

テスト作成

構成概念を適切に測定する必要性

- 教育学や心理学では構成概念を多く扱う
- 構成概念を直接、正確に測定することはできない
- できないならやらなくていい、とはならない
- 経験、信念、信条、べき論、理想論などに基づく議論には客観性、正確性、一般性に乏しく、説得力が弱い
- 教育学や心理学を研究するためには、構成概念をより良適切に測定することが必要

テストや心理尺度の作り方に関する知識や技術

テストや心理尺度の評価

試験問題や質問項目の評価

試験の妥当化の必要性

- テストで「学力」や「能力」を正確に測定・評価することはできない
- テストなどしなくてよい、で済むならならそれがいい
- そうでないなら、可能な限り正確に測定・評価することが必要
- できるだけ良い試験を実施することが求められる

構成概念の測定に関する教育・研究の必要性

- 教員養成課程において、テストとや心理尺度の作成，実施，使用，分析等について，体系的に教えられていない
- テストや心理尺度を作成，利用しているのに，その性能や品質はほとんど議論されず，結果だけが独り歩きする
- 工業製品で言えば，設計図もなく，性能検査も品質管理もしていない製品を，生産・使用している状態
- テストや心理尺度の作成，実施，採点，使用等に関する教育・研究が必要

構成概念の測定の難しさ

- 間接測定なので誤差が大きい
- 観測値がゼロでもその特性が「無い」とは限らない
- 値がいくつならどの程度というスケールがない
- 構成概念は頭の中で考えた抽象的なものであり，人によって定義や捉え方が異なる
- 測りたい構成概念を，確かに測っていると多くの人が納得できる証拠を提示することが必要
 - 尺度・テストの評価：測定の妥当性・信頼性
 - 項目・問題の評価：項目分析

用語説明

- 「多肢」を「多**枝**」，「受験」を「受**検**」と書くことが増えている
- テストにおいて「問題」「項目」「設問」はほぼ同義なので，区別しないで用いられることも多い
- 「解答」は，「問題に答えること」と「正解」の意味があるため，ここでは，問題に答えることを「回答」，正解を「正答」とする

良いテストの要件

良いテストの要件

- 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
-
- これらの1つでも欠けると、その試験すべてが適切でないものになる
 - 目的、運用も含めてテストを考えることが重要

① 目的が明確で適切である

- 何のためのテストか
資格認定，選抜，成績評価，学習支援 など
- 何を測れば目的を達成できるか
テストで測るべき能力に含めるもの／含めないものは何か
- 何を問えばその能力を測定できるか
内容の理解についての問題を単なる暗記で解けるようなものにしない
- その能力を測ることに意義はあるか
難易度を上げるためとして重箱の隅をつつくような問題を出題しない

② 適切に作成されている

- 測りたいものを測れるものになっているか
何が問われているか，何について回答すれば良いかを受検者が理解できる
- 測りたい能力以外の要素が得点に影響しないか
特定の受検者に有利（不利）な内容になっていないか など
- テスト仕様書，ブループリント，問題作成ガイドラインの利用
設計，内容，形式，言語レベル，分量，採点基準 など
- 問題作成に有効な作業
声に出して読む（不自然な箇所が顕わになる）
一度作成したあと，忘れたところに見直す

作題者と受検者の違い

- パズルは1回目は難しいが、一度解いたあとは簡単になる
- 正答を知っている作題者は、簡単に感じる
- 問題内容が頭に入っているので、問題文が分かりにくくても理解できる
- 正答を知らない受検者には、難しい問題
- 問題内容が頭に入っていないので、問題文が分かりにくいと、理解するのが困難
- 問題を忘れたところに、受検者と同様の状態で問題を読む
問題の意味を理解できるか、回答条件は整っているか、
正答は正しいか、誤字脱字はないかなどを確認

③ 適切に実施されている

- テストの目的を受検者は理解しているか

受検者が適切に能力を発揮して回答することが必要

- どの受検者も公平で適切な環境下で受検できているか

公平性，公正性の確保

- 測りたい能力以外の要素が得点に影響しないか

タブレットを使う・使わないや，実施方法や用具の違いにより，得点が異なることはないか など

④ 正しく採点されている

- 測りたいものを測る採点基準になっているか
合理的な、正答、誤答、減点、加点の基準を設定する
基準の根拠を示す
- 基準に従って採点されているか
出題者の意図に沿った回答ばかりとは限らない
採点者が基準の意味を理解していることが必要
- 記述式回答を精確に採点するのは難しい
人が段階評定できるのはおよそ 4 ～ 5 段階
採点者間のズレ → トレーニングしてなるべく調整
採点者内のブレ → 採点が安定してきたら再採点

⑤ 適切に運用されている

- 試験の目的は達成されているか
資格認定，選抜，成績評価，学習支援ができている
- 適切な結果が得られているか
試験を用いて導かれる結果が理にかなっていることが必要
- 公平・公正な解釈，運用をしているか
特定の集団や個人を有利（不利）に扱ったりしない
- 拡大解釈，過度な一般化，目的外利用をしていないか
得点は何を反映し，何を反映していないかを理解して運用

良い問題・悪い問題

良い問題とはどのような問題か？

- 測りたい能力を測っている
- 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

悪い問題とはどのような問題か？

- 測りたい能力を測っていない
- 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
- 測りたい能力が高い受検者ほど正答する， とならない
- 測りたい能力が低い受検者ほど誤答する， とならない

良い問題・悪い問題の考え方

- 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ってい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

テスト作成の準備

世の中にあるテスト・検査

- 入学試験（大学入学共通テスト，個別学力検査，etc.）
- 資格試験（医師，公認心理師，保育士，etc.）
- 能力試験（英検，TOEFL，珠算，漢字，etc.）
- 検定試験（〇〇検定，etc.）
- 学力試験（全国学力・学習状況調査，定期テスト，etc.）
- クラスルームテスト（単元テスト，小テスト，etc.）
- 実技試験（楽器演奏，ダンス，etc.）
- 適性検査（SPI，職業適性検査，中学入試，etc.）
- 知能検査（ウェクスラー式，ビネー式，etc.）
- 発達検査（新版K式，遠城寺式，etc.）
- 性格検査（Big5，MMPI，etc.）
- etc.

テストの目的と分類

目的	実施形態	実施主体	実施規模	出題範囲	フィード バック先
確認	発問	教師	クラス	トピック	児童生徒学生 教師
学習支援	小テスト	教師	クラス	トピック 単元	児童生徒学生 教師
成績評価	定期テスト	教師	クラス 学校	当該期間の学習 内容	児童生徒学生 教師
資格認定	資格試験	認定機関	受検者集団	資格取得に必要な 内容	受検者
相対評価	模試	実施団体	受検者集団	学習内容全般	受検者
選抜	入試	学校, 自治体 テスト機関	学校, 自治体	学習内容全般	学校 受検者
教育施策	調査	自治体, 国	自治体, 国	学習内容全般	自治体, 国

テストの設計 (仕様書)

- テストを作成・実施するにあたって事前に明確にしておくべきこと

- ①テストの目的
- ②対象者
- ③単元・領域
- ④能力(認知様式)
- ⑤解答形式
- ⑥所要時間
- ⑦項目数
- ⑧実施方法
- ⑨必要機材
- ⑩採点のルール など

項目	設定内容
テストの目的	学期末単位認定試験
対象者	学部2年生科目「心理学統計法」受講者
測定領域	「心理学統計法」で扱う以下の単元 記述統計, 推測統計, テスト理論
能力(認知様式)	知識: 用語説明など 適用: 統計量の算出など 応用: 分析結果の解釈など
解答形式	知識: 短答法 適用・応用: 短文法, 論述法
所要時間	60分
項目数	30問 (テストの青写真参照)
実施方法	用具: 紙筆式 形態: 集団実施
必要機材	とくになし
採点のルール	短答法問題は1問2点, 正誤評価 短文法問題は1問4点, 段階評価 論述法問題は1問5点, 段階評価 文章構成力も評価する 誤字・脱字は減点しない 判読不能解答は無解答として扱う 合計点を換算し, A+, A, B, C, C-, Fで評価
留意事項	ノート・教科書等持ち込み不可

テストの青写真 (Blue Print)

- 単元，能力，項目数の関係を表にまとめたもの
- 能力（認知様式）は，問題を解くにあたって必要な能力の種類を特定するもの
- 最近では「知識・記憶」「適用」「応用・問題解決」の3領域に分けて考えることも多くなっている
- 全体の項目数や単元ごとの項目数，また各能力への配分は，試験時間や学習内容，解答形式などを考慮して，テスト作成者が決める

単元	能力（認知様式）			小計
	知識	適用	応用	
記述統計	4	1	1	6
推測統計	4	2	2	8
テスト理論	3	1	2	6
小計	11	4	5	20

問題形式

- 論文式をパフォーマンスに分類することもある

大分類	中分類	問題形式
選択式	一枝評価式	真偽式
		二値評価式
	多枝選択式	択一式
		複数選択式
	多枝利用式	組合せ式
		並べ替え式
構築式	記述式	穴埋め式
		短答式
		論述式
		論文式
	パフォーマンス式	口述
		実技・実演
		作品

問題作成ガイドライン

問題作成ガイドライン

- 良い問題を作るための一般的な知識やスキルをまとめたもの
- マニュアルではない
- 経験や先例も重要
- 個々の問題において、すべてのガイドライン項目を満たさなくても良い
- その問題で測りたい能力は何か、テスト全体では何を測ろうとしているのかを基準に考える
- 1つの問題で複数の能力を測ろうとすると、能力を細かく識別できない「悪い問題」になりかねない → 部分点の付与

問題の内容

- 1 問いたいことは何か，問題を解くために必要な能力は何かが明確であること
- 2 重要な事柄を問うこと。些末なことや一般的過ぎる問いにならないこと
- 3 回答や正答が，受検者や作題者の個人の意見に左右されないこと
- 4 ひっかけ問題にならないこと
- 5 高次の能力を測る問題では，受検者にとって新奇的な素材を用いること

問題の形式

- 6 測りたい能力に見合った問題形式を用いること
- 7 前の問題への解答が、後の問題の正誤に影響しないこと
- 8 「あてはまるものをすべて選べ」という設問は避けること。使う場合は部分点を与えること

問題の記述

- 9 教示・本文・設問・選択枝・図表等の記述量を必要最小限にすること
- 10 言語レベルを受検者集団に合わせることに
- 11 教示・本文・設問・選択枝・図表等の文言をよく校正すること。他の人にも確認してもらうのが望ましい
- 12 教示・本文・設問・選択枝・図表・解答欄等のレイアウトや大きさを適切にすること
- 13 とくに低学年の児童に対して、選択枝は行を変えて1つずつ並べること

設問部分

- 14 問いたいことは何か，どのような形式で解答したらよいかを明確・簡潔に書くこと
- 15 本文や選択枝など他の部分を読まなくても，設問部分だけで問題の意味が分かること
- 16 否定表現を使わないこと。もし使う場合は，**太字**やアンダーラインで強調すること
- 17 特定の個人や集団に有利または不利な内容にならないこと
- 18 受検者全員には理解されない暗黙の前提を用いないこと

選択枝について

- 19 いずれの選択枝ももっともらしいこと
- 20 識別力の高い選択枝を用いること。不必要に選択枝を多くしないこと
- 21 明らかな誤答枝や、おあそびの選択枝など、余計な選択枝を入れないこと
- 22 正答枝と誤答枝が明確に区別できること
- 23 音順，数量の大きさなど，何らかの法則に従って選択枝を並べること
- 24 正答枝の位置をランダムにばらつかせること

選択枝について

- 25 「上記のいずれでもない」「上記すべてあてはまる」などの選択枝を用いないこと
- 26 「～でない」「～以外である」など否定表現を用いないこと
- 27 「絶対に」「常に」「決して」「完全に」など、強意語を用いないこと
- 28 選択枝は互いに独立であること。内容に重なりがないこと
- 29 一方が正答枝であれば他方は誤答枝であると分かるような、両立しない選択枝を入れないこと
- 30 選択枝の長さをおおむね揃えること
- 31 選択枝の内容や形式の構造を揃えること

記述式問題の採点に関して

- 32 問題作成と同時に，評価の観点，採点基準を設定すること
- 33 評価の対象とする要素，しない要素を明確にすること（誤表記は減点する，乱筆は減点しない，など）
- 34 正答となるものについて，基本的な基準を設定し，あまり細かな条件設定はしないこと
- 35 本格的に採点を始める前に，いくつかの解答を採点して，採点基準を精錬すること
- 36 最初のほうで採点した解答を後で採点し直すなどして，評定の一貫性を保つこと

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)

あてはまるものをすべて選ぶ問題の問題点

- あてはまるものをすべて選ぶ問題は，思考力を問う良い問題だと言う意見もある
- その主張の根拠はどういうものか？
- この形式の問題の問題点はどういうものか？

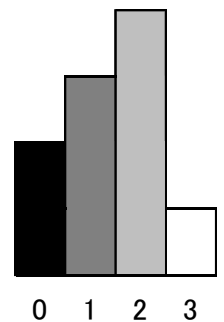
9 「あてはまるものをすべて選べ」という設問は避けること。使う場合は部分点を与えること

- 「すべて選べ」問題では、受検者を少数の正答者と大多数の誤答者の2群に分割する
- 完全に理解していないと正答できないので一見「良い問題」のように思われるが、実際は受検者の能力を細かく識別しない「悪い問題」である
- テストに対する受検者の不安・不信が増大する
- 部分点を与えたり、それぞれの選択枝の真偽を回答する問題にすれば、受検者の能力をより細かく測ることができる

完全正答方式



部分点方式



- 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
-

× 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, (1 1) 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。

正答(A)

- 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。

正答(B)

- 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

24 不必要に選択枝を増やさないこと

- 適切な選択枝の数は、多枝選択式問題なら3~5個
- 空所補充問題では、受検者が選択枝を一覧するのに困難を伴わない程度
- あて推量を避けるためとして選択枝を多くするのは、弊害のほうが大きい
- 受検者は完全なあて推量はしない
- 選択枝や、マークを塗る欄を探すという、非本質的なことに多くの時間を割かれる
- 選択枝を作成するのに苦勞する
- 出題ミス、選択ミス、マークミスのリスクが高くなる

26 否定表現を用いないこと

29 両立しない選択枝を入れないこと

× 不老町の説明として正しくないものをすべて選べ。

1. 愛知県には属さない
2. 名古屋市に属さない訳ではない
3. 千種区に属しはしない
4. 昭和区に属さないとも限らない

正しくないものを選ぶうえ、選択枝の中に否定表現があるので、とても分かりにくい。

正しくなものが何個あるか示していない。

愛知県、名古屋市、千種区と昭和区などの包含関係が、回答のヒントになり得る。

○ 不老町が属する自治体を以下の中から選べ。

1. 昭和区
2. 千種区
3. 天白区
4. 名東区

項目分析

試験問題の評価の必要性

- 「してやったり」とほくそ笑む作題者は、一体何を測定したいのだろうか？
- 正答を答えられない生徒が悪いのではなく、そんな問題を作る作題者のほうが悪いのでは？
- 「専門家」の主観的な講釈も結構だが、「専門家」と受検者との感覚がズレていることも多々ある
- **実際の回答データに基づいた実証的な検討が必要**

個々の問題の評価

- 試験を構成する個々の問題の性能が低ければ、試験全体の性能も低くなる
- 個々の問題に対する受検者の反応を調べることで、その受検者の理解度をより詳しく把握することもできる
- 個々の問題の性能を評価する必要がある

→ 項目分析 (Item Analysis)

問題例

Q. 対角線の長さが4cmの正方形の面積を以下から選りなさい。*

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

- 30問からなる算数のテストの中の問題
1問 1点, 平均 15.1点, 標準偏差 4.5点

* 例示のため多枝選択式問題としている

テスト得点の分布

- 平均正答率 50%

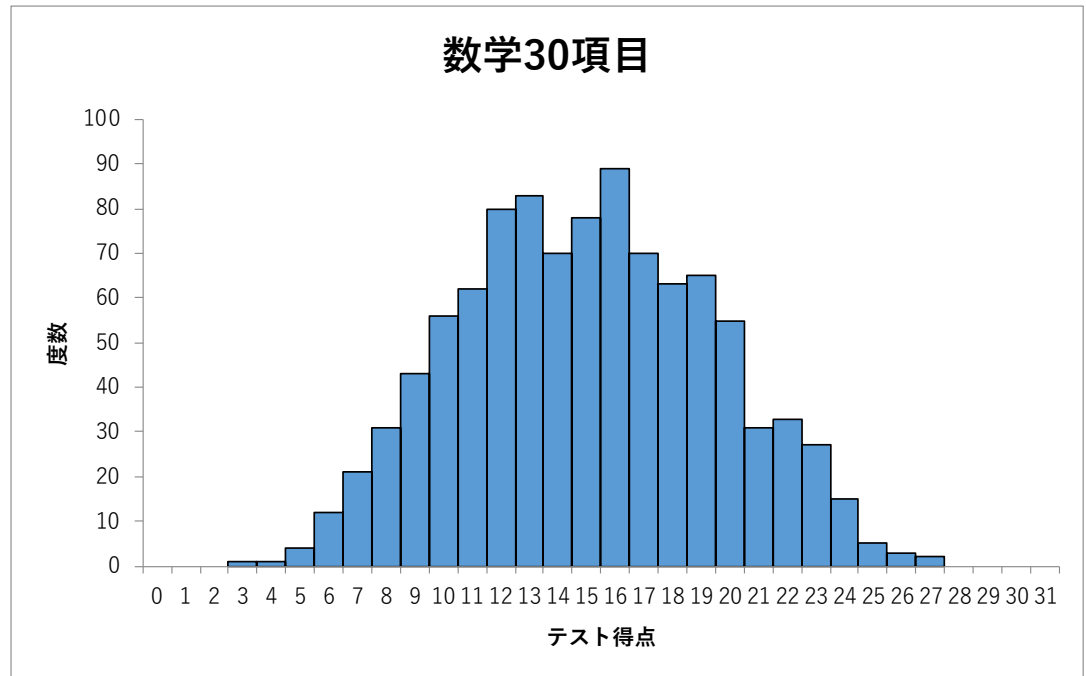
テスト得点	項目数	受験者数	平均	標準偏差	最小値	中央値	最大値	α 係数
得点	30	1000	15.01	4.45	3	15	27	0.69
得点率	30	1000	50%	15%	10%	50%	90%	0.69

- 標準偏差 15%

- 最小値 10%

- 最大値 90%

- α 係数 0.69



項目分析表

項目	受験者数	正答率	D値	I-T相関	α 係数	削除 α	正答キー	A	B	C	D	E	無答
Item1	1000	78%	34%	0.23	0.69	0.69	E	7%	7%	5%	3%	78%	0%
Item2	1000	73%	40%	0.24	0.69	0.68	A	73%	3%	2%	3%	20%	0%
Item3	1000	94%	9%	0.12	0.69	0.69	B	1%	94%	2%	2%	2%	0%
Item4	1000	84%	22%	0.17	0.69	0.69	B	10%	84%	3%	2%	2%	0%
Item5	1000	48%	70%	0.42	0.69	0.67	B	10%	48%	9%	32%	2%	0%
Item6	1000	38%	43%	0.25	0.69	0.68	D	4%	35%	10%	38%	12%	0%
Item7	1000	75%	17%	0.07	0.69	0.70	B	4%	75%	6%	6%	9%	0%
Item8	1000	31%	16%	0.05	0.69	0.70	B	3%	31%	5%	60%	1%	0%
Item9	1000	29%	14%	0.02	0.69	0.70	B	3%	29%	50%	17%	2%	0%
Item10	1000	28%	24%	0.15	0.69	0.69	D	2%	45%	12%	28%	13%	0%
Item11	1000	27%	14%	0.05	0.69	0.70	A	27%	24%	20%	20%	9%	0%
Item12	1000	64%	34%	0.22	0.69	0.69	C	6%	15%	64%	10%	6%	0%
Item13	1000	60%	16%	0.04	0.69	0.70	C	5%	19%	60%	12%	5%	0%
Item14	1000	47%	64%	0.39	0.69	0.67	E	20%	11%	14%	9%	47%	0%
Item15	1000	38%	50%	0.29	0.69	0.68	C	3%	10%	38%	34%	15%	0%
Item16	1000	44%	32%	0.17	0.69	0.69	D	3%	9%	33%	44%	11%	0%
Item17	1000	41%	68%	0.41	0.69	0.67	A	41%	12%	23%	9%	14%	0%
Item18	1000	63%	58%	0.35	0.69	0.67	C	9%	15%	63%	6%	6%	0%
Item19	1000	59%	43%	0.28	0.69	0.68	C	12%	18%	59%	8%	4%	0%
Item20	1000	21%	16%	0.09	0.69	0.69	E	4%	13%	49%	13%	21%	0%
Item21	1000	56%	46%	0.25	0.69	0.68	C	4%	12%	56%	19%	9%	0%
Item22	1000	40%	29%	0.18	0.69	0.69	B	7%	40%	18%	8%	28%	0%
Item23	1000	37%	58%	0.34	0.69	0.68	E	7%	19%	21%	15%	37%	0%
Item24	1000	55%	68%	0.40	0.69	0.67	B	5%	55%	12%	13%	16%	0%
Item25	1000	52%	36%	0.19	0.69	0.69	C	6%	18%	52%	14%	10%	0%
Item26	1000	53%	41%	0.26	0.69	0.68	E	4%	14%	20%	9%	53%	0%
Item27	1000	33%	30%	0.17	0.69	0.69	D	7%	16%	40%	33%	5%	0%
Item28	1000	58%	18%	0.06	0.69	0.70	B	6%	58%	10%	5%	21%	0%
Item29	1000	30%	50%	0.34	0.69	0.68	E	9%	28%	19%	14%	30%	0%
Item30	1000	47%	42%	0.27	0.69	0.68	D	10%	16%	21%	47%	6%	0%

詳細な項目分析の結果

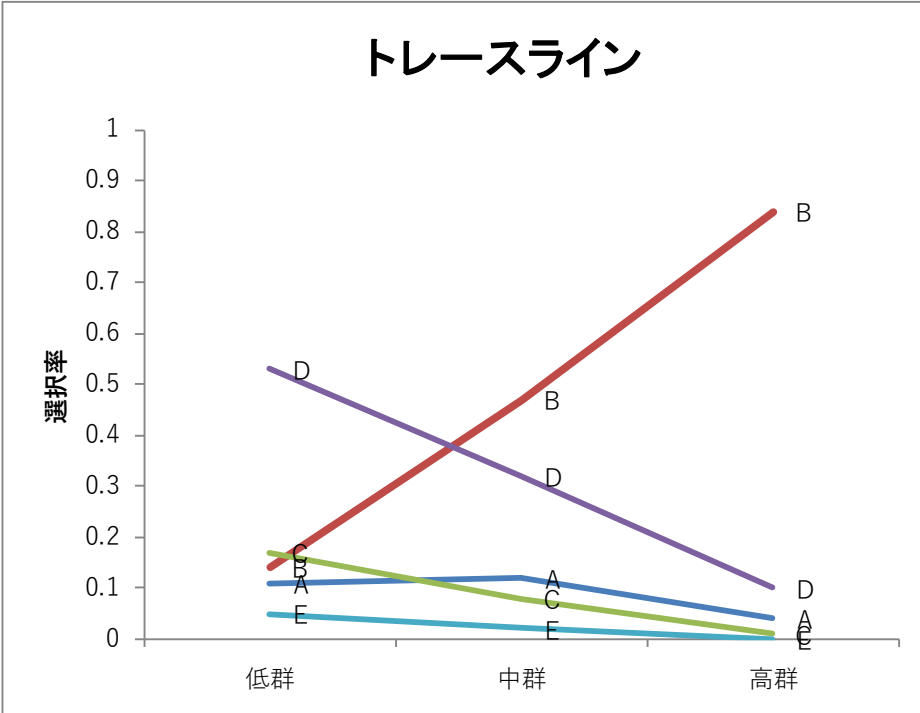
Q. 対角線の長さが4cmの正方形の面積を以下から選びなさい。

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



正答率

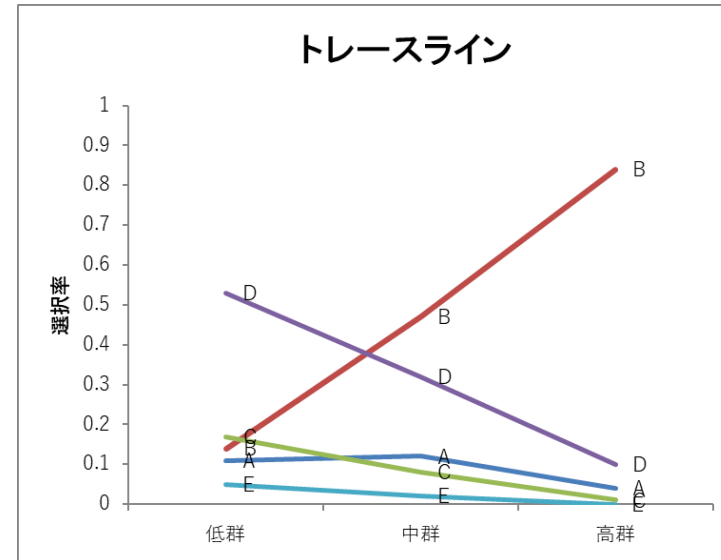
- 問題の易しさを表す指標
- その問題に正答した受検者の割合
- 0 (0%)～1 (100%)の値で表示
- 値が大きいほど易しい問題
- ほとんどの受検者が正答したり，逆に誤答する問題は，とくに相対評価を行うテストでは望ましくない
- 多くの場合，正答率は 0.1～0.9 の間の値になるように作問される
- 受検者の緊張をほぐすため，最初の問題をととても易しい問題にすることはある

選択率・無答率

- 当該の選択枝または回答類型に該当する受検者の割合
- 0 (0%)～1 (100%)の値で表示
- 選択率が0.1を下回るような選択枝または回答類型は無用
有効な選択枝は通常3枝程度
まぐれ当たりをなくすため多量の選択枝を設定しても意味がない
出題ミスや回答ミスの原因となる
- 無答率が大きい項目は，問題の意味を受検者が理解できなかった可能性がある

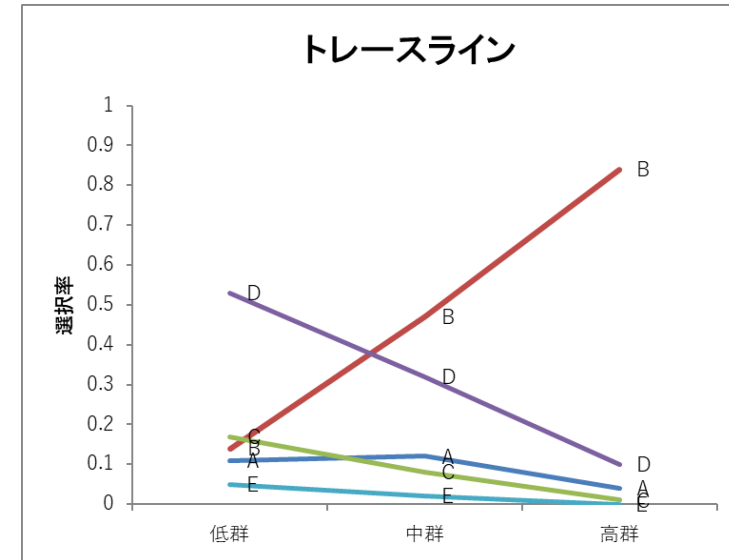
トレースライン

- 多枝選択式問題において、各選択枝の特性を表す図
- テスト得点に基づいて、受検者を群分けし(3～5群程度)、各群におけるそれぞれの選択枝の選択率を、選択枝ごとに直線で結ぶ
- 正答枝のラインは右上がりになるのが良い
- 誤答枝のラインはいろいろな形状があり得るが、正答枝のラインより上に位置したり、右上がりになるのは良くない



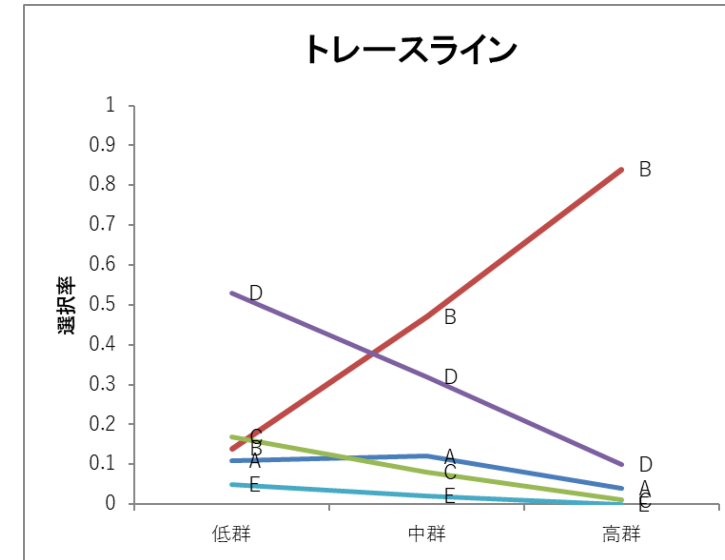
識別力

- テスト得点が高い受検者ほどその問題に正答し，テスト得点が高い受検者ほどその問題に誤答する，という傾向の強さ
- トレースラインの傾きの大きさ
- 一般に，識別力が大きいほどよい問題
- 指標として，D値，I-T相関，削除 α などがある



D値

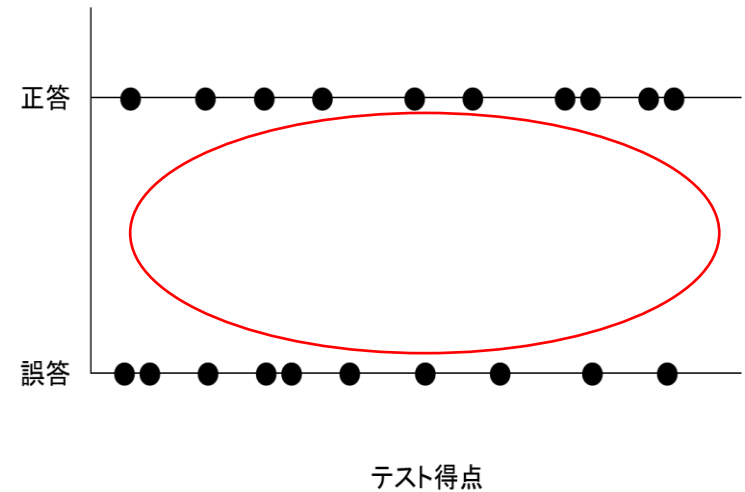
- 受検者を3群（高,中,低）に分けたときの高群と低群の選択率の差
- トレースラインにおける，高群と低群の選択率の差分
- 高群：中群：低群の割合は，およそ 0.27：0.46：0.27 になるのが良い
- 群の順位と，選択率の違いが対応しているほど，識別力は大い
- 正答枝のD値は，0.2 (20%) 以上になるのが望ましい



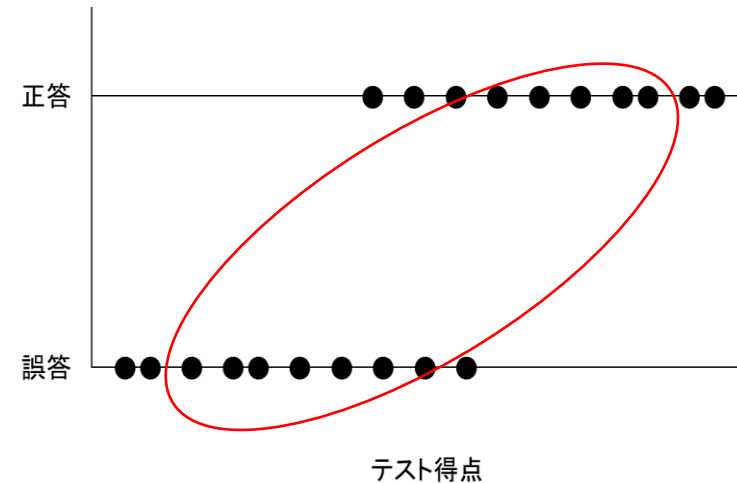
I-T相関

- 各項目について、項目得点 (Item) と、合計点から当該項目の得点を引いた得点 (Test) との相関係数
- 当該項目の得点を引かないと内部相関が生じ、相関係数が高めに出てしまう
- テスト得点の高低と、項目の正誤が対応しているほど、識別力が高い
- 正答枝のI-T相関は、 $+0.2$ 以上になるのが望ましい

識別力 小



識別力 大



削除 α

- 当該項目を削除した残りの問題からなるテストの α (アルファ) 係数

- α 係数：テストの信頼性を表す指標

能力が高い受検者はどの項目にも正答し，能力が低い受検者はどの項目にも誤答するという傾向の強さ

- テスト全体の α 係数より削除 α が大きくなる項目は，テストの信頼性を減じている，識別力の低い項目

項目分析表

項目	受験者数	正答率	D値	I-T相関	α 係数	削除 α	正答キー	A	B	C	D	E	無答
Item1	1000	78%	34%	0.23	0.69	0.69	E	7%	7%	5%	3%	78%	0%
Item2	1000	73%	40%	0.24	0.69	0.68	A	73%	3%	2%	3%	20%	0%
Item3	1000	94%	9%	0.12	0.69	0.69	B	1%	94%	2%	2%	2%	0%
Item4	1000	84%	22%	0.17	0.69	0.69	B	10%	84%	3%	2%	2%	0%
Item5	1000	48%	70%	0.42	0.69	0.67	B	10%	48%	9%	32%	2%	0%
Item6	1000	38%	43%	0.25	0.69	0.68	D	4%	35%	10%	38%	12%	0%
Item7	1000	75%	17%	0.07	0.69	0.70	B	4%	75%	6%	6%	9%	0%
Item8	1000	31%	16%	0.05	0.69	0.70	B	3%	31%	5%	60%	1%	0%
Item9	1000	29%	14%	0.02	0.69	0.70	B	3%	29%	50%	17%	2%	0%
Item10	1000	28%	24%	0.15	0.69	0.69	D	2%	45%	12%	28%	13%	0%
Item11	1000	27%	14%	0.05	0.69	0.70	A	27%	24%	20%	20%	9%	0%
Item12	1000	64%	34%	0.22	0.69	0.69	C	6%	15%	64%	10%	6%	0%
Item13	1000	60%	16%	0.04	0.69	0.70	C	5%	19%	60%	12%	5%	0%
Item14	1000	47%	64%	0.39	0.69	0.67	E	20%	11%	14%	9%	47%	0%
Item15	1000	38%	50%	0.29	0.69	0.68	C	3%	10%	38%	34%	15%	0%
Item16	1000	44%	32%	0.17	0.69	0.69	D	3%	9%	33%	44%	11%	0%
Item17	1000	41%	68%	0.41	0.69	0.67	A	41%	12%	23%	9%	14%	0%
Item18	1000	63%	58%	0.35	0.69	0.67	C	9%	15%	63%	6%	6%	0%
Item19	1000	59%	43%	0.28	0.69	0.68	C	12%	18%	59%	8%	4%	0%
Item20	1000	21%	16%	0.09	0.69	0.69	E	4%	13%	49%	13%	21%	0%
Item21	1000	56%	46%	0.25	0.69	0.68	C	4%	12%	56%	19%	9%	0%
Item22	1000	40%	29%	0.18	0.69	0.69	B	7%	40%	18%	8%	28%	0%
Item23	1000	37%	58%	0.34	0.69	0.68	E	7%	19%	21%	15%	37%	0%
Item24	1000	55%	68%	0.40	0.69	0.67	B	5%	55%	12%	13%	16%	0%
Item25	1000	52%	36%	0.19	0.69	0.69	C	6%	18%	52%	14%	10%	0%
Item26	1000	53%	41%	0.26	0.69	0.68	E	4%	14%	20%	9%	53%	0%
Item27	1000	33%	30%	0.17	0.69	0.69	D	7%	16%	40%	33%	5%	0%
Item28	1000	58%	18%	0.06	0.69	0.70	B	6%	58%	10%	5%	21%	0%
Item29	1000	30%	50%	0.34	0.69	0.68	E	9%	28%	19%	14%	30%	0%
Item30	1000	47%	42%	0.27	0.69	0.68	D	10%	16%	21%	47%	6%	0%

良い問題の分析結果の例

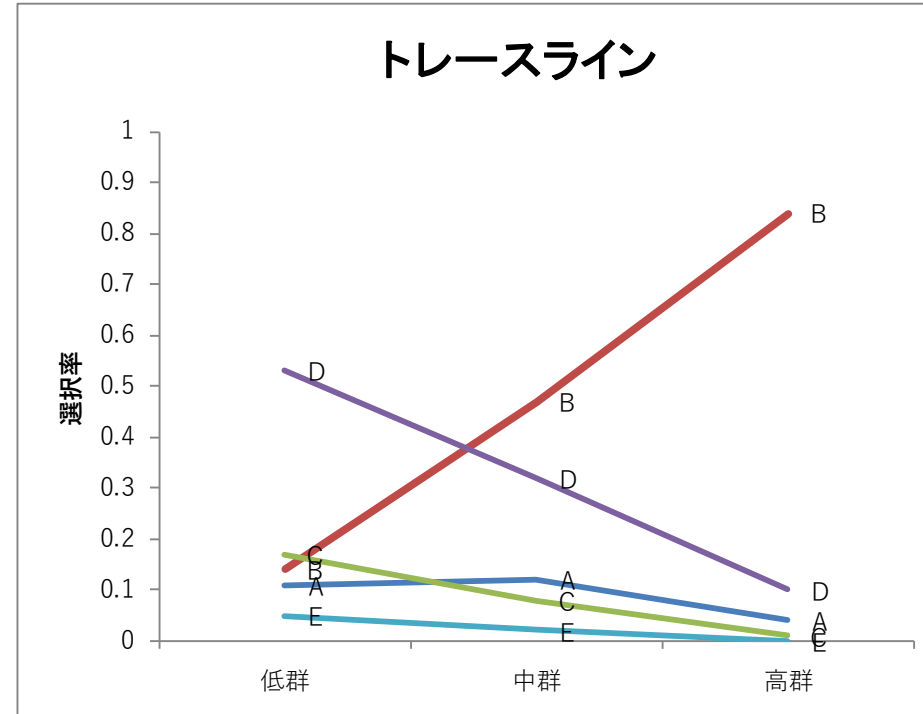
Q. 対角線の長さが4cmの正方形の面積を以下から選びなさい。

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



- 正答率は48%で適切，識別力も70%, 0.42と大きい
- 高群ほど正答枝 B，低群ほど誤答枝 D（一辺4cmの正方形と考えた）を選ぶ
- 誤答枝E以外はよく機能している

正答率が極端に高い問題

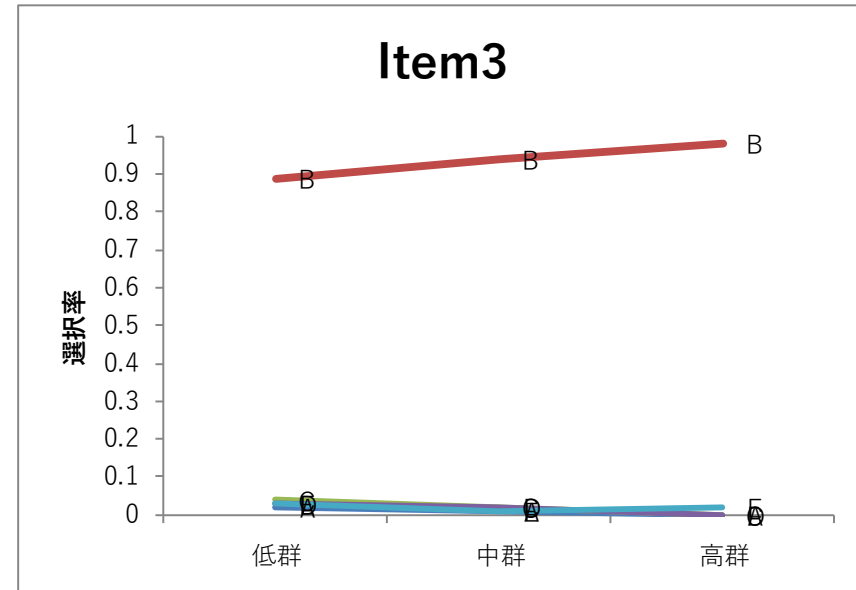
Q3. $1.5 \times 2 + 3.1 = \square$

A. 5.6 B. 6.1 C. 6.2 D. 6.6 E. 9.3

受験者数	正答率	D値	I-T相関	α 係数	削除 α	正答キー
1000	94%	9%	0.12	0.69	0.69	B

選択率	A	B	C	D	E	無答
全体	1%	94%	2%	2%	2%	0%
高群	0%	98%	0%	0%	2%	0%
中群	1%	94%	2%	2%	1%	0%
低群	2%	89%	4%	3%	3%	0%

識別指標	A	B	C	D	E	無答
D値	-2%	9%	-4%	-2%	-1%	0%
I-T相関	-0.08	0.12	-0.08	-0.06	-0.01	0.00



- 正答枝Cにほとんどの回答が集中し識別力がない
- 問題を $3.1 + 2 \times 1.5$ などに変える
- 誤答枝を受検者が間違った計算をした場合のものに変える

正答枝よりも魅力的な誤答枝がある問題

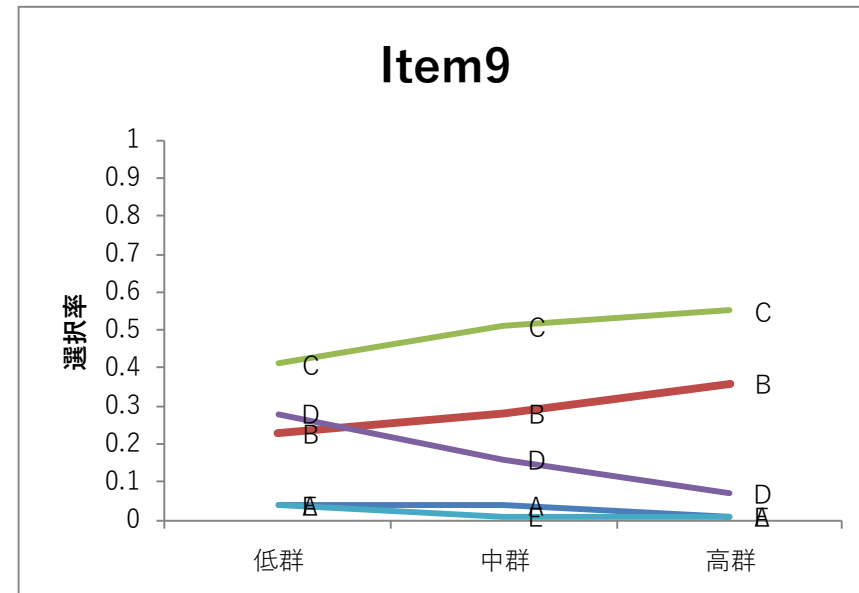
Q9. 半径4cmの半円の円周の長さを求めなさい。円周率は3.14とする。

A. 6.28cm B. 12.56cm C. 25.12cm D. 50.24cm E. 78.5cm

受験者数	正答率	D値	I-T相関	α 係数	削除 α	正答キー
1000	29%	14%	0.02	0.69	0.70	B

選択率	A	B	C	D	E	無答
全体	3%	29%	50%	17%	2%	0%
高群	1%	36%	55%	7%	1%	0%
中群	4%	28%	51%	16%	1%	0%
低群	4%	23%	41%	28%	4%	0%

識別指標	A	B	C	D	E	無答
D値	-3%	14%	14%	-21%	-3%	0%
I-T相関	-0.08	0.02	0.16	-0.18	-0.07	0.00



- 全体的に誤答が多く，識別力が小さい
- 高群ほど正答しているが，高群ほど誤答枝 C（全円の円周）を選ぶ傾向もある
- 問題文を修正する必要あり

悪い問題になりやすい問題

- 問題設定が分かりにくい問題
- 測りたい能力が分かりにくい「ひっかけ」問題
- 受検者の裏をかく「してやったり」問題
- 「あてはまるものをすべて選べ」問題
- 「上記のいずれでもない」「上記すべてあてはまる」などの選択枝がある問題
- 否定表現を含む問題