

日本教育心理学会第67回総会
会員企画シンポジウム

テスト項目の作成と統計分析
—多枝選択式・記述式と認知診断—

項目作成ガイドラインの作成と適用

2025年10月12日 (日)

石井 秀宗

名古屋大学大学院教育発達科学研究科

- いたるところに「テスト」は存在する
- ときに人生を大きく左右する
- 資格試験 : 資格を付与するか否かを判断
- 入試 : 入学を許可するか否かを判断
- 適性検査 : 適性の質や程度を評価
- 実力テスト : 広い範囲の学習の理解度を評価
- 定期考査 : 当該期間の学習の理解度を評価
- 単元テスト : 当該単元の学習の理解度を評価
- 発問 : 当該事項の理解度を把握, 授業の前準備
- 模試 : 各受検者の相対位置の把握, 試験への慣れ
- 学力調査 : 全体の実態把握, 学力像の提示, 指導改善

- 教職課程において、テスト作成に関する内容は殆ど扱われない
- 多くの教員は、テスト作成に関する教育をほとんど受けていない
- 現場に出てはじめて、自身の教育経験や先行例に基づいてテストを作成する
- テストとはどのようなものか、どのように作ればよいか、得点は何を反映しているか等がきちんと理解されていない
- よいテストが作られている例
 - 予備校等にはテスト作成のプロがいて、よい問題を作れる
 - テスト作成や実施、採点に関するマニュアルがある
 - しかし、それらが一般（学校教員等）に出回ることではない

テストの得点は何を反映しているか

- 受検者の能力

- そのテストで測っているものの程度
- 「測っているもの」と「測りたいもの」の違い

妥当性
信頼性
テスト理論

- 得点化法

- 配点, 部分点の有無, 能力値推定法
- 採点基準, 採点者の違い

項目分析
ガイドライン

- 問題の構造・性質

- 題意の分かりやすさ, 問題形式, 字数制限, 魅惑枝

- 実施方法, 実施環境

- モチベーション, テスト不安, 集中しやすさ

教育心理学

- テスト得点にはさまざまな要因が影響している

• テストの評価

- 作成されたテストについて，教科の専門家による所見等の評価はなされる
- 予備校の入試問題分析が，受検者の実際の反応と異なることも多い
- 回答データに基づいたテストの分析は殆どなされない
- なされたとしても，その後のテスト作成に活かされることは少ない

• テストの品質管理ができていない

- 作成技術の蓄積がなされず安定しない
- 設計図が曖昧なことも多い
- 品質検査が不十分である
- 品質管理の必要性の認識が十分でない

- よいテストを作成・実施するために必要なこと
 - 作成技術の獲得・向上, 品質検査・管理
- テスト問題の作成に関する指針が必要（ガイドライン）
 - 一般的な注意事項をまとめたもの
 - マニュアルではなく, もう少し抽象的なもの
- テスト問題を回答データに基づいて分析することが必要（項目分析）
 - 問題は受検者にどのように理解され, どのように反応されたか
 - 選択枝や解答類型の機能を分析
 - 問題の改善法を検討

- 実験的に収集した回答データに項目分析を適用し、テスト問題やガイドラインの性質を実証的に検討
- 検討結果に基づいてガイドラインを策定
- 項目分析ツールの開発
- 誰でも利用できるように研究成果をWebサイトで公開

- ① 目的が明確で適切である
- ② 適切に作成されている
- ③ 適切に実施されている
- ④ 正しく採点されている
- ⑤ 適切に運用されている (テスト研究ウェブサイトより)

- 本発表では主に作成・実施に着目しているが、
- 目的, 運用も含めてテストを考えることが重要

- 測りたいものを測っている
- 識別力が高い
 - 測りたい能力が高い受検者ほど正答する
 - 測りたい能力が低い受検者ほど誤答する
- 正答率が適度
- 無答率が低い
 - 何を答えたら良いのかが児童生徒に伝わっている
- 多くの選択枝が機能している
- 授業と問題がつながっている
 - 復習することにより，どうして誤答したのか受検者が理解し，学びを進めることができる
 - 受検者のつまずきを把握し，指導改善に役立てることができる

- 測りたいものを測っていない
- 識別力が低い
 - 測りたい能力が高い受検者ほど正答するとは限らない
 - 測りたい能力が低い受検者ほど誤答するとは限らない
- 無答率が高い
 - 何を答えたら良いのかが児童生徒に伝わっていない
- 少ない選択枝しか機能していない
- 授業と問題がつながっていない
 - 復習しても、どうして誤答したのか受検者が理解できず、学びが進まない
 - 受検者のつまずきを把握できず、指導改善の役に立たない

- 測りたい能力を測っているかいなか
- 何を測りたいかで決まるものではない
 - 何を測りたいかはテストの目的による
 - 思考力・表現力を問うことを狙いとしている問題が良い問題という訳ではない
 - 知識や公式をあてはめる力を測る問題が悪い問題という訳ではない
- 難しい(易しい)問題が良い(悪い)問題という訳ではない
- 測りたいものを実際に測れているかがポイント

- 別添資料参照
- 構成（39項目）
 - 問題の内容（6）
 - 問題の形式（4）
 - 問題の記述（6）
 - 設問部分（4）
 - 選択枝（14）
 - 記述式問題の採点に関して（5）
- ガイドラインに基づく修正例なども紹介

8 前の問題に対する解答が、後の問題の正誤に影響しないこと

× 次のア, イの値を求めよ。

$$(1) 3^2 + (-2)^2 = \text{ア}, \quad (2) \text{ア} \times (-3) = \text{イ}$$

正答(13, -39)

(2)の問題は正負の数の掛け算ができるかを問う問題であるが、(1)に誤答すると、正負の数の掛け算ができる受検者も誤答になってしまう。(1)と(2)の順番も変えたほうがよい。

○ 次のア, イの値を求めよ。

$$(1) 12 \times (-3) = \text{ア}, \quad (2) 3^2 + (-2)^2 = \text{イ}$$

正答(-36, 13)

14 問いたいことは何かを設問に書くこと 15 設問部分だけで問題の意味が分かること



MAKE NEW STANDARDS
東海国立
大学機構



名古屋大学

× H_2O

1. 水素 2. 酸素 3. 水 4. 思い出がいっぱい

何を問うているのか全く分からない

不適切な選択枝もある

○ 分子式 H_2O で表される物質の名前を次の中から
ら選べ。

1. 水素 2. 酸素 3. 水 4. 二酸化炭素

16 空所補充問題について、文意が分からなくなるほどの空所を設けないこと

× 以下の空所に入る語句を選択枝の中から選び、数字をマークせよ。

(1) を合計して (2) で割った値を (3) 値と言う。(1) を大きさ順に並べたとき、ちょうど (4) にくる値を (4) 値と言う。(2) が (5) のときは、(4) の2つの (1) の (3) を (4) 値とする。その値の (2) が最も大きい (1) の値を (6) 値と言う。

各 (1) から (3) 値を引いて (7) したものの (3) を取った値を (8) と言う。(8) は (3) 値と (9) が異なっている。そこで、(8) の正の平方根を算出する。これを (10) と言う。(8) も (10) も、(1) を (7) しているので、(11) 以上の値になる。

選択枝

- | | | | | | | |
|-------|--------|------|-------|------|------|-------|
| ① 0 | ② -1 | ③ +1 | ④ -2乗 | ⑤ 2乗 | ⑥ 奇数 | ⑦ 偶数 |
| ⑧ 個体数 | ⑨ 最小 | ⑩ 最大 | ⑪ 最頻 | ⑫ 時間 | ⑬ 次元 | ⑭ 中央 |
| ⑮ データ | ⑯ 標準偏差 | ⑰ 分散 | ⑱ 平均 | ⑲ 範囲 | ⑳ 標本 | ㉑ 母集団 |

パズルを解くような問題

内容理解ではなく、教科書丸暗記の問題

選択枝を探すのに手間取ってしまう

16 空所補充問題について、文意が分からなくなるほどの空所を設けないこと

- 平均値の説明として正しいものを選べ。
 - A. データを合計して個体数で割った値
 - B. データを大きさ順に並べたとき、ちょうど中央にくる値
 - C. データを大きさ順に並べたとき、中央の2つのデータを足して2で割った値
 - D. 個体数が最も大きいデータの値

- 標準偏差の説明として正しいものを選べ。
 - A. データから平均値を引いて2乗したものの平均を取った値
 - B. データから平均値を引いて2乗したものの平均の正の平方根の値
 - C. すべてのデータが含まれる区間幅の値
 - D. データを大きさ順に並べたとき、中央50%のデータが含まれる区間幅の値

正答 (A), (B)

項目分析の例

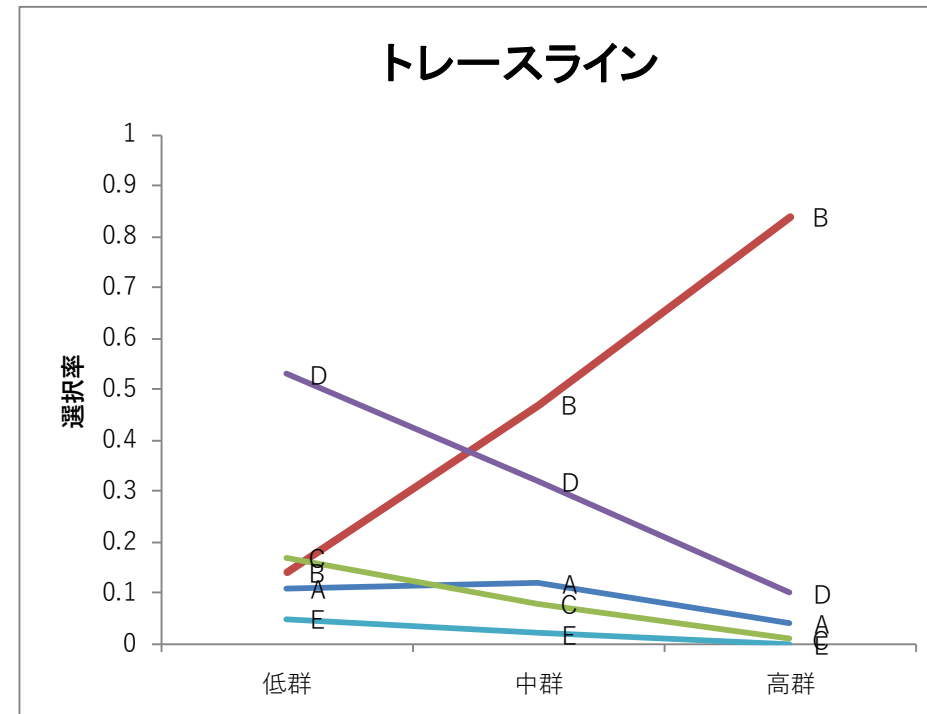
Q. 対角線の長さが4cmの正方形の面積を以下から選びなさい。

A. 4cm^2 B. 8cm^2 C. 12cm^2 D. 16cm^2 E. 20cm^2

受験者数	正答率	D値	I-T相関	α 係数	削除 α	正答
200	48%	70%	0.42	0.70	0.67	B

選択率	A	B	C	D	E	無答
全体	10%	48%	9%	32%	1%	0%
高群	5%	84%	1%	10%	0%	0%
中群	12%	47%	8%	32%	2%	0%
低群	11%	14%	17%	53%	5%	0%

識別指標	A	B	C	D	E	無答
D値	-6%	70%	-16%	-43%	-5%	0%
I-T相関	-0.05	0.42	-0.15	-0.29	-0.12	0.00





- 主なコンテンツ
 - テストの問題点
 - テストの作成
 - 項目作成ガイドライン
 - テストの評価
 - 項目分析
 - 項目分析システム
 - 項目応答理論 (IRT)

<https://www.educa.nagoya-u.ac.jp/~ishii-h/test.html>