論　文

項目難易度制約付き等質適応型テスト

岸田　若葉[†a)]　　渕本　壱真[†b)]　　宮澤　芳光[††c)]　　植野　真臣[†d)]

Item Difficulty Constrained Uniform Adaptive Testing

Wakaba KISHIDA[†a)], Kazuma FUCHIMOTO[†b)], Yoshimitsu MIYAZAWA[††c)], and Maomi UENO[†d)]

あらまし　本論では，能力推定値近傍の難易度パラメータをもつ項目に限定して出題する適応型テストを提案する．適応型テストとは，受検者の反応より逐次能力推定し，能力推定値に応じた項目を出題するテスト形式である．これにより，能力値を少ない項目数で高精度に測定できる．しかし，広範囲の能力値に対して大きな情報量をもつ項目が出題される傾向があり，このような特定項目の過度な暴露はテストの信頼性の低下につながる．この問題を解決するため，提案手法は，1 段階目では各受検者に割り当てた測定精度が等質な項目集合から出題する．能力推定値が収束したとき，1 段階目を終了する．2 段階目では能力推定値近傍の難易度パラメータをもつ項目に限定して出題する．評価実験では，提案手法は従来手法と比較して測定精度を同等に保ちつつ，暴露数の偏りを減少させることを示す．その原因として，従来手法が識別力パラメータの大きい項目を過度に暴露するのに対し，提案手法がその問題を緩和することを示す．

キーワード　適応型テスト，e テスティング，等質テスト自動構成，項目反応理論

1. まえがき

e テスティング [1], [2] の一種として，適応型テスト (CAT: Computerized Adaptive Testing) というテスト形式がある．適応型テストとは受検者の反応より逐次能力推定し，能力推定値に応じた項目を出題するテスト形式である．具体的には，各時点の能力推定値に対してフィッシャー情報量が最大の項目を出題する．これにより，適応型テストは従来の試験よりも受検者の能力値を少ない項目数で高精度に測定できる．

しかし，受検者の能力値の分布は標準正規分布に従い，その平均値付近の能力値をもつ受検者が多数存在するので，それらの能力値に対して高いフィッシャー情報量をもつ項目の暴露数が大きくなる傾向にある．過度に暴露された項目は受検対策される恐れがあり，テストの信頼性の低下につながる [3]．特定項目の暴露は実際に適応型テストが導入されている Synthetic Personality Inventory (SPI) [4] や Global Test of English Communication (GTEC) [5] においても問題となっている．

この問題を解決するために，暴露数を制御する様々な手法が提案されている [6]〜[14]．最も有名な手法として，van der Linden らは暴露数の上限を制約としたシャドーテストを逐次的に構成し，そのテストから適応的項目出題を行う手法を提案した [6]．ここで，シャドーテスト [6] とは，(1) 試験の制約 (例えば，テストの長さや暴露数の最大値) を全て満たし，(2) 既に受検者に出題した項目を全て含め，(3) 能力推定値に対して項目反応理論におけるテスト情報量が最大となる項目集合である．他にも，Sympson-Hetter 法では，事前にシミュレーションから算出した項目の出題確率を制御するパラメータを用いて，確率的に項目の出題を制御する．更に，van der Linden らは，事前にシミュレーションを必要としない手法を提案している [10], [11]．この手法では，項目の出題可能な確率を意味する適格確率 (Eligibility probability) を定義し，受検者ごとに適格確率に従ってアイテムバンクから項目を除くことで

† 電気通信大学大学院情報理工学研究科，調布市
Graduate school of Informatics and Engineering, The University of Electro-Communications, 1–5–1 Chofugaoka, Chofu-shi, 182–8585 Japan
†† 大学入試センター，東京都
The National Center for University Entrance Examinations, 2–19–23 Komaba, Meguro-ku, Tokyo, 153–8501 Japan
a) E-mail: kishida@ai.lab.uec.ac.jp
b) E-mail: fuchimoto@ai.lab.uec.ac.jp
c) E-mail: miyazawa@rd.dnc.ac.jp
d) E-mail: ueno@ai.lab.uec.ac.jp
DOI:10.14923/transinfj.2023JDP7012

暴露数を制御する．他には，分割したアイテムバンクを用いる手法が提案されている [14]．しかし，これらの手法は特定項目の暴露を抑えることができるが，受検者間でテストの長さや受検者の能力値の測定精度がばらつく傾向がある．

この問題を解決するため，宮澤・植野は等質適応型テストを提案した [15], [16]．等質適応型テストでは，事前に等質テスト構成手法により生成した等質な項目集合 (以降，等質アイテムバンクと呼ぶ) を用いる．等質テストとは，異なる項目から構成されるが受検者の能力値の測定精度が等質な項目集合である．等質テスト構成は多数研究されており，近年では大規模な数の等質テストを生成できる手法が提案され [17]〜[20]，情報処理技術者試験や医療系共用試験などで実用化されている [21], [22]．等質アイテムバンクの生成には，[15], [16] では当時最大数の等質テストを構成できた石井らの手法 [17] を用いている．等質適応型テストでは，等質アイテムバンクを各受検者に一つ割り当て，適応的項目出題を行う．これにより，受検者間の測定精度を等しく保ちつつ，暴露数を減少できた．しかし，等質適応型テストは暴露数を減少できるが，アイテムバンクサイズを縮小するので測定精度が低下するというトレードオフの問題がある．

このトレードオフを解決するため，宮澤・植野は 2 段階等質適応型テストを提案した [23], [24]．1 段階目は [15], [16] の等質適応型テストを用いる．能力推定値が収束したとき，1 段階目を終了する．2 段階目ではアイテムバンク全体から適応的項目出題を行う．これにより，高精度の能力推定値を得ることができる．結果として，この手法は従来型適応型テストと測定精度を同等に保ちつつ，暴露数を減少すると報告されている [23], [24]．宮澤・植野 [23], [24] は，1 段階目で推定値が真値近傍に収束し，2 段階目ではその能力推定値近傍の難易度をもつ項目のみが出題され，暴露数の偏りは軽減できると仮定している．この仮定は，各項目について，その難易度に一致する能力値へのフィッシャー情報量が最大となることを根拠としている．しかし，実際は項目の難易度が能力推定値と乖離していても，乖離していない場合に比較してフィッシャー情報量が大きくなる場合が多くある．結果として，広範囲の能力値に対して大きなフィッシャー情報量をもつ項目が出題されやすい．そのため，暴露数の偏りの改善は限定的かもしれない．

この問題を解決するため，本研究では，2 段階目以降は能力推定値近傍の難易度パラメータをもつ項目に限定して出題する項目難易度制約付き等質適応型テストを提案する．1 段階目では，2 段階等質適応型テストと同様に等質アイテムバンクから適応的項目出題を行う．等質アイテムバンクの生成には現在最も多くの等質テストが構成可能な Fuchimoto らの手法 [20] を用いる．能力推定値が収束したとき，1 段階目を終了する．2 段階目では，アイテムバンク全体のうち能力推定値近傍に対応した難易度パラメータ区間に属する項目に限定して適応的項目出題を行う．この難易度パラメータ区間は能力推定値の事後標準偏差に比例して決定する．以降，出題するたびに更新した難易度パラメータ区間に限定した適応的項目出題を繰り返す．前述のように，2 段階等質適応型テストでは受検者の能力値と難易度が乖離しても広範囲の能力値に対して大きなフィッシャー情報量をもつ項目は出題される傾向にあり，暴露数の偏りの原因となるが，提案手法ではこの問題の緩和が期待できる．

提案手法の有効性をシミュレーションデータと実データを用いて示す．その結果，提案手法は従来手法と比較して測定精度を同等に保ちつつ，暴露数の偏りを減少できる．

2. 項目反応理論による適応型テスト

2.1 項目反応理論

項目反応理論 (Item Response Theory: IRT) は，数理モデルを用いたテスト理論の一つであり，適応型テストの動作原理として用いられる [25]．IRT は受検者の項目への正答確率をモデル化したものである．これにより，異なる項目への受検者の反応を同一尺度上で評価できる．

適応型テストでよく用いられる IRT モデルとして 3 母数ロジスティックモデル (3-Parameter Logistic Model: 3PLM) が知られている．3PLM では，多肢選択式の問題において能力値の低い受検者が偶然正答する可能性を考慮して，能力値 $\theta \in (-\infty, \infty)$ の受検者が項目 $i \in \{1, \ldots, N\}$ に正答する確率を次の式で示す．

$$p(u_i = 1|\theta) = c_i + \frac{1 - c_i}{1 + \exp(-1.7a_i(\theta - b_i))} \qquad (1)$$

ここで，u_i は受検者が項目 i に正答するとき 1，それ以外のときは 0 である変数である．また，$a_i \in [0, \infty)$, $b_i \in (-\infty, \infty)$, $c_i \in [0, 1]$ はそれぞれ項目 i の識別力パラメータ，難易度パラメータ，当て推量パラメータで

ある．受検者が偶然正答する可能性を考慮しない場合 ($c_i = 0$)，2 母数ロジスティックモデル (2-Parameter Logistic Model: 2PLM) と呼ぶ．

適応型テストでよく用いられる他の IRT モデルとして，大問項目を扱う一般化部分採点モデル (Generalized Partial Credit Model: GPCM) がある．GPCM では能力値 θ の受検者が k 個の小問に正答する確率を次の式で示す．

$$p(k|\theta) = \frac{\exp(\sum_{m=1}^{k} \alpha_i(\theta - \beta_{im}))}{\sum_{l=1}^{K}[\exp(\sum_{m=1}^{l} \alpha_i(\theta - \beta_{im}))]} \tag{2}$$

ここで，β_{im} は項目 i の小問正答数が $m-1$ から m に遷移する難易度を示す．

2.2 フィッシャー情報量

項目反応理論において，能力推定値の標準誤差はフィッシャー情報量の逆数に漸近的に一致する [26]．能力値 θ をもつ受検者に対して項目 i が与えるフィッシャー情報量を以下の式で定義する．

$$I_i(\theta) = \frac{[\frac{\partial}{\partial\theta} p(u_i = 1|\theta)]^2}{p(u_i = 1|\theta)(1 - p(u_i = 1|\theta))} \tag{3}$$

フィッシャー情報量 $I_i(\theta)$ の大きい項目は能力値 θ 付近でその能力値をよく識別する．適応型テストでは，受検者の反応より逐次能力推定しフィッシャー情報量が最大の項目を出題することで効率の良い能力推定を実現する．本論文では情報量はフィッシャー情報量を指す．なお，テスト T を構成する項目集合の情報量の総和をテスト情報量と呼び，次のように示す．

$$I_T(\theta) = \sum_{i \in T} I_i(\theta) \tag{4}$$

テスト情報量の逆数が受検者の能力推定値の漸近分散に収束することが知られている [27]．

2.3 能力値 θ の推定

本論文で扱う適応型テストでは，受検者の能力推定に最も予測精度が高いと報告されている EAP 推定 (Expected a posteriori) を用いる [28]．受検者のそれまでの反応データ $\mathbf{u}$ と能力値 θ の事前分布 $f(\theta)$ を用いて，能力値の事後分布 $f(\theta|\mathbf{u})$ は次のように表される．

$$f(\theta|\mathbf{u}) = \frac{L(\mathbf{u}|\theta)f(\theta)}{\int_{-\infty}^{\infty} L(\mathbf{u}|\theta)f(\theta)d\theta} \tag{5}$$

ここで，$L(\mathbf{u}|\theta)$ は能力値を所与としたときに反応 $\mathbf{u}$ を返すゆう度である．EAP 推定は事後分布 $f(\theta|\mathbf{u})$ の θ に関する期待値を能力推定値 $\hat{\theta}$ とし，以下のように得られる．

$$\hat{\theta} = \int_{-\infty}^{\infty} \theta f(\theta|\mathbf{u})d\theta \tag{6}$$

2.4 適応型テスト

一般的な適応型テストでは，項目パラメータが既知のアイテムバンクを所与として，次のアルゴリズムに従って項目出題を行う．

（1） 能力推定値を $\hat{\theta} = 0$ に初期化する．

（2） 能力推定値 $\hat{\theta}$ を所与として情報量が最大の項目をアイテムバンクから選択し受検者に出題する．

（3） 反応データから能力推定値 $\hat{\theta}$ を更新する．

（4） 能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（2），（3）を繰り返す．

このように能力推定値に対して情報量最大の項目を逐次出題することで，少ない項目数で高精度な能力推定を実現する．しかし，受検者の能力値は一般に標準正規分布に従うため，受検者の能力分布の平均である $\theta = 0$ において情報量の大きい項目が過度に暴露されやすい．これらの項目は受検対策される恐れがあり，テストの信頼性の低下につながる [3]．そのため，適応型テストでは各項目を一様に出題することが望ましい．

3. 暴露数を制御する適応型テスト

本章では，従来型適応型テストの問題を解決するために提案された暴露数を制御する手法を紹介する．

3.1 整数計画問題 (Integer Programming Problem) に基づく適応型テスト (IP)

van der Linden らは，各項目の暴露数に最大暴露数 R を制約としたシャドーテストを逐次構成し，その中から適応的項目出題を行う手法を提案した (以降，IP と呼ぶ) [6]．具体的には，次のアルゴリズムに従って項目出題する．

（1） 能力推定値を $\hat{\theta} = 0$ に初期化する．

（2） 次の整数計画問題を解き，シャドーテストを構成する．

maximize

$$\sum_{i=1}^{N} I_i(\hat{\theta})x_i$$

subject to

$$r_i x_i \leq R (i = 1, \cdots, N)$$

(項目 i の暴露数 r_i，最大暴露数 R)

$$\sum_{i=1}^{N} x_i = n(\text{テスト項目数})$$

$$x_i = \begin{cases} 1 & \text{項目 } i \text{ がシャドーテストに含まれる} \\ 0 & \text{それ以外} \end{cases}$$

（3） シャドーテストから情報量が最大の項目を受検者に出題する．

（4） 反応データから能力推定値 $\hat{\theta}$ を更新する．

（5） 能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（2）から（4）を繰り返す．

3.2 整数計画問題の制約にターゲット情報量を用いた適応型テスト (TI)

整数計画問題を用いた他の手法として，Choi and Lim は制約にターゲット情報量（Target Information）を用いた手法を提案した（以降，TI と呼ぶ）[29]．具体的には，次のアルゴリズムに従って項目出題する．

（1） 能力推定値を $\hat{\theta} = 0$ に初期化する．

（2） 次の整数計画問題を用いてシャドーテストを構成する．ただし，T はターゲット情報量である．

minimize y

subject to

$$\sum_{i=1}^{N} I_i(\hat{\theta}) x_i \leq T + y$$

$$\sum_{i-1}^{N} I_i(\hat{\theta}) x_i \geq T - y$$

$$y \geq 0$$

$$\sum_{i=1}^{N} x_i = n(\text{テスト項目数})$$

（3） シャドーテストから情報量が最大の項目を受検者に出題する．

（4） 反応データから能力推定値 $\hat{\theta}$ を更新する．

（5） 能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（2）から（4）を繰り返す．

ただし，従来型適応型テストを実施したときのテスト情報量の平均をターゲット情報量 T として用いる．TI は，テスト情報量の最大化を目的関数とせず，あらかじめ設定したターゲット情報量とシャドーテストのテスト情報量の差を最小にすることで，過度に情報量の高い項目の暴露を抑えることができた．

3.3 確率的に出題項目を制御する適応型テスト (Prob)

別のアプローチとして，確率的に出題項目を制御する手法が提案されている [7]〜[11]．van der Linden らは適格確率 (Eligibility probabiliry) を用いた適応型テストを提案している（以降，Prob と呼ぶ）[10], [11]．具体的には，次のアルゴリズムに従って項目出題する．

（1） 能力推定値を $\hat{\theta} = 0$ に初期化する．

（2） 受検者 j に対する項目 i の適格確率 $P^{(j)}(E_i)$ を計算する．

$$P^{(j)}(E_i) = \min\{\frac{r^{max}}{P^{(j-1)}(A_i)} P^{(j-1)}(E_i), 1\} \tag{3}$$

ここで，r^{max} は暴露率の上限値，$P^{(j-1)}(A_i)$ は受検者 $j-1$ までの項目 i の暴露率である．

（3） 適格確率 $P^{(j)}(E_i)$ に従って，適格でない項目のみアイテムバンクから除く．

（4） アイテムバンクから情報量が最大の項目を出題する．

（5） 反応データから能力推定値 $\hat{\theta}$ を更新する．

（6） 能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（3），（4）を繰り返す．

ただし，最初の受検者 ($j = 1$) または $P^{(j-1)}(A_i) = 0$ の場合には $P^{(j)}(E_i) = 1$ とする．テスト終了後には全ての項目をアイテムバンクに戻す．

3.4 Kingsbury and Zara (1989) の適応型テスト (KZ)

Kingsbury and Zara はアイテムバンクを分割する手法を提案した（以降，KZ と呼ぶ）[14]．具体的には，次のアルゴリズムに従って項目出題する．

（1） アイテムバンクをランダムに分割し，項目集合を複数構成する．

（2） 能力推定値を $\hat{\theta} = 0$ に初期化する．

（3） 暴露数が最小の項目集合を選択し，その項目集合から情報量が最大の項目を受検者に出題する．

（4） 反応データから能力推定値 $\hat{\theta}$ を更新する．

（5） 能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（3），（4）を繰り返す．

KZ は，特定の項目群の過度な暴露を抑えることができた．しかし，各項目集合は測定精度が等質でないため，受検者間でテストの長さや測定精度にばらつきが生じる．

3.5 等質適応型テスト (UAT)

適応型テストの等質性の問題を解決するため，宮澤・植野は等質適応型テストを提案した [15], [16] (以降，UAT と呼ぶ)．UAT では，事前に等質テスト構成手法により生成した等質アイテムバンクを用いる．その等

質アイテムバンクの生成には，当時世界最大数の等質テストを構成できる Ishii et al. (2014) の手法 [17] を用いている．具体的には，次のアルゴリズムに従って項目出題する．

（1） 各受検者に等質アイテムバンクをランダムに一つ割り当てる．

（2） 能力推定値を $\hat{\theta} = 0$ に初期化する．

（3） 割り当てられた等質アイテムバンクから情報量が最大の項目を受検者に出題する．

（4） 反応データから能力推定値 $\hat{\theta}$ を更新する．

（5） 能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（3），（4）を繰り返す．

UAT は，受検者間の測定精度を等しく保ちつつ，暴露数を減少できた．しかし，UAT は暴露数を減少できるが，アイテムバンクサイズを縮小するので測定精度が低下するというトレードオフの問題がある．

3.6 2 段階等質適応型テスト (TUAT)

暴露数の減少と測定精度の向上の両立のため，宮澤・植野は 2 段階等質適応型テストを提案している [23], [24]（以降，TUAT と呼ぶ）．TUAT では，事前に等質テスト構成手法により生成した等質アイテムバンクを用いる．等質アイテムバンクの生成には，当時世界最大数の等質テストを構成できる Ishii et al. (2017) の手法 [18] を用いている．1 段階目では，各受検者に等質アイテムバンクをランダムに一つ割り当て，適応的項目出題を行う．等質アイテムバンクに含まれる項目の難易度パラメータは疎であるが均一に分布しているため，推定値が真の能力値近傍に速く到達する．能力推定値が収束したとき，1 段階目を終了する．2 段階目ではアイテムバンク全体から適応的項目出題を行う．これにより，高精度の能力推定値を得られる．この結果，従来型適応型テストと同等の測定精度を保ちつつ，従来手法よりも暴露数を減少できた．

3.7 TUAT の問題

しかし，TUAT は識別力パラメータの大きい項目に偏って暴露されるという問題がある．次の図 1 は項目の識別力と TUAT の実験における暴露数の散布図である．図 1 より，識別力の高い項目が過度に暴露されており，暴露数に偏りが生じている．宮澤・植野 [23], [24] は，1 段階目で推定値が能力真値の近傍に収束するため，2 段階目ではその能力推定値近傍の難易度をもつ項目のみが出題され，暴露数の偏りは軽減できると仮定している．この仮定は，一つの項目については，その難易度に一致する能力値のフィッシャー情報量が最大

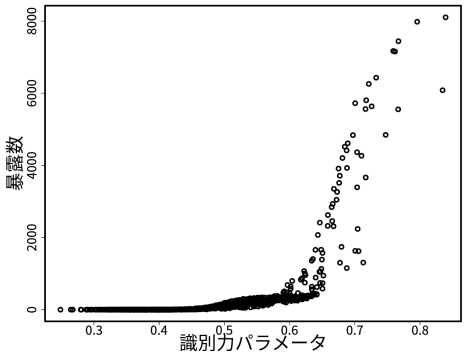


図 1 暴露数と識別力パラメータの関係

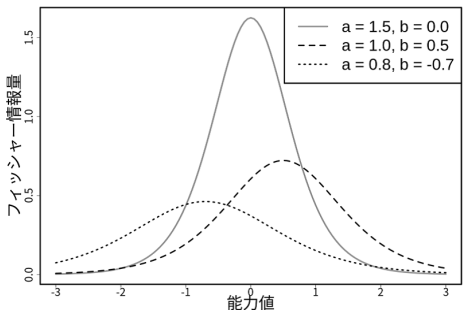


図 2 能力値とフィッシャー情報量の関係

となることを根拠としている．しかし，実際は項目の難易度が能力推定値と乖離していても，乖離していない場合に比較して，情報量が大きくなる場合が多くある．図 2 は，ある三つの項目の能力値に対するフィッシャー情報量を示した曲線である．各項目は，難易度 b と一致する能力値に対して最大のフィッシャー情報量を示す．識別力 $a = 1.5$，難易度 $b = 0.0$ である項目は，難易度に一致する能力値 $\theta = 0.0$ に対して最大値をとる．更にこの項目は，他の項目の難易度に一致する能力値 $\theta = 0.5, -0.7$ の受検者に対しても，3 項目の中で最大のフィッシャー情報量を示す．このように，他項目と比較して相対的に識別力の大きな項目は難易度が乖離していても出題される可能性が高い．図 3 は，TUAT の 2 段階目において出題された項目の難易度と能力推定値の差異を示した図である．ここで，2 段階目に出題された項目の難易度と能力推定値の差異は次式で定義される．

$$\sqrt{\frac{1}{n-l+1}\sum_{k=l}^{n}(\hat{\theta}_{k-1} - b_{i_k})^2} \tag{4}$$

また，k 問目に出題された項目を i_k，項目 i_k 出題後の

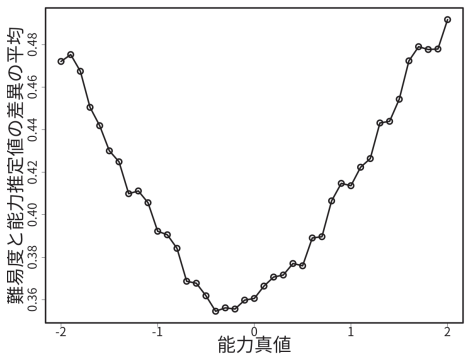


図 3　項目の難易度と能力推定値の差異

能力推定値を $\hat{\theta}_k$ とし，2 段階目で出題された項目は l 問目から n 問目とする．図 3 では，能力真値を −2.0 から 2.0 まで 0.1 刻みで離散化し，各区間の誤差の平均をプロットしている．図より，能力推定値と乖離した難易度をもつ項目が出題されていることがわかる．これは，識別力の高い項目が広範囲の能力値に対して有効であることが原因である．このように識別力の高い項目が多く出題されるため，TUAT の暴露数の偏りの改善は限定的である．

4. 提 案 手 法

TUAT の問題を解決するため，本研究では，2 段階目以降は能力推定値近傍の難易度パラメータをもつ項目に限定して出題する項目難易度制約付き等質適応型テストを提案する．

本手法では，TUAT 同様に事前に等質テスト構成技術により生成した等質アイテムバンクを用いる．その等質アイテムバンクの生成には，現在最も多くの等質テストを構成可能な Fuchimoto らの手法 [20] を用いる．この手法では，はじめに，時間計算量は小さいが空間計算量の大きい最大クリーク法 [17] により，メモリの限界まで多くの等質テストを構成する．次に，時間計算量は大きいが空間計算量が小さい整数計画法を用いた並列探索手法により，より多くの等質テストを生成する．提案手法では，TUAT 同様に離散化した能力値 $\theta_k = \{\theta_1, \theta_2, \ldots, \theta_K\}$ について，テスト情報量の下限と上限を次のように制限してテストを構成する．

$$m(\theta_k)n \leq I_i(\theta_k)z_i \leq (m(\theta_k) + sd(\theta_k))n \tag{5}$$

ここで，z_i は項目 i がテストに含まれるとき 1，含まれないとき 0 をとる．また，$m(\theta_k)$ は能力値 θ_k についてのアイテムバンクに含まれる項目の情報量の平均，$sd(\theta_k)$ は θ_k ついての標準偏差，n は等質アイテムバンクの項目数である．

更に，TUAT 同様に，1 段階目では各受検者に等質アイテムバンクをランダムに一つ割り当て，適応的項目出題を行う．等質アイテムバンクに含まれる項目の難易度パラメータは疎であるが均一に分布しているため，推定値が真の能力値近傍に速く到達する．能力推定値の更新幅がしきい値よりも小さくなったとき，1 段階目を終了する．2 段階目では，能力推定値近傍の難易度パラメータ区間に属する項目から適応的項目出題を行う．その難易度パラメータ区間は，次式に示す能力推定値の事後標準偏差に比例して決定する．

$$SD(\hat{\theta}) = \sqrt{\int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f(\theta|\mathbf{u})} \tag{6}$$

本手法では，次の区間を能力推定値近傍とした．

$$\hat{\theta} - \delta SD(\hat{\theta}) < \theta < \hat{\theta} + \delta SD(\hat{\theta}) \tag{7}$$

ただし，δ は難易度パラメータ区間に対する事後標準偏差の影響度合いを決定する重み付けチューニングパラメータである．能力推定値近傍の項目の難易度パラメータ区間は次のとおりである．

$$\hat{\theta} - \delta SD(\hat{\theta}) < b < \hat{\theta} + \delta SD(\hat{\theta}) \tag{8}$$

式 (8) は，前述の GPCM のような多段階反応を扱うモデルの複数の難易度パラメータについて工夫すれば適用できるかもしれない．しかし，その深い議論は本論文の主旨から外れるので，式 (8) の適用は項目単位で単一の難易度パラメータをもつ IRT モデルのみとし，GPCM のような多値を扱う IRT モデルには適用しない．結果として，次のアルゴリズムに従って項目出題する．

I　第 1 段階

（1）　各受検者に等質アイテムバンクをランダムに一つ割り当てる．

（2）　能力推定値を $\hat{\theta} = 0$ に初期化する．

（3）　割り当てられた等質アイテムバンクから情報量が最大の項目を出題する．

（4）　反応データから能力推定値 $\hat{\theta}$ を更新する．

（5）　能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（3），（4）を繰り返す．

能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になった後，

第 2 段階に移行する．

Ⅱ　第 2 段階

（1）　能力推定値 $\hat{\theta}$ と事後標準偏差 $SD(\hat{\theta})$ から難易度パラメータ区間を計算する．

（2）　暴露数の上限が設定されている場合は，暴露数が上限に達した項目をアイテムバンクから除く．

（3）　式 (8) に示した難易度パラメータ区間に属する項目から情報量が最大の項目を出題する．

（4）　反応データから能力推定値 $\hat{\theta}$ を更新する．

（5）　能力推定値 $\hat{\theta}$ の更新幅がしきい値 ϵ 未満になるまで手順（1）から（4）を繰り返す．

本手法により，2 段階目以降は能力推定値近傍の難易度パラメータをもつ項目に限定して出題でき，従来手法と比較して測定精度を同等に保ちつつ，暴露数の偏りを減少させることが期待される．

5. 評 価 実 験

本章では，提案手法の有効性を示すため，シミュレーションデータと実データを用いて従来手法と比較する．

5.1 ハイパーパラメータチューニング

本節では提案手法の 1 段階目から 2 段階目への切替条件である能力推定値の更新幅のしきい値パラメータ ϵ と，難易度パラメータ区間を決定するハイパーパラメータ δ について，複数候補値のうち測定精度を保ちつつ暴露数を軽減できるパラメータ値をそれぞれ選択する．

まず，切替条件を検討する．提案手法について，切替条件となるパラメータ ϵ を 0.025 から 0.5 までステップ数を 0.025 として暴露数の標準偏差と測定誤差を評価した結果を図 4 に示す．この実験では，各パラメータ a_i，b_i をそれぞれ $\log a_i \sim N(-0.5, 0.2)$，$b_i \sim N(0,1)$ から発生させた 1000 項目からなるシミュレーションアイテムバンクを用いた．測定誤差は次式により定義される能力真値と能力推定値の RMSE (Root Mean Square Error) を用いて評価した．

$$\sqrt{\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2} \tag{9}$$

図 4 は横軸が切替条件となるパラメータ ϵ，縦軸が暴露数の標準偏差及び RMSE を示す．図 4 より，切替条件となるパラメータ ϵ を大きくすると，暴露数の標準偏差が大きくなり，RMSE は小さくなる．このことから，暴露数の標準偏差と RMSE にはトレードオフの関係があることがわかる．これは TUAT と同様であるため，本実験では TUAT [23], [24] と同様に，RMSE の減少幅が 0.01 以下となったパラメータ ϵ を提案手法の切替条件とする．

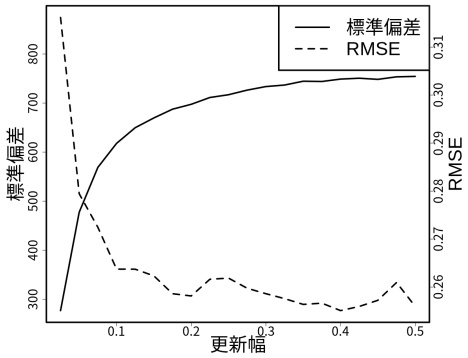


図 4　能力推定値の更新幅のしきい値パラメータ ϵ に対する暴露数の標準偏差と RMSE の変化

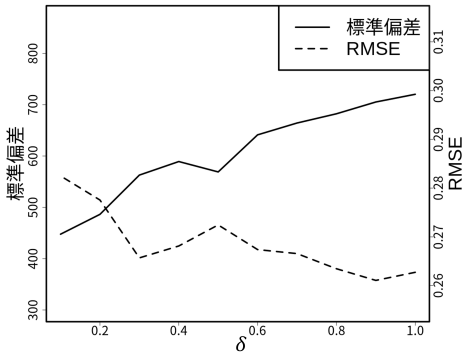


図 5　パラメータ δ に対する暴露数の標準偏差と RMSE の変化

次に，提案手法における難易度パラメータ区間を決定するパラメータ δ について検討する．パラメータ δ を 0.1 から 1.0 までステップ数を 0.1 として変化させ，暴露数と測定誤差を評価した結果を図 5 に示す．この実験は，切替条件であるパラメータ ϵ に関する評価実験と同様のアイテムバンクを用いた．測定誤差は能力真値と能力推定値の RMSE を用いて評価した．図 5 は横軸が δ，縦軸が暴露数の標準偏差及び RMSE を示す．

図 5 より，パラメータ δ を大きくすればするほど，暴露数の標準偏差が大きくなる．これは，難易度パラメータ区間が広くなることにより，広範囲の能力値に対して高い情報量をもつ識別力の高い項目が過度に暴露されるからである．一方，RMSE はパラメータ δ を大きくすればするほど小さくなるが，大きな変化はな

い．そこで，本実験では RMSE の小数点第三位を四捨五入した値が同一であるパラメータ δ の中で，暴露数の標準偏差が最小のものを採用する．

5.2 提案手法と TUAT の動作例の比較

本実験では受検者に対する提案手法と TUAT の動作例を比較する．図 6，7 はそれぞれ TUAT，提案手法の能力値 $\theta = 0.9$ をもつ受検者への 2 段階目における出題項目の難易度と能力推定値の推移を示した図である．ただし，図 6，7 中の緑色の直線は能力真値，図 7 中の青線は能力推定値とその標準偏差から計算した難易度パラメータ区間を示している．実験には，識別力パラメータ a_i，難易度パラメータ b_i をそれぞれ $\log a_i \sim N(-0.5, 0.2), b_i \sim N(0, 1)$ からサンプリングした 1000 項目からなるアイテムバンクを用いた．また，提案手法と TUAT のどちらの実験においても，第 1 段階では同一の等質アイテムバンクを割り当てた．図より，TUAT は能力推定値から乖離した難易度パラメータをもつ項目を出題していることがわかる．一方，提案手法は難易度パラメータ区間に限定することで，

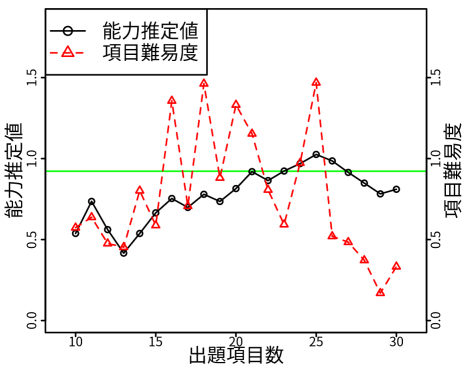


図 6　TUAT の動作例

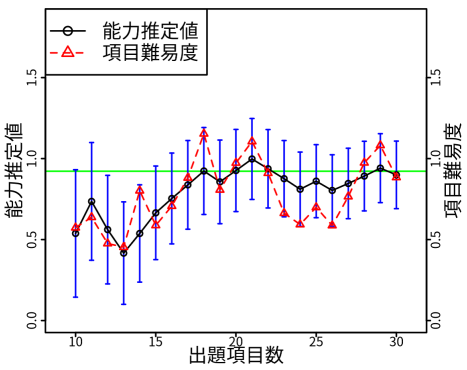


図 7　提案手法の動作例

TUAT よりも能力推定値との差異が小さい難易度パラメータをもつ項目を出題している．また，図からわかるように，提案手法の難易度パラメータ区間は能力推定値の標準偏差に比例するため，出題項目数の増加にともない小さくなる．これと同時に能力推定値と能力真値の誤差が小さくなるため，提案手法は能力真値近傍の難易度パラメータをもつ項目を出題でき，結果として能力推定精度を高めていることがわかる．

5.3 従来手法との比較実験

本実験では能力推定値に対して情報量最大の項目を逐次出題する従来型適応型テスト (以降，CAT と呼ぶ), IP [6], TI [29], Prob [10], [11], KZ [14], TUAT [23], [24] と比較する．実験の手順は以下のとおりである．

（1） 受検者の能力真値を $\theta \sim N(0, 1)$ からサンプリングする．

（2） 各手法のアルゴリズムにしたがい，項目を出題する．

（3） 能力真値と出題項目のパラメータを所与として，出題項目への反応データを生成する．

（4） 反応データを所与として，能力推定値 $\hat{\theta}$ を更新する．

（5） テスト終了条件を満たすまで手順（2）から（4）を繰り返す．

（6） 手順（1）から（5）を 10000 回繰り返す．出題項目と能力推定値から，各項目の暴露数と測定誤差 (能力真値と能力推定値の RMSE) を求める．

ただし，先行研究 [23], [24] では，暴露数が 1 以上の項目に限定して暴露数の平均値と標準偏差を計算している．本論文では，全ての項目に対して暴露数の標準偏差を計算する．実験には二つのシミュレーションアイテムバンクと，二つの実データを用いた．シミュレーションアイテムバンクは，2PLM を用いて，次の分布から各項目のパラメータをサンプリングし，1000 項目からなるアイテムバンクを生成した．

- $\log a_i \sim N(-0.5, 0.2), b_i \sim N(0, 1)$ (以降，このアイテムバンクを simu1 と呼ぶ)
- $\log a_i \sim N(-0.75, 0.2), b_i \sim N(0, 1)$ (以降，このアイテムバンクを simu2 と呼ぶ)

実データは，リクルート（株）が開発した SPI [4] のアイテムバンクと適応型テストの先行研究 [30] において使用されているアイテムバンク Science を用いた．ただし，SPI のアイテムバンクは 2PLM を用いた 978 項目から構成される．アイテムバンク Science は

GPCM を用いた 82 項目と 3PLM を用いた 918 項目から構成される．ただし，GPCM には項目単位で単一の難易度パラメータがないため，提案手法の難易度パラメータによる制約を用いることができない．そのため，提案手法は 1 段階目終了後全ての GPCM の項目を出題可能とした．

全ての実験において，暴露数の制約を設けた場合と設けない場合の 2 パターンについて実験を行った．制約を設ける実験では，各項目の暴露数の上限を 5000 とした．ただし，Prob は最大暴露数が 5000 に漸近的に一致するよう最大暴露率を 0.5 とした．テスト終了条件は先行研究 [23], [24] と同様にテストの長さとし，本実験では 30 項目とした．また，KZ，TUAT 及び提案手法について，等質アイテムバンクに含まれる項目数はテストの長さと同一とした．

表 1 は暴露数の上限を設けない場合の実験結果である．ただし，手法の横の括弧に TUAT は切替条件を，提案手法は切替条件である ϵ と δ の値を示した．

結果から，全てのデータセットにおいて TI の未出題項目数が最も小さいことが分かる．しかし，TI は暴露数の標準偏差が大きく，未出題項目数が他手法よりも大きい CAT，IP と近い値をとる．更に，CAT，IP 及び TI は暴露数の最大値と受検者数が一致するため，全ての受検者に出題された項目が存在する．このように，過度に暴露された項目はテストの信頼性の低下につながる危険性がある．

一方，提案手法は Science を除くデータセットにおいて，暴露数の最大値及び標準偏差を全ての手法の中で最小に抑えられた．特に，提案手法は Science を除くデータセットにおいて，暴露数の標準偏差及び最大値，未出題項目数を TUAT よりも小さくできた．また，提案手法は SPI では測定精度が TUAT よりもわずかに劣るが，simu1，simu2，Science では同等に保っている．

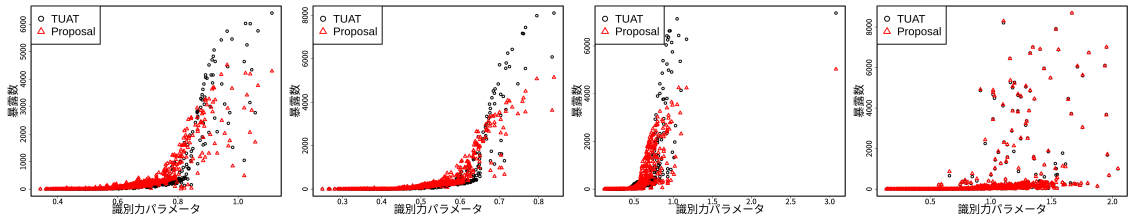
表 1　実 験 結 果

アイテムバンク	手法	暴露数の標準偏差	暴露数の最大値	未出題項目数	RMSE
simu1	CAT	1055.5	10000	832	**0.25**
	IP	1057.9	10000	833	**0.25**
	TI	1000.4	10000	**0**	0.26
	KZ	918.0	6565	779	0.26
	TUAT(0.100)	864.7	6409	188	0.26
	Proposal(0.100，0.8)	**682.3**	**4520**	68	0.26
simu2	CAT	1167.6	10000	860	**0.32**
	IP	1162.9	10000	861	**0.32**
	TI	1084.3	10000	**0**	**0.32**
	KZ	1094.9	7816	829	**0.32**
	TUAT(0.100)	932.8	8104	251	0.33
	Proposal(0.100，0.7)	**702.8**	**5145**	128	0.33
SPI	CAT	1150.3	10000	836	**0.25**
	IP	1149.2	10000	836	**0.25**
	TI	1047.6	10000	**7**	0.26
	KZ	1032.0	7364	792	0.26
	TUAT(0.075)	937.6	7381	274	0.26
	Proposal(0.100，0.6)	**672.7**	**5031**	263	0.27
Science	CAT	1229.3	10000	884	**0.20**
	IP	1230.3	10000	884	**0.20**
	TI	1154.4	10000	**0**	**0.20**
	KZ	1151.7	8892	843	0.22
	TUAT(0.075)	**1063.6**	**8687**	303	0.21
	Proposal(0.075,1.0)	1066.8	8690	307	0.21

次に，TUAT と提案手法において暴露された項目の特性を分析する．図 8 (a)，8 (b)，8 (c)，8 (d) はそれぞれ simu1，simu2，SPI，Science の実験における項目の識別力パラメータと暴露数の散布図である．これらより，Science を除くデータセットにおいて，TUAT は識別力パラメータが大きい項目を過度に暴露していることがわかる．一方，提案手法は難易度パラメータ区間に属する項目に限定して出題することで，TUAT で過度に暴露されていた識別力パラメータが大きい項目の暴露数を抑えることができている．

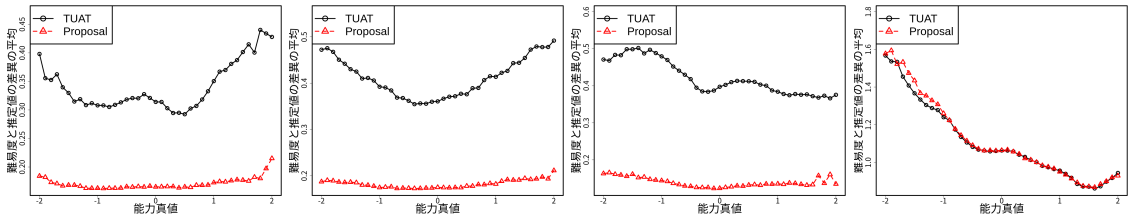
最後に，TUAT と提案手法において 2 段階目に出題された項目の難易度と能力推定値の差異を比較する．図 9 (a)，9 (b)，9 (c)，9 (d) はそれぞれ能力真値を −2.0 から 2.0 まで 0.1 刻みで離散化し，各区間の項目の難易度と能力推定値の差異の平均をプロットしている．これらの図から，Science を除くデータセットにおいて，提案手法は 2 段階目以降に出題した項目の難易度と能力推定値の差異が TUAT よりも小さい．前述のように，TUAT では能力推定値と難易度が乖離しても，広範囲の能力値に対して高いフィッシャー情報量をもつ識別力の高い項目が出題される傾向にあり，暴露数の偏りの原因と考えられている．提案手法は，1 段階目で受検者の能力値近傍を推定し，2 段階目で能力値近傍にある難易度をもつ項目に限定して出題することで，この問題を緩和している．

一方, Science のデータセットでは, 提案手法と TUAT の暴露数の標準偏差，最大値及び未出題項目数が同等となった．これは，提案手法が難易度の制約を用いることができない大問に対応する GPCM の項目が過度に暴露されたことが原因である．表 2 は，Science に含まれる項目のモデル別の暴露数である．本研究では，前述のように GPCM の項目に式 (8) の難易度の制約を課していないので，GPCM の項目の暴露数は TUAT と同等である．一方，3PLM の項目について，提案手法は暴露数の標準偏差及び最大値を TUAT よりも減少できた．図 10 は 3PLM の項目の暴露数と識別力パラメータの散布図，図 11 は 3PLM の項目の難易度と能



(a) simu1　(b) simu2　(c) SPI　(d) Science

図 8　識別力パラメータと暴露数の関係



(a) simu1　(b) simu2　(c) SPI　(d) Science

図 9　項目の難易度と能力推定値の差異

表 2　Science における項目のモデル別の暴露数

手法	モデル	暴露数の標準偏差	暴露数の最大値	未出題項目数
TUAT(0.075)	3PL	126.5	1175	303
	GPCM	725.4	8687	0
Proposal(0.075,1.0)	3PL	107.3	1134	307
	GPCM	723.4	8690	0

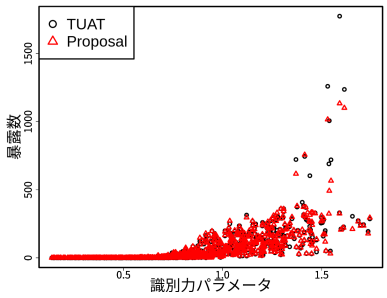


図 10　Science に含まれる 3PLM の項目の暴露数と識別力パラメータ

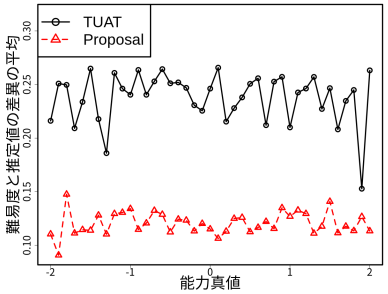


図 11　Science に含まれる 3PLM の項目の難易度と能力推定値の差異

力推定値の差異である．これらの図より，提案手法は 3PLM の項目について，能力推定値近傍の難易度パラメータをもつ項目に限定して出題することで，識別力パラメータが大きい項目の暴露を抑えたことがわかる．

暴露数の制約を設けた場合の実験結果を表 3 に示す．全てのデータセットにおいて，暴露数の制約を設けた場合，制約のない場合と比較して精度は同等または低下している．これは，識別力パラメータの大きい項目が暴露数の制約のために，過度には出題されていないことが原因である．このように，暴露数の減少と測定精度の向上にはトレードオフの関係がある．

全てのデータセットで最も精度が良いのは CAT 及び IP である．しかし，CAT 及び IP は暴露数の標準偏差及び未出題項目数が大きい．TI は全てのデータセットで未出題項目数が最小の値をとるが，暴露数の標準偏差が大きい．一方，Science を除くデータセットにおいて，提案手法はわずかに精度が劣るものの，暴露数の標準偏差及び最大値を全ての手法の中で最小に抑えられた．Science のデータセットでは，提案手法による難易度の制約を用いることができない GPCM の大問項目が過度に暴露されたことにより，TUAT と提案手法は暴露数の標準偏差，最大値及び未出題項目数

表 3　暴露数の制約を設けた実験結果

アイテムバンク	手法	暴露数の標準偏差	暴露数の最大値	未出題項目数	RMSE
simu1	CAT	983.8	5000	813	**0.25**
	IP	984.0	5000	812	**0.25**
	TI	941.4	5000	**0**	**0.25**
	Prob	987.8	5105	819	**0.25**
	KZ	897.7	5000	787	0.26
	TUAT(0.100)	823.6	5035	198	0.27
	Proposal(0.100,0.8)	**682.3**	**4520**	68	0.26
simu2	CAT	1018.3	5000	830	**0.32**
	IP	1018.3	5000	828	**0.32**
	TI	969.5	5000	**0**	0.33
	Prob	1027.8	5108	834	**0.32**
	KZ	955.9	5000	803	**0.32**
	TUAT(0.075)	769.0	5113	148	0.35
	Proposal(0.100,0.6)	**684.4**	**4911**	103	0.33
SPI	CAT	1026.3	5000	809	**0.26**
	IP	1026.4	5000	809	**0.26**
	TI	969.8	5000	**33**	**0.26**
	Prob	1034.8	5107	812	**0.26**
	KZ	968.6	5000	775	0.27
	TUAT(0.075)	851.8	5141	262	0.29
	Proposal(0.100,0.5)	**649.2**	**4627**	230	0.27
Science	CAT	1044.3	**5000**	838	**0.21**
	IP	1044.1	**5000**	842	**0.21**
	TI	1009.4	**5000**	**0**	**0.21**
	Prob	1053.4	5110	837	**0.21**
	KZ	954.7	**5000**	806	0.23
	TUAT(0.050)	**872.2**	5195	180	0.24
	Proposal(0.050,0.9)	892.4	5238	176	0.22

が近い値をとっている．

暴露数の制御により，一部のデータセットにおいてTUAT は測定精度が大きく悪化した．これは，TUATが 2 段階目にアイテムバンク全体から項目出題するため，識別力の高い項目が暴露数の制約により，出題できなくなったことが原因である．一方，提案手法は能力推定値近傍の難易度パラメータをもつ項目に限定して出題することで，能力真値ごとに異なる項目を出題できる．これにより，暴露数を一様にしながら減少させることができた．

6. むすび

本研究では，2 段階等質適応型テストの問題を解決するため，能力推定値近傍の難易度パラメータをもつ項目に限定して出題する項目難易度制約付き等質適応型テストを提案した．シミュレーションデータと実データを用いた実験により，提案手法は従来手法と比較して測定精度を同等に保ちつつ，暴露数の標準偏差及び最大値を減少できた．

しかし，本手法の適用は難易度パラメータをもつIRT モデルに限定される．項目単位で単一の難易度パラメータをもたない GPCM の大問項目を含む Science のデータセットでは暴露数の偏りを抑えることができなかった．今後，GPCM にも適用できる手法を考えたい．

謝辞　本研究は JSPS 科研費 24H00739，24KJ1124 の助成を受けた．

文　献

[1] M. Ueno, K. Fuchimoto, and E. Tsutsumi, “e-testing from artificial intelligence approach,” Behaviormetrika, vol.48, no.2, pp.409–424, 2021.

[2] M. Ueno, “Ai based e-testing as a common yardstick for measuring human abilities,” 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE) IEEE, pp.1–5, 2021.

[3] W.D. Way, “Protecting the integrity of computerized testing item pools,” Educational Measurement: Issues and Practice, vol.17, pp.17–27, 1998.

[4] Recruit, “Synthetic Personality Inventory”. https://www.recruit-ms.co.jp/freshers/spi-001.html.（最終閲覧日：2024/08/02）

[5] 株式会社ベネッセコーポレーション，Global Test of English Communication, 2014. https://www.benesse.co.jp/gtec/.（最終閲覧日：2024/08/02）

[6] W.J. van der Linden and L.M. Reese, “A model for optimal constrained adaptive testing,” Applied Psychological Measurement, vol.22, no.3, pp.259–270, 1998.

[7] R.D. Hetter and J.B. Sympson, Item exposure control in CAT-ASVAB., American Psychological Association, 1997.

[8] M.L. Stocking and C. Lewis, “Controlling item exposure conditional on ability in computerized adaptive testing,” J. Educational and Behavioral Statistics, vol.23, no.1, pp.57–75, 1998.

[9] M.L. Stocking and C. Lewis, Methods of controlling the exposure of items in CAT, Springer, 2000.

[10] W.J. van der Linden and B.P. Veldkamp, “Constraining item exposure in computerized adaptive testing with shadow tests,” J. Educational and Behavioral Statistics, vol.29, no.3, pp.273–291, 2004.

[11] W.J. van der Linden and C.A. Glas, Elements of adaptive testing, Springer, 2010.

[12] L. Swanson and M.L. Stocking, “A model and heuristic for solving very large item selection problems,” Applied Psychological Measurement, vol.17, no.2, pp.151–166, 1993.

[13] Y. Cheng and H. Chang, “The maximum priority index method for severely constrained item selection in computerized adaptive testing,” British Journal of Mathematical and Statistical Psychology, vol.62, pp.369–383, 2009.

[14] G.G. Kingsbury and A.R. Zara, “Procedures for selecting items for computerized adaptive tests,” Applied Measurement in Education, vol.2, no.4, pp.359–375, Oct. 1989.

[15] 宮澤芳光，宇都雅輝，石井隆稔，植野真臣，“測定精度の偏り軽減のための等質適応型テストの提案，” 信学論（D），vol.J101-D, no.6, pp.909–920, June 2018.

[16] M. Ueno and Y. Miyazawa, “Uniform adaptive testing using maximum clique algorithm,” Artificial Intelligence in Education, pp.482–493, 2019.

[17] T. Ishii, P. Songmuang, and M. Ueno, “Maximum clique algorithm and its approximation for uniform test form assembly,” IEEE Trans. Learning Technologies, vol.7, no.1, pp.83–95, 2014.

[18] T. Ishii and M. Ueno, “Algorithm for uniform test assembly using a maximum clique problem and integer programming,” Artificial Intelligence in Education, pp.102–112, 2017.

[19] 渕本壱真，植野真臣，“等質テスト構成における整数計画法を用いた最大クリーク探索の並列化，”信学論（D），vol.J103-D, no.12, pp.881–893, Dec. 2020.
[20] K. Fuchimoto, T. Ishii, and M. Ueno, “Hybrid maximum clique algorithm using parallel integer programming for uniform test assembly,” IEEE Trans. Learning Technologies, vol.15, no.2, pp.252–264, 2022.
[21] 谷澤明紀，本多康弘，“情報処理技術者試験における e テスティング，”日本テスト学会第 12 回大会発表論文抄録集，vol.33, no.2, pp.54–57, 2014.
[22] 仁田善雄，斎藤宣彦，後藤英司，高木　康，石田達樹，江藤一洋，“医療系大学間共用試験における e テスティング，”日本テスト学会第 12 回大会発表論文抄録集，pp.58–59, 2014.
[23] 宮澤芳光，植野真臣，“高精度能力推定を保証する 2 段階等質適応型テスト，”信学論（D），vol.J106-D, no.1, pp.34–46, Jan. 2023.
[24] M. Ueno and Y. Miyazawa, “Two-Stage uniform adaptive testing to balance measurement accuracy and item exposure,” Artificial Intelligence in Education, pp.626–632, 2022.
[25] F.M. Lord and M.R. Novick, Statistical Theories of Mental Test Scores, Addison-Wesley, 1968.
[26] F.M. Lord, Applications of item response theory to practical testing problems, Lawrence Erlbaum Associates, London, England, 1980.
[27] 植野真臣，永岡慶三，e テスティング，培風館，2009.
[28] R.D. Bock and R.J. Mislevy, “Adaptive EAP estimation of ability in a microcomputer environment,” Applied Psychological Measurement, vol.6, no.4, pp.431–444, 1982.
[29] S.W. Choi and S. Lim, “Adaptive test assembly with a mix of set-based and discrete items,” Behaviormetrika, vol.49, pp.231–254, 2022.
[30] S.W. Choi, S. Lim, and W.J. van der Linden, “TestDesign: an optimal test design approach to constructing fixed and adaptive tests in R,” Behaviormetrika, vol.49, no.2, pp.191–229, 2022.

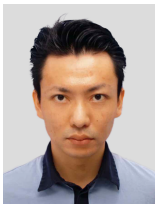
（2023 年 3 月 8 日受付，2024 年 3 月 11 日再受付，
6 月 26 日早期公開）

岸田　若葉　（学生員）

2023 電気通信大学情報理工学域卒．同年，同大学院情報理工学研究科情報・ネットワーク工学専攻博士前期課程入学．現在に至る．e テスティング，適応型テストの研究・開発に従事．

渕本　壱真

2020 電気通信大学情報理工学域・先端工学基礎課程卒．同年，同大学院情報理工学研究科情報・ネットワーク工学専攻博士前期課程入学，2022 同課程前期終了．同年，同専攻博士課程後期入学．現在に至る．e テスティング，自動テスト構成，適応型テストの研究・開発に従事．



宮澤　芳光　（正員）

2014 電気通信大学大学院情報システム学研究科博士後期課程了．博士（工学）．長岡技術科学大学，東京学芸大学を経て，2019 より大学入試センター助教に着任，現在に至る．e テスティングの研究・開発に従事．

植野　真臣　（正員）

1992 神戸大学院教育学研究科修士課程了，1994 東京工業大学大学院総合理工学研究科博士課程了．博士（工学）．東京工業大学，千葉大学，長岡技術科学大学を経て 2006 より電気通信大学勤務，2013 教授．2016 電気通信大学大学院情報理工学研究科教授，現在に至る．人工知能，e テスティング，e ラーニング，ベイズ統計の研究に従事．2008 The 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2008) Best Paper Award. 2017 本会論文賞．