

鑑定人の信用性に影響を及ぼす要因の検討^{1, 2)}

目白大学心理学部 財津 亘

【要 約】

本研究では、証人出廷時の鑑定人に対する信用性に影響を及ぼす要因を模索することを目的として、インターネットによるパネル調査を実施した（588名の男性，459名の女性，平均値44.7歳（標準偏差9.60））。調査に際しては、ある殺人事件の仮想裁判に裁判員として参加し、「DNA型鑑定」あるいは「ポリグラフ検査」を行った鑑定人がこの裁判に証人出廷するといったシナリオを想定して実施した。調査対象者は、いずれかの鑑定群に割り振られ、鑑定人が有する特徴30項目に対してどの程度信用できるか信用性の評価をもとめた。探索的因子分析（EFA）の結果によると、両群ともに2つの因子（「熟練度」因子と「覚束なさ」因子）が抽出された。「熟練度」因子は、「鑑定経験年数が10年程度である」、「今までに1,000件程度の鑑定経験がある」といった観測変数の背景となりうる因子である。それに対して、「覚束なさ」因子とは、「見た目に清潔感がない」、「下を向いて話す」などの観測変数に関連する因子である。さらに、多母集団同時分析（MGA）を実施したところ、「DNA型鑑定」群と「ポリグラフ検査」群の間に、因子平均のみ違いがみられ、鑑定種別による信用性の違いを示唆した。

キーワード：鑑定人，信用性，インターネットによる調査，探索的因子分析，多母集団同時分析

問題

裁判では、当該事件に関して鑑定を行った鑑定人³⁾が証人として出廷し、鑑定の経過および結果についての供述を求められることがある。出廷の際は、尋問を受けることとなるが、その際の鑑定人の供述内容は、鑑定手法の妥当性や信用性を反映したものと見なされる。また、その供述内容もさることながら、後述のとおり、鑑定人から受ける好感度といった印象や属性情報も、当該鑑定結果を評価する際の材料として影響する可能性は否めない。そこで、本研究では、裁判員となり得る一般市民を対象に、証人出廷時の鑑定人のどのような特徴がその鑑定人に対する信用性に影響を及ぼすのかを検討することで、鑑定人が証人出廷する際に参考となる知見を提供することを目的とする。

本邦では、2009年より裁判員制度が導入されている（最高裁判所，2021）。「裁判員の参加する刑事裁判に関する法律（通称・裁判員法）」の

第一条には、「国民の中から選任された裁判員が裁判官と共に刑事訴訟手続に関与」することが明記されており、規定された条件に該当する者を除き、成人の日本国民は裁判員として選出され、裁判審理に参加する可能性がある。また、この制度下では、裁判員制度対象となる刑事事件（「死刑または無期の懲役もしくは禁錮に当たる罪に係る事件」など）において、裁判員が「事実認定（証拠に基づいて判決に係る犯罪事実を認定すること）」および「量刑判断（有罪の場合に、科すべき刑罰を決定すること）」までを行う。刑事訴訟法では、事実認定は証拠によるといった「証拠裁判主義」が採用されているとともに、その証拠に対する証明力（事実を証明する際にその証拠を信用できる度合）は、裁判官や裁判員の自由な判断に委ねるといった「自由心証主義」が導入されている。この点に関して、従来の裁判では裁判官が鑑定書などの書面を読み込んで心証形成を行っていたものである

が、一般市民が裁判員となる裁判員制度のもとでは、書面を読み込んで心証形成を行うといったことが困難と言える。このことから、証人を法廷に出廷させて尋問を行い、その法廷でのやり取りによって心証を形成するといった変化がみられるようになったとされている（本庄，2017）。

ところで、科学警察研究所や各都道府県警察の科学捜査研究所で実施されている科学鑑定（DNA型鑑定や薬毒物鑑定，筆跡鑑定，ポリグラフ検査など）の結果は鑑定書などの書面として作成され，犯罪捜査や裁判における犯罪立証に用いられる。日本の刑事裁判では，検察側もしくは弁護側の一方が証拠採用に不同意とした場合に，裁判所が鑑定人を法廷に出廷させ，鑑定人への主尋問と反対尋問を経ることで基本的には証拠採用される（勝又，2008）。ただし，鑑定書の証拠能力が認められ，証拠の一つとして法的判断の材料となるとしても，前述のとおり，裁判に提出された鑑定書に対する証明力は，裁判官ないし裁判員の鑑定書に対する主観的な信用の度合そのものに依存する。また，その鑑定書に対する信用の度合は，出廷時の鑑定人の供述内容もさることながら，鑑定経験などの裏付けや出廷時の振る舞いなどからも少なからず影響を受けることが推察される。

この点に関して心理学的研究を概観すると，陪審員⁴⁾が（鑑定人などの）専門家証人に対して抱く信用性判断の影響を検討した海外の研究は数多く存在する。たとえば，Brodsky, Neal, Cramer, & Ziemke (2009) は，俳優に専門家証人を演じさせるといった模擬裁判場面による実験を行うことで，専門家証人の“likeability (好感度)”が，専門家証人への信用性に影響を及ぼすことを報告している。また，Brodsky, Griffin, & Cramer (2010) は，ある殺人事件に関する模擬裁判場面を大学生に提示し，その模擬裁判に登場した専門家証人に対する評価を求めるといった質問紙調査研究を行っている。因子分析の結果によると，専門家証人としての信用性に影響する4因子（理論的であるなど“knowledge (知識)”，親しみがあるなど“likeability (好感度)”，偽りがなくなど“trustworthiness (印象としての信用性)”，落

ち着いているなど“confidence (自信)”)が抽出されたとしている。ただし，これらの模擬裁判場面を用いた研究ではなく，実際の裁判における陪審員を対象に行った質問紙調査および面接研究によると，専門家証人に対する信用性については専門家証人自身の属性，中でも専門家証人の「経験年数」が最も重要であり，次いで「教育水準」が重要であることが示唆されている（Wilcox & NicDaeid, 2018）。

以上のとおり，諸外国では陪審員が専門家証人に対して抱く信用性判断に関する研究が存在するものの，本邦においては同種研究が管見の限りみられない。諸外国の先行研究では，専門家証人における特定の要因（Brodsky, et al. (2009) の“likeability (好感度)”)を操作する実験が行われているが，特定の要因に着目する実験を行うにあたり，わが国の裁判において，どのような要因が鑑定人の信用性に影響を及ぼし得るのか検討する必要がある。中でも，Wilcox & NicDaeid (2018) が主張するように，鑑定人の属性（「経験年数」や「教育水準」など）を中心に検討する必要がある。以上のことから，本研究では，裁判員となり得る日本人を調査対象として，鑑定人への信用性に影響する要因について模索することを目的とした。そのうえで，先行研究と同様に，各要因の背後に存在する因子を特定することで鑑定人への評価に係る心理的特性を検討することとした。

なお，「鑑定」とは，「特別の知識経験に属する法則またはこれを具体的事実に適用して得た判断の報告（中井，2010）」あるいは「裁判において裁判官には欠けている専門知識が必要となる場合に，専門家の力を借りてそれを補うため（本庄，2017）」に行われるものとされており，精神鑑定なども含まれる。本研究では日本において日常的に行われている警察鑑定，中でもDNA型鑑定とポリグラフ検査を題材に取り扱うこととした。DNA型鑑定は，ヒト身体組織の細胞内に存在するDNA（デオキシリボ核酸）の塩基配列を分析することによって，個人を高い精度で識別する鑑定法とされている（警察庁，2018）。DNA型鑑定は，近年における精度と活用状況から科学鑑定の中でもその代表格とも言え，一般市民の認知度も高いことが推察さ

れるために、本研究の題材の一つとして取り扱った。一方で、ポリグラフ検査は、いわゆる1923年における米国のフライ裁判⁵⁾をはじめ、長年その信用性や証拠としての価値が問われてきた経緯があり、日本も例外ではない(勝又, 2008; 辻脇, 2010; 山岡, 2000)。しかしながら、今現在日本の犯罪捜査で活用されているポリグラフ検査は、フライ裁判当時のポリグラフ検査と比較すると、測定装置や測定する生理指標も異なるうえに、米国をはじめとする欧米とは違い、日本では科学的な手法を基にした質問方法が採用されている(小林・吉本・藤原, 2009; Osugi, 2011)。また、日本のポリグラフ検査における正確性の高さについても報告されている(小川・松田・常岡, 2013, 2014)。それにも関わらず、依然としてポリグラフ検査に対する誤解は少なくない。たとえば、ポリグラフ検査は、正確には「嘘発見器」ではなく、「記憶検査の一種」であるにも関わらず、警察職員を対象とした調査によると、ポリグラフ検査が「嘘発見器である」に対して「はい」と答えた割合が51.9%と半数以上を占めていたと報告されている(建内・鈴木・山本, 2010)。また、この報告によると、ポリグラフ検査の「結果は、信用できる」に「はい」と回答した割合は72.9%であったという。この報告は、警察職員を対象とした調査から得られたものであるが、この報告からポリグラフ検査に対する一般市民の誤解の多さや信用性の低さは容易に予想される。以上の理由から、本研究では、ポリグラフ検査を題材として取り扱った。また、本研究が仮説生成型研究であること、また低コストといった利点から場面想定法(岡, 2018)を用いて、ある殺人事件の裁判に証人出廷する鑑定人について種々の特徴(鑑定経験年数や博士号の保有、外見的特徴、コミュニケーションの特徴など)がみられた場合にどの程度信用できるか評定を求めるインターネット調査を実施した。そこで、分析にあたっては、その鑑定人が実施した鑑定が「DNA型鑑定」である場合と「ポリグラフ検査」である場合に違いがみられるか否か比較検討することとした。

方法

調査対象者

株式会社マクロミルのクラウド型アンケートサービス「Questant(クエスタント)」の「GMOリサーチ」を利用して調査票の回答収集を依頼した。「Questant(クエスタント)」では、インターネット上で調査票を作成することができるほか、「GMOリサーチ」といったサービスを利用することで、作成した調査票をインターネット上で配信し、その結果を回収することができる(赤羽, 2020)。「GMOリサーチ」を利用して調査票の回答を収集するにあたり、性別の指定はしなかった。調査当時の裁判員選出対象の年齢が満20歳以上⁶⁾であったことから、20代から50代と指定した。また、後述のとおり「DNA型鑑定」群と「ポリグラフ検査」群を設定したが、それぞれ500名以上の回答を収集するように依頼した(両群で回答者が重複しないように依頼し、群間による比較を行った)。本調査は、2021年1月に実施し、1,102名の回答が得られた。年齢の記載に不備があった20名に加えて、30項目すべての質問に対して同じ評定値を記載していた35名に関しても、回答が明らかに不自然で分析には不適切と判断し、分析から除外した。除外後の調査対象者は、全体で1,047名(男性588名, 女性459名, 平均値44.7歳(標準偏差9.60), 中央値46歳, レンジ20—59歳)であった。

調査票

説明文 調査票の冒頭には、ある殺人事件およびその事件の裁判とともに、その裁判に証人出廷してきた鑑定人が行った鑑定についての説明文を記載した(詳細はFigure 1参照)。

上記説明文に続いて、次の項目を尋ねる設問となっている。

属性に関する項目 調査対象者の性別および年齢を尋ねる項目で構成されている。

鑑定人に対する信用性の評定を求める項目 以下の合計30項目で構成され、各項目において当該事件の鑑定人の特徴を説明している(各項目はTable 1を参照)。これらの項目は、Wilcox & NicDaeid(2018)で扱った専門家証人の属性「経験年数」や「教育水準」, 「研修」,

あなたは、次の事件の裁判に「裁判員(被告人に判決を下すために召集された一般人)として参加するもの」と想定してください。

(事件内容)

X市内で、当時58歳のYさんが居宅のこたつの中で血を流した状態で死亡しているのが発見されました。また、被害者の遺体には胸部に多数の創傷があり、死因は刃物で胸部を刺されたことによる失血死と推定されました。

そして、捜査を進めた結果、被害者とは面識のないXが後日逮捕されました。

(裁判)

逮捕・起訴された被告人であるXは、この殺人事件への関与を一貫して否認しています。そのため、この裁判では、被告人の犯行を裏付ける鑑定結果に焦点を置くこととなり、この事件に関して鑑定を実施した鑑定人がこの裁判で証人として出廷、鑑定書の説明をすることとなりました。鑑定書の説明に先立ち、鑑定人の紹介がされ、そのあとに鑑定人から鑑定書の説明がされました。

なお、この実施された鑑定とは「DNA型鑑定※」で、「犯行現場に遺留されていた凶器(刃物)の柄から検出されたDNA型と被告人のDNA型が合致した」という結果でした。

※「DNA型鑑定」とは、ヒト身体組織の細胞内に存在するDNA(デオキシリボ核酸)の塩基配列を分析することによって、個人を識別する鑑定法(警察庁, 2019)。

なお、この実施された鑑定とは「ポリグラフ検査※」で、「犯人しか知りえない事実(刃物の形状や侵入方法など)を被告人は知っている」という結果でした。

※「ポリグラフ検査」とは、ある事実を認識しているかどうかを、検査を受ける者の身体の反応(心拍や血管の収縮等の生理反応)の違いを手掛かりに調べる技術(警察庁, 2014)。

Figure 1 「DNA型鑑定」群で用いた説明文(実線内)

※「ポリグラフ検査」群で用いた説明文は、「なお」以下が破線内に置き換えられたもの

「資格」といった項目を中心に、また前述した先行研究を参考として作成した。加えて、証人出廷の経験を有する複数の鑑定人に意見を求め、参考とすることで作成した。

調査対象者には、「この鑑定人について次のような特徴がみられた場合に、あなたはこの鑑定人をどの程度信用しますか。次の7段階で評価してください」と教示し、7件法で評価するように求めた(「1:まったく信用できない」, 「2:信用できない」, 「3:やや信用できない」, 「4:どちらともいえない」, 「5:やや信用できる」, 「6:信用できる」, 「7:非常に信用できる」)。項目の提示順は、調査対象者ごとにランダムとした。

倫理的配慮 本調査は、事前に目白大学人文社会科学系研究倫理審査委員会の倫理審査を受けて承認を得たうえで実施した(承認番号20人-008)。

調査対象者には途中で何ら不利益なく、参加意思の拒否や中止、撤回が可能であること、調査票は無記名で個人が特定できない状態で保存されることを伝えた。そのうえで、調査票内に参加への同意回答欄を設け、同意した者のみ回答に進むように設定した。

分析方法

記述統計および効果量の算出 全サンプル(1,047名)の30項目に関する記述統計を算出した。また、本研究の目的は、各項目が示す特徴が当該鑑定人にみられた場合に、どの程度その鑑定人を信用するかを検討することである。このことから、各項目の評定値が「4:どちらともいえない」から正方向もしくは負方向からの程度離れ、信用性に関してどの程度の影響度があるのかを検討することとした。そこで、1サンプルの t 検定(基準値を「4」に設定)を実施するとともに、算出された t 値から効果量 d を算出し、各項目における「鑑定人に対する信用性」への影響度を検討した。なお、1サンプルの t 検定は、ある特定の基準値と得られたサンプルの平均値との差を検討する際に、帰無仮説「真の平均値が特定の基準値に等しい」を検討するものである(Cressie, Sheffield, & Whitford, 1984)。

因子構造の探索 「DNA型鑑定」群および「ポリグラフ検査」群ごとに、探索的因子分析(Exploratory Factor Analysis: EFA)を実施し、因子構造を探索的に検討することで仮説モデルを構築することとした。

多母集団同時分析による群間のモデル比較

「DNA型鑑定」群と「ポリグラフ検査」群に係る仮説モデルについて、平均構造・多母集団同時分析を実施し、群間でパラメータに違いがみられるか検討した。

以上の分析にはR (ver.4.0.3) を使用し、EFAの際にはpsychパッケージ (ver. 2.0.12) を、SEMおよび多母集団同時分析の際にはlavaanパッケージ (ver. 0.6-7) を用いた。

結果

全体および群別の記述統計および効果量

回収されたサンプルを群別に確認したところ、「DNA型鑑定」群で529名(男性312名, 女性217名, 平均値45.6歳(標準偏差9.55), 中央値47歳, レンジ20—59歳), 「ポリグラフ検査」群で518名(男性276名, 女性242名, 平均値43.8歳(標準偏差9.60), 中央値45, レンジ20—59歳)であった。性別と年齢に関して群間に違いがないかどうかを検討すべく, 統計的仮説検定(χ^2 検定, t 検定)を実施するとともに, 効果量を算出して比較検討した。統計的仮説検定の結果によると, 性別については群間で有意差がみられなかった($\chi^2(1, N = 1,047) = 3.22, p = .07, \phi = 0.06$)。一方で, 年齢については群間で有意差がみられた($t(1045) = 2.90, p < .01, d = 0.18$ (95%CI[0.06, 0.30]))。ただし, サンプルサイズが大きいほど, 有意差が検出されやすくなること(大久保・岡田, 2012; ラインハート・西原, 2017; 繁樹, 2002)を考慮すると, この有意差については, そもそも2群のサンプルサイズ(529名と518名)が非常に大きかったために有意差が検出されやすかったとも見受けられた。また, 得られた効果量は「0.18」であり, 効果量の基準(水本・竹内, 2011; 大久保・岡田, 2012)として提案されている「小(0.20)」を下回っている。このことから, 前述のとおり有意差がみられたものの, 実質的な差は小さいものと判断し, 本研究では両群の性別と年齢を同程度とみなして以降の分析に進んだ。

30の項目に対する評定に関して, Table 1に全体および群別の記述統計および効果量をまとめた。1サンプルの t 検定の効果量に係る数値

の右横に記載されている記号は, 評定の方向を示しており, 平均値が「4: どちらともいえない」より低い場合に「-」, 平均値が「4: どちらともいえない」より高い場合に「+」と記載している。また, 評定の大きさとして, それぞれの項目における効果量(水本・竹内, 2011; 大久保・岡田, 2012)が, 「小(0.20)」以上「中(0.50)」未満の場合に「+」「-」, 「中(0.50)」以上「大(0.80)」未満の場合に「++」「--」, 「大(0.80)」以上の場合に「+++」「---」と表記した。

続いて, 30項目において t 検定を実施し, 群間(「DNA型鑑定」群と「ポリグラフ検査」群)の比較を行ったところ, 30項目すべてにおいて有意差がみられた。しかしながら, 2群のサンプルサイズが大きいほど, 有意差が検出されやすいことを前述したとおりであることから, 30項目の効果量を算出した(Table 1)。これらの結果によると, すべての項目で有意差がみられたものの, 効果量「小(0.20)」を超えた項目は一部(「鑑定人の性別が「男性」(0.21)」や「鑑定人の性別が「女性」(0.21)」, 「鑑定人の年齢が「40代」(0.20)」)であり, 大幅に超えた項目は「質問に対する回答がちぐはぐ(0.32)」のみであった。

因子構造の探索と仮説モデルの検証

本研究では, 7件法のリッカート尺度の調査票を用いた。このような調査の場合, 心理学研究では従来から慣例的に間隔尺度とみなしてサンプルを取り扱うことが多いが, 本来であれば順序尺度として取り扱うべきであることから(小杉, 2013), 以降は順序尺度に対応したポリコリック相関行列を基にEFAを進めた。固有値を算出した結果, 「DNA型鑑定」群の第1因子で12.9, 第2因子で4.25, 第3因子で1.27, 第4因子で1.03, 第5因子以降で1.0未満となった。「ポリグラフ検査」群では, 第1因子で10.4, 第2因子で5.13, 第3因子で1.21, 第4因子で1.20, 第5因子以降で1.0未満であった。Guttman基準を目安にすると, 両群ともに4因子を採用することとなる。しかしながら, スクリーンテストや平行分析によると, 両群で2因子が妥当であることを示唆した。因子の解釈

Table 1
全サンプルおよび群別の平均値, 標準偏差, 効果量

質問項目	M	SD	Cohen's d^a	「DNA型鑑定」群		「ポリグラフ検査」群		Cohen's d^b
				M	SD	M	SD	
Q1 今までに「10件程度」の鑑定経験がある	3.86	1.12	0.12	3.98	1.16	3.74	1.06	0.11
Q2 今までに「100件程度」の鑑定経験がある	4.67	1.05	0.54	4.78	1.06	4.55	1.03	0.11
Q3 今までに「1,000件程度」の鑑定経験がある	5.17	1.08	0.74	5.28	1.08	5.06	1.07	0.10
Q4 鑑定経験年数が「1年程度」である	3.53	1.12	0.39	3.65	1.18	3.40	1.05	0.11
Q5 鑑定経験年数が「5年程度」である	4.26	0.99	0.26	4.40	0.99	4.13	0.97	0.14
Q6 鑑定経験年数が「10年程度」である	4.76	0.99	0.61	4.90	0.96	4.62	0.99	0.14
Q7 鑑定経験年数が「20年程度」である	5.07	1.07	0.71	5.19	1.04	4.95	1.09	0.11
Q8 鑑定に関する研修を「受けたことがない」	2.64	1.23	0.74	2.72	1.28	2.58	1.17	0.06
Q9 鑑定に関する研修を「5回程度」受けている	3.90	1.06	0.10	4.01	1.06	3.78	1.05	0.11
Q10 鑑定に関する研修を「15回程度」受けている	4.32	1.04	0.30	4.45	1.01	4.19	1.05	0.13
Q11 鑑定人の性別が「男性」である	4.34	0.85	0.37	4.51	0.90	4.16	0.75	0.21
Q12 鑑定人の性別が「女性」である	4.36	0.88	0.38	4.54	0.91	4.18	0.81	0.21
Q13 鑑定人の年齢が「20代」である	3.76	0.97	0.24	3.93	0.98	3.58	0.93	0.18
Q14 鑑定人の年齢が「40代」である	4.38	0.86	0.41	4.55	0.90	4.21	0.79	0.20
Q15 鑑定人の年齢が「60代」である	4.31	0.93	0.32	4.44	0.95	4.18	0.88	0.14
Q16 「博士号を持っていない」	3.81	0.98	0.19	3.95	0.99	3.68	0.95	0.14
Q17 「その鑑定に関連のない博士号」を持っている	3.88	0.99	0.12	4.01	1.03	3.74	0.93	0.14
Q18 「その鑑定に関連する博士号」を持っている	4.87	1.09	0.63	5.04	1.06	4.70	1.09	0.15
Q19 最終学歴「大学卒業」である	4.22	0.88	0.25	4.37	0.91	4.08	0.82	0.17
Q20 最終学歴「大学院(修士課程)修了」である	4.42	0.97	0.40	4.58	0.99	4.25	0.92	0.17
Q21 最終学歴「大学院(博士課程)単位取得満期退学」である	4.18	0.96	0.18	4.31	1.00	4.03	0.91	0.15
Q22 「髪の毛がボサボサ」である	3.59	1.07	0.36	3.74	1.10	3.44	1.02	0.14
Q23 「見た目に清潔感がない」	3.46	1.11	0.44	3.65	1.14	3.26	1.04	0.18
Q24 「早口で話す」	3.69	0.98	0.30	3.84	0.99	3.53	0.94	0.16
Q25 「ゆっくり話す」	4.48	0.82	0.50	4.55	0.85	4.41	0.79	0.09
Q26 「下を向いて話す」	3.25	1.10	0.56	3.37	1.14	3.14	1.05	0.11
Q27 「堂々と話す」	4.60	0.95	0.53	4.69	0.95	4.50	0.94	0.10
Q28 「話の趣旨を脱線しがち」である	2.97	1.17	0.66	3.09	1.22	2.86	1.10	0.10
Q29 「質問に対する回答がちぐはぐ」である	2.31	1.17	0.82	3.09	1.19	2.27	1.15	0.32
Q30 「質問に対する回答が的確」である	5.19	1.07	0.75	5.32	1.04	5.06	1.08	0.12

^a 1 サンプルの t 検定に係る効果量

^b 群間比較に係る t 検定の効果量

可能性も含めると、両群ともに2因子を採用することが妥当であると判断した。そこで、改めて、群ごとにEFA（最尤推定法，プロマックス回転）を繰り返した。EFAの最終結果によると、両群ともに、第1因子には、(a)「鑑定経験年数が20年程度である」、(b)「質問に対する回答が的確である」、(c)「今までに1,000件程度の鑑定経験がある」、(d)「その鑑定に関連する博士号を持っている」、(e)「鑑定経験年数が10年程度である」、(f)「堂々と話す」といった観測変数がみられた。これらの観測変数の背景に

は、調査対象者が認識した（鑑定人の）熟練度があるものと解釈し、第1因子を「熟練度」因子と命名した。また、第2因子については、(a)「質問に対する回答がちぐはぐである」、(b)「下を向いて話す」、(c)「話の趣旨を脱線しがちである」、(d)「見た目に清潔感がない」、(e)「髪の毛がボサボサである」といった観測変数が見受けられた。これらの観測変数の背景として、鑑定人としての覚束なさが関連しているものと解釈し、第2因子については「覚束なさ」因子と命名した。なお、「熟練度」因子と「覚束

なさ」因子は1成分で説明できるようにも見受けられるが、前者は鑑定人に対する信用性を高めるというプラス方向に作用する因子として、後者は鑑定人の信用性を低下させるというマイナスに作用する因子で異なるものと解釈できる。したがって、たとえ「熟練度」因子が高い鑑定人であっても、証人出廷の当日に裁判官からの質問などに対して適切な受け答えができない、または見た目には清潔感が無いようにみられるなどの状況があれば、「覚束なさ」因子によるマイナス作用が働き、信用性を低下させることになるものと解釈できることから、異なる因子として扱うことは妥当であると判断した。

多母集団同時分析による群間のモデル比較

上記のモデル（「熟練度」因子と関連する6観測変数、「覚束なさ」因子と関連する5観測変数）を用いた平均構造・多母集団同時分析を実施することで、群間（「DNA型鑑定」群と「ポリグラフ検査」群）のパラメータに差異がみられるか検討することとした。平均構造・多母集団同時分析に先立ち、Shapiro-Wilkの多変量正規性に関する検定を行ったところ、両群ともに有意差がみられた（「DNA型鑑定」群、 $W = 0.80$, $p < .01$ ；「ポリグラフ検査」群、 $W = 0.81$, $p < .01$ ）。多変量正規性が確認できなかったことから、本研究では、推定法として「対角重み付き最小2乗法（室橋，2020）」を採用した。

多母集団同時分析では、因子負荷量や観測変数の切片および誤差分散、因子の平均および分散といったパラメータの差異を検討するにあたり、各母集団の因子構造は同一という想定のもとで、各パラメータに等値制約を仮定した六つのネストモデル（入れ子状態の等値制約モデ

ル）を仮定したうえで、適合度指標を基にモデルの比較検討を行い、最適モデルを採択することとした（Hirschfeld & Brachel, 2014；池原, 2020）。Hu & Bentler（1998, 1999）の基準（ CFI 「.95以上」、 TLI 「.95以上」、 $RMSEA$ 「.06以下」、 $SRMR$ 「.08以下」）に照らし合わせた結果、ネストモデルの中でも、「因子負荷量や観測変数の切片および誤差分散が群間で等値」としたモデルがほかのモデルと比較して適合度が相対的に高かったことから、この等値制約モデルが最適であるものと判断した。続いて、採択した等値制約モデルを受けて、因子の平均構造（因子平均）が両群で異なることが示唆されたため、各群において因子得点を算出し、各因子における「DNA型鑑定」群と「ポリグラフ検査」群間の因子平均を比較検討した。「熟練度」因子に関しては、「DNA型鑑定」群で因子平均0.00（ SD 0.73）、「ポリグラフ検査」群で因子平均-0.26（ SD 0.76）が得られ、群間で有意差がみられた（ $t(1045) = 5.57$, $p < .01$, $d = 0.34$ （95%CI[0.22, 0.47]））。また、「覚束なさ」因子については、「DNA型鑑定」群で因子平均0.00（ SD 0.86）、「ポリグラフ検査」群で因子平均-0.29（ SD 0.74）が得られ、群間において有意差がみられた（ $t(1027.4) = 5.90$, $p < .01$, $d = 0.36$ （95%CI[0.24, 0.48]））。両群の因子ともに「中」程度の効果量であった。これらの結果が意味するところは、調査対象者が、鑑定人の「熟練度」および「覚束なさ」といった観点において、全体的に「DNA型鑑定」と比較して、「ポリグラフ検査」に対して低く信用性を評価していたと言える。Figure 2が最終的に得られたモデルである。

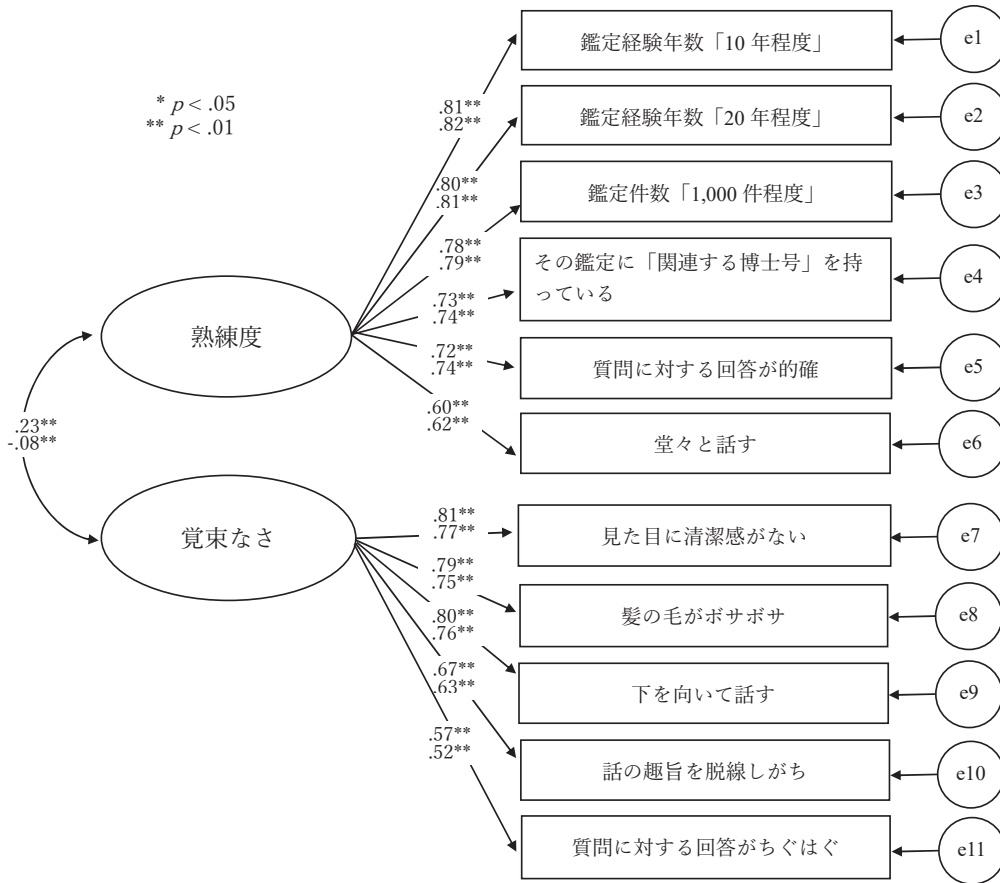


Figure 2 多母集団同時分析によるパス図

(標準化推定値, 上段が「DNA型鑑定」群, 下段が「ポリグラフ検査」群)

考察

本研究では、裁判員となり得る一般市民を調査対象とし、ある殺人事件の裁判を仮定した場面想定法を用いて、DNA型鑑定あるいはポリグラフ検査に係る鑑定人をどの程度信用できるか評定させる課題を求める調査を実施した。本調査によると、鑑定人の信用性を高める要因として、「質問に対する回答が的確である」、「今までに1,000件程度の鑑定経験がある」、「鑑定経験年数が20年程度である」、「その鑑定に関連する博士号を持っている」、「鑑定経験年数が10年程度である」、「今までに100件程度の鑑定経験がある」が挙げられた。それに対して、鑑定人の信用性を下げる要因については、「質問に対する回答がちぐはぐである」、「鑑定に関する研修を受けたことがない」、「話の趣旨を脱線しがちである」、「下を向いて話す」が確認された。

以上の結果をみると、鑑定件数や鑑定経験年数といったプラス要因の影響も大きいですが、それ以上に証人出廷時の裁判官などの質問に対する受け答えの悪さがマイナス要因として大きく作用することが分かる。「最終学歴」に関しては、「大学卒業」に比べて、「大学院（修士課程）修了」といった教育水準がより高いことで信用性が高くなることを示したものの、さらに教育水準が高いと思われる「大学院（博士課程）単位取得満期退学」は信用性を高める要因とは言えなかった。これについては、一般市民にとって大学院の博士課程に馴染み深いとは言えないこと、「退学」という文言の印象などから、信用性の判断における材料として効果がなかったものと解釈できる。また、各項目に関して、群間（「DNA型鑑定」群および「ポリグラフ検査」群）の比較を行ったところ、「質問に対する回答

がちぐはぐ」のみ効果量「小」を大幅に超える効果量が得られた。このことは、ポリグラフ検査の鑑定人が裁判員などの質問に対して適切な回答ができない場合に、DNA型鑑定人よりもマイナス要因として作用することを意味する。

EFAを行ったところ、鑑定人の「熟練度」と「覚束なさ」といった二つの因子を抽出することができたことから、鑑定人に対する信用性評定は、これらの二つの軸を基になされていることがうかがえた。「熟練度」は、鑑定人としての信用性を高める方向に作用する因子である一方で、「覚束なさ」は、信用性を低下させる因子と言える。

以上の結果を鑑みると、鑑定経験が豊富な鑑定人であれば、証人出廷時に鑑定実績などをどれだけアピールできるかが肝要となる。したがって、事前に自身の鑑定経験の年数やそれまでに実施してきた鑑定件数を整理して把握するといった証人出廷の基本的な準備が必須であることが確認できたと言えよう。また、本調査によれば、経験の豊富さを示す熟練度が高い鑑定人であっても、質問に対する受け答えが適切ではない場合などは、鑑定人の信用性を低下させる可能性があることを示した。したがって、証人出廷当日に裁判官や裁判員の質問に対する適切な受け答えが落ち着いてできるように、鑑定経験が豊富な鑑定人であっても、証人出廷に先立ち、打ち合わせを入念にするとともに、想定される質問を整理し、受け答えの練習を事前に行うといったことも重要であろう。それに対して、鑑定経験が浅い鑑定人の場合は、そもそも鑑定に関する熟練度を示す材料が少ないことが難点となる。しかしながら、本調査によれば、「鑑定経験年数」や「鑑定件数」に比べても、証人出廷当日に「的確な回答をする」ことで熟練度の低さをある程度補うことができることを示している。同様に、証人出廷時に、質問に対する回答がちぐはぐである、また下を向いて話すことで、裁判員に対して不信感を与えてしまう可能性があることから、鑑定経験が豊富な鑑定人と同様、自身が実施した当該鑑定に関する整理や振り返り、打ち合わせなどの準備を万端にすることが欠かせない。

本研究における平均構造・多母集団同時分析

の結果によると、両群（「DNA型鑑定」、「ポリグラフ検査」群）間で、「熟練度」因子・「覚束なさ」因子それぞれの因子平均に有意差がみられ、「熟練度」および「覚束なさ」といった観点からみると、全体的に「ポリグラフ検査」の信用性が「DNA型鑑定」と比較して低いことが示唆された。ただし、本調査における鑑定に係る説明は、Figure 1で示したとおり、非常に単純であるのに対して、実際の裁判では鑑定に関する詳細を説明する点で大きく異なる。緒言で述べたとおり、今現在のポリグラフ検査の正確性（小川・松田・常岡（2013）の感度86%、特異度95%）について正しい認識を有している一般市民が少ない可能性があることを考慮すると、鑑定そのものに対する信用性は詳細な説明によってその認識が変化する可能性は十分あり得る。

本研究の限界として、次の四点が挙げられる。第一に、本調査の対象となった調査対象者が、事前にパネル調査の業者に登録している者のみで構成されている点である。つまりは、本調査の母集団が日本国民を反映しているとは限らない。第二に、本調査では、ある事件の裁判を想定した場面想定法による質問紙法を用いた。心理学の研究法として一般的ではあるものの（坂元，2018）、裁判員として裁判に実際に参加する状況とは異なるため、そのような経験のない状況や場面をどれほど具体的に想像できるかは個々の調査対象者の想像力に依存する研究法と言える。したがって、場面想定法を用いた本調査については生態学的妥当性に問題がある点は否めない。この点を改善するためには、前述した諸外国の研究のように、鑑定人を演じる俳優を用いた模擬裁判場面、あるいは実際の裁判場面を用いるなどの方法を採用して、鑑定人に対する評定を求めるなどの工夫を施した研究が今後は必要と言えよう。第三に、本研究は先行研究が存在しなかったため、調査票を作成するにあたり、探索的に項目を列挙したことから、調査票に含んでいない項目については検討の対象外となっている。このことから、今後も別の要因を探索的に検討する必要があるかもしれない。第四に、調査票の質問項目に関する構成の問題である。実際の個々の鑑定人が有

する特徴というのは、「鑑定経験年数が20年程度で、博士号を持っているが、質問に対する回答がちぐはぐである」など要因が複合的であるため、より生態学的妥当性を有する研究方法を採用することが望ましいであろう。

裁判員は裁判官とは違い、法との関わりや法への関心が薄い国民が対象となることが多いとともに、鑑定書などの書面を精査し、吟味したうえで法廷に臨むということが困難と言える。そして、鑑定の内容とは別に、証人として出廷する鑑定人からの説明、そしてその鑑定人に抱く印象が影響を及ぼすこともあり得ることから、鑑定人の信用性に関する心理学的研究の意義は深い。今後は、実際の裁判に近い状況を設定し、生態学的妥当性がより高い方法で検証するとともに、本研究では扱わなかった裁判員等の性別や年齢など人口統計学的特徴の影響に関する検討も今後は必要となろう。

引用文献

- 赤羽 優子(2020). 多人数質問調査法の現在(10)個人で行うインターネット調査の諸手法 計量国語学, 32, 447-465.
- アレックス・ラインハート(著)・西原 史暁(訳)(2017). ダメな統計学—悲惨なほど完全なる手引書— 勁草書房
- Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law*, 28, 892-907. <https://doi.org/10.1002/bsl.917>
- Brodsky, S. L., Neal, T. M., Cramer, R. J., & Ziemke, M. H. (2009). Credibility in the courtroom: How likeable should an expert witness be?. *Journal of the American Academy of Psychiatry and the Law Online*, 37, 525-532. <http://jaapl.org/content/37/4/525>
- Cressie, N. A. C., Sheffield, L. J., & Whitford, H. J. (1984). Use of the one sample t-test in the real world. *Journal of Chronic Diseases*, 37, 107-114. [https://doi.org/10.1016/0021-9681\(84\)90052-3](https://doi.org/10.1016/0021-9681(84)90052-3)
- Hirschfeld, G., & Brachel, R. V. (2014). Multiple-Group confirmatory factor analysis in R: A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research and Evaluation*, 19, 7.
- 本庄 武(2017). 刑事手続における科学鑑定の現状と課題—鑑定人の地位論を中心に— 一橋法学, 16, 1-21. <https://doi.org/10.15057/28498>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. <https://doi.org/10.1080/10705519909540118>
- 池原 一哉(2020). 速習共分散構造分析 豊田 秀樹 共分散構造分析[R編]—構造方程式モデリング—(pp. 23-43)東京図書
- 勝又 義直(2008). 裁判所における科学鑑定の評価について 日本法科学技術学会誌, 13, 1-6. <https://doi.org/10.3408/jafst.13.1>
- 警察庁(2018). 警察白書(平成30年版)Retrieved from https://www.npa.go.jp/hakusyo/h30/pdf/06_dai2sho.pdf (2021年1月31日)
- 小林 孝寛・吉本 かおり・藤原 修治(2009). 実務ポリグラフ検査の現状 生理心理学と精神生理学, 27, 5-15. <https://doi.org/10.5674/jjppp.27.5>
- 小杉 考司(2013). 順序尺度の相関係数(ポリコリック相関係数)について Retrieved from <http://kosugitti.sakura.ne.jp/wp/wp-content/uploads/2013/08/polynote.pdf> (2021年1月31日)
- 水本 篤・竹内 理(2011). 効果量と検定力分析入門—統計的検定を正しく使うために— 2010年度部会報告論集『より良い外国語教育のための方法』, 47-73.
- 室橋 弘人(2020). 推定法 豊田 秀樹 共分散構造分析[R編]—構造方程式モデリング—(pp. 205-209) 東京図書
- 中井 憲治(2010). 鑑定 河上 和雄・中山 善房・古田 佑紀・原田 國男・河村 博・渡辺 咲子(編) 大コンメンタル刑事訴訟法 第3巻【第2版】(pp.254) 青林書院
- 小川 時洋・松田 いづみ・常岡 充子(2013). 隠匿情報検査の妥当性—記憶検出技法としての正確性の実験的検証— 日本法科学技術学会誌, 18, 35-44. <https://doi.org/10.3408/jafst.18.35>

- 小川 時洋・松田 いづみ・常岡 充子(2014). 隠匿情報検査の生理反応—フィールドデータの分析—日本心理学会第78回大会発表論文集, 537.
- 岡 隆(2018). 独立変数の操作 高野 陽太郎・岡隆 (編)心理学研究法—心を見つめる科学のまなざし— (pp. 42-67) 有斐閣
- 大久保 街亜・岡田 謙介(2012). 伝えるための心理統計—効果量・信頼区間・検定力— 勁草書房
- Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test*. (pp. 253-275). New York: Cambridge University Press.
- 最高裁判所(2021). 裁判員制度 Retrieved from <https://www.saibanin.courts.go.jp/index.html> (2021年4月22日)
- 坂元 章(2018). さまざまな実験法 高野 陽太郎 (編)・岡 隆(編)心理学研究法—心を見つめる科学のまなざし— (pp. 120-146) 有斐閣
- 繁榎 算男(2002). 仮説の統計的評価とベイズ統計学 佐伯 胖・松原 望(編)実践としての統計学 (pp. 147-178) 東京大学出版会
- 建内 利彦・鈴木 玲子・山本 直宏(2010). 犯罪捜査で使用されるポリグラフ検査に対する意識調査 日本心理学会第74回大会発表論文集, 455.
- 辻脇 葉子(2010). 科学的証拠の関連性と信頼性 明治大学法科大学院論集, 7, 413-443.
- Wilcox, A. M., & NicDaeid, N. (2018). Jurors' perceptions of forensic science expert witnesses: Experience, qualifications, testimony style and credibility. *Forensic Science International*, 291, 100-108. <https://doi.org/10.1016/j.forsciint.2018.07.030>
- 山岡 一信(2000). 公判廷での証拠能力 平 伸二・中

山誠・桐生 正幸, 足立 浩平(編著) ウソ発見—犯人と記憶のかけらを探して—(pp. 101-109) 北大路書房

【脚注】

- 1) 本研究は、2020年度目白大学特別研究費の助成を受けて実施されたものである。
- 2) 本論文に関して、開示すべき利益相反関連事項はない。
- 3) 刑事訴訟法では、「鑑定人」とは、「裁判所または裁判官から鑑定を命じられた者」を指し、「捜査機関から鑑定を依頼された者」は「鑑定受託者」と定義されるが、本研究では便宜上「刑事事件において鑑定に携わった者」を「鑑定人」として扱う。
- 4) 諸外国においても、国民が刑事裁判に参加する制度が導入されているが、米国や英国などにおいて導入されている制度は「陪審制」である。「陪審員」は「事実認定」を行うものの、「量刑判断」は行わないといった違いがある。
- 5) フライ裁判は、1923年米国の法廷ではじめてポリグラフ検査の証拠能力が検討され、その証拠能力が認められなかったことに加えて、その他の科学的鑑定に対する評価基準(フライ基準)が設けられた裁判として知られている。
- 6) 裁判員の選出対象年齢は、公職選挙法改正にともない2022年4月1日以降は「満十八年以上の者」であるが、調査当時は法改正前のため「20歳以上」を調査対象とした。

—2022年9.6.受稿, 2022年11.24.受理—

Factors affecting expert witness' credibility

Wataru Zaitu

Mejiro University, Faculty of Psychology

Mejiro Journal of Psychology, 2023 vol.19

[Abstract]

This study investigated factors affecting the credibility of expert witnesses through an Internet-based, self-administered questionnaire survey. Respondents were 588 men and 459 women (Mean Age = 44.7 years, $SD = 9.60$). This study assumed the respondents were participating in a murder trial as lay judges and divided into either two scenario groups: One group read an explanation that a "DNA test" has been conducted in the case while the other to a "Polygraph test". A questionnaire was composed of 30 items on the expert witness's characteristics. The credibility in each item were rated by respondents in both groups. Exploratory factor analysis (EFA) extracted two factors: the expert witness's "Skill level" and "Uncertainty". The Skill level formed the background to "ten years as a career" and "made a thousand cases until today", amongst others. By contrast, the Uncertainty factor was related to "not neat and tidy" and "looking down while explaining", amongst others. Furthermore, multiple-group analysis (MGA) indicated only a difference in factor means between the DNA and Polygraph test groups.

keywords : expert witness, credibility, internet questionnaire survey,
exploratory factor analysis (EFA), multiple-group analysis (MGA)