

保護観察におけるアセスメントツールの評定者間信頼性の検証

羽間京子^{1)*}・勝田 聡²⁾³⁾¹⁾千葉大学教育学部²⁾法務省横浜保護観察所³⁾千葉大学大学院人文科学研究院

Inter-Rater Reliability Examination of Assessment Tools for Individuals Under Supervision in Japan

HAZAMA Kyoko^{1)*} and KATSUTA Satoshi²⁾³⁾¹⁾Faculty of Education, Chiba University, Japan²⁾Yokohama Probation Office, Ministry of Justice, Japan³⁾Graduate School of Humanities, Chiba University, Japan

国際的に、犯罪をした人や非行のある少年のアセスメントと処遇に有用とされるRisk-Need-Responsivity (RNR) モデルを踏まえ、法務省は保護観察の新たなアセスメントツール (Case Formulation in Probation/Parole : CFP) を開発し、実施している。CFPは、RNRモデルが同定した7つないし6つの動的犯罪誘発性要因にそれぞれ対応する、動的犯罪誘発性要因7項目と動的保護・改善更生促進要因6項目を含む。本研究はCFPの信頼性確認のため、これらの動的要因の評定者間信頼性を検証した。保護観察官等の専門家3名による25事例の動的要因の評定について、Lightの κ 係数を算出した。 κ 係数は、動的犯罪誘発性要因全体が0.90、動的保護・改善更生促進要因全体が0.91、13項目中12項目で0.77から1.00と、適度から優れた評定者間信頼性を示し、CFPに信頼性が認められた。他方、「向犯罪的態度」の保護・改善更生促進要因の κ 係数は0.66と低く、より評定者の主観的判断が評定に伴うことの影響が考えられた。今後の研究課題が論じられた。

The Ministry of Justice in Japan developed and has used new assessment tools for individuals under supervision, referred to as *Case Formulation in Probation/Parole* (CFP), based on the Risk-Need-Responsivity (RNR) model, which globally has become the primary model in assessment and intervention for justice-involved individuals. This study examined the inter-rater reliability of the CFP's dynamic factors, which included seven dynamic risk factors and six dynamic protective and promotive factors corresponding respectively to the seven or six major dynamic risk/need factors identified in the RNR model. Light's kappa was calculated for the dynamic factors of 25 individuals under supervision to assess the inter-rater reliability of three raters who were or had been probation officers. The results showed that Light's kappa was 0.90 for all of the dynamic risk factors, 0.91 for all of the dynamic protective and promotive factors, and ranged from 0.77 to 1.00 for 12 of the 13 dynamic factors, indicating adequate to excellent inter-rater reliability. Therefore, this study demonstrated the reliability of the CFP. On the other hand, Light's kappa was 0.66 for a protective and promotive factor that was contrary to "Procriminal Attitudes." However, this dynamic factor may be influenced by the raters' judgment or may be somewhat more subjective than the other dynamic factors, which might have led to the low inter-rater reliability. Suggestions for future research were discussed.

キーワード：保護観察 (probation and parole), Risk-Need-Responsivityモデル (Risk-Need-Responsivity model), 評定者間信頼性 (inter-rater reliability), 動的犯罪誘発性要因 (dynamic risk factors), 動的保護・改善更生促進要因 (dynamic protective and promotive factors)

1. 問題と目的

1.1 刑事司法領域におけるリスクアセスメントの発展

犯罪をした人や非行のある少年に対し、より効果的な処遇を実施するためには、再犯リスクを含めた的確なアセスメントが極めて重要である。刑事司法領域でのリスクアセスメントは、国際的に次の三つの世代を経て発展

してきたとされる (Andrews et al., 2006; Bonta, 1996; Bonta & Andrews, 2017)。

まず、第一世代のリスクアセスメントは、再犯の可能性に関する、構造化されていない専門家の臨床判断を中心とするものであり、再犯予測力はあまり高くなかった。次に、第二世代のリスクアセスメントは、エビデンスに基づく保険統計的リスクアセスメントツールによる。ただし、これらのツールは、再犯予測力は第一世代のものより高いが、処遇によって変化しない、年齢等の静的リ

*連絡先著者：羽間京子 hazama@faculty.chiba-u.jp

スクを中心とするものであり、処遇に関する示唆は与えられなかった。

第三世代のリスクアセスメントは、理論とエビデンスに基づき、処遇によって変わりうる動的な犯罪誘発性要因（以下、動的犯罪誘発性要因）を中心に測定する。代表的なアセスメントツールとして、たとえば Level of Service Inventory-Revised（処遇レベル目録改訂版；Andrews & Bonta, 1995）がある。これは、犯罪をした人や非行のある少年の再犯や再非行を抑制するために有用とされているRisk-Need-Responsivityモデル（リスク・ニード・リスポンシビティモデル；以下、RNRモデル）

（Bonta & Andrews, 2017）を基盤としたものである。RNRモデルは、(a)再犯リスクの高さの程度に応じた密度により（リスク原則）、(b)犯罪を誘発する要因に焦点を当て（ニード原則）、(c)効果が実証された方法や対象者の特性に応じた方法で（リスポンシビティ原則）、処遇を行うことを中核原則とし、北米を中心に主流となっている。

さらに、21世紀からは第四世代のリスクアセスメントに至っており、そこではアセスメントとケースマネジメントの統合が強調されている（Andrews et al., 2006; Bonta & Andrews, 2017）。

1.2 リスクアセスメントツールの信頼性検証について

第三世代以降のリスクアセスメントは、動的犯罪誘発性要因の分析を踏まえて処遇の焦点を絞り、ケースプランを作成することができるため、それ以前のアセスメントツールに比べ、処遇実施上、より有益である。他方、課題として、動的犯罪誘発性要因の多くは定義が明確ではなく（Schmidt et al., 2020）、その分析と把握に評定者の主観的判断が伴う（Lowenkamp et al., 2004; Schmidt et al., 2020）ことが挙げられている。ただし、Lowenkamp et al. (2004) は、第三世代のアセスメントツールは分類やアセスメントの標準化が可能であり、適切に実施されれば、評定におけるバイアスを減少させ、ケースプラン作成に有益な情報を提供できるとした。

動的犯罪誘発性要因を的確に把握し、詳細で包括的なリスクアセスメントを行うためには、まず、担当者に、アセスメントの過程や、アセスメントツールに含まれている項目に関する専門的な知識が求められる（Lowenkamp et al., 2004）。また、アセスメントツールの妥当性と信頼性の検証も重要である。特に、信頼性はリスクアセスメントツールの最も基本的な要件であり（Douglas & Skeem, 2005）、アセスメントツールが妥当であるためには、それが信頼性のあるものでなければならない（Desmarais & Singh, 2013）。第三世代以降のアセスメントツールは動的犯罪誘発性要因の査定を含み、評定者の判断が伴うため、特に、実施された評定の信頼性に関する検証は欠かせない。

ただし、第三世代以降のアセスメントツールについては、妥当性、つまり再犯予測力に関する研究が多くなされているものの（Andrews et al., 2006; Bonta & Andrews, 2017）、信頼性の検証を行った研究は限られている（Desmarais & Singh, 2013; Duwe & Rocque, 2017; Lowenkamp et al., 2004; Rocque & Plummer-Beale, 2014; Schmidt et al.,

2020）。また、信頼性を検討したとする研究においても、信頼性係数である a 係数を計測している場合が少なくない（Baird, 2009）。しかしながら、刑事司法領域におけるアセスメントツールに関して必要とされるのは内的一貫性（信頼性）ではなく、評定者間信頼性の検証である（Baird, 2009; Duwe, 2017; Lowenkamp et al., 2004; Schmidt et al., 2020）。

加えて、Duwe & Rocque (2017) は、評定者間信頼性が低くなればなるほど、アセスメントツールの再犯予測力が低下するとし、手動で評定を行うリスクアセスメントツールでは、ICC値（級内相関係数）等の評定者間一致度の評価として、0.95以上を優れた一致、0.85-0.94を良好な一致、0.75-0.84を適度の一致、0.74以下を乏しい一致とする基準が妥当であるとした。

日本の保護観察においては、従前、再犯のリスク要因のうち、主に静的要因を査定するツール（以下、従前ツール）が使用されてきた（羽間・勝田, 2021, 2022; 勝田・羽間, 2020）。法務省保護局は、RNRモデルを基盤として、保護観察の新たなアセスメントツールであるCase Formulation in Probation/Parole（保護観察におけるケースフォーミュレーション；以下、CFP）を開発し、2018年10月からの二段階の試行を経て、2021年1月から施行した。CFPは、(a)静的要因を中心とした保険統計的リスク分析、(b)動的犯罪誘発性要因と動的保護・改善更生促進要因の把握、及び(c)犯罪・非行プロセスの分析からなる（勝田・羽間, 2020）。このうち、保険統計的リスク分析と、動的犯罪誘発性要因及び動的保護・改善更生促進要因は、再犯予測に使用することができる。

2018年10月から2019年3月のCFPの第一次試行では、全国の保護観察所において試行担当官に指名された102人の保護観察官の下で保護観察が開始された655人（家庭裁判所で保護観察に付された少年、少年院仮退院者、刑事施設仮釈放者、保護観察付全部執行猶予者と一部猶予者）が、従前のツールによるアセスメント実施群（ $n=266$ ；以下、従前ツール実施群）とCFPによるアセスメント実施群（ $n=389$ ；以下、CFP実施群）にランダムに割り付けられた^(註1)。羽間・勝田 (2021, 2022) は、両群の保護観察開始後の再犯データを分析し、従前ツールに比し、CFPに一定の再犯予測力が認められることを明らかにした。また、羽間・勝田 (2022) は、保護観察開始後2年間の再犯データの分析を通して、CFPのより高い動的犯罪誘発性要因得点とより低い動的保護・改善更生促進要因得点が、再犯を有意に予測することを明らかにした。つまり、CFPには一定の妥当性があると言えることができる。しかし、CFPの信頼性、つまり評定者間信頼性は検証されていない。

なお、RNRモデルを基盤に日本で開発されたアセスメントツールとしては、少年鑑別所において実施されている法務省式ケースアセスメントツール（Ministry of Justice Case Assessment Tool : MJCA）がある。森他 (2014) は、MJCAの再非行予測力の検証とともに、静的リスク項目と動的リスク項目の評定者間信頼性を検証するためICC値を算出し、静的領域全体についてICC値が0.92、動的領域全体に関し0.69との結果を得て、MJCAは信頼性のあるツールだとした。しかし、これら

の結果のうち、とりわけ動的領域全体のICC値は、Duwe & Rocque (2017) の基準に照らすと評価が分かれるところであり、さらに、森他 (2014) では動的領域の各因子のICC値は明らかにされていない。

1.3 本研究の目的

以上を踏まえ、本研究は、CFPの項目のうち、評定者の判断が特に求められる動的犯罪誘発性要因と動的保護・改善更生促進要因のアセスメントの評定者間信頼性を検証し、今後の課題について論じることとした。

2. 方 法

本研究は、千葉大学教育学部生命倫理審査委員会の承認を受けて実施された。

2.1 分析対象者と評定者

法務省保護局から、羽間・勝田 (2021, 2022) のために、CFP第一次試行段階の従前ツール実施群とCFP実施群の計655人について、書面により、基本属性、アセスメントの資料や結果等の匿名化されたデータが提供された。前述の通り、CFP実施群は389人であった。

評定者間信頼性を検証するための κ 係数の算出に最小限必要なサンプルサイズは、Bujang & Baharum (2017) によると、少なくとも10以上であり、さらに、 2×2 のカテゴリで、検出力80%、 $\alpha < 0.05$ であって、予想感度 (κ 係数) が0.3から0.8以上、あるいは0.3から0.7以上なら、23から39とされる。本研究は、CFP実施群から、設定した研究期間 (3年) の中で複数の人による評定が実施可能な事例数として、31人を抽出した。第二著者が、年齢、性別、保護観察の種類、罪名・非行名によりランダムに抽出し、加えて、質量ともに十分な情報を有するものを選定した。

事例担当官以外の現職の保護観察官5人、保護観察官経験者である法務省職員4人及び犯罪・非行心理学の専門家1人の計10人が、CFPの評定に参加した。2019年7月から2022年6月まで、継続的に事例検討会を実施し、上述の31事例を検討した。異動等に伴い、出席者には入れ替わりがあり、1回の事例検討会の出席者は平均6.5人だった。最も多く事例検討に参加した3人 (以下、評定者) が評定した25人を分析対象者とした。評定者はいずれも、保護観察官として10年以上の実務経験を有していた。

分析対象者の基本的属性は、平均年齢27.4歳、男性18人、女性7人であった。保護観察の種類は、保護観察処分少年3人、少年院仮退院者8人、刑事施設仮釈放者10人、保護観察付執行猶予者4人であった。主たる事件の内容は、窃盗や詐欺等の財産犯が10人、暴行、傷害、殺人等の粗暴犯が5人、強制わいせつ等の性犯罪が4人、覚醒剤自己使用を内容とする薬物犯が2人、その他が4人であった。

2.2 CFPの項目

RNRモデルでは、セントラルエイトと呼ばれるリスク/ニード要因、具体的には(a)犯罪歴、(b)向犯罪的態度、(c)向犯罪的交友、(d)反社会的パーソナリティ・パターン、(e)家族・夫婦、(f)学校・仕事、(g)物質使用、(h)余暇・レ

クリエーションを重視する (Bonta & Andrews, 2017)。このうち、犯罪歴を除く7つが動的犯罪誘発性要因である。

CFPの動的犯罪誘発性要因は14項目からなり、そのうち11項目が、RNRモデルのセントラルエイトの7つの動的犯罪誘発性要因に対応する。CFPの動的保護・改善更生促進要因は同じく14項目から構成され、そのうち12項目が、RNRモデルの7つの動的犯罪誘発性要因から反社会的パーソナリティ・パターンを除いた6要因の反対 (または強み) に位置づけられる、ポジティブな内容に対応する (羽間・勝田, 2022)。

2.3 評定手続き

それぞれの評定者において事例のデータを読み込み、CFPの動的犯罪誘発性要因及び動的保護・改善更生促進要因を評定した。その後、各々の評定を共有し、結果が異なる項目について協議を行った。

2.4 分析方法

本研究では、羽間・勝田 (2022) と同様に、RNRモデルのセントラルエイトに対応するCFPの動的犯罪誘発性要因11項目と動的保護・改善更生促進要因12項目を、それぞれの対応ごとに7ないし6項目にまとめ、該当ありを1、なしを0として分析を行った。

動的犯罪誘発性要因と動的保護・改善更生促進要因に係る評定者間信頼性を検証するため、分析対象者の25事例について評定者2人ごとの組み合わせによる κ 係数を求めた上で、Light (1971) に従い、その平均値を算出した。 κ 係数による一致度の判断はDuwe & Rocque (2017) の基準に基づいた。

統計分析には、SPSS ver. 27を使用した。

3. 結 果

Table 1にて、CFPの動的犯罪誘発性要因及び動的保護・改善更生促進要因の全体及び各項目に対する評定者3人の評定の κ 係数を示した。なお、評定者2人の κ 係数は、全て0.1%水準で有意であった。

Table 1にあるように、 κ 係数は、動的犯罪誘発性要因全体が0.90、動的保護・改善更生促進要因全体が0.91、さらに各項目は向犯罪的態度の保護・改善更生促進要因以外の12項目で0.77から1.00であり、Duwe & Rocque (2017) の基準に照らすと適度から優れた一致の範囲にあった。

他方、向犯罪的態度の動的保護・改善更生促進要因の項目については、 κ 係数は0.66であり、乏しい一致となった。向犯罪的態度の動的保護・改善更生促進要因の項目は、CFPでは「現状への危機感」「行動変化の意欲」及び「将来への希望」から構成される。評定が一致しなかった事例を確認したところ、25事例のうち5事例に評定の不一致が認められ、さらにこれらの不一致は、全て「行動変化の意欲」についてであった。4事例は、「行動変化の意欲」を表現していたが、具体的計画が乏しかったものであり、具体的計画性があるかないかで評定が異なった。また、1事例は、同じく「行動変化の意欲」を

Table 1 動的二要因における評定者間信頼性 (κ 係数)

要因の種類	項目名	κ
犯罪誘発性要因	全体	0.90
	向犯罪的態度	0.93
	向犯罪的交友	0.86
	反社会的パーソナリティ・パターン	0.94
	家族・夫婦	0.84
	学校・仕事	0.83
	物質使用	1.00
	余暇・レクリエーション	1.00
保護・改善更生促進要因	全体	0.91
	向犯罪的態度	0.66
	向犯罪的交友	0.92
	家族・夫婦	1.00
	学校・仕事	0.91
	物質使用	—
	余暇・レクリエーション	0.77

注：保護・改善更生促進要因のうち物質使用の項目は、25件全てが該当なしで一致していたため、 κ 係数は算出されなかった。この項目を除き、保護・改善更生促進要因全体の κ 係数を算出した。

言葉で表わしていたが、事件を否認し、正当化するなど意欲を疑わせる発言もあったため、評定が割れたものであった。

4. 考 察

本研究は、RNRモデル (Bonta & Andrews, 2017) を踏まえて新たに開発された保護観察のアセスメントツールであるCFP (勝田・羽間, 2020) のうち、特に評定者の判断が求められる動的要因 (Lowenkamp et al., 2004; Schmidt et al., 2020) について、評定者間信頼性を検証した。CFP第一次試行のCFP実施群389人のうち25人に関して、3名の専門家による動的犯罪誘発性要因及び動的保護・改善更生促進要因の評定を分析した。CFPの動的犯罪誘発性要因11項目と動的保護・改善更生促進要因12項目は、対応するRNRモデルのセントラルエイトの7ないし6項目にまとめて分析された。Light (1971) の κ 係数を求めたところ、動的犯罪誘発性要因全体が0.90、動的保護・改善更生促進要因全体が0.91、13項目のうち向犯罪的態度の保護・改善更生促進要因以外の12項目において0.77から1.00であり、Duwe & Rocque (2017) の基準に照らすと適度から優れた一致の範囲にあった。他方、向犯罪的態度の動的保護・改善更生促進要因の項目の κ 係数は0.66であり、乏しい一致となった。向犯罪的態度の動的保護・改善更生促進要因の不一致があった5事例を検討したところ、同項目を構成する内容のうち、「行動変化の意欲」の評定において不一致が認められた。「行動変化の意欲」の判断は、他の項目等の判断に比べ、本人の言葉に頼りがちであって、つまり、評定者の主観的判断の余地が大きい、特に具体的計画があるかどうかを考慮に入れるかについて、CFPの評定のマニュアルでは曖昧である。そのため、「行動変化の意欲」の評定者間信頼性が低くなったものと考えられる。Duwe & Rocque (2017) は、評定者の主観に頼りがちな項目は、

できるだけ判断しやすい基準に変更したほうがよいとする (たとえば、刑事施設入所中の場合であれば、「孤独かどうか」と本人に聞くのではなく、矯正施設に面接者があったかをチェックするなど)。筆者らの実務感覚からすると、「行動変化の意欲」があるかどうかの判断に、具体的計画性や反省があることを含めることは重要だと思われるが、評定の方針を明らかにするためには、これらを判断に入れるかどうかの議論と、判断に入れた場合と入れない場合の再犯予測力の検証が必要である。また、何をもって具体的計画や反省があるとするかの議論も不可欠である。

以上のように、信頼性を向上させるために、向犯罪的態度の動的保護・改善更生促進要因の評定に検討・改善の余地はあるものの、動的犯罪誘発性要因全体と動的保護・改善更生促進要因全体の評定者間一致度は、Duwe & Rocque (2017) の基準に照らして良好だった。また、各項目の評定者間信頼性も、1項目を除き、適度から優れた一致の範囲にあった。つまり、本研究の結果から、CFPは信頼性をもつことが示されたと言える。再犯予測力に加え (羽間・勝田, 2021, 2022)、信頼性も認められたことから、保護観察のアセスメントツールとして、CFPは妥当性と信頼性のある尺度と位置付けられる。

ただし、本研究にはいくつかの限界がある。まず、分析対象とした事例数についてである。先述のように、本研究では、Bujang & Baharum (2017) と研究期間を踏まえ、31事例が抽出され、そのうち25事例が分析対象となったが、Lombard et al. (2002) はサンプル全体の10%以上が評定者間信頼性の検証には必要だとした。Lombard et al. (2002) によるならば、本研究で最小限求められるサンプルサイズは39となり、分析対象者の25人は、若干少ない可能性が指摘できる。

また、本研究はあくまでも、書面のデータによる評定であり、そこから得られる情報は、実際の保護観察処遇で得られる情報量よりも少ないという点も限界として挙

げられる。

このような限界はあるが、前述の通り、非行のある少年や犯罪をした人のリスクアセスメントツールの信頼性に関する研究が世界的にも少ない中で、本研究は、保護観察におけるケースアセスメントツールであるCFPの評定者間信頼性を検討した。そして、CFPに信頼性が認められることを示し、さらに一致度が低かった1要因についてその理由を検討し、評定者間信頼性を向上させるための研究の方向性を具体的に提言した点において、本研究の意義が認められる。CFPは第二次試行段階において、動的要因の一部に項目が付加された版が使用されており、サンプルサイズを増やし更なる検証を行うことが重要である。

注

- 1 二群のサンプル数は、回答率の差によって相異が生じたものである。

付 記

本研究の分担は次の通りである。羽間は1, 2.1, 2.4及び4を執筆し、勝田は2.2, 2.3及び3を執筆した。本研究は、JSPS科研費19K03277(研究代表者:羽間京子)の助成を受けた。本研究の実施を承認し、データを提供下さった法務省保護局長に感謝申し上げる。また、事例検討参加メンバーに深謝する。本論文の議論は筆者らのものであり、法務省の見解ではない。

引用文献

- Andrews, D. A., & Bonta, J. (1995). *The level of service inventory-revised*. Multi-Health System.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7-27. <https://doi.org/10.1177/0011128705281756>
- Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system*. National Council on Crime & Delinquency. https://www.evidentchange.org/sites/default/files/publication_pdf/special-report-evidence.pdf
- Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Sage.
- Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed). Routledge.
- Bujang, M. A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Cohen's Kappa. *Epidemiology Biostatistics and Public Health*, 14, e12267-1-e12267-10.
- Desmarais, S. L., & Singh, J. P. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States*. Council of State Governments. <https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf>
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, 11(3), 347-383. <https://doi.org/10.1037/1076-8971.11.3.347>
- Duwe, G. (2017). *Why inter-rater reliability matters for recidivism risk assessment (Policy Brief Number 2017-03)*. The Public Safety Risk Assessment Clearinghouse. <https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/pb-interrater-reliability.pdf>
- Duwe, G., & Rocque, M. (2017). The effects of automating recidivism risk assessment on reliability, predictive validity, and return on investment (ROI). *Criminology & Public Policy*, 16(1), 235-269. <https://doi.org/10.1111/1745-9133.12270>
- 羽間京子・勝田聡 (2021). 保護観察におけるアセスメントツールの再犯予測力の検証. 千葉大学教育学部研究紀要, 69, 27-32. <https://doi.org/10.20776/S13482084-69-P27>
- 羽間京子・勝田聡 (2022). 保護観察におけるアセスメントツールの動的要因の再犯予測力. 千葉大学教育学部研究紀要, 70, 7-12. <https://doi.org/10.20776/S13482084-70-P7>
- 勝田聡・羽間京子 (2020). 保護観察における新たなアセスメントツール—期待される効果と課題—. 千葉大学教育学部研究紀要, 68, 317-322. <https://doi.org/10.20776/S13482084-68-P317>
- Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, 76(5), 365-377. <https://doi.org/10.1037/h0031643>
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587-604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Lowenkamp, C. T., Holsinger, A. M., Brusman-Lovins, L., & Latessa, E. J. (2004). Assessing the inter-rater agreement of the Level of Service Inventory Revised. *Federal Probation*, 68(3), 34-38. https://www.uscourts.gov/sites/default/files/68_3_6_0.pdf
- 森文弓・東山哲也・西田篤史 (2014). 法務省式ケースアセスメントツール (MJCA) の基礎的研究 I—MJCAの開発及び信頼性・妥当性の検証—. 犯罪心理学研究, 52 (特別号), 54-55.
- Rocque, M., & Plummer-Beale, J. (2014). In the eye of the beholder? An examination of the inter-rater reliability of the LSI-R and YLS/CMI in a correctional agency. *Journal of Criminal Justice*, 42(6), 568-578. <https://doi.org/10.1016/j.jcrimjus.2014.09.011>
- Schmidt, S., Heffernan, R., & Ward, T. (2020). Why we cannot explain cross-cultural differences in risk assessment. *Aggression and Violent Behavior*, 50, Article 101346. <https://doi.org/10.1016/j.avb.2019.101346>