

2024年8月 31 日

国際標準を満たす CBT 運用・実施のための ガイドライン Ver.2

植野真臣 電気通信大学

このガイドラインは、2024 年度日本学術振興会科学研究費、基盤研究(S)「信頼性向上を持続する e テスティング・プラットフォームの開発」課題番号 19H05663 で開発された Computer based Testing (CBT) の基幹技術の利用を前提に、2007 年に第 1 版として発行された ISO/IEC 23988

「Information technology A code of practice for the use of information technology (IT) in the delivery of assessments (IDT) 」を満たす CBT 運営方法について推奨事項を示す。

このガイドラインでは、1) 試験結果の重要性が高い試験（ハイステークステスト）、2) 学力調査や学習途中の形成的評価など重要性が低い試験（ローステークス テスト）の二つを対象とする。

1. CBT システムの概要

テストをコンピュータ上で実施する CBT (Computer Based Testing) が普及している。CBT は、単にペーパーテストをコンピュータに置き換えただけのものではなく、ペーパーテストでは実現できないさまざまな活用ができる。例えば、CBT の利点として、以下があげられる (植野・永岡, 2009)。

- ① ペーパーテストに比較し、テストの印刷、運搬のコストが減少する。
- ② 大量のペーパーテストの束を管理することに比較し、電子ファイル上のテストは管理が容易でセキュリティも高い。
- ③ テスト実施に際して、テストの配布、回答の回収が自動的に行える。
- ④ テスト採点が、自動化もしくは半自動化される。最新の人工知能技術を用いれば、小論文方式のテストでさえ、自動採点できる。さらに即時的にテスト結果をフィードバックできる。
- ⑤ マルチメディア技術を用いることにより、これまで測定することのできなかった能力を測定できる問題項目を構築、実施できる。
- ⑥ 回答所要時間や回答変更回数、回答見直し回数などペーパーテ

ストでは測定することができなかったデータを得ることができるようになり、問題項目評価の多様性をもたらし、補助的に用いることにより測定精度を向上させることができる。

- ⑦ 障害をもつ受検者への対策が容易に行える。例えば、字を大きくしたり、ヘッドフォンをつけさせて音を大きくしたり、特別な入力装置（手以外で入力できる入力装置）、出力装置（点字など）を用いることにより障害者の受検を可能にすることができる。
- ⑧ 複数のカメラを用いる不正を監視するセキュリティーシステムを用いることにより、遠隔地にいる受検者の評価を行うことができる。セキュリティーシステムを用いれば自宅での受検も理論的には可能となる。
- ⑨ 遠隔地に分散するテスト作成者が、共同で問題（アイテム）バンク（問題項目のデータベース）を構築し、テスト構成を行うことができる。

CBT システムの概要を図 1 に示す。国際標準を満たすためには問題項目のデータベースである問題（アイテム）バンク方式である必要がある。大量の問題項目とその属性、統計データが格納され、問題項目は繰り返し替えし出題される。そのために問題項目は非公開とな

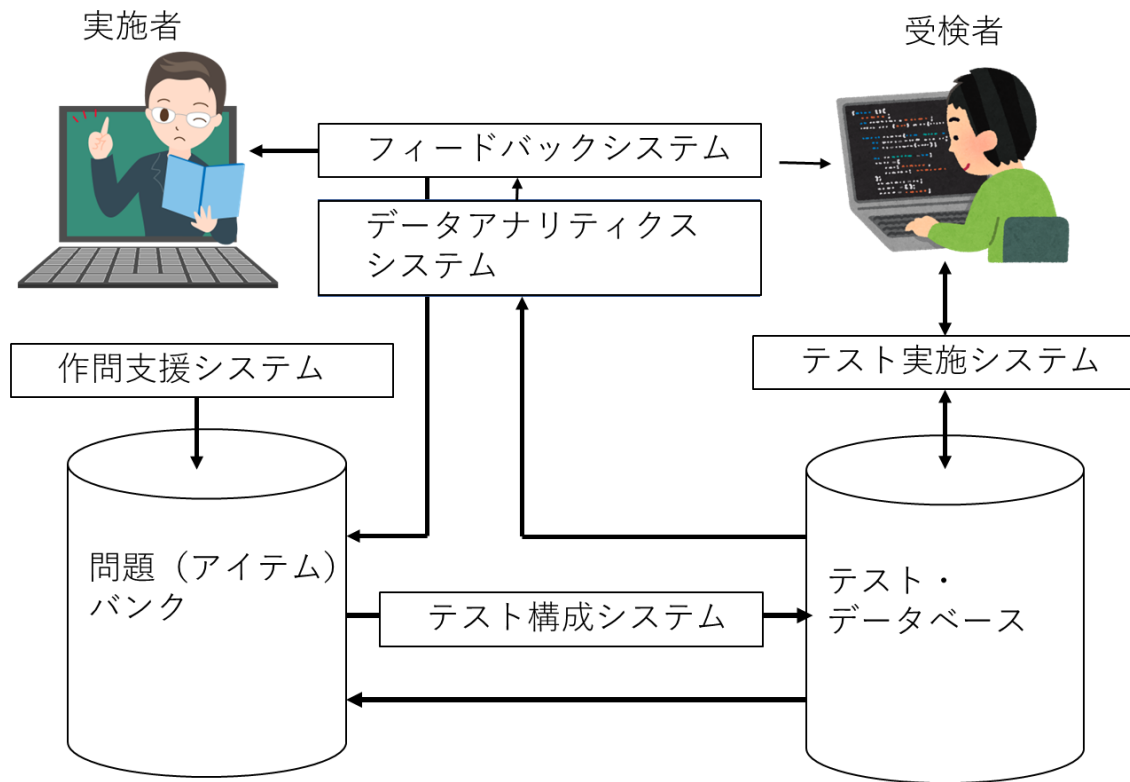


図 1. CBT システムの概要（植野、永岡 2009）

るだけでなく、受検者は問題項目の守秘義務に同意しないと受検できない。問題バンクからテスト構成システムを用いて、異なる問題項目からなるが等質のテスト（並行テスト）が生成される。

2. 問題項目の作問

作問者は作問支援システムにより問題項目を作問し、問題バンクに格納する。問題バンクでは以下のような仕様が望ましい。

- テストが測定したい能力を定義し、その能力を構成する科目・分野・領域・ブルームのタキソノミー（知識・理解・応用・分析・評価・創造）を定義した細目表を作成することが望ましい。
- 1EdTech: <https://www.1edtech.org/>が定める国際技術標準規格

表 1 QTI での問題項目出題形式例

名称	説明
単純型	
choiceInteraction	多肢選択
orderInteraction	並べ替え
associateInteraction	一群の選択肢から関連するものを選択
matchInteraction	二群の選択肢から関連するものを選択
gapMatchInteraction	選択肢による穴埋め
テキスト型	
inlineChoiceInteraction	プルダウン型の選択肢の選択
textEntryInteraction	自由記述の穴埋め
extendedTextInteraction	自由記述
hottextInteraction	テキスト内の選択肢の選択
	グラフィック型
hotspotInteraction	図の中の位置の選択. 選択肢表示あり
selectPointInteraction	図の中の位置の選択. 選択肢表示なし
graphicOrderInteraction	図の中の位置の順序付け
graphicAssociateInteraction	図の中の位置同士の対応付け
graphicGapMatchInteraction	図の中の位置と選択肢の対応付け
positionObjectInteraction	位置指定オブジェクトによる位置の選択.
その他	
sliderInteraction	スライダーによる数値の指定
mediaInteraction	音声・動画などの再生位置の指定
drawingInteraction	描画ツールによる画像などの入力
uploadInteraction	ファイルのアップロード
customInteraction	拡張用
endAttemptInteraction	受検者によるテストの終了

QTI (Question & Test Interoperability) に従った CBT システムを用い、問題項目を作成、テスト実施することが望ましい。QTI での問題項目出題形式例を表 1 に示した。

- **PCI**(Portable Custom Interaction)

ポータブル・カスタム・インタラクション (PCI) は、特定のシナリオのために開発されたインタラクションで、カスタム・インタラクションを定義し、QTI に準じた CBT システムの機能を拡張することができる。PCI は CBT システムの特定のバージョンに依存することなく動作するように開発されていなければならない。

図 2 は QTI に準拠した仕様の教科「情報」におけるプログラミングについての項目例（独立行政法人大学入試センター， 2022）である。この項目では、画面左に用意されたプログラムコードを並べ替えて所望のプログラムを右に構成することが求められる。これらの項目では、プログラムを実際に実行しながら解くことができ、プログラムの実行と解答も保存される。従来のペーパーテストでは測定できないような知識や能力を評価できることがわかる。現在、電気通信大学が開発を引き継ぎ、より高速にロバストに動作するように改良し、回答プロセス履歴も保存できるようになっている。

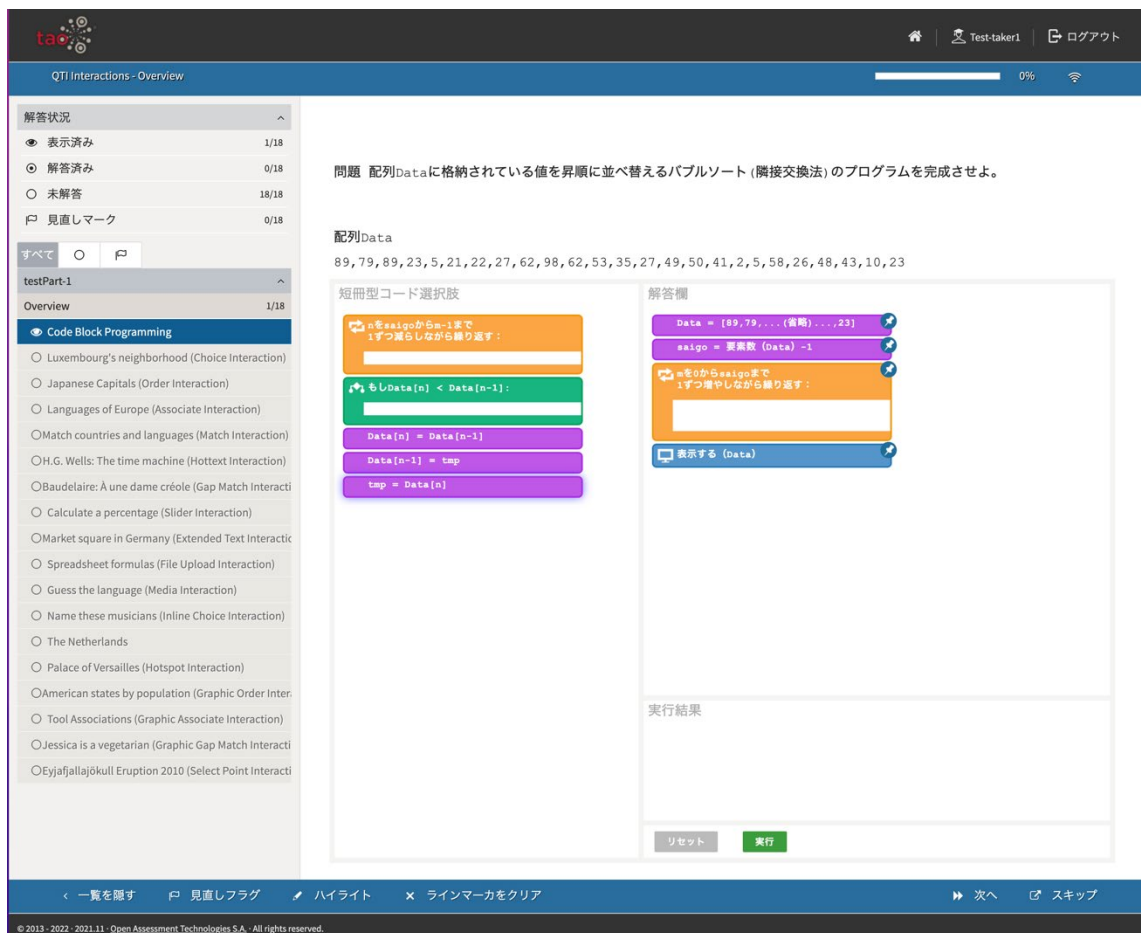


図 2. CBT における情報のプログラミング問題の PCI

- 作問者のための作問マニュアルを作成することが望ましい。
- 作問マニュアルでは、以下のような事項について記述しなければならない。以下は電気通信大学の CBT 作問で用いているものである。

■ 作問者の規定

- 以下の全てに該当する者とする

① 大学、大学共同利用機関法人又は独立行政法人（以下「大学等」という。）の教授若しくは准教授又は学識経験のある者であること。

② 高等学校等の教育課程及び問題バンク方式の CBT を使った入試に理解があること。

● 前項に規定する者のほか、地方教育行政の組織及び運営に関する法律（昭和 31 年法律第 162 号）第 18 条に規定する指導主事その他の高等学校等教育関係者で、問題バンク方式の CBT を使った入試に理解と認識を有する者とする。

■ 作問に関する秘密保持と著作権については別紙、誓約書の記載内容の通りである。

秘密保持に関する誓約書

年 月 日

国立大学法人 電気通信大学
学長殿

プロジェクト名 令和4年度 大学入学者選抜改革推進委託事業
個別大学の入学者選抜等における CBT の活用

現住所 _____
氏 名 _____
生年月日 _____年____月____日生

私は、上記プロジェクト(以下「本プロジェクト」という。)に参画するにあたり、下記事項を遵守することを誓約いたします。

記

第1条(秘密保持の誓約)

大学の許可なく、本プロジェクトに関して大学が秘密情報として指定した情報(以下「対象秘密情報」という。)を、本プロジェクトの参画者以外の者に対し開示し、又は本プロジェクト遂行の目的以外に使用しないことを約束いたします。なお、委員及び臨時委員(以下「委員等」という。)の氏名は、委員会以外で秘匿するものとする。

第2条(プロジェクト終了後の秘密保持等)

1. 対象秘密情報を、公知になったものを除き、本件プロジェクト終了後(退職後も含む。)も、不正に開示又は不正に使用しないことを約束いたします。
2. 本プロジェクトを終了するとき、本プロジェクトを担当しなくなったとき、又は大学による要求があるときには、対象秘密情報が記録等された大学の文書等(文書、図画、写真、USBメモリ、DVD、ハードディスクドライブその他の情報を記載又は記録するものをいう。以下同じ。)又は物件であって自己の保管するものを、遅滞なくすべて大学に返還し、その旨書面にて報告いたします。
3. 前項に定める場合において、対象秘密情報が自己の文書等に記録等されているときには、当該情報を消去するとともに、消去した旨(自己の文書等に対象秘密事項が記録等されていないときは、その旨)、書面にて報告いたします。

第3条(第三者に対する守秘義務の遵守)

第三者に対して守秘義務を負っている情報については、本プロジェクトにおいて知り得たかそれ以前から知っていたかにかかわらず、その守秘義務を遵守することを約束いたします。

第4条(情報の帰属)

本プロジェクトの業務の成果である情報は大学に帰属することを確認し、異議を述べません。

以上

■ 問題の受け渡し方法

- 作問者は、「【作問者】CBT 入試作問フォーマット」に従って問題情報や設問等を記載する。

- 作問者：作問者の氏名を記載する。
- 番号：当該作問者の作成した問題の通し番号を記載する。
- 作成日：問題を作成した日を記載する。
- 教科、分野、領域：教科ごとに事前に作成されている細目表ファイルに記載された内容から選択する。
- 測定する知識/技能、測定する思考力・判断力・表現力：細目表に記載された内容から選択する。
- 問題タイプ：別紙ファイルの問題タイプ一覧に記載された内容から選択する。
- 種問題番号：作問では、特定の問題から派生させ、新しい問題を作成することがある。この場合、派生元の問題を種問題と呼ぶ。派生された問題には、種問題の番号を記載する。
- 敵対問題番号：同一の試験に入れてはいけない問題番号の集合を指す。ある問題のヒントになるような情報を含む問題は

敵対問題となる。

- 想定難易度：作問者が想定する難易度であり、5段階として1が易しい問題であり、5が難しい問題とする。
- 想定回答所要時間：作問者が想定する回答所要時間を記載する。
- キーワード：問題に関連するキーワード。キーワードは複数入力できる。
- 問題確認者：問題を確認した人の氏名を記載する。
- 設問：問題の設問を記載する。句読点については、文化庁の『文化審議会建議「公用文作成の考え方」について』に従い、句点に「.」、読点に「、」を用いる。
(文化庁、文化審議会建議「公用文作成の考え方」について、令和4年1月7日、https://www.bunka.go.jp/koho_hodo_oshirase/hodohappyo/93650001.html)
- 選択肢：問題が選択肢から選ぶ形式である場合、選択肢を記述する。
- 正答：問題の正答を記載する。
- 正答の解説：正答の解説を記述する。ただし、必須項目ではない。

■ 作問での注意

- 作問者、問題確認者は問題項目が測定しようとする能力が定義されている細目標に対応しているかを内容的に吟味し、十分な妥当性を持たせなければならない。
- CBT 入試作問では、特定の教科書や参考書、問題集を学習した受検者が有利になるような問題項目を作成してはいけない。
一般的で基礎的な問題項目の構造は既存の出版物に掲載されているものと同じでもよいが、適宜問題設定などを変更し、表現も変えて作問することが望ましい。
- CBT 入試作問フォーマットへの記入では、試験問題・正答・属性情報の作成時のチェック表に書かれた以下の項目に留意しなければならない。

【問題内容】

- ✓ 難易度は適当か
- ✓ 募集要項に示した出題範囲からの出題となっているか
- ✓ 問題内容の妥当性は適切か

【問題表記】

- ✓ 問題の表記に間違いはないか
- ✓ 特に以下の点について、チェックしてください

- ・誤字、脱字、スペルミス、数字・年号・単位の使用方法や表記

- ・別ページ等にヒントや答えになる部分はないか、択一問題の場合、複数正解はないか、また、同一の選択肢はないか

- ・句点に「。」、読点に「、」を用いているか

【図・表】

- ✓ 図・表はきちんとできているか

【正答】

- ✓ 出題者以外が問題を解き、正答に間違いがないことを確認したか

- ✓ 想定される正答はすべて正しく採点されるように設定されているか

【問題の属性情報】

- ✓ 問題の属性情報(CBT 入試作問フォーマットの「問題情報」)に間違いはないか.

- ✓ 特に以下の点についてチェックしてください.

- ・問題内容は 教科、分野、領域などの属性情報と一致す

るか

- ・ 問題タイプは適切か
- ・ 想定難易度と想定回答所要時間は妥当か
- ・ キーワードは適切か

■ 問題確認者のチェック

問題確認者はUSBメモリなど安全な方法で問題を受け取り、次ページのチェック表を用いて問題確認を行う。チェック表もUSBメモリなど安全な方法で作問者に受け渡す。

2023年度試験問題・正答・属性情報の作成時のチェック表

チェック項目	試験問題・正答の作成時の確認		
チェック方法	CBT入試作問フォーマットで書かれた問題を確認する。		
科 目 名	□数学 □物理 □化学 □英語 □情報	作問者+番号	()

区分	チェック項目	チェック者			
		①	②	③	④
問題内容	難易度は適当か	☐	☐	☐	☐
	募集要項に示した出題範囲からの出題となっているか	☐	☐	☐	☐
	問題内容の妥当性は適切か	☐	☐	☐	☐
問題表記	問題の表記に間違いはないか 特に以下の点について、チェックしてください ・誤字、脱字、スペルミス、数字・年号・単位の使用法や表記 ・別ページ等にヒントや答えになる部分はないか、択一問題の場合、複数正解はないか、また、同一の選択肢はないか	☐	☐	☐	☐
図・表	図・表はきちんとできているか	☐	☐	☐	☐
正答	出題者以外が問題を解き、正答に間違いがないことを確認したか	☐	☐	☐	☐
	想定される正答はすべて正しく採点されるように設定されているか	☐	☐	☐	☐
問題の属性情報	問題の属性情報（CBT入試作問フォーマットの「問題情報」）に間違いはないか。 特に以下の点についてチェックしてください。 ・問題内容は 教科、分野、領域、などの属性情報と一致するか ・問題タイプは適切か ・想定難易度と想定回答所要時間は妥当か ・キーワードは適切か	☐	☐	☐	☐

- 注 1. チェック者は、氏名を記入してください。
 2. 科目名、コース・日程の□の該当するものにレを記入してください。
 3. チェックした内容にレを記入してください。
 4. 問題のチェックを問題ごとに行った（「問題 1」、「問題 2」と分けて行った）場合は、問題番号を記入してください。
 5. チェックは複数の委員が行ってください。

チェック者①氏名	チェック者②氏名
チェック日 (/)	チェック日 (/)

■ 作問者へのデータ解析結果のフィードバック

また、テスト実施後の問題項目へのデータ解析結果を作問者にフィードバックし、作問プロセスの改善に活用させる機会を与えることが望ましい。そのためには、作問者も I R T などのデータ解析手法についてあらかじめ理解していることが望ましい。

3. 問題（アイテム）バンク

CBT では、あらかじめ構築された問題項目のデータベース、問題（アイテム）バンクを構築する必要がある。問題バンクでは、コンピュータ・データベース上に細分化された教育目標と各目標に対応した項目、それに関するテスト実施後の統計データなどが格納される。異なる説明の仕方をすれば、あらかじめテスト領域を満遍なくカバーできるような大量の項目を用意し、項目反応理論（[item response theory](#)；以後 IRT と呼ぶ）などの統計処理によりテスト出題の品質保証をする。原則、問題バンク中の項目は非公開で、受検者も項目内容について口外しないことを誓約してから受検する。項目内容が外部に漏れることで、受検対策に用いられる危険性があり、本来、測定

したい能力にバイアスがかかってしまい、結果として測定精度の意味で項目劣化を導いてしまうからである。

問題バンクより、各テスト版が異なる項目にもかかわらず等質となるように項目の組合せを自動的に抽出し、テスト構成を行う。

問題バンクの利点として、項目のアクセス性、共有性、再利用性があげられる。アクセス性は、コンピュータで項目を管理することにより、簡単に必要なテスト項目を検索できる利点を示し、共有性は複数のテスト開発者によってテスト項目を共有することが非常に容易になる（植野・永岡，2009）。これらの利点を実現するために、QTI（Question and Test Interoperability）規格（IMS Global Learning Consortium，2020）とよばれる CBT の標準規格に従って項目を作成することが一般的である。例えば、QTI が扱う出題形式は膨大ですべてを紹介することはできないが基本的な出題形式だけでなく、より複雑な処理を可能にする形式ももつ。

QTI では出題形式のみが定義されているが、著者らが開発している問題バンクでは、項目内容に関わるより詳細な項目属性も格納する。以下に問題バンクの属性例を示した。

□ 項目 ID

- ☐ 作成年月日
- ☐ 作問者
- ☐ 問題確認者
- ☐ 教科名
- ☐ 対象年齢・学年の範囲
- ☐ 分野名
- ☐ 領域目標名
- ☐ 目標タイプ（知識，理解，応用，分析，結合，評価）
- ☐ 項目出題形式
- ☐ 項目メディア（文章，図，音声，ビデオ）
- ☐ 正答もしくは正答例
- ☐ 問題バージョン
- ☐ 問題趣旨・解説
- ☐ 関係する URL
- ☐ 親項目 ID（他の項目を改正して作問した場合の元項目 ID）
- ☐ 敵対項目 ID（同じテストに含めない項目 ID 集合。親項目も項目の要素）
- ☐ 露出回数
- ☐ 項目正答率

☐ 項目正答率の標準誤差

☐ 項目平均所要時間

☐ 項目所要時間標準誤差

☐ IRT におけるパラメタ a , b , c (世界標準では3パラメタ・ロジスティック・モデルが推奨されている)

項目 ID は、項目を識別するために割り当てられる数字やアルファベットによって構成される文字列であり、一般に教科番号、領域番号、目標タイプ番号、出題形式番号、出題メディア番号などを組み合わせて設定されることが多い。また、問題バンクでは、教科名、領域名、目標タイプを記述する。表 1 のようにブルームの細目表 (ブルーム, 1973) を簡易化して目標タイプを記述している。さらに、項目出題形式 (多枝選択式項目, 真偽式項目, 単答式項目, 論述式項目) とマルチメディア・テストにも対応した項目が用いているメディア (文章, 図, 音声, ビデオ) を記述する。最も重要な項目内容と正答 (正答例) の記述は当然として、項目の解説になるような URL を記述しておく
とフィードバックが容易になる。項目の下位項目 (その項目に正答しなければ対象項目に正答できない項目で、対象項目に隣接したもの) を記述している。

テスト実施後に得られる項目に関する統計情報も項目改善のための分析材料としてだけでなく、構成されたテストの予測得点や予測時間を計算してテスト構成を支援できる。IRT のパラメタは、CBT における異なる問題項目からなる等質なテスト集合である平行テスト自動構成に用いられる。

データ解析の専門家はテスト実施ごとに IRT のパラメタやその他のテスト統計データを推定し、問題バンクで適切に更新されていることを確認しなければならない。

問題項目および受検者の回答データの機密性が守られるような処置が必要である。

4. 問題バンク構築のための CBT 運用モデル

前述のように IRT パラメタを推定するためには事前にテスト項目をターゲットとなる受検者と同程度の能力の受検者に実施しなければならない。電気通信大学での CBT 入試の枠組みでは新入学生全員に UEC 検定という基礎学力調査を実施し、問題項目は公表しないで

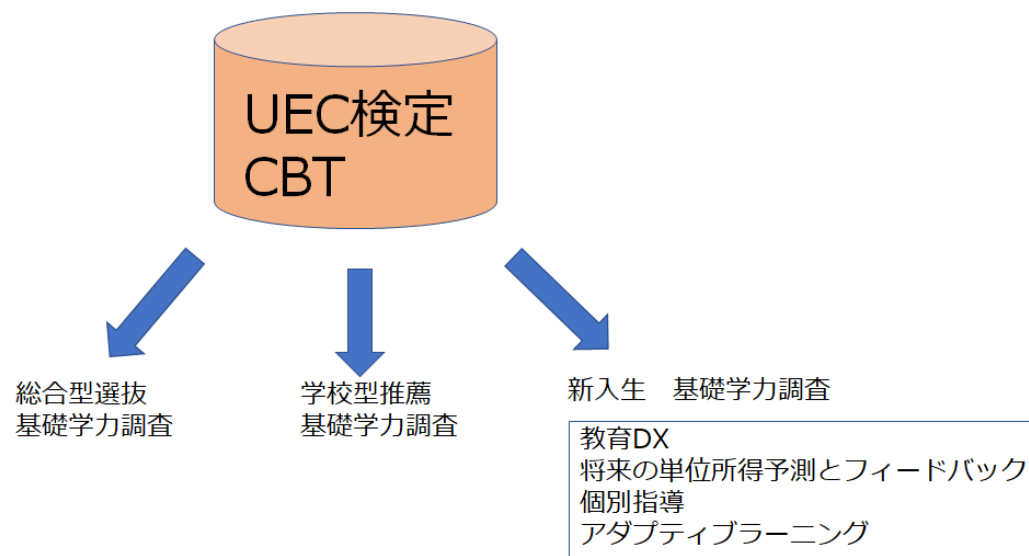


図 3. 電気通信大学での CBT 入試の枠組み

各領域の習熟度をフィードバックしたり、高校数学の補講の受講を勧めたりしている。ここで IRT パラメタを推定し、問題バンクの構築を行い、図 3 のように UEC 検定を用いて総合型選抜入試、学校型推薦入試にも用いている。一旦、問題バンクが構築されると新作問題をすべての問題数に対して 10%となるように UEC 検定に加えて試験を実施する。なお、新作問題はテスト素点、スコアの計算には用いられないが、実施したテストデータからパラメタ推定され問題バンクに格納される。入試には新作問題は含まれない。

5. CBT 実施環境

- 回答に必要なのない標準搭載機能やアプリケーションの利用を禁

止できるように受検者端末のハードウェア設定の一時変更等を行えるロックダウンシステムを利用することが望ましい。電気通信大学では端末に CBT 用 ID を入力すると自動的にロックダウンシステムの機能が作動するように設定している。

- CBT 実施のための端末は、公平性のために画面サイズを含め同一に統一することが望ましい。
- CBT のためのモニタはできる限りスクロールせずに問題文全体を読めるように 22 インチ以上あることが望ましい。
- のぞき見防止フィルタやのぞき見防止カバーなどをディスプレイに装着して近隣のコンピュータ画面をのぞき見できないように工夫することが望ましい。図 4 は電通大での CBT 環境である。



図 4. 電気通信大学での CBT 受験

- 各受検者に提示されるテストフォームは異なるものかもしくは同一でも出題順、選択枝の順序をランダムに提示することが望ましい。
- CBT のためのサーバ・ネットワーク環境は同時受検者数の負荷を考慮して冗長に設定しておかなければならない。想定する受検者数に対して負荷テストをあらかじめ行うことが望ましい。
- 停電電源の設置、クラウドもしくは複数台サーバによりサーバが何らかの要因でダウンした場合にも継続して動作できるための工夫が望ましい。
- 実施中は予備の端末を準備しておき、トラブルのあった受験生を移動させる準備をしておくことが望ましい。
- 障害を持つ受検者には対応をあらかじめ申し出てもらい、CBT 実施組織で検討して認められれば、障害を補完するのに必要なソフトウェアもしくはデバイスをその受検者のための端末に準備しておく。

6. CBT 実施

- 以下の専門スタッフが CBT 運営・保守のために従事する。

1. 作問者
2. 問題確認者
3. CBT 実施責任者
4. CBT サーバ管理者
5. CBT システム技術者
6. テスト理論専門家
7. データサイエンティスト
8. 事務作業従事者

以下のスタッフによって CBT が実施される

1. CBT 実施責任者：当日の実施マニュアルの作成、CBT 実施方法を設計し、当日は設計された方法、実施マニュアルを評価し、その改良に活かす
2. 主任監督者：実施マニュアルに従い、受検者への指示を読み上げ、不測の事態にも対応し、受検教室での CBT 実施に責任を持つ
3. サーバ管理者：当日は待機し、サーバの不具合に対応する
4. CBT システム技術者：当日の CBT の不具合への対応や CBT 設定の調整を行う

5. 技術支援員：試験監督を行いつつ、試験当日、現場での受検者の CBT 操作への質問、端末の不具合時に予備端末への移動を支援する、などにも対応する。電気通信大学では CBT システムに精通するように訓練された学生バイトが担当している

6. 事務作業従事者：当日の受付、受検者の照合などを行う

- CBT の説明会

定期的に、CBT の考え方、利点、IRT を用いたスコア計算、平行テスト、について理解を得るための説明会を開催することが望ましい。

- CBT 受験前にシステム操作の練習の機会

CBT 受検者には 受験前に CBT システム操作の練習の機会を与えなければならない。電気通信大学の CBT 入試では、志願した後に受検予定者のみ Web 上で実際の CBT システムを用いてサンプル問題を通じて操作の練習を行う機会を与えている。

- 受検者への同意書

図 5 のように CBT 実施前に受検者に CBT における回答データ利用に関する同意書、試験問題の守秘義務の同意書に同意してもらい、同意できない場合には受検できない。

UEC検定

試験の開始前に、以下の同意事項に同意していただく必要があります。同意していただけない場合は受験辞退となり、受験することができません。

CBTにおけるデータ活用の同意事項

本CBTの受験において収集したデータは、以下の目的のために利用いたします。

①本人への成績フィードバックや履修指導など、学生の指導管理に関する業務のため
②CBTや問題の評価・分析に関わる業務のため
③本学における教育・試験の改善や施策の推進のため

上記の目的のための私のデータの利用について同意します。

試験問題の非公開（漏洩禁止）の同意事項

本試験の試験問題は非公開です。受験者は、以下の事項に同意していただく必要があります。同意していただけない場合は受験できません。

①試験問題の全部又は一部(試験問題中に登場する文字に限らず、出題内容を示唆する表現も含む)を第三者に開示(漏洩)しないこと
②上記(1)を開示(漏洩)した場合、関係法令等に基づき損害賠償請求等の措置が取られること

なお、試験問題は著作権法で保護されています。出題に関する SNS への投稿、インターネット掲示板への書き込みなどについても、上記の開示(漏洩)とみなす場合がありますので、くれぐれもご注意ください。

「CBTにおけるデータ活用の同意事項」と「試験問題の非公開（漏洩禁止）の同意事項」に同意する場合は、以下のボタンを押してください。

同意 >>

図 5. CBT における回答データ利用に関する同意書、試験問題の守秘義務の同意書

7. データ解析とフィードバック

テストデータ分析はデータサイエンティストが行い、データ分析の考察はテスト理論の専門家が行う。テスト理論専門家がデータサイエンティストを兼ねてもよい。テストデータ分析では IRT のパラメ

試験名
UEC検定 第1回

科目
数学

表示

平均予測正答率	習熟度	解説
81~100%	A	内容を十分習熟できています。
61~80%	B	内容を習熟できています。
41~60%	C	習熟度がやや不足しています。自習を行いましょう。必要に応じて補習に参加しましょう。
21~40%	D	習熟度が不足しています。補習に参加しましょう。
0~20%	E	知識が不足しています。基礎から自習を行った後、補習に参加しましょう。

① あなたのスコア偏差値は50です。判定はC、平均予測正答率は59%です。

②

数学I

▼

数学II

▼

数学III

▼

数学A

▼

数学B

▼

数学C

▼

図 6. IRT スコアのフィードバック

タ推定を行い、さらに基本統計、教育用に必要な各受検者へのフィードバックを生成する。

図 6, 7, 8 は電気通信大学の新入生への UEC 検定でのフィードバックの例である。図 6 は IRT スコアからの各受検者へのフィードバックであり、該当科目のスコア偏差値、判定及びあなたの回答から算出した平均予測正答率が表示され、クリックすると、図 7 の分野別・図 8 の領域別のフィードバックが表示される。

大学には入試選抜型別平均スコア、スコアと入試得点との相関、各領域スコアと各授業単位の成績との相関などがフィードバックされ、教学指導、カリキュラムの改良に役立ててもらう。

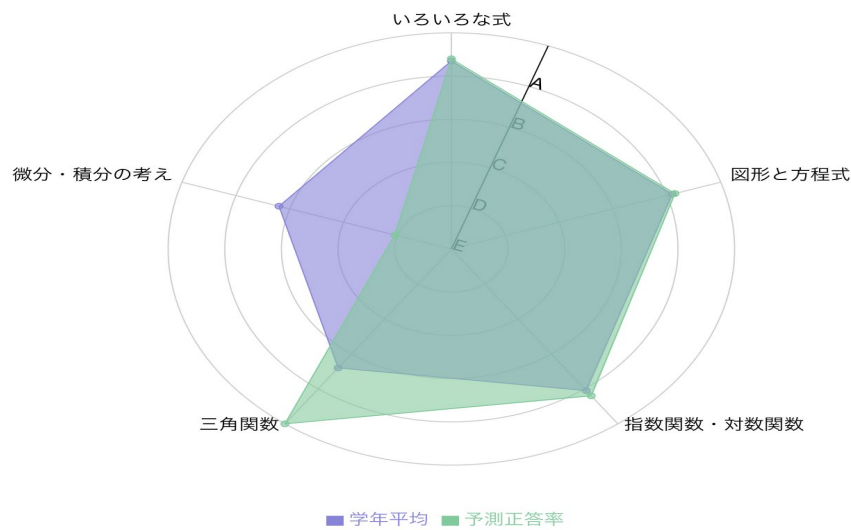


図 7. ある受検者の各分野の習熟度と受検者全体の平均値との比較

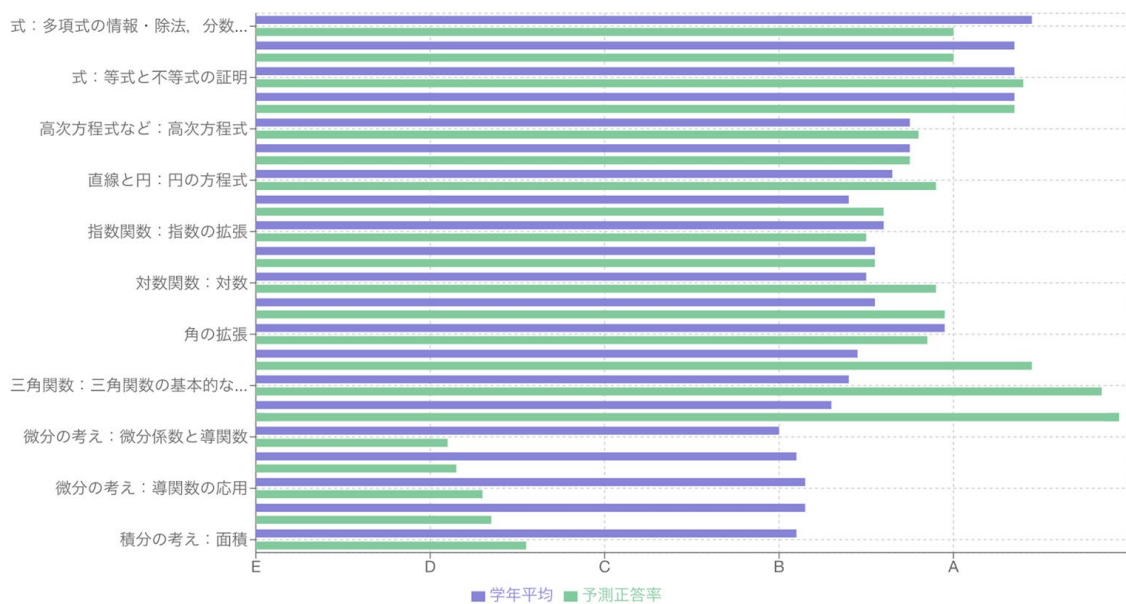


図 8. ある受検者の各領域の習熟度と受検者全体の平均値との比較

各作問者には大学へフィードバックした情報以外に各問題の基本統計、IRT のパラメタ値、テスト理論専門家による良問、悪問のフィー

ドバックが与えられ、作問プロセスの改善に役立ててもらおう。

8. テスト構成システム

8.1. 平行テスト

CBT では異なる問題項目で構成されるが、同一精度の測定を実現できる複数のテストフォーム（並行テストと呼ぶ）を用意することが望ましい。平行テストにより、同一能力の受検者が異なるテストを受検しても同スコアとなる保証がある。そのために、受検者は同一精度の複数回受検することや、受検者の居住区に合わせて、実施日時と場所を分けて実施することも可能となる。これまで、平行テストの自動構成法が数多く提案されている（van der Linden & Adema, 1998; van der Linden, 2005; Songmuang & Ueno, 2010; Sun et al., 2008）。それぞれの提案手法では、平行テストの具体的手法として「シャドーテスト」や「等質テスト」、「ユニフォームテスト」などの命名がされている。

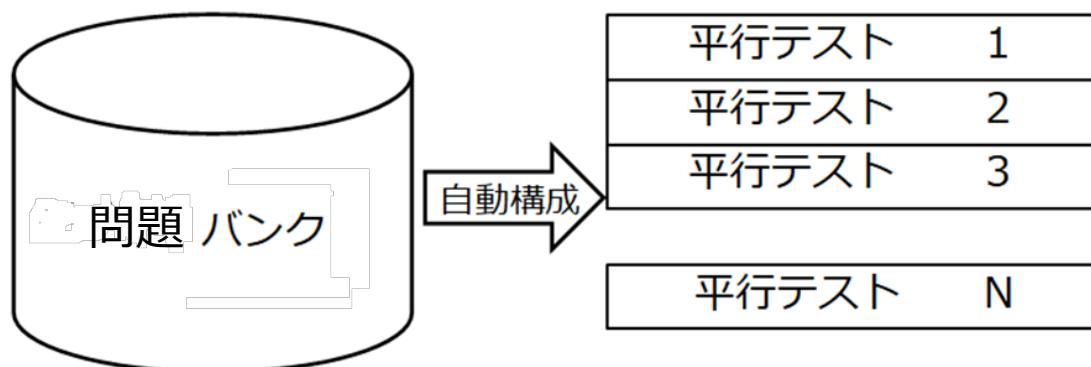


図 9. 平行テストの概念図

図 9 は平行テスト自動生成手法の概念図である。平行テストを実現するためのテストの自動構成は数理最適化問題として解かれる。

van der Linden & Adema, 1998; van der Linden, 2005 は、構成されるテストと問題バンクに残る項目集合全体の性質が同じなるように逐次的にテスト構成する Shadow Test を提案している。この後も様々な Shadow Test が提案されているが (van der Linden and Ueno (2022))、逐次的に最適化する貪欲アルゴリズムを用いており、テスト構成するごとに精度が劣化していく問題がある。

Shadow Test のような貪欲法と対照的に、テストの等質性を条件として最大数のテストを生成する人工知能アプローチが e テスティング分野で注目されている。このアプローチでは、生成されるテストの等質性が保証されたままテスト構成数を最大化できる。例えば、

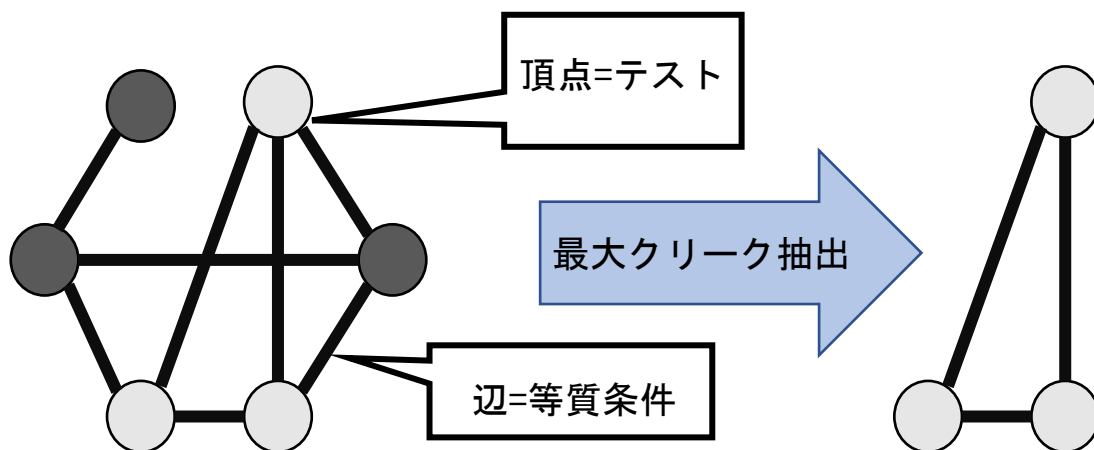


図 10. 最大クリーク法による自動テスト構成の概要

Songmuang & Ueno (2010) は最適化問題の解探索手法の 1 つである Bees Algorithm を用いてテスト構成を提案した。この手法は情報処理技術者試験をはじめとして、日本の国家試験で実際に使用されている。Ishii, Songmuang & Ueno (2014) はグラフ上で定義される最大クリーク問題に帰着してテスト構成を行う手法を提案した。具体的には与えられた問題バンクからテスト構成条件を満たすテストを頂点、2 つのテスト（頂点）が等質条件を満たす場合に辺を引いたグラフ構造から最大クリーク探索することで平行テストを構成する。

図 10 は最大クリーク法の概要を示している。まず、テスト項目数条件や情報量の条件を満たすテストを生成し、各テストをグラフの頂点とする。各二つのテストを比較し、二つのテストが等質条件を満た

せば、対応する頂点間に辺を引く。こうすることにより、（実際にはより大規模であるが）図 10 左のグラフが生成できる。このグラフですべての頂点間に辺が引かれている最大の図 10 右の部分グラフ（グラフ理論で「最大クリーク」と呼ばれる）は、最大の等質なテスト集合、すなわち最大の平行テストであることがわかる。コンピュータサイエンス分野では、グラフから高速に最大クリークを求めるアルゴリズムが多く研究されており、それらを利用することにより大規模な平行テストが生成することができる。この手法は理論的に最大のテスト構成を保証するが、問題バンクの項目数に対して構成可能な頂点（テスト）数が組合せ爆発的に増加するため、最大クリーク探索が困難である。そこで Ishiii, Songmuang & Ueno (2014) はグラフからランダムに部分グラフを選択し、最大クリーク探索を繰り返すことによりグラフ全体の最大クリークを近似的に探索する平行テスト構成手法（RndMCP 法）を提案した。本手法により、当時の既存研究よりも 10～100 倍以上多くのテストを構成できた。最大クリーク探索はグラフの頂点集合を V とすると、最先端のクリーク探索手法を用いても膨大な空間計算量を必要とするため、最大で 10 万程度のテスト構成が限界であるという問題があった。RndMCP 法の空間計算

量を緩和するために、Ishii & Ueno (2017) は第 1 段階で RndMCP 法を用いてメモリ限界まで大きな最大クリークを探索した後、第 2 段階目で第 1 段階目で求めたクリークの全頂点 18 と隣接する頂点を整数計画法により逐次的に探索することで必要な計算量を大幅に削減させる手法を提案した。これにより 10 万を超える平行テストを構成することが可能になった。しかし、Ishii & Ueno (2017) の手法の第 2 段階目は時間計算量が膨大で改善は限定的であった。この問題を解決するために淵本・植野 (2020) , Fuchimoto, Ishii & Ueno (2022) は、時間計算量の大きい整数計画法での最適化を並列化できるようにし、計算コストを大幅に削減させた。この結果、40 万を超える平行テストの生成に成功している。Ishii & Ueno (2017), 淵本・植野 (2020) , Fuchimoto, Ishii & Ueno (2022) の手法は医療系大学共用試験の CBT 用平行テスト構成に用いられている。

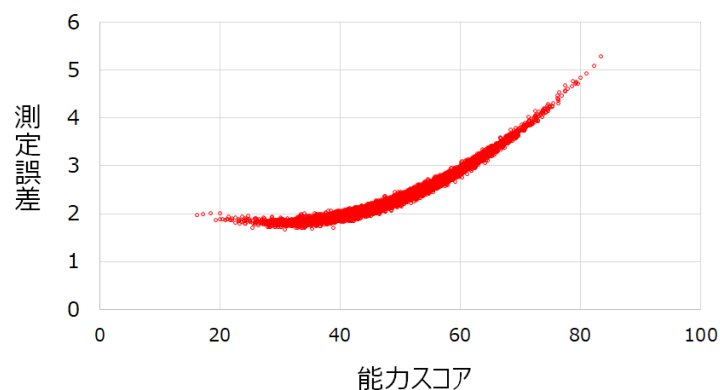


図 11. 異なる受検者の異なるテストのスコアとその誤差

また、図 11 はある試験において異なる受検者が受検したそれぞれ異なる平行テストのスコアを横軸に、縦軸にスコアの測定誤差を示している。このテストでは 40 点を合格基準としているためにスコアの測定誤差が 40 点付近でおよそ 2 点と小さくなっていることがわかる。誤差に幅があるが、この幅が大きいほどテストの等質性が低いことになり、逆に幅が小さいほどテストの等質性が高いことになる。このテストでは非常に高い等質性が実現されていることもわかる。

さらに、淵本・湊・植野 (2022) , Fuchimoto, Minato, Ueno (2023) は、Zero-suppressed Binary Decision Diagrams (ZDD) を用いて、従来手法が 1 カ月を要しても成績できなかった 50 万以上の平行テストを 24 時間以内に生成している。

8. 2. 問題項目数露出数の制御

5. 1. で紹介した平行テストの自動テスト構成手法では、ある項目がテスト構成全体で出題される回数（以降、露出数とよぶ）に偏りが生じる問題がある。

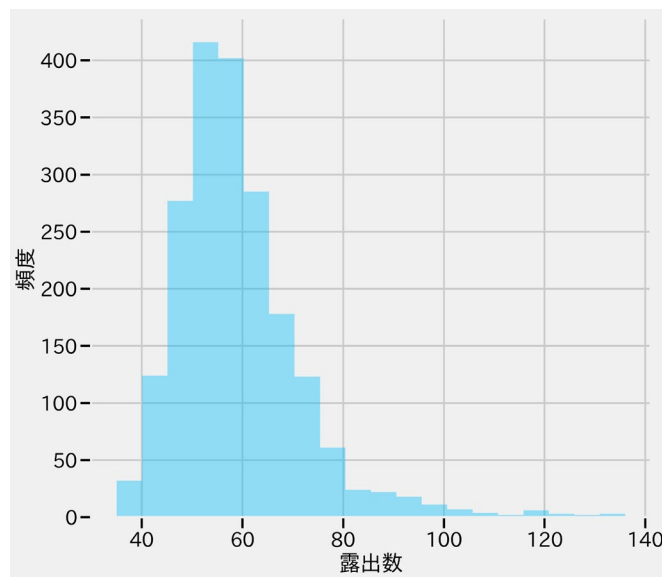


図 12. 平行テストにおける項目露出のヒストグラム

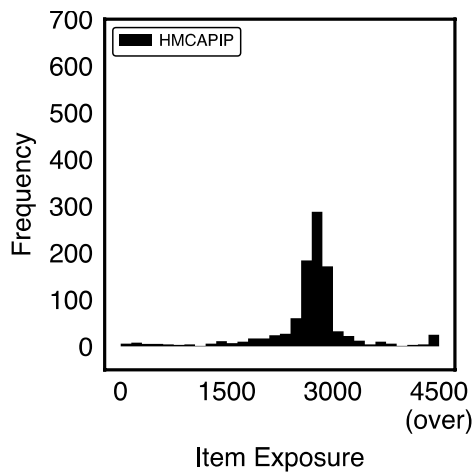
図 12 は、平行テストで出題された項目の露出数のヒストグラムである。一回も出題されていない項目もある反面、100 回を超えて出題されている項目もある。このような露出の偏りはさまざまな弊害を生じる。最も深刻な問題は、露出数が多い項目は受検者間で共有されやすく、その項目のみについてあらかじめ受検対策をされてしまい、その項目の信頼性低下につながることである。項目露出の偏りを軽減するために、Ishii & Ueno (2015) では RndMCP 法と整数計画法を用いてテスト構成し、その中から最も露出率（＝露出数の最大値／テスト構成数）が小さいテスト構成を選択する手法を提案した。具体的には、探索したすべてのクリークを平行テストの候補として保存しておき、最後にその候補で最も露出率が小さい平行テストを出力す

る。これによって、従来手法よりも露出率を軽減することができた。

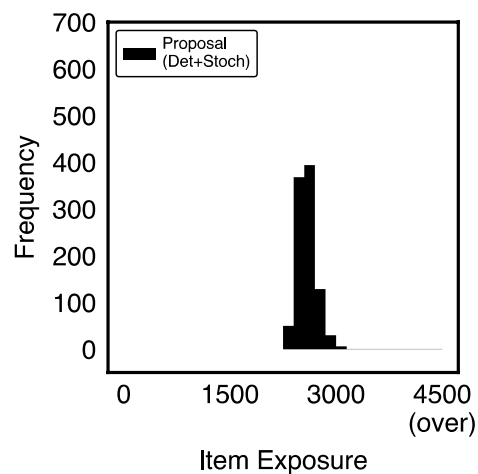
Ishii & Ueno (2015) は RndMCP 法および整数計画法で問題バンクの全項目から頂点を生成しているため、露出率の高い項目を含む頂点を生成する問題があった。植野・淵本・植野 (2021, 2022) は RndMCP 法の部分グラフの頂点選択の際に、整数計画法を用いてグラフの頂点に含まれる回数が最大の項目以外から項目を選択し頂点を生成することで、クリーク探索後のテスト構成の露出率を RndMCP 法よりも抑える手法を提案している。また、上記の手法で露出率を抑えることで、RndMCP 法よりも密な部分グラフを得ることができ、その結果より大きなクリークを見つけることができた。さらに、第 2 段階目の整数計画法でも最大露出数の項目以外から項目を選択し、この整数計画法を逐次的に解くことで、項目露出を抑えつつテストを構成できる。表 2 に提案手法と RndMCP との比較を行った。提案手法がテスト構成数を増加させる、もしくは大きく減少させずに露出率を減少させていることがわかる。

表 2 従来手法と提案手法の露出率の比較

問題バンク サイズ	OC	RndMCP			提案手法		
		テスト 構成数	最大 露出数	露出率	テスト 構成数	最大 露出数	露出率
500	0	10	1	10.0%	11	1	9.1%
	5	4371	378	8.6%	5015	375	7.5%
	10	99981	13006	13.0%	99970	5370	5.4%
1000	0	17	1	5.9%	18	1	5.6%
	5	46190	3374	7.3%	50779	1616	3.2%
	10	100000	8767	8.8%	99998	2847	2.8%
2000	0	32	1	3.1%	32	1	3.1%
	5	96773	3833	4.0%	97320	1397	1.4%
	10	100000	4044	4.0%	100000	1418	1.4%
978 (実データ)	0	18	1	5.6%	19	1	5.3%
	5	45790	5177	11.3%	54800	1683	3.1%
	10	100000	16310	16.3%	100000	2754	2.8%



Fuchimoto et al. (2022)



淵本・植野(2024)

図 13. 淵本・植野(2024)による露出数の統制

さらに、最先端の手法では、淵本、植野(2024)では, Fuchimoto, Ishii & Ueno(2022)における整数計画による問題生成の目的関数に露出数をペナルティを加えた手法を提案し、現在、露出数の分散を抑えて平行テスト生成数を最大化する手法を提案している。図 13 は Fuchimoto, Ishii, Ueno(2022) を用いて横軸に露出数、縦軸に項目数をとったヒストグラムであり、左図の従来手法では出題されない項目や過度に出題されている項目が多いのに右図の提案手法では露出数の分散が抑えられていることがわかる。

平行テスト構成では、その時代の最先端アルゴリズムを用いることにより、効率的で劣化しにくいテストを実施することができるので、テスト構成の最新技術を調査し用いることが望ましい。

9. 問題バンクのメンテナンス

問題バンクの問題項目は継続的に評価し、過度な露出による劣化などが確認された場合は問題バンクから削除しなければならない。そのためには各項目の正答率や IRT パラメタを経時的に分析し、基準としている集団に比較し、正答率が高くなってきたり難易度が低くなってくると露出による劣化が起こったと判断できる。その場合

には問題バンクから当該項目を削除していく。また、削除される問題項目数の予測値を計算し、新作問題はその予測値を超えるように作問し、問題バンクに格納していくようにすることが望ましい。

10. Q&A

1. 質問：タブレットやノートパソコンでなくコンピュータ演習室を用いるメリットは何ですか？

回答：

- ・ 国際標準に従うことが容易になります。
- ・ ロックダウンシステムを設定できるので安全にセキュリティの高い CBT 実施ができます。
- ・ 安全なネットワークを用いて受検者が割り当てられた受験番号を入力すれば CBT サーバに自動的にアクセスでき準備が少ないので試験実施者の負担が減ります。

2. 質問：試験中に機器の不具合が出ればどうしますか？

回答：

すぐにあらかじめ準備してある予備の端末に移動してもらい、途中で再開できる場合は試験時間を担保して試験を修了することができます。ただしこの場合でも移動時間を考慮して、試験終了を遅らせます。これまでそのようなケースはなかったのですが、万一途中で再開できない場合は、異なるテストフォームのテストを別日時で再試験してもらいます。

3. 質問：CBT の問題の間違いが試験中に発覚したらどうしますか？

回答：その場で問題訂正は行いません。その問題番号を記録し、採点に用いなくします。問題数が比較的多いので 1 問程度の間違いでは大きな影響はありません。

4 質問：問題項目の漏洩はどのように防ぎますか？

回答：

- 作問者・確認者・管理者と守秘義務の契約書を交わしています。

- 受検者は CBT 受験前に守秘義務に同意した方のみ受検できます。
- 配布する計算用紙は受験番号、氏名を書いて回収しています。
- ビデオカメラで受検者が問題を撮影したり書き写していないかを監視しています。

5. 質問：CBT の得点として IRT の能力推定値を用いるのでしょうか？

回答：英語や医療系テストのように IRT スコアが世界的に標準的な科目では IRT の能力推定値をスコアに用いることが多いですが、日本の入試ではまだなじみはないでしょう。平行テストでは、すべてのテストの等質性が保証され IRT 能力推定値を素点に変換できるので、あるスコアで合否を決める場合には IRT 能力推定値を用いても素点を用いても構いません。それよりも合否を決めるクリティカル・ポイントとなるスコアに対応する IRT 能力推定値での測定誤差が小さくなっていることを確認しておくことが重要です。

6. 質問：IRT を用いた場合にはスコアに誤差があるようですが、誤差のない素点のほうが良いのではないのでしょうか？

回答：IRT は推定値に誤差が大きいので使えないという人がいます。

しかし、信頼性の概念は 同一能力の受検者が異なるテストを受験しても同等のスコアを返すことです。ですので一般のテスト得点としての素点は 必ず誤差があります。むしろ誤差を考えずにテスト構成しているので相当大的な誤差を持っていると考えられます。このことは世界標準から見ると致命的な欠点に見えてしまいます。その誤差を小さくするために IRT や問題バンクを用いるのです。また、IRT は現在では十分に確立された技術でもあります。正しいモデルと推定法を用いれば真の同時確率分布を漸近的に推定でき、予測精度を非常に高く推定できるのです。ただ、誤用してしまうと誤差が大きくなってしまう危険性があります。正しい知識を持ちましょう。

7. 質問：CBT はテスト技術なので教育に役立たないのではないのでしょうか？

回答：それも大きな誤解でしょう。一般に IRT における能力推定値は一つの値を返すだけで学習者へのフィードバックとしては弱く見

えるかもしれません。しかし、CBT は、IRT などの受検者の各問題項目への学習者の反応の予測精度の高いモデルを採用していることを思い出してください。IRT を含むこれらのモデルは学習者の未知の項目への理解度の予測精度が高く、学習者がどこで分からなくなっているかなどを精度高く予測できるモデルなのです。これらは学習者に適応して教育するアダプティブラーニングのエンジンとして採用されています。

8. 質問：IRT モデルはどのモデルを使えばよいですか？

回答：現在は 国際標準では 単問では 3 パラメタ・ロジスティック・モデル（Lord, F. M. and Novick, M. R. (1968)）が推奨されています。ただし、3 パラメタ・ロジスティック・モデルはベイズ推定でしかロバストにパラメタ推定できないことが知られているので気を付けましょう。ラッシュモデルや 1 パラメタ・ロジスティック・モデルは基本テスト素点と同じ情報で見た目を IRT スコアに見せているだけで予測精度が低く、結果として信頼性も低くなるという欠点があります。CBT では等質テストに予測精度の高い手法が必要で IRT にこだわらず Deep Learning など様々な手法が提案されています。ま

た、複数の問題によって成り立つ大問では一般化部分的採点モデル generalized partial credit model (GPCM) (Muraki, E. (1997)) が用いられることが多いです。

9 . 質問 : IRT モデルのパラメタ推定は何を用いればよいでしょうか？

回答 : 質問 8 への回答でベイズ推定でしかロバストに推定できないことを述べました。ベイズ推定の中でも予測誤差を最小にする EAP (Expected a posteriori) ベイズ事後分布におけるパラメタの期待値が推奨されます。この推定のためには、MCMC 法 (Markov chain Monte Carlo 法)、変分ベイズ法、ガウス近似法などが知られています。

10 . 質問 : アイテムバンク構築のためにリンケージ (等化) はどのようにすればよいですか？

回答 :

線形変換による等化手法 (Mean/Sigma 法、Mean/Mean 法、項目特性曲

線法など)が知られています。しかし、これらの線形変換は数学的に IRT 能力値全体には成り立たないことが知られています。実際には、テスト構成時に全体のテスト項目数の 5%から 10%に匹敵する新規項目を付け加えてテスト出題し、採点には用いず推定にのみ用いる方法が多く採用されています。既知の項目パラメタのみを用いて受検者の能力値パラメタを推定します。新たに推定された能力値パラメタを所与として、テストデータより未知の新項目のパラメタを推定します。これらのパラメタの推定誤差が十分小さく望ましい値になるまで異なる受検者に実施し、パラメタ推定精度が十分になるとその項目は問題バンクに格納します。採点に用いない項目が出題されていることは受検者には知らせないようにしています。

11. 質問：CBT の世界標準のためにどのようなシステムを用いていますか？

回答：

CBT の問題コンテンツの国際標準規格 (IMS Question & Test Interoperability (QTI)) としてオープンソースの TA0 (内田洋行) を用いています。基本無料で用いることができます。

平行テスト生成のためのプログラムは植野真臣研究室で公開されていますが簡単には使えないかもしれませんので研究室まで直接ご相談ください。

12. 質問：CBT の実施後の問題バンクの評価はどのように行いますか？

回答：

以下のような評価を行って改善を行っています。

- 問題項目の IRT 特性値や統計値から各出題項目の評価を行い
作題者にフィードバック
- 平行テストの事後のスコア誤差推定値を評価し，テストの等
質性，誤差を評価
- 問題項目の正答率、IRT 特性値の経時的分析から大きく変化し
ている項目を調査して劣化している場合には問題バンクから
削除

参考文献

独立行政法人大学入試センター (2022). CBT での「情報特」の出題

に関する調査研究について,

https://www.dnc.ac.jp/research/cbt/cbt_houkoku.html

渕本壱真・植野真臣 (2020). 平行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. 電子情報通信学会論文誌,

J103-D(12), 881-893.

Fuchimoto,K., T. Ishii,T. & Ueno,M. (2022). Hybrid Maximum

Clique Algorithm Using Parallel Integer Programming for

Uniform Test Assembly. *IEEE Transactions on Learning*

Technologies, doi: 10.1109/TLT.2022.3163360.

渕本壱真・湊 真一・植野真臣 (2020). Zero-suppressed Binary

Decision Diagrams を用いた自動テスト構成, 人工知能学会論文

誌, **37**(5), A-M23_1-11.

Fuchimoto, Kazuma, Shin-ichi Minato, and Maomi Ueno (2023).

Automated Parallel Test Forms Assembly using Zero-suppressed

Binary Decision Diagrams." *IEEE Access*.

渕本壱真・植野 真臣 (2024) 項目露出ペナルティを用いた整数計

画法による自動並行テスト構成, 統計数理

IMS Global Learning Consortium, Inc. (2020). *IMS Question and Test Interoperability Version 3.0 Public Draft*.

Ishii, T., Songmuang, P., & Ueno, M. (2014). Maximum clique algorithm and its approximation for uniform test form assembly. *IEEE Transactions on Learning Technologies*, 7(1), 83–95.

Ishii, T., & Ueno, M. (2015). Clique algorithm to minimize item exposure for uniform test forms assembly. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education 17th International Conference Proceeding* (pp. 638–641). Springer.

Ishii, T., & Ueno, M. (2017). Algorithm for uniform test assembly using a maximum clique problem and integer programming. In E. André, R. Baker, X. Hu, Ma. M. T. Rodrigo, & B. du Boulay (Eds.), *Artificial Intelligence in Education 18th International Conference Proceeding* (pp. 102–112). Springer

石井隆稔・赤倉貴子・植野真臣 (2018). 複数平行テスト構成における整数計画問題を用いた最大クリーク探索の近似法. 電子情報通信学会論文誌 D, *J101-D*(1), 725–728.

Lord, F. M. and Novick, M. R. (1968) *Statistical theories of mental test scores*, IAP

Miyazawa, Y., & Ueno, M. (2020) Computerized adaptive testing method using integer programming to minimize item exposure. *Advances in Artificial Intelligence: Selected Papers from the Annual Conference of Japanese Society of Artificial Intelligence* (JSAI 2019), (pp.105–113). Springer.

宮澤芳光・宇都雅輝・石井隆稔・植野真臣 (2018). 測定精度の偏り軽減のための等質適応型テストの提案. 電子情報通信学会論文誌 D, *J101-D*(6), 909–920.

宮澤芳光・植野真臣 (2021a). 項目暴露を軽減する二段階等質適応型テストの提案. 日本行動計量学会第 49 回大会付録集, pp. 118-121.

宮澤芳光・植野真臣 (2021b). 最大暴露数を制約に用いた二段階等質適応型テスト. 日本テスト学会第 19 回大会発表論文抄録集, pp. 138-141.

Muraki, E. (1997). A Generalized Partial Credit Model. In: van der Linden, W.J., Hambleton, R.K. (eds) *Handbook of Modern Item*

Response Theory. Springer, New York, NY.

https://doi.org/10.1007/978-1-4757-2691-6_9

Samejima, F. (1977) Weakly parallel tests in latent trait theory with some criticisms of classical test theory, *Psychometrika*, Vol. 42, No. 2, pp. 193–198.

Sun, K. T., Chen, Y. J., Tsai, S. Y., & Cheng, C. F. (2008). Creating IRT-based parallel test forms using the genetic algorithm method. *Applied measurement in education*, 21(2), 141–161.

ソンムアン ポクポン・植野真臣 (2008). 統合型 e テスティングの開発と実践. 日本テスト学会誌, 4(1), 54-64.

Songmuang, P., & Ueno, M. (2010). Bees algorithm for construction of multiple test forms in e-testing. *IEEE Transactions on Learning Technologies*, 4(3), 209–221.

植野晶・瀧本壱真・植野真臣 (2021). 項目露出を考慮した整数計画法による平行テスト構成. 科学研究費基盤研究(S) 19H05663 (研究代表者 植野真臣) 報告論文集, 18-38,

<http://www.ai.lab.uec.ac.jp/symposium21/symposium21.pdf>

植野晶・瀧本壱真・植野真臣 (2022). 項目露出を考慮した整数計画

法による平行テスト構成. 電子情報通信学会論文誌 D, J105-D(8), 485-498.

植野真臣 (2009a). e テスティング——最先端テスト技術. 電子情報通信学会誌, 92(12), 1017-1021.

植野真臣 (2009b). e テスティング——先端理論と技術. 教育システム情報学会, 26(2), 204-217.

Ueno, M. (2021). AI based e-Testing as a common yardstick for measuring human abilities. *18th IEEE International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 1-5, doi: 10.1109/JCSSE53117.2021.9493810.

Ueno, M., Fuchimoto, K., & Tsutsumi, E. (2021). e-testing from artificial intelligence approach. *Behaviormetrika*, 48(2), 409–424.

植野真臣・永岡慶三（編著） (2009). e テスティング. 培風館.

Ueno, M., & Miyazawa, Y. (2019). Uniform adaptive testing using maximum clique algorithm. *20th International Conference, AIED 2019*, 482–493.

Ueno, M., & Miyazawa, Y. (2022). Two-stage uniform adaptive testing to balance measurement accuracy and item exposure. *23rd*

International Conference, AIED 2022, 626-632..

van der Linden, W. J. (2005). *Linear models for optimal test design*. Springer.

van der Linden, W. J., & Adema, J. J. (1998). Simultaneous assembly of multiple test forms. *Journal of educational measurement*, 35(3), 185–198.

van der Linden, W. J., & Ueno, M. (2022) Shadow-Test Approach to Adaptive Testing, *Behaviormetrika*, 49(2), 1-3

Wainer, H. (2000). CATs: Whither and whence, *Psicológica*, 21(1), 121–133

Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, 17(4), 17–27.