

5月23日

テストと評価

第2章 英語学力測定論 担当 pp. 30-38

## 1. 英語学力の測定

### 1.1 測定の基本概念

□英語学力の測定は教育機関や実社会において様々な目的や方法により実施されている。

□測定の基本概念

①measurement：明確な規則と手順に従って、受験者の特性を数量化する方法

※数量化：対象となる特性に数値を割り当てること

②assessment：統計的かつ十分な根拠をもった手続きに従って、関心ある対象についての情報を収集するプロセス

③evaluation：アセスメントの結果を利用して何らかの判断や決定を行うこと（価値判断を含む）

### 1.2 測定の手続き

□測定の手続きは③段階で説明され、この手続きを経て得られたテスト得点は変数と呼ばれる。

第一：特性または構成概念を概念的に定義すること

第二：構成概念を操作的に定義すること

第三：観察されたものを数値化すること

□第一段階について

測定は「誰に対して」「どのような目的で」実施するのかを明確化し、それに適した方法で対象となる構成概念を定義する必要がある。構成概念の定義方法は①シラバスに基づく場合、②理論モデルに基づく場合の2通りが考えられる（Bachman & Palmer, 1996）。

□第二段階について

測定したい構成概念の推論を可能にする言語運用を、観察したり導きだしたりする手順や条件を明示する事が不可欠。手順や条件を明示することで構成概念を操作的に定義することができ、どのようなタスクを課しどのように採点するのかといった測定の手続きが明確化される。測定したい構成概念の指標として変数を解釈するための論理的根拠を与えるのが操作的定義であると言える。

□第三段階について

観察された言語運用に対する数値化の手順を定める段階。テストや質問紙、面接などを実施する場合は以下2通りの数値化の方法がある。

①あらかじめ定義された段階別の評定尺度に基づき、観察された言語運用の質や水準がどの段階に相当するかを判断する方法（受験者に言語産出を求めるタスク：ライティング・スピーキングテスト）

②個々のタスクや項目に対する応答に正答・誤答を与えたり、部分点を与えたりすることで得点化し、合計する方法（短答式の項目の場合：多肢選択問題・穴埋めなど）

□テスト得点を用いた統計分析結果が意味のあるものとなるためには、測定の信頼性・解釈の妥当性が確保されている必要があるが、それを判断するためには測定の手続き段階が明確化されていることが必要である。

### 1.3 測定尺度の種類

- 観察された言語運用を数量化すると、その数値は測定した構成概念をあらわす変数となる。
- 構成概念をどのように定義し、どのような規則・手順で測定したかによって数値の種類と情報量は異なる。数値の種類と情報量に応じて定義されたのが以下 4 種類の尺度であり、①→④の順に情報量が増す。
  - ① 名義尺度：集団や個人の属性を区別するための尺度
  - ② 順序尺度：測定したい特性の度合いの異なる水準に、数値を順番に割り当てた尺度  
水準の区別に加え、「どちらが大きいか」という順序性の情報を含む
  - ③ 間隔尺度：ある特性の水準間の間隔が等しい場合に、それらの異なる水準を数値化した尺度  
水準の区別・順序性に加え、「どのくらい大きいか」という情報を含む
  - ④ 比率尺度：区別性・順序性・等間隔性に加えて絶対原点（その特性が全くない状態）が存在する尺度。ある数値が別の数値に対して何倍大きいかという情報が得られる。

## 2. テスト得点の記述統計

### 2.1 得点分布

- テスト得点を統計処理する前に度数（ある階級に含まれるデータの個数）の分布図を描き、形状を確認しておくことが重要。
- 得点分布にはさまざまな形状があるが、低得点から高得点まで左右均等に分布し、中央の度数が高くなっている分布の代表的なものが正規分布である。
  - 中央の度数が大きく正規分布より分布の形状がとがっている場合→プラスの尖度
  - 中央の度数が小さく分布がなだらかな場合→マイナスの尖度
  - 高得点に得点集中・低得点に裾を引く→マイナスの歪度・分布は負に歪んでいる
  - 低得点に得点集中・高得点に裾を引く→プラスの歪度・分布は正に歪んでいる
- 分布が正規分布にどれくらい近似しているかということが相関係数を解釈したり、テスト得点分布の違いについて推論したりする場合に重要となる。
  - 集団基準標準偏差テスト：正規分布であることが予想される
  - 目標基準標準偏差テスト：尖度がプラスで負に歪んだ分布であることが予測される

### 2.2 代表値と得点の散らばり

- 度数分布図で得点分布の形状を確認した後、より正確に得点分布の傾向を記述するために、代表値と得点の散らばりの程度に関する統計量を求める。
- 主な代表値
- 代表値：分布を代表する値のことで、得点の中心傾向を表す。代表値としては最頻値・中央値・平均値が良く用いられる。
- 最頻値：もっとも多くの受験者が得た得点
  - 全ての受験者情報は反映していない。名義尺度・順序尺度の値に適切な代表値の指標。
- 中央値：受験者を得点の低い方から高い方に順に並べたときの中央の得点。受験者が偶数の場合は中央値が 2 つとなるので、その得点を合計し 2 で割った数を中央値とする。
  - 全ての受験者情報は反映していない。順序尺度・間隔尺度の値に適切な代表値の指標。

分位数：テストを得点順に並べ、低い方から 25%の値→第一四分位数、50%の値→第二四分位数、75%の値を第三四分位数という。  
→順序尺度や歪みが強い分布に有効活用可能

《補足：分位数はどのような場合に使用されるのか》

■四分位数は、平均値や標準偏差のように複雑な計算をせずに、データの分布状況を調べたいときに便利である。四分位数の求め方は複数存在し、統計ソフトや書籍によって様々である。当然結果として求められる数値も計算方法によって微妙に異なるが、データの分布状況を調べるという目的に対して負の影響をおよぼすほどの差異はない。特にデータが多数ある場合には全く問題はないと考えられている。

平均値：受験者の得点を全て合計して受験者数で割った値

→全ての受験者情報を反映しており、外れ値の影響を強く受ける。間隔尺度・数値計算をする場合・推測統計に最適

□分布に歪みが有る場合には最頻値・中央値・平均値のすべてを記述しておくことが望ましい。

□代表値に加え、得点の散らばり具合の程度も得点分布を記述する上で重要な指標であり、最もよく用いられる散らばりの程度の指標が標準偏差である。

□標準偏差：間隔尺度の値の散らばりを表す最適な指標。標本の分布を記述する場合に用いる。

※標準偏差→偏差の二乗和を受験者数  $N$  で割った標本分散の平方根

□不偏分散：標本をもとに母集団の分布を推測したい場合に用いる。

※不偏分散→偏差の二乗和を受験者数  $N-1$  で割った標本分散の平方根

□標準得点：同じ集団に実施された異なるテストの得点を相互に比較したい場合

※標準得点：得点から平均値を引いて標準偏差で割った値。標準得点を 10 倍して 50 を加えた値は  $Z$  得点と呼ばれる（日本では偏差値得点とも呼ばれる）。

### 2.3 テスト得点間の関係

□2 つ以上のテストを同じ集団に実施し、得られた 2 つ以上の得点分布の関係を調べる場合もある。一方の変数の増加につれて他方の変数も増加する場合を正の相関関係があると言い、逆に一方の変数の増加が他方の変数の減少に対応している場合を負の相関関係があるという。

□変数間の関係を表す統計量が相関係数であるが、その種類は様々なので測定尺度の種類と分布の状況により、適切な相関係数を用いる必要がある（代表例：ピアソンの積率相関係数、スピアマンの順位相関係数）。

□積率相関係数：

①2 つの変数が間隔尺度であること

②対になっている変数はそれぞれ他の変数から独立していること

③正規分布であること

④線形性があること

□順位相関係数：2 つの変数間に線形の関係はあるが、どちらかの変数が順序尺度で構成されている場合に用いられる。

□いずれの相関係数も、 $-1$  から  $1$  までの値となり、相関係数の絶対値が大きくなるほど関係が強いことを示す。

□相関係数の解釈の際の注意点

①相関係数と因果関係は別物。相関係数が高いからといって必ずしも 2 つの変数間に因果関係があることを意味するわけではない。

②第 3 の変数の存在で 2 つの変数間に見かけ上の相関がみられる場合、偏相関係数を用いる。

③データが正規分布ではない場合・測定誤差が存在する場合・2 つの変数間が線形関係に無い場合・一部の数段だけを見ることにより選抜効果または切断効果が存在する場合、外れ値が有る場合、層別化されたデータの場合、受験者の能力の範囲に制限がある場合などの相関係数の値には注意が必要

□相関係数を意味あるものとして解釈するためには、2 つの変数間にどのような関係があるのかについて、理論や先行研究、経験に基づいて論理的な理由が説明されることが重要である。