

## 意見陳述参考資料

NPO全国検定振興機構  
理事長 吉田博彦

日本人のほとんどが学生時代にペーパーテストを受けた経験があると思います。ところが、受検したテストの品質、妥当性や信頼性（たとえば、その測定の誤差・測定の精度・問題の適正な困難度・その困難度の配分など）について、ほとんどの人が疑問を抱きません。工業製品に対する厳しい「品質管理」や「信頼性工学」を積極的に取り入れ、JIS規格の自主的評価を行って、世界の信頼を得てきた国民性にしては不思議なことです。

自費で受検する民間検定試験であっても状況は同じです。これには日本の教育界の特殊性が影響していると思います。日本に限ったことではなく、世界中の国々で「教育には権威が必要だ」とする考え方がありますが、日本の場合、過度に「教育」や「教師」という言葉に権威主義的なイメージを国民が抱いていることがその原因ではないでしょうか。

こうした「教育する側の権威」がテストの信頼性や内容的妥当性をしっかりと評価し、その情報が公開されることで得られているのであればまだいいと思います。しかし、過去に行われている形式を踏まえたもので、経験的に妥当性があれば、信頼性や内容的妥当性の検査・測定がされていない学力テストの結果を、受検者が何の疑いもなく受け入れるということは、自ら学ぶことを基本とする生涯学習の推進にとって、大きな課題といえると思います。

日本には千を越える民間検定試験があり、行政においても多くのテストや資格試験が実施されています。しかし、受検者の不正受検が社会的な問題になることはあっても、その多くのペーパーテストの「テストの品質」について、これまで日本社会では問題になったことはありません。テストの作成側、実施側が自己評価を行っているのかどうかも公開されず、ましてや第三者評価もまともには行われていないという状況が続いているのです。

特に、現在進行しているアクティブラーニングの教育における評価の課題として、従来の主力であったペーパーテストでは測れないパフォーマンスを評価する必要性が出てきています。そうした評価・測定にかかわる課題を克服していかないと教育におけるPDCAサイクルは確立されず、アクティブラーニングの教育が行き詰ってしまう可能性も出てきます。

そうした時に、本部会で検定試験の評価の問題が論議されていることに敬意を表したいと思います。我々全国検定振興機構も民間のテスト・検定試験の質的な向上を目指し、社会から信頼されるテスト・検定試験の改善に向けて活動を続けており、今後はパフォーマンスにかかわる学習活動をどのように評価するのかという点について具体的な評価法について研究し、教育活動の現場での試行錯誤を行っていく所存です。

以下、検定試験の質の向上等について、意見をまとめておきたいと思います。

## 1. 教育評価の意味と検定団体の自己評価・第三者評価について

教育評価は学習者の知能・学力等の変化を、教育目的に照らして判定することである。そのため、検定試験の評価の目的は受検者を指導する団体等の教授計画改善や受検者の学習の動機づけを通して、教育効果の向上を図ることである。

こうした「教授計画改善」「学習の動機づけ」「教育効果の向上」という学習者の育成のために行われる評価というイメージは日本社会では弱く、教育評価と言えば、入試に代表されるように、選別のために行われる評価のイメージが強い。

こうした日本の教育風土を理解したうえで、民間の検定試験は「選別のため」ではなく、「学習の質の向上」を目的としていることを肝に銘じ、学習者の育成のためにより良い学習を実現するための手段として教育評価を行うということを理解しなければならない。

こうした学習者の側の教育評価と同じ問題意識で検定試験実施団体の評価や、その団体が行う検定試験の評価のあり方を考えれば、今回の検定試験の質的向上のためにはしっかりと評価活動が必要であることは理解されるはずである。そして、「検定試験の改善・質的向上」を目的として評価を行うのだから、評価の基本が「検定試験の改善・質的向上」を実践する検定試験実施団体の自己評価であることは自明である。

その自己評価の質を改善するためには、自己評価を行う方法や能力を各検定試験実施団体が保持していることが必要であり、そのことを当該団体とは利害関係を有しない第三者が評価することで、検定試験実施団体の自己評価に社会的な認知が得られることになる。こうした「自己評価を補完する機能」が第三者評価なのである。

## 2. 自己評価のあり方

自己評価をより良くするためには、自己評価を行うための労力が検定試験実施団体の負担になり、そのために検定試験の費用が高騰したり、運営の質的な低下を招いたのでは意味がないので、自己評価の頻度については、試験実施毎が理想だが最低年度内に一回とするのが妥当であると考えます。

## 3. 第三者評価のあり方

### (ア) 第三者評価を行う対象範囲

社会から信頼される検定試験であるためには、自己評価として検定試験団体の財務運営、組織運営、試験運営等のような定型的運営業務関係について評価（定型的評価）するのは最低限のことであることは誰もが理解できることである。そして、第三者評価が必要だということを前提とするならば、第三者評価としても定型的評価が最低限の必要事項であ

ることは議論の余地がない。

第三者評価で議論となるのは、民間の検定団体が多様な分野・内容の検定試験を実施しているため、その試験問題・検定内容や試験結果・検定結果についての評価に踏み込んでいいかどうかという点である。前に述べた「我が国の評価の文化」が一般にあるため、第三者評価が「第三者が選別すること」という感覚から、問題内容等を評価対象にすれば、「その分野の高度な専門性を保持している我々に対して、第三者がその専門性の内容の評価するというのはプライドが許さない」という感情を多くの民間検定団体が抱くだろうし、ましてや国がそれを直接実施するとなれば、「民間教育事業に対する国家管理だ」という議論も起こる可能性が高い。

しかし、この議論は「テストの品質」ということについての理解不足があるための誤解から生じたものである。

テストの品質の良否はその検定試験の知識や技能の質ではなく、その検定試験を構成する一つ一つのテスト問題の識別力やテスト全体の誤差の程度や設計の妥当性をもって判定するものである。つまり、良質なテスト問題とは、測定しようとする知識や技能を問題にするのではなく、そのテストがしっかりとした識別力を持ったもので、一定の誤差内で測定できたものを言うのである。

誤解を恐れずに言えば、野球で投手が160kmを超える凄い速球を4球続けて投げても、その投球がストライクでなければ、ファールボールで出塁されてしまうように、野球のルールではストライクゾーンの投球でなければ「良い投手」とはいえない。これと同じように、「テスト理論」から見て、妥当性があり、信頼度が担保されていなければ、専門性の高い問題が出題されていても、良質なテスト問題とはいえないのである。

検定試験を実施する団体の中にはそうした「テストの品質」に対する理解がなく、「専門性が高ければ良い検定だ」と勘違いしているところも多く存在するので、自己評価ではテストの品質が保証されない可能性が高い。そのため、第三者評価では試験問題・検定内容や試験結果・検定結果に対して、テストの専門家による妥当性、信頼度などについて評価を行い、その情報を実施団体に提供して、検定試験の質的向上を図ることが必要である。

#### (イ) 第三者評価は何をどう評価するのか…「信頼性」と「妥当性」の問題

##### [評価の信頼性]

評価・測定に対する信頼性とは、「評価者に対する信頼性」と「評価方法に対する信頼性」とに分けて考えることができる。「評価者に対する信頼性」とは、評価者が評価を行うだけの十分な技術を持っているかということである。また、「評価方法に対する信頼性」とは、評価する内容に対して、適切かつ十分な方法で評価を行っているかということである。

たとえば、「学力テスト」を例にしてみると、学力テストには問題の作り方や小問の組み合わせなどによって総合得点に必ず「誤差」が出てくる。どんなに優れたテスト問題でも測定に誤差はある。つまり、極端な言い方をすれば「誤差のないテスト」は存在しない。

評価には誤差がつきもので、信頼性が高い評価とは誤差がないものではなく、誤差を少なくするための方法論があり、それが常に作動していることを意味している。そのため、そうした作問体制、テスト結果の分析体制によって、「誤差を徐々に改善するためにどのような方法がとられているか」という点を第三者評価の視点とするのが良いと考える。

#### [評価の妥当性]

「妥当性」とは評価の内容面をとらえる概念である。たとえば、学力テストを使ってある領域の学力を測定したいとき、その学力テストの内容がいかに妥当的な問題群で構成されているかというのが「妥当性」の概念である。

妥当性の第一はそのテストが何を目的にしたものであるかということである。経験主義的に作問し、時間内で解けることや100点満点になる様に形式を整えること等に意識がいてしまい、「妥当性」という考え方が薄れてしまうことがある。そのため、「評価の目的と評価の基準がいかに妥当な関係で構成されているか」という点を第三者評価の視点とするのが良いと考える。

#### (ウ) 第三者評価の評価者について

第三者評価を行うには定型的評価とテストの品質評価の2側面が必要である。そして、評価者とテストの品質を評価する評価者は別のチームにする必要がある。

財務運営、組織運営、試験運営などを評価する定型的評価は主に会計士や弁護士などの財務・法令の専門家と検定試験関係者があたり、自己評価をベースにそれを検証するという形式が妥当であると思う。また、テストの品質の評価についてはテスト学会の専門家を中心となり、メンバーの中に第三者評価の対象となっている当該検定団体のスタッフが参加し、評価活動を行うのが良い。それによって検定団体の特性を踏まえた検証が可能となり、作問の現状、過程が説明されることで評価活動の助けになり、評価の後の実際のテスト改善に役に立つ。

以上のような形で第三者評価を進めていく場合、第三者評価を行う実施機関が必要となるが、この第三者評価機関の質の問題を担保するためにも、第三者評価を行う実施機関が2つ以上あることで、その評価活動を各検定団体が比較できるようにしたほうが良い。また、第三者評価機関の公益性の観点から法人格が議論されることになるが、公益法人や非営利法人であるかどうかよりも、良質な評価サービスを提供できる組織であることが肝要なので、法人格がそれを担保するわけではない。従って、株式会社であろうが、NPOであろうがそれが第三者評価機関の可否を決定するものではない。

ただ問題なのは、第三者評価機関はビジネスとしては成立するのが極めて困難な事業である。なぜなら、評価者にはかなりの数の専門家が当たらなくてはならず、その費用は安くはない。しかし、第三者評価を受ける一つの検定団体が負担する「審査料」は、大多数の民間検定団体は中小規模の事業体で、経営的には安定しているところが少なく、審査の費用が高額だと、その費用を負担できない。こうしたことを勘案すると、第三者評価機関の運営事業に参入する株式会社は少ないだろう。

#### (エ) 第三者評価と国のかかわり

民間の検定試験が大学入試に活用されるという時代が目の前に迫っていることから、第三者評価の重要性が理解されるようになった現在、この問題を担当する文部科学省が第三者評価機関になる、または、何らかの形で直接に関与するなどのことが検討されるかもしれない。

しかし、「新しき公共」ということが言われて久しい我が国において、民間の検定試験の第三者評価を文部科学省が直接行うことは、2005年度末に廃止した民間の検定試験の「認定」を復活させることと同義となり、時代に逆行することになりかねない。

また、国が直接第三者評価に関わるのではなく、一定の基準等を文部科学省が提示するという形でこの事業の質的向上を図るという方法が考えられる。確かに、第三者評価がスタートする段階でこの問題の方向性を示し、民間検定団体の自己評価・第三者評価への取り組みを促進することは検定試験の質的向上には大切なことだが、この第三者評価の質的向上を図るには(ウ)で述べたような実施機関が切磋琢磨する状況を作り出す方がより有効であると思う。

そして、こうした状況を作り出すために、第三者評価の審査料の一部を国が負担し、ビジネスとして成立する環境を作り上げる方が、この事業の質的向上を図るより有効な手段であると思う。そうすることでこの事業に参入する事業者が増えて、その第三者評価の質が比較できれば、より良い第三者評価機関が育成されることになる。

この時、第三者評価機関を文部科学省が認定し、その認定された第三者評価機関の審査料を国が負担するという方策や良質の第三者評価機関を育成するために機関助成を行う方策などが考えられるが、こうした国の関与はうまく行ったためしがない。本気で「良い第三者評価を受けたい」と願う検定団体は自己評価をしっかりとて、その補完として第三者評価を受けようとするので、第三者評価機関の選定は各検定団体に任せたほうが良い。

検定試験の質的向上を目指して、国がこの評価問題で関与すべきは、方向性を示すこと、良い第三者評価機関が育成されるための事業環境を整備すること、そして最後に、その評価結果を検定試験を受検しようとしている受検者に、検定試験の結果を活用しようとしている大学や企業に公表することである。

#### (オ) 第三者評価の試行報告

今年度受託した第三者評価の試行においては、文部科学省が平成22年にまとめた「検定試験の評価ガイドライン(試案)」を基本に検討を行ったが、今後のコンピュータの普及の可能性、および検定試験の大学入試等での活用も視野に入れて検討したため、自己評価シートの評価項目に「コンピュータを使って行う試験」に関する4項目、および「試験の実施運営を(検定試験実施団体でなく)当該試験を受検する団体(法人・学校・塾等)の監督下で行うことを認めている場合」「不正受検」に関する3項目を加えた58項目を審査項目として試行を実施している。

追加した項目は以下の7項目である。

**【コンピュータを使って行う試験の場合】**

- ① PCの稼働状況等に関し、試験の前までに十分なチェックを行う体制が整えられている。
- ② IDとパスワード等で本人確認が行われている。
- ③ システムの冗長化等、機器に不具合が生じてても試験が継続できる体制が整えられている。
- ④ 何らかの理由で試験が途中で止まっても、停止箇所から試験が再開できる等、バックアップリカバリ体制が整えられていること。

**【試験の実施運営を個別会場(法人・学校・塾等)の監督下で行うことを認めている場合】**

- ① 試験実施運営の管理が適切になされている。
- ② 受検手続きに関する共通理解を図り、厳正公平・適切に試験実施を遂行できる体制がとられている。

**【すべての検定試験に対して】**

- ① 受検者の不正行為・迷惑行為防止に対する対応がなされている。

第三者評価方法

最終的な評価・・・「A評価」「B評価」「C評価」「不適合(不合格)」の4区分とする(呼称に関しては検討中)

第三者評価に関する基本的な考え方

- ① 各小項目評価(58項目)の総合得点における段階評定を行う
- ② 大項目毎の段階評定(評価分野は「実施主体」「実施内容」「実施手続」「検定結果の活用促進及び継続的な学習支援」「情報公開」の5項目)を行う
- ③ 最終的な評価の判定方法の概要は以下のとおり  
各小項目を内容の重要度に応じて、「◎・○・△」の3段階に区分して「重み」を付ける。【◎(5点) ○(3点) △(1点)】
- ④ 各小項目への回答に対して、以下a～eの5段階で評価を行い、a、bの評価を「達成(合格)」と判定し、c、d、eの評価を「未達成(不合格)」と判定する。【a:「十分達成されている」 b:「おおむね達成されている」 c:「一部達成されている」 d:「あまり達成されていない」 e:「ほとんど達成されていない」】
- ⑤ 小項目の合計点数(満点)の一定ライン(例:「85%」・「75%」・「65%」)を「各認定ライン」の閾値として設定する。

- ⑥ 大項目評価毎の点数（満点）の一定ライン（例：「90%」・「80%」・「65%」）を「各評価ライン」の閾値として設定する。
- ⑦ 「総合得点（小項目の合計得点）」に基づき評価区分を定め、「大項目毎の評価」の要因と不適合条件を加味して総合的な判定を行う。

※不適合条件の例

- ① 「大項目「実施主体」「実施内容」「実施手続」で下位評価があった場合の評価区分を下方に修正する」等
- ② 「小項目の「◎」の重みの項目に1つでも未達成（不合格）があった場合、評価区分を下方に修正する」等

以上

〔用語解説〕

【妥当性】テストが本来測ろうとしている受検者の知識・技能等を正確に測定しているか、の度合い

【信頼性】測定対象が変化しない限り安定したテスト結果が得られる度合い（測定の一貫性の度合い）

【識別力】テスト項目が測ろうとする知識・技能等の個人差をどれだけ識別できるかの指標