2 つの大規模研究により、108 個の候補検査から最終的に 6 つの検査が精神疾患の社会認知機能検査として推奨された。特に Hinting task（Hinting）、Bell Lysaker Emotion Recognition Task（BLERT）、Penn Emotion Recognition Task（ER-40）の 3 つが強い計量心理学的特性と機能予後との関連を示し、臨床試験での使用が推奨された。SCOPE 研究は、社会認知機能障害の評価において大きな進歩をもたらし、暫定的なテストバッテリーを提供することとなった。

1.4　社会認知機能検査における文化差、言語差について

SCOPE 研究は今後の社会認知機能に関する研究において使用されうる社会認知機能検査を提案したが、この結果は米国のみで収集されたデータに基づいており、他の文化的背景への一般化可能性は不明であった。社会認知機能検査は、神経認知機能検査よりも文化的・言語的な差異に敏感であるとされ、社会認知機能には基本的な文化的差異があると考えられている (Hajduk *et al*, 2020)。また言語的な違いは、社会認知機能検査の計量心理学的特性にさらに影響を与える可能性があり (Frank, 2010)、英語で作成された測定法が他言語において同等にその性質を引き継ぐかは不明確である。このことから各検査の文化的妥当性や翻訳可能性は各文化に対して個別に検討される必要があると考えられる。さらに SCOPE 研究では 4 つの社会認知領域のうち、原因帰属バイアスおよび社会知覚の領域を測定する検査が推奨されなかったことも限界点としてあげられており、今後の課題であった (Pinkham *et al.*, 2018)。

近年行われたアジア圏における社会認知機能検査の研究は、特定の社会認知機能領域に限定されるか、測定法の選択に一貫性がなかった (Chen *et al*, 2017; Lee *et al*, 2018; Mehta *et al*, 2011)。Lim らがシンガポールでの統合失調症をもつ患者と健常対照者を対象に、SCOPE 研究の枠組みと同様に社会認知機能検査の計量心理学的研究（SCOPE Asia）を実施したところ、BLERT と ER-40 が最も良好な計量心理学的特性を示し、SCOPE 研究と同様の結果が得られた一方で、Hinting は再検査信頼性が低く、異なる結果を示した (Lim *et al*, 2020)。このことは言語が共有されていても、社会認知機能検査が文化的背景の異なる集団の間では異なる計量心理学的特性を示す可能性があることを改めて示唆した。

以上のことから、SCOPE 研究の結果を異なる言語圏、文化圏で適用するためには、改めて各地域における社会認知機能検査の信頼性と妥当性を検討することが重要であると考えられた。翻訳された課題や英語以外の言語で作成された検査を含む、非英語圏の人々を対象にした包括的な計量心理学的研究は、私たちの知る限りではまだ行われていない。

1.5　本研究の目的

本研究は、SCOPE 研究の枠組みを用いて、非英語圏かつアジア圏にある本邦において、社会認知機能検査の計量心理学的特性を包括的に評価した初めての研究である。SCOPE 研究と同様に、事前に専門家パネルが結成され、SCOPE 研究で推奨された 6 つの検査と日本で開発された 2 つの検査を含む 15 個の候補検査の適合性を評価・議論し、最終的に社会認知機能の 4 つの中核領域を代表する 9 つの検査を選定した (Okano *et al*, 2021)。本研究の目的は、パネル会議で選定された 9 つの社会認知機能検査の日本語版について、計量心理学的特性と社会機能との関連を検討するとともに、各検査の信頼性・妥当性の検討を行うことである。

2. 研究方法

2.1 研究デザイン

本研究は多施設による横断観察研究であり、北海道大学病院、国立精神・神経医療研究センター、東邦大学医療センター大森病院、国立病院機構帯広病院で行われた。この研究は、人を対象とする医学系研究に関する倫理指針に基づいて計画され、国立精神・神経医療研究センター倫理委員会および北海道大学病院独立臨床研究審査委員会の承認を得た。実施期間は 2021 年 3 月～2023 年 3 月であり、すべての被験者から書面で同意を取得し、実施された。

2.2 対象

研究の被験者は以下の組み入れ基準および除外基準を満たすことを確認した上で、口頭および書面によるインフォームド・コンセントを行った。患者群は北海道大学病院、国立精神・神経医療研究センター、東邦大学医療センター大森病院、国立病院機構帯広病院に通院中であり、統合失調症と診断され、病状が安定している外来患者から募集された。健常者は、上記の研究実施施設に掲示または配布されるチラシやメーリングリスト、臨床試験管理会社を通じて募集された (Kubota *et al*, 2021)。

患者群：組み入れ基準

本研究を実施する施設に外来通院中であり、

1）評価時に主診断が Diagnostic and Statistical Manual of Mental Disorders Fifth Edition（DSM-5）の診断基準に基づいて統合失調症と診断される者

2）過去 2 ヶ月にわたって入院がなく、過去 6 週間にわたって向精神薬の種類に変更がなく、過去 2 週間にわたって向精神薬の用量の変更がない者

3）同意取得時において年齢が 20 歳以上 59 歳以下の者

4）本研究の目的・内容を理解し、研究への参加同意能力のある者

組み入れ基準の設定にあたっては、一般的に病状が不安定な患者に対して、負荷のかかる神経心理学的検査は実施しないことが多いことを考慮した、本研究においても病状の安定した患者を選別するために、入院の有無や薬剤の処方状況に関して上記の通りとした。また将来的に社会認知機能を評価する目的として、精神疾患をもつ患者のリハビリテーションに活かし、良好な対人関係や社会参加を促進することにあるため、対人関係や社会参加の改善を目指して行われるリハビリテーションが主に 20 歳以上 59 歳以下の外来患者を対象に行われることから、上記の基準を設定した。

健常対照者群：組み入れ基準
1）同意取得時において、20 歳以上 59 歳以下の者
2）同意取得時において、精神疾患の診断を受けていない者
3）本研究の目的・内容を理解し、研究への参加同意能力のある者

患者群/健常対照者群：除外基準
1）評価時において、検査が実施できないような身体疾患・精神疾患が認められる者
2）質問内容を十分に理解したうえで、自記式心理検査に回答ができる日本語能力がない者
3）その他、主治医や研究スタッフが参加不可能と判断した者
除外基準の設定にあたっては、検査に耐えられない疾患がある場合、研究参加に伴う被験者のデメリットが大きいことや自記式の検査項目は被験者が自分自身あるいは他者の状態をどのように評価しているかを測定する目的で実施されるため、基本的な日本語能力が必要となることを考慮した。また本人は参加を希望しているが、病状を鑑みると主治医が参加を認められない等、特別な事情がある場合も想定されるため、上記の除外基準を設定した。

2.3 調査項目および実施する評価尺度

・背景情報：患者群、健常対照者群
性別、年齢、本人および両親の学歴、既往歴、アレルギー歴、現在の内服薬などをカルテ、もしくは参加者への聴き取りにより収集した。

・精神疾患簡易構造化面接法（Mini-International Neuropsychiatric Interview: M.I.N.I.）：患者群、健常対照者群
精神疾患を診断するために作成された構造化面接法である (Otsubo *et al*, 2005; Sheehan *et al*, 1998)。本研究では版権管理者の許可を得て、DSM-V に対応した版を用いた。
患者群および健常対照者群の参加者に対して同意取得後にそれぞれの組入基準を満たすかどうかを確認するために実施した。

・Japanese adult reading test-25（JART-25）：患者群、健常対照者群
JART-25 は、25 個の二字熟語を音読することで言語性 IQ を測定する検査である。統合失調症をもつ患者においては、JART-25 の結果は病前の言語性 IQ を反映すると考えられている (Matsuoka *et al*, 2006)。

・陽性陰性症状評価尺度（Positive and Negative Syndrome Scale: PANSS）：患者群のみ
統合失調症の精神症状全般を面接により評価するものである (Kay *et al*, 1987)。この尺度は陽性症状 7 項目、陰性症状 7 項目、一般精神症状 16 項目の計 30 項目からなる。本研究では、日本若手精神科医の会（Japan Young Psychiatrists Organization: JYPO）が翻訳し信頼性妥当性が検証されている PANSS の日本語版を使用した (Hashimoto *et al*, 2020)。本研究で半構造化面接である Structured Clinical Interview for the PANSS（SCI-PANSS）を使用し、情報提供者がいる場合は Informant Questionnaire for the PANSS（IQ-PANSS）を用いた。使用にあたり、版権管理者から許可を得た。

神経認知機能：患者群のみ
・BACS-J（Brief Assessment of Cognition for Schizophrenia, Japanese Version）
統合失調症の認知機能障害を評価するために開発された、標準的な検査バッテリーである (Keefe *et al*, 2004)。本研究では信頼性、妥当性が検証されている日本語版を使用した (Kaneda *et al*, 2007)。本検査は言語記憶、ワーキングメモリ、運動速度、言語流暢性、注意、実行機能の下位項目で構成されている。

社会機能：患者群のみ
・日常生活技能簡易評価尺度（UCSD Performance-based Skill Assessment-Brief: UPSA-B）
社会機能のうち、機能的能力を評価する検査である (Mausbach *et al*, 2007)。日常生活を想定したロールプレイングを行い、「金銭管理」と「コミュニケーション（電話）」の 2 つの領域を測定する。本研究では信頼性、妥当性が検証されている日本語版を使用した (Sumiyoshi *et al*, 2014)。

・特定機能レベル評価尺度（Specific Levels of Functioning Scale: SLOF）
社会機能のうち、実世界の機能障害を評価する尺度であり (Schneider & Struening, 1983)、信頼性、妥当性が検証されている日本語版を使用した (Sumiyoshi *et al*, 2016)。SLOF は自己評価版、情報提供者版、評価者版の 3 つを含む。そのうち、情報提供者版と自己評価版の結果を踏まえ採点される評価者版の得点が最も適切に実世界の機能障害を反映すると考え (Sabbag *et al*, 2011)、本研究の主な評価値とした。

2.4　社会認知機能検査

社会認知機能検査：患者群、健常対照者群

・情動認知

1.Bell Lysaker Emotion Recognition Task（BLERT）

被験者が動画を視聴し、俳優の表情、声のトーン、声色、上半身の動きなどを手がかりに、どの感情を表現しているかを選択式で回答する検査である。21 本のビデオで構成され、俳優が 3 つの異なる台詞を使って幸福、悲しみ、恐怖、嫌悪、驚き、怒り、無感情の 7 つの異なる感情を演じている (Bryson *et al*, 1997)。被験者は各動画で表現されている感情を 1 つ回答し、本研究ではその正解数を評価値とした。原著者の許可を得て、日本人の俳優を起用し、翻訳した台本で新たに日本語版を作成した。

2.Facial Emotion Selection Test（FEST）

性別や年齢の異なる日本人の顔写真を見て、表情の感情を選ぶ検査である (Hagiya *et al*, 2015)。この検査は、Facial Emotion Identification Task （FEIT） (Kerr & Neale, 1993)を参考に日本語版が作成され、信頼性および妥当性が検証されている。「喜び」「悲しみ」「怒り」「恐怖」「驚き」「嫌悪」「感情なし」の 7 つのいずれかの感情を示した 21 枚の顔写真が順に提示され、被験者はそれぞれの写真を見て、どの感情に該当するか回答した。本研究ではその正答数を評価値とした。

・心の理論

3.Hinting Task（Hinting）

2 人の登場人物の対話を読んで聞き、一方の登場人物の発言の真意を推測して回答する検査である (Corcoran *et al*, 1995)。本研究ではすでに存在する日本語版の検査をもとに、原作者の許可を得て、以下の手順で参加者自身がパソコンで検査を受ける仕様とした。原版では、最初の指示で不正解だった場合のみ、発言の意図を明確にする追加の会話文が提示され、被験者は再度回答するが、本研究では最初の回答の正誤にかかわらず、追加の会話文を提示し、被験者は再度回答を行った。被験者は声に出して質問に答え、その回答は録音され、一度目で正答すれば 2 点、追加教示を受けて正答すれば 1 点とし、10 通りのシナリオの評点の合計を評価値とした。各施設の評価者でオンライン会議を行い、各施設での評価基準は統一した。

4.Metaphor and sarcasm scenario test（MSST）

比喩や皮肉を含む文章を読み、その文章の意図を最も正確に表現している答えを 5 つの選択肢から選ぶ検査である (Adachi *et al*, 2004)。この検査は自閉スペクトラム症の小児に対する評価のために日本で開発された検査であり、原著者の許可を得て使用し

た。検査の構成は、比喩文 5 問、皮肉文を 5 問の計 10 問であるが、本研究では皮肉文の正答数を評価値とした。

・原因帰属バイアス

5.Ambiguous Intentions and Hostility Questionnaire（AIHQ）

自身に不利益が生じる各シナリオにおいて、相手の行動を意図的に感じる程度、怒りを感じる程度、相手のせいに感じる程度をそれぞれリッカート尺度で回答する検査である (Combs *et al*, 2007)。本研究では原著者の許可を得た日本語版を使用した (Kanie *et al*, 2014)。不利益が偶然によるシナリオと意図的か偶然か曖昧なシナリオの 2 種類 5 通りずつ合計 10 通りのシナリオを用い (Buck *et al*, 2017)、その各回答のリッカート尺度の平均を評価値とした。

6.Intentionality Bias Task（IBT）

制限時間内に人間の行動を表す 24 個の短文に回答し、その行動が意図的なものか偶発的なものかを答える検査である (Rosset, 2008)。原著者の許可を得て、新たに IBT の日本語版を作成した (Kubota *et al.*, 2021)。日本語に翻訳する際、英語と日本語の文法構造や平均読解速度の違いを考慮して制限時間を変更した (Trauzettel-Klosinski *et al*, 2012)。具体的には、速い条件では 2.4 秒、遅い条件では 5 秒であった制限時間を、それぞれ 3 秒、6 秒に延長した。全体の回答数に対する「意図的」と回答した比率を算出し、本研究での評価値とした。

・社会知覚

7.Social Attribution Task – Multiple Choice（SAT-MC）/ 8.SAT-MCII

簡単なストーリーに沿って、擬人化した図形が「喧嘩する」「逃げる」といった社会的行動を思わせる動きをするアニメーションを見て、その動きの意味や動機を回答する検査である (Bell *et al*, 2010)。被験者は動画を 2 回閲覧した後、19 個の質問に対して 4 つの選択肢から回答する。SAT-MCIIは、SAT-MC の代用あるいは同等のものとして、後に開発され (Johannesen *et al*, 2013)、動画の長さ、質問数、評価方法は SAT-MC と同じである。SCOPE 研究では、これら 2 つの検査を 1 つの検査の 2 つの版として、初回テストと再テストでそれぞれ実施されたが、2 つの版は基本的な難易度に差があり、そのために再検査信頼性などの基本的な計量心理学的指標で劣る結果を示した (Pinkham *et al.*, 2018)。そこで本研究では、これら 2 つの検査をそれぞれ独立した検査として実施した。本研究での実施にあたり SAT-MC および SAT-MCIIの日本語版を新たに作成し、原著者の許可を得た。19 個の質問に対する正答数を本研究での評価値とした。

9.Biological motion task（BM）
人の動作が十数個の光点の動作として表現される映像（Biological Motion; BM）と、同じ動作について光点の初期位置をランダムにすることで非生物的な動きに見える映像（Scrambled Motion; SM）の 2 つがランダムに提示され、被験者は毎試行その映像が BM と SM のどちらであるかボタン押して回答する検査である (Grossman *et al*, 2004)。言語的な影響を受けない検査であり、本研究では、Hashimoto らの研究で使用された動画をもとに (Hashimoto *et al*, 2014)、以下の手順でランダムに動く光点の数を増減させた。参加者が正解した場合は、次の試行でランダムに動く 12 個の光点が追加され、不正解の場合は、次の試行でランダムに動く 12 個の光点が減らされた。正解から不正解、もしくは不正解から正解の回答をした回数（“reversals”）を数え、13 回 reversals に達した後、増減する光点の数を 6 個に減少させた。28 回 reversals に達した時点で検査を終了とし、最後の試行から 10 回の試行におけるランダムに動く光点の数の平均を弁別閾値として本研究での評価値とした。

上記のうち、日本語版がない SAT-MC、SAT-MCIIおよび IBT については、原著者の許可を得て日本語に翻訳した。その他の検査については、すでに入手可能な日本語版を使用した。BM を除くすべての検査はオンラインで利用できるようにし、BM は各施設のコンピュータに実装した。日本語版の検査がオンラインで実施されることについて、原著者から許可を得た。各施設を訪れた参加者は、心理士の監督のもと自ら端末を操作し検査を実施した。社会認知機能検査の結果は、本研究のために構築されたプラットフォームに直接登録された。

2.5 研究の手順

リクルートされた被験者に対して、各施設の担当者が文書を用いて説明し、文書同意を取得した。同意取得後にそれぞれの適格性の確認のために、医師が M.I.N.I.を実施した。背景情報をカルテ、もしくは参加者への聴き取りにより収集するとともに、JART-25 と PANSS を実施した。神経認知機能の評価として BACS-J を、社会機能の評価には UPSA-B と SLOF を実施した。年齢、性別などの背景情報および JART-25、PANSS、BACS-J、UPSA-B、SLOF は紙媒体で収集及び実施され、その結果は各施設から本研究のために構築されたプラットフォームに登録された。初回の社会認知機能検査は同意取得から 1 週間以内に実施され、2～6 週間後に再検査が行われた。各社会認知機能検査の実施順は疲労や集中力の低下の影響が特定の検査の成績に影響しないように、乱数表を用いカウンターバランスをとった。また、すべての社会認知機能検査について快適性と忍容性が評価された。快適性は参加者がその検査の主観的な満足度を 1（非常に不快）から 7（非常に快適）までのリッカート尺度で評価し

た。忍容性は説明を含む各検査の実施に要した時間とし、15 分以内の検査であれば忍容性ありと判断することとした。各社会認知機能検査はコンピュータ上で動作する仕様であり、各施設に訪れた被験者自身で端末を操作し、検査を実施し、その結果はそのままプラットフォームに登録された。

2.6　統計解析方法

本研究では以下の計量心理学的指標が算出されるが、その中でも当該社会認知機能検査が社会機能を独自に予測できる検査かどうかを重視することとした。主要評価項目は階層的重回帰分析による社会機能の増分妥当性、すなわち神経認知機能検査を独立変数としてあらかじめ組み込んで決定係数を算出後、社会認知機能検査を独立変数に追加して新たに解析を行なった場合に決定係数が有意に増加すること、と定義した。先行研究では (Pinkham *et al.*, 2018)、社会機能に対する神経認知機能の効果量は 0.25 であり、本研究では、社会認知機能検査の追加により効果量が 0.35 程度に引き上げられる社会認知機能検査を検出するためのサンプル数を設定した。α=0.05、β=0.2 で検出するための必要サンプル数は 99 例であったが、先行研究では 40%と高い脱落率であったことを考慮し、本研究でも 30%程度の脱落率を想定し、患者群の目標症例数を 140 例とした。健常対照者群については、各社会認知機能検査の一般集団における標準値を探索的に推定するために、被験者の年代および性別の比率を一致させるように、20 代男女、30 代男女、40 代男女、50 代男女の 8 区分に対してそれぞれ 8 人前後のリクルートを行ない、脱落も考慮し、目標症例数を 70 例とした。

参加者の背景情報については、患者群と健常対照者群に分けて要約統計量を算出し、群間で比較した。再検査信頼性は Pearson's r および級内相関係数（ICC）で評価した。Pearson's r≥0.6 および ICC>0.5 を許容範囲とした (Koo & Li, 2016)。反復測定としての有用性は Cohen's d を算出し、学習効果を評価し、さらに床・天井効果（偶然レベルまたは 100%得点した参加者の数）のそれぞれの頻度を示し評価した。なお Cohen's d は 0.5 未満を許容範囲とし、床・天井効果はそれぞれ全体の 10%以下であることを許容範囲とした。年齢と性別を説明変数に加えた重回帰分析を行い、各社会認知機能検査が社会機能をどの程度予測するかを総合的に判断した。また、各社会認知機能検査について、社会機能を従属変数とした階層的重回帰分析を行い、社会機能の増分妥当性を検討した。さらに、各社会認知機能検査に要した時間（忍容性）、および各検査に対する参加者の主観的評価（快適性）の平均値と標準偏差（SD）を患者群と健常対照者群に分けて算出した。各社会認知機能検査の平均値と標準偏差を患者群と健常対照者群に分けて算出し、両群間の平均値を比較し、Cohen's d を示した。すべての解析は、R バージョン 4.0.4（https://www.r-project.org）を用いて実施した。

3. 研究結果

3.1 対象者の背景

研究参加者は合計 200 名で、そのうち 129 名が統合失調症をもつ患者、71 名が健常対照者だった。適格基準の確認のため M.I.N.I.を実施し、診断が統合失調症でない患者 1 名、およびアルコール使用障害が疑われた健常者 1 名が除外された。また同意撤回した 7 名の患者も解析から除外され、最終的な解析対象者は 191 名（患者群 121 名、健常対照者群 70 名）であった（図 1）。

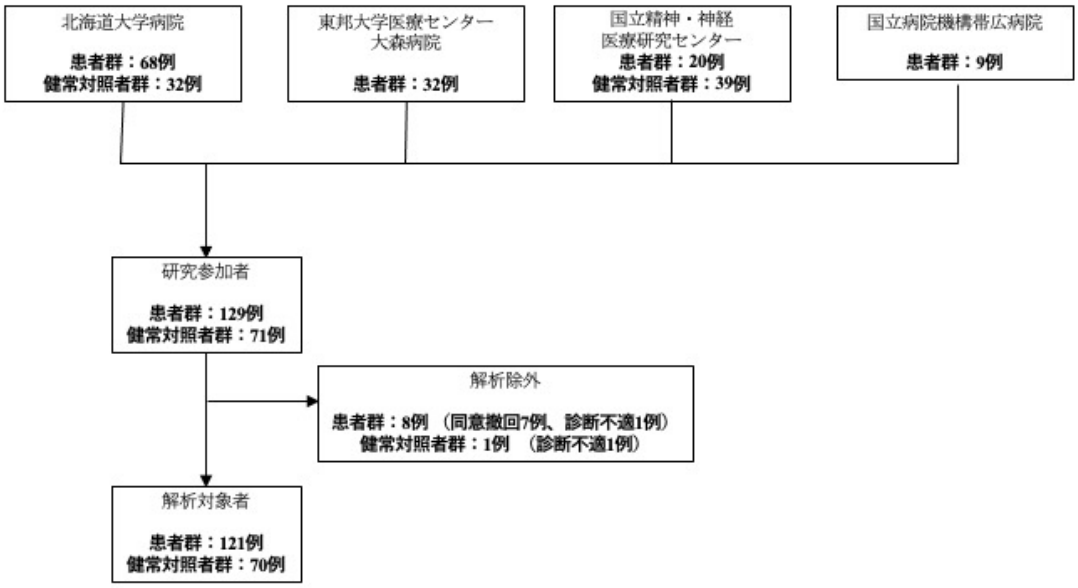

図 1：研究参加者の流れ

両群間に年齢・性別の有意差は認めなかった。本人の教育歴および JART−25 で有意差を認めたが、両親の教育歴には有意差は認めなかった。また患者群の PANSS 総得点の平均は 71.1 点であり、抗精神病薬のクロルプロマジン換算は 555.9mg だった（表 1）。

表 1：背景情報

性別以外の項目では、括弧内は標準偏差（standard deviation: SD）を示す。患者群と健常対照者群に対して性別については $\chi 2$ 乗検定を行い、年齢・教育歴・JART-25 についてはt検定を行った。それぞれの検定について有意水準（p 値）は 5%未満とした。* : $p < 0.001$

略語: JART-25 = Japanese Adult Reading Test-25; FSIQ = Full scale intelligence quotient; PANSS = Positive and Negative Syndrome Scale; CPZ = chlorpromazine

	患者群（121 例）		健常対照者群（70 例）	
性別（女性、%）	64	(52.9)	39	(55.7)
年齢（歳）	39.5	(10.4)	39.3	(11.2)
本人の教育歴（年）*	13.8	(2.3)	15.7	(2.3)
父の教育歴（年）	13.3	(3.0)	14.0	(3.0)
母の教育歴（年）	12.9	(2.1)	13.3	(2.0)
JART-25（FSIQ） *	104.6	(9.7)	109.6	(6.9)
PANSS 総得点（点）	71.1	(14.7)		
CPZ 換算値（mg/日）	555.9	(424.1)		

3.2 再検査信頼性

再検査信頼性は、大半の検査で許容範囲内であった。患者群においてはIBT（r = 0.418, ICC=0.419）とBM（r = 0.377, ICC = 0.370）が不十分な値を示した。健常対照者群では、BLERT、FEST およびHintingでPearson's r が 0.6 を下回ったが、ICC は 0.5 以上のため、再検査信頼性は保たれていると考えられた（表 2）。

表 2：再検査信頼性

有意水準（p 値）は 5%未満とし、***: p < 0.001 で示した。

略語: r = Pearson's r; ICC = intraclass correlation coefficients; BLERT = Bell Lysaker Emotion Recognition Task; FEST= Facial Emotion Identification Test; Hinting = Hinting Task; MSST = Metaphor and Sarcasm Scenario Test; AIHQ = Ambiguous Intentions and Hostility Questionnaire; IBT = Intentionality Bias Task; SAT-MC = Social Attribution Test-Multiple Choice; SAT-MCII = Social Attribution Test-Multiple Choice II; BM = Biological motion task

	患者群			健常対照者群		
検査	N	r	ICC	N	r	ICC
BLERT	89	0.654***	0.647***	36	0.531***	0.529***
FEST	89	0.608***	0.592***	36	0.529***	0.535***
Hinting	89	0.748***	0.729***	35	0.598***	0.527***
MSST	89	0.742***	0.733***	36	0.899***	0.896***
AIHQ	89	0.861***	0.858***	36	0.792***	0.796***
IBT	88	0.418***	0.419***	36	0.618***	0.580***
SAT-MC	89	0.737***	0.724***	36	0.834***	0.816***
SAT-MCII	89	0.751***	0.680***	36	0.657***	0.653***
BM	89	0.377***	0.370***	35	0.617***	0.560***

3.3 反復測定の有用性

患者群では、SAT-MCIIは学習効果が大きく（Cohen's d = 0.83）、反復測定で結果が大きく改善することが示唆された。健常対照者群では Hinting、IBT、BM において高い効果量が観察された。また MSST は、床効果が 1 割程度の患者（89 名中 10 名）が床効果を認めた一方で、半数以上で天井効果（89 名中 51 名）を示した。SAT-MC では 1 割程度の患者（89 名中 14 名）で天井効果を認めた（表 3）。

表 3：反復測定の有用性

略語: SD = standard deviation; BLERT = Bell Lysaker Emotion Recognition Task; FEST= Facial Emotion Identification Test; Hinting = Hinting Task; MSST = Metaphor and Sarcasm Scenario Test; AIHQ = Ambiguous Intentions and Hostility Questionnaire; IBT = Intentionality Bias Task; SAT-MC = Social Attribution Test-Multiple Choice; SAT-MCII = Social Attribution Test-Multiple Choice II; BM = Biological motion task

		初回検査（T1）		再検査（T2）		床/天井効果の人数				
検査	N	平均	(SD)	平均	(SD)	T1	T2	t	p	Cohen's d
患者群（n=89）										
BLERT	89	16.3	(3.3)	16.9	(3.4)	1/1	0/4	1.92	0.059	0.29
FEST	89	14.4	(2.8)	15.1	(2.8)	0/0	1/0	2.68	0.009	0.40
Hinting	89	12.7	(3.8)	13.6	(3.5)	0/1	1/0	2.94	0.004	0.44
MSST	89	4.0	(1.4)	4.3	(1.4)	10/51	8/61	2.18	0.032	0.33
AIHQ	89	6.9	(2.1)	7.1	(2.0)	-	-	1.74	0.085	0.26
IBT	88	0.5	(0.2)	0.6	(0.2)	-	-	0.68	0.497	0.10
SAT-MC	89	13.7	(4.9)	14.4	(4.3)	4/14	2/13	1.97	0.052	0.29
SAT-MCII	89	10.5	(3.6)	12.0	(3.0)	4/0	0/0	5.56	<0.001	0.83
BM	89	21.3	(16.2)	24.4	(18.3)	6/0	10/0	1.65	0.103	0.25
健常対照者群（n=36）										
BLERT	36	18.8	(1.7)	18.6	(1.9)	0/7	0/2	-0.85	0.401	-0.20
FEST	36	16.6	(2.1)	16.6	(2.2)	0/0	0/0	0.00	1.000	0.00
Hinting	35	16.0	(2.5)	17.0	(1.9)	0/1	0/2	3.02	0.005	0.72
MSST	36	4.7	(0.8)	4.8	(0.9)	1/30	2/33	0.81	0.422	0.19
AIHQ	36	5.9	(1.4)	5.9	(1.4)	-	-	-0.33	0.744	-0.08
IBT	36	0.5	(0.2)	0.5	(0.2)	-	-	-2.66	0.012	-0.63
SAT-MC	36	16.4	(3.2)	16.8	(2.6)	1/6	0/5	1.04	0.307	0.24
SAT-MCII	36	14.7	(2.3)	15.1	(2.2)	0/0	0/0	1.14	0.262	0.27
BM	35	24.8	(16.3)	32.5	(20.3)	2/0	3/0	2.76	0.009	0.60

3.4 社会機能との関連

説明変数として年齢及び性別を加えて各社会認知機能検査および神経認知機能の各下位項目について社会機能に対する重回帰分析を行ない、それぞれの標準化偏回帰係数 b を示した（表 4）。神経認知機能は BACS-J の下位項目の Z-scores を用いて評価した。UPSA-B に対して神経認知機能のすべての下位項目で有意な関連を示した。これらの関係の大きさは、中程度（0.23-0.43）であった。一方で、SLOF に対して、有意な関連を示した下位項目はなかった。社会認知機能検査では、Hinting、MSST、SAT-MC、SAT-MCII、BM が UPSA-B に対して有意な関連を認め、FEST が SLOF に対して有意な関連を示した。BLERT、AIHQ、IBT は UPSA-B、SLOF どちらの社会機能に対しても有意な関連が得られなかった。

表 4：社会機能に対する各神経認知機能および各社会認知機能検査の重回帰分析

有意水準（p 値）は 5%未満とし、*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ で示した。

略語: UPSA-B = UCSD Performance-based Skill Assessment-Brief total score; SLOF = Specific Levels of Functioning Scale total evaluator score; b = standardized partial regression coefficient

	UPSA-B			SLOF		
	N	b		N	b	
社会認知機能						
BLERT	120	0.169		121	-0.036	
FEST	120	-0.009		121	0.209	*
Hinting	120	0.502	***	121	0.125	
MSST	120	0.288	**	121	0.174	
AIHQ	120	0.060		121	-0.043	
IBT	118	-0.076		119	0.009	
SAT-MC	120	0.345	***	121	0.154	
SAT-MCII	120	0.248	**	121	0.143	
BM	120	0.186	*	121	-0.059	
神経認知機能						
言語記憶	120	0.394	***	120	0.073	
ワーキングメモリ	120	0.309	***	120	0.156	
運動速度	120	0.227	*	120	0.159	
言語流暢性	120	0.328	***	120	0.026	
注意	120	0.425	***	120	0.061	
実行機能	119	0.302	***	119	0.059	

妥当性は、各社会機能と有意に関連する社会認知機能検査を予測因子とする回帰モデルによって評価した。社会認知機能検査は、UPSA-B の分散の 24%（調整済み R^2 = 0.245, F（7,109）= 6.38, p < 0.001）、SLOF の 4%（調整済み R^2 = 0.044, F（3,113）= 2.78, p = 0.044）を有意に占めた（表 5）。

表 5：社会機能に対する社会認知機能検査全体での重回帰分析

略語: aR^2 = Adjusted R^2; UPSA-B = UCSD Performance-based Skill Assessment-Brief total score; SLOF = Specific Levels of Functioning Scale total evaluator score; b = standardized partial regression coefficient; sr^2 = squared semi-partial correlation coefficient

	R^2	aR^2	F	p	b	t	p	sr^2
UPSA-B	0.291	0.245	6.377	<0.001				
Hinting					0.430	4.533	<0.001	0.115
MSST					0.053	0.556	0.579	0.002
SAT-MC					0.133	1.281	0.203	0.009
SAT-MCII					-0.022	-0.217	0.829	0.000
BM					0.046	0.538	0.591	0.002
SLOF	0.069	0.044	2.782	0.044				
FEST					0.199	2.169	0.032	0.039

社会認知機能検査の増分妥当性は、神経認知機能検査以上の分散を有意に予測するかどうかで評価した（表 6）。神経認知機能は BACS-J の下位項目の Z-scores を用いて評価した。神経認知機能検査のみで、UPSA-B の分散の 20%（調整済み R^2 = 0.198, F（8,108）= 4.58, p < 0.001）を有意に占めた。第 2 ブロックとして神経認知機能検査の後に社会認知機能検査を変数に加えたところ、UPSA-B スコアの分散の 11%（R^2 変化量 = 0.108, p = 0.006）を追加的に説明した。なかでも Hinting の寄与が大きかった。SLOF に対して年齢及び性別を説明変数に加えて行った重回帰分析では神経認知機能検査の下位項目で有意な関連を認めた項目はなかったが、同様に階層的重回帰分析を行った。神経認知検査のみで、SLOF の分散の 2%（調整済み R^2 = 0.02, F（8,108）= 1.29, p = 0.255）を占めたが有意ではなかった。第 2 ブロックとして FEST を加えたところ SLOF の分散の 3%を示したが、そのモデルおよび変化量ともに統計学的に有意ではなかった（R^2 変化量 = 0.02、p = 0.113）。

表 6：階層的重回帰分析

有意水準（p 値）は 5%未満とし、*:p < 0.05, **:p < 0.01, ***:p < 0.001 で示した。

略語: aR^2 = Adjusted R^2; UPSA-B = UCSD Performance-based Skill Assessment - Brief total score; SLOF = Specific Levels of Functioning Scale total evaluator score; b = standardized partial regression coefficient; sr^2 = squared semi-partial correlation coefficient

	UPSA-B				SLOF	
	b		sr^2		b	sr^2
Block1-神経認知						
言語記憶	0.162		0.008		-0.058	0.001
ワーキングメモリ	0.020		0.000		0.237	0.018
運動速度	-0.025		0.000		0.176	0.016
言語流暢性	0.106		0.003		-0.117	0.003
注意	0.240		0.015		-0.017	0.000
実行機能	0.155		0.019		-0.019	0.000
Block2-社会認知						
FEST	-		-		0.154	0.021
Hinting	0.358	***	0.065	**	-	-
MSST	0.025		0.000		-	-
SAT-MC	0.063		0.002		-	-
SAT-MCII	-0.051		0.001		-	-
BM	0.051		0.003		-	-
Overall Model						
aR^2	0.280	***			0.034	
R^2 change	0.108	**			0.022	

3.5 忍容性と快適性

忍容性はすべての検査で 15 分以内であり、許容範囲であった。Hinting、SAT-MC、SAT-MCII、BM では検査の実施に平均 10 分以上の時間がかかったが、残りの検査はおおむね 5 分前後で実施可能であった（表 7）。快適性は健常対照者群における BM 以外はどの検査も平均が 4（普通）以上の結果であった。

表 7：検査の快適性と忍容性

忍容性は検査の実施に要した時間（単位：分）で示し、快適性は被験者の各検査に対する主観的な満足度を 1（非常に不快）から 7（非常に快適）までのリッカート尺度で評価した。略語: SD = standard deviation

	忍容性		快適性	
	患者群 (n=120)	健常対照者群 (n=70)	患者群 (n=120)	健常対照者群 (n=70)
検査	平均 (SD)	平均 (SD)	平均 (SD)	平均 (SD)
BLERT	6.8 (0.9)	6.4 (0.5)	4.4 (1.2)	4.7 (1.2)
FEST	2.6 (1.1)	2.1 (0.6)	4.4 (1.2)	4.9 (1.3)
Hinting	13.6 (2.8)	12.7 (1.6)	4.3 (1.3)	4.5 (1.3)
MSST	3.9 (1.6)	3.1 (1.0)	4.6 (1.3)	5.1 (1.1)
AIHQ	4.8 (1.9)	4.1 (1.3)	4.5 (1.1)	5.1 (1.2)
IBT	3.9 (1.0)	3.4 (0.5)	4.1 (1.1)	4.5 (1.4)
SAT-MC	10.8 (2.9)	10.0 (2.5)	4.4 (1.2)	5.0 (1.2)
SAT-MCII	10.1 (2.6)	9.4 (1.9)	4.2 (1.3)	4.7 (1.2)
BM	13.9 (2.4)	13.2 (2.3)	4.0 (1.2)	3.9 (1.2)

3.6 患者群と健常対照者群の比較

患者群では、ほとんどの検査において健常対照者群よりも有意に低い結果であった（d = 0.42-1.03）。一方で IBT、BM の評価値は 2 群間で有意な差を認めなかった（表 8）。

表 8：患者群と健常対照者群の比較

	患者群		健常対照者群				
検査	N	平均（SD）	N	平均（SD）	t	p	Cohen's d
BLERT	121	16.4（3.3）	70	18.4（1.9）	5.23	<0.001	0.79
FEST	121	14.7（2.7）	70	16.2（2.1）	4.32	<0.001	0.65
Hinting	121	12.6（3.8）	69	15.2（2.4）	5.69	<0.001	0.86
MSST	121	3.9（1.5）	70	4.6（1.0）	4.01	<0.001	0.60
AIHQ	121	6.9（2.0）	70	6.2（1.6）	-2.78	0.006	-0.42
IBT	119	0.5（0.2）	70	0.5（0.2）	-0.24	0.810	-0.04
SAT-MC	121	13.5（4.8）	70	16.3（3.2）	4.71	<0.001	0.70
SAT-MCII	121	10.5（3.8）	70	13.7（2.8）	6.85	<0.001	1.03
BM	121	20.9（17.0）	69	24.0（16.0）	1.26	0.209	0.19

3.7 各社会認知機能検査の特徴のまとめ

ここまでの結果を表9にまとめた。

表 9：各検査の結果の概要

検査	感度	再検査信頼性	学習効果 (-)	天井/床効果 (-)	社会機能の予測性	社会機能の増分妥当性	快適性	忍容性
BLERT	○	○	○	○			○	○
FEST	○	○	○	○	○ (SLOF)		○	○
Hinting	○	○	○	○	○ (UPSA)	○ (UPSA)	○	○
MSST	○	○	○		○ (UPSA)		○	○
AIHQ	○	○	○	○			○	○
IBT			○	○			○	○
SAT-MC	○	○	○		○ (UPSA)		○	○
SAT-MC II	○	○		○	○ (UPSA)		○	○
BM			○		○ (UPSA)		○	○

4. 考察

4.1 各社会認知機能検査の主な結果とその解釈

Hinting は十分な再検査信頼性や社会機能との強い関連を示し、神経認知機能の影響をコントロールした階層的重回帰分析においても社会機能を有意に予測した。さらに、日本の臨床現場でも許容される程度の短時間で実施可能であり、被験者の忍容性も良好で、患者群と健常対照者群で有意な差を認めた。過去には Hinting の天井効果が報告されているが (Davidson *et al*, 2018)、本研究で使用した日本語版において、問題となる天井効果は示されなかった。SCOPE Asia では Hinting は低い推奨度であり、文化差の影響を受ける可能性が示唆されていたが、本研究の結果からは文化差の影響よりも厳格な採点基準 (Klein *et al*, 2020)を用いたことなど方法論の違いがより影響したのかもしれない。SCOPE 研究と同様に Hinting の日本語版においても再検査信頼性および社会機能に対する収束的妥当性が検証されたことで、今後日本の臨床現場での利用のみならず、国際治験でも統一的な評価方法として期待される。

FEST は十分な再検査信頼性を示し、社会認知機能検査の中で唯一、社会機能における実世界の機能障害を示す SLOF との関連を示した。増分妥当性は明らかではなかったが、明らかな学習効果もなく、最も短い時間で実施可能であり、患者群と健常対照者群で有意な差を認めていることから、臨床現場での利用可能性は高いものと思われる。FEST は SCOPE 研究や SCOPE Asia では実施されていない検査であり、先行研究では同様の顔表情を用いた情動認知の検査として ER-40 が用いられ、高く評価されている。これら 2 つの検査は課題の基本構造が同様であることから、先行研究における ER-40 の結果を本研究では FEST が引き継ぎ再現されたと考える。なお本研究の前段階のパネル会議においては、本研究で実施する候補検査として ER-40 も候補に挙がったが、実施にあたり原著者から使用の許可が得られなかったため、改めてパネル会議の許可を得て実施を見送った経緯がある。結果として我々は、顔表情を用いた検査として FEST のみを実施することとなったが、日本人の被験者に対して日本人の顔に対する情動認知を測定することで、根本的に文化差を排除できた可能性がある (Gendron *et al*, 2014)。今後 ER-40 のような多様な人種の顔表情を用いた課題と FEST の結果を比較することで、情動認知に対する文化差の影響がより明確になるかもしれない。

一方で同じく情動認知を評価する BLERT は本研究では社会機能との有意な関連が示されず、その妥当性が不十分であった。我々は BLERT についても FEST と同様に、文化差や人種差の影響を弱めるために日本人の動画を使用したが、先行研究のような高い評価が得られなかった。FEST のような静的な顔写真から情

動を認知する検査と、BLERT のような感情の手がかりの多い動画から情動を判別する検査では、コミュニケーションにおける言語的および文化的な違いの影響があることが考えられる。すなわち、FEST よりも BLERT の方が言語的なシナリオや口調、身振り手振りといったジェスチャーに影響され、情動認知のプロセスが複雑となったことで、被験者にとって検査の難易度が高くなり、かえって実世界での社会機能との関連を乏しくさせたものと推測される (Frank, 2010)。しかしながら、患者群と健常対照者群との間に有意な差を認めていることから、上記特徴を踏まえつつ、社会認知機能障害を同定する検査として今後の検討が望まれる。

MSST は元々日本で自閉スペクトラム症（ASD）の小児のために開発されたものであり、今回の研究で初めて統合失調症患者に使用された。本研究において MSST は忍容性および快適性が高いこと、増分妥当性は示せなかったものの社会機能に対する有意な関連を示したこと、患者群と健常対照群とで成績に有意差を認めたことなど良好な結果を残した一方で、被験者の 6 割以上が天井効果を示し、その難易度の低さが課題となった。これは MSST が小学校の教科書の文章を題材に作成されており、成人の被験者には簡単すぎたためと考えられる (Adachi *et al.*, 2004)。一方で床効果を示す被験者も一定数存在しており、結果が二極化していた点が興味深い。本研究では M.I.N.I.によって併存疾患を確認しているが、ASD を合併した患者は含まれていない。成人期の ASD は他の精神疾患の合併が多いうえ(Mazzone *et al*, 2012)、発達早期からの症状の存在を確認するのがしばしば困難であることから、今回の方法では背景に存在する ASD の併存を正確に評価できなかった可能性が考えられる。今後 ASD の併存の有無も厳密に区別したうえで、統合失調症における心の理論の障害に関するさらなる検証が望まれる。

AIHQ は社会機能の予測においては有意ではなかったものの、健常対照者群と患者群で有意な差を認め、再検査信頼性も高かった。SCOPE 研究では AIHQ は「推奨しない」と分類されたが、SCOPE Asia は比較的良好な結果を示している。本研究では Buck らが報告したように(Buck *et al.*, 2017)、自由形式の質問を削除する形式で実施したことで、SCOPE Asia に続き、良好な結果を示したものと考えるが、社会認知機能における文化差をみている可能性も考慮が必要かもしれない。

一方で IBT は実施にあたっては 5 分未満と簡便に実施できたものの、患者群と健常対照者群で有意な差を認めず、社会機能との関連も認めなかった。検査を日本語に翻訳する際に回答の制限時間の設定に工夫を要したが(Trauzettel-Klosinski *et al.*, 2012)、それでもなお制限時間が短く、被験者の社会認知機能を正確に反映できなった可能性が考えられる。SCOPE 研究では高い推奨度を得た検

査であったが、本研究では良好な結果は得られず、検査における言語差の影響を大きく受けた結果と考えられる。

今回 SAT-MC と SAT-MCIIはそれぞれ独立した検査として実施し、どちらも社会機能との関連を認めた。SCOPE 研究では社会知覚領域の検査で推奨された検査がなかったが、今回言語的な影響が少ないこれらの検査で良好な結果が得られたことは有意義であった。このことの要因の一つとして、本研究では以前より難易度の差を指摘されていた SAT-MC の 2 つの版をそれぞれ独立した検査として評価したことで再検査信頼性が向上し、結果に影響した可能性がある。ただし SAT-MC は天井効果が高く、SAT-MCIIは高い学習効果を認めたことには留意が必要である。

BM は忍容性や快適性は許容範囲内であり、UPSA-B との有意な関連が得られたが、患者群のみならず健常対照者群においても床効果を認め、患者群と健常対照者群の間に有意差を認めなかった。SCOPE 研究において評価値として弁別閾値を用いることの有用性について考察されており、今回その閾値を指標と用いたことが社会機能との関連につながった可能性が考えられる。一方で、生物学的な動きを評価するため、完全に非言語的な検査であり、文化差や言語差の影響を受けないことが期待されたが、本研究における BM はその難易度の高さから患者群における社会認知機能障害をうまく検出できなかったのかもしれない。今後検査の難易度を調整することでより良好な計量心理学的な結果が得られることが期待される。

4.2 本研究の問題点およびその限界

本研究にはいくつかの限界がある。第一に、患者群と健常対照者群では一部の背景情報に差があり、本人の教育歴や知的能力の違いが結果に影響を与えた可能性がある。しかし、両親の学歴に差がないことから、健常対照者群と患者群の病前の知的能力の差はなかったと考えられる。また、社会認知機能検査の妥当性評価にあたっては統計学的に神経認知機能検査の結果を考慮しているため、最終的な検査の評価にあたっては群間の学歴や知的能力の差の影響は取り除かれていると考えられる。

第二に、患者群のほとんどが安定した外来患者であるため、選択バイアスの可能性がある。したがって、今回の結果は日本のすべての統合失調症をもつ患者に一般化できるものではないかもしれない。今後、精神症状がより重篤な患者を対象とすることで、各検査の特徴がより明確になる可能性がある。

第三に本研究では社会認知機能に対する薬剤の影響については考慮しなかった。向精神薬の中には神経認知機能に影響を与えるものがあることが示されており(Ogino *et al*, 2011)、今回の患者群における結果に影響を与えた可能性がある。

今後、本研究で信頼性妥当性が検証された社会認知機能検査を用いて、薬剤の社会認知機能に対する影響について検討する研究が待たれる。

5. 総括および結論

本研究全体から得られた新知見

本研究において我々は、以下の新たな知見を得た。

・神経認知機能の影響を考慮した上で、日本語版の Hinting の社会機能に対する増分妥当性が認められた。

・BLERT, FEST, Hinting, MSST, AIHQ, SAT-MC, SAT-MCIIについて再検査信頼性が確認された。

・SLOF に対して FEST の、UPSA-B に対して Hinting, MSST, SAT-MC, SAT-MCII, BM の収束的妥当性が確認された。

・本研究で新たに日本語版を作成した IBT や SAT-MC、SAT-MCIIをはじめ、日本語版の各検査の特徴が明らかとなった。

新知見の意義

社会認知機能を 4 つの中核領域により評価するという枠組みにおいて、各領域に対応する社会認知機能検査の社会機能との関連および信頼性妥当性が評価された。各検査の日本語版の信頼性妥当性が評価されたことで、今後の国際臨床試験で使用できる統一的な評価法を提供した。

また天井効果や床効果、学習効果、実施に要する時間などもあわせて評価しており、各検査の特徴が明らかとなったことで、今後日本の臨床現場で利用する際の参考になるものと思われる。本研究の結果をもとに、日本で初めての社会認知機能検査の推奨水準の提案につながっており、今後、各検査の特徴を踏まえながら日本の臨床現場で活用され、社会認知機能障害に関する知見がさらに蓄積されることが期待される。

Hinting の良好な計量心理学的特性が確認された一方で、先行研究に比し、BLERT と IBT の計量心理学的特性は劣る結果であった。BLERT は文化的差異によるコミュニケーション方法の違いが、IBT は言語の読字時間という要素が結果に影響した可能性がある。

本研究から今後どのような研究が展開されうるか

まず本研究で信頼性妥当性が評価された検査を用い、統合失調症にとどまらず疾患横断的に社会認知機能障害の有無や重症度を評価することが検討される。また社会復帰に向けたリハビリテーションの介入効果の判定として、社会認知機能検査を実施し評価することが考えられる。また認知機能障害を標的とした治療薬の国際治験での活用も予想される。

今後の課題

社会認知機能に影響を与える薬剤や陽性症状や陰性症状など他の症状との関連性について検討が必要である。今後、社会認知機能を総合的に評価するために、どのような組み合わせで 1 つのバッテリー検査とするのが最も適切かについて検討が望まれる。

謝辞

本研究の計画、実施、また論文作成や学会発表にあたりご指導、ご助言をいただいた北海道大学大学院医学研究院精神医学教室　橋本直樹准教授に深く感謝をいたします。同様に研究の立案、実施、論文作成にあたりご指導をいただいた国立病院機構帯広病院　大久保亮先生に厚く御礼を申し上げます。研究全体の統計解析およびその解釈につきましてご指導をいただいた北海道大学病院データサイエンスセンター 伊藤陽一教授に感謝をいたします。各施設における本研究の遂行にご尽力いただいた共同研究者の豊巻敦人先生、宮崎茜先生、服部幸子先生、野原万梨子先生、佐々木洋平先生、久保田涼太郎先生、岡野宏紀先生、高橋奏海先生、長谷川由美先生、和田泉先生、内野敬先生、竹田和良先生、池澤聰先生、根本隆洋先生に感謝申し上げます。また、本研究のデータ収集や論文作成にあたり、貴重なご助言やご協力を下さった北海道大学大学院医学研究院精神医学教室　久住一郎教授をはじめとする精神医学教室のスタッフの皆様に改めて感謝をいたします。最後に、本研究に参加を下さった被験者の皆様とそのご家族、介護者の皆様に最大限の感謝を表します。

利益相反

本研究に関し、開示すべき利益相反はありません。

引用文献

Adachi T, Koeda T, Hirabayashi S, Maeoka Y, Shiota M, Wright EC and Wada A (2004) The metaphor and sarcasm scenario test: a new instrument to help differentiate high functioning pervasive developmental disorder from attention deficit/hyperactivity disorder. *Brain Dev* 26, 301-306

Addington J, Penn D, Woods SW, Addington D and Perkins DO (2008) Facial affect recognition in individuals at clinical high risk for psychosis. *Br J Psychiatry* 192, 67-68

Bell MD, Fiszdon JM, Greig TC and Wexler BE (2010) Social attribution test--multiple choice (SAT-MC) in schizophrenia: comparison with community sample and relationship to neurocognitive, social cognitive and symptom measures. *Schizophr Res* 122, 164-171

Brekke J, Kay DD, Lee KS and Green MF (2005) Biosocial pathways to functional outcome in schizophrenia. *Schizophr Res* 80, 213-225

Bryson G, Bell M and Lysaker P (1997) Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit. *Psychiatry Res* 71, 105-113

Buck B, Iwanski C, Healey KM, Green MF, Horan WP, Kern RS, Lee J, Marder SR, Reise SP and Penn DL (2017) Improving measurement of attributional style in schizophrenia; A psychometric evaluation of the Ambiguous Intentions Hostility Questionnaire (AIHQ). *J Psychiatr Res* 89, 48-54

Chen KW, Lee SC, Chiang HY, Syu YC, Yu XX and Hsieh CL (2017) Psychometric properties of three measures assessing advanced theory of mind: Evidence from people with schizophrenia. *Psychiatry Res* 257, 490-496

Combs DR, Penn DL, Wicher M and Waldheter E (2007) The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cogn Neuropsychiatry* 12, 128-143

Corcoran R, Mercer G and Frith CD (1995) Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophr Res* 17, 5-13

Davidson CA, Lesser R, Parente LT and Fiszdon JM (2018) Psychometrics of social cognitive measures for psychosis treatment research. *Schizophr Res* 193, 51-57

Frank CK (2010) Linguistic effects on the neural basis of theory of mind. *Open Neuroimag J* 4, 37-45

G. B. D. Disease Injury, Incidence Prevalence, Collaborators (2017) Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390, 1211-1259

Gendron M, Roberson D, van der Vyver JM and Barrett LF (2014) Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion* 14, 251-262

Green MF, Kern RS, Braff DL and Mintz J (2000) Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? *Schizophr Bull* 26, 119-136

Green MF, Penn DL, Bentall R, Carpenter WT, Gaebel W, Gur RC, Kring AM, Park S, Silverstein SM and Heinssen R (2008) Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. *Schizophr Bull* 34, 1211-1220

Grossman ED, Blake R and Kim CY (2004) Learning to see biological motion: brain activity parallels behavior. *J Cogn Neurosci* 16, 1669-1679

Hagiya K, Sumiyoshi T, Kanie A, Pu S, Kaneko K, Mogami T, Oshima S, Niwa S, Inagaki A, Ikebuchi E *et al* (2015) Facial expression perception correlates with verbal working memory function in schizophrenia. *Psychiatry Clin Neurosci* 69, 773-781

Hajduk M, Achim AM, Brunet-Gouet E, Mehta UM and Pinkham AE (2020) How to move forward in social cognition research? Put it into an international perspective. *Schizophr Res* 215, 463-464

Hashimoto N, Takahashi K, Fujisawa D, Aoyama K, Nakagawa A, Okamura N, Toyomaki A, Oka M, Takanobu K, Okubo R *et al* (2020) A pilot validation study of the Japanese translation of the Positive and Negative Syndrome Scale (PANSS). *Asian J Psychiatr* 54, 102210

Hashimoto N, Toyomaki A, Hirai M, Miyamoto T, Narita H, Okubo R and Kusumi I (2014) Absent activation in medial prefrontal cortex and temporoparietal junction but not superior temporal sulcus during the perception of biological motion in schizophrenia: a functional MRI study. *Neuropsychiatr Dis Treat* 10, 2221-2230

Jaaskelainen E, Juola P, Hirvonen N, McGrath JJ, Saha S, Isohanni M, Veijola J and Miettunen J (2013) A systematic review and meta-analysis of recovery in schizophrenia. *Schizophr Bull* 39, 1296-1306

Jauhar S, Johnstone M and McKenna PJ (2022) Schizophrenia. *Lancet* 399, 473-486

Johannesen JK, Lurie JB, Fiszdon JM and Bell MD (2013) The Social Attribution Task-Multiple Choice (SAT-MC): A Psychometric and Equivalence Study of an Alternate Form. *ISRN Psychiatry* 2013, 830825

Kaneda Y, Sumiyoshi T, Keefe R, Ishimoto Y, Numata S and Ohmori T (2007) Brief assessment of cognition in schizophrenia: validation of the Japanese version. *Psychiatry Clin Neurosci* 61, 602-609

Kanie A, Hagiya K, Ashida S, Pu S, Kaneko K, Mogami T, Oshima S, Motoya M, Niwa S, Inagaki A *et al* (2014) New instrument for measuring multiple domains of social cognition: construct validity of the Social Cognition Screening Questionnaire (Japanese version). *Psychiatry Clin Neurosci* 68, 701-711

Kay SR, Fiszbein A and Opler LA (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13, 261-276

Keefe RS, Goldberg TE, Harvey PD, Gold JM, Poe MP and Coughenour L (2004) The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res* 68, 283-297

Kerr SL and Neale JM (1993) Emotion perception in schizophrenia: specific deficit or further evidence of generalized poor performance? *J Abnorm Psychol* 102, 312-318

Klein HS, Springfield CR, Bass E, Ludwig K, Penn DL, Harvey PD and Pinkham AE (2020) Measuring mentalizing: A comparison of scoring methods for the hinting task. *Int J Methods Psychiatr Res* 29, e1827

Koo TK and Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15, 155-163

Kubota R, Okubo R, Akiyama H, Okano H, Ikezawa S, Miyazaki A, Toyomaki A, Sasaki Y, Yamada Y, Uchino T *et al* (2021) Study Protocol: The Evaluation Study for Social Cognition Measures in Japan (ESCoM). *J Pers Med* 11

Kurtz MM, Moberg PJ, Gur RC and Gur RE (2001) Approaches to cognitive remediation of neuropsychological deficits in schizophrenia: a review and meta-analysis. *Neuropsychol Rev* 11, 197-210

Lee HS, Corbera S, Poltorak A, Park K, Assaf M, Bell MD, Wexler BE, Cho YI, Jung S, Brocke S *et al* (2018) Measuring theory of mind in schizophrenia research: Cross-cultural validation. *Schizophr Res* 201, 187-195

Lim K, Lee SA, Pinkham AE, Lam M and Lee J (2020) Evaluation of social cognitive measures in an Asian schizophrenia sample. *Schizophr Res Cogn* 20, 100169

Matsuoka K, Uno M, Kasai K, Koyama K and Kim Y (2006) Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test. *Psychiatry Clin Neurosci* 60, 332-339

Mausbach BT, Harvey PD, Goldman SR, Jeste DV and Patterson TL (2007) Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophr Bull* 33, 1364-1372

Mazzone L, Ruta L and Reale L (2012) Psychiatric comorbidities in asperger syndrome and high functioning autism: diagnostic challenges. *Ann Gen Psychiatry* 11, 16

McGrath J, Saha S, Chant D and Welham J (2008) Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev* 30, 67-76

Medalia A and Choi J (2009) Cognitive remediation in schizophrenia. *Neuropsychol Rev* 19, 353-364

Mehta UM, Thirthalli J, Naveen Kumar C, Mahadevaiah M, Rao K, Subbakrishna DK, Gangadhar BN and Keshavan MS (2011) Validation of Social Cognition Rating Tools

in Indian Setting (SOCRATIS): A new test-battery to assess social cognition. *Asian J Psychiatr* 4, 203-209

Nuechterlein KH, Subotnik KL, Green MF, Ventura J, Asarnow RF, Gitlin MJ, Yee CM, Gretchen-Doorly D and Mintz J (2011) Neurocognitive predictors of work outcome in recent-onset schizophrenia. *Schizophr Bull* 37 Suppl 2, S33-40

Ogino S, Miyamoto S, Tenjin T, Kitajima R, Ojima K, Miyake N, Funamoto Y, Arai J, Tsukahara S, Ito Y *et al* (2011) Effects of discontinuation of long-term biperiden use on cognitive function and quality of life in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 35, 78-83

Okano H, Kubota R, Okubo R, Hashimoto N, Ikezawa S, Toyomaki A, Miyazaki A, Sasaki Y, Yamada Y, Nemoto T *et al* (2021) Evaluation of Social Cognition Measures for Japanese Patients with Schizophrenia Using an Expert Panel and Modified Delphi Method. *Journal of Personalized Medicine* 11

Otsubo T, Tanaka K, Koda R, Shinoda J, Sano N, Tanaka S, Aoyama H, Mimura M and Kamijima K (2005) Reliability and validity of Japanese version of the Mini-International Neuropsychiatric Interview. *Psychiatry Clin Neurosci* 59, 517-526

Pinkham AE, Harvey PD and Penn DL (2018) Social Cognition Psychometric Evaluation: Results of the Final Validation Study. *Schizophr Bull* 44, 737-748

Pinkham AE and Penn DL (2006) Neurocognitive and social cognitive predictors of interpersonal skill in schizophrenia. *Psychiatry Res* 143, 167-178

Pinkham AE, Penn DL, Green MF, Buck B, Healey K and Harvey PD (2014) The social cognition psychometric evaluation study: results of the expert survey and RAND panel. *Schizophr Bull* 40, 813-823

Pinkham AE, Penn DL, Green MF and Harvey PD (2016) Social Cognition Psychometric Evaluation: Results of the Initial Psychometric Study. *Schizophr Bull* 42, 494-504

Pinkham AE, Penn DL, Perkins DO, Graham KA and Siegel M (2007) Emotion perception and social skill over the course of psychosis: a comparison of individuals "at-risk" for psychosis and individuals with early and chronic schizophrenia spectrum illness. *Cogn Neuropsychiatry* 12, 198-212

Rosset E (2008) It's no accident: Our bias for intentional explanations. *Cognition* 108, 771-780

Sabbag S, Twamley EM, Vella L, Heaton RK, Patterson TL and Harvey PD (2011) Assessing everyday functioning in schizophrenia: Not all informants seem equally informative. *Schizophrenia Research* 131, 250-255

Schneider LC and Struening EL (1983) SLOF: a behavioral rating scale for assessing the mentally ill. *Soc Work Res Abstr* 19, 9-21

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R and Dunbar GC (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 59 Suppl 20, 22-33;quiz 34-57

Sumiyoshi C, Takaki M, Okahisa Y, Patterson TL, Harvey PD and Sumiyoshi T (2014) Utility of the UCSD Performance-based Skills Assessment-Brief Japanese version: discriminative ability and relation to neurocognition. *Schizophr Res Cogn* 1, 137-143

Sumiyoshi T, Nishida K, Niimura H, Toyomaki A, Morimoto T, Tani M, Inada K, Ninomiya T, Hori H, Manabe J *et al* (2016) Cognitive insight and functional outcome in schizophrenia; a multi-center collaborative study with the specific level of functioning scale-Japanese version. *Schizophr Res Cogn* 6, 9-14

Trauzettel-Klosinski S, Dietz K and Group IRS (2012) Standardized assessment of reading performance: the New International Reading Speed Texts IReST. *Invest Ophthalmol Vis Sci* 53, 5452-5461

Uchino T, Okubo R, Takubo Y, Aoki A, Wada I, Hashimoto N, Ikezawa S and Nemoto T (2022) Perceptions of and subjective difficulties with social cognition in schizophrenia from an internet survey: Knowledge, clinical experiences, and awareness of association with social functioning. *Psychiatry Clin Neurosci* 76, 429-436

Wykes T, Huddy V, Cellard C, McGurk SR and Czobor P (2011) A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *Am J Psychiatry* 168, 472-485