

★評価の目的

★評価の機能

★評価の分類

A) 時期的評価分類

- ① 診断的評価
- ② 形成的評価
- ③ 総括的評価

B) 評価主体分類

- ① 内在的評価
- ② 外在的評価

C) 基準分類

- ① 相対的評価
- ② 絶対的評価

D) 性格分類

- ① 選別評価
- ② 測定評価
- ③ 認定評価

★評価の方法

テスト アンケート 面接 観察 レポート ルーブリック評価
ポートフォリオ

★テストの種類

A) 用途分類

適性テスト

プレースメントテスト
アチーブメントテスト
プロフィシエンシーテスト

B) 目的分類

目標準拠テスト CRT
集団準拠テスト NRT

C) 評価する側からの分類

①主観テスト

論述 書き取り 短文作成 作文 口頭試験 スピーチ

②客観テスト

再認法

真偽式

多肢選択法

組み合わせ法

序列法

穴埋め式

再生法

完成法

訂正法

質問文作成法

変換法

質問法

クローズ法

D) 内容別分類

①項目別テスト

②全体テスト

E) 分野別分類

聴解テスト 表記テスト（漢字テスト）語彙テスト 文法テスト
作文テスト 口頭発表 読解テスト ディクテーションテスト

<客観テストの具体例>

1 真偽式

上の文章の内容にあてれば○を、間違っていれば×を

() の中に書きなさい。

() 最も駅から近いスーパーはAスーパーである。

() Aスーパーより、Bスーパーのほうが野菜が新鮮だ。

2 多肢選択法

A ご主人はいらっしゃいますか。

B (1) はい、いらっしゃいます。

(2) はい、まいります。

(3) はい、おります。

(4) はい、います。

3 組み合わせ法

ズボンを ・ かぶります

洋服を ・ まきます

帽子を ・ しめます

眼鏡を ・ かけます

ベルトを ・ きます

スカーフを ・ はきます

4 序列法

() から () います () あした () はじめ () 思って

() よう () 仕事 () を () と

5 穴埋め式

この道 () まっすぐ行ってください。

東京 () 電車を乗り換える。

子供は池のそば（ ）います。

6 完成式

頭が痛いから、もう_____

宝くじが当たったら、_____

7 訂正法

次の文には間違いがある。訂正しなさい。

「面白いの本を貸してください」

「大学から卒業しました」

8 質問文作成法

Q _____。

A かぎは机の引き出しにあります。

Q _____。

A デパートへ行きました。

9 変換法

() の言葉を適当な形にしなさい。

旅行の前にホテルを(予約する)_____おきます。

昨日の映画はあまり(おもしろい)_____です。

10 質問法

上の文章について次の質問に答えなさい。

(1) 下線部 A そのこととは何ですか。文中から5文字で書き抜きなさい。

--	--	--	--	--

(2) 筆者がその仕事を選んだ理由は何ですか。

<クローズ法 例>

1) 定間隔法

ジャパニーズスマイルということばがある。日本人は良く訳のわからない笑いをするとと言われる。外国人は薄気味悪く思い、これ（ ）について悪口を言うが、それ（ ）日本人の気質を知らない（ ）である。日本人は昔から、（ ）、悪いという判断を直接（ ）はっきりという事を避ける。（ ）の立場を考えて、その（ ）を傷つけることを恐れるから（ ）ある。

（秦喜美恵 1990「Japanese for Today」学習研究社 p212）

2) 変則間隔法

ジャパニーズスマイルということばがある。日本人は良く訳のわからない笑いをするとと言われる。外国人（ ）薄気味悪く思い、これ（ ）について悪口を言うが、それは日本人の気質を知らない（ ）である。日本人は昔から、よい、悪い（ ）という判断を直接にはっきり（ ）いう事を避ける。相手（ ）立場を考えて、その人（ ）傷つけること（ ）恐れるからである。

（秦喜美恵 1990「Japanese for Today」学習研究社 p212）

3) 変型Cテスト

ジャパニーズスマイルということばがある。日本人は良く訳のわからない笑いをするとと言われる。外国人は薄気味悪く思い、これ○について悪口を言うが、それ○日本人の気質を知らない○○である。日本人は昔から、○○、悪いという判断を直接○はっきりという事を避ける。□□の立場を考えて、その□を傷つけることを恐れるからである。

（秦喜美恵 1990「Japanese for Today」学習研究社 p212）

良いテストの条件

★テストの妥当性

①内容的妥当性

テストの問題が測定しようとしている内容を適切に代表しているか

領域代表性 出題範囲の中で偏りなく作られているか

領域適切性 出題範囲内で作られているか

指導目標とあっているか

難易度

②基準関連妥当性

テストの得点と、他の信頼できる外的基準との関連を示す

③概念的妥当性

語学力がそれを構成する下位能力を予測してはかれているかどうか

真正性^{しんせいせい} テストの素材が現実の言語使用を反映しているかどうか

④表面的妥当性

テストの見た目

★テストの信頼性

信頼度係数 $0 < \rho < 1$

個人 0.85以上 団体 0.65以上

信頼度係数の求め方

再テスト法 平行テスト法 折半テスト法

クダー・リチャードソンの公式

クロンバックの α 係数

★テスト結果の処理

ヒストグラム（度数分布図）

折れ線グラフ（度数分布曲線）

累積度数表 度数分布表

代表値

平均値 最頻値（モード）中央値（メジアン）

散布度

範囲（レンジ） 標準偏差

得点

粗点 標準点（偏差値）

相関係数

$-1 < x < 1$ $0.7 \leq x \leq 1$ の場合相関関係が高い

項目分析

難易度＝正答率 20%～80%

識別力＝弁別度 弁別指数0.4以上

項目分析一覧図（縦軸に識別力、横軸に正答率を示したグラフ）

項目特性曲線（縦軸に正答率、横軸に能力差を示したグラフ）

S－P表

<テスト結果処理の実例1>

テスト得点例

98	96	88	85	80	78	76	75	73	73
69	67	65	64	64	62	61	59	57	55
55	49	46	45	45	43	38	36	34	25

Q上のテスト得点例を見て累積度数表を作ってみよう

級間	度数	累積度数	級代表値
91～100			95.5
81～90			85.5
71～80			75.5
61～70			65.5
51～60			55.5
41～50			45.5
31～40			35.5
21～30			25.5

Q 平均値 モード メジアン レンジを出してみよう

<テスト結果処理の実際例2>

あるクラスの文法テストの結果

学 生 名	粗点 X	$x = (X-M)$	$x^2 = (X-M)^2$
P1	93	44	1936
P2	67	18	324
P3	58	9	81
P4	52	3	9
P5	48	-1	1
P6	45	-4	16
P7	39	-10	100
P8	35	-14	196
P9	29	-20	400
P10	24	-25	625
N=10	M=49		$\Sigma x^2 = 3688$

$$\text{標準偏差 } S = \sqrt{\frac{\sum X^2}{N}} = \sqrt{\frac{3688}{10}} = 19.20$$

$$\text{偏差値 } Z\text{スコア} = \frac{\text{粗点} - \text{平均点}}{\text{標準偏差}} = \frac{X - M}{S}$$

$$\text{偏差値 } T\text{スコア} = \frac{\text{粗点} - \text{平均点}}{\text{標準偏差}} \times 10 + 50 = \frac{10(X - M)}{S} + 50$$

Q P1 の学生の Z スコアと T スコアを出してみよう

Q P8 の学生の Z スコアと T スコアを出してみよう

<SP表分析>

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	得点
1	1	1	0	0	1	0	1	1	1	1	7
2	1	0	1	0	1	1	0	0	0	0	4
3	1	0	1	0	1	1	1	0	1	0	6
4	1	1	0	0	1	0	0	0	1	1	5
5	1	1	1	0	1	0	0	0	1	0	5
6	1	0	0	0	0	1	0	1	1	0	4
7	1	1	0	0	1	1	0	0	0	1	5
8	1	1	0	0	0	0	0	1	0	1	4
9	1	0	0	0	0	1	0	0	1	0	3
10	1	0	0	0	1	0	0	0	0	0	2
11	0	0	0	1	0	1	1	0	0	0	3
12	1	1	1	1	1	1	1	1	1	1	10
13	1	1	0	0	1	1	0	1	0	1	6
14	1	1	1	0	1	0	0	0	0	0	4
15	1	1	1	0	1	1	0	1	1	0	7
16	1	1	1	0	1	0	0	0	0	0	4
17	0	1	0	0	1	1	1	1	1	1	7
18	1	1	0	0	1	0	1	1	0	1	6
19	1	1	1	0	0	0	0	0	0	0	3
20	1	1	0	0	1	0	0	0	1	1	5
正答数	18	14	8	2	15	10	6	8	10	9	

	Q1	Q5	Q2	Q6	Q9	Q10	Q3	Q8	Q7	Q4	得点
12	1	1	1	1	1	1	1	1	1	1	10
1	1	1	1	1	0	1	1	0	1	1	7
15	1	1	1	1	1	1	0	1	1	0	7
17	0	1	1	1	1	1	1	0	1	1	7
3	1	1	0	1	1	0	1	0	1	0	6
13	1	1	1	1	0	1	0	1	0	0	6
18	1	1	1	0	0	1	0	1	0	1	6
4	1	1	1	0	1	1	0	0	0	0	5
5	1	1	1	0	1	0	1	0	0	0	5
7	1	1	1	1	0	1	0	0	0	0	5
20	1	1	1	0	1	1	0	0	0	0	5
2	1	1	0	1	0	0	1	0	0	0	4
6	1	0	0	1	1	0	0	1	0	0	4
8	1	0	1	0	0	1	0	1	0	0	4
14	1	1	1	0	0	0	1	0	0	0	4
16	1	1	1	0	0	0	1	0	0	0	4
9	1	0	0	1	1	0	0	0	0	0	3
11	0	0	0	1	0	0	0	0	1	1	3
19	1	0	1	0	0	0	1	0	0	0	3
10	1	1	0	0	0	0	0	0	0	0	2
正答数	18	15	14	10	10	9	8	8	6	2	

Q 11番の学生と問題3について分析しよう

★テスト採点の誤差

後光効果

寛容効果

ステレオタイプ

系列効果

中心化傾向

初頭効果

対比誤差

★項目応答理論（IRT）（Item Response Theory）

⇔ CTT（Classical Test Theory）

本テスト作成前に予備テストを行い、さまざまな推定値パラメータを測定する。そこから、順序尺度ではなく間隔尺度に基づいて得点等化を行う。

テストを問題や受験者集団にかかわらず、本人の能力を評価できる客観的な評価である。

★統計処理

1つのテスト結果の事例が他にも当てはまるかどうか、一般化できるかどうかを証明するために統計処理を行う。

t 検定

ふたつの集団の平均に意味差があるかどうかを測る

分散分析

ふたつ以上の集団の平均の差を比較するための分析

カイ二乗検定

測定値と期待値の誤差を測る計算

数値ではない名義尺度のデータを使って相関を測ることもできる

「漢字に興味があるかないか」

★日本語のテスト

JLPT 日本語能力試験

国際交流基金と日本国際教育支援協会

1984年 開始

2010年 新能力試験スタート

7月、12月年2回

N5～N1

日本語の文字 語彙 文法の知識 →言語知識

その知識をどれだけコミュニケーションに使えるかの運用力

→ 聴解 読解

「高度人材ポイント制」において、N1合格者は15ポイント付与。

日本留学試験（日本語）日本学生支援機構

外国人留学生として、日本の大学（学部）等に入学を希望する者について、日本の大学等で必要とする日本語力及び基礎学力の評価を行うことを目的に実施する試験

アカデミックジャパニーズ

6月、11月の年2回

海外での実施

インド（ニューデリー）、インドネシア（ジャカルタとスラバヤ）、韓国（ソウルとプサン）、シンガポール、スリランカ（コロンボ）、タイ（バンコク）、台湾（台北）、フィリピン（マニラ）、ベトナム（ハノイとホーチミン）、香港、マレーシア（クアラルンプール）、ミャンマー（ヤンゴン）、モンゴル（ウランバートル）、ロシア（ウラジオストク）の14の国と地域の17都市。

聴解 聴読解 読解 記述

BJT ビジネス日本語能力テスト

日本漢字検定協会

ビジネスコミュニケーション能力の測定

コンピュータ受験

J 5 ～ J 1 + の 6 段階 8 0 0 点満点

聴解 聴読解 読解

海外での実施

**中国（含 香港）、台湾、韓国、タイ、ベトナム、マレーシア、
インドネシア、ミャンマー、インド、シンガポール、メキシコ、
ブラジル、フランス**

**「高度人材ポイント制」において、BJT 4 8 0 点を取得していれば、
15 ポイントが付与。※「在留資格認定証明書交付申請」**

OPI オーラル・プロフィシエンシー・テスト

ACTFL

2 0 分間の口頭インタビュー試験

初級、中級、上、超級の 4 つの級。初中上級にはそれぞれ上中下がある

ウォーミングアップ レベルチェック 突き上げ ロールプレイ

機能タスク、場面内容、正確さ、テキストの型の 4 つの要素がある