

テストをより良いものにするために

2026年6月10日(水)
名古屋大学大学院教育発達科学研究科
石井 秀宗

担当している授業の内容

- 大学1年生全員を対象とした、データサイエンス科目の統計領域(記述統計, 推定, 検定)
- おもに文系の大学1年生を対象とした, Excel及びRの演習(基本操作, 統計分析, 初級プログラミング)
- 教育学部の学部生及び大学院生を対象とした, 統計分析・テスト理論・Rの使い方
- 内容の一部は, 高等学校におけるデータサイエンスの授業用にアレンジ・実践されている
 - 名古屋大学教育学部附属中・高等学校(著)・名古屋大学数理・データ科学・人工知能教育研究センター(監修)・石井秀宗(執筆協力)(2025). THE FIRST STEP データサイエンス 大修館書店



3

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好ましくない
- 正答率は 0.1~0.9 の間の値になるのが望ましい
- 受検者の緊張をほぐすため, 最初の問題をとて易しい問題にすることはある

5

本日の内容

- テストデータから得られる基本情報
- テスト問題の分析例
- テスト問題作成ガイドライン
- 良い問題・悪い問題
- 項目反応理論 (IRT)

2

テストデータから得られる基本情報

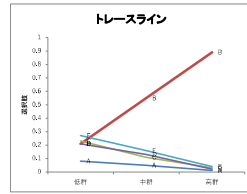
選択率・無答率

- 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, 機能していない
- ある程度選択される, 有効な選択枝は, 通常3枝程度
 - まぐれ当たり防止として多量の選択枝を設定するのは不適切
 - 出題ミスや回答ミスのもととなる
- 無答率が大きい項目は, 何を問うているかが受検者に伝わらなかった可能性がある

6

トレースライン (解答率分析図)

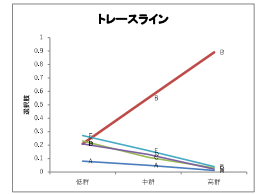
- 各選択枝や回答類型の特性を表す図
- テスト得点に基づいて受検者を群分けし (3～5 群程度), 各群におけるそれぞれの選択枝 (回答類型) の選択率 (出現率) を, 選択枝 (類型) ごとに直線で結ぶ
- 正答枝のラインは右上がりになるのがよい
- 誤答枝のラインはいろいろな形状があり得るが, 正答枝のラインより上に位置したり, 右上がりになるのはよくない



7

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
- 識別力が大きいほど良い問題
 - 正答率の高低よりも識別力の大小のほうが重要



8

テスト問題の分析例

問題

英文が日本文と同じ内容になるように, 下に与えられた語から最も適当なものを選んで () に補いなさい。

「あなたはあの少女を知っていますか？」

「はい, 以前彼女に会ったことがあります。」

Do you know that girl?

Yes, I have met () before.

① she ② that ③ her ④ him

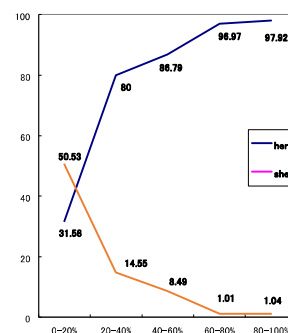
10

英語Q1 正答率・選択率

- 正答率・選択率
 - ① she 0.15 ② that 0.03
 - ③ her 0.79 ④ him 0.03
- 約 8 割の生徒が正答, 約 2 割の生徒が誤答
- 誤答する生徒の 7 割 (15/21) は she と回答
- 「彼女」に関連する単語が she, her であることは概ね理解されている (15+79=94)
 - that や him を選ぶ生徒は極端に少なく, 実質二択

11

英語Q1 トレースライン



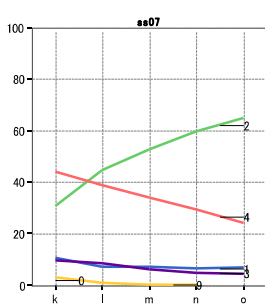
- 上位群の生徒はほぼ正答できる
- 中～下位群では約 1 割の生徒が she と誤答
- 最下位群の正答率は約 3 割で, she を選ぶ生徒が半数にのぼる

12

小学校 社会 Q7

- 問題
自動車工場のいろいろな工程で、機械を使って作業をしていることを確認するために、調べる場所を選択
- 解答類型
1：プレス
2：溶接
3：塗装
4：組み立て

社会Q7 トレースライン



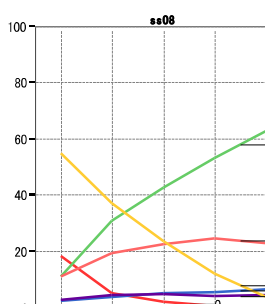
Line graph showing response percentages for Social Q7 across grade levels k to o. The y-axis represents percentage (0-100%). The x-axis represents grade levels (k, l, m, n, o). Five lines represent options 0, 1, 2, 3, and 4. Option 2 (welding) shows a significant increase from k to o, while option 4 (assembly) shows a decrease.

- 上位群ほど「2.溶接」を選択
• 30～65%
- 下位群ほど「4.組み立て」を選択
• 45～25%
- 「1.プレス」「3.塗装」を選択する児童は、どの層でも 1 割程度

小学校 社会 Q8

- 問題
Q7で選んだ場所で機械を使う良さを考え説明する
- 解答類型
1：プレスを選択し整合性のある考えを記述
2：溶接を選択し整合性のある考えを記述
3：塗装を選択し整合性のある考えを記述
4：組み立てを選択し整合性のある考えを記述
9：その他
0：無回答

社会Q8 トレースライン

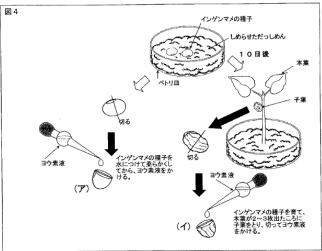


Line graph showing response percentages for Social Q8 across grade levels k to o. The y-axis represents percentage (0-100%). The x-axis represents grade levels (k, l, m, n, o). Five lines represent options 0, 1, 2, 3, and 4. Option 2 (welding) shows a significant increase from k to o, while option 4 (assembly) shows a decrease.

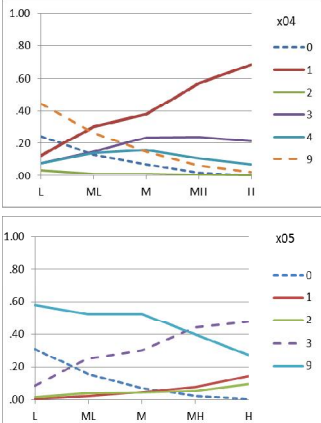
- 高群では、ほとんどの児童が、整合的な記述ができています
- 最下位群の 7 割以上の児童が、整合的な記述ができていない
• 約 2 割は無回答
- 下位群ほど「4.組み立て」で正答となる児童が減少
- 「4.組み立て」より「2.溶接」のほうが、整合的な考えを書きやすかったか？

小学校 理科 Q4, Q5

- Q4. (ア) は青紫色に変わったことから、どんなことが分かるか書きなさい
- 解答類型
1 デンプンがある
2 ヨウ素反応があった
3 養分がある
4 栄養がある
9 その他 0 無回答
- Q5. (イ) は反応がほとんど見られなかったことから、どんなことが分かるか書きなさい
- 解答類型
1 デンプンが使われた
2 デンプンがなくなっている
3 デンプンがない
9 その他 0 無回答



理科Q4, Q5 トレースライン



Two line graphs showing response percentages for Science Q4 and Q5 across grade levels L to H. The y-axis represents percentage (0-100%). The x-axis represents grade levels (L, ML, M, MII, H). Five lines represent options 0, 1, 2, 3, and 4. Option 2 (starch is present) shows a significant increase from L to H, while option 4 (starch is not present) shows a decrease.

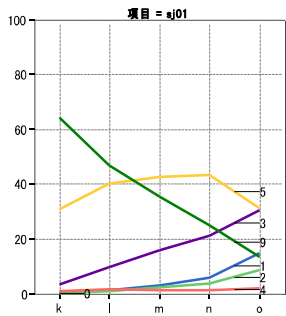
- Q4では、上位者ほど正しい回答をしている
- 「3.養分がある」「4.栄養がある」と解答する児童も 3 割程度いる
- Q5では、中位群以下で「その他」の解答が 5 割を超える
- 下位群で 3 割以上が無回答
- 上位群ほど「デンプンがない」が多い
- 作題者の狙いである「使われた」「なくなった」と、経過を解答している児童はわずか

中学校 国語 Q1

- 問題
文章で筆者が主に伝えたいことを3点まで挙げる
- 解答類型
1：主要点 3
2：主要点 2， 周辺 1
3：主要点 2， その他文中から 1
4：主要点 2
5：主要点 1， 周辺 1 以下
9：その他

19

国語Q1 トレースライン



- 主要点を3つ挙げられる生徒は、最上位群でも2割未満
- 上位群ほど主要点を2つ以上挙げられる
- 主要点を1つしか挙げられない生徒が、どの層でも多い
- 下位群では、その他の解答が5割以上
- 要求水準が高かった？

20

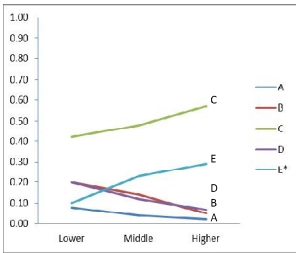
小学校 算数 旅人算

Q. 2000mの道のりを妹は60m/分で進む。15分後、姉は自転車で出発し、120m/分で妹を追いかける。妹は出発してから何分後に姉に追いつかれるか求めなさい。
A. 5分 B. 10分 C. 15分 D. 20分 E. 30分

	選択率	
A	0.04	
B	0.13	
C	0.49	←正答枝Eよりも誤答枝Cのほうが選択率が高い
D	0.13	
E*	0.21	←正答枝

21

旅人算 トレースライン



- 多くの受検者が、正答枝Eより、誤答枝Cを選択
- 上位群ほどその割合が高い
選択率を見ているだけでは分からなかった
- 設問に何か問題があった
誰が出発してから時間を求めるのが不明確

旅人算 修正例

Q. 2000mの道のりを妹は60m/分で進む。15分後、姉は自転車で出発し、120m/分で妹を追いかける。妹は出発してから何分後に姉に追いつかれるか求めなさい。

Q. 2000mの道のりを妹は60m/分で進む。妹が出発してから15分後に、姉は妹と同じ地点から自転車で出発し、120m/分で妹を追いかける。妹が出発してから何分後に、妹は姉に追いつかれるか求めなさい。

23

テスト問題作成
ガイドライン

24

ガイドライン項目の例

- 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
- 14 教示文・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
- 17 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
- 22 高得点者と低得点者をよく区別できるような、識別力の高い選択枝を用いること
- 24 不必要に選択枝を増やさないこと
- 35 問題作成と同時に、評価の観点、採点基準を設定すること

25

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

× フレイルについて説明せよ。

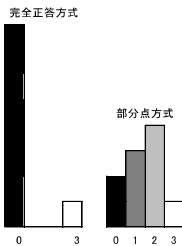
- 問題文が大雑把すぎて、何を問うているのか、何を回答したら良いのか、受検者にはよく分からない。
- 語源、原因、症状、構成要素、基準、サイクル、予防、etc
- 「○○の△△について(□□の観点から)説明せよ」程度の教示が必要

○ フレイルの症状と判断基準について説明せよ。

27

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



29

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よい
- その問題で測りたい能力は何か、テスト全体で何を測ろうとしているのかを基準に考える

26

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

正答(-36, 13)

28

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)しているため、(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
- 選択枝を探すのに手間取ってしまう

30

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び。 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び。 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

31

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

33

良い問題・悪い問題

35

17 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

× 円周率は（ ）である。

- 小学校6年生算数のテスト問題
- 教師の想定している正答は3.14
- 何を問うているのか、児童には分からない
- 値？ 性質？ 表記？
- 児童の回答：3.14, 3.14159..., 無理数, π など

○ 円周率の値として用いられる数値を答えなさい。

32

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

34

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
- 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
- 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

36

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
- 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
- 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
- 受検者のつまづきを把握できず，指導改善の役に立たない

37

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
 - 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や個々の設問の目的による
- 識別力が高いか
 - 測りたい能力が高い受検者ほど得点できる問題が良い問題
 - 難しい問題が良い問題， 易しい問題が悪い問題，ではない
 - 正答率よりも識別力のほうが重要
- 思考力や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

38

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

39

項目反応理論

40

項目反応理論

- 困難度や識別力などの項目特性と，受検者の能力を分離して扱うモデル（関数）を用い，共通の尺度上で複数のテストを実施することを考えるテスト理論
 - 項目応答理論とも言う
 - Item Response Theory : IRT
- 項目特性と受検者の能力が分離されているので，
 - 受検者の能力を表す値（能力パラメタ）は，項目（テスト）に依存しない
 - 項目の特性を表す値（項目パラメタ）は，受検者集団に依存しない

41

項目反応理論

- 同じ能力を測っているテストであれば，テストが異なっても，
- 能力値を比較することが可能
 - 複数回の異なるテストの結果を，相互に比較することができる
- 能力値の変化を捉えることが可能
 - 個人や集団の年次進行による能力の変化を捉えられる
 - 同一学年の成績を年度間比較できる

42

項目特性値・能力値推定の概要

- 困難度や識別力などの項目特性値が分かっている**既存項目**と、それらの値をこれから推定する**新規項目**を収録した**テスト**を作成する
- 既存項目について、ある能力値の受検者がその項目に正答する確率を表す関数 P_j を準備する
- テストに含まれる既存項目の関数 P_j を掛け合わせ、ある能力値の受検者が、それらの既存項目に回答したときの、各回答パターンの出現確率を表す関数 P_T を構成する

43

項目特性値・能力値推定の概要

- テストを実施し、**既存項目および新規項目への各受検者の反応（データ）**を収集する
- **既存項目に対する受検者のデータ**(回答パターン) を関数 P_T に代入し、**各受検者の能力値を推定**する
- ある能力値の受検者が、新規項目に正答する確率を表す関数 P_n を設定する
- 各受検者の能力値をその関数 P_n に代入し、**新規項目の項目特性値を推定**する

44

適用例

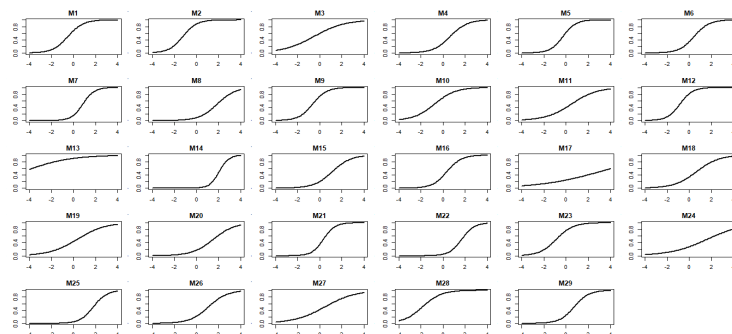
状況設定

- 共通問題のある2つの数学テストを作成し、県内の中1の生徒全員に、2回に分けてテストを実施
- 各受検者はいずれか一回のテストを受検
 - 群1：第1回受検者、群2：第2回受検者
- 各テストともに、共通問題11問、独自問題9問の、合計20問で構成
 - 全部で11+9+9=29問
- 全項目の項目特性、全受検者の得点を比較可能にしたい
 - 共通問題だけでなく、各回のみの問題も対象

45

46

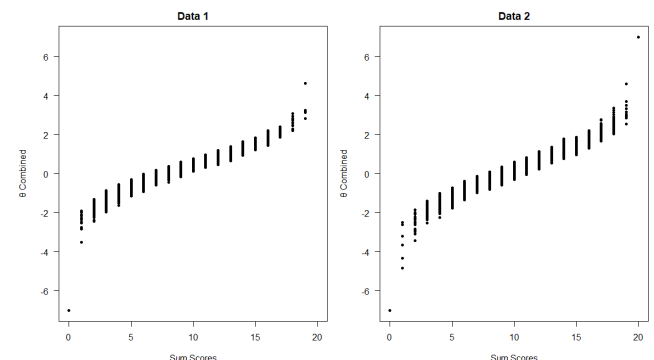
項目特性曲線



- 易しい項目：M2 など
- 識別力の大きい項目：M14 など
- 難しい項目：M8 など
- 識別力の小さい項目：M17 など

47

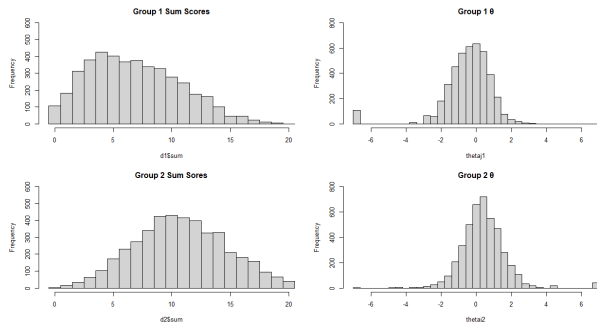
合計得点と能力パラメタの対応



- 基本的に、合計得点と能力パラメタの大小関係は対応する
- 合計得点と同じでも、回答パターンの違いにより、能力パラメタの値は異なる

48

合計点及び能力パラメタの分布



- テストが異なるので、合計点は直接比較できないが、等化された能力パラメタは直接比較できる
- 同じ得点でも、回答パターンにより能力パラメタの値が異なるので、分布の形が変わる

49

項目反応理論を用いたテストに求められる要件

- 大規模データが必要
- 基本的に多枝選択式問題
- 項目間に局所独立性という性質が求められる
- 項目の性質を変えないようにするため、問題は非公開
- 大量の項目を蓄積したデータベースが必要

50

おわりに

テストを評価する必要性

- 多くの教育現場でテストを利用しているのに、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している
- 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・出荷し、使用されている状態
- テストの作成、実施、採点、運用等に関して、客観的な評価も含めた検討が必要

52

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

- 目的、運用も含めて考えることが重要

- これらの1つでも欠けると、そのテストすべてが適切でないものになる

テスト研究ウェブサイト



- <https://www.educa.nagoya-u.ac.jp/~ishii-h/test.html>
- テストに関する諸事項をまとめたサイト
- 問題作成ガイドライン、項目分析ツール、Rのスク립ト集などをダウンロード可能

53

54

テスト問題作成ガイドライン

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 正解が問題作成者の価値観に左右されるような問いにならないこと
- 4 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
- 5 ひっかけ問題にならないこと
- 6 高次の能力を測る問題では、受検者にとって新奇な素材を用いること

問題の形式

- 7 測りたい能力に見合った問題形式を用いること
- 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
- 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
- 10 読解力や思考力を測る記述式問題では字数制限を設けないこと

問題の記述

- 11 言語レベルを受検者集団に合わせること
- 12 教示文・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
- 13 教示文・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
- 14 教示文・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
- 15 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
- 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

設問部分

- 17 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
- 18 本文や選択枝など他の部分を読まなくても、設問部分だけで何を問われているかが分かること
- 19 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
- 20 一部の受検者にしか分からないような暗黙の前提を用いないこと

選択枝

- 21 いずれの選択枝ももっともらしいこと
- 22 高得点者と低得点者をよく区別できるような、識別力の高い選択枝を用いること
- 23 正答枝と誤答枝が明確に区別できること
- 24 不必要に選択枝を増やさないこと
- 25 明らかな誤答枝やお遊びの選択枝など、余計な選択枝を入れないこと
- 26 五十音順、数量の大きさ順など、何らかの法則に従って選択枝を並べること
- 27 正答枝の位置をランダムにばらつかせること
- 28 「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
- 29 「～でない」「～以外である」など否定表現を用いないこと
- 30 「絶対に」「常に」「決して」「完全に」など、強意語を用いないこと
- 31 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
- 32 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
- 33 選択枝の長さをおおむね揃えること
- 34 選択枝の内容や形式などの構造を揃えること

記述式問題の採点に関して

- 35 問題作成と同時に、評価の観点、採点基準を設定すること
- 36 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）
- 37 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
- 38 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
- 39 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

石井秀宗・荒井清佳・坪田彩乃・安永和央・寺尾尚大（2021）. テスト問題作成ガイドラインの開発（1）—日本での普及に向けた整理—. 日本テスト学会第19回大会

https://www.educa.nagoya-u.ac.jp/~ishii-h/test_guideline.html