

II 基本条項の解説

0章 序

「テストの開発，実施，利用，管理にかかわる規準（以下，本規準と略記する）」の目的は，テストの開発，実施，利用，管理の過程において留意すべき事項について，目指すべき方向性を具体的に示すことである。

そのために，本章においては，まず，0.1で本規準の目的について述べる。次いで，0.2では本規準で対象とする「テスト」の定義を試みる。本規準が適用できる領域は日常用語の「テスト」よりも広く，テストという名称が用いられていなくとも，本規準の対象に含まれる場合があることがわかるであろう。0.3，0.4では，そのテストの対象領域および対象者の範囲を明らかにしておく。

0.5では本規準における用語の統一方針を示す。テストに関連して実際に用いられている用語はさまざまである。本規準では，同一の概念に対して可能な限り同一の用語を使う方針をとった。

これらの条項は，1章以降のガイドラインに基本的な方向づけをするものである。ただし，テストの対象領域や対象者を狭く限定したり，テストの営みをいたずらに制限しようという試みではない。本規準は日本の現状を踏まえた上で，テストにかかわる多様な局面においてテストの適切な使われ方を志向するものである。そのため，それぞれの状況に応じた柔軟な理解と本規準の適用が望まれる。

0.1 テスト規準の目的

テストに関係する者は，テストの結果がもたらす社会的影響の大きさを認識し，テストの開発，実施，利用，管理にいたるすべての過程で適切な方策を考える必要がある。本規準はテスト関係者がそうした社会的責任を果たすための指針を示すものである。

今日「テスト」という名で使われているものには、自分ひとりでも試せるクイズ形式のような簡単なものから、何万人の人が受ける大規模な入学試験や検定試験などさまざまな形がある。おびただしい数のテストの中で、どれがよいテストでどれがそうでないかは一見しただけでは簡単に判断できない。ていねいに作られたテストもあれば、見かけだけのものもある。

テストがもたらす結果は人に与える影響が大きいだけに、その扱いは慎重でなければならない。テストが社会で広く使われるようになるにつれ、テストが与える弊害もまた無視できない状況になっている。

よいテストを作り、実施し、利用し、また管理していくには、テストに関係する人がそれぞれの段階でさまざまな条件を考慮し、よいテストはどうあるべきかを考え、責任をもって適切に対処することが重要である。またそうすることで、テストが本来もつ効用が活かされてくる。そのためには、指針を与えるための規準を定めることが必要である。なお、本規準の対象となるテストとは、人間の特性を測定するものである（「0.2 テストの定義」参照）。機械やプログラムの動作の確認など、日常的にテストということばが使われていても人間の特性の測定にかかわらないものは本規準の対象としない。

本規準は、日本テスト学会が目指す社会的活動のひとつとして、日本テスト学会の責任において作成したものである。すでに多くの先進諸国がこのような趣旨の規準を用意しているのを参考にしつつも、日本の現在の状況を深く考慮し、日本においてテストの実践がより機能的に行える環境を整備することを目的としている。本規準がすべてのテスト関係者が目指すべき方向性と目標をはっきりさせるのに役立ち、テストが日本社会に適切に貢献することを期待する。

0.2 テストの定義

本規準で対象とするテストとは、能力、学力、性格、行動などの個人や集団の特性を測定するための用具であり、実施方法、採点手続、結果の利用法などが明確に定められているべきものである。したがって、本規準は心理学的なテスト、学力・知識試験はもとより、行動評定、態度評定などの評定手法、調査のほか、構造化された面接、組織的観察記録

II 基本条項の解説

にも適用され得るものである。

テストはある目的を遂行するために、人間の**特性**を測定するための用具である。ここで挙げるように、能力、学力、性格、行動など、人間の特性は多様である。これらに加えてさらに、興味や態度の特性を加えることもできる。原則として、テストはある程度の期間安定して持続する人間の特性を測定するものであり、それがどのような名前と呼ばれようと意味のある目的で測定されるならばそのための用具はテストと呼ぶことができる。

人間の特性を測定する際の測定対象はもちろん人間であるが、集団を測定対象とし、集団の特徴を明らかにするためのテストを用いることもある。たとえば、組織や文化の比較などに関心がある場合には、関心の対象は集団である。また、実験的方法で人間について知識を深めようとする場合には、人為的に設定した実験条件間の差を問題とするが、この場合も個人よりも集団の特徴に関心がある。個人を対象にするにしろ、集団を対象にするにしろ、テストを実施するに際しては目的を明確にすべきである。

目的を明確化することによって、テストを作るにしても使うにしてもその良し悪しを論ずることができる。逆にいえば、テストの開発、実施、利用などにかかわる複数の選択枝*について、それぞれがよりよい方向へ向かうか否かの判断をするためには、その前提としてテストを開発する意図やテストを利用する目的が明確になっている必要がある。

何をテストと呼ぶかは多分に状況依存的であり、テストの定義をすることには、理論的な難しさというよりも、もともと境界がはっきりしないところに境界を置くような難しさがある。

本規準で意図していることは、学問的な定義ではなく、実際のテストについての指針を示すことである。したがって、テストを厳密に定義することはここでは重視しないが、すべてのテストは実施方法、採点手続、利用方法が明確に

* 「選択枝」という用語が広く使われていると思われるが、「肢」が当初当用漢字になかったことから枝が使われたが、身体用語を避けたいということ、英語で設問部分を「幹」(stem)といい、選択枝が「枝」(branch)にあたることから、本規準では「選択枝」という用語を採用することにした。

定められているべきであると考える。

一般的にはテストとは呼ばれない測定や評価のプロセス、たとえば、行動評定、態度評定などの評定手法、調査のほか、構造化された面接、組織的観察記録に対しても、本規準は適用されるべきである。

0.3 テスト規準の対象領域

本規準は、すべての領域のテストを対象とする。すなわち、教育機関における選考、評価、指導、企業などの組織における選考、人事評価、能力開発・指導および資格認定、臨床診断、カウンセリング、各種ガイダンス、公共的施策決定など、多様な領域のテストのすべてを対象とする。

本規準では、テストとは、原則として、ある程度の期間安定した人間の特性を測る手段として広くとらえられている（「0.2 テストの定義」参照）。また、本規準は、実施方法、採点手続、結果の利用法などが明確に定められているテストを望ましいものとして、そのための努力をすべきであると主張している。しかし、具体的に手続が明示されていない測定用具が、そのことを根拠として本規準が示す方向性や示唆の適用範囲から外れるわけではない。

通常、テストとは呼ばれないような評定、評価、アンケート、査定、調査なども、本規準の適用範囲に含まれる。実際、日常的に「テスト」と呼ばれていないものにこそ、順守してもらいたい条項もある。

テストが使われる領域や分野を大きく分けると、教育場面、臨床場面、企業や公共の場面における施策決定などが考えられる。テストということば自体は教育場面で最もよく使われるであろう。教授活動とテストを使う評価活動とは一体となっており切り離せない。教師の日々の評価は固定した枠組みによらず、自由になされるべき側面があり、教室場面で日常的に行われる評価などでは本規準がそのままでは適用されにくい。教育の成果を、より客観的で大規模に実施される標準テストによって評価したい場合も少なくない。標準テストが満たすべき要件は本規準が要請する要件に一致する。また、高校や大学で行われ

II 基本条項の解説

る入学試験も、それぞれの教育を享受すべき学生を選抜する重要な公共的役割があり、本規準の主要な対象のひとつと考えられる。国や地方公共団体などで教育などに関する施策の検討のために使われるデータも、施策に反映させようとするならば、データの意味が明確で正しい解釈を導くものでなければならず、本規準の要請を満たすべきである。

次に、企業では、採用試験のような選抜の目的や、職務に適切な人材を配置するための情報を得る目的でテストが使われることが多い。また、テストは社会全般の重要な施策決定のための情報を提供する。テストは心理臨床場面や医療機関でもカウンセリングや診断の参考資料などとして利用するために頻繁に使われる。これらのテストも本規準は対象とする。

テストは以上のような実際の目的のためだけに使われるものではない。心理学などの研究においては、心理学的な**構成概念**を確認するのにテストが使われる。また、テスト開発のための理論的・実験的研究もある。

テストの実施目的や利用状況は、このようにさまざまである。本規準は、個々の状況に共通する要素を考慮し、一般的に実現すべき方向性を示そうとするものである。

0.4 テスト規準の対象者

本規準は、テストにかかわる研究者、開発者（設計・開発する者）、頒布者（提供・頒布する者）、実施者（実施・採点する者）、利用者（結果を利用する者）、管理者（結果の情報管理に携わる者）、そして受検者などすべての関係者を対象とする。

ひとつのテストが作られて役目を果たすまでには、研究、開発、頒布、実施、利用、管理、そしてふたたび研究という循環する流れがある。学校の先生が教室内で自作のテストを使う場合のように、規模が小さく、すべての作業をひとりで担うものもあるが、**知能検査**など多くの人の協力で作られるテストのように、開発者、頒布者、実施者、利用者、管理者、研究者が、作業を分担して行うものもある。

開発者とは、テストという測定用具を、何も無いところから完成された形に作り上げる者を指す。質問項目を作成する者も開発者に含まれる。開発者は、テストの基本設計（「1.1 テストの基本設計」参照）を立案し、具体的に、質問形式や回答形式はどうするのが適切か、採点手続はどうするのか、尺度化の方法はどうするのかなどを決め、質問項目を作成していかなければならない。さらに、**尺度**の性質を詳しく説明する手引を作成することも必要である。いったん完成したテストを後に改訂する作業も、開発者の仕事である。

頒布者とは、完成したテストを複製し、それを実施者および利用者に提供する者やテスト、手引および解説書などを編纂し、出版する者などを指す。また、インターネットなどを通してテストを提供する者も含まれる。頒布にあたっては、テストの基本設計（「1.1 テストの基本設計」参照）にそって適切に実施、利用が行われるように十分に配慮する。

実施者とは、開発者が作成したテストを実際にとり行う者や、実施後に採点作業を行う者を指す。実施者は、会場での実施手順を順守し、採点にあたっては、基本設計にそって作業を遂行するようにする。

利用者とは、受検者に対して何らかの判断を行うためにテストの結果を利用する者を指す。たとえば、生徒の学習指導に役立てるための情報を必要としている教師や教育関係者、入学試験の可否や入社試験の採否を決定する者、テスト結果をもとに受検者の処遇を決める者がそれにあたる。また、利用者がテストの開発者や実施者と同一であるとは限らない。たとえば、医師や相談員がみずから直接テストを実施せずに、他の者が実施したテストの結果を診断や処遇の決定の参考として用いるような場合には、医師や相談員が利用者にあたる。公的な資格試験の結果を受検者から提出してもらい、それを利用して判断することもある。利用者はテストの目的や内容を熟知し、基本設計にしたがってテスト結果の解釈を行う必要がある。

管理者とは、テストにかかわる情報を保管する立場の者を指す。テストの情報には受検者に関する情報と、使用した質問項目に関する情報などが含まれる。両者は相互に関連性がある。管理者はそれらの保全と管理に責任をもたなければならない。

研究者とは、テストそのものを対象に研究を行う者である。テストの開発や

II 基本条項の解説

実施、テスト用具などにかかわる研究、テストの理論的研究やテスト結果の分析などテストの一般的な技術にかかわる研究のほか、特定のテストの測定内容にかかわる研究を行う者も含まれる。研究成果を開発者に還元することにより、より良いテストの開発につなげていくことができる。

ここにあげる基本条項のひとつひとつがこれらすべての人に等しく関係するわけではないとしても、テストの健全な運用を可能にするには、テスト全体の流れを把握し、その中でそれぞれが担う役割を理解しておくことが、どの人にも求められることである。

さらに、テストの受検者も本規準の対象に含まれる。テストを受ける際には、受検者自身も、みずからがもつ権利と義務を理解し自覚して行動し、諸々の規則および規範を順守することが求められる。

本規準は、そうした意味でテストにかかわるすべての人にとっての指針となるべきものを目指して作られている。

0.5 テスト用語の表記

テストの使用場面や状況はさまざまで、適用分野や発展経緯の違いによって、同じ意味をもつ概念に異なった用語が用いられているものがある。本規準では、できるだけ共通性をもち、偏りの少ない、平易なひとつの用語によってその概念を表記するものとする。

テストと類似した用語には、検査、考査、試験、クイズ、調査、アンケートなど、使われる状況や用途で、さまざまな言い方がある。また、テストを受ける人の呼び名も、入学試験や検定試験の状況では受検者が一般的であろうし、心理実験の場面では実験者に対する被験者や実験参加者、実験協力者という言い方が一般的である。また、病院や相談所で心理検査を受けるような場合は検査者に対して受検者と呼ばれることが多い。しかし、特定場面で使われるテストを想定するのではなく、テストに共通する一般的な話として「テストを受ける人」を考える場合、こうしたことばの不統一はテスト全般に通じる規準を考えようとする試みにとってひとつの障害となる。そこで、複数の用語を代表して

何かひとつのことで表現することが望ましい。本規準ではいくつかの用語を選び、それぞれ同じ概念を表す複数の用語の代表語として用いている。慣用的な用語を聞きなれた人にはやや違和感を覚えるものがあるかもしれないが、できるだけ偏らず、意味もわかりやすいという基準で用語を選定した。

そのような例として、以下のようなものがある。

- テスト（一定の質問対回答の形式を通じて、判断に必要な情報を収集する手段）：検査，考査，試験，クイズ，調査，アンケートなどを含む。
- 予備テスト：予備調査，試行テスト，プリテスト，プレテスト，パイロットテストなどを含む。
- 受検者（テストを受ける人）：受験者，被検者，被験者，実験参加者，実験協力者，クライアントなどを含む。
- 質問項目（受検者に提示される質問）：問，問題，設問，質問，項目，課題，アイテムなどを含む。
- 回答（質問に対する受検者の反応）：答，答案，解答，応答などを含む。
- 選択枝（受検者が正答として選ぶべき複数の候補）：選択肢，カテゴリーなどを含む。
- 評定者：採点者，評価者，判定者，審査員などを含む。
- 面接員：面接者，面接官，面接担当者などを含む。
- 素点（回答を直接点数化した数値，得点）：点数，得点，素得点，粗点，スコア，測定値，実測値，観測値，観察値などを含む。
- 尺度得点（回答から意味ある尺度に換算した数値）：尺度値，換算点，変換点，偏差値，パーセンタイル順位，等化尺度値，個人特性値，特性値，尺度特性値，モデル準拠得点などを含む。素点も尺度得点のひとつとみなす。
- コンピュータテスト（コンピュータを利用したテスト）：コンピュータ基盤テスト，CBT（Computer-Based Testing），IBT（Internet-Based Testing），WBT（Web-Based Testing），e テスティングなどを含む。

また、原語を和訳する際に複数の訳語が存在するものもある。訳語についても同様の方針により、ひとつの代表的な訳語によって表記する。

1章 開発と頒布

本章では、テストの開発および頒布にかかわる重要事項に関して15の条項にまとめている。本規準の各章の中で最も条項数が多い。理論的な部分が多く、やや難解で必ずしも理解しやすすくない内容もあるが、本規準の根幹をなす重要な章である。1.1～1.5は主としてテストの開発および設計にかかわる内容について述べている。1.6～1.9は回答結果の得点化および尺度化にかかわる条項である。1.10～1.14はテストの品質保持にかかわる条項であり、1.15は開発者および頒布者と実施者および利用者をつなぐ手引の役割にかかわっている。

1.1はテストを開発する際の基本設計について、1.2ではテストの測定内容の定義とその構造化について方針を示している。この2つの条項はテスト開発の基本である。1.3は質問項目の設計にかかわる留意点について、1.4は回答方法の設計、1.5は採点手続の設計について述べている。

1.6はテストの結果の表示方法である尺度得点の構成について、1.7ではその各論として尺度の標準化について解説している。1.8、1.9は複数の尺度を共通化して、尺度得点を比較可能にする方法について方針を示す。

1.10は尺度得点の信頼性、1.11は尺度得点の妥当性という、テストの品質の検証にかかわる重要な2つの柱となる概念に関する解説とこの概念を使ってテストをどのように評価するかを示す。1.12は繰り返し実施されるテストの改訂について述べている。1.13は質問項目の内容開示というテストの品質保持に関連する重要な問題について方針を示す。1.14は翻訳の問題を扱っている。

1.15では手引、解説書に記載すべき事項について述べている。

本章の15の条項は、テストの開発にかかわる基本的な問題を取り上げており、テストの品質向上とそれに伴う合理的な意思決定には欠くことのできない重要な内容である。

1.1 テストの基本設計

開発者は利用目的や場面にあわせて、測定内容、測定形式、実施方法・手続、結果の利用方法、適用を想定する対象者の範囲などを明確に定め、基本設計を行う。

テストの開発にあたっては、基本設計を行うことが必要である。基本設計を行わずにテストが開発されると、そこで測定される特性が曖昧になり、目的外の利用をされたり、開発意図から外れた変更がなされたりする可能性がある。そうした誤用を防止する意味からも、しっかりとした基本設計を行うことが重要となる。

テストの基本設計にあたっては、以下の諸項目について明確に定めておくことが必要である。

1 テストの目的

テストを作成する場合には、どのような目的と場面でそのテストが利用できるのかを明確にする。選抜や資格付与など、ある特定の明確な目的をもって特定の場面で利用されることを想定したものなのか、一定の範囲で汎用性をもたせた設計になっているのかを区別することは重要である。テストの利用目的と場面を明確に規定することにより、実施者、利用者が目的に合致したテストを適切に選択することが可能となる。

2 測定対象と測定内容

テストの実施によって利用の目的に合致した有用な情報が得られるように、測定対象となる特性を明確に規定し、それに合致した測定内容をもつテストを構成する必要がある（「1.2 測定内容の定義と構造化」参照）。

3 受検者層の想定

想定される受検者層によって、適切な測定内容、測定形式、実施方法などが異なる場合がある。障害者や外国人など、特別な配慮を必要とする受検者を含

II 基本条項の解説

む場合には、測定結果に対するそれらの属性の影響を可能な限り小さくするように留意する（「2.5 回答方法の説明」、 「2.6 障害者などへの配慮」、 「2.7 少数受検者への配慮」 および 「5.3 テスト画面や操作方法のわかりやすさ」 参照）。

4 測定形式

テストの目的の達成に適した測定形式を選定する。測定形式を回答形式から分類すると、たとえば、記述式、選択式、口述式などがある。また、コンピュータの発達により、従来にはなかった新たな測定形式や回答方法も利用可能になってきている（「1.3 質問項目の設計」、 「1.4 回答方法の設計」 および 「5.1 コンピュータテストの基本的な性質」 参照）。

テストの特性は、質問項目の内容だけではなく、ことば、図、音声、画像などの質問項目の提示方法、記述式、選択式、口述式などといった回答形式、質問項目数、選択枝数、回答時間、配点など、さまざまな要因によって変化する。質問項目が繰り返し利用されることを想定している場合には、毎回のテストの特性の等質性、連続性を確保するため、質問項目数、選択枝数、回答時間などの開発者が決められる要因は、与えられた時間や設備などの条件の制約の許容する範囲で、一定の形に整えておく必要がある。

テストの目的、方針などの変更により、テストの内容や測定形式が変わり、従来のテストとの等質性や連続性が保たれないと考えられるとき、テストは改訂されるべきである。このときは、同じ目的を継承していたとしても、実質的に新しいテストを開発しているといえる。テストの基本設計を改めて見直すとともに、改変について実施者、利用者および受検者などに知らせ、さらに、新しい測定形式についての解説も加える（「1.12 テストの改訂」、 「1.15 手引、解説書の内容」 および 「2.3 実施前の受検者への説明」 参照）。

5 テストシステムとしての配慮

テストの開発にあたっては、質問項目の内容はもちろんのこと、測定形式、回答時間、実施方法、回答の収集方法、採点方法、質問項目や回答結果の管理および各種情報の公開までを含めて、全体のシステムを配慮した設計を行う。

テストの開発に必要な情報を集め、現実的に実現可能な条件のもとでテストの品質を保って目的にそった利用が可能となるように、さまざまな要因について十分に検討をした上で、テストの骨格を設計するべきである。とくに、質問項目を一度だけ用いるのか、繰り返し用いるのかという区別は基本設計に大きな影響を与える。質問項目を繰り返し用いる場合には、測定形式、テスト用具および質問項目の構造とその管理から実施時の不測の事態への対処、不正行為への対応や実施後の質問項目内容の秘匿にいたるまで、細心の配慮のもとでテストシステム全体の設計を行わなければならない（「1.13 質問項目の内容開示の是非」、「2.2 テスト用具の管理」、「2.4 手順どおりの実施と不測の事態への対処」、「4.1 テスト結果の保管手続」、「5.4 機密保持や実施管理への対策」、「6.4 著作権の尊重」および「6.5 受検者の義務」参照）。

同じ目的で繰り返し行われるテストでは、内容面だけではなく、テストとしての性能においても各回のテストができるだけ等質となるような尺度化の方法も考慮して作成する（「1.6 尺度化の方法」参照）。得点分布、信頼性や妥当性の評価などにかかわる統計情報を用いて等質性を容易に検討することができる形式が望ましい（「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。たとえば、客観式テストであれば、質問項目数や選択枝数などの形式を統一する。

学習や発達の状況を比較するテストでは、各回の実施結果に見られる差異が意味のある変化としてとらえられるように、結果が共通尺度として表現されるような構成が必要である。テストの作成にあたっては、そのことを配慮した上で事前に必要な手立てを設計に組み込む（「1.3 質問項目の設計」、「1.4 回答方法の設計」、「1.6 尺度化の方法」、「1.7 尺度の標準化」、「1.9 複数のテストの共通尺度化」および「3.3 判定基準の個別検証の必要性」参照）。

6 ミスの防止

テストの実施および利用の過程で起こり得るミスを想定し、事前に防止する方策を立てておかなければならない。質問項目の作成時には、項目内容の表現、語法、提示順序およびレイアウトのみならず、実施時の受検者への回答方法の説明、不注意によって誤った回答が導かれることを防止するための教示、採点

II 基本条項の解説

者に対する指示など、ひとつひとつの具体的な場面で誤解や誤解に基づくミスが起こらないように配慮をすべきである（「2.3 実施前の受検者への説明」および「2.5 回答方法の説明」参照）。また、想定外のミスが起こった場合の対応も考慮したシステムを開発段階から準備しておく（「2.4 手順どおりの実施と不測の事態への対処」参照）。

7 結果の利用方法

開発者は、テストで得られる結果をどのように利用すべきかについて、あらかじめ明記しておくこと、および、テストにかかわる統計的な情報をなるべく公開し、利用目的を逸脱しないように利用方法の限定をすることが必要である（「1.15 手引、解説書の内容」、「3.2 拡大解釈の防止」および「3.5 統計情報の公開」参照）。

1.2 測定内容の定義と構造化

開発者は、測定しようとする能力、学力、性格、行動などの特性を明確に定義し、それが表現できるような適切な尺度を構成する。なお、測定しようとする特性が複数の下位の特性で構成される場合は、その構造を明らかにし、それらも測定できるように複数の下位尺度を設計する。

テストを作成するにはまずテストが測定しようとする特性を定めることが原則である。測定内容の定義が曖昧なままに作成されたテストの結果は適切に解釈することができず、結果を有効に利用することができない。

1 特性の構造

テストを作成する際には、まず、測定対象とする**特性**、すなわち、測定内容を明確に定義する。測定対象は階層的な構造として考えられる場合もある。たとえば、「英語力」という特性を「日常生活のさまざまな場面で英語を運用する能力」と定義し、これを「英語を読む力」、「英語で書く力」、「英語を聞く力」、「英語で話す力」の4つの下位特性で構成されると考えた場合には、「英

語力」を測定するテストを設計するには、4つの下位特性を明確に定義し、さらに特性間の関係を把握することが重要となる。「英語力」はこれら4つの下位特性をそれぞれ適切に測定する下位テストの結果を総合することで評価可能となる。測定対象を決定するに際し、上位特性（先述の例では「英語力」と下位特性（「英語を読む力」など）との間の関係や構造が、調査、研究によってあらかじめ明確にされていることが望ましい。

測定対象に関係する既存のテストなどに関する記録について調査し、参考資料とする（「3.5 統計情報の公開」参照）。

2 尺度構成と尺度得点

多くの場合、それぞれの下位特性を測定する複数の質問項目が用意され、回答をもとに**尺度**が構成される。各受検者の回答結果は、**尺度得点**としてその尺度上に位置づけられる。尺度の構成、それにいたる採点手続などを考慮の上、質問項目の構成を考えておく（「1.5 採点手続の設計」および「1.6 尺度化の方法」参照）。

3 測定対象と特性値との1対1対応

1つの特性または下位特性は、1つの尺度で表すことを原則とする。そうでなければその尺度得点が何を意味するか不明瞭になるおそれがある。1つの尺度を構成する質問項目が幅広い範囲の内容を含む場合もあれば、比較的狭い範囲の内容を詳しく問うている場合もある。原則として、尺度の名称は測定しようとする特性にふさわしいものでなければならない（「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照）。

4 尺度の性質

テストの利用者がテストの結果を適切に利用できるように、平均、**分散**、**標準偏差**、**得点分布**、信頼性および他の尺度との**相関係数**などの尺度得点の性質を示す資料を作成し、公開する（「1.10 尺度得点の確からしさの推定と公開」および「3.5 統計情報の公開」参照）。

1つの特性を1つの尺度で表現することが難しいときには、さらに複数の下

II 基本条項の解説

位尺度に分解することも考えられる。それぞれの下位尺度に対して、それが何を測定目的として作られているかを確認する。

下位尺度の尺度得点を総合して上位尺度が定義されているときには、それら下位尺度と上位尺度間の関係および下位尺度の尺度得点相互の関係を示す資料も用意すべきである（「3.5 統計情報の公開」および「4.2 テスト実施後の調査と検討」参照）。

下位尺度に分割することによって、下位尺度の性能が著しく低くなる場合には、下位尺度への分割は避けた方がよい。とくに、少数の質問項目からなるテストの場合、そのような危険性がある。

5 下位尺度の互換性

学力テストにおいて回答する科目の選択を許すなど、各科目の得点、すなわち、下位尺度を相互互換的に扱うのであれば、各下位尺度の尺度得点が基本的な共通性をもつようにテストを設計しなければならない。通常、質問項目の作成段階の工夫のみによってこれを実現するのは難しいので、尺度得点が比較可能になるように共通尺度化する必要がある（「1.9 複数のテストの共通尺度化」参照）。共通尺度化を行わないまま下位尺度を相互互換的に扱うことは避けなければならない。

6 測定内容の組み換え

異なる特性を測定するテストとして使われていた質問項目を組み合わせて新たに別のテストを設計する場合がある。その際には、新たなテストを設計する理由や元のテストおよび尺度との異同がわかるように、新たな測定対象、測定内容、尺度を定義し、テストの名称も新しくすべきである。また、従来からのテストとの関係がわかるように新旧の尺度得点の関係を分析して両者の関連を明示することも必要である（「3.5 統計情報の公開」参照）。

7 テストの内容にかかわる情報の公開

測定対象、測定内容、尺度の性能および尺度得点の分布などは、テスト実施後に調査や分析を行い、状況に応じて適切な時期に適切な方法、適切な範囲で

公開する（「4.2 テスト実施後の調査と検討」参照）。ただし、質問項目の内容は実施後においても開示できない場合があることに留意する（「1.13 質問項目の内容開示の是非」参照）。

1.3 質問項目の設計

質問項目の作成には、測定内容に関する十分な知識はもちろんのこと、その内容を測定するにふさわしい質問形式や表現についての十分な知識と経験が必要である。開発者は、質問項目作成のための手引を用意する。また、質問項目は回答者が質問の意図や回答方法を正しく理解できるような明快で統一のとれた表現を用いる。

1 質問項目と作成者

質問項目はテストで測定したい特性を十分に測定できる内容でなければならない。たとえば、文章の大意を要約する力を測定するためには、ある程度長い文章を読ませる必要があるが、漢字の読みの知識を測るために長い文章を読ませる必要はない。テストが測定対象とする特性に応じて適切な内容と形式の質問項目を作成することが大切である。原則として、1つの質問項目は、1つの特性のみを測定するように作成する。そうでないと、質問項目が多義的になり、回答者が答えにくく、その項目を含めた尺度の信頼性を低めることになる。

測定対象となる特性、測定内容の構造、結果の利用目的、実施規模、想定される受検者などに応じて、適切な質問項目の形式や具体的な質問項目の内容を決める必要がある。したがって、質問項目作成者は、測定内容の専門家であると同時に、質問項目の選定、作成に対する知識、経験をもつ者であることが望ましい。

2 質問項目作成の手引

開発者は、目的に合致したテストを作成し、質問項目の特性を維持していくために、質問項目作成のための規準を整備し、質問項目作成に関する手引を用意することが望ましい。質問項目作成者は手引を活用し、担当者によって解釈

Ⅱ 基本条項の解説

が異なることがないように心がける。手引は質問項目作成の過程で随時改良し、作業の実情に合致した内容とする（「1.15 手引, 解説書の内容」参照）。

3 質問項目作成者の選定とトレーニング

質問項目の作成においては、質問項目の内容および表現が、テストの目的に照らし合わせて適切であることが求められる。

質問項目の作成者が、測定対象や内容の専門家であっても、質問項目の作成経験がない場合もある。そのような状況では、質問項目作成者に対して質問項目作成の手引にしたがったトレーニングを行うことも必要である（「1.15 手引, 解説書の内容」参照）。

4 テストに用いる測定形式

テストの目的に即した測定を行うには、測定形式が重要である。たとえば、ことば、図、音声、画像などの質問項目の提示方法、記述式、選択式、口述式などといった回答形式、質問項目数、選択枝数、回答時間、配点など、さまざまな仕様を決める必要がある（「1.4 回答方法の設計」参照）。コンピュータを利用したテストにおいては、多様な測定形式が可能となっている（「5.1 コンピュータテストの基本的な性質」参照）。これらの決定にあたっては、受検者の人数や特性、テスト実施環境、また、利用の仕方によっては複数のテストの互換性などを考慮しなければならない。

測定形式の違いによる測定結果の差異に対する考慮も必要であり、そのための研究の蓄積が求められる（「1.10 尺度得点の確からしさの推定と公開」, 「1.11 尺度得点の適切さの確認」, 「4.2 テスト実施後の調査と検討」および「5.6 コンピュータテストと通常のテストの併用」参照）。

5 誤解の防止

質問項目の作成時には、テストの過程にかかわる人たちに誤解を与えない工夫を心がけるべきである。たとえば、回答指示文の表現、質問項目の表現、提示順序、レイアウトおよび採点者への指示などをわかりやすく工夫し、明確な表現で手引に記載する。同じ回答方式を用いる質問では指示文の表現を同じ形

に統一することが望ましい（「1.15 手引，解説書の内容」，「2.5 回答方法の説明」および「2.8 採点手順の順守」参照）。

6 質問項目の評価

質問項目の作成に際して，実際に測定される特性が測定目的の特性の内容を忠実に反映するように努力をするべきである。そのため，質問項目はあらかじめ余分に作成しておき，それらが適切かどうか，実施前にあらかじめ評価を行って，実際に使用する質問項目の選定を行う必要がある。

質問項目の評価には，いくつかの方法がある。最低限，当該の質問項目の作成に直接関与した者以外の専門家によって質問項目の内容の吟味を行うべきである。テストの内容が事前に漏れることを防止するには，複数の質問項目作成者が自分以外の者が作成した質問項目について相互に評価を行ってもよい。

実施可能であれば，基本設計で想定されている受検者層に対応する少人数の集団による**予備テスト**を行ってデータを得ることが効果的である（「1.1 テストの基本設計」参照）。質問項目に対する実際の回答をもとに，受検者がどのように質問を理解し，どのように回答にいたるのか，その認知過程を探ることも質問項目の改良のためには重要な情報となり，指示文の適切さや質問内容が測定する特性について詳細な情報を得ることができる。

質問項目ごとに**通過率**などを求め，質問項目の**難易度**を確認する。テスト全体の回答結果から算出した得点と個々の質問項目の得点との相関係数などから**識別力**を求め，質問項目が受検者の特性の個人差を識別することに寄与しているかどうかを確認する。識別力が低すぎる場合や通過率が極端に高かったり低かったりする場合には，その項目は取り除くか，改良しなければならない。このような質問項目の取捨選択のプロセスは**項目分析**と呼ばれる。また，識別力や難易度などの質問項目の性能にかかわる指標はまとめて**項目特性**と呼ばれている。

コンピュータを利用したテストの場合，予備テストを実施して項目分析を行うプロセスを開発過程に組み込んだテストの基本設計がとくに有効である（「5.1 コンピュータテストの基本的な性質」参照）。項目分析によって質問項目の特性を特定し，利用できる質問項目を**項目プール**に加える。

II 基本条項の解説

質問項目の評価の過程で漏洩が起こらないように細心の注意を払うことが必要である（「1.13 質問項目の内容開示の是非」，「4.1 テスト結果の保管手続」，「4.4 テスト結果の再利用のための保管情報」，「5.4 機密保持や実施管理への対策」，「6.4 著作権の尊重」および「6.5 受検者の義務」参照）。

1.4 回答方法の設計

質問項目への回答方法の設計は、読み誤りや見落としなどによって無用な戸惑いや回答の過誤が生じないように、回答者の立場に立ってわかりやすく表現し、質問項目の順序や配置にも留意する。必要ならば、例題による解説や練習問題を用意する。

1 回答形式の種類

回答方法とは、受検者が質問に対し応答する具体的な手段を指す。回答形式とは、それらの様式の種類を指す。一般に、回答があらかじめ提示されているかどうかによって、回答構築式と選択式の2つに大別できる。

回答構築式には、筆記によって回答する**記述式**と筆記以外の方法による回答形式がある。前者の代表的な方法として、**論文式**、**論述式**、**短答式**、数式などを用いた回答方式および図や絵による回答、あるいは、語句、文章、数値、数式、または固有名詞などを空欄に補充する**穴埋め式**がある。後者の例として面接などの口述による回答、絵画などの作品の創作、音楽や運動などの実技および演技などが挙げられる。

また、**選択式**には**真偽式**（**○×式**）と**多枝選択式**がある。多枝選択式には与えられた選択枝の中から1つを選ぶ**択一式**、複数の選択枝を選ぶことが許される**複数選択式**、その他**組み合わせ式**および**並べ替え式**などがある。

さらに、コンピュータを用いたテストでは、回答方式として複数の選択枝の画像やテキストを表示されている画面上に動的に配置するドラッグ・アンド・ドロップ、**数量選択式**または**スライダックス式**、図面の中の位置を選択する**座標選択式**といった方法を用いることも可能である（「5.1 コンピュータテストの基本的な性質」参照）。

2 回答形式の設計

回答形式の設計においては、選択枝の配置や回答の仕方、記述欄の大きさなど、テストで測定しようとする受検者の特性とは関係のない要因が回答に影響しないように細心の注意を払わなければならない（「1.3 質問項目の設計」参照）。

たとえば、**マークシート**方式による回答は採点の効率性と正確さに優れているが、測定したい特性だけではなく受検者によるマークミスの有無も回答に影響を与える。「正しくマークする能力」を見るのがテストの目的ではないとすれば、マークミスはテストの妥当性を低下させる（「1.2 測定内容の定義と構造化」, 「1.5 採点手続の設計」および「1.11 尺度得点の適切さの確認」）。他にも、選択枝の配置や長さなどがヒントとなっていないか、質問と回答欄との対応は明瞭か、回答の仕方が複雑すぎないかなど、留意すべき点は多くある。

多枝選択式が論述式の質問に比べて劣っているかのような議論がなされることがあるが、回答形式だけを単独に取り上げて質問項目やテストの優劣を論じるのは短絡的である。現実的なさまざまな制約条件のなか、テストの測定目的の特性を適切に把握できる形式を選択する姿勢が重要である（「1.2 測定内容の定義と構造化」参照）。

1.5 採点手続の設計

テストの開発においては、その採点手続を詳細かつ具体的に示すべきである。客観式テストにおいては、採点手続の理論的根拠を明らかにする。主観的評価においては、評価基準および評価手順を明確に設定する。

1 採点手引による採点

採点とは、質問項目への受検者の回答に対して、採点手引にしたがって適切な評価を与えるプロセスと考えることができる。評価結果は、いくつかの段階、順位、あるいは得点などで表される。評価結果は、そのまま判定などに用いられることもあれば、換算して使用されることもあるが、本規準ではそのいずれの場合も尺度得点と呼ぶ（「1.6 尺度化の方法」参照）。採点は、テストに

II 基本条項の解説

よる特性の測定において欠かせない手続である。

採点手引には、具体的な採点手続が記述されている（「1.15 手引，解説書の内容」）。採点は採点手引に則って行われる。採点者によって採点手続に違いが生じないように、採点手続は詳細に明確に定められていなければならない（「2.8 採点手順の順守」参照）。

採点手引は多様な回答の出現を想定して作成すべきであるが、すべての起こり得るパターンをあらかじめ網羅することはできないこともある。とくに、記述式など回答構築式のテストでは難しい場合が多い（「2.9 主観的な評定による採点」参照）。採点手引の記載内容だけでは適切に採点できない場合は、その事態への対応を記録に残し、関係者すべてに周知することが重要である（「4.1 テスト結果の保管手続」参照）。

障害や疾病などに対する特別措置で、通常とは異なる特別な採点手続を必要とする場合については、とくに細心の注意が必要である（「2.6 障害者などへの配慮」参照）。

2 客観式テストにおける採点

正答を客観的に示すことができるテストを**客観式テスト**という。選択式などの回答形式による場合には、一般に正答を定めることができる（「1.4 回答方法の設計」参照）。なお、正答を1つに定める場合、正答の一意性、つまり、他に正答とみなせる選択肢がないことを十分に確認すべきである。正答が定まったならば、客観的に採点を行い、正確な結果を記録することができる。客観式テストにおける採点手引では、何を正答、あるいは、部分的に正しい回答とするか、そして、それぞれにどのような得点を与えるかを定める。また、受験番号を記入していないなど、テストの回答方法におけるミスの取り扱いについてあらかじめ決めておく。誰が、あるいは、どのような機械や採点プログラムを用いて採点しても同じ結果が得られるように、採点手続を詳細に決めておく（「1.15 手引，解説書の内容」参照）。

3 客観式テスト以外の採点

回答形式が回答構築式などのように、正答が客観的に定められない場合には、

回答に対して人為的に数値を与える。通常、その行為を評定、評定に携わる者を評定者と呼ぶ。評定者の主観によって採点が行われる場合においても、評定者によって評定基準の変動という問題はあるが、公正で適正な採点結果が得られるような手続が望ましい（「2.9 主観的な評定による採点」参照）。採点手引には、テストの目的に即した採点が可能なように、明確な評定基準、評定手続が明記されていなければならない。

主観的評定においては、同一の回答に与えられる得点の評定者間のばらつきを調整する必要がある（「1.9 複数のテストの共通尺度化」および「2.9 主観的な評定による採点」参照）。

1.6 尺度化の方法

採点結果から尺度得点を求める場合は、さまざまな尺度化の方法の中から、最も適切な方法を吟味して選択する。その選択根拠は、求めに依って説明できるように記録しておくべきである。

1 質問項目に対する回答と尺度得点

質問項目に対する受検者の回答は、一定の手続を通じて尺度得点と呼ばれる1つの数値にまとめられる。つまり、テストで測定される特性が1つの数値に集約され、大小に配列されて尺度が構成されるのである。本規準では、叙述の一貫性のために、**素点**も特殊な尺度得点として、そのひとつとしている。ここで尺度得点に関する説明は、素点にも適用される。しかし、素点を尺度得点のひとつとして採用するには、素点にも尺度得点としての性能が求められる。

尺度得点は多義的であってはならず、利用者に共通の理解が得られるものであることが望ましい。しかし、その実現は容易ではなく、しばしば誤解を生む（「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照）。

後述のように、尺度の作り方によって、同じ回答に異なる尺度得点を与えられることがある。どのような回答をどう評価するのかという意味で、尺度化の方法は測定内容の定義にも密接にかかわるといえる。尺度化の手続はテストの

II 基本条項の解説

基本設計上きわめて重要であり、テストの根幹をなす部分のひとつである（「1.1 テストの基本設計」および「1.2 測定内容の定義と構造化」参照）。

2 尺度の水準

通常の尺度得点の場合、厳密に言えば尺度の水準はせいぜい順序尺度と考えられる。すなわち、大小の比較ができるという程度の意味である。それを暗黙のうちに間隔尺度として扱っていることが多い。間隔尺度とは、たとえば、同じ10点の差であれば、50点と60点の10点差と70点と80点の10点差は同等であり、同じ尺度上での数値の足し算や引き算が意味をもつ。テスト得点を実際に適用する場面では、尺度得点を間隔尺度として扱うことに大きな問題はない。しかし、テストの尺度得点は、比尺度としては扱わない。長さや重さの測定値は比尺度であり、原点の意味がはっきりしており、かつ、2つの測定値の比を取ることができる。一方、尺度の「0点」という得点に「無」という意味はない。たとえば、あるテストで0点を取った受検者に測定しようとしている特性が全く存在しないことを意味しているわけではないのである。

3 尺度化の方法の具体例

主観的評定による採点を行うテストでは、個々の回答に対して段階や数値を割り当てる作業が評定者の判断に任される（「2.9 主観的な評定による採点」参照）。客観式テストでは一定のルールが決まれば自動的に得点が与えられる。いずれの場合でも、採点とは、受検者が産出した多様な回答に対して、採点手引にしたがって採点手続を適用し、1つの結果として示す情報集約の形といえる（「1.5 採点手続の設計」および「1.15 手引、解説書の内容」参照）。

得点の算出にはさまざまな方法がある。テストの利用者は、テストの基本設計上許される範囲内で、利用目的に応じて適した尺度化の方法を選ぶ。

従来の尺度には、回答を直接尺度得点とする素点方式と、素点を何らかの形で尺度得点に変換する（して求める）**換算点**方式がある。コンピュータの発達により、テスト得点の統計モデルを作り、そのモデルに基づき複雑な計算を経て求められる尺度得点（**モデル準拠得点**と呼ぶ）も可能となっている（「1.1 テストの基本設計」および「5.1 コンピュータテストの基本的な性質」参照）。

(1) 素点方式

①重みつき加算方式

全問正答のとき、満点となるように、質問項目ごとの正答に与える得点を決め、それを合計して尺度得点とする。各質問項目または質問項目のまとまりごとに与えられる得点を**配点**と呼ぶ。この方法は**重みつき加算方式**と呼ばれる。わかりやすいが、配点によって結果が変わることに留意すべきである。配点は何らかの根拠に基づいて決めなければならない。また、新たにテストを設計する際に参考とするために記録を残しておく（「4.4 テスト結果の再利用のための保管情報」参照）。

配点を定めずに質問項目ごとに正答ならば「1」、そうでなければ「0」として合計する素点方式の尺度化もある。**単純加算方式**と呼ばれる。この場合、尺度得点は正答した質問項目数に一致する。

②段階的加算方式

回答の正誤だけで判断するのではなく、それぞれの質問ごとの回答にいくつかの段階を決め、その点数を合計して尺度得点を定める方式である。段階分けはあらかじめ決められていることが望ましい。部分的に正しい回答に得点を与える方式はこの考え方にしたがっているとみなしてよい。質問紙調査のように、各質問項目に対し、「5＝強くそう思う、4＝そう思う、3＝どちらでもない、2＝そうは思わない、1＝決してそうは思わない」といった得点を与える方式もある。**段階的加算方式**は、適切な重みづけで満点を定めれば、重みつき加算方式となる。

(2) 換算点方式

素点方式による得点は、たとえば、一定の受検者集団が共通の質問項目に回答し、ひとりの評定者が全回答の採点を行うという場合のように、同一条件下で求めた得点に対してのみ比較可能な尺度となる。評定者はもちろんのこと、質問項目の構成や受検者集団などが異なると、素点結果をそのまま比較することはできないので換算点方式による尺度化が必要である。

①順位方式

受検者集団内部での順位を基礎とする尺度である。得点を低い方から並べたときに何%の位置にいるかを示す**パーセンタイル順位**などが用いられる。順位

II 基本条項の解説

方式は順位を算出する母体となる集団の中だけで意味をもつ。したがって、属する集団によって値が変わるので、異なる集団の数値と比較することはできない。

なお、単純な順位ではなく正規分布などをあてはめた上で順位を尺度化する方式もある。

②標準得点方式

平均値をある値に固定し、規準集団の標準偏差を1単位として、そこから何単位くらい上下に隔たっているかを示す尺度である。規準集団とは、標準得点を算出するために回答結果が用いられた集団を指す。想定される受検者層を代表していることが望ましい。たとえば、**偏差値**は、規準集団の平均が50、標準偏差が10となるように設定された標準得点方式による尺度化である。標準得点方式で尺度化された得点を標準得点という（「1.7 尺度の標準化」参照）。

なお、正規分布などをあてはめる正規化された標準得点もある。

(3) モデル準拠得点

素点方式にしろ、換算点方式にしろ、尺度得点は各質問項目に付与された得点を加算して出された得点をもとにしている。しかし、同じ得点でも実際はどの質問に正答し、どの質問に正答しなかったかという回答パターンはいろいろである。

統計的モデルとコンピュータの数値的計算を利用して、ひとつひとつの質問項目への回答パターンの違いを考慮した尺度得点を得ることもできる。従来の標準得点方式だけでは難しい共通尺度の作成が可能であることが最大の利点である。

異なる集団で実施された異なる質問項目からなる同一特性の測定を目的としたテスト結果を比較するには、共通の統計モデルに準拠して共通尺度を作成する必要がある。すなわち、尺度得点はこの共通統計モデルのパラメータとして推定するので、複数の尺度得点が比較可能になる。

しかし、モデル準拠得点を用いるためには、事前に予備テストに基づく項目分析を行って、個々の質問項目の性能を把握しておかなければならない（「1.3 質問項目の設計」参照）。テストの設計から実施後のデータ管理段階にいたるまで、緻密な計画と周到な手立てが必要であり、安易に利用することはできな

い。なお、成長や発達など、個人の変化を見るためには共通尺度を構築することは欠かせない手続である（「1.9 複数のテストの共通尺度化」参照）。

利用者は、各尺度の特徴を認識し、結果の利用目的と状況から判断して、最もふさわしい尺度を選択しなければならない。

なお、障害や疾病などに配慮して一部の質問項目を免除するなどの特別な措置をとった場合には、それに応じた尺度化が必要である（「2.6 障害者などへの配慮」参照）。

1.7 尺度の標準化

汎用されるテストは、規準とする集団を明確に定め、その集団における相対的位置づけによって尺度化することが望ましい（この手続を「標準化」という）。標準化においては、用いた標本と標準化手続について記録し、公開する。また、標準化の結果は定期的にその有効性を確認し、改訂の必要性の有無を検討する。

1 標準化の必要性

テストの素点は、質問項目や適用される受検者集団が異なると、尺度得点として比較するのに必要な性質が満たされず、得点の意味づけや解釈ができない（「3.1 テストの趣旨や尺度の性質の理解」参照）。そのため、素点から適切な尺度得点への変換が求められる。

汎用的なテストでは、個人のテスト結果が受検者の母集団の分布における相対的な位置で表現されることが多い。標準得点方式はそのひとつの方法である（「1.6 尺度化の方法」参照）。素点から標準得点への変換を尺度の標準化と呼ぶ。

2 標準化の手続

そのテストを受けた受検者集団がそのまま規準集団に用いられることも多いが、その場合の標準得点の有効範囲はその集団だけに限られる。

知能検査や**性格検査**などには、汎用的に繰り返し利用されることを想定して

II 基本条項の解説

いるものがある。テストの開発時に想定されている受検者層全体を**母集団**としてそこから**無作為標本**を抽出してテストを実施し、それを規準集団として標準化を行う。しかし、現実的に母集団から無作為標本を選ぶことが難しいときには、協力可能な学校や企業などから抽出された集団を規準集団とみなすことも多い。したがって、テストの手引には、利用者が結果の解釈や判断をする際の助けとなるように、標準化に利用した規準集団の選出規準、方法、人数、年齢構成などととも、標準化を行った年や場所などの情報を詳しく、正確に記録しておく。もちろん個人情報保護に関しての慎重な配慮がなされていなければならない（「1.15 手引、解説書の内容」、「3.6 研究目的でのテスト結果の利用」、「4.1 テスト結果の保管手続」および「4.4 テスト結果の再利用のための保管情報」参照）。

3 尺度の再構成とテストの改訂

テストが開発されてから長い時間が経過すると、標準化に用いられた規準集団から得られた結果が時代や社会環境にそぐわなくなることがある。開発時に想定された受検者層を超えて、テストの利用範囲が広がる場合もある。頒布者はテストの実施結果に関連する情報を集め、そのような可能性の有無を定期的に確認する必要がある。何らかの変化の兆候が見出された場合には、規準集団を新たに選び直し、尺度の再標準化を試みる必要がある。また、質問項目の文言の中に不適切な用語や時代に合わない内容がないかどうか、見直す必要もある（「1.12 テストの改訂」および「6.2 テストの公平性」参照）。

1.8 複数の尺度得点の比較

複数のテストによって同一受検者の特性比較を行う場合には、素点のままでの比較は正確な判断を損なうおそれがあるので、素点を比較可能な尺度得点に変換すべきである。

1 尺度得点の利用

素点が比較可能なのは、基本的には受検者群が同一で、用いられた質問項目

も同一の場合である。同じ質問項目が利用できないときには平行テストを利用できればよいが、実際にはその実現は困難である。そのため、素点を比較できるように変換した尺度得点を用いる必要がある（「1.6 尺度化の方法」および「1.7 尺度の標準化」参照）。

2 すべての受検者が複数個のテストのすべての質問項目に回答する場合

複数科目を受けた受検者の得点を比較して、どの科目の成績が良かったのか、あるいは、いくつかの下位検査からなる**心理テスト**の得点を比較してどの領域が優れているかをみる場合などが該当する。この場合には、標準得点やパーセントイル順位を用いるとよい。ただし、比較対象となる尺度得点の分布形が著しく異なる場合は、正規分布の頻度に対応して数値を割り当てる正規化された標準得点を用いてもよい。心理テストのプロフィール得点などはそのようにして尺度化されていることが多い。

3 受検者が同一であっても回答する質問項目が異なる場合

複数の下位テストからどれかのテストを選択して回答したり、同一の科目の中で質問項目を選んで回答するようなテストもある。そのような場合、選択する質問項目が何であっても結果に影響しない設計になっている場合を除いて、比較には特別な工夫が必要になる。

受検者全員が共通に回答する質問項目を一定数以上含んだテストを作成し、それを**アンカー項目**としてその回答結果を手がかりに共通尺度を定義し、他の質問項目に対する得点をそれに合わせて調整する方法がある。このような得点の共通尺度化は、テストの**水平等化**と呼ばれる。ただし、すべてのテストが同一の特性の測定を目的としている場合以外は、同じ方法を用いても**リンク**と呼ばれ、区別される。リンクは厳密な意味では等化とはいえない（「1.9 複数のテストの共通尺度化」参照）。

4 同じ受検者が同一の質問項目を回答しても実施時期が異なる場合

同じ受検者が時間をおいて同一の質問項目に回答し、その間の発達や変化を見るようなテストがある。一見簡単なようだが、受検者集団全体の特性が変化

II 基本条項の解説

しているため、標準得点では受検者個人の絶対的な変化を見ることができない（「1.9 複数のテストの共通尺度化」参照）。

5 共通受検者を用いた共通尺度化

共通尺度を作成するために、通常のテスト受検者とは別に共通尺度化の対象となるすべての質問項目を受ける受検者群を用意し、その回答結果を比較分析することもある。このような手続は、共通受検者を用いた共通尺度化と呼ばれる。質問項目数が膨大であるなど負担が過大なときには、受検者の特性の分布が等質になるように無作為に割り付け、各群にアンカー項目を含む質問項目の一部を受検させて、その回答結果を比較分析することもある。

なお、アンカー項目と共通受検者を**実験計画法**において発展した方法を用いて組み合わせ、データ収集を効率化することも可能である。

6 共通尺度化の手続の記録

共通尺度化を行った場合には、その考え方や手続に関して可能な限り詳細な記録を残す。それらの記録は、新たなテストの開発やテストの改訂に際して有益な情報となる（「4.4 テスト結果の再利用のための保管情報」参照）。

1.9 複数のテストの共通尺度化

同一の特性を測定する2種以上のテストの採点結果を比較する場合や、実施時期が異なるテストの採点結果を比較する場合には、それらの結果が相互に比較できるように尺度得点に変換されている必要がある。

1 想定されている状況

テストは、回答結果として得られた特性の程度が尺度得点として適切に表現されるように設計する（「1.1 テストの基本設計」および「1.2 測定内容の定義と構造化」参照）。適切な尺度化がなされていない場合には、複数のテストの採点結果はそのままでは相互に比較できないことが多い。正しく設計され、実施されたテストにおいては、質問項目の構成や受検者に依存することなく、

同一の特性は同一の尺度得点として表現されなければならない。たとえば、年に複数回行われる資格試験で質問項目も受検者も異なっているような場合には、素点や標準得点を尺度得点として一律に合否の判定を行うことは避けた方がよい（「3.3 判定基準の個別検証の必要性」参照）。

2 受検者もテストも異なるが共通尺度が必要な場合

2つの異なるテスト得点を比較可能な共通尺度で表すには、両者をともに受検している共通受検者の存在が必要である。共通受検者の回答結果から、一方のテストの尺度得点が高方のテストの何点に相当するかを計算し、双方の尺度得点の対応表を用意する。パーセンタイル順位を用いたり、**回帰推定**を行うのが一般的であるが、テストの基本設計によってはモデル準拠得点を利用することも考えられる（「1.6 尺度化の方法」参照）。

共通受検者が得られないときには、質問項目の一部にすべての受検者が受検するアンカー項目を含むテストを作成し、その回答結果を手がかりに他の質問項目の尺度得点を推定する方法が取られる。さらに、多数の受検者が得られる場合には、受検者を等質な群に分けられるよう無作為に割り付ける方法もある（「1.8 複数の尺度得点の比較」参照）。これらの共通尺度化の方法は**等化**と呼ばれている。

等化の理論は、共通尺度化されるテストがともに同一の特性の測定を目指し、ある程度等質な質問項目から構成されたテストであることを前提としている。しかし、現実にはその条件を満たさないなかで共通尺度化を行わなければならない状況もある。テストに含まれる質問項目群が同一の特性を測定しようとしていると考えるのは無理がある場合や、アンカー項目となる質問項目が、調整の対象となる質問項目と同じ特性を測定しているとは言いがたい場合である。そのような異種のテストの回答を利用して共通尺度化を図る方法は、厳密な意味での等化ではない（厳密ではない等化の方法をリンクと呼ぶ）。しかし、共通尺度化が行われる質問項目の尺度とアンカー項目として扱われる質問項目の尺度の間に十分とみなされる程度の相関関係が見られるならば、素点をそのまま尺度得点とするよりは合理的である。

II 基本条項の解説

3 同一受検者の特性の変化量を測定する場合

同じ受検者集団に対して、実施時期の異なる2つのテストの尺度得点から受検者の特性の変化を知ろうとする場合、各テストを標準得点化しても変化量を知ることはできない。こうしたケースで2つの尺度得点を比較するには、両得点が同一の原点と単位をもとにした共通尺度上にある必要がある。質問項目の構成が異なるテスト間では尺度得点の意味が異なるので、結果を直接比較できないからである。

たとえば、学年ごとに別の質問項目で構成される学力テストを用いて学年間の連続的成長変化の様子を知ろうとしても、そのままでは難しい。学年が変わると質問項目の内容が急に難しくなり、学力が伸びているのに得点が下がる、というような例はしばしば見られる。比較を可能にするためには、やはり、アンカー項目を用いて共通尺度化する試みが考えられる。それは**垂直等化**と呼ばれている（「1.8 複数の尺度得点の比較」参照）。水平等化でも同様であるが、垂直等化を行う場合には、とくにアンカー項目の選定が難しい。多様な難易度や識別力をもつ質問項目を十分な数だけ用意する必要がある。

4 発達、変化量測定の難しさ

テストで受検者の成長や発達の変化をとらえるには、共通尺度化が実現されていても、単純に尺度得点の差を比べるだけでは、その差が本当の意味での変化量を示すとはいえない。とくに、テストの信頼性が高くない場合には差得点の誤差が大きく、誤差を含んだ値を実際の変化量と見間違えてしまう危険が大きい（「1.10 尺度得点の確からしさの推定と公開」参照）。

さらに、尺度得点の差を相対的に比較することには十分な注意が必要である。たとえば、学年当初と終了時に同一の学力テストを実施し、その間にどれだけの得点増加が見られたかについて調べてみるとする。年度当初に低い得点を得たの方が高い得点を得た者よりも、終了時のテスト結果は見かけ上高い伸びを示す。この現象は回帰効果または平均への回帰と呼ばれている。2つの尺度得点の相関係数が1でない限り必ず見られる現象であり、「成績の低い多くの生徒が奮起してがんばった」などという解釈は誤解である（「3.1 テストの趣旨や尺度の性質の理解」参照）。

5 開発者、頒布者および利用者の責務

開発者は尺度得点の性質を的確に理解し、利用者が適切に尺度得点を解釈し、利用できるような設計を行わなければならない。頒布者は、利用者および受検者がテストの結果を適切に解釈するために整備された手引を用意すべきである（「1.15 手引、解説書の内容」参照）。さらに、開発者、頒布者および利用者には、テストの設計上の特徴と尺度得点の性質を理解して、誤用を防止する責任がある（「3.1 テストの趣旨や尺度の性質の理解」, 「3.2 拡大解釈の防止」および「6.1 テスト関係者の責務」参照）。

1.10 尺度得点の確からしさの推定と公開

開発者は、構成された尺度得点がどの程度安定しているかを、しかるべき統計指標を算出して検討し（この過程を「信頼性の確認」という）、その結果を公開すべきである。なお、テストが複数の下位テストから構成される場合は、それぞれの下位テストごとに検討し、その結果を公開すべきである。

1 尺度得点の安定性

テストには、必ず測定しようとする特性がある（「1.2 測定内容の定義と構造化」参照）。本来は測定しようとする同一の特性に対して、常に同一の尺度得点が与えられるテストが理想的である。同一の受検者の特性が変化していなければ、尺度得点は何度測っても同一であるべきである。その値を**真値**と呼ぶ。**測定誤差**とは、テストの結果として得られた尺度得点の値と真値との差である。ここでは本規準の全体の統一性のために尺度得点と真値との差という表現を用いているが、測定誤差は最初に得られた測定値に対して検討することが多く、その場合には、素点とその真値との差が測定誤差である（真値は統計的には想定される個人得点全体の平均値であり、特性の本当の値というわけではない）。

測定誤差は小さいほど良い。しかし、回答結果に測定誤差がどの程度含まれているかを示すのは容易なことではない。測定誤差の程度を示す統計指標およびそれを導くモデルや算出方法が必要である。その代表的な指標が信頼性係数

II 基本条項の解説

である。開発者は信頼性係数などの統計指標を算出し、テストの手引や解説書などによって公開すべきである。

2 信頼性係数の定義

改めて測定誤差を定義する。測定誤差とは、テストの回答結果から得られた尺度得点と真値との差を意味する。

信頼性係数とは、ある集団の受検者の尺度得点の分散に占める真値の分散の割合を意味する。たとえば、信頼性係数が0.5であるということは、尺度得点の分散の半分が真値の分散であり、後の半分は測定誤差の分散であることを意味する。信頼性係数は、尺度得点の安定性を表す指標である。

測定誤差が大きいテストは役に立たない。信頼性が低いことはテストにとって致命的な欠点であり、その指標である信頼性係数を報告することは重要である。

信頼性係数という表現は、日常的に「信頼できる」という場合の信頼の意味とは異なっている。あるテストやその尺度得点が「信頼できる」という場合には、「それを使っても良い」ないしは「使うべきである」と理解されるのが普通である。しかし、テストの専門用語としての信頼性にはそこまでの意味はない。テストにとって信頼性が高いことは必要だが、それだけでそのテストを利用すべきだということにはならない。テストを実施して得られる尺度得点が、測定すべき特性を内容的に反映していることが重要である（「1.11 尺度得点の適切さの確認」参照）。

3 信頼性係数の推定方法

個々のテストの信頼性係数を推定する方法について説明する。信頼性係数の推定には、伝統的によく用いられる方法がいくつかある。

(1) 再テスト法

同じテストを2回実施して、その間の相関係数を信頼性係数の推定値とする方法である。この方法は信頼性係数を測定する上で有用な方法である。ただし、実際にこの方法によって信頼性を推定する場合の問題点も多いので注意を要する。たとえば、2回の実施時期の間隔が短いと最初に受検した回答の記憶が2

回目の受検結果に影響する。この場合には、信頼性係数の推定値は本来の値よりも高くなる。逆に、長い時間が経過すると受検者の特性が変化してしまう可能性がある。この場合には、信頼性係数の推定値は本来の値よりも低く推定される。

(2) 平行テスト法

テストを構成する質問項目は異なるが、すべての受検者の真値が同一で、測定誤差の分散も等しいテストを2つ作成したとき、それらのテストは**平行テスト**と呼ばれる。同一の受検者に双方のテストを実施し、得られた相関係数を信頼性係数の推定値とする。理論的には再テスト法で求めた信頼性係数の推定値と同等である。記憶の効果や時間の経過による特性の変化を避けるには、平行テスト法は一見有効に見えるが、実際には2つの平行テストを作成することは難しく現実的ではないことがある。平行テストを新たに作る作業を回避するひとつの方法は、テストの質問項目を無作為に2つに分割して、平行テストとみなすことがある。分割された2つの部分に含まれる質問項目の合計点の相関から信頼性係数を推定する方法である。これは**折半法**と呼ばれる信頼性係数の推定法であるが、どのように分割するかによって信頼性係数の推定値が左右される。また、一定時間内にできた作業量を得点にするようなスピード・テストには、折半法は適用できない。

(3) 内的一貫性による方法

内的一貫性による信頼性係数の推定法でよく使われている指標はクロンバックのアルファ係数である。アルファ係数を算出するための前提条件はそれほど厳しくないので、アルファ係数が多用されるひとつの要因となっている。また、アルファ係数を一般化した方法も提案されている。

類似した内容の質問項目を多数集めると内的一貫性による信頼性係数の推定値は高くなるが、測定内容として定義された特性をくまなくカバーしている保証はない。逆に、内容的に多様な質問項目からテストを構成すると、信頼性係数の推定値は低くなる。また、テストに含める項目数が多くなるにつれ信頼性係数の推定値は大きくなることや、項目間の首尾一貫性を反映した値であるが再テスト法による指標と異なり、時間的な安定性の指標ではないことなどに留意する必要がある（「1.11 尺度得点の適切さの確認」参照）。実際の適用場面

II 基本条項の解説

ではそれぞれの状況に応じて適切なバランスを図る必要がある。

4 信頼性係数の推定にかかわる留意点

(1) 信頼性係数と質問項目の質の判断

質問項目の良し悪しはその内容や表現のみで判断することはできない。測定内容にかかわる分野の専門家から見て、素晴らしい内容の質問項目に思えたとしても、測定されている特性にかかわる受検者の個人差が尺度得点に反映されなければテストとしては不適切である。その適否を判断するには、実施後の回答結果を利用した項目分析がきわめて重要となる（「4.2 テストの実施後の調査と検討」参照）。

(2) 信頼性係数推定のための規準集団

信頼性係数の推定値は、相関係数に基づいて算出されるが、その際、どのような特徴をもつ規準集団において算出されたものであるかを明確にする必要がある。受検者に偏りがあり、テストの適用対象とする母集団の代表とは考えられない場合や規準集団が小さい場合には、安定した結果が得られない。

学力テストなどの場合、たとえば、大学入試センター試験全体の受検者層と比べると、個別大学の受検者層の学力差は一般的に狭い範囲に分布し、真値の分散が小さい等質な集団となっていることが考えられる。この場合、高い信頼性係数は得られにくい。等質な集団では信頼性係数は小さく推定され、多様な集団では信頼性係数は大きく推定されることに留意すべきである。

(3) 下位尺度の信頼性係数

テストがさらに下位尺度に分かれており、それらが独立に使われる場合には、それぞれの下位尺度について信頼性係数を推定すべきである。それぞれの下位尺度が異なる特性を測定していると考えられてテストが設計されているからである（「1.2 測定内容の定義と構造化」参照）。

(4) テストの基本設計と信頼性係数

内的一貫性による信頼性係数で暗黙のうちに前提とされているのは、相互に内容的な文脈に左右されない独立した質問項目を集めた**細目積み上げ方式**のテストである。したがって、ある問の回答が次の回答に影響を与えるような形式のテストでは、内的一貫性の指標を用いて信頼性係数の推定を避ける

なければならない。同一のリード文に複数の関連する質問項目が提示されるような形式も厳密には問題があるので、推定値の解釈には十分注意すべきである（「1.3 質問項目の設計」参照）。

(5) 利用目的と求められる精度

テストの利用目的によって求められる精度が変わることも重要である。研究などのため、集団の特性がテストの利用目的である場合と、テストの結果が個人の処遇に直接結びつく場合を比較すると、後者の方がより厳しい精度を求められる。すなわち、測定誤差をより小さくする努力が求められる。

(6) 評定者の信頼性

主観的な評定による採点が行われる回答形式のテストにおいては、評定の安定性が問題となる（「2.9 主観的な評定による採点」参照）。複数の評定者が同一の回答を評定する場合、与えられた得点の平均、標準偏差などの統計的指標に見られる評定者の個人差とともに、異なる評定者を平行テストとみなして信頼性係数を算出し、記録しておくことが望ましい（「4.2 テストの実施後の調査と検討」）。

5 テストの信頼性に対する認識の重要性

以上のことから、テストに求められる信頼性の程度も達成可能な程度も状況によってまちまちであり、また、実施可能な信頼性係数の推定方法も条件によって異なっていることがわかる。したがって、信頼性係数の数値目標を一律に設定して、テストの質の保証をすることは適当ではない。

しかしながら、開発者および頒布者は、テストの性能を示すひとつの指標としてテストの信頼性にかかわる情報を記録し、積極的に公開すべきである（「3.5 統計情報の公開」，「4.4 テスト結果の再利用のための保管情報」参照）。また、研究者および利用者は、テストの品質にかかわる情報を適切に理解するための識見をもって、公開された信頼性の指標を理解し、解釈しなければならない（「3.1 テストの趣旨や尺度の性質の理解」および「6.1 テスト関係者の責務」参照）。

1.11 尺度得点の適切さの確認

開発者は、構成された尺度が測定内容として定義された特性をどの程度適切に測定しているかを多面的に検討し（この過程を「妥当性の確認」という）、その結果を公開すべきである。

1 尺度得点の適切さ

テストの測定内容は、明確に定義されている必要がある（「1.2 測定内容の定義と構造化」参照）。しかしながら、テストの測定内容ははじめから自明なものとして存在するものではない。人間の特性を直接的に観測することができない**構成概念**として言語で表現し、定義するものである。開発者はテストの目的と明確に定義された測定内容を把握し、それを尺度得点に忠実に反映させるように、質問形式、回答形式および尺度化の方法などのテストの具体的な骨格を設計し、質問項目の内容、指示文および手引などを作成するよう努める（「1.1 テストの基本設計」, 「1.3 質問項目の設計」, 「1.4 回答方法の設計」, 「1.6 尺度化の方法」および「1.15 手引, 解説書の内容」参照）。

テストで得られる尺度得点が安定した結果をもたらすかどうかという観点からの品質は、テストの信頼性と呼ばれる（「1.10 尺度得点の確からしさの推定と公開」参照）。それに対して、テストが測定しようとする特性をきちんと測定する質問項目を備え、それが尺度得点に反映されているかどうかという観点から見た品質は、テストの**妥当性**と呼ばれる。物理的実体がない概念を定量的に析出しようとする以上、質問項目として表現された内容が本来測定すべき構成概念を忠実に反映しているかどうかの検証は欠かせない。安定した測定が可能なテストでも、測定する内容が目的とずれているならば、良質なテストとはいえない。同時に、測定目的となる特性が尺度得点に安定して反映されない場合にも良質なテストとはいえない。妥当性と信頼性はテストの品質にとって車の両輪のような関係である。

いかに周到に設計されたテストでも、テストの基本設計にそった尺度得点が得られる保証はない。実施時に発生した不測の事態や不正行為によって結果が歪むおそれもある（「2.4 手順どおりの実施と不測の事態への対処」, 「6.2

テストの公平性」および「6.5 受検者の義務」参照)。測定しようとする内容が忠実にテストの質問項目に反映されているかどうか事前に確認するのが難しい場合もあるので、実施後にそれが達成されたかどうかを確認するための分析を行い、その情報を適切に保存して結果の利用に反映させるべきである（「4.2 テストの実施後の調査と検討」および「4.4 テスト結果の再利用のための保管情報」参照）。

2 尺度得点の適切さの検討の重要性

尺度得点の適切さの検討は、開発者がテスト尺度を構成する質問項目の文言や内容を検討するだけではなく、多方面からの検討が必要となる。テストの目的、受検者、実施場面および利用目的によって、開発者は、頒布者および利用者など、多方面にわたる関係者の協力を得て、適否を総合的に判断する（「6.1 テスト関係者の責務」参照）。尺度得点の適切さの検討には専門的な識見が必要となるので、その結果が重大な影響をもたらすテストの場合は、開発者にとって専門家の助力が得られる協体制の整備が求められる。

汎用的な利用目的をもった、標準化されたテストのように繰り返し利用される場合、検討時点で適切とされたとしても、テストを取り巻く社会的条件の変化によって、質問項目が有効でなくなることもある。一度検証がなされた後も継続的な検討を繰り返す必要がある（「1.12 テストの改訂」参照）。

実施機会ごとに新たな質問項目が作成される場合、過去のテストと同様の構造であるからといって、質問項目および尺度得点の性質が同様である保証はない。その場合でも、開発者には多面的な検討と検討結果の関係者による共有や議論が可能な協体制を整備し、適切なテストの開発と利用を可能にする努力が求められる。

3 質問項目の内容的な検討

(1) 第三者からの意見の聴取

開発者は、一連のテストの開発過程を文書として記録し、第三者の専門家にその適切さについて意見を求める。開発の当事者は、思い込みによって見方が偏ったり、テストの質問項目としては不適切な点を見逃したりするおそれもある

II 基本条項の解説

る。また、質問項目の内容に注意が集中するあまり、質問項目が受検者にどのように受け取られるかについて、すべての可能性を想定しきれないこともある。そのような点を第三者の視点から検証することは重要である。

(2) テストの関係者による相互検討

開発者はテストの実施後に管理者から回答結果および採点結果などの情報を得て専門的に検証する。利用者からは、テストの基本設計や質問形式、質問内容が実施目的、適用方法や受検者の特性などの視点からみて適切であったかどうか、広く意見を収集する。このように開発者は他の関係者と分析結果を広く共有し、協力してテストの品質向上のための検討を行う（「3.5 統計情報の公開」および「4.2 テストの実施後の調査と検討」参照）。

(3) 利用後の追跡的な検討

テスト結果の利用が終了した後に、テストがどのような利益を生じたか、また反省すべき点は何かについて追跡的な検討を行うとよい（「3.5 統計情報の公開」参照）。実施者はテストの実施そのものの障害にならない限り、その後の分析に資するために必要な情報をテストの実施場面で収集することができる。ただし、利用目的をあらかじめ受検者に明示して、利用の許諾を得ておくことが望ましい。管理者は受検者一人ひとりの尺度得点も含めて実施後のテスト結果に関連した情報を実施者から受領して保管する（「3.6 研究目的でのテスト結果の利用」および「4.1 テスト結果の保管手続」参照）。開発者は、テストの実施結果および関連する情報を利用者と共有しつつ、テストにかかわる研究者の協力を得ながらテスト結果と他の関連データとの関係を分析する。

4 統計的分析による妥当性の検討

基本的に統計的な指標で数値的に検討される信頼性とは異なり、妥当性に関する検討は定量的分析が困難な側面もある。しかしながら、いくつかの観点から統計的に検討する方法が提案されている。

同じ特性を測定している基準となる尺度の存在を前提に、相関係数を計算し、その相関係数をもって妥当性の証拠とする。得られた指標は**妥当性係数**と呼ばれる。等化のための共通受検者と同様に、双方のテストを受検する受検者が必要である。この方法の問題点は、基準となる尺度も通常はテストによって得ら

れるものであり、それ自体が本規準の適用を免れないことである。基準となるテスト自体の検証が必要となる。

複数の下位テストの尺度得点が同時に得られるような構造のテストの場合、それらの相関関係が妥当性の間接的な証拠になる。同じ特性を測る質問項目の得点は相互に高い相関関係を示すべきであり、異なる特性を測る質問項目の得点は同じ特性を測る質問項目と比べて低い相関関係を示すべきである。

テストの結果を用いて何らかの決定を行った場合、テストの情報がどのように決定に役立ったかを検証する必要がある。尺度得点が決定にどのような寄与をしたかを確認することができれば、それが妥当性のひとつの証拠となる。これは、妥当性の証拠を探す最も一般的なやり方であるが、具体的な方法はそれぞれの適用場面において工夫しなければならない。

いわゆる**追跡調査**と呼ばれる方法も妥当性検討のための有効な方法である。追跡調査において、テストを利用して選抜が行われている場合には、**選抜効果**の問題が生じる。選抜効果は構造的で多かれ少なかれ避けられないものであるが、一般にその存在に対する認識は薄い。このような選抜効果の存在を踏まえて追跡調査の結果を解釈する必要がある。テスト結果の統計的分析による妥当性の検証にはこういった難しい側面が伴うとはいえ、テストの品質の検証にかかわるすべての研究者、開発者、頒布者、管理者および利用者は十分にその性質を理解した上で検証の計画を立て、分析結果を適切に解釈するように努力しなければならない（「3.1 テストの趣旨や尺度の性質の理解」参照）。

5 多面的な検討の大切さ

テスト得点の適切性の評価は、ひとつの側面からのみ判断すべきでなく、このようにさまざまな事実やデータを収集、保管し、統計的な解析などの証拠も参考に、関係者が多面的に検討することが重要である。テストの質問項目の適切さに関する判断は、開発者のみの課題ではなく、頒布者および利用者を含む関係者全体の課題である。開発者は、専門的な技術の研鑽に努めるとともに、テストの専門家からの支援を得ながら継続的な議論を通じてテストの内容、実施方法、適用方法などの工夫を繰り返す努力が求められる（「6.1 テスト関係者の責務」参照）。

II 基本条項の解説

また、妥当性の指標を公開する場合には、その指標を得た方法を具体的に提示する必要がある。妥当性を確かめるにはさまざまな困難があるが、たとえそれが十分に行えない場合でも、その事実を手引や解説書に示しておくことが望ましい（「1.15 手引，解説書の内容」参照）。

1.12 テストの改訂

開発者は、尺度得点の解釈に影響を与える種々の社会的条件の変化を考慮し、必要に応じて質問項目の改訂や尺度の再標準化を行う。

1 改訂と再標準化

開発者および頒布者は、テストの項目、手引および解説書などに関して定期的に保守点検を行い、改訂や再標準化の必要性を検討すべきである（「1.15 手引，解説書の内容」参照）。開発者は、諸々の社会的条件の変化によって尺度得点の解釈が従来どおりに行えない可能性が生じたとき、質問項目の修正、入替えおよび提示順序の変更などによりテストを改訂し、また、尺度の再標準化を行う。それによって、目的にかなう適切な測定用具としてのテストの性能や測定精度を維持、向上させることができる（「1.1 テストの基本設計」および「1.2 測定内容の定義と構造化」参照）。

2 改訂や再標準化の必要性

(1) 改訂の必要性

質問項目の内容やその構成要素が時代遅れとなった場合には、開発者は速やかに改訂を行う。たとえば、学校で履修する学習達成度の測定を目的としたテストならば、学習指導要領の改訂があればそれに応じてテストも改訂する必要がある。また、当該分野の研究や利用可能な機器装置などが発展し、測定可能な対象領域や適用可能な質問形式が変わることもある（「5.1 コンピュータテストの基本的な性質」参照）。状況に応じて質問項目の内容や順序の変更、尺度の再構成など、テストの品質向上を図ることが望ましい。

テストを取り巻く環境に明確な変化がない場合であっても、項目分析を行い、

通過率や識別力などの項目特性の値が開発時と大きく異なっている質問項目があれば、その内容を検討するなど、定期的な見直しをする（「4.2 テストの実施後の調査と検討」参照）。

(2) 尺度の再標準化の必要性

質問項目の修正、テストの改訂、尺度の再構成など、尺度得点の意味に以前と変化が生じたときは原則として尺度の再標準化を行う。一般に、質問項目の変更前後では尺度が同等であるとはいえない。ただし、モデル準拠得点を用いられている場合、その項目プール内の項目がいつも最新の項目パラメータ値で維持されているならば、共通の項目プールから選択された質問項目によって作成された新しいテストには改めて再標準化の手続きを行う必要はない（「1.1 テストの基本設計」および「1.6 尺度化の方法」参照）。

質問項目や尺度構成に変更がない場合においても、規準集団の位置づけに開発当初から変化が見られたとき、再標準化が求められる（「1.7 尺度の標準化」参照）。たとえば、知能検査には世代が変わると尺度得点の平均が上昇するという現象が見られている。しかし、知能そのものの向上に関する説得力のある研究がない限り、そのような傾向が見出された場合には速やかに再標準化を行う（「3.2 拡大解釈の防止」参照）。

頒布者は、開発者と協力し、汎用的なテストの実施、利用の円滑な運用のために見直しを定期的に可能にする仕組みを作り、常に尺度の再検討が可能となるシステムを構築しておくことが望まれる。

3 改訂および再標準化後のテストの実施と結果の利用

テストが改訂または再標準化された場合には、原則として改訂後や再標準化後のテストや尺度を使用する。改訂や再標準化に際しては、その特徴や範囲を示し、改訂前後の版、再標準化前後の尺度の比較資料を示すべきである。改訂版や再標準化された尺度が利用できる環境にありながら改訂前の版を実施したり、再標準化以前の尺度得点を利用したりする場合にはそのことを明示し、実施者や利用者は、改訂前の版や再標準化以前の尺度値を利用する根拠を示すべきである（「3.5 統計情報の公開」参照）。

1.13 質問項目の内容開示の是非

質問項目の内容を開示することの是非は、開示によって測定の質を維持することが難しくなるなどの影響を考慮し、また、類似問題の代替可能性、非開示期間を設定する方式なども検討した上で慎重に判断すべきである。

1 質問項目の開示についての検討

受検者に対し、事前にテストの目的、実施方法、結果の適用方法などを説明し、理解を得ることはテストを行う上できわめて大切なことである（「2.3 実施前の受検者への説明」参照）。しかしながら、それは、質問項目の内容まで開示しなければならないという意味ではない。

質問項目の開示の是非は、テストの基本設計、実施方法および利用の仕方にかかわるさまざまな状況を慎重に検討した上で議論されなければならない。それには、どの時点で開示を行うかというタイミング、質問項目をすべて開示するのか、例題を開示するのかという開示の範囲などを含めた検討が必要である（「1.1 テストの基本設計」参照）。

2 質問項目の内容開示が不適切な状況

(1) テストの実施前

入学試験や資格試験、採用試験の質問項目など、正解または望ましい回答が存在し、処遇に大きくかわかるテストの質問項目の内容を事前に開示することは難しい。それは、質問項目を知り得た受検者が正解または望ましい回答をあらかじめ記憶してテストに臨むことを許す状況であり、情報へのアクセス可能性によって著しい不公平が生じる。さらに、テストが測定している特性を歪め、テストの品質を著しく低下させることになる（「1.2 測定内容の定義と構造化」および「1.11 尺度得点の適切さの確認」参照）。その上、受検に備えて、予想問題とその正解の記憶という、テストの目的とは異なる準備行動を促進してしまう副作用からも免れえない。

非開示の質問項目の内容を事前に知ろうとする行為は、悪質な不正行為とみ

なされる（「6.5 受検者の義務」参照）。質問項目などの漏洩は、社会問題化することもある。したがって、質問項目の内容が非開示のテストは、質問項目の漏洩を未然に防止するために質問冊子などの厳重な管理、保管が求められる（「2.2 テスト用具の管理」参照）。

(2) テストの実施後

テストの実施後であっても、質問項目を開示することが認められない場合がある。それは、それらの質問項目が繰り返し利用されることを前提にして設計されている場合である。

病気や障害の診断など、臨床医学的な目的などに使用されるテストでは、受検者に質問項目が知られてしまうことによって、事前に入手した知識に影響を受けて受検者の反応が歪み、適切な判断が不可能となるおそれがある。質問項目の形式によっては、質問項目とともに判定基準、診断基準が知られてしまったならば、受検者が自らの反応を自ら望む形に自在に歪めることが可能となる。判定基準、診断基準が長年の繰り返し実施によって蓄積されたテスト結果に基づいて構築されている場合、質問項目の開示はテストの性能に致命的な打撃を与える。

正解や望ましい回答が存在する形式のテストでも、質問項目が繰り返し使われる場合には、実施後も質問項目の開示は許されない。それは、受検予定者にとっては質問項目が実施前に開示されていることによって、必然的に不公正な状況やテストの目的とは異なる準備行動が引き起こされるからである（「6.2 テストの公平性」参照）。

3 質問項目の内容開示が可能な状況

(1) テストの実施前

テストの目的に照らして、質問項目の内容を事前に開示してもかまわない場合もある。それは、たとえば、一定の学習効果を促進することを主たる目的として実施され、測定結果の評価よりも学習過程そのものが重要となるようなテストが該当する。通常、レポートやプレゼンテーションなどと呼ばれている測定形式では、受検者を一堂に集めて制限時間内に回答させるなど、場所や時間に一定の制約を設けている場合であっても、事前に質問項目の内容が提示され、

II 基本条項の解説

受検者が回答を準備することが可能な場合もある。

なお、テストの実施前に質問項目の内容開示を行う場合には、テスト実施の管理コストは小さくなるが、質問項目の性質によっては、単なる一時的な記憶力など、テストの目的とは異なる特性を測定する結果になることに留意しなければならない。

(2) テストの実施後

テスト実施後に質問項目の内容を開示することが適切な場合がある。現在、実施されている大規模なテストの多くは、テストの実施後に質問項目の内容の開示を行っている。たとえば、高校入試や大学入試の主な質問項目は、実施の翌日に新聞紙上に掲載され、一般に公開されるのが慣例となっている。正解、あるいは、想定される正解例が同時に発表されることも多い。また、資格試験の一部などでは、実施後に Web 上で質問項目が公開されている。実施後に開示する場合、質問冊子などを試験会場から持ち帰ることが認められている場合もある。

実施後に開示された質問項目は、受検者が自らの処遇にかかわるテストについて知るには大切な情報である。また、受検予定者や受検者をサポートする者にとっては、受検準備の内容を検討する上で貴重な資料となる（「2.1 テストの選択」参照）。

一方、実施後に質問項目を開示しているテストでは、原則として毎回のテストで初出の質問項目、すなわち過去に出題されたことのない項目を使用することが必要となる。このため、質問項目作成のコストが大きくなるほか、継続的に実施されている資格試験などではテストの共通尺度化が難しく、結果的に各回の合否水準を一定に保つことが困難になる（「1.9 複数のテストの共通尺度化」および「3.3 判定基準の個別検証の必要性」）。

4 公正なテストを受ける権利と質問項目の開示

質問項目の開示の是非の判断は、質問項目を一度だけ用いるのか、繰り返し用いるのか、という利用の仕方に依存する。さらに、受検が終わった者とその後の受検予定者という、異なる立場の受検者間の相互に矛盾する利害の調整の問題と考えられる。開発者、頒布者および実施者は、受検予定者の受検準備の

便宜を図ることが望ましい場合には、実施予定のない模擬問題や一定の非開示期間の後に使用をとりやめた過去の質問項目を開示して、情報提供を行うべきである。また、実施者は、受検者がテスト結果を適切に解釈することが可能であり、コストなどの面で大きな障害がない場合には、受検者本人からの請求に応じて、本人のテスト結果を事後に開示することが望ましい（「3.4 受検者への結果報告」参照）。

いかなる立場の者であっても、非開示とされている質問項目を勝手に他者に教えたり、一般公開することは許されない（「6.5 受検者の義務」参照）。

5 大規模な項目プールが用意されている場合

繰り返し利用する質問項目であっても、著しく多くの質問項目が項目プールに保存されており、実質的に単なる記憶力では対処できないような場合には、テストの品質に対する悪影響を軽微なものと考えて開示できる可能性もある。しかし、その場合においても、どのようにして適切な質問項目を必要な数だけ確保するか、また、開示によってテストで測定しようとする特性が歪むことはないか、といったテストの性能に本質的にかかわる諸問題について、あらゆる角度から慎重に検討することが必須の前提である（「3.5 統計情報の公開」、「4.2 テストの実施後の調査と検討」、「4.4 テスト結果の再利用のための保管情報」および「6.1 テスト関係者の責務」参照）。

1.14 テストの他言語への翻訳

開発されたテストを他言語に翻訳して用いる場合は、文化的差異に影響されず意味内容が同一になるように努力する。また、翻訳の後、適用しようとする規準集団の標本から得られた資料に基づいて、あらためて標準化を行う必要がある。

1 他言語への翻訳における検討事項

開発されたテストを他言語に翻訳して実施、利用する場合、以下のような諸点に関して検討を行う。

II 基本条項の解説

(1) 訳語の正確性

基本的な用語や慣用的表現についての対訳対照規準を作成し、最適な訳語が使用されるようにする。その際には、翻訳された質問項目をさらに原語に翻訳しなおし、元の同じことばに戻るか、意味が的確に把握できるかどうかを確認することが必要である。

(2) 文化的差異にかかわる配慮

用語の概念、表現方法、あるいは設定された状況に文化的に特有な文脈が想定される質問項目は可能な限り利用しない。利用する場合には、文化的差異を補う翻訳を心がける必要がある。また、質問項目の背景にある文化的差異の影響により、性別、宗教および文化などの違いから一部の受検者が不利益を被らないように留意すべきである。なお、質問項目の一部削除や文化的背景を考えて一部変更する場合には、著作権の問題や変更によるテスト構造の変化に留意する必要がある。外国で開発されたテストの指標や尺度を利用する場合には、正当な手続による開発者の許諾を求める必要がある（「2.7 少数受検者への配慮」、「6.2 テストの公平性」および「6.4 著作権の尊重」参照）。

2 他言語に翻訳されたテストの再標準化

(1) 再標準化の必要性

他言語に翻訳されたテストの実施結果は、原語のテストに定められた手続が順守されている場合でも、元の尺度をそのまま利用することはできない。それは、適切な翻訳により質問項目の意味内容が的確に表現されるように努めていても、翻訳の適否とは別に文化や習慣の違いなどから質問項目の意味内容が変わってくるおそれがあり、テストの利用にあたってその影響についての検証が必要となるからである。また、適用を想定されている受検者層も翻訳前と同等である保証がない。そのため、以下の検討事項について考慮し、翻訳後のテストの規準集団を無作為抽出して、元のテストと同じような分析を行い、再標準化を実施する必要がある（「1.7 尺度の標準化」参照）。

(2) 再標準化の手続

翻訳後のテストの各質問項目について、項目分析を行い、元のテストの項目特性と比較して差がないかどうかを検討する。項目特性に著しい違いが見出さ

れた場合には、翻訳の質について精査する。翻訳に問題点が見つかった場合には、翻訳を修正するか、その質問項目を使用しないこととするべきである。

また、翻訳後のテストの信頼性や妥当性にかかわる指標に、元のテストと比較して大きな差が生じていないことを確認することが望ましい（「1.10 尺度得点の確からしさの推定と公開」参照）。

1.15 手引、解説書の内容

開発者および頒布者は、テストの手引や解説書を作成し、関係者がテストを実施して解釈するために必要なことがらを明らかにしておく。たとえば、テストの目的、測定内容、テストの適用が想定されている集団、標準化の時期や方法などの背景、実施方法と手続、結果の利用法、さらに、実施者などに求められる資格、経験やトレーニングなどである。

1 手引、解説書作成の基本姿勢

テストを実施し、その結果を適切に利用するには、そのテストについての十分な理解が必要である。実施者は、手引、解説書を通してテストの実施方法について習熟することはもちろん、テストの基本設計についても理解しなければならない（「1.1 テストの基本設計」参照）。開発者、頒布者はテストの実施、利用にあたって必要な情報を手引や解説書にわかりやすくまとめ、基本設計にそった利用目的および望ましい利用方法を示すべきである（「2.1 テストの選択」参照）。

適切に作成された手引、解説書をテストにかかわるすべての人の規準として利用することで、統一的で適正な手続でテストが実施され、利用目的にそったテストの水準を保つことができる。また、それまでの研究や開発過程での経験により、テストの実施や解釈などにおいて誤用が想定される場合には、具体的な誤用例やそれを防止するための留意事項を示すべきである（「6.4 著作権の尊重」参照）。

Ⅱ 基本条項の解説

2 手引、解説書に掲載する内容

(1) 目的

テストの基本設計にそった目的を記載する。たとえば、入学試験における選抜、採用選考における適性把握、資格の認定、学校教育における到達度の評価、企業内教育における理解度の把握など、テスト結果がどのような場面でのよう使用されることを想定しているのかを明記する（「1.1 テストの基本設計」参照）。

(2) 測定内容

テストの測定内容を記載する。測定対象となる特性が階層的な構造をもっている場合には、その下位領域の特性とともにテストの尺度構成との関係を説明する（「1.2 測定内容の定義と構造化」参照）。

これらの情報は、テスト結果を適切に利用するために必要である。テストの目的と測定内容との関係について理論的な背景も踏まえてわかりやすく説明する。

(3) 適用が想定されている対象者

テストの基本設計にそって、適用が想定されている対象者を示す。たとえば、年齢、学年、学習履歴、取得資格および技能など、テストの目的に対応して想定される受検者集団の特徴や範囲を明記する（「1.1 テストの基本設計」参照）。

(4) 標準化の時期や方法など

テストの結果が標準得点で示される場合には、規準集団、標準化の時期や方法などについて明記する（「1.6 尺度化の方法」および「1.7 尺度の標準化」参照）。また、再標準化およびテストの改訂のタイミングや方法についても必要に応じて説明することが望ましい（「1.12 テストの改訂」参照）。

(5) 実施方法と手続

テストが適切に利用され、期待される役割を果たすために、定められた手続にそって実施できるよう、実施手順を具体的に記述する。実施環境、実施に必要な道具の準備、実施体制、配布、回収、教示および受検者からの質問や不測の事態への対応方法など、実施者が手続を誤らないようにわかりやすく簡潔に表現する。とくに、多くの人が監督者などの役割で実施に加わる大規模テスト

においては、不測の事態を未然に防ぎ円滑にテストが実施されるために、可能な限り手続を明確化し、手引に記載しておくべきである（「2.2 テスト用具の管理」、「2.4 手順どおりの実施と不測の事態への対処」、「6.3 受検者の疑義への誠実な対応」および「6.5 受検者の義務」参照）。

また、障害がある受検者への対応や、帰国児童・生徒や外国人などの少数受検者などへの対応についても、必要に応じて説明する（「2.6 障害者などへの配慮」および「2.7 少数受検者などへの配慮」参照）。

(6) 結果の利用法

利用者に結果の拡大解釈などの誤用が生じないように、テストによってわかることの範囲と限界について、わかりやすく説明する（「3.2 拡大解釈の防止」参照）。たとえば、利用者や受検者がテストの結果を理解したり解釈したりするためには、尺度得点の確からしさや適切さについて情報を公開することが必要である。尺度得点や下位尺度の得点について、信頼性係数もしくは標準誤差が、利用目的に応じて十分なレベルであるかどうかを判断できるように説明することが望ましい。また、テストが測定すべき内容を測っているかどうかに関する検討を行い、その結果も提示すべきである（「1.10 尺度得点の確からしさの推定と公開」、「1.11 尺度得点の適切さの確認」および「3.5 統計情報の公開」参照）。

テストの結果の表示方法については、正確な表現を心がける。理解しやすいことは重要であるが、わかりやすさを優先するあまり、不正確な表現となりテストの適用範囲を超えて誤解されたりすることがないように留意する。

そのテストを繰り返し使用することが想定される場合、複数の実施機会にわたって同一の質問項目を使用することの影響がどのようなものかについても、手引や解説書に記されていることが望ましい（「1.13 質問項目の内容開示の是非」および「6.2 テストの公平性」参照）。

(7) テスト実施者などに求められる資格、経験やトレーニング

テストによっては、内容が専門的で、実施者や利用者に一定の資格、経験やトレーニングを求めるものもある。そのような場合、求められる資格、経験やトレーニングなどについては、以下の事項を明示する。

①必要な資格、経験やトレーニングなどの内容

II 基本条項の解説

- ②資格、経験などを得る手続
- ③トレーニングや研修が求められる場合の実施方法
- ④トレーニングや研修過程で起こり得る問題への対応方法、など
- (8) その他

上記以外にも実施者および利用者に必要な情報を適宜記載する。また、手引、解説書の内容に変更が必要な場合には、変更前後の内容の違いを明示する。手引、解説書の表現にも留意して、テストの利用にあたって混乱や誤解が生じないように努める（「2.1 テストの選択」参照）。

3 質問項目の作成および回答の採点にかかわる手引、解説書

(1) 質問項目作成の手引

質問項目作成の手引には、以下のような事項を明示する。

- ①テストの目的と場面、測定内容とその階層構造、受検者層、質問項目の提示方法、回答方法、テスト時間などの基本設計
- ②出題範囲、テスト全体の質問項目数、分野別質問項目数、質問項目の形式、難易度の設定などの質問項目作成の規準
- ③質問項目作成者の選定規準（資格、経験など）
- ④質問項目作成者に求められるトレーニングの方法
- ⑤質問項目作成の具体的な手順や方法
- ⑥質問項目の性能の分析方法
- ⑦質問項目の作成にあたっての参考資料、など

継続的に実施されているテストでは、長年かけて経験的に質問項目作成の技術が練り上げられてきている。質問項目の質を維持してテストの性能を保持するためには、このような技術が組織として伝達されることが大切である。そのため、質問項目作成の経験の少ない担当者にも理解できるように、質問項目作成の方法や留意点について具体的に解説した手引の作成が望まれる（「1.3 質問項目の設計」参照）。

(2) 採点手引

客観式テストの場合には、採点手引に以下のような事項を明示する。

- ①正答や部分的に正しい回答、各質問項目の配点および得点の算出方法、判

定や判断を行う場合の基準となる得点などの算出方法

- ②質問項目の選択方法などの回答方法のミスに対する取り扱い
- ③採点上のミスを防止するための確認方法（複数回のチェック、クロスチェック法など）
- ④秘密保持

また、主観的評定の場合には、採点手引に以下のような事項を明示する（「2.9 主観的な評定による採点」参照）。

- ①評価の観点、採点の段階または点数の範囲、各段階や点数のめやす（相対的なレベルや採点割合など）、判定や判断を行う場合の基準となる段階や点数などの採点の基準
- ②採点者数と各採点者の採点を総合する方法
- ③採点者の選定基準（資格、経験など）
- ④予想される回答のポイントや回答例
- ⑤回答方法のミスに対する取り扱い（「2.5 回答方法の説明」参照）
- ⑥秘密保持
- ⑦採点結果記入表の様式
- ⑧採点にあたっての参考資料

採点は、採点手引にそって実施されなければならない（「1.5 採点手続の設計」参照）。

(3) コンピュータテスト運用の手引

コンピュータテスト運用の手引には、以下のようなことを明示する（「5.2 コンピュータや実施環境の仕様」、「5.4 機密保持や実施管理への対策」および「5.5 不測の事態への対応」参照）。

- ①受検者が使用するコンピュータ等の仕様
- ②コンピュータの利用環境
- ③テスト実施環境
- ④質問項目やテスト結果の機密保持
- ⑤システム障害および対応
- ⑥その他関連事項

2章 実施と採点

1章に示されたように、テストは、利用目的や場面にあわせて、測定内容や実施方法などの基本設計が構想され、その基本設計に基づいて開発される。テスト結果を有効に利用するためには、このテストの開発が適切に行われることに加え、その利用目的にあわせて選択されたテストが、適切に実施されることが求められる。また、本規準では受検者の回答を採点して尺度得点に変換し、テスト結果として利用可能な状態になるまでの一連のプロセスも「実施」と呼んでいる。したがって、「採点」も「実施」の重要な一部である。

本章では、このテストの実施と採点にかかわる重要事項について、10の条項にわたって述べている。

2.1は、テストの実施に先立ち、さまざまなテストの中から適用場面にあったテストを選択する際に、テスト関係者が留意すべき基本事項について示したものである。

2.2では、テストの準備段階を中心にテスト用具の厳重な管理の重要性と、管理にあたっての必要事項について述べている。

2.3～2.7は、テストを実施する際の指針についての条項である。2.3はテストの実施者が行うべき受検者への説明について、2.4は、テストを手引に定められた手順にしたがって実施することと不測の事態への対処について、2.5は回答方法の説明について留意すべき事項を示す。また、テストの実施にあたっては、公平性が求められるが、2.6および2.7には、それぞれ障害者への配慮と少数受検者への配慮の必要性について述べている。

2.8、2.9および2.10は、採点にかかわる条項であり、それぞれ開発者の設計にそった採点手順を順守すること、主観的評定を行う場合の評定者のトレーニングの重要性、採点方法が変更される場合の留意事項が示されている。

開発されたテストを利用する場合でも、自らがテストを開発して利用する場合でも、本章に示された条項にそって実施と採点を行うことが望まれる。

2.1 テストの選択

テストの適用場面それぞれにおいては、テストの手引や解説書を参照し、テストの適用にかかわる諸条件を十分に吟味して、使用目的に合った適切なテストを選択する。

1 テストをする目的

実施者、利用者は、既存のテストの中からテストを選択するにあたって、まず、なぜテストを実施または利用するのかという「テストをする目的」をはっきりさせておくことが重要である（「0.2 テストの定義」参照）。選抜のための入学試験か、採用のための**適性テスト**か、臨床場面で使うテストか、学校内の**到達度評価**のためのテストかといった大枠からはじめて、さらに具体的にどのような特性を測定したいのかを明確にする（「1.2 測定内容の定義と構造化」参照）。テストの目的や測定しようとする特性を明確にすることは、どのテストを選択すべきかについて、テストの適用にかかわる諸条件を吟味するための前提となるものである。

2 テストを選択する際の留意点

既存のテストの中からテストを選択する際には、どれが最も目的に合うかを判断する。テストの基本設計において想定されている受検者層の属性が、テストを実施しようとする集団の属性と合致していること、実施方法、手続、実施時間などが可能なものであること、などを確認して選択する。また、テストの信頼性や妥当性に関する情報も、判断材料として意味をもつ（「1.10 尺度得点の確からしみの推定と公開」, 「1.11 尺度得点の適切さの確認」および「3.5 統計情報の公開」参照）。

とくに想定されている受検者層の属性が適用しようとする集団の属性と合致しているかどうかの判断は重要である。標準化されているテストの場合、標準化の際の規準集団と、実施の対象となる受検者集団との間で、年齢、文化的背景、その他の特性などの属性が異なっている場合、意味のあるテスト結果が得られないことがある（「1.1 テストの基本設計」, 「1.10 尺度得点の確からし

II 基本条項の解説

さの推定と公開」および「1.14 テストの他言語への翻訳」。知名度の高いテストだからという理由だけでテストを選ぶと、受検者集団がテストで想定されている受検者層と異なっていることもあるので、注意が必要である。

また、テストを選択する際には、テストの名称や見かけの形式などだけで判断して選択することのないようにする。名称と測定する内容とは必ずしも一致しているとは限らず、とくに、臨床診断に利用するためにテストを選択する際には、その点に留意すべきである（「1.11 尺度得点の適切さの確認」参照）。さらに、学力テストの場合には、利用しようとするテストの出題範囲、**難易度**などもよく見極める必要がある。学校場面では、到達度を確認する目的で市販のテストを利用することがあるが、そのときには、学習した内容とは関連のない出題が含まれていないか、難易度は適切かという点も確認する。

既存のテストで適切なものがない場合、また、測定したい特性があるけれども既存の尺度の中には適当なものがないという場合には、新たに独自のテストを開発、作成する必要がある。そのときには、本規準にしたがって、テストの開発を行うべきである。

また、利用するテストに関して、テスト結果からわかることの範囲と限界をあらかじめ理解しておくことは、テストの選択に寄与するだけでなく、拡大解釈をしないためにも大切なことである（「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照）。

3 手引や解説書の重要性

テストの手引や解説書には、テストの目的、測定内容、想定されている受検者層、標準化の時期や方法、実施方法と手続、結果の利用法などのテストの選択に必要な情報が、わかりやすく明記されている必要がある（「1.15 手引、解説書の内容」参照）。

また、手引や解説書に書かれている情報を正確に理解できる知識を身につけておくことは、テストの実施者と利用者の責務であるといえる（「6.1 テスト関係者の責務」参照）。

2.2 テスト用具の管理

テスト実施責任者は、質問冊子などのテスト用具を、関係者以外が扱ったり触れたりすることがないように厳重に管理する。

1 テスト用具を扱う者に関する制限

質問項目の内容が事前に受検者にわかってしまうと、あらかじめ正答や望ましい回答を用意してからテストに臨むことが可能になる（「1.13 質問項目の内容開示の是非」参照）。その結果、選抜を目的としたテストでは公平な選抜ができず、診断に利用することを目的としたテストにおいては正しい診断が困難になる（「6.2 テストの公平性」および「6.5 受検者の義務」参照）。また、テストの回答などの受検者が記した情報に対して、テスト後に改変が加えられることがあってはならない。したがって、機密保持の観点から、テスト用具へのアクセスを厳格かつ適切に制限する必要がある。

テストの実施責任者は、テスト用具の管理業務全体を把握し、輸送、仕分けおよび保管などの管理業務の各過程においてテスト用具に触れる作業に従事する者を制限するとともに、従事する者に対して守秘義務を順守すること、安全管理に努めることを確認する。また、実施規模の大きなテストの場合は作業量が多く、すべての管理業務を少数の者で行うことが難しいので、管理業務全体の責任を負う統括責任者を置くとともに、各過程にそれぞれ責任者を置き、各過程での責任の所在を明確にしておく（「6.1 テスト関係者の責務」参照）。

また、とくに、入学試験のような選抜を目的としているテストの場合には、親族に受検者がいる者を業務の責任者や作業員から除外する措置がとられることが多い。

2 テスト用具の管理、保管

テスト用具の不足や汚損・破損は、円滑なテスト実施の妨げとなり、また、実施後の答案の回答内容の読み取りを機械で行う場合、機械のメンテナンスが不十分であると不正確な採点を生じるおそれがある。テストの円滑、公平な実施という観点から、テスト用具の管理は重要であり、テスト用具の管理、保管

II 基本条項の解説

の方法については、実施のための手引を準備しておく必要がある（「2.4 手順どおりの実施と不測の事態への対処」参照）。

(1) テスト用具の保管、作業管理

テスト用具は、盗難防止や災害への備えなどに配慮し、保管庫などに施錠して厳重に保管する。鍵の管理は、責任者などの必要最低限の者に制限する。たとえば、実施前の質問冊子などのとくに厳重な管理が要求されるテスト用具に関しては、テスト実施にいたるまで封印をほどこすことなどもひとつの方法である。

テスト用具にかかわる作業を行う際には、紛失や流出を防止するため、関係者以外の立ち入りを禁止するとともに、作業の開始時と終了時にテスト用具の数量を確認し、記録に残す。また、作業日時をその作業や管理者以外の者が知ることのないようにする。

また、テスト用具にかかわる作業を行った日時や保管場所の開閉時刻、開閉者名、責任者名、作業者名などを記録して、実施責任者が常に把握できる状態にし、これらの記録は必要に応じて少なくともテストの終了が確認されるまで一定の期間は保管する（「4.1 テスト結果の保管手続」参照）。

(2) テスト用具の輸送

大規模なテストでは、テストは複数の会場で実施され、保管場所から各会場へ、また各会場から採点を行う場所へと、テスト用具を輸送する業務が生じる。輸送にあたっては、方法、日程、経路、携わる人員などを確認し、引き渡しの際の責任体制や警備体制、危機管理体制について綿密に打ち合わせを行い、盗難防止に努めることはもちろんのこと、途中で紛失したり、汚損・破損したりすることのないよう留意する。

(3) テスト用具の回収、管理

回答用紙などを回収する際には、回収した枚数が受検者数と一致しているかどうかをテスト会場で確認し、その確認が済んでから、受検者の退室を許可する。回収した回答用紙などは改変が加えられることのないよう、厳重に管理する。

また、質問冊子や質問項目についても、持ち帰りが認められている場合は別として、回答用紙などと同様に、質問項目の記載された冊子をすべて回収した

ことを確認してから受検者の退出を許可する。回収されたものについては、未使用のものも合わせて、再度利用する場合も廃棄する場合も、実施前と同様の厳重な管理が求められる（「1.13 質問項目の内容開示の是非」参照）。

なお、テスト結果や用具の回収、管理については、テストの実施にかかわる人が正しく理解し、適正な実施が行われるように、手引に詳細かつ明確に記載されていないなければならない（「1.15 手引、解説書の内容」参照）。

2.3 実施前の受検者への説明

テスト実施者は、実施に先立って受検者に実施目的や結果の適用方法をわかりやすく説明し伝えなければならない。また受検者は、これらについて十分な説明を求めることができる。

1 公開する情報

テスト実施者は、受検者がそのテストを受検するかどうかを判断するための情報として、テストの目的や結果の適用方法などをわかりやすく受検者に明示する必要がある。そのテストがどのような目的で実施されるのか、結果をどのように利用するかについては、たとえば入学試験や資格試験であれば、受検希望者に配布される**実施要領**に書かれているのが一般的である。実施要領には、実施の目的、方法や回答形式、科目ごとの実施時間、出題範囲、合格予定者数または定員、合格者発表期日と発表方法などの具体的な細目に加えて、できれば、テストの回答を採点する方法や尺度化を行う場合にはその尺度化の方法、合格判定基準などが含まれていることが望ましい（「1.6 尺度化の方法」, 「1.7 尺度の標準化」, 「2.5 回答方法の説明」および「3.3 判定基準の個別検証の必要性」参照）。さらに、繰り返し実施されているテストにおいて内容や測定形式に変更があるときには、その変更点がわかるように実施要領に明記することが望ましい（「1.12 テストの改訂」参照）。また、障害がある者に対する特別な措置をどのように行うかという情報も、受検者がそのテストを受検するかどうかを判断するために必要である（「2.6 障害者などへの配慮」参照）。なお、受検者がテストに関して何らかの疑問を感じた場合の問い合わせ

II 基本条項の解説

先をあらかじめ設け、明示しておくことも大切である（「6.3 受検者の疑義への誠実な対応」参照）。

教育や指導を目的として実施されるテストにおいても、事前に実施者から受検者に対してテストの目的や結果の取り扱いについての説明を行う。**心理テスト**の場合には、受検者にテストの目的や結果の開示についての説明を行い、受検者の同意を得る。

また、受検にあたって実施者から提供される情報が不足であると思われる場合には、受検者は実施者に対しさらに説明を求めることができる。

2 開示が不適当な情報

たとえば、選抜や採用のための**面接**では、質問のねらいを受検者に示すと、そのねらいに合うように受検者が意識的に振る舞うために、実施者が本当に得たい情報を得にくくなる。また、心理テストの中には、その本来の目的を説明すると、受検者の回答に影響が及び、それが結果を歪め必要な判断ができなくなるおそれのあるものがある。このように、テストの実施目的や結果の処理方法などを明らかにすると、受検者の回答を歪めてしまうおそれがあるときには、テスト実施者は、それらを受検者に事前に説明することなくテストを実施することができる。質問項目が繰り返し用いられるテストについては、実施後であっても質問項目の内容は開示できない（「1.13 質問項目の内容開示の是非」参照）。受検者から質問などがあつた場合には、テストの実施などに支障があるためそれらについての情報を開示できない旨の説明を行う。また、必要性に応じて、実施後にテストの実施目的や結果の処理方法を説明する機会を設けることなどについて配慮する。

2.4 手順どおりの実施と不測の事態への対処

テスト実施者は手引に定められたとおりの手順にしたがって実施しなければならない。不測の事態が起きた場合、あるいは受検者に不正行為やテストの円滑な実施に対する妨害行為が認められた場合は、その状況について速やかに実施責任者に報告し、実施責任者は適切に対処しなけ

ればならない。

1 基本設計にそった実施

開発時に想定した条件のもとでテストを実施しないと、そのテストの信頼性や妥当性が保たれなくなるおそれがある（「1.1 テストの基本設計」，「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。基本設計にそってテストを実施するには、手引で指定された条件が守られなければならない。また、手引に決められたとおりに実施することによって、手引に合わせて結果を解釈することが可能になる（「1.15 手引，解説書の内容」，「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照）。

手引は、テスト会場の実施責任者やテストの監督者が事前に読んで、その内容を理解できるように作成する。また、監督者など、実施にかかわるすべての者は手引を熟読して理解しておかなければならない。複数の実施者や監督者によりテストを実施する場合には、手引に書かれた実施手順の理解を促進するために、説明会を開催するのも有効である。

2 同一条件での実施

実施責任者は、開発者が想定した条件と同一の条件のもとでテストが実施できるように、手引にしたがって実施環境を整える必要がある。

大規模な試験の実施が予定される場合、手引に記載すべき主な内容を以下に挙げる。

①試験会場および試験室の設定

試験会場の収容人数、机の間隔や静かな環境の確保などの試験室の条件、試験関係者以外の立ち入り制限などの管理の方法、病人・けが人のための救護室の設定。

②試験の監督者の選定

監督者の選定基準（一監督者あたりの受検者数、監督者の資格条件、監督者の間での役割分担）。

③配布，回収および保管

II 基本条項の解説

配布，回収および保管の対象となる質問冊子や回答用紙（機材なども含む）と，それらの方法。

④実施日程およびタイムスケジュールと受検者への指示の内容

監督者が行う作業についての実施日程およびタイムスケジュールと受検者への指示の内容。

回答時間延長などの特別な措置を行う試験室については，通常のタイムスケジュールとは別のタイムスケジュールを作成して，混同を避ける。

手引には，通常，監督者によって受検者への指示が異なることのないように監督者などの発言事項を記載する。監督者は，手引にしたがって発言し指示を出すようにする。発言事項の変更は，監督者自身の解釈や思い込みで指示を出すことにつながるおそれがあるので避けるべきである。

また，手引に記載された手順どおりに実施したことが確認できるよう，テストの開始時刻ならびに終了時刻，準備状況やテスト用具の数量確認，受検者の出欠，回答用紙や質問冊子の回収状況などの当日の実施状況，事後の処理を記録に残しておく（「4.1 テスト結果の保管手続」参照）。実施規模や重要度によっては，手引の段階で実施状況の記録用紙の書式をあらかじめ決めておく必要がある（「2.2 テスト用具の管理」参照）。

3 不測の事態への対処

(1) 事故などへの対処

交通機関の混乱，地震，停電，事故などの不測の事態が発生した際は，受検者の安全を守ることを最優先するとともに，受検者間の公平性が保たれるように対処する（「6.2 テストの公平性」参照）。

これらの事態については，手引に対処方法を明記しておき，実施者が事前に読んでおくことで，実施者間で統一のとれた方法で速やかに対処することが可能になる。以下に，想定される事態と記載内容の例を挙げる。

①交通機関の混乱への対処

タイムスケジュールの変更の判断基準，変更方法，影響を受けた受検者への対処方法など。

②受検者の発病，遅刻などへの対処

試験会場での発病・けが、遅刻者および受検票などを忘れた者への対処方法など。

③テスト用具の事故への対処

質問冊子や回答用紙に外見上の不備（乱丁・落丁、汚れ）が発見された場合、配布物が不足した場合、回答中にテスト用具が汚損した場合およびテストに用いる機器が故障した場合の対処方法など。

④火災、地震、停電、著しい騒音への対処

テストを中断する場合の判断基準、中断したテストの再開方法、回答時間の延長の有無など。

⑤出題ミスの発見、受検者からの疑義への対処

テストの実施中に受検者から質問があった場合の対処方法、受検者の指摘によって出題ミスが発見された場合の対処方法など（「6.3 受検者の疑義への誠実な対応」参照）。

⑥事故などの発生と対処内容とを記録するための書式

不測の事態が発生した際にその状況とそれに対する処置結果がすぐ記録できるよう書式と用紙を常に準備しておく。

なお、受検者に直接かかわる事態が発生した際には、速やかに受検者に連絡し、受検に支障がないように配慮しなければならない（「2.3 実施前の受検者への説明」参照）。また、不測の事態により、受検者の回答結果の取り扱いに事前に通知しておいた内容から変更が生じた場合には、その内容を受検者に公開する（「3.4 受検者への結果報告」参照）。

(2) 不正行為や妨害行為への対処

不正行為や妨害行為の例として、以下のものが挙げられる（「6.5 受検者の義務」参照）。

- ①カンニングペーパー・参考書、他の受検者の答案などを見ること、他の受検者から回答を教わること
- ②テスト時間中に回答を教えるなど、他の受検者を利する行為をすること
- ③テスト時間中に通信手段を用いて他人と連絡をとること
- ④監督者の指示にしたがわないこと
- ⑤辞書、計算機、携帯電話などの使用を禁じられた用具を使用すること

II 基本条項の解説

⑥故意に他の受検者の迷惑になる行為をすること（奇異な行動をとったり、騒音を発生させたりすることなど）

⑦故意に試験の公平性を損なう行為をすること

受検者による不正行為や妨害行為に対しては、座席配置の工夫、実施前の注意喚起などの予防措置を講じておくことが必要である。テスト実施中に不正行為や妨害行為の疑いが生じた場合は、必ず複数の監督者や実施に関係する者で確認を行い、確認がとれた後、手引にしたがい、該当する受検者を不正行為者または妨害行為者として措置を講じる。

不正行為者や妨害行為者に対する措置として、それ以後の受検を続けさせるかどうか、不正が発覚したテスト結果の扱いやそれ以外に受検したテスト結果の扱いなどは、手引に明記しておくべきである。また、不正行為者や妨害行為者に対して講じた措置は、必ずその状況とともに記録しておく（「1.15 手引、解説書の内容」および「4.1 テスト結果の保管手続」参照）。

手引に想定されていない事態が発生した場合や緊急時には、実施責任者の迅速な判断が求められる。実施責任者の適切な対処を可能にするため、実施における情報連絡システムの整備が必要である（「6.1 テスト関係者の責務」参照）。

2.5 回答方法の説明

テスト実施者は、テストの実施に際して、基本設計にそった方法で回答できるよう、回答方法について受検者の理解力に合わせていねいに説明しなければならない。

1 基本設計にそった回答方法

受検者が回答方法を十分に理解しないまま実施されたテストでは、不正解であることが回答方法の理解不足や誤解によるものか、能力不足や知識の欠如によるものかわからない（「3.1 テストの趣旨や尺度の性質の理解」参照）。また、テスト結果に基本設計で意図した特性とは異なるものが反映されてしまうおそれがある（「1.1 テストの基本設計」参照）。

とくに、受検者にとって不慣れな回答方法を含むテストでは、回答方法に習熟させるための配慮が必要となる場合がある（「1.4 回答方法の設計」）。たとえば、**マークシート**を利用したものやコンピュータ上での回答の場合、テストの実施前の周知、マーク例や練習課題を用意する。また、職業適性検査や資格試験などの場合のように多様な受検者を想定する場合には、説明をとくにしていねいに行う。記入誤りなどの可能性にも配慮して、基本設計にそった回答方法が受検者に理解されるように、あらかじめ受検者層に応じた説明と導入方法を工夫しておく（「2.3 実施前の受検者への説明」参照）。

2 質問項目への回答の仕方に関する説明

基本的な点として、回答を質問冊子に記入するのかあるいは回答用紙に記入するのか、監督者の指示に応じて1問ずつ、あるいは、数問ずつ回答するのか、受検者自身のペースで回答するのか、ということなどを明確に説明する。

通常、テストにおける質問項目に対する回答方法は、回答構築式と選択式に大別される（「1.4 回答方法の設計」参照）。

回答構築式は、受検者自身が作成した語句、文章、数値、数式、図表を回答させるものである。受検者の能力を正確に評価するためには、受検者が回答方法で迷うことなく応答できるように説明する必要がある。具体的には、回答を記入する場所、回答の字数制限の有無や回答用紙が足りなくなった場合の措置、回答の導出過程の記入の要不要などについて、質問冊子の冒頭に記載したり、実施者または監督者が指示したりすることが適当である。また、回答の下書きや導出段階の記述も採点対象とする場合には、そのことも受検者に知らせておく必要がある（「1.5 採点手続の設計」および「2.3 実施前の受検者への説明」参照）。

選択式は、複数の選択枝から回答を選んだり、記述内容の真偽を判定したり、選択枝の正しい組み合わせを作成したりするものである。受検者が回答方法に関して迷う余地がないように、用意された選択枝以外の回答の可否を含めて選択する選択枝の範囲、複数選択の可否や選択すべき選択枝の数、複数の質問項目で共通の選択枝を用いる場合に同じ選択枝を複数の質問項目で選択することの可否、選択した選択枝の記入方法などについて、受検者が理解できるように

II 基本条項の解説

する必要がある。回答方法の理解を助けるような例題を示すのも有効である。

とくに、複数の選択枝から回答を選ぶ多枝選択式では、マークシートによる回答用紙が用いられることもある。その場合は、マークに用いる筆記用具の種類やマークの仕方、訂正の仕方について、受検者が戸惑わないよう、マーク例を示すなど十分な説明を行う。

また、質問項目への回答の仕方に関しては、出題方式に加え、問題のレイアウトや表記にも配慮し、受検者の誤解の余地を最小限に抑えるようなテストの開発段階での工夫も重要である。たとえば、設問の位置と回答する選択枝の位置が別ページになるようなことは避けるべきである（「1.3 質問項目の設計」参照）。

2.6 障害者などへの配慮

障害がある者などが受検する場合は、その障害などに応じた適切な器材や環境を提供し、標準的な手続を変更して実施するなどの措置をとる。具体的には、テスト時間の延長、障害などを補完する設備・器材の利用、介助者による支援などである。また事後の検討と将来の参考のために、実施の詳細を記録しておく。

1 障害などを補完する措置に関する情報提供

障害がある者などがテストを受検するかどうかを判断するためには、そのテストにどのような具体的な配慮があるのかという情報が必要である。実施者および利用者は、想定される障害などの種類や程度に応じてとりうる具体的な措置について、専門家の助言のもとに検討し、その内容を基準や要項などとして定めておくことが望ましい。便宜供与が可能な措置については、受検案内や受検要項などに掲載し、そのような情報を必要とする者に確実に伝わるように情報提供を行う（「1.1 テストの基本設計」および「2.3 実施前の受検者への説明」参照）。

また、受検にあたって特別な措置を希望する者は、テストの出願時に障害などの種類や程度に関する情報を実施者に申請するように促す。実施者は、その

ような申請があった場合、必要に応じて障害者手帳の提示、診断書の提出などを求めた上で、障害や疾病などの種類や程度にあった適切な措置を決定する。決定された具体的な措置については、受検者に通知するとともに、テスト会場の実施責任者などにもその実施に必要な具体的な指示を出して準備を進める（「2.2 テスト用具の管理」参照）。

2 具体的な措置の例

(1) 回答時間の延長

多くのテストで実施されているのが、回答時間の延長である。回答時間を延長する場合は、テストの公平性に留意することが必要である（「6.2 テストの公平性」参照）。一般には、延長が必要な理由、同様のテストで時間延長措置が認められたかどうかなどの情報をもとに開発者、利用者が検討し、適切な延長時間を決定している。また、テストの開始時刻や終了時刻が異なることによって混乱が生じないように、他の受検者とは別室で実施されている。

(2) テストの一部免除

たとえば、外国語試験に含まれる聴解問題など、テストの一部にある特定の障害がある受検者にはどうしても回答不可能な部分がある場合は、受検者の申請により、診断書などに基づいて審査して、その部分の受検を免除する場合がある。そのような特別な場合の尺度得点の算出方法やテスト結果の扱いについては、不公平が起こらず、またそれによるミスが生じないように万全の手立てを講じておく必要がある（「1.5 採点手続の設計」, 「1.6 尺度化の方法」, 「2.8 採点手順の順守」および「3.1 テストの趣旨や尺度の性質の理解」参照）。

(3) 障害や疾病などに応じた具体的な措置

障害や疾病などに応じて考えられる具体的な措置としては、以下の例が挙げられる。

① 視覚障害者への具体的な措置

- 拡大鏡や点字器などの障害を補完する器材の使用を認める。
- 座席を窓側の席など明るい席とする。
- 試験室に備え付けのもの以外に照明器具の使用を認める。

II 基本条項の解説

- 文字を拡大した質問冊子や回答用紙を使用する。
- 回答記入欄を通常よりも大きくした回答用紙を使用する。
- 点字による質問冊子や点字による回答用紙を使用する。

②聴覚障害者への具体的な措置

- 座席を試験室の前列の席など監督者の指示を聞き取りやすい席とする。
- 補聴器の使用を認める。
- 試験官の発言内容などを文書で知らせる。
- テストの開始、終了を肩をたたくなどの口述以外の方法で知らせる。
- 口述による部分を筆談により実施する。
- 手話通訳者を手配し、口述による部分を手話により実施する。

③身体に障害がある者などへの具体的な措置

- アクセスしやすさや障害者用トイレなどに配慮して試験室および座席を決定し、場合によっては他の受検者と別室とする。
- 杖や車椅子の使用を認める。
- 特別な机やタイプライターの使用を認める。
- 試験室入口までの付添者（介助者）の同伴を認める。

④テストの当日にけがや病気を発生した者への具体的な措置

- 別室での休息を認める。
休息時間をテスト時間に含めるかどうかをあらかじめ検討しておく必要がある。
- 別室による受験を認める。

なお、これらの措置については、なるべく具体的かつ詳細に手引に記載しておくことが望ましい（「1.15 手引、解説書の内容」参照）。

3 留意事項

障害や疾病などへの措置が、適切であったかどうかを検討し、公平な措置を確保するためには、どのような申請に対してどのような措置を行ったかに関する情報を記録しておく（「4.1 テスト結果の保管手続」参照）。

なお、テストの出願時に判明している障害などへの対応の他に、出願からテスト開始までに負傷や疾病が生じたという申請があった場合の対応についても

あらかじめ検討しておくことが望ましい（「2.4 手順どおりの実施と不測の事態への対処」参照）。

2.7 少数受検者への配慮

受検者集団の中に主要受検者集団と異なる一部の少数受検者（たとえば、帰国児童・生徒、外国人）が含まれており、かつ一部の少数受検者について正当な理由なく不利な扱いが予想される場合は、それぞれの集団に適したテストや判定方法を用意すべきである。

1 配慮の必要性

受検者が共通の属性を備えた集団であることを前提として開発されたテストは、そのような集団とは属性の異なる受検者にとって不利となることがある。そのため、帰国児童・生徒や外国人受検者などへの配慮がなされる場合がある。

たとえば、ある受検者にとってそのテストの使用言語が外国語として習得したものであって、その言語を学習した期間も短いという場合、そのテストが測定しようとする知識や能力があっても、内容を読み取って回答する過程でそのテストの使用言語が母語である受検者と比べて不利となる。さらに、テストの使用言語をその文化の中で日常語としてきた人でなければ理解できない言いまわしが、そのテストの回答の鍵となっているとその影響は大きい。

2 措置の具体案

テストの目的が、出題内容自体についての知識や理解力を測定することであって、テストに使用されている言語を活用する能力の測定を含まないのであれば、その言語が母語ではないことが不利な条件とならないようにすることが望ましい。そのような措置のひとつとして代替措置を検討する場合、その代替措置の対象者が逆に有利になるようなことのないように、十分に吟味することも大切である。

テストにおいて、使用する言語に影響されない出題を目指そうとする工夫は、言語ではなく図や絵や道具を用いる**知能検査**の開発などにも古くから見られる。

II 基本条項の解説

しかし、必ずしもこのような代替措置の開発がすべてのテストにおいて可能とは限らない。受検者の母語で出題するといっても、母語としての言語の範囲の決定や翻訳作業に相当の困難を要する場合がある。受検者の属性を事前に判定し、辞書の使用を認めたり、回答時間を延長したりするといった対応が現実的である場合もある。

また、内容面においても、テストの目的に直接関係しない部分において、育った文化的環境が異なることによって一部の受検者が答えにくいような内容を出題することは望ましくない（「1.14 テストの他言語への翻訳」参照）。

大学入試では、帰国者のための入試、外国人留学生のための入試などを別に設けて、一般入試とは別の試験内容、判定方法で行っていることが多い。なお、その場合、一般入試の結果と単純に比較することができないことに留意する。

辞書の使用や時間延長などの具体的な措置を認めることが適切であるかどうかも含め、開発者は、可能な限りさまざまな受検者の属性を念頭に置きながら開発を進めることが望ましい（「1.1 テストの基本設計」参照）。

2.8 採点手順の順守

テストの採点者は、基本設計にそって採点する。採点の責任者は定められた手順どおりの採点が行われるよう点検管理し、誤りが発見された場合は速やかに実施責任者に報告し、実施責任者は適切な措置を講じる。

1 採点手引による採点

標準化されたテストには、開発者によって採点手引が用意されている。採点者は、採点手引に記載された指示にしたがい、採点や採点後の得点の集計に誤りが生じないように、手引に定められた手順や要領を順守して採点を行わなければならない。さらに、カウンセリングなどに利用される**投影法**による性格検査のような、専門的な知識、技能を必要とするテストの採点は、あらかじめ十分なトレーニングを受けた者が行わなければならない（「1.5 採点手続の設計」および「1.15 手引、解説書の内容」参照）。

また、主観的な評定による採点を行う場合の採点手引には、回答を評価する

ための観点、採点の方法や手順などが定められ、必要に応じて評価のポイントや回答例などが記載されていることが望ましい（「1.5 採点手続の設計」および「2.9 主観的な評定による採点」参照）。採点者は、評定の基準が異なったり、加点や減点の幅が異なったりすることのないように、採点手引にしたがって採点しなければならない。さらに、複数の採点者が採点作業を分担する場合には、事前に採点についてのトレーニングや打ち合わせを行うことが望ましい。ただし、初出の質問項目、すなわち過去に出題されたことのない項目を出題するようなテストなどで事前に採点手引の作成が困難な場合には、テスト実施前の段階では採点の際のポイントをおおまかに挙げる程度にとどめ、テスト実施後に採点作業にとりかかりながら、採点基準などの調整を行う手続の方が実際的に有効な場合もある。

なお、複数の採点者が採点作業を分担する場合、採点者を統括する立場にある採点の責任者は、採点手引にそって採点が行われるように、各採点者への資料の配布や回収、情報提供、採点結果の点検管理などを行う（「2.9 主観的な評定による採点」参照）。

2 採点作業におけるミスの防止

採点作業においてミスが生じないようにするためには、採点手引に定められた手続どおりに作業を遂行することが重要である。

選択式の回答の採点を行う場合の採点手順は開発者によって明確に定められているが（「1.5 採点手続の設計」参照）、採点作業に誤りが生じる可能性もあるので、採点責任者の監督のもとで作業ミスの有無を確認する手続が必要である。この手続には正答が誤答扱いになっていないか、誤答が正答扱いになっていないか、集計ミスはないかなどを確認する手順を入れるべきである。

マークシート方式の回答用紙などを機械が読み取ってコンピュータによる採点が行われる場合には、採点手引に記載された採点手続を遂行するためのプログラムを実行して採点が行われる。採点プログラムについては、不具合がないかどうかを事前に点検して万全を期すようにしなければならない。とくに、回答用紙の文字や数字を正確に読み取っているか、**組み合わせ採点**や部分的に正しい回答に与える得点が適切か、など念入りに点検することが必要である。

II 基本条項の解説

なお、回答構築式テストの回答などをコンピュータプログラムによって自動的に採点する場合もある。回答の読み取りが正しく行われるならば、主観的な評定のように事前に設定された採点基準が変動するおそれはない（「2.9 主観的な評定による採点」参照）。しかし、本来測定しようとした受検者の特性が採点結果に反映されているかどうかについては、プログラムの開発段階から専門家の判断をまじえて具体的かつ慎重に検討しなければならない（「1.11 尺度得点の適切さの確認」および「3.2 拡大解釈の防止」参照）。

3 採点の誤りへの対応

採点手順どおりの採点を実施されなかったことが判明した場合は、速やかに実施責任者に報告する必要がある。実施責任者は、採点の誤りに対して速やかに対応する。

採点結果に基づいて何らかの判断がなされた後に、採点の誤りが発見された場合、具体的にどのような措置が必要であるかは、テストの公平性、特定の受検者に不利な扱いとならないことなどに配慮して、実施者や利用者などが慎重に協議して判断する（「2.4 手順どおりの実施と不測の事態への対処」および「2.10 採点方法の変更」参照）。

2.9 主観的な評定による採点

主観的な評定によって採点する場合は、基本設計にそって採点できるように、評定者のトレーニングをする。また、採点後には複数評定者の評定の整合性などを分析し、必要に応じてさらに調整を加える。

1 主観的な評定に固有の問題点

多枝選択式テストとか、名称や数値を回答するテストなどでは評定者の主観をまじえずに機械的に採点することが可能であるが、通常の**論述式**テストや面接、実技テストなどの場合には、評定者の主観が採点に影響するので、採点基準の変動という問題に注意を払わなければならない。採点基準の変動は、評定者間の変動と同一評定者内の変動の2つに整理することができる。**主観的評定**

における採点基準の変動は、テストの測定目的からみて誤差といえるもので、この変動を最小限に抑える努力をすべきである（「1.1 テストの基本設計」および「1.10 尺度得点の確からしさの推定と公開」参照）。

(1) 評定者間の採点基準の変動

複数の評定者が評定作業を分担する場合、どのような答案を高く評価するかという基準や、最高点や最低点を何点にして**尺度得点**の分布をどのようにするかなど、採点基準の認識に違いがあれば、その差が採点結果に現れることは明らかである。また、評定者の中に一貫して甘すぎる評定をする評定者と辛すぎる評定をする評定者がいれば、同一の回答を採点した結果に大きな違いが生じる。

(2) 同一評定者内の採点基準の変動

採点基準の変動は、同一評定者内でも生じる。たとえば、200人の論述式テストの回答を採点したとき、最初の20人では辛い採点基準で厳しく減点していたものが、後半に近づくにつれて評定が甘くなる場合である。もちろん、その逆になる場合もある。

また、採点順序も採点基準に変動を与える大きな要因である。評価の高い回答の直後の回答に対しては厳しい評価を与えがちである。

2 採点基準の変動を抑える方法

(1) 採点基準の共有

主観的な評定によって採点を行うテストでは、テストの基本設計にそった採点基準を定め、すべての評定者がそれを理解し共有することが重要である。採点基準の共有化を図るためには、チェックリストや評定票を用意したり、必要に応じてあらかじめ回答例を準備したりといった工夫が考えられる。しかし、論述式テスト、面接、実技テストなどでは、多種多様な回答が得られることが多く、事前に回答を予測し定型化することには限界がある。したがって、実際には採点前のある程度の採点基準を定め、提出された回答を概観した後に、必要に応じて採点基準を追加・修正することになる。さらには、ある程度採点が進んだ段階で、再度採点基準の確認を行い、評定者間および同一評定者内の採点基準のばらつきを修正する作業も有効である（「2.8 採点手順の順守」参照）。

II 基本条項の解説

(2) 評定者のトレーニング

採点基準にそって安定した評定を行うことは、テストにかかわる専門的な技術のひとつである。採点基準が妥当なものであり、適切な指導法でトレーニングを受けるならば、ある程度安定した採点基準を共有することができる。また、この技術は、実際の採点作業を経験することにより熟達するものである。

論述式テストにおける評定者のトレーニング法としては、比較的少数の回答に基づき、採点基準の共有化を図る方法が現実的である。この方法では、まず、トレーニングを実施する者が当該テストの答案の中から10～20程度の答案を抽出する。トレーニングを受ける評定者は、抽出された各答案について評定を行い、その際、それぞれどのような理由でその評定を行ったのかを具体的に示す。これらの評定は、その理由とともに各評定者に示され、評定者間における採点基準の変動や評定者内での変動について評定者自身に認識させる。次に、同じ答案に対する、熟練し安定した評定ができる者の評定例を示し、その評定結果の変動やその理由を認識させることにより、採点基準の理解と共有化を図る。このような手続を経ることにより、評定の信頼性を高めることができる。また、熟練し安定した評定者が存在しない場合には、評定者間で議論し、採点基準をできる限り明確化させた上で再度採点を行うことが望ましい。なお、現実には十分なトレーニングの時間がとれない場合には、採点者間に採点基準の変動が極力生じないようにするため、評定者間の打ち合わせを十分に行う必要がある。

また、論述式テストなどの答案であれば、採点基準を見ながら繰り返し評定することや追加または修正された採点基準にそって再度評定することが可能であるが、面接や実技テストでは、通常、回答を繰り返し見て評定することはできない。したがって、採点基準を明確化し共有するために、事前にビデオなどによるモデル事例をもとにしたトレーニングを行うことが望ましい。さらに、回答が回答用紙に記載されるなどして残る場合と比べて、よりいっそう入念な事前の打ち合わせが求められる。

採点者のトレーニング方法については採点手引に具体的に記載されることが望ましい（「1.5 採点手続の設計」および「1.15 手引、解説書の内容」参照）。

3 採点後に行うこと

(1) 複数の評定結果のまとめ

主観的な評定を含むテストでは、採点手引をもとに採点が行われたとしても、複数の評定者による評定結果が完全に一致するとは限らないため、同一の答案を複数の評定者が採点することが望ましい。それが不可能な場合、評価観点を一貫させるには、同じ評定者が同一の処遇にかかわるすべての回答を評定することが望ましい。たとえば、大学入試に2つの学部の受検者の回答を2名の評定者が処理するような場合、それぞれの評定者が担当する学部を定め、一貫して担当の学部のすべての回答を評定すべきである。各学部について半分ずつの回答を分担すると、評定者による観点の相違が直接的に尺度得点に反映されることになる。

複数の評定結果が得られた場合には、それをひとつにまとめる必要が生じる。その方法として、すべての評定者の評定結果を平均する方法以外に、極端な評定結果の影響を排除するために評定結果の**中央値**を利用する方法、あるいは最低点と最高点を除いて評定を平均する方法などがとられる。なお、あらかじめ許容範囲となる評定結果のズレの範囲を決めておくべきである。評定者間で許容できないほどの差異が生じた場合は、さらに第三者が評定した結果により得点を確定させたり、合議の結果で得点を確定させたりする方法もある。

すべての評定者がすべての回答を採点するのではなく、評定者の一部が回答の一部のみを評定するような場合には、評定者間の変動の影響を小さくするために、個人ごとに平均と分散をそろえるなどの統計的調整を行うこともある（「1.9 複数のテストの共通尺度化」参照）。

(2) 評定の整合性の分析

複数の評定者間の評定の整合性については、**算術平均**などの採点基準の高低（甘辛）の指標、**標準偏差**などの散らばりの指標とともに、評定の類似度の指標として**相関係数**、**順位相関係数**などの情報をもとに分析することができる。どのような答案を高く評価するかといった採点基準の認識に大きな違いがあれば、評定の類似度は低くなる。評定者間の評定の整合性については統計的な指標により分析することができる。

II 基本条項の解説

4 主観的な評定の利用

主観的な評定を行うテストでは、評定から主観的な要素を完全に排除することは困難であるので、実際には、主観的な要素によるバイアスが含まれるおそれがあることを認識した上で評定を行うことになる。

どのような方法であれ、採点において重要なことは、測定されるべき特性に関する受検者の真の状態を可能な限り正確に尺度得点として表現することである。「がんばったから点を与える」、「ボーナス点」というような報償的な扱いなど、測定対象とする特性以外の要因に基づく加減は行わない。数値の与え方の決定に際しては、テストが測定しようとしている**特性**が尺度得点に正確に現れるのが理想であることを常に念頭に置くべきである（「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。

正確な測定という観点からすると、主観的な評定は、多枝選択式のようなテストでは十分に測定できない特性を測定目的としたテストで用いることが望ましい（「1.3 質問項目の設計」および「1.4 回答方法の設計」参照）。

2.10 採点方法の変更

不適当な質問項目の発見など、何らかの理由で、事前に説明された採点方法を変更する必要がある場合には、その理由やそれへの対応などについて開示すべきである。

1 不適当な質問項目への対応

ある程度の規模で**予備テスト**が可能な場合には、予備テストのデータに基づいて不適当な質問項目を発見し、そのような質問項目を除去したり、あるいは改善して再度の予備テストを実施するといった措置を講じることもできる（「1.3 質問項目の設計」参照）。予備テストによる検討が可能な場合には、誤りを含む質問項目や不適当な質問項目が実際に出題される可能性は低い。それでも、実施後の分析によって、何らかの意味で不適当な質問項目が発見された場合は、そのような質問項目を除いて、標準化の手続をやり直す場合がある（「1.7 尺度の標準化」参照）。

予備テストが行えず、初出の質問項目を出題するテストでは、実施前のチェックで検出できなかった出題ミスが実施中や実施後に判明する可能性を無視できない。たとえば、テストの開発者が全く想定していなかった回答が出現することによって、その質問項目の不具合が明らかになるような場合も考えられることから、テストの実施後に質問項目の不具合が生じた場合の対応を検討しておくことが望ましい。

また、採点手順を順守した場合でも、採点の途中で採点方法を変更する必要性が生じる可能性もある（「2.8 採点手順の順守」参照）。採点方法について、テストの実施前に十分な検討が必要なことはいうまでもないが、事前に定めた採点方法が不適當であることが判明したときには、状況に応じた採点方法の変更を行わなければならない（「2.4 手順どおりの実施と不測の事態への対処」参照）。

2 採点方法の変更と受検者への影響

受検者にあらかじめ採点方法が説明されており、採点方法を変更することが受検者にとって重大な意味をもつと考えられる場合には、受検者に対して、採点方法を変更する理由と変更による受検者への影響などを知らせるべきである。

具体的な変更措置については、受検者にどの程度の影響が及ぶのか、どの程度の混乱が予想されるかなどを慎重に考慮した上で、対策を検討する必要がある。なお、現状では出題ミスや不適當な質問項目が発見された場合、その質問項目を採点対象から外す、あるいは受検者全員にその質問項目分の得点を与えるなどの変更措置がとられることが多いが、そのため新たな不公平が生じないかどうか、個々の事例について慎重に検討することが必要である（「3.4 受検者への結果報告」および「6.2 テストの公平性」参照）。

3章 結果の利用

テストが実施された後、その結果は何らかの目的のために利用される。本章では、テストの実施結果の利用に関する条項が6つ提示されている。これらの条項は、テストの実施結果として用いる情報の種類という観点から、2つに大別できる。

ひとつは、個々の受検者のテスト結果の利用にかかわるものである。3.1～3.4は、テスト利用者および管理者が、テスト結果をもとに適切に個人の処遇に関する決定を行うために留意すべき基本事項である。3.1はテスト結果を利用する前提として理解すべき事項が述べられている。3.2はテスト結果の解釈に関する注意事項、3.3はテスト結果を何らかの判定に用いる際の判定基準についての条項である。3.4は受検者へのテスト結果の報告についての条項である。

もうひとつの観点は、個々の受検者のテスト結果ではなく、テストの統計数値結果の利用にかかわるものである。3.5、3.6がそれにあたる。3.5は統計情報の公開、3.6はテスト結果の利用に関する条項である。テストにかかわる研究者および開発者が、テストの統計的な分析結果をテストの品質保持および向上に適切に反映させるために留意すべき基本事項が述べられている。

また、本章は、受検者にとっても、テストを受けることで得られる利益や公正にテストを受ける権利を理解する上で大切である。

3.1 テストの趣旨や尺度の性質の理解

テスト利用者は、利用に先立ってテスト設計の意図、尺度得点の意味、テスト結果の解釈法などを理解しておかなければならない。

1 テストの基本設計の理解

テストは、その目的や利用場面にあわせて設計される（「1.1 テストの基本

設計」参照)。特定の目的や利用場面のために開発されるテストもあるが、多様な目的や利用場面に適用することを前提として開発されるテストもある。利用者は、テストの基本設計を正しく理解した上でテストを選択し、利用方法を検討する必要がある。学校における授業履修評価や資格試験など、特定の状況や目的に応じて開発されるテストでは、想定される受検者層や実施場面、結果の利用法などが明示されている場合が多いが、学校や進路指導における**適性テスト**や企業における人事テストなどは、多様な目的のために、適用する実施場面もある程度の幅をもたせて設計されているものが多い。

利用者は、テストの選択と利用を適切に行うため、そのテストの開発目的、想定されている実施状況、受検対象者や結果の解釈法などについて、理解しておかなければならない。テストの開発者、頒布者によってテストの実施と解釈のための手引、解説書が用意されていれば、利用者はそれらをもとにテストの目的やテストから得られる情報を検討し、場合によっては複数のテストを比較した上で利用するテストを選択する（「1.15 手引、解説書の内容」および「2.1 テストの選択」参照）。

一般に、広く頒布されているテストでは、テストの設計仕様が明示されて、事前に確認することができる場合が多い。設計仕様が明らかではない場合や、想定された利用場面が示されていない場合は、利用者自身が事前に**予備テスト**を行ってテストの適切さについて検討することが望ましい（「1.3 質問項目の設計」参照）。また、利用者が自身で開発したテストを用いる場合も、テストの性質をあらためて確認して適用の仕方を検討することが望ましい。

2 尺度得点の意味

尺度得点は、テストの開発目的を達成するためにさまざまな工夫がなされているべきである（「1.6 尺度化の方法」および「1.7 尺度の標準化」参照）。尺度得点の意味が明確な場合もあるが、標準化されたテストでは、基本設計や標準化の方法などによって意味が異なる。テストの利用に先立ち、手引や解説書などによって、尺度得点の意味を正しく理解しておくことが、テスト結果を適切に解釈して利用する前提となる。テストの信頼性や妥当性にかかわる情報も尺度得点の適切な理解には大切である（「1.10 尺度得点の確からしさの推

II 基本条項の解説

定と公開」および「1.11 尺度得点の適切さの確認」参照)。

また、標準化されたテストで、**主観的評定**によって採点されるテストを利用する場合にも、評定の手順や方法を確認し、尺度得点の意味を理解した上で利用法を検討することが大切である（「1.5 採点手続の設計」および「2.9 主観的な評定による採点」参照）。

3 テスト結果の解釈法

テスト結果の適切な解釈には、尺度得点の意味を正確に理解しておくことが前提となる。標準得点が用いられるテストにおいては、手引や解説書などに加え、開発者または頒布者が公開しているデータを確認すべきである。また、予備テストの結果、過去の利用から得られたデータの解析結果、参照可能な外部情報などから、テスト結果の解釈に役立つ知見を事前に得おくことも大切である（「1.3 質問項目の設計」、 「3.5 統計情報の公開」および「4.4 テスト結果の再利用のための保管情報」参照）。

3.2 拡大解釈の防止

テスト利用者は、テスト開発において想定された受検者層の範囲を超えて実施しないように、また、テストの手引で指示されていない過剰な解釈をすることがないように留意する。

1 基本設計にそった解釈

テストは、どのような目的でどのような受検者層に対して実施するかを定めて開発されており、その枠組みの中で尺度得点の意味づけがされている（「1.1 テストの基本設計」参照）。利用者はテストの基本設計を正しく理解し、それにそってテスト結果を解釈しなければならない（「3.1 テストの趣旨や尺度の性質の理解」）。利用者は、手引や解説書を読んで、開発目的や想定されている受検者層、尺度得点の性質、テスト結果の利用法について理解しておくべきである（「1.15 手引、解説書の内容」参照）。

手引や解説書に書かれた尺度得点の意味づけは、そのテストが開発時に規定

された目的で実施され、かつ、想定された受検者層に対して手順どおりに実施される枠組みを前提としてなされたものである。すなわち、尺度得点の適切さは、テストが基本設計にそって手順どおりに実施された場合に保証されているのであって、テストが単独でもつ属性ではないことに留意すべきである（「1.11 尺度得点の適切さの確認」、 「2.4 手順どおりの実施と不測の事態への対処」および「2.8 採点手順の順守」参照）。また、手順どおりの実施は、安定した測定結果を得ることには寄与するが、それだけでは尺度得点の適切な解釈に結びつかない。目的や受検者層が想定外である場合は、尺度得点の意味づけに用いた枠組みから外れることになり、テスト結果を適切に解釈することができない（「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。

2 テストが適用可能な受検者層

尺度得点が標準化されたテストでは、規準集団における**素点**の分布などを参照して、素点と標準得点の対応づけが行われる（「1.6 尺度化の方法」および「1.7 尺度の標準化」参照）。規準集団は、想定された受検者層から偏りなく抽出されているのが前提である。したがって、利用者は、テストの利用に先立ち、実施を予定している受検者が基本設計で想定されている受検者層に含まれるかどうか、そのような受検者におけるテスト結果の解釈の範囲について手引や関連文献などで確認しなければならない（「2.1 テストの選択」参照）。

想定範囲外の受検者がテストを受けた場合には、参照すべき根拠がないことから、素点と標準得点の対応づけは基本設計にそったものにはならず、尺度得点から適切な解釈が導かれるとは限らないことになる。もし、想定範囲外の受検者に対してテストを実施してしまった場合には、その新たな受検者に対する解釈の可能な範囲を、手引や関連文献などをもとに慎重に検討する必要がある（「1.15 手引、解説書の内容」参照）。

3 テストの結果の過剰な解釈の防止

(1) 測定誤差に関する適切な解釈

テストの結果を解釈する際には、尺度得点には**測定誤差**が含まれることを考

II 基本条項の解説

慮すべきである。テストの結果が数値で表されると、その数値が絶対的なものとみなされ、本来は本質的な意味のない小さな差であっても、それが不当に強調されたり、それによって序列化されたりすることがある。測定誤差や**信頼区間**の大きさなどの指標を参考に、信頼性を考慮して尺度得点の差異がもつ意味について適切な解釈をく다すべきである（「1.10 尺度得点の確からしさの推定と公開」および「3.5 統計情報の公開」参照）。

(2) テスト結果による予測可能な範囲の順守

テストの結果をもとにした予測や推論は、その限界を適切に理解した上で行うべきである。予測可能な範囲を超えた解釈は誤りである。たとえば、職業適性検査の得点が良好で、いくつかの職業に向いているとされた場合でも、**妥当性**の証拠が示されていない職業にまで適性があるかのような解釈は避けなければならない（「1.11 尺度得点の適切さの確認」参照）。また、習熟や加齢によって変化し得る得点に対して、不変のものであるかのような解釈はできない。

テストの結果は、受検者の**特性**の一部を表しているにすぎない。受検者の人格を表したものであるが、まして、人間の価値を表したものではない。受検者の特性の一部しか測定していないにもかかわらず、他のさまざまな特性をも測定しているかのような解釈や誤解を与えるような表現に気をつけなければならない。

(3) 先入観による歪曲の回避

テストの結果を解釈する際に、受検者に対する先入観に左右されることがあってはならない。たとえば、**学力テスト**の結果が思わしくない生徒に関して、その生徒の**知能検査**の結果が良かったという事前情報を得ていると、得点の低さを一時的な体調不良をその原因とするなど、テストとは本質的に無関係な推論が入り込む可能性がある。的確な情報が付加されることで、テスト結果の解釈がより適切になる場合もあるが、テストの手引などに記載されている事項とは異なる外部からの情報は、先入観として一定のバイアスを生む場合も少なくないので、このような情報を加えて解釈を行う場合には慎重さを要する。

4 テスト利用者の研鑽

テストの結果から、利用目的に都合のよい歪んだ解釈や過剰な解釈を行ってはならない。テスト結果を適切に解釈するには、手引に示された基本設計の枠

組みでテスト得点を解釈していくことが基本となる（「3.1 テストの趣旨や尺度の性質の理解」参照）。実施や利用にあたり、手引の解説を理解した上で、さらに一定の経験が必要とされるテストも少なくない。そのようなテストについては、研修や研究会に参加したり、開示されている事例を参考に解釈の検討を行ったりするなど研鑽を積むことが望まれる（「1.15 手引、解説書の内容」および「6.1 テスト関係者の責務」参照）。

3.3 判定基準の個別検証の必要性

テスト結果を合否などの判定に利用する際は、理論的、実証的な研究に基づいて個々の具体的な状況を考慮し判定基準を設計する。他の状況で検証された基準や一般的な基準を適用する場合は、その適否を専門家の支援を得ながら検討する。

1 判定基準設計の考え方

入学者選抜や入社試験、あるいは技能検定や資格試験などにおける合否判定では、ある判定基準にそってテストの結果を利用し、合否判定や選考が行われることが多い。テストの結果以外の資料が判断材料として加味される場合もあるが、判定のいずれかの過程でテストの結果を用いた判断がなされる。判定基準を設計する際は、合格のための要件を明確にし、それに応じた判定のルールを定めるべきである。判定基準は受検者にも開示されることが望ましい（「2.3 実施前の受検者への説明」参照）。

合否の境界値の決定方法や複数の**下位尺度**の尺度得点の使い方などは、設計を誤るとテストの目的に照らして大きな問題が生じることがある（「1.2 測定内容の定義と構造化」、 「1.7 尺度の標準化」、 「1.8 複数の尺度得点の比較」および「1.9 複数のテストの共通尺度化」参照）。たとえば、複数の下位尺度の尺度得点を、各下位尺度の平均値と**標準偏差**（ばらつき）をそろえないまま合計すると、特定の下位尺度による順位が合計得点に強く影響して、複数の下位尺度を用いて判定する効果が低くなる場合がある。

また、同じ目的で複数回実施される試験では、実施回ごとに合格者群の質が

II 基本条項の解説

大きく変わらないように、合否判定基準の性質を十分に理解した上で、慎重に判定基準を設計しなければならない（「1.1 テストの基本設計」、 「1.6 尺度化の方法」および「3.1 テストの趣旨や尺度の性質の理解」参照）。

ここではテストの結果で合否が決定される状況を想定して、合否判定基準設計の考え方を説明する。

2 合否の境界値の決定方法

(1) 合格者数が一定の場合

合格者数、または、定員があらかじめ決まっている場合には、各受検者のテストの結果を比較することで合否が決まる。いわゆる、競争試験の状況である。辞退者が出ると予想される場合は、合格者数は定員より多く設定するのが一般的である。

競争試験では、合格の相対的な難しさは、合格者数に対する志願者数の比（競争率）で決まる。テストの**識別力**は、合否の境界付近で高くなるように設計されるのが望ましい。そのためには、競争率が高い場合は、テストの**難易度**を高く、逆に、競争率が低い場合は、テストの難易度を低くするのが望ましい。テストの質問項目の作成を行う前に競争率が予測できる状況であれば、質問項目を作成する際に、難易度の調整を行うことができる（「1.3 質問項目の設計」参照）。

(2) 合格者数が不定の場合

ある一定の基準を満たすと判定された者が合格とされる試験は、通常、資格試験と呼ばれている。他の受検者の結果とは無関係に本人の結果だけで合否が決まるが、合否判定基準の設計の面では合否の境界値を定めるのが難しく、経験豊かな専門家の支援が欠かせない。

資格試験の合否判定基準は、質問項目の難易度や内容が一定であることを前提として、尺度得点にある境界値を定め、それ以上の者を合格者とする方法が一般的である。同一の受検者層を想定した規準集団による予備テストに基づき、**項目分析**を行って毎回のテストの等質性を保つことが望ましい。予備テストが不可能な状況では、テストの質問項目作成の際に細心の注意を払って内容や難易度について検討して、テストの品質が安定するように努力する必要がある。

質問項目の作成時点で難易度の調整ができない場合には、各回の合否の境界値が上下に大きく変動することもありうる（「1.1 テストの基本設計」および「1.3 質問項目の設計」参照）。

合否判定基準が一定でも、各実施回の受検者層が異なれば、合格者数や合格率に変動が生じる可能性がある。また、毎回のテストの尺度得点が共通尺度化されていないければ、このような変動が質問項目の難易度や内容の揺れによるものなのか、受検者の特性分布の変動に由来するものなのか、区別するのは困難である。

3 複数の下位尺度を用いた合否の判定方法と問題点

複数の下位尺度の尺度得点を用いて合否判定を行う場合、代表的な合否判定基準には以下のようなものが考えられる。テストの目的を考慮して、どの方式を用いるのが適切かを決めることになる。

(1) 総合得点方式

総合得点がある値以上であれば合格とする方法である。一部の下位尺度の得点が著しく低くても、別の下位尺度の得点で補って合格に達する可能性がある。

(2) 最低基準値方式

各下位尺度に最低基準値を設け、合格に達するためにはすべての尺度得点で基準を超えなければならない。最低基準値方式を用いる場合、合否を決する基準値を低めに設定しないとすべてのテストで基準を満たす者の数が少なくなり、結果的に合格者数が少なくなることに注意する必要がある。

(3) 最高基準値方式

いずれかの下位尺度または特定の下位尺度の尺度得点がある基準値以上であれば合格とする方法である。最高基準値方式では特定の下位尺度の結果だけが合否にかかわることになり、他の下位尺度の情報が合否に反映されない。合否を決する基準値を高めに設定しないと、いずれかの下位尺度で基準を満たす者の数が多くなり、結果的に合格者数が多くなることに注意する必要がある。

4 下位尺度の尺度得点の総合

総合得点方式によって合否判定を行う場合には、各下位尺度の素点方式によ

II 基本条項の解説

る尺度得点の合計を総合得点とする方式と、各テスト得点を**換算点**方式で尺度得点化してから合計して総合得点とする方式が考えられる（「1.6 尺度化の方法」および「1.7 尺度の標準化」参照）。いずれの方法をとるかで結果が変わってくることに注意しなければならない。

総合得点化するには、各下位尺度の重要性を考慮して**重みつき加算方式**によって総合する場合がある。その場合、重みが本当に利用者の意図を反映したのものになっているかどうか、データに基づいて確認することが望ましい。

5 判定の不完全性

どんなテストでも、尺度得点には測定誤差が含まれる。したがって、テストを用いた判定結果も常に不完全なものであることを認識すべきである。判定基準を厳格に適用しても、本来、合格の要件を満たしているはずの者がたまたま実力を発揮できなくて不合格になったり、逆に合格の要件を満たさない者が運よく合格したりする可能性は常に存在する。テストの信頼性や妥当性が高ければ、判定の不完全性は小さくなる（「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。

資格試験の場合、合格の要件を満たす者が不合格になる危険性と、合格の要件を満たさない者が合格する危険性は二律背反の関係にある。基準を下げて合格者を多めに出せば、要件を満たす者が不合格になる危険性は少なくなる代わりに要件を満たさないものが合格になる危険性が高くなる。基準を上げて合格者を絞り込むと、その逆のことが起こる。判定基準を設計する際には、どちらの危険性を減らすことを重視するかを考慮に入れる必要がある。

6 合否判定基準の周かな準備および検証の必要性

合否判定基準を決める際には、各方式の特徴を考慮して設定しないと、不自然な判定結果をもたらすので注意が必要である。また、試験の実施ごとに判定方式を変更することは望ましくない。合否判定基準の決定には、専門家の意見を参考にした周かな準備が必要である。

合否判定基準の適切性については、受検者数、定員、使用するテストの尺度得点、総合得点に関して、異なる判定基準を用いた場合の合否判定結果につい

でも比較し、検討を行うことが望ましい。そうした分析結果の積み重ねを記録に残し、公開することは、判定基準を再検討する際にも大変有益な資料となる（「3.5 統計情報の公開」, 「4.1 テスト結果の保管手続」, 「4.2 テスト実施後の調査と検討」および「4.4 テスト結果の再利用のための保管情報」参照）。

3.4 受検者への結果報告

学校教育、企業内教育、臨床診断、カウンセリング、職業ガイダンスなどにおいて、受検者にテスト結果を伝える場合には、テスト結果が受検者に適切に解釈されるよう努める。利用者は、そのために必要な知識や技能を事前に修得しておく必要がある。

1 受検者への結果報告の状況依存性と研修の必要性

テストは、さまざまな状況においてさまざまな目的をもって利用されている（「1.1 テストの基本設計」参照）。学校や企業などにおいては、その集団や組織の成員の教育、指導のために必要不可欠な道具として利用されるほか、成員の選抜や配置などの目的でも利用されている。臨床診断やカウンセリングの場面では、クライアントの状態を把握して、適切な治療や支援を選択するための参考としてテストが用いられる。

多くの場合、受検者は、学習する立場、選抜される立場、あるいは、治療や支援を必要とする立場の者である。利用者は、状況や目的に合った適切なテストを選択し、テストの結果を適切に解釈することによって意思決定を行うのももちろんであるが、それと同時に、受検者に対してテスト結果を適切に伝え、テスト結果が受検者に正確に解釈されるように十分に配慮する（「3.1 テストの趣旨や尺度の性質の理解」参照）。利用者は、受検者などへの結果報告のあり方がテストの目的と状況によって異なることに、常に留意する。

2 臨床診断やカウンセリングのためのテストの場合

臨床診断やカウンセリングのためのテストでは、受検者の状況によっては、受検者本人がテストの結果を知ることが治療やその後の学習の妨げになる場合

Ⅱ 基本条項の解説

もあるので注意が必要である。利用者は、テストの特性について十分に把握しておくとともに、状況に応じた使い分け、テスト結果の解釈の多様性や結果の伝え方などに関して一定の研修を積み重ねておくことが望ましい（「3.1 テストの趣旨や尺度の性質の理解」、「3.2 拡大解釈の防止」および「6.1 テスト関係者の責務」参照）。

3 教育や指導を目的としたテストの場合

一般に尺度得点化されたテスト結果だけが強調されることが多いが、教育や指導の目的で用いられるテストについては、第一義的に重視すべきことは、テスト結果の解釈が利用者や受検者にとって有効な意思決定に結びつくことである。そのため、各受検者の特徴が明らかになるような形でテストの結果が示されるべきである。たとえば、学習達成度を確認する目的のテストでは、受検者は返却されたテストの答案を検討することで、その後の学習への示唆を得ることができる。利用者は、受検者に対して尺度得点のみに着目するのではなく、個々の質問項目の正誤にも注目することの重要性を指導するとともに、正誤の理由が個々の受検者に明確に伝わるように回答例を示すなどの工夫が望まれる（「1.13 質問項目の内容開示の是非」参照）。また、テストの結果を受検者にわかりやすく表現する工夫として、順位や**偏差値**などの尺度の個人間の差を示す情報だけでなく、学習領域別の達成度のような個人内プロフィールを示す情報を提供することもよい。ただし、受検者が自らの個人内プロフィールについて正しい解釈を行うことは必ずしも容易ではないので、ていねいな解説を添えるべきである（「1.8 複数の尺度得点の比較」参照）。テストの結果の解釈について、利用者が受検者に適切な指導や助言を行うことで、受検者にとってテストの結果が有効に活用されることが期待できる。

4 選抜や配置を目的としたテストの場合

選抜や配置を目的として用いられるテストについては、個々の受検者に詳細な結果が返却されることはまれであろう。たとえば、選抜の場面では、テスト結果を利用した合否の結果だけが受検者に伝えられ、詳細なテスト結果が伝えられることは少ない。しかし、尺度得点の統計的な指標など、テストの特性を

示すための値は公表されることが望ましい（「3.5 統計情報の公開」参照）。それらの情報により、個々の受検者は受検者集団内の自分の相対的な位置をある程度把握できる。また、テストの品質についての情報を得ることができる。次回のテストに臨む受検者にとっては学習意欲が高まることも期待される。さらに、実施者や利用者にとっては公正な選抜が行われたことを受検者に理解してもらい、選抜結果に納得してもらうことも期待できる（「6.2 テストの公平性」参照）。

個々の受検者にそれぞれのテストの結果が通知されない場合でも、個人情報としてテストの結果の開示請求があった際には、利用者および管理者は慎重かつ的確に対応することが求められる。とくに、選抜の場合には、不合格になった受検者から異議が申し立てられたり、テスト結果の開示が要求されたりすることも少なくない。そのような状況を想定して、開示および不開示の判断基準の設定、テスト結果の蓄積と保管方法、**情報セキュリティ**の確保、開示請求などに関する受付窓口の体制、開示の様式や方法、本人確認手続の詳細など、開示請求に速やかに対応する準備を整えておく必要がある（「4.1 テスト結果の保管手続」および「6.3 受検者の疑義への誠実な対応」参照）。

なお、採点手続の変更など、受検者にとって自らのテスト結果にかかわる情報が変更される事態が生じたときには、その旨を公表すべきである（「2.4 手順どおりの実施と不測の事態への対処」および「2.10 採点方法の変更」参照）。

3.5 統計情報の公開

テストの品質を検証し、また改善の方向を見出すために、テスト結果の統計情報が公開され、活用されるべきである。

1 統計情報公開の目的

テストの実施結果にかかわる統計情報を公開する目的は、利用者にテストの基本設計や測定内容を示し、テストの選択のために有益な情報を提供することである（「1.1 テストの基本設計」、「1.2 測定内容の定義と構造化」および

II 基本条項の解説

「2.1 テストの選択」参照)。また、開発者、頒布者、実施者、利用者および管理者などの当該テストの関係者以外の第三者によって、テストについての研究、あるいは、テストを利用した研究が促進されることにより、テストの品質が検証され、改善の方向が見出されることにある（「1.12 テストの改訂」参照）。したがって、公開される統計情報は、第三者が入手することを前提として、受検者の個人情報を含まない、統計的に集約されたものでなければならない（「3.6 研究目的でのテスト結果の利用」、 「4.1 テスト結果の保管手続」、 「4.2 テスト実施後の調査と検討」および「4.4 テスト結果の再利用のための保管情報」参照）。

2 品質の検証のための統計情報

尺度得点として数値化されたテストでは、テストの基本的性質を示す統計情報として、たとえば、尺度得点の分布や平均値、標準偏差などを公開する。複数の下位尺度で構成されるテストでは、下位尺度間の**相関係数**なども有益な情報である。また、受検者集団の性質に関する情報も、統計情報を解釈する上で有益である（「1.2 測定内容の定義と構造化」参照）。

尺度得点がどの程度安定しているかを示す**信頼性係数**や、尺度得点が測定目的とする特性をどの程度適切に測定できているかという妥当性の根拠となる統計情報も、テストの品質を検証する上で重要な情報である（「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。

3 公開された統計情報の解釈に関する注意点

研究のために実施されるテストでは、個人間の得点の違いよりも、集団の平均値や分布を把握したり、集団間の平均値や分布の違いを検討したりすることを目的とすることが多い。それに対して、人の処遇にかかわるテストでは、各受検者の得点そのものが本人のその後の意思決定に重大な意味をもつ。したがって、人の処遇にかかわるテストでは、適切な判断を下すために、より測定精度の高いテストが用いられることが望ましい（「1.10 尺度得点の確からしさの推定と公開」参照）。

翻訳されたテストについては、新たに**項目分析**を行ってテストの性能を確認

しておくべきである。原語で分析された統計情報をそのまま利用することはできない（「1.3 質問項目の設計」および「1.14 テストの他言語への翻訳」参照）。

公開された統計情報は、その数値を鵜呑みにするのではなく、テストの目的、質問項目の内容や構成、受検者集団の性質、受検時期などを考慮に入れた上でテストの良し悪しを総合的に判断することが大切である。

3.6 研究目的でのテスト結果の利用

テスト結果を研究、再標準化、適用法の設計などの目的で利用する場合は、個人が特定できないような形で統計的に分析する。

1 テストの品質向上のための研究の重要性

テストの品質を向上させるためにはテストの実施結果をさまざまな観点から分析することが重要であり、研究目的でのテスト結果の利用は積極的に奨励されるべきものである（「4.2 テスト実施後の調査と検討」参照）。たとえば、尺度得点の信頼性の確認（「1.10 尺度得点の確からしさの推定と公開」参照）には、受検者の回答データが必要である。また、尺度得点の妥当性の確認（「1.11 尺度得点の適切さの確認」参照）も、理念的・概念的な検討にとどまらず、データを分析しながら実証的に行うべきである。さらに、開発後の社会的な状況の変化に伴うテストの改訂の必要性については、テスト結果を用いた分析や研究に基づいて検討することになる（「1.12 テストの改訂」参照）。

2 個人情報の保護

テストを実施する目的は、受検者個人の処遇にかかわることのほか、研究、調査などさまざまである。テストの実施にあたって受検者から取得する情報やテスト結果の内容、形式についても多様なものが考えられる。テストの目的、内容および形式にかかわらず、氏名などのように直接的に得られた個人情報はもとより、他の情報（氏名、生年月日、住所等）との照合により個人が識別される場合には、テストの回答そのものも基本的に保護されるべき個人情報である。

II 基本条項の解説

る。

管理者をはじめとするテスト関係者は、テストの実施から、保管、廃棄にいたる過程で個人情報外部に流出し、個人の権利や利益が侵害されることのないようにしなければならない。とくに、テスト結果については電磁的記録として保有されることが多い。テストの管理者には、情報管理のための規定を作成するとともに、第三者のテスト結果へのアクセスを防止し、アクセスの権限を有する特定の立場の者以外が情報に接触することがないようにするため、具体的な方法を講じることが求められる（「4.1 テスト結果の保管手続」参照）。

3 研究目的での個人情報の利用

テスト自体の研究やテストが測定している特性などの研究のため、受検者の回答などのテスト結果を用いて分析が行われる。そのような場合には、個人の特定につながる情報を取り除いた上で、統計的に分析することが求められる。研究の必要性から個人を識別する情報を利用する場合には、発表される分析結果の中に、受検者の氏名、生年月日、住所等の個人の特定につながる手がかりが残らないようにしなければならない。この点は、個別の事例を用いた研究結果を公表する場合にはとくに注意する必要がある。

また、そのテスト結果を広く共有し、テスト結果の情報提供を行うべきケース（「4.3 公的な性格が強いテスト結果の共同分析利用」参照）や、テスト結果を外部の機関や個人に委託して研究を行うケースもある。前者の場合、回答などについて個人が特定できないように匿名化することが必須であり、後者の場合にも匿名化が基本である。いずれの場合も、個人の権利や利益が侵害されることのないよう、情報提供から研究、研究結果の公表にいたる過程において個人情報の管理には細心の注意を払う必要がある。

4章 記録と保管

テストにかかわるさまざまな情報は、テストそのものの改善やテストを有効に利用するための資料として貴重なものである。目的にそったテストの利用は当然のこととして、将来への資産として十分に活用すべきである。一方、テストの結果にかかわる情報の利用が活発になるにつれて、個人情報の保護と情報漏洩を防ぐ手立てもいっそう重要になる。本章では、そのような観点からテスト情報の管理および保管に関する留意事項について述べる。

本章では、テスト情報の管理は、受検の受付、テスト実施による情報の取得、テスト結果の利用を経て、最終的に情報を廃棄することにより完了すると考える。その一連の流れを念頭に置きつつ、具体的には、以下の4つの条項に分けて解説を行った。

4.1では、テストを実施した結果として得られた情報の保管手続にかかわる留意事項について述べている。4.2では、テストを実施した後に行うべき調査や分析の内容、4.3では、テストの中でも、とくに個人の特性の測定を目的としたものではなく、もともと公共性が高い目的で設計されたテスト結果の活用の必要性について述べている。4.4では、さまざまな目的でテストの結果を再利用するために保管すべき情報の内容などについての条項となっている。

本章は他の章と比べて基本条項が4項目と少なくなっているものの、とくに実施者、管理者および研究者にとってはおろそかにできない内容である。

4.1 テスト結果の保管手続

テストの管理者は、テスト結果の保管に際して、保管するデータ、その保管期間、保管方法、利用目的、利用方法、期間経過後の得点利用などに関する規定を作成しておき、関係者に順守させる。なお、このテスト結果の保管においては、個人情報が漏洩しないよう万全を期さねばならない。

II 基本条項の解説

1 テスト結果の保管手続に関する規定の作成と順守

テスト情報の管理は、受検を受け付け、テストを実施して情報を「取得」することから始まり、テスト結果の「利用」を経て、最終的に情報を「廃棄」または「消去」することにより完了する。

テストの内容や結果には、適切なテストの選択、結果の理解、テストの改善などに役立つ情報もあれば、受検者が知ることによりテストの実施や利用に支障が生じる情報もある（「1.13 質問項目の内容の開示の是非」、 「2.1 テストの選択」、 「3.4 受検者への結果報告」および「3.5 統計情報の公開」参照）。テストで得られた情報には、テストの性能の分析や、今後の改善のために不可欠なデータが含まれており、実施回ごとにすべての情報を廃棄すべきではない（「1.3 質問項目の設計」および「1.12 テストの改訂」参照）。ただし、特定の集団や個人に関する情報も含まれるため、適切な保管手続が求められる。志願者や受検者の属性などを保存する際には個人情報をも可能な限り除去し、保管手続や利用規定を明文化しておくべきである。管理者は実施責任者と十分に協議を行いつつ、個人情報の保護に関する法律（**個人情報保護法**）、関係する組織や団体における規定やガイドラインなどを踏まえて、テストの実施に伴って入手した個人情報の管理を含めたテスト結果の保管手続に関する規定を作成する。さらに、この規定を関係者に周知、順守させることが求められる。また、管理者は必要に応じて規定を見直し、管理体制を整備し、関係者の教育、研修などを通じて、テスト結果の安全な保管体制の継続に努める（「2.2 テスト用具の管理」、 「4.2 テスト実施後の調査と検討」および「6.1 テスト関係者の責務」参照）。

2 利用目的の範囲内での利用

(1) 保管するデータ、保管期間

テストの結果として保存すべき情報には、**尺度得点**や判定結果のほか、**多枝選択式**において選ばれた選択枝、各質問項目の正誤回答、**記述式**のテストにおける記述回答、**面接**における受検者の発言内容などの回答状況、質問項目の一部を選択して回答する形式の場合には回答した質問項目の選択状況などが挙げられる。また、各実施回の質問冊子や回答用紙、正解表なども保存する。それ

らを記録したファイルの内容には、回答、回答用紙、質問冊子、採点結果一覧表、採点者、質問項目や下位尺度ごとの採点結果一覧表、マークシートの読み取り結果および機械採点の電磁的記録、面接などにおける個別の評定票、カルテなどに記入する検査結果や診断結果の記録票、成績通知書、プロフィール表などを含むことが考えられる。しかし、個人情報としてのテストに関する情報を何でも漫然と保存しておくことは、個人情報の適切な保護という観点からは望ましくない。個々の情報について保存するか消去するか、また、保存する場合にはどのような形態でどのくらいの時期にわたって保存するか、という判断は、調査、分析および研究のための情報としての必要性、個人情報としての機密性および管理のしやすさなどの観点から慎重に行うべきである（「4.2 テスト実施後の調査と検討」参照）。

保存すべきデータについては保存形式や期間および利用規定、消去すべきデータについては消去時期およびその方法などについての規定を作成することが必要である。また、テストの実施および採点時に何らかの不測の事態が生じた場合には、その状況と対応について詳細に記録し、類似のアクシデントを未然に防止するための資料として活用する（「2.4 手順どおりの実施と不測の事態への対処」参照）。

テストの**妥当性**に関する検討などのために当該テスト以外の情報と連結した場合には、連結に利用した個人情報を削除した状態で分析を行うのが望ましい（「1.11 尺度得点の適切さの確認」および「3.6 研究目的でのテスト結果の利用」参照）。

(2) 利用目的の明示と公表

利用者は、テストの趣旨や性質を理解した上で、あらかじめ受検者の同意を得た場合および法令により例外として認められた場合（「3.6 研究目的でのテスト結果の利用」参照）を除き、明示または公表した利用目的の範囲内でのみテストの結果を取り扱う。利用目的、利用者の範囲、問い合わせ窓口などの必要な情報を明示し、同意を得た上でテスト情報を取得するよう努める（「6.3 受検者の疑義への誠実な対応」参照）。

質問項目の内容については、開示可能な場合と開示してはならない場合がある（「1.13 質問項目の内容開示の是非」参照）。開示してはならない場合には、

II 基本条項の解説

テスト結果の情報に準じた厳重な管理が必要である。

3 保管方法

(1) 安全管理措置

管理者は、テスト結果に関する情報の漏洩や滅失の防止など安全管理のために必要かつ適切な措置を講じなければならない。適切な管理のために、保管場所、保管期間、廃棄の手続、管理責任者の指定、電磁的記録の取り扱いなどについての規定を設け、情報が記録されている文書またはファイルの作成から保存、消去までの管理体制について整備し、管理簿などによる一元的な管理を行うことなどが求められる。

規定として明記しないまでも、個人情報の含まれた文書などのコピーを安易に作成しない、会議などで配布した個人情報の含まれる資料を回収する、文書の送付は親展扱いとする、電子メールを利用する場合には誤送信のないようにするとともに暗号鍵やパスワードなどにより秘密保持を図る、保管の必要のない個人情報や個人データは速やかに廃棄する、などの情報漏洩対策を励行することが望ましい。

テスト結果にかかわるデータの消去についても、保管と同様な視点から検討し、規定を作成することが望ましい。

(2) 電磁的記録の保管およびバックアップの手続

電磁的記録として保管されるテストの結果については、事故などに備えてバックアップを行っておくことが望ましい。また、バックアップされたデータについても、オリジナルデータと同様の利用方法、保管手続を適用する。

①手続および責任の明確化

保管されたテスト結果の安全性を確保し、不正アクセス、不正使用などを防止するため、テスト結果の情報に応じて重要性のレベルを設定し、保管およびバックアップの手続、保管場所、テスト結果の保管管理者、バックアップの作業責任者およびその権限の範囲などを明確にすべきである。

②管理簿などによる一元的な管理

保管すべきテスト結果は、必要に応じていくつかのファイルに分け、それぞれのファイルごとに、利用や保管の状況、バックアップの申請、承認および作

業の実行経緯，アクセスの承認，などの事項を管理簿に記録する。

(3) 不正アクセスの防止

①保管場所の管理

テスト結果を記録している媒体は，保管庫や金庫などの施錠できる場所で保管する。保管庫などについては，鍵の保管責任者を明確にし，「鍵管理簿」ないしは「鍵使用記録簿」などを設け，その鍵の使用状況などを記録する。

②電磁的記録へのアクセスの制限

サーバなどに電磁的記録としてテスト結果を保管する際には，テスト結果を扱うアクセス権限者以外の者がテスト結果にアクセスできないようにするため，暗号鍵やパスワードを使用し，それらの登録，変更および抹消の手続を明確に定める。保管管理者ならびにアクセス権をもつ人に付与する権限は適切な範囲に限定することが求められる（「5.4 機密保持や実施管理への対策」参照）。

4.2 テスト実施後の調査と検討

テストの開発者，実施者，利用者は，テストが意図された特性を測定していたか，実施目的，対象者，適用場面に適切であったか，テストの利用がどのような影響を及ぼしたかなど，過去の資料と照合しつつその結果について調査分析し，得られた知見も開示するよう努める。

1 テスト実施後の調査，検討の意義

(1) テストの性能の確認

尺度得点の確からしさと適切さを確認することはきわめて重要である（「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。しかし，テスト実施前に**予備テスト**を行うことなどによって事前の確認ができない場合，テストの実施後に調査，検討を行う必要がある。このために，必要なデータを利用できるように保管しておかなければならない。テストの尺度得点などの情報以外の情報についても必要に応じて適切に保存，蓄積し，適宜公開することで，テストの特性や得点の確からしさ，適切さ，テスト利用の状況などに関する情報分析が可能になる（「3.5 統計情報の公開」

II 基本条項の解説

および「4.1 テスト結果の保管手続」参照)。

(2) データの蓄積

繰り返し実施されるテストにおいては、実施のたびにデータを蓄積していく体制を整えておくことが望ましい。質問項目ごと、あるいは、下位尺度ごとの尺度得点の分布を調べるだけでも、**測定誤差**がどのようなところで大きな影響を与えているのか検討することができる(「1.2 測定内容の定義と構造化」、「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照)。たとえば、多枝選択式のテストの場合、選ばれた選択枝の情報まで保管しておけば、選択枝レベルの細かな**項目分析**ができる(「1.3 質問項目の設計」および「1.4 回答方法の設計」参照)。一定の判定基準を設けているテストの場合、判定基準が合理的で安定したものであったかどうかを吟味することも重要である(「3.3 判定基準の個別検証の必要性」参照)。このような分析は、利用者が適切なテストを選択するためには重要なプロセスである(「2.1 テストの選択」参照)。

(3) 初めて実施されたテストの分析

質問項目を一度しか利用しないテストや初めて実施されたテストの場合、総合得点や下位尺度ごとに、得点の分布をさまざまな視点で検討することによって、**外れ値**など特異なデータの存在や、分布の歪みの状況などから予期せぬ測定誤差の混入の可能性を検討することができる。たとえば、入力ミスなどによる異常なデータが発見される可能性もある。必ずしも統計的な指標だけから原因を同定することはできないが、項目作成や実施の工夫で測定誤差を減らす手がかりが得られるかもしれない。テストの実施とともに受検者に質問紙調査を行ったり事後に面接調査を行うことができれば、測定誤差の原因を探る有力な手がかりが得られる可能性もある。このような分析は、テストにかかわる研究者に期待される役割のひとつである。

2 テストの社会的影響

テストは、その基本設計の段階で考えられた枠組みを超えて社会的に大きな影響力をもつ場合がある。**知能検査**、**適性テスト**、性格検査など、利用場面の範囲が広いテストでは、それらのテストが適切に実施、利用されているか、常

に確認することが必要である。テストの社会的影響力についての調査研究が行われ、その成果や検討の結果得られた知見について開示することも重要である。入学試験や各種国家試験など、大規模で、かつ、受検者の処遇に対して重大な影響を及ぼす可能性があるテストにおいては、テストそれ自体のみならず、それを必要とする社会的制度面も含め、多角的な視点からテストがもたらす社会的影響について慎重に研究、分析を重ねていくべきであろう（「4.3 公的な性格が強いテスト結果の共同分析利用」参照）。

なお、このような調査、検討の過程で個人情報外部に流出することがないように厳格に管理することは重要である（「3.6 研究目的でのテスト結果の利用」、 「4.1 テスト結果の保管手続」および「4.4 テスト結果の再利用のための保管情報」参照）。

4.3 公的な性格が強いテスト結果の共同分析利用

特定個人の処遇を問題にしない公的な性格が強い大規模テストの結果情報は、公共的情報資産として提供されるべきである。

1 公的な性格が強い調査、テスト結果の情報公開の考え方

施策を立案したりその方向性を定めたりする際の参考とするために、全国、あるいは、地域ごとに、世論調査や学力調査などの大規模な調査やテストが実施されている。この種の大規模な調査やテストは、多くの人々に影響力をもつ公的な性格が強いテストであり、設計、実施から記録、保管にいたる一連のプロセスにおいて、本規準にそった厳格な品質管理が求められる。なかでも、調査やテストの実施結果は公共的な財産であり、テストに直接携わる関係者以外の者にも利用可能な形で一般に公開されるべきである。公開された情報は、多様な目的による**二次的分析**に利用可能な形で供される必要がある（「4.2 テスト実施後の調査と検討」参照）。

2 提供情報とその利用

公的な性格が強い調査やテストにおいて、第一の利用者は実施主体であるが、

II 基本条項の解説

同時に政策の影響が直接及ぶ者やそれ以外の一般の人々も利用者と考えられる。したがって、調査やテストの品質についても一般の利用者が検証可能な形で情報公開が行われる必要がある。すなわち、テストに関して一定の専門知識と技術をもつ研究者であれば、公開された情報に基づいてテストの品質の検証が可能となる状況を用意すべきである。

個人の処遇にかかわらない公的な性格が強い調査やテストでは、回答者個人が特定されないという条件のもとで、各種統計指標の算出のために必要となる詳細なデータそのものを公開すべきである。さらに、テスト結果の情報のみならず、テストや調査の目的、実施プロセスなどの関連情報も合わせて一般の利用者が共有できるような配慮が望まれる（「4.4 テスト結果の再利用のための保存情報」参照）。

3 集約された結果の公表における配慮

一方、公的な性格が強い調査やテストの結果は、一般の利用者が専門的知識を必要とせずに調査やテスト結果の概要を理解でき、また、そこから導き出される結論や調査の意義も理解できるように、集約された形として公開される必要がある。その際、専門知識の不足による無用な誤解を防ぐため、統計量の読み方に関する簡単な解説や結果の解釈にかかわる留意点を合わせて提示しておくことが望ましい。

個人の処遇を直接的な目的としない公的な調査やテストでは、テスト結果が特定の個人に直接的な影響を及ぼすことは少ない。その反面、結果の解釈が誤った形で広まったときの悪影響は甚大である。可能な限り多くの人が結果を正しく理解するための解説の提示とその普及活動はテストリテラシーを高めるためにきわめて重要である（「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照）。

4 テスト結果の保管

管理者は、公的な性格が強い大規模テストの結果情報やデータが調査対象者の特性を把握するための貴重な財産であることを認識し、一連の分析が完了した後においても、将来、別の視点で再吟味することが可能な状態でデータを保

存すべきである（「4.4 テスト結果の再利用のための保管情報」参照）。適切な保管手続をとるとともに、データの記録媒体が時代遅れのものにならないように保守管理を行い、バックアップを作成して貴重な知的財産が失われないように努力する必要がある（「4.1 テスト結果の保管手続」参照）。

なお、受検者が特定できる尺度得点などの個人情報にアクセスすることができないように保守管理することは、当然の措置である。保管に際しては、受検者個人のプライバシー保護には細心の注意を払うことはいうまでもない（「3.6 研究目的でのテスト結果の利用」参照）。

4.4 テスト結果の再利用のための保管情報

テストの開発者、頒布者、利用者は、協力して効果的な利用事例や利用後の実証的なデータの解析結果を蓄積し、テストや尺度の意味内容に関する知識を共有するよう努める。また、これらのデータが正しく解釈されるよう、実証の過程を詳細に記録しておく必要がある。

1 実施にかかわる情報の蓄積

(1) データ収集、蓄積の理由

テストの質問項目や受検者の構成、実施環境は常に同一とは限らない。標準化されたテストでも、目的にそって適切に利用するためには、実施回ごとに開発者が想定した設計にそって実施されているか、開発者や頒布者が利用に対して十分な説明責任を果たすことができるか、利用者が利用目的を達成する手段として適切であるかといった点について、必要な資料を収集し、蓄積していくことが求められる（「1.1 テストの基本設計」参照）。

(2) 蓄積された情報の利用

テストを実施して得られた情報を目的にそった意思決定に利用するのは当然のことであるが、それに加えて、事後の調査と検討が可能な形で保管されるべきである（「4.1 テスト結果の保管手続」および「4.2 テスト実施後の調査と検討」参照）。尺度得点の確からしさや適切さの確認や、テストの改訂には、蓄積、保管されたデータや解析結果が必要となる（「1.10 尺度得点の確から

II 基本条項の解説

しさの推定と公開」, 「1.11 尺度得点の適切さの確認」および「1.12 テストの改訂」参照)。

2 テスト結果にかかわる実証的データの分析と蓄積

(1) 統計資料の算出

テスト結果を実証的に分析するには、最低限、各種の統計指標を算出することが必要である。たとえば、**信頼性係数**の算出や妥当性の検討を目的とした他の指標との**相関係数**などを算出しておく。実施回ごとの一貫性について総合的に検討するには同じ統計指標を算出し、それを蓄積することが必要である。また、それを公開することが望ましい(「3.5 統計情報の公開」参照)。

(2) 統計指標の分析と総合

特定の実施回において受検者に偏りが生じたり、受検者数が少なかつたりすると、統計指標そのものが不安定になることがある。その場合、複数の実施回で得られたデータを合併して、単一の統計指標を算出できるようにも配慮し、保存のフォーマットを統一しておくとうい。

分析結果が期待されていた結果と異なっていた場合、それが、受検者数が少ないために統計指標が安定しないのか、テストの設計や実施の段階に何らかの原因があって不都合が生じたのか、判断できるだけのデータを収集することが重要である。また、複数回実施されたテストの同じ種類の統計指標群を要約、統合し、より明確な結果を得るアプローチもある。それは、**メタ分析**といわれている。

3 テスト利用にかかわる周辺情報の重要性

テストの特徴を統計指標のみで表すことは難しい。テスト結果に影響を及ぼす測定誤差の要因は複雑に相互作用しており、単独で一定の影響を及ぼすとは限らない。また、指標化することのできない要素がテスト結果に対して無視しえない影響をもつこともある。統計指標に集約しきれない要因については、実施状況、利用状況などについて記述された蓄積記録が重要な役割をもつ。また、尺度化の方法や配点の根拠、標準化されたテストの場合には標準化の方法や規準集団などについても記録を残すべきである(「1.6 尺度化の方法」および

「1.7 尺度の標準化」参照)。共通尺度化を行った場合には、その考え方や方法についての記録を残すべきである(「1.8 複数の尺度得点の比較」および「1.9 複数のテストの共通尺度化」参照)。それらの記録は、新たなテストの開発やテストの改訂に際して有益な情報となる。

実施にかかわる周辺情報は、統計指標の適切な解釈の助けとなる。適切に記録された周辺情報の蓄積は、テストの実施と利用の改善につながる(「1.12 テストの改訂」, 「1.15 手引, 解説書の内容」および「4.2 テスト実施後の調査と検討」参照)。保存された情報は、短期的にはテストの開発過程における質問項目の改善に役立つ。長期的には類似のテスト開発を行い、別の条件で実施した結果と比較する際にも役立つ(「1.1 テストの基本設計」, 「1.2 測定内容の定義と構造化」および「1.3 質問項目の設計」参照)。

4 保存情報の活用

テスト結果がどのように利用され、どのような効果をもたらしたのか、事後の検討が可能な状態で保存されていれば、尺度得点の適切さを確認するのに役立つ(「1.11 尺度得点の適切さの確認」参照)。さらに、単独のテスト得点や尺度を用いたのか、複数のテスト得点や尺度を組み合わせたのか、複数のテストや尺度を用いた場合にはテストや尺度間の関係はどうだったのか、などの情報も保存する必要がある。保存情報の適切な分析は、テスト関係者の相互研鑽に役立ち、テストの利用方法の改善にも結びつく(「2.1 テストの選択」および「6.1 テスト関係者の責務」参照)。

テストデータの解析結果や利用事例は、開発者、頒布者、利用者それぞれが蓄積している場合もある。それらを有効に活用するためには互いに参照可能となる仕組みを確立する必要がある。その際、目的外利用とならないように注意し、また、個人情報漏洩しないよう万全の手立てを講じなければならない(「3.2 拡大解釈の防止」および「4.1 テスト結果の保管手続」参照)。蓄積された保存情報の活用は、公正なテスト利用を推進するためにも必要である(「6.2 テストの公平性」参照)。テストによる判断の誤りは、他の蓄積された関連情報を参照せずに、そのテストで得られたひとつの情報だけを手がかりに判断するために生じることが少なくないからである。

5章 コンピュータを利用したテスト

1章から4章まで、テストの開発から実施、採点、テスト結果の利用に関する重要事項について示してきた。これらの重要事項は序章で定義した通常のテストに普遍的に適用できる。一方、受検機会の拡大、時間や場所の制約からの解放、採点結果の即時報告、従来のテストでは実現が難しい特性の測定などを目的として登場してきた、情報通信技術を利用したコンピュータテストにおいては、特別に考慮すべき事項がある。

本章では、コンピュータテストと通常のテストとの特徴を比較し、コンピュータテストに特有の実施上の留意点について述べる。それらは、コンピュータテストの技術的な観点から3つに大別できる。

1つ目は、コンピュータテストの基本的特徴に関するものである。5.1に概略を示す。

2つ目は、コンピュータテストの実施環境やコンピュータの操作性や情報セキュリティなど、コンピュータやインターネットなどの利用に伴う実施管理における重要事項、さらには、コンピュータ操作や通信経路で生じるトラブルへの対応などについて述べている。5.2から5.5までの4つの条項に分けて記述する。

3つ目は、コンピュータテストと通常のテストを併用する場合の方針を示す。これは、5.6で述べる。

コンピュータテストは、利便性が向上するという利点もあるが、使用するコンピュータの性能など、テストの測定用具の環境に関する制約もあることから、利用にあたっては十分な検討が必要である。

5.1 コンピュータテストの基本的な性質

コンピュータを利用したテストにおいても、得られる尺度の基本的な特徴は通常のテストと変わるものではなく、その開発や結果の利用にお

いては通常のテストの場合と同様の努力や配慮が必要である。

1 コンピュータを利用したテストの特徴

「コンピュータを利用したテスト」は、単にテストをコンピュータで実施することのみではなく、コンピュータを用いてテストの作成を支援することや、テストの実施およびテストの結果データを統合的に管理することを意味している。

コンピュータを利用したテストは、CBT (Computer-Based Testing), IBT (Internet-Based Testing), WBT (Web-Based Testing), e テスティング, などさまざまな呼称がある。これらの用語は、それぞれにコンピュータシステムや実施方法に関する違いを反映している。本章では特別に区別する必要がない限り、すべてを総称して「コンピュータテスト」と呼ぶ。

(1) コンピュータテストの特徴

①コンピュータテストでは、即時採点が可能になる。

たとえば、多肢選択式項目におけるコンピュータテストでは、実施後、回答結果の採点はコンピュータプログラムにより即時実施され、採点結果を即時に示すことも可能である（「1.5 採点手続の設計」参照）。また、**モデル準拠得点**もすばやく計算できる。

②コンピュータテストで使用できる豊富なメディアにより、これまで筆記によるテストでは測定することができなかった受検者の特性を測定できるようになる。

たとえば、ビデオや音声を用いた質問項目によって、測定できる**特性**の範囲が大きく拡張した（「1.2 測定内容の定義と構造化」, 「1.3 質問項目の設計」および「1.4 回答方法の設計」参照）。

③障害がある受検者に対して高度に対応する機能をもつ。

たとえば、視覚障害者にはオーディオを用いたり、手が不自由な受検者にはキーボード以外の特殊な文字入力装置を用いたりすることができ、さらに、コンピュータの機能を複合的に活用することもできる。これらの機能により、障害者への対応が行いやすくなる（「2.6 障害者などへの配慮」参照）。

④受検者から得られるデータが回答結果だけでなく、これまで通常のテスト

II 基本条項の解説

では得られなかった，受検者の認知プロセスを解明するためのデータが得られるようになる。

たとえば，最終回答にいたるまでに受検者が示した途中経過の応答記録から，多くの質問項目で受検者が間違えやすいタイプの傾向や理解不足の個所などを知り，学習診断や教育指導に役立つ情報を得ることができる（eラーニングとの融合）（「1.11 尺度得点の適切さの確認」，「4.2 テスト実施後の調査と検討」および「4.4 テスト結果の再利用のための保管情報」参照）。

⑤コンピュータで受検者の特性を逐次測定しながら，受検者に最適な質問項目を出題する**適応型テスト**を実現することができる。

測定精度を下げずに出題する質問項目数を減らすことができたり，受検者ごとに出題する質問項目や出題順を変更できたりするなどの利点があり，試験時間の短縮やカンニングなどの不正行為を抑制する効果が期待できる（「1.1 テストの基本設計」，「1.6 尺度化の方法」，「1.9 複数のテストの共通尺度化」，「2.4 手順どおりの実施と不測の事態への対応」および「6.5 受検者の義務」参照）。

(2) コンピュータテストの種類

コンピュータテストは，テスト実施中に他のコンピュータとの通信を行っているかどうかにより，以下のように分類することができる（「1.1 テストの基本設計」参照）。

①テスト実施中に通信を行う。

テストの実施中に受検者の使用するコンピュータが，サーバなどの別のコンピュータとインターネットやLANなどの通信回線を介して質問項目を表示したり，回答結果を送信するなどしたりして，相互に逐次データ通信をしながらテストを実施する。

②テスト実施中は，通信を行わない。

テスト実施にあたり，LAN等のネットワークを介してサーバ等の別のコンピュータから質問項目をまとめてダウンロードし，終了後に回答結果を送信する。もしくは，最初から受検者の使用するコンピュータ上に質問項目が格納されており，回答結果もそのコンピュータに記録されるなど，テスト実施中には他のコンピュータとは通信しない。

上記の①と②を比較すると、①は受検者に応じたテスト内容や質問項目の選択が統合的に制御可能となる。テスト内容の柔軟性が高くなる一方で、通信、ネットワーク障害の影響を受けるリスクが相対的に高くなり、システムの安定性では一般的には②の方が優れている。また、②は質問項目や回答結果を受検者のコンピュータ上に記録するため、コンピュータ上での情報漏洩や改ざんへの対応が必要となる（「2.2 テスト用具の管理」、「4.1 テスト結果の保管手続」および「5.4 機密保持や実施管理への対策」参照）。

それぞれ利点と欠点が存在するため、開発者は、実施するテストの目的や性質、受検者の利便性を配慮し、適切な方法を選択する必要がある。

そのほか、インターネットなどを利用したeラーニングにおいて、自宅で学習結果を確認したり、企業の採用選考において、**適性テスト**を事前に実施し、その後面接で利用したりする場合とか、適性テストを自宅などからインターネットを介して実施することも行われている。実施場所や時間の自由度が高まるため受検者の利便性は高くなるものの、技術的に本人認証を行うことが難しいことを理解した上で実施する必要がある（「5.4 機密保持や実施管理への対策」、「6.2 テストの公平性」および「6.5 受検者の義務」参照）。

コンピュータテストでは、受検者の使用するコンピュータの仕様がテスト結果に影響を及ぼす要因になることが想定されるため、開発者は、コンピュータに要求される性能などの仕様を明示すべきである（「5.2 コンピュータや実施環境の仕様」参照）。また、受検者のコンピュータに関する知識の程度がテスト結果に影響を及ぼすことも想定されるため、開発者、実施者は、テストの仕様や実施環境に留意する必要がある（「1.2 測定内容の定義と構造化」、「1.4 回答方法の設計」、「2.5 回答方法の説明」および「5.3 テスト画面や操作方法のわかりやすさ」参照）。

2 通常のテストと同様の努力や配慮の必要性

コンピュータテストにはさまざまな種類があり、その形態によって配慮すべき点に違いが生じる。しかし、テストの回答から得点および**尺度得点**を求め、結果を利用する点では通常のテストと同様であり、テスト開発と頒布、テスト実施と採点、結果の利用、記録と保管において、通常のテストの場合と同様な

II 基本条項の解説

努力や配慮が必要である。

3 コンピュータテスト方式とそのテスト開発、実施への影響

コンピュータテストにおいては、受検者の回答に基づいて次に出題される質問項目を自動的に選択する適応型テスト、音声や動画などのマルチメディアを使用したテスト、コンピュータソフトウェアの擬似環境を設けて、受検者に操作を行わせる**シミュレーション型テスト**、実際のアプリケーションソフトウェアなどを使って受検者の技能を測定する**パフォーマンステスト**など、さまざまなテスト方式が出現しつつある。これらのテストにおいては、通常のテストとは異なる実践技術が必要となる場合もあり、研究者や開発者によるさらなる研究が期待される（「1.1 テストの基本設計」および「1.3 質問項目の設計」参照）。

コンピュータを利用したテストの開発においては、質問項目の開発以外に、提示方法および回答方法をコントロールするためのソフトウェアあるいはプログラムを開発するか、あらかじめ作成されたソフトウェアを利用することになる。そのために、研究者および開発者は、提示方法や回答方法の違いによる回答への影響を研究し、結果の利用についても適切な方法を提示することが必要である（「1.4 回答方法の設計」, 「2.3 実施前の受検者への説明」, 「2.5 回答方法の説明」, 「3.5 統計情報の公開」および「4.2 テスト実施後の調査と検討」参照）。

また、テスト開発と同時にソフトウェアを開発する場合には、仕様どおりに動作することを確認するために十分なシステム検証が必要である。

5.2 コンピュータや実施環境の仕様

開発者は、使用するコンピュータや実施環境について必要な仕様を明示し、その条件を満たす質問項目を作成する。実施者は、開発者から示された仕様を満たすコンピュータや実施環境を用意した上で実施する。

1 テスト実施に必要なコンピュータに関する仕様の明示

開発者は、テスト実施用のソフトウェアが機能的、性能的に問題なく稼動するような受検用コンピュータのハードウェアおよびオペレーティングシステムなどの最低限の仕様を規定し、実施者および受検者に提示する必要がある（「2.3 実施前の受検者への説明」参照）。可能であれば、最低限の仕様とともに推奨する仕様を提示することが望ましい。とくに、音声や動画像などのマルチメディアを使用する場合には、通信環境やコンピュータの処理能力の影響が大きいので、推奨環境を明示することが必要である（「1.15 手引、解説書の内容」および「5.1 コンピュータテストの基本的な性質」参照）。

規定すべき受検用コンピュータおよび利用環境の仕様として、以下のものがある。

(1) 受検用コンピュータなどの仕様

- ① CPUの種類と性能（クロック数）
- ② オペレーティングシステム（OS）の種類とバージョン
- ③ メインメモリの容量
- ④ ハードディスクの空き容量
- ⑤ モニタディスプレイのサイズ、解像度（画素数）および色階調
- ⑥ 必要な入力機器（キーボードの種類、マウスの必要性など）
- ⑦ その他の必要な周辺機器の仕様
- ⑧ テスト実施に必要なとなるソフトウェアの種類とバージョン

(2) コンピュータの利用環境

- ① サーバの性能
- ② ネットワーク接続を必要とする場合は、ネットワーク接続の種類と性能、通信プロトコルと通信速度
- ③ その他、使用するコンピュータに関するハードウェアおよびソフトウェア、セキュリティの条件

2 実施環境について必要な仕様を明示する

開発者は、受検者に対するテスト実施環境の公平さを確保するために、コンピュータテスト利用の前提条件を明確にし、必要となる実施環境を明示する必

II 基本条項の解説

要がある。また、受検者が保有するコンピュータを用いて受検することが認められる場合を除き、テスト実施のためのすべてのコンピュータは同一の仕様であることが望ましい（「6.2 テストの公平性」参照）。

規定すべき実施環境としては、以下のものがある（「2.4 手順どおりの実施と不測の事態への対処」参照）。

- ①受検者のテスト実施場所（受検会場か、学校や会社の自席か、自宅か）
- ②使用するコンピュータなどの設備
 - コンピュータや機の配置、机の大きさ、机間の間隔
 - 机間の目隠し（パーティション）の必要性および仕様（高さなど）
 - テスト実施ソフトウェアの**インストール**や設定方法など
- ③通信およびネットワーク環境（「5.4 機密保持や実施管理への対策」参照）
 - テスト実施中に逐次データ通信を行う場合は十分な帯域のあるネットワーク回線を使用する
- ④テストの実施方法（「2.3 実施前の受検者への説明」および「2.4 手順どおりの実施と不測の事態への対処」参照）
 - 受検者の監督方法（たとえば、監督者の試験室内の巡回による監視や、TV モニタでの監視など）
 - テスト開始の方法（一斉に開始するか、随時開始してよいか）
 - 受検者の本人確認およびテスト実施システムでの**認証方法**
 - テストのために持ち込み可能なもの、あるいは使用禁止のものの有無（たとえば、メモ用紙、ノート、辞書、計算器付き時計、電卓、携帯電話、テストのための参考資料など）
 - 音を使用する場合には、音量の調節方法に対する指示
- ⑤障害者対応の優遇措置が設けられている場合には、その優遇措置の内容（「2.6 障害者などへの配慮」参照）

3 要求仕様を満たすコンピュータや実施環境の用意

開発者は、規定されたコンピュータや実施環境の制約条件を満たすように質問項目を開発する必要がある。しかし、なんらかの理由によって、開発者が規

定する条件を満たしていないコンピュータや実施環境でテストが実施された場合、コンピュータの性能の低さによる動作の不安定さ、表示画像のわかりにくさ、音声の聞きにくさなど、測定対象とする特性とは別の要因から、テスト結果が大きく影響を受ける可能性がある（「1.2 測定内容の定義と構造化」、「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。したがって、実施者は、開発者から示された仕様を満たすコンピュータや実施環境を用意し、テストを実施する必要がある（「2.2 テスト用具の管理」および「2.4 手順どおりの実施と不測の事態への対処」参照）。

5.3 テスト画面や操作方法のわかりやすさ

開発者は、テスト画面や操作方法を受検者にわかりやすいものとし、例題や練習問題を用意するなど、テスト画面や操作方法に慣れる機会を設けるようにする。また、受検者の誤操作があってもテストの実施に支障が生じないような設計しておくべきである。

1 テスト画面や操作方法をわかりやすいものとする

受検者がテストに集中することができ、テスト結果が受検者の特性以外の要因にできるだけ影響されないようにするためにも、テスト画面や操作方法はわかりやすいものとする（「1.2 測定内容の定義と構造化」および「1.4 回答方法の設計」参照）。

受検者にわかりやすい画面や操作を提供するための指針としては、**ユーザビリティ**（使い勝手、使いやすさ）、**アクセシビリティ**（障害者や高齢者など、さまざまな人々が利用しやすくする思想および利用しやすさの度合い）、**ユニバーサルデザイン**（障害の有無、年齢、性別、国籍、人種などにかかわらず多様な人々が気持ちよく使えるように設計すること）に留意する。とくに、開発しようとしているテストの受検者として、幼児や児童、高齢者、障害者が想定される場合は、受検者の身体的あるいは心理的特性を十分に配慮する必要がある（「1.1 テストの基本設計」参照）。テスト画面や操作性のわかりやすさを実現するためには、以下の項目に配慮する（「5.2 コンピュータや実施環境の

Ⅱ 基本条項の解説

仕様」参照)。

- ①テスト画面のレイアウト，操作ボタンの表示順序を質問項目に依存せず，統一する。また，その操作性は，一般的な Web サイトやアプリケーションソフトウェアのものとなるべく同様であることが望ましい。
- ②質問項目の選択や回答は，マウスやキーボードなどの限られた操作だけで行う。
- ③マウスの利用が困難である受検者に対してキーボードショートカットなどでの選択を可能とする。
- ④コンピュータリテラシーの差がテスト結果に影響を及ぼさないような工夫をする（「1.2 測定内容の定義と構造化」参照）。
- ⑤読み取りに支障のない文字の大きさを設定し，画面などの配色は色弱者のことを考慮して決める。
- ⑥質問項目が複数画面にわたる場合には，質問項目の続きがある旨を表示する。または，質問項目をすべて参照した上で回答させるなどの工夫をする（「1.4 回答方法の設計」参照）。
- ⑦質問文と参考資料（図表など）を参照し，回答する必要がある場合には，質問文と参考資料（図表など）が同一ウィンドウまたは同一画面内の異なるウィンドウ上で同時に参照できるようにする。同時に表示不可能な場合には，できる限り迅速に切り替え，参照できるようにする。
- ⑧回答の見直しが許可されている場合には，回答状況の一覧参照画面や回答の見直し方法を提供する。可能である場合には，操作方法のヘルプ機能を用意するか，操作方法を説明する資料を提供する（「1.15 手引，解説書の内容」参照）。
- ⑨テストの終了など，前の状態に戻ることができないような操作を必要とする場合には，注意を促す表示をし，受検者に確認させるようにする（「2.5 回答方法の説明」参照）。

上記の事項を検討する際には，受検対象者の属性を考慮することが必要である。たとえば，受検者が幼児や低学年の児童である場合には，画面や文字の配色，表示対象の点滅間隔などに留意するなどである。

2 例題や練習問題の用意

コンピュータテストに慣れるために、例題や練習問題を事前に提供し、受検者に体験させる機会を設けることが望ましい（「2.5 回答方法の説明」参照）。事前の提供が困難である場合には、テスト実施前の説明・練習時間を設け、受検者に実施させるか、テストの画面配置や操作方法を説明した用紙を配布し、テスト実施中に参照できるようにすることが望ましい。

3 受検者の誤操作の防止

受検者が誤って関係のないマウス操作やキーボード操作を行ったとしてもテストが中断したり、元の状態に戻せなくなったりするような事態を防止すべきである（「1.1 テストの基本設計」, 「1.4 回答方法の設計」, 「1.5 採点手続の設計」および「2.2 テスト用具の管理」参照）。具体的には、以下のような事項を検討する必要がある。

- ①テスト実施システムにおいて、テスト実施中に使用しないキーボード操作やマウス操作が発生しても、操作を無効にして影響を及ぼさないようにする。
- ②文字入力させる個所では、入力文字に対応した操作の制限を行うことが望ましい。たとえば半角英数字のみの入力であれば日本語入力プロセッサを起動しないようにするなどである。
- ③キーボードに不慣れな受検者が多いとき、入力文字数が多くなければ、画面上にソフトウェアキーボードを用意し、マウス操作で文字を選択させる方法を採用することも考えられる。
- ④キーボードを全く使用しない場合には、あらかじめ、受検者が操作できないようにキーボードを取り去る、または、操作できないようにカバーを設けるなどの措置をとる。
- ⑤不要なキーボード操作やマウス操作を無効にできない場合には、誤ってこれらのキーボード操作やマウス操作をした場合の解除方法を説明する資料を用意する。

5.4 機密保持や実施管理への対策

コンピュータテストでは、質問項目やテスト結果の機密を保持するためのシステムや通信上の対策を講じる。また、インターネットを利用するテストの場合には、機密保持のほかに、不正行為の防止や本人確認の方法について、とくに配慮が必要とされる。

1 質問項目やテスト結果の機密保持の必要性

コンピュータテストでは、質問項目やテスト結果はコンピュータ上に置かれて利用される。とくに実施中に通信を行わない場合は、テスト終了後もコンピュータ上に記録されることがある。また、ネットワークを用いて質問項目や回答などのテスト結果を送受信するようなテスト実施システムでは、利便性のためにインターネットが利用されることが多い。コンピュータやインターネットからの情報漏洩は重大な問題となっており、テストにおいても例外ではない。とくに、通信およびネットワークを利用する場合には、受検者が特定できる個人情報に関して**情報セキュリティ**に留意しなければならない（「1.13 質問項目の内容開示の是非」，「4.1 テスト結果の保管手続」，「5.1 コンピュータテストの基本的な性質」および「6.4 著作権の尊重」参照）。

2 質問項目やテスト結果の機密保持の対策

質問項目やテスト結果の機密保持の対策としては、以下のものがある。

(1) ネットワーク上での情報漏洩や改ざんへの対策

インターネットを利用する場合には、**通信経路の暗号化**を行う必要がある。**SSL**（Secure Socket Layer）による通信経路の暗号化が広く利用されている。資格試験や認定試験など、より高いセキュリティを求める場合には、送受信するファイル自体を暗号化することが望ましい。また、**電子署名**を採用することも有効である。

(2) コンピュータ上での情報漏洩や改ざんへの対策

質問項目やテスト結果を格納するコンピュータ（たとえばサーバ）を、受検者や不特定多数の人がアクセスできない場所（マシンルームや施錠されたコン

ピュータラックなど)に格納し、物理的にアクセスできないようにすることが望ましい。受検者が使用するコンピュータ上に質問項目やテスト結果を置く場合には、それぞれのテストが要求するセキュリティレベルに対応して、質問項目やテスト結果を受検者からは簡単に読めないようなファイル形式にしたり、テスト実施ソフトウェアが動作している間は受検者が他のアプリケーションに切り替えられないようにしたりする。さらに、質問項目やテスト結果のファイル自体を暗号化したりするなどの対策を講ずる必要がある(「2.2 テスト用具の管理」,「4.1 テスト結果の保管手続」および「4.4 テスト結果の再利用のための保管情報」参照)。

(3) 運用におけるセキュリティ確保

コンピュータテストでの機密保持は、システムや通信経路のセキュリティだけでは不十分であり、運用におけるセキュリティ指針の策定と実施が必要である。

たとえば、テスト実施システム(たとえばサーバ)の設置場所への入退室や施錠などの管理方法を明確にしたり、受検者が使用するコンピュータのテスト実施準備の手順を明確にして受検者の操作を限定したりするなどがある(「2.2 テスト用具の管理」および「4.1 テスト結果の保管手続」参照)。

(4) 法的な措置などの導入

質問項目やテスト結果に関して、より厳しい機密保持を必要とする場合には、システム管理者、運用者および監督者との間で機密保持契約を締結したり、コンピュータテスト実施に関する運用研修を実施し、許可された運用者のみがコンピュータテストの管理、運用に携わるようにしたりするなどの対応が必要である(「6.1 テスト関係者の責務」参照)。また、受検者に対しても、テスト実施前に質問項目やテスト結果に関する機密保持契約を提示し、同意した場合のみ受検を許可するようにすべきである(「6.5 受検者の義務」参照)。

3 インターネットを利用する場合の注意点

インターネットを利用するコンピュータテストでは、受検者の便宜性のために、受検者の在籍する学校や企業にあるコンピュータや、自宅のコンピュータを使用して受検するテストがある。この場合、監督者がいない状態で、テスト

II 基本条項の解説

が実施されることが多い。

これらのテストにおいて、テスト実施システムへの認証（ログイン ID またはパスワードでの認証や IC カードでの認証など）により、ある程度の本人認証（主体者認証）を行うことはできる。しかし、受検者本人が受検していること、受検者が不正行為をせず、正しく受検していることを技術的に完全に保証するものではないことに留意しなければならない。また、テストに対する回答についても監督者のもとでのテストと比較して、十分に制御されたものでなく、回答の信憑性についても必ずしも保証されないことに留意しなければならない（「5.1 コンピュータテストの基本的な性質」、 「6.2 テストの公平性」 および 「6.5 受検者の義務」 参照）。したがって、本人による適正な受検を担保する必要がある試験の場合には、インターネットを利用するテストにおいても監督者の立ち会いのもとで実施すべきである。

一方、受検者本人の実力を把握するためのテスト、受検者への利便性提供が実施や結果の厳密さよりも優先されるテスト、監督者のもとでの二次テストが別途用意されているテストなどの場合、監督者のいない状態でテストを実施することになるが、そのテストの長所と短所を十分に把握した上で利用を検討する必要がある（「1.1 テストの基本設計」、 「1.2 測定内容の定義と構造化」 および 「1.13 質問項目の内容開示の是非」 参照）。

5.5 不測の事態への対応

開発者は、コンピュータテストシステムをできる限りシステム障害に強いものにするとともに、障害が発生しても速やかに復旧、再開できるような対策を講じておくべきである。

1 システム障害に強い仕組み

受検者がコンピュータテストを安心して利用できるようにするためには、万全な対策が必要である。たとえば、**システム障害**に対する以下のような対応を検討する（「2.4 手順どおりの実施と不測の事態への対処」、 「5.1 コンピュータテストの基本的な性質」 および 「5.2 コンピュータや実施環境の仕様」

参照)。

- ①ハードディスクやネットワークの性能条件に依存しすぎないようなシステムとする。たとえば、容量の大きなデータを通信する場合には、テスト実施前にダウンロードするか、容量の小さなデータに分割して通信するなどの方法を採用する。
- ②テスト実施に無関係なキーボード操作やマウス操作を無効にし、誤操作などによる障害のおそれを減らす（「5.3 テスト画面や操作方法のわかりやすさ」参照）。
- ③一時的な通信、処理エラーに関して再試行（エラーリトライ）を行うなど、耐障害性に優れたシステムとする。
- ④必要に応じて、コンピュータの二重化やハードディスクなどの装置の冗長化を行い、耐故障性に優れたものとする。

2 障害発生時の速やかな復旧、再開のための対策

テスト実施システムに障害が発生したとしても、できる限り速やかに障害から復旧するような対策を講ずる。そのための対策として、たとえば、以下ののような事項を検討する。

- ①システムの復旧、再開後に、障害発生時以前に回答した情報が保存され、残りの質問項目からテストを再開できるようにする。
- ②受検者が利用できる予備のコンピュータを準備し、障害時に代替機として利用できるようにする（「2.2 テスト用具の管理」参照）。
- ③想定されるシステム障害の内容および対応方法を記述した運用の手引を準備する。システム管理者、運用者および監督者は、運用の手引にしたがって速やかにシステム復旧、再開の手順を実施できるようにする（「1.15 手引、解説書の内容」参照）。
- ④復旧、再開が不可能な場合に備えて、代替のテストの実施方法についても規定しておき、対応できるようにする。

5.6 コンピュータテストと通常のテストの併用

同じ特性を測定するために、コンピュータテストと通常のテストが併用される場合には、得点の比較可能性を検証し、換算表などを用意する。検証ができていない場合や換算が難しい場合は別のテストとして位置づけるべきである。

1 コンピュータテストと通常のテストの相違点

受検機会の拡大や結果報告の即時性をねらいとして、コンピュータを利用したテストが増えていく傾向にある。この場合、通常のテストで使用していた項目を使って、コンピュータテストを実施することがある。同じ質問項目を使用したとしても、コンピュータテストでは、問題文の提示および回答手続が通常テストとは異なるため、テスト結果に影響を及ぼす可能性がある（「1.2 測定内容の定義と構造化」、 「1.3 質問項目の設計」 および 「1.4 回答方法の設計」 参照）。たとえば、テスト画面の見やすさ、回答入力などの操作性、システムの応答速度、受検者のコンピュータに対する知識、技能の有無などが異なるような場合である（「2.5 回答方法の説明」、 「5.1 コンピュータテストの基本的な性質」、 「5.2 コンピュータや実施環境の仕様」 および 「5.3 テスト画面や操作方法のわかりやすさ」 参照）。

コンピュータテストと通常のテストの形態でテストを実施し、両者の得点を統一的に評価する必要がある場合には、コンピュータの使用が得点に有意な影響を与えないことを検証するか、影響がある場合には、換算表などを使用し、両者の受検者のいずれも不利益を被ることがないように考慮すべきである（「1.6 尺度化の方法」、 「1.8 複数の尺度得点の比較」、 「1.9 複数のテストの共通尺度化」、 「1.11 尺度得点の適切さの確認」、 「3.5 統計情報の公開」、 「4.2 テスト実施後の調査と検討」 および 「6.2 テストの公平性」 参照）。

2 得点の比較可能性の検証

得点の比較可能性を検証するためには、コンピュータテストと通常のテストにおいて以下のような方法でテスト結果を比較する。

(1) 相関関係の確認

2つのテスト得点間における相関関係を分析する。これらには、高い正の相関関係が必要である（「1.10 尺度得点の確からしさの推定と公開」）。

(2) 換算表の用意

高い相関関係が確認され、2つのテスト得点が比較可能と判断された後に、両者のテスト得点を同一尺度上で比較できるように換算表を用意する（「1.9 複数のテストの共通尺度化」参照）。

たとえば、テストの順位を2つのテスト間で変えないように、等パーセントイル法などを用いて換算表をつくる。また、コンピュータテストの得点を通常のテスト得点に換算し、通常のテストでの評価基準を適用することも可能となる（「1.6 尺度化の方法」参照）。

3 検証が困難な場合の対応

比較可能性の検証ができない場合や換算が困難な場合には、同じテスト内容であったとしても、コンピュータテストと通常のテストとは別のテストとして扱うべきであり、単純に得点を比較するべきではない。受検者に対しても、別のテストであることや両方の得点を比較してはいけないことを説明する必要がある（「1.1 テストの基本設計」, 「2.3 実施前の受検者への説明」および「3.2 拡大解釈の防止」参照）。

6章 テスト関係者の責任と倫理

本規準の0章から5章までの記述でも明らかなように、テストは開発から研究にいたるまでのさまざまな役割の人が連携して初めて正常に機能する営みである。本章では、テストにかかわるすべての人たちが有機的に連携し、テストをより機能的に活用するために、開発者、頒布者、実施者、利用者、管理者、受検者および研究者といった相互に異なる立場でテストにかかわる関係者の責任と倫理に関する事項を5つの条項に分けて記述している。

6.1では開発者、頒布者、実施者、利用者、管理者および研究者を含むテスト関係者の相互理解と責任、研鑽の必要性について記述している。時代や環境による社会的条件の変化、テストにかかわる研究の進展などにより、ある時期に通用した認識が時代とともに移り変わる可能性は十分ある。テストにかかわる関係者は変化を敏感に察知し、機敏に対応することが求められる。6.2はテストの公平性を保つために開発者、実施者、利用者および管理者が留意すべき事項について記述されている。6.3では志願者、受検者がテストに関して疑義を抱いた場合の対応について、主として実施者の責任とは何か、を記述したものである。

6.4では著作権とのかかわりについての記述である。テストに対する既製の著作物の利用および著作物としてのテストの扱いについて、主として開発者、実施者、利用者、受検者の責任について記述している。テストを取り巻く環境という意味では、著作権に関する考え方は、現状では最も急激に変化しつつあることがらのひとつである。

6.5は受検者の倫理についての記述である。

テスト関係者の主張や利害は、ときにその立場によって矛盾する場合がある。本章の5つの条項は、その問題を認識しつつも、すべてのテスト関係者が相互の利益を侵害することなく円滑にテストを活用するために、大切だと考えられる。

6.1 テスト関係者の責務

テストにかかわるすべての人は、開発、頒布、実施、さらに利用による影響など、テストの過程全体のあり方について相互に研鑽し、識見を深め、それを一般に周知するよう努める。

1 テスト関係者の相互研鑽と情報の共有の重要性

テストは測定しようとする**特性**に対する具体的な測定方法として質問項目が設計され、それらの質問項目によってテストが構成される。テストが活用されるまでには、テスト全体の設計、質問項目の原案作成、**予備テスト**を含む検討、編集、頒布、実施、採点、利用および利用後の検証などの過程に加え、それらの過程の制御、調整および管理が必要となる（「1.1 テストの基本設計」、「1.2 測定内容の定義と構造化」、「1.3 質問項目の設計」、「1.4 回答方法の設計」、「1.5 採点手続の設計」、「2.1 テストの選択」、「2.2 テスト用具の管理」、「2.4 手順どおりの実施と不測の事態への対処」、「2.8 採点手順の順守」、「3.3 判定基準の個別検証の必要性」、「3.4 受検者への報告」、「4.1 テスト結果の保管手続」、「6.2 テストの公平性」および「6.3 受検者の疑義への誠実な対応」参照）。すべての過程が有効に機能して、初めて成果が得られる。各過程での情報がつながりをもって有機的に集積されることで、テストの有用性が高められていく。テストは複数の人々がかかわる一連の過程として成立する体系的システムである。

したがって、テスト関係者には、それぞれの過程の中で各自が責務を全うするように努めるとともに、他の過程にも関心を持ち、必要な情報を適切に保存、公開、共有し、課題を相互検討する協働の姿勢と努力が望まれる（「3.5 統計情報の公開」および「4.4 テスト結果の再利用のための保管情報」参照）。

一般に、ある特定のテストで生じた問題は、類似した他のテストにおいても生じる可能性が高い。学術研究を利用目的とするテストや公共性の高いテストにおいては、各過程を分担する者が相互に連携し、研鑽に努めることが重要である。また、相互研鑽の中で得られたテストの改善に有益な情報は、関係者全体で共有されることが望まれる（「4.3 公的な性格が強いテスト結果の共同分

II 基本条項の解説

析利用」参照)。

とくに、テストの結果は**尺度得点**として表されるが、その適切な解釈は意外に難しい場合がある(「1.6 尺度化の方法」, 「1.7 尺度の標準化」, 「1.8 複数の尺度得点の比較」, 「1.9 複数のテストの共通尺度化」, 「1.10 尺度得点の確からしさの推定と公開」, 「1.11 尺度得点の適切さの確認」, 「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照)。尺度得点にかかわる事項の適切な理解がテストの誤用を防ぎ、テストの結果を有効に活用するための要である。

2 テストの各過程におけるテスト関係者の責務

ここでテスト関係者と呼ばれる人たちには開発者(テストを設計・開発する者)、頒布者(提供・頒布する者)、実施者(実施し採点する者)、利用者(結果を利用する者)、管理者(結果の情報管理者)、そして研究者(テストについての研究者)などが含まれる。それぞれの人はテストの各過程で必要とされる役割を担い、そこで求められる責務を果たしていかなければならない。それにはどのようなものがあるかを以下に述べる。

(1) 開発者

開発者は、測定の技術的側面のみではなく、実施環境、採点過程、結果の利用の全過程に関心をもち、テスト結果の利用までの過程全体に配慮した上で、開発に臨む必要がある。実施および利用について各過程の関係者から情報を収集し、さらに利用後の検証に関する情報を収集してテストの改善に努めることが求められる(「4.2 テスト実施後の調査と検討」参照)。

(2) 頒布者

頒布者は、開発の目的、意図および測定しようとする特性のみならず、測定用具としてのテストがもつ性能を十分に理解し、それを実施者、利用者に的確に伝える役割を担っている。テストの各過程の状況に注目し、助言や指導を行いながらテストが適切な方法により円滑に実施、利用されるように心がける。また、必要に応じて関係者間での情報の共有を図り、テストの改良に参考となる情報を収集する(「1.12 テストの改訂」参照)。また、頒布者と開発者が相互に協力しながら適切な手引、解説書を作成し、頒布することも重要である

(「1.15 手引, 解説書の内容」参照)。

(3) 実施者

実施者は、手引や解説書に記載されたテストの基本設計を理解し、手順どおりに定められた役割を全うするとともに、実施および採点状況にかかわる内容でテストの利用に必要な情報を利用者に伝える(「1.15 手引, 解説書の内容」参照)。とくに、設計の意図や手順と、実施や採点状況との間に食い違いが見られた場合は、関係者への速やかな情報提供と実施責任者による適切な対処が求められる。また、必要に応じて受検者の反応や採点上の問題点など、テストの開発や改良に有益と思われる情報を開発者に提供することも大切である。

(4) 利用者

利用者は、適用が想定されている対象者の範囲やテストの性能を十分に理解し、テストを適切に利用する責任がある(「3.1 テストの趣旨や尺度の性質の理解」および「3.2 拡大解釈の防止」参照)。利用に際しての疑問や問題が生じた際には速やかに開発者または頒布者に相談し、状況に応じた適切な利用を心がける。また、利用上の課題や利用から得られたデータなどを開発者など他の過程を担当する人と共有し、相互に研鑽を積む努力が期待される。

(5) 管理者

管理者は、テストの実施結果として得られた情報について責任をもって維持、管理し、適切に利用できる状態を保つ責任がある。調査、分析に必要な情報は廃棄せずに保管する必要がある。一方、個人情報管理には十分な注意を払わなければならない(「3.5 統計情報の公開」, 「3.6 研究目的でのテスト結果の利用」, 「4.1 テスト結果の保管手続」, 「4.2 テスト実施後の調査と検討」および「4.4 テスト結果の再利用のための保管情報」)。

(6) 研究者

テストにかかわる研究者には、テストの仕組みおよび性質を理解した上で理論的な可能性と現実的な制約条件の双方を考慮しつつ、現実に開発、実施および利用されるテストの品質向上に貢献する研究成果をあげることが期待される。テストの実務にかかわる関係者に対し、研究成果に基づく有益な示唆を与えることが望ましい。

テストそのものを研究対象とする研究者は、テストの性質にかかわる基礎知

II 基本条項の解説

識が一般の人々に十分に理解されるように努めるべきである。そのための教育活動、啓蒙活動も研究者に期待される重要な役割である。限定された条件における研究成果のみを根拠としてテストの実務にかかわる関係者に実現不可能な要求をしてはならない。また、特定のテストの内容領域にかかわる研究者は、そのテストの基本設計のみならず、テストの基本的な仕組みおよび性質を十分に理解した上で、当該内容領域に関する専門的な見地からテストの改善に結びつく適切な助言を与えることが期待される。

3 テスト過程全体の制御および管理の必要性

以上のように、テストの各過程の担当者は、相互に情報を共有し協働して課題の検討にあたることが大切である。しかし、担当者が相互に面識のない場合も多い。したがって、とくに大規模に実施されるテストの場合、テストの実施過程全体を制御、調整および管理する役割を担う統括責任者をあらかじめ定めておくことが必要となる。統括責任者はテストに関して専門的知識、技能を有する者があたることが望ましい（「2.2 テスト用具の管理」および「2.4 手順どおりの実施と不測の事態への対処」参照）。

テスト過程全体の統括責任者を設けることが困難である場合は、開発者または頒布者がそうした役割を担えるように体制を整えておくことが望まれる。

6.2 テストの公平性

受検者は、テストのすべての過程において年齢、性、国籍、障害の有無などによって差別されてはならない。また、質問項目の表現は、特定集団の成員に不快感を生じさせないように配慮されなければならない。

1 受検資格における公平性

(1) 受検資格

資格試験や採用試験などでは、一般にその目的達成のために必要な受検者の要件が受検資格として定められている。受検資格を設ける場合には、公平性について十分に配慮した上でその要件を明確化し、問い合わせ先などとともに募

集案内、受検要項などに記載して周知する必要がある（「2.3 実施前の受検者への説明」参照）。また、テストには基本設計によって設定された対象者の範囲がある（「1.1 テストの基本設計」および「1.2 測定内容の定義と構造化」参照）。

(2) 募集、申し込み

受検にあたって申し込みを必要とするテストにおいては、受検希望者がそのテストの募集などに関する情報を得ることができること、そして、受検資格を有する者であれば、実際に問題なく申し込みができる環境が用意されていることが必要である。一部の者に対してだけ情報の提供が行われたり、特定の集団に情報が伝わらなかつたりするということがないように留意しなくてはならない。たとえば、募集、申し込みをインターネットで行う場合には、インターネットを使用できることが受検資格としての意味をもつ場合を除き、郵送等の代替となる方法による募集案内、受検要項などの送付および申し込みについても準備しておくことが必要である。また、一部の者に実施前に質問項目の内容が一部でも知られている状況では、テストが十分に機能しないことに留意しなければならない（「1.13 質問項目の内容開示の是非」、「6.4 著作権の尊重」および「6.5 受検者の義務」参照）。

2 実施方法の違いへの配慮

テストの回答方法や実施方法には、テストの目的、測定する特性、実施規模などによってさまざまなものがある。一般に、テストは開発段階で口述式にするか、**記述式**にするか、**選択式**にするかなど、質問項目や回答形式が検討され、その方法にそってテスト時間を含めた基本的な実施事項が定められる（「1.3 質問項目の設計」および「1.4 回答方法の設計」参照）。実施者は、基本設計にそってテストを実施する必要がある。その際、テストの目的、趣旨を歪めることのない範囲で、特定の受検者に対して実施方法による不利が生じないように配慮する（「2.4 手順どおりの実施と不測の事態への対処」参照）。

たとえば、大規模なテストにとって**マークシート（OMR）**方式や**OCR**方式による回答は欠かせないものとなっているが、それは大量のデータを正確かつ効率的に処理するために必要な過程であって、マークを塗りつぶしたり、指

Ⅱ 基本条項の解説

定された枠内に数字を記入したりする作業能力を測定するものではない。特定の受検者にとって回答方法による不利が生じないように、可能な範囲で代替のテスト方式を用意すべきである。また、代替法が選択可能となる条件などについて、十分に周知する必要がある（「2.6 障害者などへの配慮」および「2.7 少数受検者への配慮」参照）。テストにコンピュータを用いる場合には、テストに利用される機器やソフトに対する慣れが結果に対して影響しないように、テストの実施前に説明・練習時間を組み入れるなどの配慮が必要である（「5.2 コンピュータや実施環境の仕様」および「5.3 テスト画面や操作方法のわかりやすさ」参照）。

実施者は、受検場の環境などの実施条件の違いによって著しい有利、不利が生じないように心がける。また、受検者に対する監督者の発言事項については、実施のための手引の中に定めておくことが望ましい（「1.15 手引、解説書の内容」参照）。

面接などでは、高度に構造化された方式を除けば、事前に面接員のすべての発言事項や質問を準備することはできない。その場合、差別的表現や誤解を生じる発言がないように、面接員に対して十分に事前研修などを行うことが望ましい（「2.9 主観的な評定による採点」参照）。個人の能力や適性をみることが目的の面接では、思想、信条、収入等の人権にかかわる事項などについての質問を行ってはならない。さらに、テストの目的や趣旨に無関係な質問をすることや質問や評価とは関連のない資料を面接員が保持しないなど、無関係な質問や資料に基づいて評価がなされたという誤解を受けないようにすることが必要である。

3 質問項目の表現上の配慮

テストの質問項目などの表現に差別的な表現があってはならない。特定集団の成員に不快感を生じさせるような表現や無用な誤解を生じるおそれのある表現がないように配慮されなければならない。開発者は、質問項目などの内容に誤りのないようにすることはもちろんであるが、差別的な表現や不快感を生じさせる表現や用語がないように留意する。さらに、障害や疾病、文化的な背景の違いなどにより、特定の受検者が著しく不利にならないような措置を講じる

べきである（「2.6 障害者などへの配慮」および「2.7 少数受検者への配慮」参照）。

とくに、外国で開発されたテストの質問項目などを翻訳して使用したり、国内外の文献などを引用して質問項目などを作成したりするような場合、時代や社会的背景などの相違によって特定の国や集団に対する差別的表現や不快感を生じさせる表現が用いられていないか注意が必要である（「1.14 テストの他言語への翻訳」参照）。また、差別的表現ではなくとも、たとえば、「新聞を見る」という表現を「新聞を読む」という表現にするなど、障害がある人にとってあてはまらないような状況については、質問項目の趣旨を踏まえた上で他の表現に置き換える可能性を検討することが望ましい。

4 結果の取り扱い

テスト結果の利用において不適切な取り扱いがあってはならない。また、得点の算出方法、結果の利用法などについて、受検者が疑念を抱くことのないように、テストの利用目的に支障のない範囲でのテスト情報の開示を検討することが望ましい（「3.3 判定基準の個別検証の必要性」、「3.4 受検者への結果報告」、「3.5 統計情報の公開」および「6.3 受検者の疑義への誠実な対応」参照）。

6.3 受検者の疑義への誠実な対応

受検者からテスト実施方法、テスト結果や利用方法に関する疑義が提示された場合には、開発者、実施者および利用者はそれに対して誠実に説明するよう努める。

1 テストに対する疑義

テストの実施者および利用者は、受検者がテストの目的、実施方法、結果、結果の利用方法について十分な情報が得られるようにする（「2.3 実施前の受検者への説明」）。しかしながら、情報が適切に公開されている場合でも、テストの実施方法、結果や利用方法に関して受検者から疑義が提示されることがあ

II 基本条項の解説

る。その原因として、質問項目やテストの内容に誤りがある、説明資料に誤植などがある、用語が間違えて使用されている、実施方法に曖昧な点がある、説明不足、受検者に勘違いや誤解があるなどが考えられる。いずれの場合も、実施者および利用者には、受検者が理解できるように、支障を生じない範囲でいいいな説明が求められる。

2 疑義に対する誠実な対応

疑義への円滑な対応を図るため、あらかじめテストについての問い合わせ先を明示しておく（「2.3 実施前の受検者への説明」および「4.1 テスト結果の保管手続」参照）。実施者は、疑義が提示された場合における開発者、実施者および利用者間の連絡調整、情報の共有化を図るシステムを整備しておくなど、あらかじめ十分な危機管理体制を敷いておくべきである（「2.4 手順どおりの実施と不測の事態への対処」参照）。

疑義の内容によっては、テストの内容、実施方法、結果の処理などに広範な影響が生じる可能性もある。いかなる場合でも、誠実な対応は受検者のテストに対する理解を深め、不安を軽減し、結果的に実施を円滑化し、テストに対する信頼感を高める効果がある。また、誤解が生じやすい点に関する分析や疑義に対する説明を通じて、テスト関係者に見逃されていた問題が発見され、テストの改善に役立つ場合もある（「1.12 テストの改訂」参照）。

受検者の勘違いや誤解によって生じた疑義については、早期に回答して誤解を解く。

実施中にテストの内容に関する疑義が提示された場合には、実施者は指摘された内容について早急に検討しなければならない。検討の結果、指摘が妥当であり、改善が必要であると判断された場合には、その影響を勘案した上で早急に対応策を決定しなければならない。さらに、影響が及ぶすべての受検者に対して、可能な限り迅速に対応策について連絡する必要がある。疑義を提示した受検者だけに情報が伝わり、他の受検者に伝わらないことで、有利、不利が生じてはならない（「6.2 テストの公平性」参照）。

6.4 著作権の尊重

テストの関係者は、質問項目、手引や解説書、結果表示様式などの著作権を尊重し、適切に管理する。利用者は、これらを頒布者との契約に基づいて入手して利用しなければならず、許可なく著作物の複製やインターネットでの送信などを行ってはならない。

1 著作物としてのテスト

著作権法では、著作物は「思想又は感情を創作的に表現したものであつて、文芸、学術、美術又は音楽の範囲に属するもの」と規定され（第2条）、著作物として「小説、脚本、論文、講演その他の言語の著作物」、「音楽の著作物」、「地図又は学術的な性質を有する図面、図表、模型その他の図形の著作物」、「写真の著作物」などが例示されている（第10条）。また、編集物でその素材の選択または配列によって創作性を有するものは、著作物として保護される（第12条）。

一般に、開発されたテストは学術的なアイデアが創作的に表現されたものであり、質問項目集、図版、手引や解説書、評価票や結果票などは著作物と考えられる（「1.15 手引、解説書の内容」参照）。著作者の権利には、財産的利益を保護する著作権（財産権）と、人格的利益を保護する**著作者人格権**とがあるが、本条項でいう著作権はそのどちらも含めたものである。テストに関しても、著作権は尊重され、保護されなければならない。著作権法により例外的に著作物などの許諾なしで利用ができるとされている場合を除き、利用にあたり著作権者（著作権を有する者）の許諾が必要である。通常、開発者や頒布者との契約に基づき、そのテストの利用形態に応じて必要部数入手することなどが求められる（「1.13 質問項目の内容開示の是非」参照）。

2 テストにおける著作物の利用

質問項目の作成などのテストの開発にあたって、出版されたものやホームページに掲載されたものなど公表された著作物の一部を利用する場合がある（「1.3 質問項目の設計」参照）。著作権法では「入学試験その他の学識技能

II 基本条項の解説

に関する試験又は検定の目的上必要と認められる限度において、当該試験又は検定の問題として」公表された著作物を複製することが認められている（第36条）。また、テストに著作物を利用した場合、出所を明示する慣行があるときには、著作物の出所を明示しなければならないとされている。なお、外国の著作物をわが国で試験問題に利用する場合にもわが国の著作権法が適用される。

3 著作物利用の制限

「試験問題」であれば無制限に著作物を利用、複製してもよいわけではない。著作権者の利益を不当に害することとなる場合はこの限りではない。また、テストが営利を目的としている場合には、補償金を著作権者に支払うことが定められている。開発しようとするテストが「目的上必要と認められる限度」を越えて著作物を利用、複製する場合、著作権者の許諾が必要となる。許諾の条件は、特別の場合を除き、原則として著作権者、あるいは、著作物によっては著作権を集中的に管理している団体と著作物の利用者である開発者との交渉により決定される。

さらに、たとえば、国語の読解問題で質問項目に小説の一部を利用する場合のように、公表された著作物の一部を利用して開発されたテストには、この場合の小説の一部のような、テストに利用された著作物についての著作権、すなわち、原著作者の著作権がある。したがって、このようなテストには、テスト自体の著作権と原著権が存在することとなり、そのテストを複製したり、インターネットなどで送信したりする場合には、テストの著作権者と原著作者から許諾を得る必要がある。

また、著作者は、著作者人格権のひとつとして著作物の内容や題号を無断で改変されない権利である同一性保持権を有している。著作物を利用する場合には原作のまま利用することを原則とする。改変を行う場合には同一性保持権を侵害しないよう、利用の目的などに照らしてやむをえないと考えられるものにとどめるべきである。

6.5 受検者の義務

公正な実施と正確な測定を期するため、受検者は実施者の指示にしたがって受検し、不正行為や妨害行為をしてはならない。また、テスト問題の守秘義務に同意して受検した場合には、質問項目を許可なく公開したり、配布したりしてはならない。

1 実施者の指示の順守

テストは、特性の正確な測定を前提として利用される（「1.1 テストの基本設計」, 「1.2 測定内容の定義と構造化」, 「1.10 尺度得点の確からしさの推定と公開」および「1.11 尺度得点の適切さの確認」参照）。入学試験や採用試験、資格試験のようにテストの結果が受検者個人の処遇の決定に利用されるテストでは、テストが公正に実施され、テスト結果が適正に利用されなければならない。テストの公正な実施とテスト結果の適正な利用は、テストの過程が定められた手順どおりに進められて初めて確保される（「2.4 手順どおりの実施と不測の事態への対処」参照）。受検者は、実施者の指示にしたがわなければならない。カンニングやいわゆる替玉受検のような不正行為、他の受検者の受検環境を損なう妨害行為が許されないのはいままでもない。なお、実施者には不正行為や妨害行為が起こった場合には、手引にしたがって迅速に対応することが求められる（「1.15 手引、解説書の内容」および「2.4 手順どおりの実施と不測の事態への対処」参照）。

2 テストにおける守秘義務

テストの質問項目には、性格検査のように繰り返し使用されることを前提としているものや、**項目プール**にデータベースとして蓄積され、抽出されて複数回使用されるものもある。新たなテストの開発のために予備テストが行われる場合もある（「1.3 質問項目の設計」参照）。

毎回の実施において新規に項目が作成されるテストでは、公正な実施と特性の正確な測定を損なうおそれがないと判断される範囲内で、質問冊子の持ち帰りが認められ、質問項目が公開されている場合もある。しかし、質問項目の再

II 基本条項の解説

利用を前提として設計されたテストでは、質問項目の開示によってその後の実施に支障が生じる。そのため、質問冊子を含むテストにかかわる情報の持ち帰りが認められず非公開とされている場合もある（「1.13 質問項目の内容開示の是非」参照）。

質問項目などが非公開のテストでは、受検に際してテスト問題の守秘義務への同意を求めることがある（「4.1 テスト結果の保管手続」参照）。テスト問題の守秘義務に同意し受検した場合には、守秘義務に記述された内容を順守し、質問項目の内容を許可なく他人に伝達したり、ホームページなどに公開したり、配布したりしてはならない（「6.4 著作権の尊重」参照）。