【原著】 大学入試研究ジャーナル第 34 号，98-103，2024

完全情報最尤法を用いた令和 5 年度理科②の周辺平均

荘島 宏二郎，橋本 貴充，宮澤 芳光，石岡 恒憲，大津 起夫，前川 眞一（大学入試センター）

令和 5（2023）年 1 月に実施された第 3 回大学入学共通テストの理科②において得点調整が実施された。本研究では，20 点差以上ついた物理・生物の平均点差のうち，選択科目における受験者集団の学力差によって，どれくらいの割合が説明できるのかを統計分析によって検討した。具体的には，モデルとして多変量正規分布を，欠測メカニズムとしてランダム欠測を仮定し，完全情報最尤法によってモデルの母数（多変量正規分布の平均ベクトルと共分散行列）を推定した。用いたデータは，中間集計データ（22 万人，全体の約 45%）であった。その結果，平均点差は，13.73 点（57.4%）に縮小した。

キーワード：共通テスト，理科②，得点調整，完全情報最尤法，ランダム欠測

1 はじめに

令和 5（2023）年 1 月に実施された第 3 回大学入学共通テスト（以下，共通テスト）では，理科②において得点調整が実施された（大学入試センター，2023a）。得点調整は，原則として，得点調整対象科目間に 20 点以上の平均点差が生じ，これが試験問題の難易差に基づくものと認められる場合には，得点調整を行う（大学入試センター，2022, pp.52–53）とされている。ただし，受験者数が 1 万人に満たない科目については，得点調整対象科目から外れるという付則がある。

表 1 は，2023 年 1 月 18 日（水曜日）付で（大学入試センター，(2023b）が公表した各科目テストの基本統計量のうち，理科②の 4 科目について抜粋したものである。水曜日時点で公表される統計資料は，中間発表と呼ばれ，全体の半分弱である約 22 万人の集計結果となっている。このうちの 56,212 人が物理を受験し，彼らの平均点が 64.46 点であった，ということである。その他の科目も同様である。

表 1 理科②の基本統計量（中間発表 221,659 人）

科目	n	平均	Max	Min	SD
物理	56,212	64.46	100	0	22.80
化学	67,665	49.95	100	0	20.14
生物	21,500	40.55	96	0	15.14
地学	656	49.12	100	0	21.50

表 1 より，物理と生物の平均点差が約 24 点であることが分かる。20 点以上の差は，最終結果まで維持され，この差が科目の難易差に基づくと判断されたことにより得点調整が実施された。なお，地学は，受験者数が 1 万人に満たなかったため，得点調整の対象から外れ，物理・化学・生物（物化生）の 3 科目のみで調整が行われた。

ところで，平均点差には，科目間の難易差のほかに受験者の学力差も混入している。一般的に，物理選択者の学力は高いので，物理（および化学）の平均点が高かったのは，物理選択者の学力が高いことも一因である。図 1 は，物化生選択者別の国語・数学・英語（国数英）の平均点を示している。物理選択者（と化学選択者）の国数英の平均点は生物選択者のそれらよりも高いことが分かる。

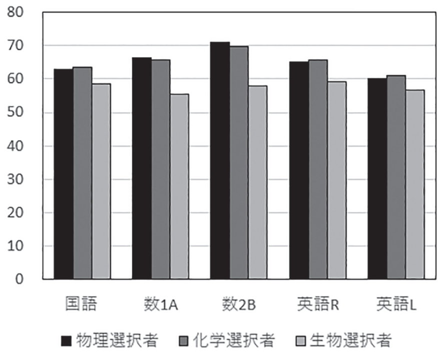

図 1 物化生選択者別の国数英の平均値
（ただし 200 点満点の国語は 0.5 を掛けて表示）

規則で「平均点差 20 点以上」が得点調整実施の前提条件となっているのは，平均点差が素朴な意味で科目間の難易差を代弁しているからである。しかし，実際には，科目間の平均点差には，科目間難易差と選択受験者間学力差が混交しているため [1)]，平均点差から科目間難易差を見積もることは難しい。本研究では，完全情報最尤法（full-information maximum likelihood, FIML; Arbuckle, 1996）を用いて，「仮に全員が物化

生を受けたら平均点がそれぞれ何点になるのか」ということを推測することで，平均点差に含まれる選択受験者間学力差の除去を試みる[2)]。

共通テストは，全体で10の時間区分（コマ）に分けられて各教科科目が実施されている。受験者は，各自必要な科目を選択して受験するため，受験しなかった科目の得点は欠測データとなる。平均点は、選択した（観測された）受験者の得点から算出するため，選択受験者の学力が高い（低い）と平均点が高く（低く）なるという性質がある。

この性質を修正する１つのやり方として，欠測となっている受験者の分も含めて平均点を推測することがある。例えば，物理を選択しなかった受験者を含めた全受験者の物理の得点の平均値を推測することがある。仮に全員が物化生を受けたときのそれら３科目の平均値を知ることができれば，そこで得られた平均点差は，学力差が除去されて残った科目間の難易差と思うことができる。

欠測を考慮した分析という意味で，同様な試みを行った先行研究として，ランダムフォレストを用いた方法（石岡, 2011），非線形因子分析を用いた方法（大津, 2011），加算モデルを用いた方法（前川, 2020），傾向スコアを用いた方法（吉村・荘島, 2004; 荘島・橋本・宮澤・石岡・前川, 2022）などがある。なお，共通テストの前身である大学入試センター試験に対して，FIML を適用した例は，荘島・大津・田栗（2006）や Shojima, Otsu, Mayekawa, Taguri, & Yanai（2007）にあるが，得点調整における平均点差に着目した分析ではなく，科目間の相関構造や因子構造に着目した分析であった。また，荘島・石塚・橋本（2007）では，FIML を用いた科目の平均点にも触れられているが，全体として得点調整の枠組みを論じており，FIML はそのうちの手法の１つとして取り上げられているのみで詳述されていない。本研究では，FIML を用いた科目間平均点差に焦点化して論じることを目的とする。

2 方法

2.1 データ

１月 18 日（水曜日）付けで公表された 221,659 人の中間集計データを用いる。これは，最終集計の約 45% の受験者に関するデータである[3)]。

また，分析に用いる科目は，表２に挙げた 19 科目とした。共通テストは全部で 31 科目あるのに，なぜこれら 19 科目を選択したのかについては，これらは受験者数の多い主要科目だからである。英語以外の外国語など，中間発表の時点で受験者数が 2,000 人に満たない 12 の少人数科目は除いた。

表２ 分析に用いる 19 科目

教科	科目
国語	国語
地歴	世界史 B，日本史 B，地理 B
公民	現代社会，倫理，政治経済，倫理政治経済
数学	数学 1A，数学 2B
理科①	物理基礎，化学基礎，生物基礎，地学基礎
理科②	物理，化学，生物
外国語	英語 R，英語 L

したがって, 分析対象は, 中間集計 221,659 人のうち, 19 科目のどれか１科目以上を受験した者で 221,497 人であった。つまり，データのサイズは，221,497（人）× 19（科目）である。ただし，１人当たりの平均受験科目数は 5.65 科目（SD は 2.32）であり, 68.9% のデータは欠測している。

2.2 モデル

FIML を適用するにあたり，２つの仮定を必要とする。１つはモデル，もう１つは欠測メカニズムである。本節ではモデルについて説明する。本研究では，多変量（19 変量）正規分布を仮定する。したがって，多変量正規分布における平均ベクトルと共分散行列（SD と相関行列）が母数である。多変量正規分布の妥当性については，第 4.1 節で議論する。

2.3 欠測メカニズム

FIML を適用する上で必要な２つ目の仮定である欠測メカニズム（星野 , 2009; 高井・星野・野間 , 2016）について本節で説明する。欠測メカニズムとは，欠測データの起こり方について分類したものであり，それらは完全ランダム欠測（missing completely at random, MCAR），ランダム欠測（missing at random, MAR），非ランダム欠測（missing not at random, MNAR）の３タイプである。

MCAR は，欠測がどこで起こるのか，あたかもサイコロで決めたかのように（得点とは無関係に），完全にランダムに起こる状況である。病気や忌引き，災害を理由に，ある科目の受験を取りやめたならば欠測となり，MCAR に該当するであろう。

MAR は，ある科目の得点が観測か欠測かに関する確率が，観測されている得点に依存している場合である。一般に，共通テストは学力が高いほど多科目受験型になる傾向がある。学力が低いほど少数科目受験型となる傾向がある。共通テストの一因子性は高く，実

際に本データを固有値分解したときの第 1 固有値の寄与率は 64.2% である [4]。したがって，ある科目が欠測するかどうかは，他の科目（特に，英数国などの主要科目）の得点からある程度推測できる。すなわち，観測された得点が低いほど，全体的に少数科目型となって欠測しやすくなる。完全に MAR が成立するとは考えにくいが，部分的に成立すると考え，本研究では，欠測メカニズムに MAR を仮定する。FIML は，MAR のときに母数を偏りなく推定する方法である。

MNAR は，欠測されている得点自体が原因で，欠測が生じた場合である。例えば，状況としては少し考えにくいが，学力が高い受験生がいてマークシートを桁ずれ解答していたとする。受験終了間際に桁ずれに気づき，みじめな結果を残すくらいならば，と解答用紙を持って受験会場から去ったならば [5]，これは MNAR であろう（MAR ではないだろう）。解答用紙が残っていないので欠測であるし，学力が高いのに欠測しているので観測データから当該科目の得点（の低さ）を推測できないからである。

本研究では，ひとまず MAR を仮定する。また，MAR を仮定することの妥当性については，第 4.2 節で議論する。

2.4 FIML

FIML は，欠測パタンごとに尤度を構築し，尤度を最大化する母数を得る手法である。まず，多変量正規分布の密度関数を $p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ とする。ここで，$\boldsymbol{x}$ は，確率変数ベクトル，$\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ は，多変量正規分布の母数である平均ベクトルと共分散行列である。

いま，欠測パタン数を G とし（本データでは G=1842），受験者 i の欠測パタンが g 番目であれば 1，そうでなければ 0 となる欠測パタン指示子

$$m_{ig}=\begin{cases}1 & \text{受験者 } i \text{ の欠測パタンが } g \text{ 番目} \\ 0 & \text{それ以外}\end{cases}$$

を考える。そして，データ全体の尤度を

$$p(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{N}\prod_{g=1}^{G} p\left(\boldsymbol{S}_g\boldsymbol{x}_i\middle|\boldsymbol{S}_g\boldsymbol{\mu},\boldsymbol{S}_g\boldsymbol{\Sigma}\boldsymbol{S}_g'\right)^{m_{ig}} \quad (1)$$

と構成する。ここで，N(=221,497) は受験者数，x_i は受験者 i の得点ベクトル，$\boldsymbol{X}$= $\{\boldsymbol{x}_i\}$（$N \times 19$）はデータ行列である。

また，$\boldsymbol{S}_g$ は，欠測パタン g に対応した得点や母数を取り出す選択行列である。例えば，受験者 i が国語・日本史 B・英語 R・英語 L の 4 科目のみを受験したとすると，受験者 i の得点ベクトルは，

$$\boldsymbol{x}_i = [x_{i1}\ x_{i2}^{(0)}\ x_{i3}\ x_{i4}^{(0)}\ x_{i5}^{(0)}\ x_{i6}^{(0)}\ x_{i7}^{(0)}\ x_{i8}^{(0)}\ x_{i9}^{(0)}$$

$$x_{i10}^{(0)}\ x_{i11}^{(0)}\ x_{i12}^{(0)}\ x_{i13}^{(0)}\ x_{i14}^{(0)}\ x_{i15}^{(0)}\ x_{i16}^{(0)}\ x_{i17}^{(0)}\ x_{i18}\ x_{i19}]'$$

となる。下付き添え字の 1 ～ 19 は，表 2 の国語～英語 L までの最初行左から最終行右までの科目順に対応している。また，上付き添え字の（0）は当該の科目が欠測していることを表している。このとき，この欠測パタンに対応する選択行列は，

$$\boldsymbol{S}_g = \begin{bmatrix}1000000000000000000\\0010000000000000000\\0000000000000000010\\0000000000000000001\end{bmatrix}$$

となる。この $\boldsymbol{S}_g$ を用いると，

$$\boldsymbol{S}_g\boldsymbol{x}_i = [x_{i1}\ x_{i3}\ x_{i18}\ x_{i19}]'$$

となり，観測得点のみを取り出すことができる。同様に，平均ベクトルと共分散行列に対してもこの選択行列を適用することで，

$$\boldsymbol{S}_g\boldsymbol{\mu} = [\mu_1\ \ \mu_2\ \ \mu_3\ \ \mu_4]'$$

$$\boldsymbol{S}_g\boldsymbol{\Sigma}\boldsymbol{S}_g' = \begin{bmatrix}\sigma_1^2 & & & sym\\ \sigma_{3,1} & \sigma_3^2 & & \\ \sigma_{18,1} & \sigma_{18,3} & \sigma_{18}^2 & \\ \sigma_{19,1} & \sigma_{19,3} & \sigma_{19,18} & \sigma_{19}^2\end{bmatrix}$$

のように，観測得点に対応する平均と共分散のみを取り出すことができる。

このように構成されている（1）式の尤度は，欠測パタン数だけグループを用意して，全ての母数にグループ間等値の制約を入れた多母集団分析と考えることが可能である。そして，（1）式の尤度を最大化して母数を推定する方法が FIML である。最大化の方法は，Newton-Raphson 法や，Fisher のスコアリング法など，古典的な方法でよい。本研究では後者を用いた。

FIML は MAR のもとで不偏な推定量であるため，仮に欠測があっても母集団の母数（ここでは，平均と共分散行列）を推定することができる。したがって，語弊を恐れずに言えば，「仮に全員が 19 科目を選択受験したら，そのときの平均ベクトルと共分散行列がどうなるか」ということを 2 つの仮定の下で推測したものと言える。

3 結果

多変量正規性と MAR のもとで，FIML を用いて 19 変量正規分布の平均ベクトルと共分散行列を推定した。このうち，理科②の得点調整に関わる物化生の平均のみを表２に示す。

表２ FIML による理科②物化生の平均・SD

科目	平均	SE（平均）	SD
物理	55.18	0.070	23.60
化学	42.76	0.061	20.95
生物	41.45	0.076	15.77

多変量正規性と MAR の２つの仮定の下で，「仮に全員が受けたら」物化生間の最大平均点差は 13.73 点になると分かった。つまり，見かけの平均点差 23.92 点（表１）のうち，10.19 点は選択者集団の学力差であり，残る 13.73 点が科目間難易差であると示唆される。

得点調整の真意は，科目間難易差について調整するものであろう。そして，実受験者に対する得点調整の処置として，物化生の最大平均点差の約 24 点は，化学と生物選択者に対して加点することで約 15 点まで縮小した[6)]。すなわち約９点を詰めたことになる。科目間難易差は 14 点くらいであると見積もられているので，64%（9/14）くらいは調整によって難易差が軽減されたと思うことができよう。しかし，これはあくまでもモデルに基づく試算の１つにすぎないので，本分析結果によって，現実の得点調整を「点差の縮小幅が小さすぎる／大きすぎる」などと単純に評価することはできない。

4 結論

結果において，見かけの平均点差 23.92 点のうち，学力差で説明できる部分は 10.19 点であると述べた。しかし，結果の解釈には，注意が必要である。これは，①モデルが多変量正規分布であること，②欠測メカニズムが MAR であること，という２つの仮定に依拠した結果ということである。そこで，これら２つの仮定が，どれくらい妥当であるかについて議論しておく必要がある。

4.1 多変量正規性の検討

まず，①多変量正規性の仮定である。多変量（19 変量）正規分布は，個々の周辺分布も１変量正規分布に従うので，各科目の得点分布について Kolmogorov-Smirnov（KS）検定を行ったところ，19 科目の全てについて有意になった。すなわち，正規分布から逸脱していると判断された。とはいえ，KS 検定は，標本サイズ（受験者数）が大きいとき，どんなに正規分布に近似していても有意になってしまう。そこで，正規性の視覚的な確認のため，図２に物化生の３科目のみの P-P プロットを示した。

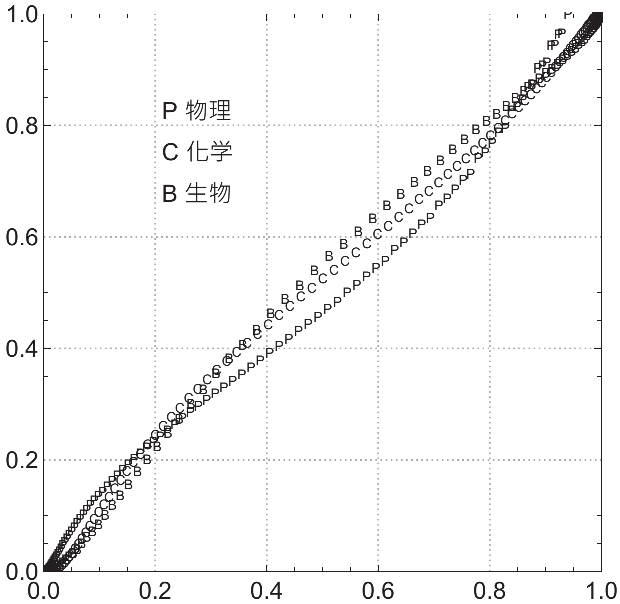

図２ 物化生の P-P プロット

P-P プロットは，プロットが対角線に沿って配置されていれば，分布が正規分布に近いことを意味している。図より，特に，物理は正規分布から大きく逸脱していることが分かる。１科目でも１変量正規分布に従わない科目が存在することは，全体として 19 科目の得点分布が多変量正規分布に従っていないことを示す証左である。しかし，統計分布を仮定するとき，多変量正規分布以外に扱いやすい多変量分布がないのも事実であり，今後の検討が必要である。

4.2 MAR の検討

実際は，中間発表データの約 22 万人の受験者が，さまざまな理由で欠測を起こすと考えられる。約 421 万（22 万人× 19 変量）のデータのうち，約 296 万（約７割）のデータは欠測している。その中には，MCAR によって起こった欠測もあるだろうし，MAR や MNAR で起こった欠測もそれぞれあるのが事実であろうが，多くは単純に（A）出願に必要ないから，（B）選択できないから，ということによる欠測であろう。

（A）については，例えば，理系学部に出願する予定の受験者が理科①を欠測するのは，理系学部の出願

に不要だからである。その代わり，理科②科目が観測されているので，理科①科目の欠測は，理科②科目の得点が観測されていることによって分かる。

また，(B) については，例えば，世界史Ｂと地理Ｂの受験者が日本史Ｂを欠測しているのは，地歴公民教科は２科目までしか受験できないからである。このとき，日本史Ｂの欠測は，世界史Ｂと地理Ｂが観測されたことによって分かる。また，日本史Ａと日本史Ｂのように，同一名称を含む２つの科目を同時に選択できないため，例えば，日本史Ｂを選択した受験者の日本史Ａの得点は欠測となる。したがって，日本史Ｂの観測によって日本史Ａの欠測が分かる。

つまり，(A) と (B) は，構造的に表れる欠測パタンであるが，いずれも欠測するか否かの情報を観測された科目群が保持しているので，これら欠測パタンの欠測メカニズムとしては MAR と見なすことができる。むろん，このような欠測の全てを観測された科目群によって説明できるわけではないが，「部分的に MAR が成立する」「MAR を仮定しても全く的外れではない」とは言ってよいだろう。

このように，欠測の全てではないが，大部分は，大学入試センターの受験案内が定めるところの科目選択に関する規則（大学入試センター , 2022, pp.3–4）が素直に適用されたことによる結果であるから，欠測の相当部分は MAR と見なすことができるだろう。また，仮に病気や災害などによる欠測（MCAR）が起きたとしても，FIML は，MCAR のもとでも正しく母数を推定することができる。

4.3 今後の課題と終わりに

全体の欠測のうち，いったい MNAR の割合がどれくらいなのか，ということが問題として残っている。定義上，１つでも MNAR によって起こった欠測があるならば，データ全体の欠測メカニズムは MNAR となる。そして，約 275 万の欠測データのうち，MNAR による欠測が１つもないという可能性はないので，共通テストデータ全体の欠測メカニズムは MNAR ということになる。この点，どのように欠測が起きたかについて，受験生に面接をするなどして MNAR の種類や割合などを調査し，欠測メカニズムとして組み込むことで，さらに精緻な分析を行うことができる可能性があるので今後の課題としたい。

なお，MAR か MNAR かの検定として Hausman 検定（Hausman, 1978）がある。しかし，この検定を実行するためには，まず，欠測データメカニズムをモデリングして一致性を持つ推定量を構築する必要がある（高井・星野・野間 , 2016）ため，MNAR に関して未調査の現段階では実行不可能である。今後の課題としたい。しかし，仮にできたとしても観測されなかった欠測データがなければ厳密な意味で MAR か MNAR の判別はできない（松山 , 2004）。

少なくとも，観測されたデータから平均・SD・相関を求めるという従来の統計資料は，欠測メカニズムが MCAR であるという厳しい条件を暗に仮定されていることについて，自覚的であるべきである。本研究によって，従来，暗に仮定されていた MCAR を MAR に緩和し，従来よりも真の母数に接近した結果を示すことができた。

注

1）その他，原因の特定できない様々な細かい要因も混入していると思われるが，本論文では扱わない。

2）本論文の見解は，著者らの個人的見解であり，所属組織の公式見解ではない。

3）分析に投入する受験者数が一定数いる科目は，中間集計データの平均点は，最終集計データの平均点と大きくても１点ほどしか違わない。したがって，中間集計データが十分に最終集計データを代表していると見なして分析する。

4）ただし，倫理政治経済を除いた 18 科目の共分散行列を固有値分解したときの結果である。倫理政治経済は，倫理および政治経済と同時受験ができないため，この２科目との共分散は計算できないためである。

5）マークシート（解答用紙）の持ち帰りは違反行為である。

6）得点が離散値であり，かつ，同一の得点をとる受験者がいるため，正確に 15 点差にならない。

謝辞

本研究は，大学入試センター理事長裁量経費研究「共通テストに対応した得点調整方法と成績提供法の検討」の助成を受けたものである。

参考文献

Arbuckle, J. L.（1996）Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker（Eds.）, *Advanced Structural Equation Modeling*. Lawrence Erlbaum Associates, Inc.（pp. 243–277）.

大学入試センター（2022）.「令和５年度大学入学共通テスト（本試験）の得点調整について」大学入試センター（https://www.dnc.ac.jp/albums/abm.php?d=34&f=abm00000322.pdf&n= 令和５年度「受験案内」（一括ダウンロード）.pdf）（2023 年１月 23 日）.

大学入試センター（2023a）.「令和５年度大学入学共通テスト（本試験）の得点調整について」大学入試センター（https://www.dnc.ac.jp/news/albums/abm.php?d=230&f=abm00003367.pdf&n=【プレス発表】令和５年度大学入学共通テスト（本試験）の得点調整について.pdf）（2023 年１月 23 日）.

大学入試センター（2023b）.「令和５年度大学入学共通テスト（本試験）平均点等一覧（中間集計）」大学入試センター（https://www.dnc.ac.jp/news/albums/abm.php?d=225&f=abm00003349.pdf&n=（プレス発表）令和５年度大学入学共通テスト（本試験）平均点等一覧（中間集計）_.pdf）（2023 年１月 23 日）.

Hausman, J. A.（1978）. Specification tests in econometrics. *Econometrica*, **46**, 1251–1271.

星野崇宏（2009）.『調査観察データの統計科学』岩波書店.

石岡恒憲（2011）.「Random Forest を用いた欠測データの補完に基づく大学入試センター試験科目間得点差」『応用統計学』**40**, 193–209.

前川眞一（2020）. 成績データから見たセンター試験 . 大学入試センター（編）センター試験」をふり返る, 164–187.（https://www.dnc.ac.jp/albums/abm.php?f=abm00040328.pdf&n=「センター試験」をふり返る.pdf）（2021 年８月８日）.

松山裕（2004）.「経時観察研究における欠測データの解析」『計量生物学』**25**, 89–116.

大津起夫（2011）. 大学入試センター試験における科目別得点の非線形因子分析による比較. 大学入試センター研究紀要, **40**, 1–23.

Rubin, D. B.（1976）. Inference and missing data. *Biometrika*, 63, 581–592.

荘島宏二郎･橋本貴充･宮澤芳光･石岡恒憲･前川眞一（2022）.「傾向スコアを用いた令和３年度共通テスト公民科目間差の試算―国語・英語リーディング・英語リスニングを共変量に用いた場合―」『 大学入試研究ジャーナル』**32**, 122–129.

荘島宏二郎・石塚智一・橋本貴充（2007）.「多重指標モニタリングによる得点調整手続きの試案」『大学入試センター研究紀要』**36**, 53–70.

Shojima, K., Otsu, T., Mayekawa, S., Taguri, M., & Yanai, H.（2007）. Factor structure of the National Center Test 2005 by the full-information pseudo-ML method. *Behaviormetrika*, **34**, 131-156.

荘島宏二郎・大津起夫・田栗正章（2006）.「自己組織化マップと完全情報最尤法から見た平成 17 年度センター試験の教科学力構造」『日本テスト学会誌』**2**, 75–89.

高井啓二・星野崇宏・野間久史（2016）.『欠測データの統計科学』岩波書店.

吉村宰・荘島宏二郎（2004）.「本追モニター調査の結果を利用した傾向スコア加重法による本追試験間の難易度比較」『大学入試センター研究紀要』**33**, 19–28.