

# テスト作成ハンドブック翻訳完成と その活用

---

東北大学教育情報学研究部・教育部教授  
村木英治

# 日本テスト学会

JART (The Japan Association for Research  
in Testing)

---

- 日本テスト学会は、現時代に即応したテスト技術に関する研究者を育成し、科学的基礎と教育的理念に裏付けられた新しいテスト法の考え方と評価技術の研究開発・実用化、産官学を交えた共同のインフラ作りと社会的コンセンサスの構築、国内外の評価技術関連情報の収集・提供を目的に2003年5月23日に設立された団体です。  
今後、年次大会、セミナー、ワークショップ等を通じて評価技術に関する研究交流・普及啓蒙活動を行うとともに、評価技術に関するガイドラインの作成や標準化活動支援等に取り組んでいきます。

# JART大会の歴史 I

---

- 2003年7月 日本テスト学会WEBサイトオープン
- 2003年8月 日本テスト学会 第1回大会開催
  - 東京大学
- 2004年8月 日本テスト学会 第2回大会開催
  - 東京工業大学
- 2005年8月 日本テスト学会 第3回大会開催
  - 早稲田大学: E-learningの高度化とテスト理論
- 2006年8月 日本テスト学会第4回大会開催
  - 東北大学: テストの科学と日本的風土

# JART大会の歴史 II

---

- 2007年8月 日本テスト学会 第5回大会開催
  - 聖路加看護大学：医療分野におけるテスト利用の現状と展開
- 2008年8月 日本テスト学会第6回大会開催
  - 成蹊大学：テストの科学とその果たす役割
- 2009年9月 日本テスト学会第7回大会開催
  - 名古屋大学：社会に貢献するテスト科学

# JART学会誌の歴史

---

- 2005年3月 日本テスト学会誌 (Japanese Journal for Research on Testing) 創刊
- 2009年3月 日本テスト学会誌Vol.5 発刊

テスト規準作成委員会委員長 繁柘算男

日本テスト学会編 テスト・スタンダード 日本のテストの将来に向けて 金子書房 2007

# Handbook of Test Development

---

- ❑ Edited by Steven M. Downing & Thomas M. Haladyna
- ❑ In 2006, Published by Lawrence Erlbaum Associates, Publishers
- ❑ 46 Authors (The total of 778 pages, 33 chapters)

# Handbook of Test Development

---

Steven M. Downing, Robert L. Linn, E. Roger Trent & Edward Roeber, Robert J. Mislevy & Michelle M. Riconscente, James C. Impara & David Foster, Linda Crocker, Michael Kane, Norman L. Webb, Mark Raymond & Sandra Neustel, Gregory J. Cizek, C. David Vale, Steven M. Downing, Catherine Welch, Stephen G. Sireci & April L. Zenisky, Rebecca A. Baranowski, Michael Zieky, Jamal Abedi, Ronald K. Hambleton & Stephen J. Jirka, Samuel A. Livingstone, Cathy LW Wendler & Michael E. Walker, Michael J. Young, Paul Jones & Russell W. Smith & Diane Talley, Gale H. Roid, Tim Davey & Mary J. Pitoniak, Richard M. Leucht, Dan Champion & Sherri Miller, Rose C. McCallin, Martha L. Thurlow & Sandra J. Thompson & Sherl S. Lazaus, Joseph M. Ryan, Douglas F. Becker & Mark R. Pomplun, Chad W. Buckendahl & Barbara S. Plake, Thomas H. Haladyna, Cynthia Board Schmeiser

産: Blue  
官: Green  
学: Black

# テスト作成ハンドブック

## 発達した最新技術と考え方による公平妥当なテスト 作成・実施・利用のすべて

---

□ 日本語版監訳 池田央

□ 2008年 教育測定研究所

□ 翻訳執筆者31名

安部哲弥、荒井清佳、池田央、岡部満帆子、韓太哲、  
喜岡恵子、小泉利恵、近藤悠介、佐々木典彰、  
佐藤喜一、佐野真、柴山直、清水裕子、菅原良、  
筒井英一郎、永井圭、中畝菜穂子、中野美知子、  
中村洋一、野上康子、林規生、廣瀬英子、藤田智子、  
松本佳穂子、村木英治、村田佳奈子、森昭彦、  
藪田由己子、吉村宰、渡邊誠一、渡辺光明



# 構成

---

1. 基礎	6章
2. 内容	4章
3. 問題作成	9章
4. テストのデザイン	6章
5. テストの製作と実施	3章
6. テスト実施後の活動	4章
7. エピローグ	1章

# 第1部 基礎

---

1. 効果的なテスト作成にいたる12のステップ
2. 教育・心理検査法のスタンダード:テスト開発の指針
3. テスティングサービスとの契約
4. 証拠中心の評価デザイン
5. テストの不正行為を最小限にするための問題項目とテストの作成方略
6. 受験者のためのテスト準備:テスト開発者と使用者へのガイドライン

## 第2部 内容

---

1. テスト開発における内容関連妥当性の証拠
2. 生徒用学力テストの内容確定
3. 資格認定試験内容の決定
4. 規準の設定

## 第3部 問題作成

---

1. コンピュータ化された項目バンキング
2. テスト作成における解答選択式問題の形式
3. パフォーマンステストの問題作成
4. コンピュータ式テストによる革新的な問題形式： 構成概念の代表性改善への探求
5. 問題の編集と再検討
6. テストによるアセスメントの公平性の審査
7. テスト問題作成における言語問題
8. アンカー基準法を基にした項目統計量の判断による推定
9. 項目分析

## 第4部 テストのデザイン

---

1. 大規模プログラムのための複数テストフォームの設計と維持に関する実際問題
2. 垂直尺度
3. 小規模学力テストシステムのためのテストフォームの作成
4. 能力テストのデザイン
5. コンピュータ適応型テストのデザイン
6. 項目応答理論を用いた合否決定のためのテストデザイン

## 第5部 テストの製作と実施

---

1. テスト作成が妥当性に及ぼす影響
2. テストの実施
3. 特別な必要性のある学生に対するテスト実施への配慮： 調整、修正、そしてさらなるもの

## 第6部 テスト実施後の活動

---

1. 生徒のテスト得点を報告する際の実際と課題、傾向
2. 技術報告および文書化
3. テストの評価
4. テスト作成における妥当性研究の役割と重要性

# 第7部 エピローグ & 付録

---

1. エピローグ
2. 一般用語索引



## 第1章

# 効果的なテスト作成に至る12のステップ

Steven M. Downing

---

- Test Standards(教育・心理検査法のスタンダード: AERA, APA, & NCME, 1999)との連携
  - テストの目的の明確な定義の重要性
  - 妥当性の強調と妥当性の証拠収集
    - 意図されたテスト得点の解釈が妥当かどうかは、テストシステムの技術的な品質に関連する可能なすべての証拠に依存している。これは注意深いテスト構築に向けての証拠の確保を含んでいる。
  - 報告書の作成
-

# 効果的なテスト作成に至る12のステップ

---

1. 全体計画
2. 内容の定義
3. テストの仕様
4. 問題作成
5. テストデザインと組み立て
6. テストの作成

# 効果的なテスト作成に至る12のステップ

---

7. テストの実施
8. テスト解答の採点
9. 合格点の設定
10. テスト結果の報告
11. 問題項目バンキング
12. テストの技術報告書

# 効果的なテスト作成に至る12のステップ

## 1. 全体計画

---

- 全テスト作成活動についての体系的ガイダンス
    - 構成概念
    - 望ましいテストの解釈
    - テスト形式
    - 妥当性を示す証拠の主な資料
    - 明確な目的
    - 望ましい推論
    - 心理測定モデル
    - 作成日程
    - 安全保持
    - 品質管理
-

## 効果的なテスト作成に至る12のステップ

### 2. 内容の定義

---

- 内容領域の明確で詳細な定義とその描写、  
そして内容定義の方法
  - 内容領域/母集団からの標本抽出計画
  - 評価目的に関連した諸方法
  - 内容関連妥当性の証拠を示す基礎資料
  - 構成概念の描写

効果的なテスト作成に至る12のステップ

### 3. テストの仕様 (test specifications) : テストの青写真 (test blueprint) 作り

---

- テスト仕様は問題、課題あるいは質問の形式、解答の形式、あるいは解答に当たっての条件、また採点手続きを詳述すること。
  - 内容の操作的定義
    - テスト形式の型(解答選択型、解答構築型、作業式)、テスト問題の総数、認知様式の型、採点ルール、解答制限時間
    - 主なる内容領域と従なる内容領域に配分されるテスト問題の数とその割合
  - 内容領域の抽出に関連して、体系的に説明できる妥当性の証拠を示す枠組み
  - 集団規準か目標規準か？
  - 望ましい項目特性は？
-

## 効果的なテスト作成に至る12のステップ

### 4. 問題作成

---

- 効果的な質問の作成
- テスト形式
- 証拠中心の原理に忠実な妥当性の証拠
- 問題執筆者と検閲者の訓練
- 効果的な問題編集
- 問題の不備からConstruct-irrelevance variance (CIV)

# 効果なテスト作成に至る12のステップ

## 5. テストデザインと組み立て

---

- テスト形式のデザインと創作
- 特定形式のテスト問題の選択
- 計画された青写真による抽出作業
- 事前テストの配慮
- 組み立て
  - 選択枝と正解位置のバランス
  - アンカー項目や共通項目の配置



# 効果なテスト作成に至る12のステップ

## 6. テストの製作

---

- 出版活動
- 印刷またはCBT化
- 安全保持問題
- 品質管理についての妥当性問題

## 効果なテスト作成に至る12のステップ

### 7. テストの実施

---

- 標準化に関する妥当性問題
- 米国障害者法(ADA)問題
- 試験監督
- 安全性問題
- 時間設定問題

効果なテスト作成に至る12のステップ

## 8. テスト解答の採点(test scoring)

---

解答そのものは測定ではない。むしろある採点ルールやアルゴリズムの応用、あるいは解答に対する規定(rubic)の適用であり、それが測定値を生むことになる。

- 妥当性の議論
- 品質管理
- 正解の妥当性確認
- 項目分析

## 効果なテスト作成に至る12のステップ

### 9. 合格点の設定

---

- 説明できる合格点の設定
- 相対基準か絶対標準か？
- 分割点の妥当性問題
- 規準の比較可能性
- 得点尺度の恒常性の維持(等化、リンク付け)

## 効果なテスト作成に至る12のステップ

# 10. テスト結果の報告

---

- 妥当性の議論
  - 正確さ、品質管理
  - 時期の適切さ
  - 意義
  - 誤用問題
  - チャレンジ出願
  - 再受験

効果なテスト作成に至る12のステップ

## 11. 問題項目バンキング

---

- 安全性問題
- 有用性、柔軟性
- 効果的な項目バンキングの原理

効果なテスト作成に至る12のステップ

## 12. テストの技術報告書

---

- 体系的で綿密に記録された妥当性の証拠
- 12ステップの構成
- 推奨できること

テストの妥当性の証拠を最大にすること

# ハンドブックの特色と意義

---

- テスト業務実務の全般にわたる具体的かつ実  
際的な情報
- アメリカ版テストスタンダードとの連携
- 内容妥当性の強調
- 証拠中心の評価デザイン(Evidence-  
centered design: ECD)
- コンピュータ版テストに関する章の充実



# ハンドブックのこれからの活用

---

## 検定試験の評価

	資格	任意
国家	医師国家試験 <del>看護師国家試験</del> 薬剤師国家試験 歯科国家試験 美容師国家試験 管理栄養士国家試験	
民間	税理士資格試験 行政書士資格試験 保育士資格試験 日本語教師資格試験 医療事務資格試験	実用英語検定 漢字検定 秘書検定 コンピュータ技能各種検定 簿記検定 ご当地検定

# テスト

---

- テストとは測定の実行である。
- Psychometricianは測定技術の専門家である。
- テスト事業
  - サービス (消費者に見えやすい)
  - 技術 (消費者に見えにくい)
  - こころざし・コミットメント (消費者にもっと見えにくい)

# 認定試験のもたらす価値 受験者にとって

[www.noca.org](http://www.noca.org)

---

- 第三者による認知
- 信用性
- 専門性についての評価の向上
- 個人的な達成感
- 専門性についての持続的なサポート

# 認定試験のもたらす価値 受験者にとって

[www.noca.org](http://www.noca.org)

---

- 一定のレベルの知識や技術を保持していることの実証
- 職業 (career) における向上と、それにとともなう収入の増加
- 保持している技術と知識を高め、また高められた技術と知識を確認する
- 雇用者あるいは政府機関からの必要条件

# 認定試験のもたらす価値 雇用者にとって

[www.noca.org](http://www.noca.org)

---

- 顧客の満足度の向上
- 安全性の向上
- 被雇用者全体の職業における能力の向上
- 採用あるいは雇用における判断材料
- 従業員の専門性を高めるための機会の提供

# 認定試験のもたらす価値 雇用者にとって

[www.noca.org](http://www.noca.org)

---

- 職業的知識と技術の常時啓発
- 従業員自身の自己の能力に対する自信を高める
- 仕事場における適性や能力に対しての雇用者の重視を実証
- 産業界や政府において定められた規則や必要資格を遵守する

# 認定試験のもたらす価値 社会全体にとって

[www.noca.org](http://www.noca.org)

---

- 信頼できるサービス提供者を特定するため
- サービス提供者に対しての信頼の向上
- 不平申し立てにおける懲戒行為のため



# 各種検定

---

- 検定試験は社会全体の利益を守るためにある。
  - しかしその公共における機能は潜在的でかつ複雑である。
  - 顕在的な受験者へのサービスをのみ強調しすぎることは危険である。
  - 不正な検定ビジネスは排除しなければならないが、不成熟な検定ビジネスは啓蒙していかなければならない。
  - 各種検定試験事業の啓蒙・評価のための第三者機関の設立の必要性。

# 各種検定

---

- 検定試験はカテゴリー化できる。
  - 検定の意図によってそれぞれの存在意義がある。
  - 余暇・趣味のために民間任意検定があってもよい。
  - なんらかの基準が必要。同じ“検定”という言葉をつかうより、“ご当地趣味検定”、“民間資格試験”とか、分けたらどうか。

# 第三者啓蒙・評価機関の活動

---

## □ NOCAの紹介

- National Commmision for Certifying Agencies: Standards for the Accreditation of Ceritication Programs (添付資料)

## □ ATPの紹介

## □ NCMEの紹介

## □ The Buros Instituite of Mental Measurementsの紹介

## □ JARTの役割

- 日本版テストスタンダード
- テスト作成ハンドブック

# NOCA

National Organization for Competency Assurance

[www.noca.org](http://www.noca.org)

---

- 資格審査関連の組織 (credentialing organizations) においての高い行動指針 (quality standards) を決定保持するためのリーダー的役割を果たす目的で1977年組織された。
  - 年次総会、各地域で行われるセミナー、各種出版事業を通して、NOCAメンバーに認定試験、資格試験、人材育成 (certification, licensure, and human resource development) に関するさまざまな情報を提供する。
-

# NCCA

National Commission for Certifying Agencies

---

- **資格試験の認定基準 (Standards for the Accreditation of Certification Programs)**
  - 消費者にも認可することの価値をPR。
  - NCCA認証マーク (seal) を与える。
  - 世界的標準にまで高める運動
  - 各資格試験プログラムは任意で加盟、また認可を受ける (application)。

# 認証活動をとおしてのNOCAの 役割と意義

- 公共の保護
- 専門的知識、技術、実務についての標準を設定
- 各専門家が実務能力において標準に達していることを実証
- 各専門家が政府の定める資格条件を満たしていることを実証
- 政府機関との協力を通して専門的職業についてのさまざまな規則について定めることの補助
- 新しい資格の創造、あるいはすでに存在している資格からの新資格の分離
- 一定の知識や技術を持つ個人を特定
- 専門性を高める
- 専門性を持つ個人にプライドと達成感を与えること
- 専門性に対しての個人のコミットメントを保持し、そして専門性に関する生涯教育を支援<sup>46</sup>

# Psychometricianの役割

---

- 資格試験によっては規定(regulatory)された条件を基に設定されたものや随意的な条件を基に作られたものがある。
- しかしどんな資格試験でも法のもとにおいて守られるべきであり、またPsychometricsが提供する標準に沿って作られているべきである。
  - *Standards for Educational and Psychological Testing* (1999, American Psychological Association, American Educational Research Association, National Council on Measurement in Education)
- “Engaging the services of a **psychometrician** is necessary to interpret and implement these standards as part of a psychometrically sound and legally defensible credentialing program.”
  - Psychometricianの役割を強調している。
  - しかし日本ではこのPsychometricsという学問の存在もよく知られていない。

# ATP

Association of Test Publishers

[www.testpublishers.org](http://www.testpublishers.org)

---

- ATPは1992年に教育、雇用、資格、認定、臨床関係のテストや測定技術を提供する機関を代表する非営利団体で1992年に設立された。
  - Education Division
  - Industrial/Organizational Division
  - Certification and Licensure Division
  - Clinical Division
- The Association of Test Publishers is the intelligent voice for testing.
  - 年次総会
  - Journal of Applied Testing Technology



# The Buros Institute of Mental MEASUREMENTS

[www.unl.edu/buros](http://www.unl.edu/buros)

---

- このInstituteは心理測定の分野における発展とそして公共の利益の保護を目的にこの70年間テストに関する分析情報を民間に提供してきた。
  - Mental Measurements Yearbook
  - Tests in Print series
  - the Buros-Nebraska Symposium on Measurement and Testing
  - the journal Applied Measurement in Education
  - direct professional consultation.

# NCME

National Council on Measurement in Education

[www.ncme.org](http://www.ncme.org)

---

- 教育の分野における測定評価に関する科学的進歩と学問的發展を促進するために設立された。
  - 教育分野における試験関係に働く心理測定のスペシャリスト
  - 政府機関に働く専門家
  - 大学関係の心理測定を専門とする研究者
- 新しい心理測定の技術はここで討議される。
- 実務関係者と研究者の交流の場
- 日本でのNCMEを目的にJARTが設立された。

The National Council on Measurement in Education (NCME) is a professional organization for individuals involved in assessment, evaluation, testing, and other aspects of educational measurement. Members are involved in the construction and use of standardized tests; new forms of assessment, including performance-based assessment; program design; and program evaluation. NCME members include university faculty; test developers; state and federal testing and research directors; professional evaluators; testing specialists in business, industry, education, community programs, and other professions; licensure, certification, and credentialing professionals; graduate students from educational, psychological, and other measurement programs; and others involved in testing issues and practices.

# JART

## 日本テスト学会

Japanese Association for Research on Testing

[www.jartest.jp](http://www.jartest.jp)

---

- 科学的で正しいテスト理論に基づき、十分な妥当性・信頼性が保証されたテスト法を一般に定着させる
  - コンピュータと情報通信の先端技術を駆使した「テスト」の研究開発、およびそれらに対応した「評価」(assessment)技術の研究開発
  - 米国などにおけるテスト理論に基づく評価技術理論の進展を導入
  - テストについてのガイドライン
  - テスト評価の団体
  - 日本における心理測定学者 (psychometrician) や評価技術者などの専門家の育成
  - このための産官学による相互の協力関係の緊密化、共同のインフラ作り、社会的コンセンサスの構築
-

# 第三者啓蒙・評価機関の活動

---

- 検定事業者からの独立中正な機関
  - 消費者や事業者への啓蒙活動
    - テストとは
    - テストに関する技術
  - 格付けあるいは審査は現段階では無理
  - Psychometrics技術の普及(JARTとの協力)
    - Psychometrician のためのWorkshop, Seminar
    - 検定業者のためのWorkshop, Seminarなどを主催
    - 消費者のためのWorkshop, Seminarなどを主催
    - Guidelineの作成
    - 検定のためのテスト技術のManual
    - 検定の実態調査
  - テストスタンダードの紹介

# 検定実務業者の組織化

---

- ATPのような団体
- 第三者啓蒙・評価機関との連携
- JARTとの協力活動
- 自己のテストに関する研究活動
  - NCMEのような活動
- 将来NOCAのような組織づくりのための母体作り
  - オープンな組織作り
  - 民主的な運営
  - 検定試験における測定技術を科学的に捉える
  - Psychometricianの採用

## これからの行動プラン

---

- Psychometricianを含めた検定試験に関する第三者独立中立な啓蒙・評価機関を設立。
- ガイドラインを実際の運営
- 啓蒙・評価機関、JART、検定実務団体の3者協力体制で日本版NOCAを育成する。
- 日本版テスト作成ハンドブックの発行