

より良いテスト作成・実施のために！ 知っておきたい「テストに関する基礎知識」

2023年3月

石井 秀宗

名古屋大学大学院教育発達科学研究科

本日の内容

1. はじめに
 2. 「良いテスト」の要件
 3. テストの目的と分類
 4. 問題の形式
 5. 問題作成ガイドライン
 6. 本日のまとめ
- Ex. 事前に頂いたご質問への回答

次の問いの答えは何でしょう？

円周率は

である。

先生が想定した答え

円周率は

3.14

である。

(小学校 6 年生算数のテスト問題)

児童の回答

円周率は

3.14

である。

円周率は

3.14159...

である。

円周率は

無理数

である。

円周率は

π

である。

どう採点する？

- 3.14を○, 3.14159...と無理数と π を×
- 3.14と3.14159...を○, 無理数と π を×
- 3.14と π を×, 3.14159...と無理数を○
- 4つとも○

など

どうしていろいろな回答が出てくるのか？

- 円周率の何を書くのか明確でない。 値？ 性質？ 表記？
- 授業で習ったことを書くのか、知っていること、正しいと思うことを書くのかが示されていない
- 何を書けば正解とされるか、先生の考えを推測できる能力が求められている
- 修正例

円周率の値として用いられる数値を答えなさい。

「良いテスト」の必要性

- 児童生徒の能力を適切に評価する
- テスト問題に不備があり，児童生徒が回答に迷うことがある
- 児童生徒には不明な出題意図や，合理性の低い採点基準に基づいて採点しては，児童生徒は納得できないことがある
- テストは児童生徒とのコミュニケーションの1つ
- 児童生徒と良いコミュニケーションを保つためにも良いテストが必要

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

本日の内容

1. はじめに
 2. 「良いテスト」の要件
 3. テストの目的と分類
 4. 問題の形式
 5. 問題作成ガイドライン
 6. 本日のまとめ
- Ex. 事前に頂いたご質問への回答

「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

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
- 何をすれば目的を達成できるか
テストで測るべき能力は何か
- 何を問えばその能力を測定できるか
内容の理解 → 単なる丸暗記で解くような問題にしない
- その能力を測ることに意義はあるか
重要事項の理解 → 重箱の隅をつつくような問題にしない

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
何を問うているか，何について回答すれば良いかを
受検者が理解できること
- 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
- 問題作成ガイドラインの利用
問題内容，問題形式，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
- 作題者は、正答が分かっているので簡単に感じる
- 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
- 受検者は、正答を知らないので難しく感じる
- 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は困難
- 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・使わないなど，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

- 測りたいものを測れる採点基準になっているか
正答・誤答の判断基準は何か
何を減点 (加点) 対象とするか
- 正答, 誤答の合理的な根拠を示せるか
出題者の意図通りに受検者が問題を理解するとは限らない
- 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か
- 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は理にかなっているか
- 公平な解釈，運用をしているか
特定の集団や個人を有利（不利）に扱っていないか
- 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が正しく理解されているか

良いテストであるために

- 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

本日の内容

1. はじめに
 2. 「良いテスト」の要件
 3. テストの目的と分類
 4. 問題の形式
 5. 問題作成ガイドライン
 6. 本日のまとめ
- Ex. 事前に頂いたご質問への回答

テストの分類を知る必要性

- テストは目的に応じて備えるべき性質が異なる
- テストの目的や性質を意識してテストを作成・運用する必要がある
- 良いテストを行うことにつながる

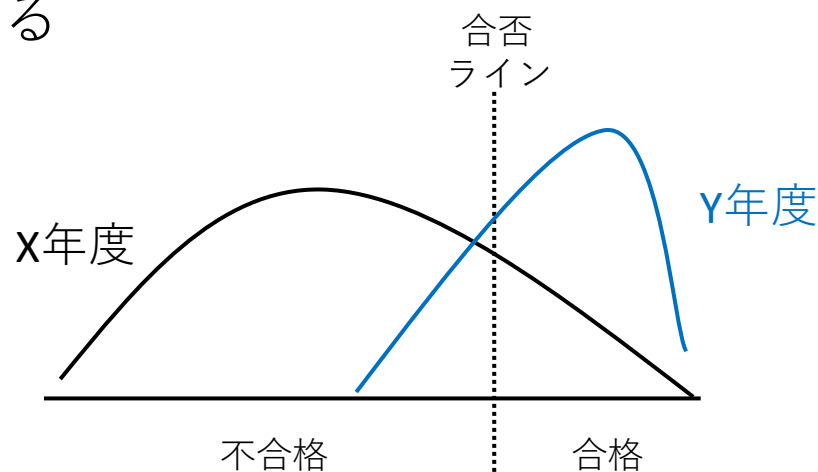
→ テストの分類

テストの目的と分類

- 資格試験：英語能力検定試験，運転免許試験，医師国家試験
- 選抜試験：入試，採用試験
- 成績評価のための試験：定期試験，模試
- 学習支援のための試験：授業におけるテスト，小テスト

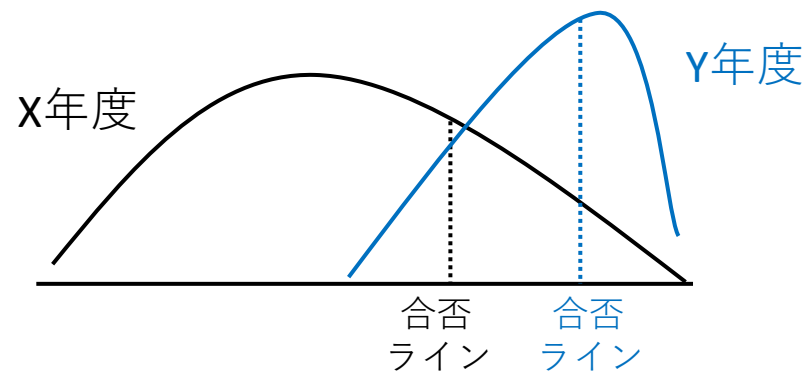
資格試験の目的・性質

- 領域の内容について一定の基準を満たしているか否かを判定
- 領域の内容全般をカバーすることが必要
- 受検者の絶対的位置を把握
- 受検者集団（受検回）によらず，合格になる能力水準は一定でなければならない
- 合理的な基準の設定が求められる



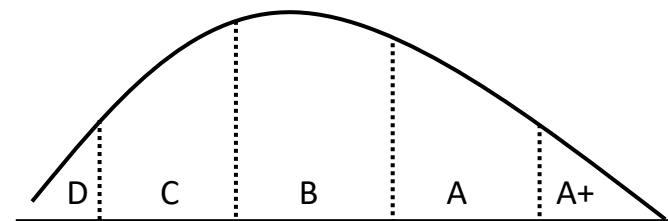
選抜試験の目的・性質

- 受検者集団を上位（合格），下位（不合格）の2群に分割
- 公平性，公正性が特に強く求められる
- 広い領域から出題される
- 受検者集団における各受検者の相対的位置を把握
- 受検者を識別するため，得点は広く細かく分布するのが良い
- 受検者集団（受検回）や定員によって合格になる能力水準は変動する



成績評価のための試験の 目的・性質

- 学期末, 学年末等における受検者の成績評価
- 当該期に学習した重要事項をカバーすることが必要
- A, B, C, D などの成績を付与する
- 資格試験と選抜試験の両方の性質を持つ
- 受検者を識別するため, 得点は広く細かく分布するのが良い
- 合理的な基準の設定が求められる



学習支援のための試験の 目的・性質

- 診断的評価：児童生徒の準備状況の確認
 - 形成的評価：児童生徒の理解度の確認・促進
 - 総括的評価：児童生徒の習得度の確認
 - 指導と評価の一体化：評価を授業改善に生かす
-
- 授業におけるテストや小テストで実施されることが多い
 - 授業にあわせて作成・実施される
 - 全国学力調査など大規模試験を利用して、直接的または間接的に学習支援することもある

留意点

- 目的を意識してテストを作成，実施する
- 1つのテストを複数の目的に用いようとすると，中途半端なテストになりかねない
- 学習支援で，前の単元の総括的評価が，次の単元の診断的評価になるということはある
- そのテストでできること，できないことを明確にした上でテストを運用する

本日の内容

1. はじめに
 2. 「良いテスト」の要件
 3. テストの目的と分類
 4. 問題の形式
 5. 問題作成ガイドライン
 6. 本日のまとめ
- Ex. 事前に頂いたご質問への回答

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

- 構成式

記述式：穴埋め，短答，論述，論文*

パフォーマンス式：口述，実技・実演，作品

* 論文をパフォーマンス式に分類する考え方もある

選択式・記述式問題の 長所・短所 1

●選択式問題の短所

選択枝にない情報は得られない

選択枝が回答のヒントになり得る

あて推量でも正答になり得る

思考過程が限定される

思考力・表現力の評価が難しい

⇒ 「真の能力」を反映しにくい



○記述式問題の長所

記述内容から多くの情報を得ることができる

回答のヒントとなる情報を与えにくい

あて推量で正答になることはまずない

思考過程が限定されない

思考力・表現力を評価できる

⇒ 「真の能力」を反映しやすい

選択式・記述式問題の 長所・短所 2

○選択式問題の長所

1問あたりの回答時間が比較的短い

多くの問題を出題できる

広い領域をバランスよくカバー

合計点に対する1問あたりの影響小

採点が客観的，正確，短時間でできる

⇒公平性・公正性を確保しやすい



●記述式問題の短所

回答に時間がかかる

問題数が限られる

狭い領域しかカバーできない

合計点に対する1問あたりの影響大

採点が主観的，不精確，時間がかかる

⇒公平性・公正性の確保が難しい

選択式・記述式の使い分け

- 大規模テスト（受検者の多い資格試験，大学入学共通テストなど）

選択式のほうが向いている

大量の答案を，短時間で，正確に採点することが可能

大量の記述式回答を公平・公正・正確に採点するのは大変

- 中小規模テスト（学期末試験，個別大学の入試など）

どのような能力を測りたいかによって選択

選択式で知識・技能，記述式で思考力・表現力を測るなど
それぞれの長所を生かし互いに補完

本日の内容

1. はじめに
 2. 「良いテスト」の要件
 3. テストの目的と分類
 4. 問題の形式
 5. 問題作成ガイドライン
 6. 本日のまとめ
- Ex. 事前に頂いたご質問への回答

問題作成ガイドラインの意義

- テストが良いテストであるためには、個々の問題が良い問題であることが必要
- 良い問題とはどのような問題か？
- それを実現するために問題はどのように作成すべきか？

→ 問題作成ガイドライン

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児童生徒ほど正答する
- 測りたい能力が低い児童生徒ほど誤答する
- 復習することにより，どうして誤答したのか児童生徒が理解し，学びを進めることができる
- 児童生徒のつまずきを把握し，指導改善に役立てることができる

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
- 測りたい能力が高い児童生徒ほど正答するとならない
- 測りたい能力が低い児童生徒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児童生徒が理解できず，
学びが進まない
- 児童生徒のつまずきを把握できず， 指導改善の役に立たない

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いか
 - 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
 - 測りたいものを実際に測れているかがポイント
-
- 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は限らない
 - 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は限らない
 - 難しい(易しい)問題が良い(悪い)問題な訳ではない

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
- 研究室ウェブサイトでガイドラインや修正例などを紹介

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



問題作成ガイドライン 1

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

問題作成ガイドライン 2

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

問題作成ガイドライン 3

問題の記述

- 11 言語レベルを受検者集団に合わせることに
- 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

問題作成ガイドライン 4

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

問題作成ガイドライン 5

選択枝（1／2）

- 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

問題作成ガイドライン 6

選択枝（2／2）

- 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

問題作成ガイドライン 7

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、あまり細かい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、評価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)

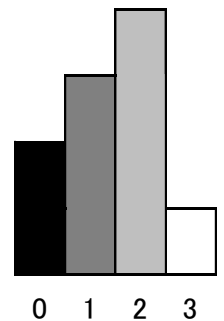
9「あてはまるものをすべて選べ」という設問は避けること。使う場合は部分点を与えること

- 「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



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

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

29 「～でない」「～以外である」など否定表現を用いないこと

× 愛知県の説明として正しくないものをすべて
選べ。 正答(?)

- A. 西日本には属さない
- B. 本州に属さない訳ではない
- C. 東海3県に属しはしない
- D. 中部地方に属さないとも限らない

理解が困難
設問も選択枝も否定型だと
頭がこんがらが

○ 愛知県が属する地域を以下の中から選べ。 正答(C)

- A. 東北地方 B. 関東地方 C. 中部地方 D. 関西地方

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

ガイドラインの利用にあたって

- 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
- 多くの（雑多な）資料から有用な情報を読み取る
→ 本文や図表を複雑（冗長）なものにする
- 公式を用いる能力を問う → 問題設定は簡潔にする
- 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

本日の内容

1. はじめに
 2. 「良いテスト」の要件
 3. テストの目的と分類
 4. 問題の形式
 5. 問題作成ガイドライン
 6. 本日のまとめ
- Ex. 事前に頂いたご質問への回答

本日のまとめ

- 「良いテスト」とは，測りたい能力を測っているテスト
- テストの目的を明確にして，適切に作成，実施，採点，運用することが必要
- 作題にあたっては，ガイドラインを参照する，声に出して読む，忘れたところに見直すなどの作業を行う
- テストをより良いものにしていくために，テストや問題の性質・性能を評価することも重要

本日の内容

1. はじめに
2. 「良いテスト」の要件
3. テストの目的と分類
4. 問題の形式
5. 問題作成ガイドライン
6. 本日のまとめ

Ex. 事前に頂いたご質問への回答

問題の難易度はどの程度に設定したら良いか？

- 多くの試験において、得点は広く細かく分布するほうがよい
 - 下位層の受検者の状況をよく知るには、易しい問題が必要
 - 中位層の受検者の状況をよく知るには、中程度の問題が必要
 - 上位層の受検者の状況をよく知るには、難しい問題が必要
 - 下位層や上位層の受検者は少なく、中位層の受検者が多いと考えられるから、問題数もその割合に合わせるのが良い
-
- 易しい問題 2～3割， 中程度の問題約 5 割， 難しい問題 2～3 割くらいにするのが良い

各問題の配点はどのように設定したら良いか？

- 多くの試験において、得点は広く細かく分布するほうがよい
- 下位層の分布を広くするには、易しい問題の配点を大きくすればよい
- 中位層の分布を広くするには、中程度の問題の配点を大きくすればよい
- 上位層の分布を広くするには、難しい問題の配点を大きくすればよい

- 結局、 **どの問題の配点も均等にするのが良い**
- 難しい問題の配点を大きくすると、中～下位層の分布は団子状になり、細やかな能力評価ができなくなってしまう
- 問題の配点を大きくする場合は、部分点を付与すること

複数の資料を用いる問題を作り続ける工夫は？

- 複数の資料を参照して解く問題は作るのが大変
- 作り続けることはもっと大変
- 資料の構成や、組み合わせる資料の種類のパターンをいくつか決め、それに沿って作問することがある
- 問題のパターン化という弊害
- 児童生徒に問題を作らせる
- 能力評価の1つの方法として作問法というものがある
- よく理解している児童生徒ほど適切な問題を作る

主体性を測るテストを作ること はできるか？

- 主体性をテストで測るにはまず、主体性とは何かを定義することが必要
- 主体性が高ければ正答し、低ければ誤答する、という傾向を良く示す問題を作成すればよい → 実際に作れるか？
- 主体的であれば学力は高くなるだろうと考えれば、**学力試験の結果から、主体性の一部の側面を推測**できる
- 主体性は、学力よりは行動特性に近い
- **観察などにより行動を見るほうが主体性を捉えやすい**

コンピュータを利用したテスト は今後どうなるか？

- 資格試験，一部の入試（AOなど），自治体テスト，教材等で導入が進められている
- 全てではないが，今後さらに普及していくと考えられる

利点

オンデマンドでも実施可能

動画などの素材も利用可能

簡潔な回答形式であれば，採点，集計等の作業を効率的に行える

データを蓄積・参照しやすい（どのように活用するかという問題はある）

課題（これがかなり大変）

公平・公正で安定した実施環境の確保

実施者側のスキルや体制整備の問題