

平成 27 年度 文部科学省委託調査

「検定試験における第三者評価に関する調査研究」報告書

特定非営利活動法人全国検定振興機構

平成 28 年 3 月

目次

I. 調査の概要	1
1. 調査の目的	1
2. 調査の内容	1
3. 「第三者評価の在り方に関する検討委員会」委員	2
II. 第三者評価における評価項目及び評価方法について	3
1. 第三者評価における評価項目について	3
2. 具体的な評価項目	4
3. 第三者評価の評価方法の基本的な考え方	8
4. 第三者評価の具体的な評価方法	8
5. 第三者評価の具体的な運用方法	10
III. 第三者評価の試行について	11
1. 第三者評価の試行のプロセス	11
2. 申請に必要な書類	11
3. 第三者評価の試行結果	12
4. 第三者評価の試行から抽出された課題等	19
IV. 問題内容の評価の在り方に関する分科会（英語部会）について	20
1. 分科会（英語部会）開催と検討委員	20
2. テスト作成手順に係る信頼性と妥当性	20
3. 評価項目について	21
4. 今後の課題等	23
V. 第三者評価の推進に資する研修会	24
1. 第1回：効果的なテストの作成方法	24
2. 第2回：ICTを活用したテスト導入	24
3. 主なアンケート結果	25
資料編	26
【資料1】 第三者評価の在り方に関する検討委員会及び開催経過	27
【資料2】 研修会アンケート結果	28

1. 調査の概要

1. 調査の目的

検定試験は、個人の自己実現の手段として、また、個人の学習成果を評価する手段として有効であるが、民間検定試験の実施は特段の法規制等がないことから、その信頼性の確保や質の保証を図ることが課題となっている。このため、民間検定試験の質を保証するための評価手法の有効性、安定性、継続性等を確保するための仕組みとして、検定試験における第三者評価の在り方等を検討する。

2. 調査の内容

生涯学習振興行政関係者、検定事業関係者、学校関係者、企業等関係者、学識経験者等で構成する「第三者評価の在り方に関する検討委員会」を設置し、以下の調査研究を実施した。

(1) 第三者評価の試行

調査期間 平成 28 年 1 月～平成 28 年 2 月

調査対象 検定試験の効果が全国的に通用し、実績や受検者数が一定程度ある検定試験の中、分野に関して偏りのない 3 団体

調査方法 書類審査（財務諸表等）及び現地調査（評価項目の確認、聞き取り等）

(2) 問題内容の評価の在り方に関する分科会（英語部会）

検定試験問題の信頼性や妥当性等を評価する方法について検討し、第三者評価に必要なツール（評価項目シート）を作成した。

(3) 第三者評価の推進に資する研修会の開催

第 1 回：『効果的なテストの作成方法』

～効果的なテストの作成方法とは/問題内容の信頼性と内容の検証/合否判定基準について～

日 時：平成 27 年 12 月 9 日（水）

会 場：国立オリンピック記念青少年総合センター

講 師：東北大学大学院教育学研究科教授 柴山直氏

内 容：昨今、個人の価値観やライフスタイルが多様化する中で、多種多様な学習の機会が提供されている。各個人の学習成果が社会全体で幅広く通用し、評価され、活用できる民間検定試験に求められるテストとはどのような条件を備えたテストを指すのか。テストの作成・実施における一連の作業手順について、全体目標と計画作りから始まり、テスト内容の決定、問題作成、結果の集計、分析、判定など段階ごとに解説した。社会から信頼され、安心して受検できる検定試験の在り方について検討した。

第 2 回：ICT を活用したテスト導入

～ICT を活用したテスト導入に向けて/ICT を活用したテストの現状/導入のメリット～

日 時：平成 28 年 1 月 22 日（金）

会 場：国立オリンピック記念青少年総合センター

講 師：電気通信大学教授 植野真臣氏

内 容：情報通信技術の進展により学習スタイルが劇的に変化中、ICT活用教育の進展と相まって、民間検定試験にもICTを活用した試験実施を行うことにより受検者の利便性を高めることが考えられる。この現状を踏まえ、民間検定実施団体はその導入に向けてどのように対応していく必要があるか、ICTを活用したテスト実施に向けて、について考察した。また、テスト実施業務の標準化、テスト結果の分析や評価の簡素化、多様な実施形態のテストの実施が可能になるなどICTの活用による検定実施団体側のメリットについても解説した。

(4) その他（検討委員会での主な意見）

- 第三者評価は良質な検定試験が広く普及するよう育成・検定試験の改善を促す観点から行うことが大切。
- 第三者評価では「検定試験の運営・組織に関する審査」とともに、企業・大学入試での検定試験の活用等を視野に入れた場合には、検定試験問題の質を第三者が評価することにより、各検定実施団体の自律的な質の改善と向上が促されていくことが必要。
- 今回は実績が一定程度ある知識・技能を測る検定試験について第三者評価の試行を行った。今後は社会的に通用する検定試験と学習到達度の確認や趣味的な検定試験、及び小規模な検定試験やご当地検定等のあらゆる分野・規模・特徴を持った検定試験に対しての第三者評価システムの確立（コスト面での考慮も必要）が必要になっていくと思われる。また国の検定及び実技を測る検定試験の審査が必要になることも考えられる。
- 分科会（英語部会）で検討した試験問題の質の評価では、受検者側が適切な判断を行えるよう、試験目的に見合った活用がされているかどうかの審査が要求される。日本の学校教育（指導目的・進学・入試・単位取得等）で活用されている検定試験については学習指導要領との整合性を審査することが不可欠と思われる。

3. 「第三者評価の在り方に関する検討委員会」委員

梅澤 健	（弁護士・竹町法律事務所）
佐々木 康浩	（株式会社三菱総合研究所・主席研究員）
惣脇 宏	（京都大学大学院総合生存学館・教授）
中野 達也	（東京都立白鷗高等学校・主任教諭）
萩原 民也	（特定非営利活動法人日本語検定委員会・事務局長）
村木 英治	（東北大学・名誉教授） （敬称略、五十音順）

Ⅱ. 第三者評価における評価項目及び評価方法について

第三者評価の評価項目については、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）」を基本に作成し、また昨年度実施した第三者評価の試行結果を踏まえ、第三者評価の在り方に関する検討委員会において、必要なツール（「審査項目記入シート」等）の検討を行った。

1. 第三者評価における評価項目について

前記述のとおり、今年度の評価方法は第三者評価の在り方に関する検討委員会において、昨年度実施した第三者評価の試行結果を踏まえて検討し、次の評価項目で行うこととした。

表 1-1 第三者評価における評価項目の概要

大項目	中項目	小項目数	当該のみ	重み◎	重み○	重み△
1 実施主体	組織	5	0	1	1	3
	財務	4	1	0	2	2
	その他	2	0	1	0	1
2 実施内容	目的	1	0	1	0	0
	内容	5	4	1	1	3
	手段	2	1	1	0	1
	その他	1	0	0	1	0
3 実施手続	事前準備	6	2	2	2	2
	試験実施	14	9	4	2	8
	事後対応等	2	0	0	1	1
4 検定結果の活用促進 継続的な学習支援	検定結果の活用促進	3	0	1	0	2
	継続的な学習支援	2	0	0	0	2
5 情報公開	実施主体	3	0	0	3	0
	実施内容	3	0	0	3	0
	実施手続	3	0	0	3	0
	検定結果の利用促進	1	0	0	1	0
	継続的な学習支援	1	0	0	1	0

2. 具体的な評価項目

第三者評価における具体的な評価項目の内容は、以下の 58 項目である。なお網掛けで記された項目は、昨年度の第三者評価の試行結果を踏まえ、今回付け加えた項目である。

1 実施主体

1-1 組織

- (1) 検定試験を実施する組織としての理念や目的が明確である。
- (2) 検定試験を実施することの社会的使命が明確である。
- (3) 検定試験を実施する組織（役職員体制、事務処理体制、作問体制、危機管理体制等）が、検定試験の目的、使命を達成するための組織として適切に構成されている。
- (4) 手続等に関する問い合わせ、試験後の問い合わせ先が設置され、受検願書やホームページ等に適切に公開されている。
- (5) 目標（P l a n）－実行（D o）－評価（C h e c k）－改善（A c t i o n）という P D C A サイクルに基づき、組織的・継続的に事業を改善している。

1-2 財務

- (6) 実施主体の財務経理情報を備えている。（検定試験を継続して実施している場合には、複数年分。）（財務経理情報の例） 収支計算書、貸借対照表、財産目録等
- (7) 財務経理担当者等に対する内部牽制体制が確立されている。
- (8) 財務経理に関して、定期的または適宜監査を受けている。
- (9) 【公益性のある実施主体の場合】（該当団体のみ）
検定事業とその他の事業との関係や財務経理情報等の区分けが明確である。

1-3 その他

- (10) 情報公開する項目や方法などを明確に定めている。
- (11) 受検者の個人情報保護に関する方針やマニュアル等が整備されている。

2 実施内容

2-1 目的

- (12) どのような知識・技能を測るかという検定試験の目的が明確である。

2-2 内容

- (13) 検定試験の内容が以下の点で明確で、検定試験の目的にかなっている。
<検定試験で測る具体的な知識・技能とその水準>
領域（分野）／対象層（受検資格等）／試験範囲／難易度／その他
- (14) 【他に類似試験がある場合】（該当団体のみ）
類似試験との関係性を学習者や利用者に分かりやすく示している。
- (15) 【学校の単位認定や進学、就職等の際に活用される検定試験の場合】（該当団体のみ）

当該検定試験と学校教育や職業能力との関係性が示されている。

- (16) 【学校教育で活用される検定試験の場合】(該当団体のみ)
学習指導要領等に準拠している明確な説明等がなされている。
- (17) 【受検資格を制限する試験の場合】(該当団体のみ)
年齢制限や事前の講座受講の有無等によって受検資格が制限されている場合には、その合理的な理由が示されている。

2-3 手段

- (18) 知識・技能の測定手法(筆記試験、C B T試験、実技試験、面接試験等)、審査・採点の基準が明確である。
- (19) 【特にITなどの技術革新が著しい分野の場合】(該当団体のみ)
内容や手段、認定基準(合格基準)等の見直しを一定期間ごとに行うなどの取組を、継続的に実施する体制となっている。

2-4 その他

- (20) 試験結果から得られるデータ等に基づき、検定試験の内容や測定手法、審査・採点基準等について、質の確保や継続的な改善を図っている。

3 実施手続

3-1 事前準備

- (21) 試験の実施規則・要項等の受検手続が定められている。
- (22) 試験実施前の情報管理対策(情報管理マニュアルの整備や担当者への研修・注意喚起など)が講じられている。(例) 試験問題、解答等の作成、印刷、搬送、保管
- (23) 受検者に対し十分な出願期間が確保されている。
- (24) 受検料の適正性・妥当性について点検・検証されている。
- (25) 【学校の単位認定や進学・就職等の際に活用される検定試験の場合】(該当団体のみ) 児童・生徒等が不利益を被らないように、配慮がなされている。
- (26) 【コンピュータを使って行う試験の場合】(該当団体のみ)
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。

3-2 試験実施

- (27) 試験監督業務のマニュアル等が定められており、試験実施会場・機関に事前に配付されており、試験監督者等の共通理解が図られている。
- (28) 受検者の本人確認は、顔写真を添付した受検票の用意や身分証による照合など、本人確認が確実に行われるよう講じられている。
- (29) 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。
- (30) 試験当日、天災や交通機関の遅延等があった場合には、試験開始時刻の変更や再

受検の容認など、受検機会の確保について配慮されている。

- (31) 障がい者について、検定試験の目的や内容、規模等に応じた一定の配慮が考慮されている。
- (32) 【試験監督者の外部委託等を行う場合】（該当団体のみ）
外部受託者等（法人・個人等）を総括する責任者が配置されている。
- (33) 【試験監督者の外部委託を行う場合】（該当団体のみ）
受検手続きに関する共通理解を図り、円滑に試験監督業務を遂行できるよう試験監督者への説明・研修が実施されている。
- (34) 【試験を個別会場（法人・学校・塾等）で行うことを認めている場合】（該当団体のみ）
試験実施運営の管理が適切になされている。
- (35) 【試験を個別会場（法人・学校・塾等）で行うことを認めている場合】（該当団体のみ）
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。
- (36) 【コンピュータを使って行う試験の場合】（該当団体のみ）
IDとパスワード等で本人確認が行われている。
- (37) 【コンピュータを使って行う試験の場合】（該当団体のみ）
システムの冗長化等、機器に不具合が生じても試験が継続できる体制が整えられている。
- (38) 【コンピュータを使って行う試験の場合】（該当団体のみ）
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ー体制が整えられていること。
- (39) 【児童・生徒や学生を対象とした検定試験の場合】（該当団体のみ）
受検者の利便性確保の観点から、公平性を保った上で、試験実施会場として学校や民間教育施設等が活用されている。
- (40) 【試験実施会場として学校や民間教育施設等を活用している場合】（該当団体のみ）
公平性・公正性が確保（試験日、試験監督の体制、試験会場に試験に関する掲示物が無いか等）されている。

3-3 事後対応等

- (41) 試験結果に関する一般情報（受検者数及びその構成、合格者数、合格率等）が、適切に公開されている。
- (42) 受検者への学習支援の観点から、試験問題や正答が公開されている。
（ただし、試験の性質上、公開することができないものを除き、事後の作問等に影響を及ぼさない範囲で）

4 検定結果の活用促進・継続的な学習支援

4-1 検定結果の活用促進

- (43) 検定試験の結果を証明する合格証や認定証等が発行されている。
- (44) 合格証や認定証等が発行されている場合には、受検者が獲得した又は保持している知識・技能の内容が一見して判断し得る記載がある。
- (45) 試験結果の活用に係るニーズを把握するための取組（受検者や利用者（学校・企業等）に対するアンケート調査等）が行われている。

4-2 継続的な学習支援

- (46) 段階的・継続的な学習をサポートする過去問題、類似問題などの参考となる資料が提供されている。
- (47) 学習意欲の促進のため、受検者の知識・技能レベルなどの情報を提供している。

5 情報公開

5-1 実施主体

- (48) 実施主体の「組織」に関する情報が公開されている。
- (49) 実施主体の「財務」に関する情報が公開されている。
- (50) 実施主体の「情報公開や個人情報保護の方針等」に関する情報が公開されている。

5-2 実施内容

- (51) 検定試験の「目的」に関する情報が公開されている。
- (52) 検定試験の「内容」に関する情報が公開されている。
- (53) 検定試験の「手段」に関する情報が公開されている。

5-3 実施手続

- (54) 検定試験の「事前準備」に関する情報が公開されている。
- (55) 検定試験の「試験実施」に関する情報が公開されている。
- (56) 検定試験の「事後対応」に関する情報が公開されている。

5-4 検定結果の利用促進

- (57) 「検定結果の利用促進」に関する情報が公開されている。

5-5 継続的な学習支援

- (58) 検定試験の「継続的な学習支援」に関する情報が公開されている。

3. 第三者評価の評価方法の基本的な考え方

第三者評価の評価方法の基本的な考え方については以下のとおり。

- (1) 昨年実施した第三者評価の試行結果を踏まえ、最終評価を「合格」「不合格」の2区分から「A評価」「B評価」「C評価」「不適合（不合格）」の4区分とした。
- (2) 前表 1-1 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）を行い、総合的な評価を行う。（詳細は、「4. 第三者評価の具体的な評価方法」を参照）

4. 第三者評価の具体的な評価方法

(1) 昨年度実施した第三者評価の試行における評価方法

最終的な評価・・・「合格」「不適合（不合格）」の2区分とする。

【第三者評価に関する具体的な評価方法】

- 1) 各項目評価の総合得点における評定（閾値を境に合格・不合格とする）
 - ①各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）： 必須項目であり、全ての項目で要件を満たすことが「合格」の条件。
 - （3点）： 80%以上の項目で要件を満たすことが「合格」の条件。
 - △（1点）： 特に条件を設けない項目。
 - ②各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」
 - ③ 全ての項目を満たした点数（満点）の「80%」を「合格ライン得点」として設定する。
なお、項目数と得点は、検定試験の分野や形態等によって異なる。

(2) 今年度の第三者評価の評価方法

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。

【第三者評価に関する具体的な評価方法】

- 1) 各小項目評価（58項目）の合計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) 「合計得点（小項目の合計得点）」に基づき評価区分を定め、「小項目の評価」と「大項目毎の評価」の2要因を不適合要件として加味し総合的な最終判定を行う。

4) 最終的な評価の判定方法は以下のとおり。

- ① 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)
- ② 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」
- ③-1 総合得点 (すべての小項目の合計得点・満点) の「85%」・「75%」・「65%」を「各評価ライン」の閾値として設定する。

表 1-2 【項目数と得点は、検定試験の分野や形態等によって異なる】

満点	A 評価	B 評価	C 評価	不適合 (不合格)
148	85%	75%	65%	95 以下
	126 以上	111 以上	96 以上	

※得点誤差を想定し「60% (89 点)」以上獲得し不適合 (不合格) になった検定試験は再評価を行うこととする。

- ③-2 大項目評価毎の点数 (満点) の「90%」・「80%」・「65%」を「各評価ライン」の閾値として設定する。

なお大項目評価毎の段階評定は「評価 1」「評価 2」「評価 3」「不合格」とする。

表 1-3

		評価項目							得点率 (閾値)			
		重み別項目数			重み付得点				評価1	評価2	評価3	
		◎	○	△	◎	○	△	Total (満点)				
1 実施主体	組織	1	1	3	5	3	3	11	25	23	20	16
	財務		2	2		6	2	8				
	その他	1		1	5		1	6				
2 実施内容	目的	1			5			5	25	23	20	16
	内容	1	1	3	5	3	3	11				
	手段	1		1	5		1	6				
	その他		1			3		3				
3 実施手続	事前準備	2	2	2	10	6	2	18	56	50	45	36
	試験実施	4	2	8	20	6	8	34				
	事後対応等		1	1		3	1	4				
4 検定結果の活用促進 継続的な学習支援	検定結果の活用促進	1		2	5		2	7	9	8	7	6
	継続的な学習支援			2			2	2				
5 情報公開	実施主体		3			9		9	33	30	26	21
	実施内容		3			9		9				
	実施手続		3			9		9				
	検定結果の活用促進		1			3		3				
	継続的な学習支援		1			3		3				

※小数点以下は四捨五入。大項目評価が 65% に届かない場合は「不合格」とする。

- ③-3 「合計得点（小項目の合計得点）」に基づき評価区分を定め、「小項目の評価」と「大項目毎の評価」の2要因を不適合要件として加味し総合的な最終判定を行う。

表 1-4

評価区分	総合得点	得点率	不適合条件
A 評価	126 点 以上	85% 以上	1) 大項目評価 1 から 5 で「評価3」が1つでもあった場合は「B 評価」とする 2) 小項目評価の「◎」の重みの項目に1つでも不合格の場合は「C 評価とする」 3) 大項目評価 1 から 5 で「不合格」が1つでもあった場合は「不適合（不合格）」とする。
B 評価	125 点～ 111 点	75% 以上	1) 大項目評価 1 から 3 で「評価3」が2つあった場合は「C 評価」とする 2) 大項目評価 1 から 5 で「評価3」が3つ以上あった場合は「C 評価」とする 3) 小項目評価の「◎」の重みの項目に1つでも不合格の場合は「C 評価とする」 4) 大項目評価 1 から 5 で「不合格」が1つでもあった場合は「不適合（不合格）」とする。
C 評価	110 点～ 96 点	65% 以上	1) 大項目評価 1 から 5 で「不合格」が1つでもあった場合は「不適合（不合格）」とする。
不合格	95 点以下		

5) 評価区分の呼称・・・「評価A」「評価B」「評価C」「不適合」とする。

(呼称に関しては引き続き検討が必要)

6) 各評価の位置付けは以下のとおりとする。

「評価A」・・・公的な活用が可能な検定。

「評価B」・・・企業・大学入試等での活用が可能な検定。

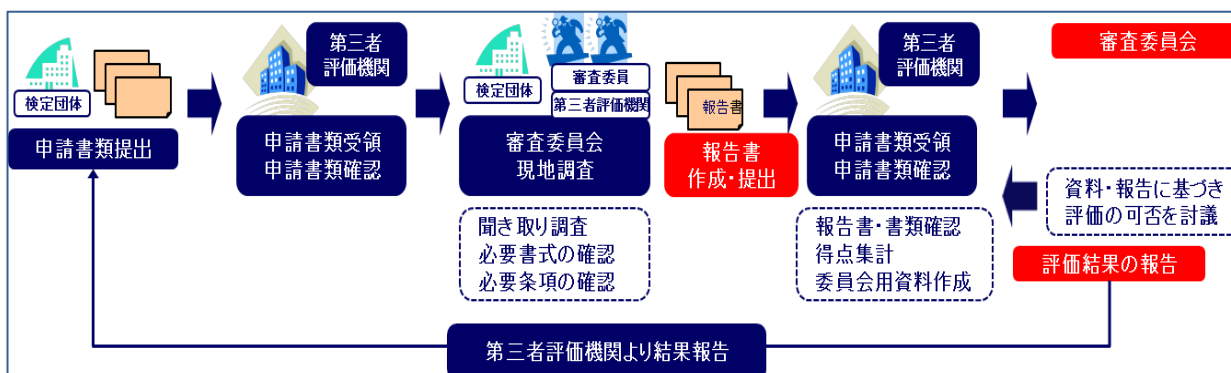
「評価C」・・・検定試験を実施する上での最低条件は満たしている。

「不適合（不合格）」・・・検定試験の実施運営上での欠陥があるため、現時点で検定試験として活用していくには不適合。

5. 第三者評価の具体的な運用方法

第三者評価の具体的な運用方法について、「第三者評価機関」は、検定実施団体からの申請を受けるとともに、評価結果を検定実施団体に報告する。また「第三者評価機関」の中に、申請のあった検定実施団体に対する現地調査を行い、調査結果のまとめ、評価の可否を審査する「審査委員会」を設置する。運用イメージは以下のとおり。

図 2-1



Ⅲ. 第三者評価の試行について

今回の検定試験の第三者評価の試行に際しては、実際の運用そのものも検証対象となるため、前記述「5. 第三者評価の具体的な運用方法」の運用イメージをシミュレートし、第三者評価の各審査過程における検証を行った。具体的な審査過程は、以下のとおりである。

また第三者評価に必要とされる書類等も整備し、実際の運用を試みた。

1. 第三者評価の試行のプロセス

今回実施した第三者評価の試行プロセスは、以下のとおり。

- ① 申請書類等の提出
検定実施団体 → 全国検定振興機構事務局
- ② 申請書類等の確認
全国検定振興機構事務局
- ③ 申請書類等の試行審査及び現地調査での審査事項の確認
第三者評価の在り方に関する検討委員会（現地調査委員）及び全国検定振興機構事務局
- ④ 現地調査
第三者評価の在り方に関する検討委員会（現地調査委員）及び全国検定振興機構事務局
- ⑤ 審査結果に基づく評価の検討（試行審査（案）の作成）
第三者評価の在り方に関する検討委員会（現地調査委員）及び全国検定振興機構事務局
- ⑥ 試行審査（案）の報告・検討
第三者評価の在り方に関する検討委員会（全員）

※ 今回は試行のため、第三者評価の在り方に関する検討委員会に報告・検討を行い終了。
実際の第三者評価においては、検定実施団体に対して報告を行う。

2. 申請に必要な書類

第三者評価を申請する検定実施団体が作成し、提出する書類の概要は以下のとおり。

(1) 検定試験認証制度申請書

検定試験団体に作成を課した。申請する検定実施団体名（あるいは機関名）、連絡先、申請する検定試験名等を記入し、申請する検定実施団体の責任者の捺印（代表印、役職印等。個人印は不可）を課した。

(2) 基本情報シート

申請する検定試験に関する基本的な情報を記入する書類。検定試験名、試験実施概要、実績等を記入。

(3) 審査項目記入シート

審査対象となる項目（表 1-1 の具体的な評価項目等）を示した書類。検定実施団体の組織に関する情報、検定試験の運営状況等について記入。

(4) 付属資料（原則として最新版）

- 1) 定款（申請時点）
- 2) 役員名簿
- 3) 組織図
- 4) 実施団体概要
- 5) 自己評価シート
- 6) 受検案内・願書・受検票等、資料一式
- 7) 資格・検定試験についての情報
- 8) 審査基準・問題・解答用紙
- 9) 試験実施者・試験監督者への説明資料
- 10) 結果通知票・フィードバック案内等
- 11) 合格証書・証明書等
 - 1) ～11) 必須
- 12) 付属説明資料（提出可能な場合）
 - (ア) 申請の日の属する事業年度の直前の事業年度末における事業報告書及び収支決算書
 - (イ) 業務分掌
 - (ウ) 問題作成委員名簿
 - (エ) 評価基準（採点基準）
 - (オ) 学習指導要領等に準拠している根拠となる資料

3. 第三者評価の試行結果

当面、検定試験の効果が全国的に通用し、実績や受検者数が一定程度ある検定試験から第三者評価の取組が進展することが期待されるため、第三者評価を試行的に実施するに際しては、検定試験の分野に関して偏りのない下記の 3 団体を抽出した。第三者評価を試行した検定実施団体の概要及び評価の概要は次のとおり。

表 1-5 第三者評価を試行した団体及び検定試験

項番	検定実施団体	検定試験	年間受検者数	検定分類
1	公益財団法人	X 検定試験	約 1 万人～3 万人	コンピュータ
2	特定非営利活動法人	Y 検定試験	約 5 万人～10 万人	語学
3	公益社団法人	Z 検定試験	約 3 万人～5 万人	教育・学術

(1) 公益財団法人が主催するX検定試験

検定分類がコンピュータに属する検定実施団体に対する調査は、「X検定試験」を実施する公益財団法人に協力を依頼した。年間受検者数では中規模の検定である。「X検定試験」に対する第三者評価の試行結果は以下のとおり。

表 1-6 X検定試験の第三者評価試行結果

【試行結果】		X検定試験											大項目評価ライン					
大項目	中項目	評価項目(満点)								X検定採点			得点	評価	評価1	評価2	評価3	不合格
		重み別項目数			重み付得点					重み付得点								
		◎	○	△	◎	○	△	合計	◎	○	△	点数						
1 実施主体	1-1 組織	1	1	3	5	3	3	11	5	3	3	11						
	1-2 財務	0	2	2	0	6	2	8	0	6	1	7						
	1-3 その他	1	0	1	5	0	1	6	5	0	1	6						
		2	3	6	10	9	6	25	10	9	5	24	評価1	23	20	16		
2 実施内容	2-1 目的	1	0	0	5	0	0	5	5	0	0	5						
	2-2 内容	1	0	2	5	0	2	7	5	0	2	7						
	2-3 手段	1	0	1	5	0	1	6	5	0	1	6						
	2-4 その他	0	1	0	0	3	0	3	0	3	0	3						
		3	1	3	15	3	3	21	15	3	3	21	評価1	19	17	14		
3 実施手続	3-1 事前準備	2	1	2	10	3	2	15	10	3	2	15						
	3-2 試験実施	3	2	6	15	6	6	27	15	3	5	23						
	3-3 事後対応等	0	1	1	0	3	1	4	0	3	1	4						
		5	4	9	25	12	9	46	25	9	8	42	評価1	41	37	30		
4 検定結果の活用促進 継続的な学習支援	4-1 検定結果の活用促進	1	0	2	5	0	2	7	5	0	1	6						
	4-2 継続的な学習支援	0	0	2	0	0	2	2	0	0	2	2						
		1	0	4	5	0	4	9	5	0	3	8	評価1	8	7	6		
5 情報公開	5-1 実施主体	0	3	0	0	9	0	9	0	9	0	9						
	5-2 実施内容	0	3	0	0	9	0	9	0	9	0	9						
	5-3 実施手続	0	3	0	0	9	0	9	0	6	0	6						
	5-4 検定結果の利用促進	0	1	0	0	3	0	3	0	3	0	3						
	5-5 継続的な学習支援	0	1	0	0	3	0	3	0	3	0	3						
		0	11	0	0	33	0	33	0	30	0	30	評価1	30	26	21		
最終評価		11	19	22	55	57	22	134	55	51	19	125	A評価					
													得点率	93.3%				

X検定試験に対する第三者評価の試行により集計された得点結果は、表 1-6 のとおり。X検定試験については、今回の評価項目 58 項目の内、◎11 項目、○19 項目、△22 項目の合計 52 項目が評価対象となっている。表 1-6 の集計結果をもとに、前記述「4. 第三者評価の具体的な評価方法」の判定基準に基づき判定した結果、X検定試験は「評価A」と判定された。

具体的な内訳は、必須項目である◎の 11 項目については、すべて合格の判定で、11 項目×5 点=55 点を獲得した。また○の該当 19 項目の内、17 項目が合格の判定で、17 項目×3 点=51 点を獲得した。また△の該当 22 項目の内、19 項目が合格の判定で、19 項目×1 点=19 点を獲得した。この結果、X検定試験は、満点 134 点の内 125 点（得点率 93.3%）を獲得した。また大項目 1～5 ともすべて「評価 1」となり、A評価基準 85%（114 点）、及び前表 1-4 で記した不適合条件もクリアしたため、「評価A」と判定された。

【評価の概要】

X検定試験の第三者評価の試行については、A評価基準 85%をクリアし、該当する項目に対しては、大項目 1～5 ともすべて「評価 1」、◎の必須項目は全項目合格したため、「評価A」と判定された。全般的な評価の傾向を見ると、管理・監督面での改善点が多い傾向が認められた。例えば、試験実施運営に関する業務等を外部団体に委託した際の管理体制の確立や準会場（企業・学校等を会場として当該受検団体が自ら監督を行う会場）での試験実施運営の管理・監督プロセスが、今回の課題として指摘された。

今後、これらの管理・監督体制を具体的に確立することにより、X検定試験は、より信頼性が増すものと考えられる。

(2) 特定非営利活動法人が主催するY検定試験

検定分類が語学に属する検定実施団体に対する調査は、「Y検定試験」を実施する特定非営利活動法人に協力を依頼した。年間受検者数では中規模の検定である。「Y検定試験」に対する第三者評価の試行結果は以下のとおり。

表 1-7 Y検定試験の第三者評価試行結果

【試行結果】		Y検定試験											大項目評価ライン				
大項目	中項目	評価項目(満点)							Y検定採点			得点	評価	評価1	評価2	評価3	不合格
		重み別項目数			重み付得点				重み付得点								
		◎	○	△	◎	○	△	合計	◎	○	△	点数	90%	80%	65%	65%以下	
1 実施主体	1-1 組織	1	1	3	5	3	3	11	5	3	3	11					
	1-2 財務	0	2	2	0	6	2	8	0	6	2	8					
	1-3 その他	1	0	1	5	0	1	6	5	0	1	6					
		2	3	6	10	9	6	25	10	9	6	25	評価1	23	20	16	
2 実施内容	2-1 目的	1	0	0	5	0	0	5	5	0	0	5					
	2-2 内容	1	1	1	5	3	1	9	5	0	1	6					
	2-3 手段	1	0	1	5	0	1	6	5	0	1	6					
	2-4 その他	0	1	0	0	3	0	3	0	3	0	3					
		3	2	2	15	6	2	23	15	3	2	20	評価2	21	18	15	
3 実施手続	3-1 事前準備	2	1	2	10	3	2	15	10	3	1	14					
	3-2 試験実施	3	2	6	15	6	6	27	15	0	5	20					
	3-3 事後対応等	0	1	1	0	3	1	4	0	3	1	4					
		5	4	9	25	12	9	46	25	6	7	38	評価2	41	37	30	
4 検定結果の活用促進 継続的な学習支援	4-1 検定結果の活用促進	1	0	2	5	0	2	7	5	0	0	5					
	4-2 継続的な学習支援	0	0	2	0	0	2	2	0	0	2	2					
		1	0	4	5	0	4	9	5	0	2	7	評価2	8	7	6	
5 情報公開	5-1 実施主体	0	3	0	0	9	0	9	0	9	0	9					
	5-2 実施内容	0	3	0	0	9	0	9	0	9	0	9					
	5-3 実施手続	0	3	0	0	9	0	9	0	9	0	9					
	5-4 検定結果の利用促進	0	1	0	0	3	0	3	0	3	0	3					
	5-5 継続的な学習支援	0	1	0	0	3	0	3	0	3	0	3					
		0	11	0	0	33	0	33	0	33	0	33	評価1	30	26	21	
最終評価		11	20	21	55	60	21	136	55	51	17	123	A評価				
													得点率	90.4%			

Y検定試験に対する第三者評価の試行により集計された得点結果は、表 1-7 のとおり。Y検定試験については、今回の評価項目 58 項目の内、◎11 項目、○20 項目、△21 項目の合計 52 項目が評価対象となっている。表 1-7 の集計結果をもとに、前記述「4. 第三者評価の具体的な評価方法」の判定基準に基づき判定した結果、Y検定試験は「評価A」と判定された。

具体的な内訳は、必須項目である◎の 11 項目については、すべて合格の判定で、11 項目×5 点=55 点を獲得した。また○の該当 20 項目の内、17 項目が合格の判定で、17 項目×3 点=51 点を獲得した。また△の該当 21 項目の内、17 項目が合格の判定で、17 項目×1 点=17 点を獲得した。この結果、Y検定試験は、満点 136 点の内 123 点（得点率 90.4%）を獲得した。また大項目については 1・3 が「評価 1」、2・4・5 が「評価 2」となり、A評価基準 85%（116 点）、及び前表 1-4 で記した不適合条件もクリアしたため、「評価A」と判定された。

【評価の概要】

Y検定試験の第三者評価の試行については、A評価基準 85%をクリアし、該当する項目に対しては、大項目については 1・3 が「評価 1」、2・4・5 が「評価 2」、◎の必須項目は全項目合格したため、「評価A」と判定された。

今回の第三者評価の試行結果では、平等・公平な試験実施運営を行うにあたり、統一した対応をするための内規・文書等が欠如している点や準会場（企業・学校等を会場として当該受検団体が自ら監督を行う会場）での試験実施運営状況の確認が不十分な点が指摘されている。またY検定試験合格者への情報提供に一部改善の余地があること等も指摘された。今回の第三者評価の試行より、課題として浮かび上がってきたこれらの事項を改善することにより、Y検定試験は、より信頼性が増すものと考えられる。

(3) 公益社団法人が主催するZ検定試験

検定分類が教育・学術に属する検定実施団体に対する調査は、「Z検定試験」を実施する公益社団法人に協力を依頼した。年間受検者数では中規模の検定である。「Z検定試験」に対する第三者評価の試行結果は以下のとおり。

表 1-8 Z検定試験の第三者評価試行結果

【試行結果】		Z検定試験							Z検定採点				大項目評価ライン					
大項目	中項目	評価項目(満点)							Z検定採点			得点	評価	評価1	評価2	評価3	不合格	
		重み別項目数			重み付得点				重み付得点			点数		90%	80%	65%	65%以下	
		◎	○	△	◎	○	△	合計	◎	○	△							
1 実施主体	1-1 組織	1	1	3	5	3	3	11	5	3	3	11						
	1-2 財務	0	2	2	0	6	2	8	0	6	1	7						
	1-3 その他	1	0	1	5	0	1	6	5	0	1	6						
		2	3	6	10	9	6	25	10	9	5	24	評価1	23	20	16		
2 実施内容	2-1 目的	1	0	0	5	0	0	5	5	0	0	5						
	2-2 内容	1	0	2	5	0	2	7	5	0	2	7						
	2-3 手段	1	0	0	5	0	0	5	5	0	0	5						
	2-4 その他	0	1	0	0	3	0	3	0	3	0	3						
		3	1	2	15	3	2	20	15	3	2	20	評価1	18	16	13		
3 実施手続	3-1 事前準備	2	1	1	10	3	1	14	10	3	1	14						
	3-2 試験実施	3	2	6	15	6	6	27	15	6	5	26						
	3-3 事後対応等	0	1	1	0	3	1	4	0	3	1	4						
		5	4	8	25	12	8	45	25	12	7	44	評価1	41	36	29		
4 検定結果の活用促進 継続的な学習支援	4-1 検定結果の活用促進	1	0	2	5	0	2	7	5	0	1	6						
	4-2 継続的な学習支援	0	0	2	0	0	2	2	0	0	2	2						
		1	0	4	5	0	4	9	5	0	3	8	評価1	8	7	6		
5 情報公開	5-1 実施主体	0	3	0	0	9	0	9	0	9	0	9						
	5-2 実施内容	0	3	0	0	9	0	9	0	9	0	9						
	5-3 実施手続	0	3	0	0	9	0	9	0	9	0	9						
	5-4 検定結果の利用促進	0	1	0	0	3	0	3	0	3	0	3						
	5-5 継続的な学習支援	0	1	0	0	3	0	3	0	3	0	3						
		0	11	0	0	33	0	33	0	33	0	33	評価1	30	26	21		
最終評価		11	19	20	55	57	20	132	55	57	17	129	A評価					
									得点率		97.7%							

Z検定試験に対する第三者評価の試行により集計された得点結果は、表 1-8 のとおり。Z検定試験については、今回の評価項目 58 項目の内、◎11 項目、○19 項目、△20 項目の合計 50 項目が評価対象となっている。表 1-8 の集計結果をもとに、前記述「4. 第三者評価の具体的な評価方法」の判定基準に基づき判定した結果、Z検定試験は「評価A」と判定された。

具体的な内訳は、必須項目である◎の 11 項目については、すべて合格の判定で、11 項目×5 点=55 点を獲得した。また○の該当 19 項目についても、すべて合格の判定で、19 項目×3 点=57 点を獲得した。また△の該当 20 項目の内、17 項目が合格の判定で、17 項目×1 点=17 点を獲得した。この結果、Z検定試験は、満点 132 点の内 129 点（得点率 97.7%）を獲得した。また大項目 1～5 ともすべて「評価 1」となり、A評価基準 85%（112 点）、及び前表 1-4 で記した不適合条件もクリアしたため、「評価A」と判定された。

【評価の概要】

Z検定試験の第三者評価の試行については、A評価基準 85%をクリアし、該当する項目に対しては、大項目 1～5 ともすべて「評価 1」、◎の必須項目は全項目合格したため、「評価A」と判定された。全般的な評価の傾向としては、試験実施運営を遂行する過程での判断や試験実施時にマニュアル外の事態が起こった際の判断等が一部の役職に集中しており、文書化された内規等の整備体制に一部改善の余地があること等が指摘された。また問題や試験実施運営上の管理体制で軽微な改善点が複数箇所指摘されたが、大きな課題は見あたらなかった。今回の第三者評価の試行より、指摘されたこれらの事項を改善することにより、Z検定試験は、より信頼性が増すものと考えられる。

4. 第三者評価の試行から抽出された課題等

- 1) 今回、検定試験の問題内容の妥当性と信頼性の評価に関しては試行の対象外としている。問題内容の第三者評価を行う場合は、試験問題の作成にあたる体制等の審査を検定試験の運営面を主とした定型的評価（外形的評価）で行い、問題内容の妥当性と信頼性の評価は別途行うことが考えられる。今後、これらの第三者評価をどのような体系で行うことが適切か等の検討が必要となる。
- 2) 複数の級等の実施を行っている検定試験の場合、級等の違いによって実施方法・対応が異なるが、今回の試行は級全体で捉えて総合的な評価を行った。例えば本人確認で、上位級は厳密に行う反面、奨励級に対しての本人確認が甘い検定試験がある。今後は取得することで社会的な価値が認められる級と自己の学習の到達度の確認等で利用される級に分けて考えて評価を行うこと等が考えられる。
- 3) 今回の試行では受検者数が一定程度ある検定試験の第三者評価を行ったが、試験の実施運営が小規模な検定試験やご当地検定等でまちおこし（町興し）的な要素が強い検定の審査を行う際は、審査項目を削減し審査費用を抑えた簡易版の評価を行うことが必要になってくること等が考えられる。
- 4) 今回の試行ではより広範囲な検定試験を適切に審査できるよう、最終評価を「評価A」「評価B」「評価C」「不適合」の4区分にし、大項目毎の評価を加える等、細分化したことにより、各項目の得点や配分点等が最終評価に与える影響が増した。審査のテクニカル的な面で、今後、大項目間での得点のバランスを修正（小項目の見直し）することや各小項目で合格としている「a」「b」で、差をつける（部分点等）ことにより第三者評価の信頼が高まると考えられる。
- 5) 第三者評価機関の人員体制は検定実施団体からの申請書類の精査人員、及び弁護士、会計士、有識者等で構成される実地審査委員・評価検討委員会等が不可欠と考えられる。第三者評価を実際に行う場合は、これら人員の時間と労力に相応するコストが発生する。

IV. 問題内容の評価の在り方に関する分科会（英語部会）について

1. 分科会（英語部会）開催と検討委員

「検定試験における第三者評価に関する調査研究」においては、検定試験の運営面を主とした定型的評価(外形的評価)だけでなく、各検定実施団体の実施している検定テストの質やそのテストを構成する試験問題としての質を評価することが考えられる。今回は民間の検定試験として多くの受検者を抱え、テスト問題の分析データが確保できる英語関係の検定試験を対象にして第三者評価に必要なツール(評価シート等)を作成するため分科会(英語部会)を開催した。

本件の検討は、テストの信頼性と妥当性の評価方法の仕組みを構築するため「英語・言語テスト」における専門家と「テスト理論」の専門家の参加が必須となるため、「英語・言語テスト」の専門家として藤田 保 先生（上智大学教授）と中野 達也 先生（東京都立白鷗高等学校）、「テスト理論」の専門家として服部 環 先生（法政大学教授）を構成委員とした。

開催日時：平成 28 年 1 月 29 日（金）午前 10 時 00 分～午前 11 時 30 分

開催場所：特定非営利活動法人 全国検定振興機構（東京都港区）

委 員：

中野 達也（東京都立白鷗高等学校）

服部 環（法政大学教授）

藤田 保（上智大学教授）

（敬称略、五十音順）

2. テスト作成手順に係る信頼性と妥当性

テストの質を評価するに際しては、信頼性と妥当性の概念が重要である。テストの信頼性とは、テストで得られた結果が、どの程度信頼のあるものであるかを示す概念となり、一般的には「信頼性係数」という指標をもって表される。テスト得点には必ず「誤差」が含まれるが、テスト得点に対する誤差の混入が少ない場合に、そのテストは信頼がおけると考えることができる。逆に言うと、テスト得点に対して真の得点の割合が大きくなる場合に、そのテストの信頼性（reliability）は高くなるということができる。これら信頼性の概念は、主に「テスト得点の安定性」と「測定の一貫性」の観点から評価される。

一方テストの妥当性とは、そのテストが『測定しようと考えている特性』を的確に捉えているかどうかを評価するものとなる。テストの観点からは、テストが適切に設計され、適切に使用されているかを評価することになる。妥当性の概念にはいくつかの側面があるが、一般的には「内容的妥当性」、「基準関連妥当性」、「構成概念妥当性」等が主なものとして知られている。

【用語解説】

信頼性

[テスト得点の安定性]

同一人が同じテストを複数回受検した時、どれ位同じ結果となるかの指標。この結果が安定していれば、信頼性の高いテストと考えられる。

[テストの一貫性]

テストは多くの問題項目から構成されるが、それらがテスト全体として同一の特性を測定しているかの概念を数量化して表したものの。

妥当性

[内容的妥当性]

テストで測定しようとしている特性を、テストに含まれる問題項目がそれに相応しいものとなっている。

[基準関連妥当性]

テストから得られた結果が、そのテストで測定しようとしている特性と関連性が強いと考えられる外的基準との相関によって評価される。

[構成概念妥当性]

テストから得られた結果の高低が、テストの構成概念（直接観察することはできないが、理論の構成上想定される概念のこと）の強弱を支持しているかどうかを表す。

3. 評価項目について

第三者評価に必要な問題内容に関するツール（評価項目シート）を作成するにあたり、以下の点を留意してまとめた。

- 1) テストの信頼性及び妥当性の概念の範囲は広いため、重要度が高く総合的に見て適正な評価が行える評価項目を取り上げてまとめる。
- 2) 多種多様な検定であっても汎用的に審査できるようテストの基本設計を中心とした評価項目にする。

表 1-9 検定試験の問題内容に関する評価項目シート

大項目	中項目	小項目			評価方法			評価軸	評価重み	
		項番	評価項目	評価の観点	聞き取り調査	媒体有無	その他			
1. テストの妥当性	1. テスト設計	1		測定目的を達成するために、現行の測定手法が妥当であることの明確な理念あるいは証拠がある。	○	○		検定目的	◎	
		2	測定・解答形式の基準が明確である。	記述式、選択式、口述式といった解答形式、解答時間、配点等が、仕様(テスト設計図・・・測定領域・測定形式・設問数・問題内容等が記載された一覧表等)通りに設計されている。	○	○		検定目的	○	
		3		問題内容の表現、語法、提示順序およびレイアウト、回答方法の説明のほか、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が具体的に示されている。	○	○		検定目的	△	
	2. 測定内容	4		測定しようとする特性(上位概念)が明確に定義されていて、テストがその定義に従って制作されている。	○			検定目的	◎	
			5 該	テストの測定対象が明確である。	測定しようとする特性(上位概念)が、複数の下位の特性で構成されている場合は、その構造が明確である。	○			検定目的	○
		6		出題領域(分野)が、当該検定試験で定義されたものに合致している。	○	○		問題内容	◎	
			7	テスト全体の測定内容が明確である。	出題範囲が、当該検定試験で定義されたものに合致している。	○	○		問題内容	○
		8		各問題項目は、当該問題項目で測定したい特性を適切に測定している。	○	○		問題内容	○	
			9	問題項目の測定内容が明確である。	原則として、1つの問題項目は1つの特性のみを測定するように設計されている(2つ以上の特性を測定する際は検定の目的に則り適切に設計されていることが必要)。	○			問題内容	○
			10 該	【日本の学校教育(教育・指導目的・進学・入試・単位取得等)で活用される検定試験の場合】学習指導要領等に準拠している明確な説明等がなされている。	学習指導要領と検定試験の構成概念が合致していることが明確である。	○	○		問題内容	◎
		11 該		学習指導要領の各事項が、検定試験の問題項目において適切に測定されている。	○	○		問題内容	○	
	3. 尺度構成	12	判定基準が明確である。	審査・採点・合否の基準が明確であるとともに、可視化した文書として存在する。	○	○		検定目的	○	
		13		テスト得点あるいは合否基準等の水準の一貫性が保持されている。	○			問題内容	○	
			14	尺度構成が適切である。	出題の困難度尺度が、当該検定試験で定義されたものに合致している。	○	○		問題内容	○
		15 該	【コンピューターを使って行う試験の場合】通常のテストと同様の結果が得られるような配慮がなされている。	同じ検定試験をPPTとCBT等、異なる媒体で実施する場合は、両者の結果の等質性が保証されていること。(システムの応答速度、受検者のコンピューターに対する知識等が異なっても同じ結果が得られるかどうか)	○	○		検定体制	○	
2. テストの得点信頼性	1. テスト設計	16	テストの信頼性に関する証拠が明確である。	信頼性係数が数量的に明示されている、あるいはそれに準ずる証拠が明示されている。	○	○	回答データ分析	検定体制	◎	
	2. 問題制作の信頼性	17	問題項目作成者の選定・トレーニングを適切に行っている。	具体的な人員選定・トレーニングの方式が定義されている。(問題の内容・表現が、テストの目的に照らし合わせて適切に作成するためのトレーニング等)	○	○		検定体制	△	
	3. 採点手続きの信頼性	18	採点手続きの設計が適切に行われている。	具体的な採点手続きに関する方式が定義されている。(特に主観的な評定によって採点を行うテスト[記述式試験・実技試験・面接試験等]において、採点基準が明確で、すべての評定者が理解し共有しているかどうか等)	○	○		検定体制	△	

※評価小項目「該」・・・該当する検定試験のみ審査する項目

【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特性を正確に測定している。
- 〔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）。
- 〔特性〕 認識能力、物理的能力、技術、知識、態度、人格特徴等のようにテストで測定される人間の属性特性の総称。
例：特性（上位概念）…『英語力』
下位の特性…英語力は『読む力』『書く力』『聞く力』『話す力』の4つの下位の特性で構成される。
- 〔尺度〕 テストで測られた特性について、一定のルールに基づいて定量的に示すために設けられる基準のこと。特性は基本的に一元的な尺度上に数値として表すことが求められる。
- 〔PPT〕 紙ベースのテスト（Paper and Pencil Test）
- 〔CBT〕 コンピュータベースのテスト（Computer-Based Testing）
- 〔信頼性係数〕 テスト得点の誤差が含まれる割合を1から除いた値。信頼性係数が高いほど、測定対象が変化しない限り、テスト結果は同様の結果になる。

※一般的にテスト理論での用語解説は専門用語が多いため、本用語解説は少々の誤差があってもわかりやすい表現を使用することとした。

4. 今後の課題等

- 今後、問題内容の質に関する第三者評価を行うにあたり、実際に表1-9で示した「評価項目」、「評価の観点」で試行審査等を行い平等・公平な審査が可能かどうかの検証、及び各小項目の審査結果と最終評価をどう結び付けていくかの仕組みの構築等が必要となる。
- 「信頼性係数」は、実技試験等で算出することが困難である。また「信頼性係数」の計算方法のノウハウがない、機器・技術・ソフト等の装備がない検定実施団体も多いと推測されるため、今後、外部団体へ計算・分析を依頼できる環境を整備する等、審査項目にするための条件を引き続き検討・検証していく必要がある。

V. 第三者評価の推進に資する研修会

第三者評価等による民間検定試験の質的向上（質や信頼性の確保）に関する、検定実施団体等を対象とした下記2回の研修会を開催した。

1. 第1回：効果的なテストの作成方法

開催日時：平成27年12月9日（水） 午後2時00分～4時00分

開催場所：国立オリンピック記念青少年総合センター センター棟 311室（東京都渋谷区）

講師：柴山直教授 東北大学大学院教育学研究科

テーマ：『効果的なテストの作成方法』

参加者数：82名（57団体）

内 容：

～効果的なテストの作成方法とは／問題内容の信頼性と内容の検証／合否判定基準について～

昨今、個人の価値観やライフスタイルが多様化する中で、多種多様な学習の機会が提供されている。各個人の学習成果が社会全体で幅広く通用し、評価され、活用できる民間検定試験に求められるテストとはどのような条件を備えたテストを指すのか。テストの作成・実施における一連の作業手順について、全体目標と計画作りから始まり、テスト内容の決定、問題作成、結果の集計、分析、判定など段階ごとに解説した。社会から信頼され、安心して受検できる検定試験の在り方について検討した。

2. 第2回：ICTを活用したテスト導入

開催日時：平成28年1月22日（金） 午後2時00分～4時00分

開催場所：国立オリンピック記念青少年総合センター センター棟 402室（東京都渋谷区）

講師：植野真臣教授 電気通信大学

テーマ：『ICTを活用したテスト導入』

参加者数：54名（37団体）

内 容：

～ICTを活用したテスト導入に向けて／ICTを活用したテストの現状／導入のメリット～

情報通信技術の進展により学習スタイルが劇的に変化する中、ICT活用教育の進展と相まって、民間検定試験にもICTを活用した試験実施を行うことにより受検者の利便性を高めることが考えられる。この現状を踏まえ、民間検定実施団体はその導入に向けてどのように対応していく必要があるか、ICTを活用したテスト実施に向けて、について考察した。また、テスト実施業務の標準化、テスト結果の分析や評価の簡素化、多様な実施形態のテストの実施が可能になるなどICTの活用による試験実施団体側のメリットについても解説した。

3. 主なアンケート結果

1) 試験問題の作成、信頼性・妥当性について

- 試験問題の作成、信頼性・妥当性については、テスト理論に基づいたテストの作成方法について初めて耳にした検定実施団体も多く「テストの作成に最低限必要な考え方が分かった」「試験問題の質的向上について考えることがなかった」「スタートラインにも立っていないと感じた」等の意見が寄せられた。
- 今回のテスト作成方法の研修内容を踏まえてテスト問題の作成にあたらうと考えている参加者の割合は72%（「全くそう思う」「まあそう思う」の合計）に達した。

2) 検定試験へのICTの活用について

- 検定試験にICTを導入した場合「信頼性が高まる」「論理的観点による評価が容易になる」等のように、ICTの必要性は理解しているものの「どちらともいえない」と考えている参加者の割合が37%と最多になった。このようにICT導入への課題として「導入コスト」「PC端末等の環境整備」が大きなハードルとなっていることが浮き彫りになった。コスト・環境整備をいかにクリアしていくかが今後の大きな課題となっている。

3) 検定試験における第三者評価について

- 「検定試験のみならず検定実施団体、全体の信頼性の確保に繋がると思う」等の意見で、第三者評価が必要と肯定的に考えている割合が第1回76%、第2回77%（「全くそう思う」「まあそう思う」の合計）となった。また「どちらともいえない」は第1回21%、第2回16%の割合であった。
- 問題内容の質を第三者評価機関が審査することに関する意見としては「外部機関の審査は必須で全ての検定試験に対して必要」と肯定的（「全くそう思う」「まあそう思う」の合計）に考えている参加者が最多割合ではあるものの、一方で「どちらともいえない」では「入試等、人生を左右する選択に利用される試験でなければ不要と思う」「問題が外部に漏れる恐れがある」等の意見が寄せられた。

資 料 編

【資料 1】 第三者評価の在り方に関する検討委員会及び開催経過

【資料 2】 研修会アンケート結果

【資料 1】 第三者評価の在り方に関する検討委員会及び開催経過

「第三者評価の在り方に関する検討委員会」委員

梅澤 健（弁護士・竹町法律事務所）

佐々木 康浩（株式会社三菱総合研究所・主席研究員）

惣脇 宏（京都大学大学院総合生存学館・教授）

中野 達也（東京都立白鷗高等学校・主任教諭）

萩原 民也（特定非営利活動法人日本語検定委員会・事務局長）

村木 英治（東北大学・名誉教授）（敬称略、五十音順）

「第三者評価の在り方に関する検討委員会」開催経過

第 1 回検討委員会

開催日：平成 27 年 12 月 15 日（場所：東海大学校友会館）

議 事：評価項目と評価方法の見直し、分科会（英語部会・試験問題の質の評価について）の設置等

第 2 回検討委員会

開催日：平成 28 年 3 月 1 日（場所：東海大学校友会館）

議 事：第三者評価の試行（課題等）、試験問題の質の評価項目等

（検討委員会での主な意見）

第三者評価全般について

- 第三者評価は良質な検定試験が広く普及するよう、育成・検定試験の改善を促す観点から行うことが大切。
- 第三者評価では「検定試験の運営・組織に関する審査」とともに、企業・大学入試での検定試験の活用等を視野に入れた場合には、検定試験問題の質を第三者が評価することにより、各検定実施団体の自律的な質の改善と向上が促されていくことが必要。
- 今回は実績が一定程度ある知識・技能を測る検定試験について第三者評価の試行を行った。今後は社会的に通用する検定試験と学習到達度の確認や趣味的な検定試験、及び小規模な検定試験やご当地検定等のあらゆる分野・規模・特徴を持った検定試験に対しての第三者評価システムの確立（コスト面での考慮も必要）が必要になっていくと思われる。また国の検定及び実技を測る検定試験の審査が必要になることも考えられる。

分科会（英語部会）について

- 英語の検定試験は 10 種類以上あり、学校ではどの検定試験を受検することが適切かわからない。試験の目的と活用方法が異なれば、学校と受検者は混乱する。学校で行った授業内容が適正に評価できる検定がどれかわかるような評価をしてもらいたい。
- 試験問題の質の評価では、試験の活用目的に見合った内容かどうかを審査する項目を設定することが必要。日本の学校教育（指導目的・進学・入試・単位取得等）で活用されている検定試験については学習指導要領との整合性を審査することが不可欠と思われる。

【資料2】 研修会アンケート結果

第1回：『効果的なテストの作成方法』

参加者数 82人 有効回答数 71通 (回収率 87%)
注：理由等のコメント最後尾に回答した方の選択番号を記載

(1) 今回の研修を受講して今後の貴団体のテストに活かそうと思えますか

①全くそう思う	②まあそう思う	③どちらともいえない	④あまりそうは思わない	⑤全くそうは思わない	⑥検定試験を実施していない	未回答
20	31	12	2	0	4	2
28%	44%	17%	3%	0%	6%	3%

【理由等】

- ・テストの信頼性・妥当性で具体的に活かす必要性を感じた。非常に理解し易い講演だった。(①)
- ・スタートラインにも立っていないと感じた。きちんと整備していきたい。妥当性・信頼性があるかどうかチェックしたい。(①)
- ・今まで、検定の質向上等について考えることがなかったので、今後考えるきっかけになった。(②)
- ・きちんとした理論展開がいかに必要かを感じた。(②)
- ・作文の評価など、デジタルで割切れないものをいかに評価し信頼性を確保していくか。見識を深めたい。(②)
- ・問題作成時の基準やどんな能力を受検者に求めているか、しっかり定める必要があると認識した。(②)

(2) 効果的なテストの作成方法についてどのような印象をお持ちになりましたか(自由記述)

- ・妥当性・信頼性という観点、尺度を持つことなど、知ってはいたがやはり重要だと再確認できた。
- ・今後、テストを作成していく上で、非常に参考になった。
- ・全体的な試験運営、及び作成について参考になった。
- ・問題作成に最低限必要な考え方、人員等が分かり、これからの弊法人の運営方針に活用できそうと思った。
- ・目的や意義について必要性を感じたが、実際にどのように取り入れていくのか、取り入れればどのような成果がでるのか、と、言うところは、正直まだイメージ出来ない。

(3) 貴団体で実施している検定試験について、自己評価を実施していますか

①「検定試験の自己評価シート」を活用し自己評価を行っている	②団体独自に自己評価を実施している	③実施していない	④検定試験を実施していない	未回答
22	15	25	5	4
31%	21%	35%	7%	6%

【理由等】

- ・質的向上の確立を目指す上では、当然のことと思われる。(①)
- ・検定試験の品質保証や信頼性の確保に必要である。(①)
- ・他業務に追われ自己評価している余裕がない。(③)

- ・人的・費用的負担が大きい。 (③)
- ・自己評価という考えを今回初めて聞いた。 (③)

(4) 検定試験の信頼性を担保するために、第三者評価機関が検定試験を評価する仕組みが考えられますが、このような仕組みは必要だと思いますか

①全くそう思う	②まあそう思う	③どちらとも いえない	④あまりそうは 思わない	⑤全くそうは 思わない	未回答
26	28	15	1	0	1
37%	39%	21%	0%	1%	1%

【理由等】

- ・第三者評価がなければ、その検定実施団体の中で閉鎖的に運営されるだけなので危険だと感じた。 (①)
- ・試験の信頼性の高低があることがもっと広く知れ渡る必要がある。 (①)
- ・検定の規模によってある程度評価が決まってしまう、普及や導入が小規模の検定にとっては難しいと感じる。第三者によって評価し、認知度を高めるなどしていただける機会があれば良いと思う。 (①)
- ・問題の質・受検の公平性を担保するためにも第三者評価機関の設置を検討したいと思う。 (①)
- ・自己評価の取組みの先に第三者による客観的な視点が加わることは、検定試験のみならず検定実施団体、全体の信頼性の確保に繋がることと思う。 (①)
- ・コストを考えると、第三者評価を受ける検定は基準を設けて絞り込んでよいと思う。 (②)
- ・検定試験の活用を広げるためには、第三者による定数的な格付けが行われることが有効と思う。 (②)
- ・目的を見失った検定になるのを防ぐ必要がある。 (②)
- ・技術試験を評価されるのはどうか考えてから。 (③)
- ・自分達の検定に必要だとは今のところ考えていない。 (③)

(5) 検定試験に対する信頼性を担保するために、検定試験の問題内容の信頼性や妥当性について、第三者評価機関が審査することは必要だと思いますか

①全くそう思う	②まあそう思う	③どちらとも いえない	④あまりそうは 思わない	⑤全くそうは 思わない	未回答
18	25	26	1	0	1
25%	35%	37%	1%	0%	1%

【理由等】

- ・今後、大学入試や、就職の場面で弊協会の資格が導入されることがあれば、外部機関の審査は、必須であると思う。全ての検定実施団体に対し必要なのではないか。 (①)
- ・客観的視点で審査することで、内部からは見えない点も見えてくると思う。 (②)
- ・当事者（作成側）では信頼性や妥当性が測りづらいし、そのような技術や知識も全ての検定実施団体が持ち合わせていないと思うため。 (②)
- ・検定試験の活用を広げるためには、第三者による定数的な格付けが行われるのが有効と思う。 (②)
- ・最初は検定実施団体が自己分析を行い公表することが普及し、その後考えるのがよいと思う。 (③)
- ・入試等、人生を左右する選択に利用される試験でなければ第三者評価は不要と思う。 (③)

第2回：『ICTを活用したテスト導入』

参加者数 54名 有効回答数 44通 (回収率 81%)
注：理由等のコメント最後尾に回答した方の選択番号を記載

(1) 今回の研修を受講して、今後の貴団体の検定試験にICTを導入しよう(活用しよう)と思いますか

①全く そう思う	②まあ そう思う	③どちらと もいえない	④あまりそう は思わない	⑤全くそう は思わない	⑥既に 導入済	⑦検定試験を実 施していない	無回答
3	4	15	5	0	10	5	2
7%	9%	34%	11%	0	23%	11%	5%

【理由等】

- ・大学入試改革の流れの中で我々予備校として必須課題。(①)
- ・必要性を感じるが、実際の準備のハードルは高い。(②)
- ・実際に導入するにはコスト等の敷居が高い。(③)
- ・検定の性質上、ICTの導入が難しく大きな課題である。(③)
- ・弊団体の実施している資格試験は対象となる受検者数が少なくICT導入コストを吸収できないと思う。(④)

(2) ICT化により検定試験の信頼性は高まると思いますか

①全くそう思う	② まあそう思う	③どちらとも いえない	④あまりそう 思わない	⑤全くそう は思わない	無回答
8	20	13	1	0	2
18%	45%	30%	2%	0	5%

【理由等】

- ・IRTを使用し信頼性をコントロール出来る。(①)
- ・何時受けても誤差が少なく同じ結果になるということは、信頼性が高まると思う。また、数字でそれが示され透明性もある。ただ、一般に伝えるには工夫が必要。(②)
- ・第三者評価が伴うことでより高まるだろう。受検者からの信頼を得るには受検者のICTに関する理解が必要になりそうである。(②)
- ・客観的・論理的観点による評価が比較的容易になる。受検者への説明・理解の獲得が、必須だが、これはICT化で達成される。(②)
- ・古典的テストでは生徒の能力を正しく判定しきれてない事を痛感している。(②)
- ・最初、信頼性は高まるが、攻略法がいずれ開発され対策とのイタチごっこになると思われるため。(②)
- ・ICT化の技術課題が理解出来た。実現的なコストを含めた体制構築が絶対的に必要。(③)

(3) 貴団体の検定試験でICTを活用するとしたらどのような課題がありますか(自由記述)

- ・検定の規模。コスト。検定内容など。
- ・「2万項目」の作問コスト。項目の継続的作成。
- ・現場の社員に対して認知させ生徒に理解させること。PC端末など環境整備。
- ・指導者・学生に対する説明体制をいかに作っていくかが課題。

- ・多くの検定実施団体の声として項目を大量に作成するコストが問題になるという話を良く聞く。コストをいかに抑えるかが課題。
- ・コスト。会場確保。機材（PC、ネットワーク）の確保、セキュリティー、不正防止。本人認証。
- ・統一のテストにおいてC B Tで同一にならない点。問題が変わるため同じ状況でないと思われる場合がある。
- ・思考力を測る論述・記述系の検定なのでいかにI C Tを導入するかが大きな課題である。
- ・メリットの理解。試験という重要なものを新しい方法にすることへの抵抗感。
- ・本人のみで受検したか確認できない。アイテムバンクの設計。

(4) 検定試験の信頼性を担保するために、第三者評価機関が検定試験を評価する仕組みが考えられますが、このような仕組みは必要だと思いますか

①全くそう思う	②まあそう思う	③どちらともいえない	④あまりそうは思わない	⑤全くそうは思わない	無回答
11	23	7	1	0	2
25%	52%	16%	2%	0	5%

【理由等】

- ・「私は～が出来ます」といってもそれがどの程度出来るかを測る仕組みが検定であり、その検定がどのくらいの信頼性があるかを積極的に評価してもらいたい。(①)
- ・現状はハイスタータスな試験でも質が悪い。又は、質をどのように担保しているかが見えないものが多い。(①)
- ・受検をしようとする時の判断基準にもなるだろうし、検定、教育の分野自体の発展にもなるだろう。(②)
- ・レベルアップのため。(②)
- ・全ての検定試験で行う必要はなく、重要なもののみで良いと思います。(②)
- ・専門分野の方をお願いしている。(③)

(5) 検定試験に対する信頼性を担保するために、検定試験の問題内容の信頼性や妥当性について、第三者評価機関が審査することは必要だと思いますか

①全くそう思う	②まあそう思う	③どちらともいえない	④あまりそうは思わない	⑤全くそうは思わない	無回答
5	21	15	1	1	1
11%	48%	34%	2%	2%	2%

【理由等】

- ・問題内容の質の評価は難しいが、作問の背景や社会の変化への対応、作問委員はどのような方かなど、調査し評価して広く社会に検定取得を促してほしい。(①)
- ・但し、検定実施団体のポリシーによるのでは？(②)
- ・必要であるが現実にはコストとの折り合いが必要になると思います。(②)
- ・専門性が高いため。(③)
- ・必要だと思うが内容、やり方が問題。(③)
- ・問題の中身について審査する場合問題が外部に漏れることはないか、また基準を満たさないものがあつた場合プールできる問題の不足に繋がることが心配。(③)