

基礎編

第1章 項目バンク構築の利点

1.1 項目バンクとは

- **項目バンク**：「**項目 (item, テストにおける個々の問題)**」を保存しているデータベースのこと。バンク内の項目は難易度や問題形式によって整理・分別して格納され、テスト作成に際して適切に利用できるようになっている。
- コンピュータ適応型テスト (Computer-Adaptive Test)の開発とともに発達したため、項目バンクを利用しているのは主に大規模なテスト機関や教育機関だが、個人による小規模な環境でも利用される。学校教育におけるテストや授業でも、項目バンクを利用すれば、良質なテスト問題や教材を保存し、再度利用することができる。
- 本書では、このような個人ベースの使用目的において活用できる **TEASY 項目バンク**の理論や仕組みを説明していく。

1.2 項目バンク利用の利点

- 良い項目の作成には時間と労力を要する。そこで、良質な項目を、再利用可能なかたちでバンク内に整理、保存しておくことで、少ない労力で高い学習効果を得ることができる。
 - また、広範な難易度をカバーする数多くの項目をバンク内に備えておけば、個々の受験者の能力に応じた出題が可能となり、測定の正確さを高めることができる。
 - さらに、項目数が多ければ、異なる項目で同じ難易度のテストフォームを複数取り揃えることができる。複数のテストフォームを使う利点は、情報漏れを防ぎ、安全性を高めることである。
 - 学校のテストにおける項目バンク利用の利点を考える。項目を作成 → 項目バンクに良問を登録 → 項目を作成…を繰り返し、項目バンク内に良問を蓄積していけば、作成の労力を徐々に削減でき、いずれは項目バンク内にある良問だけでテストを作ることができる。
 - このように作られたテストを **Form1** とする。**Form1** は項目バンク内の良問のみから構成されているので、学年が変われば再度利用できる。そしてこの **Form** を複数用意しておけば、しばらくは新しい項目を作成する必要がなくなる。
- ☆ 例えば、国内の TOEIC 公開テストでは、およそ 6 割の問題が過去の試験から再利用されたものである。テスト機関として世界有数の規模を誇る ETS は莫大な項目バンクを蓄えていると考えられるが、今後は情報機器の発達と普及に後押しされ、学教教育の現場でもこのような項目バンクの活用が広まることが期待される。

1.3 TEASY 項目バンクの主な機能と特徴

- 項目に関する様々な情報 (e.g., 難易度、問題形式)を整理して保存できる。(保存、登録は Microsoft Word 形式で行えるため一般の教員でも扱

いやすい。)

- 登録した項目を組み合わせて、バランスの良いテストフォームを作成できる。
- 項目やテストフォームの使用履歴、統計をとっておくことができる。

1.4 項目再利用の流れ

- 簡単に項目を組み合わせて教材やテストを作成したい場合は、**Step1, Step2**のみで可。
- 項目バンクに、ある程度項目がたまれば、**Step5, Step6**のみで可。

表1. 項目再利用の流れ

Step1. 項目の作成：目的に合ったテスト問題や教材の作成
Step2. 項目の保存と登録：1項目1ファイルとして保存し、バンクに一括登録
Step3. 項目の分類：項目のカテゴリを登録して分類（任意）
Step4. 付属情報の登録：項目の難易度や採点基準などの記入（任意）
Step5. テストフォームの作成：項目を選択して、新しいテストの自動作成
Step6. テストフォーム履歴の保存：選択した項目の履歴保存（任意）

第2章 Step1. 項目の作成

2.1 項目の妥当性

- 各項目が測定しようとしている能力を、正しく測定できているかどうかを、項目の妥当性という。
- 例えば文法テストにおいて、どんな文法的知識（e.g., 前置詞の知識を問いたいのか、関係詞の知識を問いたいのか）を測定するのか明らかにすれば、高い妥当性をもった項目を作成しやすくなる。そこで、各項目の測定対象名を明記しておく、項目バンクからそれぞれの用途に合った項目を引き出しやすくなる。

2.2 集団規準準拠テストと目標基準準拠テスト

- テストは**集団規準準拠テスト (Norm-Referenced Test: NRT)**と、**目標基準準拠テスト (Criterion-Referenced Test: CRT)**に大別される。
- **集団規準準拠テスト (NRT)**とは、特定の集団内における、個人の相対的な位置を観測するためのテストである。このような、集団の中での個

人の相対的位置に基づいた評価方法を**相対評価**という。

- 相対評価には自分の位置を把握できる反面、得点の向上が必ずしも評価の向上には繋がらず、受験者の動機を削ぎかねないという短所がある。
 - 対して**目標基準準拠テスト (CRT)**は、個人が学習内容をどのくらい習得したかを測定する。あくまでも個人の習得の度合いを判断するもので、そのような評価方法を**絶対評価**という。
 - しかし、異なる集団間で絶対評価の基準を統一することは難しい。また評価に評価者の主観が入りやすいという短所がある。
 - 相対評価と絶対評価のもつこのような短所に対して、項目応答理論を使い、項目に難易度 (困難度) をつけて分析することで、克服しようとする努力がなされている。これは、テストの難易度や集団の優劣に左右されずに受験者の能力を算出することを可能にし、どのテストを受けても受験者の能力がある程度正確に算出されるので、過去の自分との比較を行いやすくなる。
 - さらに最近では、**能力記述表 (Can-do statements)**を示すテストも増えつつある。能力記述表とは、ある得点を取れば、何ができるようになるかを具体的に表す指標である。また、NRT と CRT を融合した **CNRT (Criterion-Norm-Referenced Test)**というテストも存在する。
- ☆ TOEIC では、受験後に *equating* と呼ばれる統計的処理を行なって点数を調整している。そのため、母集団に関わらず自分の実力を同じ基準で測ることができるとされている。しかし実際には、スコアの出やすさはテストフォームによって異なっており、数十点程度の誤差はどのスコアレンジでも散見される。受験後には *Abilities Measured* という記述書によって項目別正解率を知ることができるが、そこからスコアとその得点層をよく観察すると、特定のスコアレンジには、毎回同じくらいの割合の受験者が含まれていることが分かる (e.g., 900-990 点は受験者全体の 3%程度)。よって、素点にある程度基づいた絶対評価に相対評価の特性を加えた処理を行なっていると考えられる。合否を出す英検などの他の英語資格試験では、どのように合格点を算出しているのかも非常に興味深い内容だと感じる。

2.3 テスト項目の形式

この節ではテスト項目の形式 (question type)を、解答時の応答形式によって 3 タイプに大別して紹介する

- 1つ目のタイプは、解答時に表現力を要求しない**選択式応答問題 (selected-response questions)**である。これは理解能力 (受容能力) を問う理解様式応答テスト (**receptive-response test**) において使用される。また、各項目が比較的限定された能力を測定する部分的測定テスト (**discrete-point test**) 項目が多い。
- 2つ目のタイプは、理解したことを 1 語から 1 文で発表することを求める**制限つき表現能力問題 (limited-response questions)**である。理解能力と表現能力の両者を測る総合的測定テスト (**integrative test**)で使われる。
- 3つ目のタイプは、トピックに対して自由に発表させる**自由表現応答課題 (extended-production tasks)**である。タイプ 3 に進むほど、表現能力の比重が重くなる。

- **タイプ 1. 選択式応答問題 (selective-response questions)** : 複数の選択肢から正解を選ばせることで、リーディングやリスニングの理解度を測定する形式。部分的測定に適する

(1) **多肢選択問題 (multiple-choice questions)** : 最もよく使用される形式。質問部分である基幹 (stem)部分と、正答と錯乱肢である選択肢部分からなる。

- ◆ 長所 : ・ 良問かつ十分な項目数があれば、高い信頼性をもつテストを作成できる。
 - ・ 採点が容易
- ◆ 短所 : ・ 良質な錯乱肢を作成することは非常に難しい。
 - ・ テスト・テーキングストラテジー (test-taking strategy)の影響を受けやすく、妥当性が損なわれる可能性がある。

☆ テストテーキング・ストラテジーの是非

→テストテーキング・ストラテジーが有効なテストは、本来想定している能力以外の影響に左右され、妥当性を確保できないという欠点が指摘されるが、その存在がプラスに働く場合もある。例えば、TOEIC や英検などの英語資格試験における長文問題では、あらかじめ設問に目を通していき、正答に必要な情報のみを本文から拾い読みする、というストラテジーが有効な場合がある。これは、内容理解の妥当性を損ねている可能性も考えられるが、必要な情報のみを探し出し読み取るという scanning (スキヤニング) の訓練にもなり得る。このことは、テストテーキング・ストラテジーが有効なテストをあえて課すことで、生徒に効率的な読解、聴解法を身に付けさせることができる可能性があることを示している。

(2) **正誤判定問題 (true-false questions)** : 本文の内容に関する文が、正しいか誤りかを判断させる問題

- ◆ 長所 : ・ 問題作成や採点が容易
 - ・ 本文のどの程度の情報を問うかで難易度を調整可能
- ◆ 短所 : ・ 信頼性を確保するためには、多くの項目を作成する必要がある。
- ◆ 留意点 : ・ 曖昧な表現、間違いだと分かりやすい絶対的な表現 (e.g., all, always, never など)は避けたほうが良い (Brown, 2005)
- ◆ その他 : 誤りを修正させるなどの追加課題を行えば、より正確に理解を測定することが可能。

(3) **マッチング・組み合わせ問題 (matching questions)** : 情報を、前提と選択肢の 2 列に分けて提示し、前提に合う情報を選択肢から選んで結びつける形式

- ◆ 長所 : 作成、採点が容易

- ◆ 短所：前提と選択肢の数が同じ場合、各項目が独立しておらずマイナスに働くことがある (i.e., 1 つ間違えると必ずどこかでもう 1 つ間違いが出る。)

(4) **整序問題、文構築問題 (word rearrangement, sentence builds)**：ばらばらの一文を、正しい語順に並び替える問題。文構成能力が測定される。

- ◆ 長所：作成、採点が容易
- ◆ 短所：組み立てるパーツが多過ぎると、解答に時間がかかり過ぎたり、可能な並び替えが正解以外にも出てしまう可能性がある。
- ◆ 留意点：ポイントとなる文法項目や構文をあらかじめ把握した上で出題すべきである。

(5) **間違い探し問題 (error identification, error detection)**：誤文の間違い箇所を選ぶ問題

- ◆ 長所：文法、構文の知識が異なる学習者を弁別できる項目を多く作成しやすい。
- ◆ その他：誤文の間違い箇所を修正させる課題を課せば、当て水量での解答を防ぐと同時に、文の校正能力を見ることもできる。

(6) **空所なし補充問題 (missing-word identification / detection)**：あらかじめ文章中の単語をいくつか抜いておき、脱落箇所と語を答えさせる形式

- ◆ 長所：
 - ・ 抜く語を変えることで、文法、語彙、読解能力等を柔軟に測定できる。
 - ・ 読ませながら、本文の音声を聞かせることで、正書法(orthography) やプロソディーの訓練もできる。
 - ・ 一度学習した文章からも容易に難易度が高い問題を作ることができるので、復習テストとして活用しやすい。
- ◆ 留意点：抜いた語が別の箇所にも挿入可能なことが多々あるので注意が必要である。

■ **タイプ 2. 限定表現様式問題 (limited-production questions)**：問いに対して、1 語から 1 文程度の発表 (筆記 / 発話) を求める問題。

(1) **空所補充問題 (gap-filling questions)**：文や文章中の空所を穴埋めさせる問題

- ◆ **クローズテスト (cloze test)**：テキストから等間隔に語を抜き、穴埋めさせる、空欄補充問題の代表的な形式。
 - 構成概念妥当性や信頼性が盛んに議論されてきた (総合的能力を測ることができる ⇔ もっと狭い能力を測定している)。
 - **意図的削除法 (rational deletion)**で、測定したい能力に絞って、対象語のみを抜く形式もある。
 - 採点には、原文通りの答えのみを正解とする正語法 (exact-word method) と、文脈に合っていれば正解とするか部分的を与える適語法 (acceptable-word method) がある。前者は採点が容易で、後者は高い波及効

果が期待できる。

→ クローズテストの実用性、信頼性をより高めたものとして **C-test** がある。これは 空所の単語の一部を与えて、解答のバリエーションを制限する。

◇ 英語熟達度テストとしてクローズテストを利用する際は、(a) テキストのレベル、(b) 項目数、(c) 採点基準に特に注意する必要がある。

→ (a) 床面効果や天井効果が出る可能性を極力抑えるために、あらかじめ母集団の熟達度を予測し、熟達度の異なる協力者を適切に弁別できる難易度を見極めること、

→ (b) 信頼性確保のために 30~50 項目は用意すること、

→ (c) 正語法では、原文では代名詞で記述されている語を (e.g., it)、指示対象の名詞で答えている場合 (e.g, test) など、正誤の判定が難しくなる可能性がある。文脈から文法的、意味的にも適当だと判断できれば正解とする適語法の方が上手く弁別できる場合が多い。

→ また、調査に用いたクローズテストは必ず **Appendix** に載せて、客観性を示すことも必要である。

(2) **短文応答問題 (short-response / short-answer questions)** : 質問に対して 1 語から短文で口頭または筆記で応答する形式。理解能力と、適切な応答ができるかという発表能力が問われる。

◆ 長所 : 実際のコミュニケーションに則った形式のため、**真正性 (authenticity)** が高く、理解能力と産出能力との橋渡しとなる形式である。

◆ 短所 : テストテキング・ストラテジーの影響を受けやすい。

(3) **書き取り、ディクテーション (Dictation)** : 聞こえてくる文やその一部を書き取るテスト。聴解能力、語彙力、文法能力、スペリングの知識などの総合的な能力が求められる。

◆ 長所 : ・作成が容易

・採点が大きくずれることがないので、生徒同士での採点や、学習目標の自己採点へ効果的な活用が可能

◆ 留意点 : 測定したい能力によって、採点基準を変化させる必要がある。例えば聴解力を見たいのであれば、少しのスペリングのミスなら許容するなど。

(4) **英文和訳 (English-Japanese translation, translation from English into Japanese)** : 対象の英文を和訳する問題。英文理解の程度を測定する。

◆ 長所 : 作成が非常に容易

◆ 短所 : 採点が非常に難しい

◆ 留意点：様々な訳が考えられるため、採点基準を明確にすることが必須。

(5) 音読 (**reading aloud**)：英文を音読させる問題で、発音、強勢、流暢さなどの音声面に加え、語彙知識や読解能力も見ることができる。

◆ 長所：ペアでの活動が行い易く、授業やテストにインタラクティブな楽しみを与えることができる。

(6) 復唱 (**repeat**)、リテンション (**retention**)：1文が読まれた後に、聞いたとおりに繰り返す課題。聴解能力及び文章構成能力を測定する。

◆ 長所：・文を見ずに発音するので、より自然な発音や流暢さなどを測定できる。

・ ペアでの評価、採点が可能で、作成、実施、採点の効率が総じて高い。

(7) 口頭文構築問題 (**oral sentence builds**)：ランダムに聞こえてくる語句を並び替えて意味の通る文を作る課題（整序問題の口頭版）。文章構成能力、発音、流暢さなどを測定する。

◆ 長所：ペアによる訓練、評価を行い易い。

◆ 留意点：比較的難しい問題になりやすい。組み合わせるピースは意味のかたまりで区切ったものを3つ程度作成すること。

■ **タイプ 3. 自由表現様式応答問題 (extended-production questions)**：与えられたタスクに対し、筆記か口頭で自由に答える形式。現実のコミュニケーションの形式に近いものも多く真正性が高い反面、評価、採点が採点者の主観に影響されやすい。

(1) 英訳、英作文 (**Japanese-English translation, English composition**)：1文から数文の和文を英語に訳す問題。発表語彙、文法、構文能力を測定する。

◆ 留意点：さまざまな解答が考えられるので、採点基準を明確にすることが必須。

(2) 要約 (**summary writing**)：テキストの要点をまとめるテスト。読解力（または聴解力）と作文能力が総合的に測定される。

◆ 長所：理解能力と表現能力、さらには情報をまとめる能力を総合的に見ることができる。

◆ 短所：英文をコピーして繋げただけの要約が多く、採点が非常に難しい。

◆ 留意点：読解や聴解における理解能力に焦点を当てたいのならば、母語で要約させた方が理解度を特定しやすくなる。

(3) **談話完成タスク (DCT : discourse completion task)** : 短い場面描写と、それに続く言語行為部分が空欄になった対話が提示され、筆記または口頭で文脈に合った応答を求めるタスク。以下のバリエーションがある。

(a) 解答を選択式にした

選択式談話完成タスク

(multiple-choice discourse completion task: MDCT)

(b) プロンプト (prompt: 正解を引き出すためにヒントや手助け)の後、口頭で回答する

口頭談話完成タスク

(oral discourse completion task: ODCT)

(c) プロンプトを読み、受験者と母語話者が演じる

談話ロールプレイングタスク

(discourse role-play task: DRPT)

(4) **ストーリー・リテリング (story retelling)** : ストーリーを読んだ後、もしくは聞いた後にその内容を語るタスク。

◆ 長所 : ・見聞きしたことを誰かに伝えるという形式から、真正性が高い。

・ 作成が容易

・ 使用テキストを工夫することで、読解力、聴解力、話す能力、文法知識などさまざまな能力を見ることができる。

(5) **プレゼンテーション (presentation)** : 与えられたトピックについて話すタスク。準備期間や原稿の有無によって主に4つのスタイルがある (表2参照)。

◆ 長所 : 音声面 (e.g., 発音、強勢、イントネーション)に加え、スピーチ内容から文章能力、論理性、説得性、創造性など多様な能力を見ることができる。

◆ 留意点 : 即興、即席スピーチの場合、その場でトピックを与えるためトピックの熟知度に影響されやすい。

表2 準備期間と原稿の有無によるスピーチスタイル

スタイル	準備期間	原稿の有無	備考
1. 即興スピーチ	1分程度で準備する	無し	手軽に実行でき、短時間

(impromptu speech)			で評価可能
2. 即席スピーチ (extemporaneous speech)	即興スピーチよりも少し長く準備する	簡単なメモ or 無し	
3. 原稿スピーチ (manuscript speech)	原稿をしっかりと読み込む	有り	原稿を持って行うが、読んでいるという印象を与えないよう準備する。
4. 暗唱スピーチ (memorized speech)	原稿を完全に暗記する	無し	レシテーションコンテスト (recitation contest) で行われる。

(6) **ロールプレイ (role play / simulation tasks)** : 2人の受験者同士、あるいは面接官と受験者で役割を演じるタスク。

◆ 長所 : ・真正性が非常に高い。

・そのため演じる役割やタスクによって、問題解決能力や語用論的知識 (pragmatic knowledge)、会話を円滑に運ぶための方略的能力 (strategic competence) などを測ることができる。

◆ 短所 : 受験者間で行う場合、発話が受験者間のレベルに大きく影響される。

(7) **質疑応答 (question & answer)、討論 (discussion)** : 質問に対して口頭で答える対面式のテスト。短い時間に適切に応えられるかという即時性、適切さ、及び流暢さが測定される。以下3つのバリエーションがある。

(a) 面接形式 (interview) : 面接官と受験者の1対1で行う。

(b) 集団面接 (group interview) : 面接官1人 vs. 複数の受験者で行う。

(c) 集団討論 (group discussion) : 受験者同士が話し合いを行う。

→集団討論では、積極性や協調性なども測定される。

(8) **絵の描写 (picture description, story telling)** : 絵を口頭で説明したり、4コマの絵のストーリーを作って話すタスク。

→英検やアルク社のSST (Standard Speaking Test)でも採用されている、スピーキングテストでは最もよく使われるタスクの1つ。

(9) エッセイ (essay)、作文 (composition) : 与えられたトピックに対して、ある程度の長さの文章を書かせる形式。

- ◆ 長所 : 内容、文法、構成、作文の技術的側面 (e.g., 意見の展開の仕方)などを総合的に見ることができる。
- ◆ 短所 : ・トピックの熟知度に影響されやすい。
 - ・採点に採点者の主観が入りやすく、時間もかかる。

(10) タスク活動 (task-based activities) : 主に教室内活動として使用されるさまざまなタスク活動がある。

- ◆ インフォメーション・ギャップタスク (information gap task) : ペア (あるいは数名) でどちらか一方しか持っていない情報を、お互いに質疑応答しながら補完させていく形式。積極性、問題解決能力を見ることができる。

このように項目にはさまざまな形式があるので、それぞれの特徴や長所、短所を把握し、複数の項目を組み合わせることで目的に合ったテストを作成することが大切である。