

日本テスト学会講習会

データに基づくテスト分析とテスト作成、新しい時代のテスト
—教師力向上のためのヒント—

新しい時代のテスト

2016年10月30日

熊谷龍一(東北大学大学院教育学研究科)

本発表の目的

CBT および IRT についての
概要を説明する。

高校生の基礎学力の定着に向けた学習改善のための研究開発事業 別紙 4

平成29年度概算事業額 288百万円
平成28年度予算額 109百万円

1 目的

「経済財政運営と改革の基本方針2016(平成28年6月)」及び「教育再生実行会議第9次提言(平成28年5月)」で示された高校教育・大学教育と入学者選抜を通じた高大接続改革を推進するため、高校生の学習意欲の喚起とともに基礎学力の着実な定着を目指し、高等学校基礎学力テスト(仮称)の導入検討等に取り組むこととする。この観点から、具体的には、

- ① 学校現場においてきめ細やかな学習指導に取り組むための指導体制の在り方の検討や教材開発等とともに、
- ② 上記指導を受けた生徒の基礎学力の定着度や学習状況を的確に把握し、更なる指導改善に活かすためのテスト手法等に関する仕組や実施方法等を調査研究することをもつて、高等学校現場におけるPDCAサイクルの確立を目指す。

2 実施内容(委託事業 対象:都道府県教育委員会等)

(1) 高等学校基礎学力テスト(仮称)の導入を念頭に置いた基礎学力の着実な定着に取り組む実践研究における調査研究

- 基礎学力の定着に向けた学校現場の専門性向上の策定
- 指導改善に向けた教材開発計画の策定・変更 など
- 2 階層の教員や支援スタッフ等からなる指導体制の構築
- 「行動」(教、習、練、観)等の学習活動の実施 など

(2) 基礎学力の定着度を把握するためのテスト手法等の開発

- 1 試行実施に向けてのアンケート調査等に関するテスト手法の開発
- 2 テスト実施手法に関する研究開発
- 3 試験問題の作成・収集・検査・選別 等

図1 テストのイメージ

試験作成 → 委託先 → 試験の送付 → 試験実施 → 結果の送付 → 指導改善

試験実施のイメージ

試験の実施方式やテスト内容の課題(CBT及びIRT導入の実現可能性、記述式問題の作成、英語の「話す」に関する出題や解答方法、採点方法、結果の指導改善への活動方策など)を洗い出し、技術的な課題とその解決方法を探るため、上記取組を通じて実証的・専門的な検討に資するデータ収集や分析等を行う。

文部科学省(2016)高大接続改革の進捗状況について
報道発表日 平成28年08月31日

文部科学省(2016)高大接続改革の進捗状況について
報道発表日 平成28年08月31日

CBT と IRT って何？

CBT … Computer Based Test

IRT … Item Response Theory

② CBT、IRTの活用

CBT、IRTの活用については、現時点では安定的・継続的に活用可能と判断できる段階ではなく、引き続き専門的・技術的な研究・検討が必要。当面はPBT(紙による実施)を基本とする。

文部科学省(2016)高大接続改革の進捗状況について
報道発表日 平成28年08月31日

Computer Based Test

コンピュータを使ったテスト



7

CBTの特徴

- データ管理, 採点の自動化
- 印刷, 回収などのコスト
←→ 開発, 維持コスト
- マルチメディア活用
- 適応型テスト

8

適応型テスト

- 受験者の解答状況に応じて, その受験者に適切な問題を出題する。

9

適応型テスト

メリット

- 少ない項目数(短い時間)で精度よく測定できる。
- 受験者への負担少(難しすぎず, 易すぎず)。

10

適応型テスト

受験者ごとに違う問題が提示される。

→ 不公平は起こらないのか?

IRT

11

IRT (Item Response Theory)

IRTの前に…

一般的なテスト得点の算出

正答した項目数の合計

(各項目には配点が付されていることが多い)

「合計得点」

12

偏差値

合計得点が80だった

だけではよく分からないので

受験者集団の中での位置を示す指標

$$\text{標準得点} = \frac{\text{合計得点} - \text{平均}}{\text{標準偏差}}$$

$$\text{偏差値} = \text{標準得点} \times 10 + 50$$

13

合計得点だけじゃだめなの？

14

こんなことを考えてみよう！

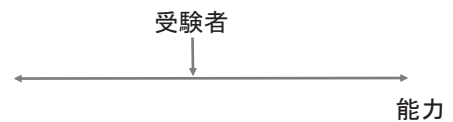
1. テストで測りたい能力を左右方向の直線で表す。



15

こんなことを考えてみよう！

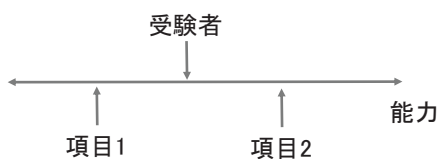
2. 受験者を能力水準に応じて直線上の一点に位置付ける。



16

こんなことを考えてみよう！

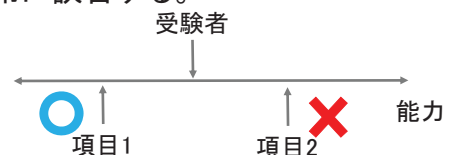
3. 各項目を困難度に応じて直線上の一点に位置付ける。



17

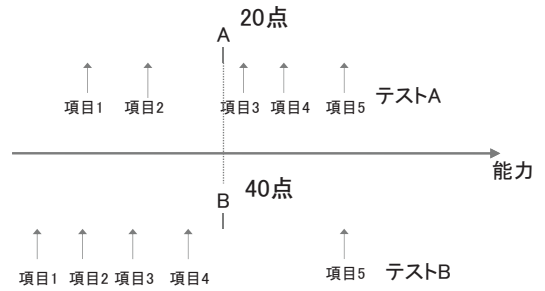
こんなことを考えてみよう！

4. 自分の位置よりも左側にある項目には常に正答し、右側にある項目には常に誤答する。



18

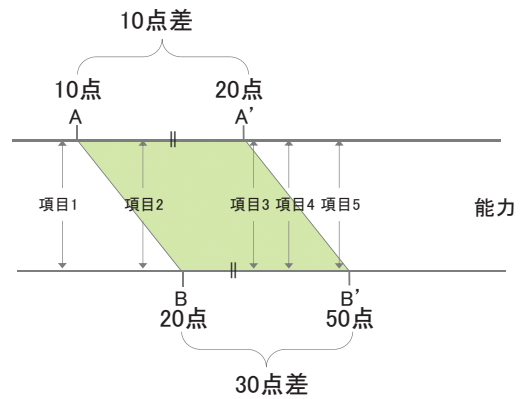
- 能力が等しくても, 受けるテストが異なると, テスト得点が異なる!



19

20

- 同じテストを受けて, 同じくらい能力が伸びていても, その伸びが公平にテスト得点に表れないことがある!

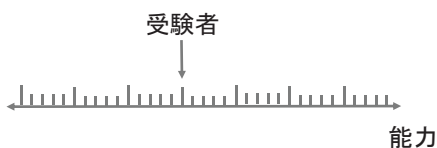


21

22

解決策は...

- 能力の直線に, 目盛りをつけてあげればいい!



23

項目反応理論 (IRT)

- この直線上で示された数値が項目反応理論(IRT)で「能力」を示すものになります。
- 「潜在特性尺度値」, 「特性値」, 「尺度値」, 「 θ 」などと呼ばれます。

24

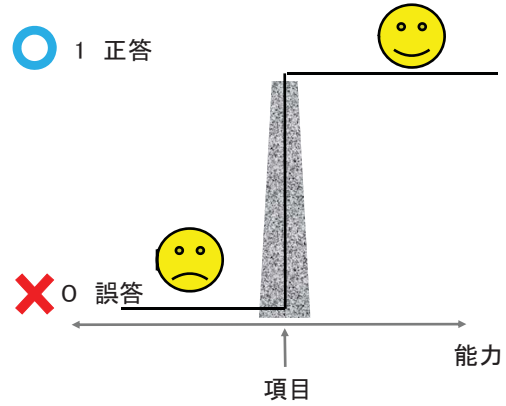
ただし…

4. 自分の位置よりも左側にある項目には常に正答し、右側にある項目には常に誤答する。

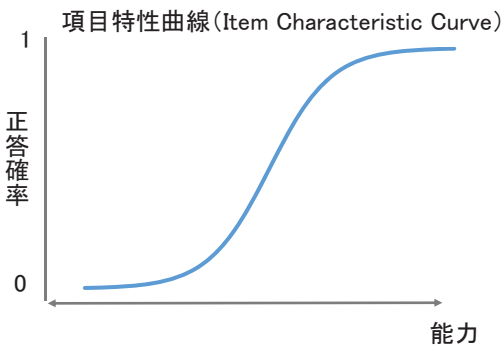


4. 自分の位置よりも左側にある項目ほど正答しやすく、右側にある項目ほど誤答しやすい。

25



26



27

項目特性関数

$$P(\theta) = \frac{1}{1 + \exp\{-a(\theta - b)\}}$$

潜在特性尺度(能力) ↓
 ↑ 問題の識別力 ↑ 問題の困難度

2パラメタ・ロジスティック・モデル

28

どうやってパラメタを求めるの？

- コンピュータ(専用ソフトウェア)を利用した数値計算
 - BILOG-MG, IRTPRO
 - WinStep, Facets
 - ConQuest
 - EasyEstimation(熊谷) …

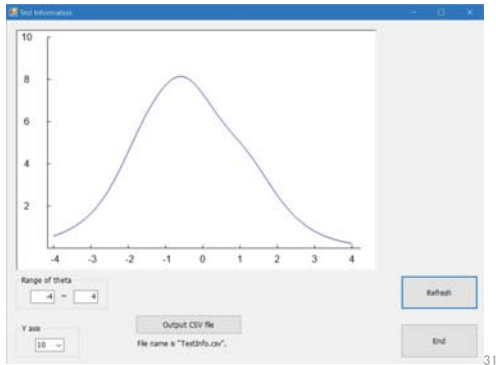
29

IRTのメリットは？

- より詳細なテスト精度(信頼性)の表示
- 等化による複数テストの比較
- コンピュータ適応型テスト
(Computerized Adaptive Test: CAT)

30

テスト情報量



31

適応型テストにおいて

どの問題を出題するのか

- 易しすぎず
- 難しすぎず

「正答できるかどうか」

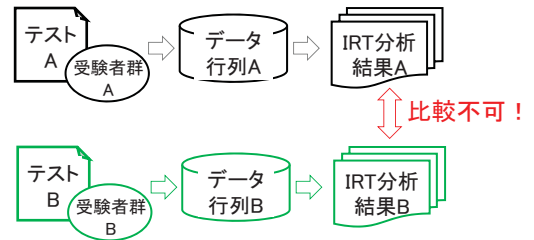
32

等化 (equating)

テスト得点を
実施時期・実施地域などが異なる
複数の版で比較可能にする
→ IRTとの親和性
(必ずしもIRTでなくてもできる！)

33

等化をするためには

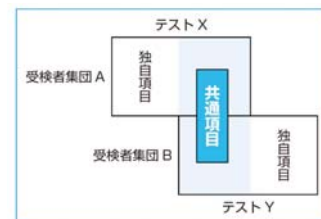


34

等化をするためには

- 共通な項目群
- 共通な受験者集団

共通項目デザイン

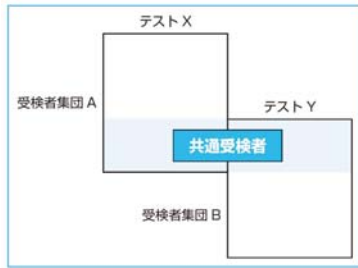


出典: 熊谷龍一・荘島宏二郎 2015 教育心理学のための統計学
—テストでココロをはかる— 誠信書房

35

36

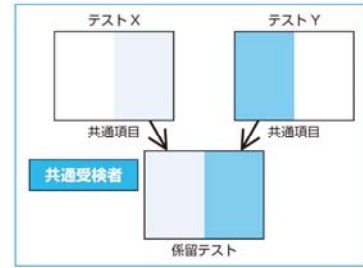
共通受験者デザイン



出典:熊谷龍一・荘島宏二郎 2015 教育心理学のための統計学
-テストでココロをはかる- 誠信書房

37

アンカーテストデザイン



出典:熊谷龍一・荘島宏二郎 2015 教育心理学のための統計学
-テストでココロをはかる- 誠信書房

38

等化用のデータを収集したら

- 等化の計算
 - 等化係数の推定
 - 等化係数以外の方法での等化
 - 項目固定法
 - 共時等化法

39

IRTの前提条件

- 尺度の1次元性
テストで測定しようとしている能力が「1本」
- 項目間の局所独立
 - ある項目の正誤が、他の項目の正誤に影響を与えない

40

IRTを用いる際の注意点

- 基礎分析をしっかり行なう。
 - 1次元性のチェック
- サンプルサイズは？
- 局所独立は成り立っているか？
- 変な数値はないか？

41

IRTを用いる際の注意点

- 等化(デザイン)をどのように実現するか。
 - 項目露出, 初出主義
 - コスト
- 教育測定・心理測定の専門家
そもそもIRTが必要ですか？

42