

機関番号：13801

研究種目：基盤研究（B）

研究期間：2008～2010

課題番号：20330020

研究課題名（和文）テスト理論による法学試験の質の向上の研究

研究課題名（英文）A Study for Improving Legal Test based on Test Theory

研究代表者

藤本 亮 (FUJIMOTO AKIRA)

静岡大学・法務研究科・教授

研究者番号：80300474

研究成果の概要（和文）：法律学試験の択一式試験において、これまでテスト理論の観点からのテストの質の検討は行われていない。古典的テスト理論における識別力や信頼性といった観点で問題項目を検討するだけでもより質の高い、すなわち能力を正確にかつ適切に測定するテストを実現できることが明らかとなった。さらに、項目反応理論を応用することで、より公平な法律学資格の試験設計が可能となることが示唆される。

研究成果の概要（英文）：Legal examinations in the form of multiple choice tests have not been reviewed with the test theory. This study shows that even within the classical test theory framework, checking the test with the discriminating power of each item and some measures of reliability will enable to administer more accurate and appropriate test. Furthermore, our experimental survey data indicates that the modern test theory based on the Item Response Theory allow us to design fairer legal examination.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	2,100,000	630,000	2,730,000
2009年度	2,700,000	810,000	3,510,000
2010年度	1,500,000	450,000	1,950,000
年度			
年度			
総計	6,300,000	1,890,000	8,190,000

研究代表者の専門分野：法社会学

科研費の分科・細目：法学・新領域法学

キーワード：①法学教育・法曹論 ②法知識 ③テスト理論 ④IRT ⑤法学試験

1. 研究開始当初の背景

(1) 本研究の方法論である項目反応理論は、合衆国の LSAT、法科大学院入学試験判定材料としても用いられている TOEFL や TOEIC のみならず、日本の医学教育モデル・コア・カリキュラムにもとづいて臨床実習開始前に全ての医学部で実施されている共用試験（歯学、看護学等でも導入中）でも利用されている。

(2) 法学分野の資格試験では、テスト理論に

もとづく試験の質の管理がなされていない。各種の法律関係の資格試験は数多く毎年実施されているが、そこでは得点による能力値の表示ではなく、単に合否判定がなされている。これ自体が不合理なわけではない。「等化」がなされていないため同年度の受験者間の能力は合否判定できても、その判定された能力が異なる年度間で同一である保障はない。つまり、試験の公平性という点で実は重大な欠陥が存在するのである。

(3) テスト理論に基づき心理測定や教育学分野で当然のごとく用いられている Item Response Theory (IRT=項目反応理論) を本格的に法学分野に応用し、法学試験や法知識の測定の精度を高める可能性を探ることにある。これにより異なる測定機会(法学試験の異なる実施回相互や複数の法知識調査)の間の比較が可能となる。

2. 研究の目的

テスト理論に基づいてテスト間の「等化」がされると、受験したのがいつ実施された試験であっても(=複数のテスト・フォーム間)同じ得点であればそれは同じ能力を示すものとして扱われるようになっている。したがって、一種の資格試験(いつ受けても同一の能力とみなされる)として扱うことができる。その可能性を法律テストの実データと、それに基づく実験データによって示すことが研究の目的である。

3. 研究の方法

(1) 古典的テスト理論と項目反応理論を応用し、多肢選択式法学試験の質について検証する。古典的テスト理論の観点からは、困難度(正解率)、識別力、信頼性などの指標を用いて分析する。IRTの特性は下記のようにまとめられる。

- ① 各項目の特性(困難度や識別力)を表すパラメータが、特定の受験者群とは独立に定義される。
- ② 各受験者の能力尺度値が当該受験者の回答した項目群とは独立に定義される。
- ③ 測定精度を受験者ごとに表現できる。
- ④ 複数のテスト・フォーム間の結果についても相互比較可能な共通尺度上に表すこと(得点の等化)ができる。

(2) 法学検定委員会の協力により、匿名化された2006-08年度の「法科大学院既修者試験(既修者試験)」と「法学検定試験2~4級の実試験データ」の提供を受け、古典的テスト理論の枠内で試験問題(項目)を検討した。通過率(正解率)のみならず識別力や信頼性の点からも検討をした。

(3) さらに、それらの項目特性値をベースとして、憲法・民法・刑法の3科目につき問題分野のばらつきを確保しながら、識別力の一定程度高い問題を中心に問題セット構成した。また、オリジナル出題年ごとに通過率が低くなっていくように5問ずつ問題を選定した。この問題セットを用いて、ウェブ調査により、古典的テスト理論の枠内での問題

改善の可能性を検証した。調査対象者は、WEB調査サイトへの事前登録者のうちから、一定の法律試験に合格しているあるいはそのための受験勉強を一定期間行っているという者をスクリーニング質問群によって選別することとした。調査実施は中央調査社に依頼し、250ケースのサンプルを得た(調査協力者に対して後日解答と解説、正答数得点をウェブ上で公開した)。

(4) ウェブ調査データを用い、IRTによる等化が可能であることを検証した。ウェブ調査は

4. 研究成果

(1) 「法科大学院既修者試験(既修者試験)」の試験データ提供を受け、2006-08年度分の困難度(正解率)と識別力の算出をした。その結果から、極端に識別力が低い問題項目を中心に、内容的妥当性を検討した。識別力が低い問題は、法律学の分野を問わず、法的知識の枝葉末節にこだわりすぎていたり、問題文の文意が注意深く読まないとい義的に定まらなかったりといった傾向にあることが明らかとなった。さらに、正誤問題と多肢選択式では、後者の識別力が高い傾向にあることも確認できた。【図1】

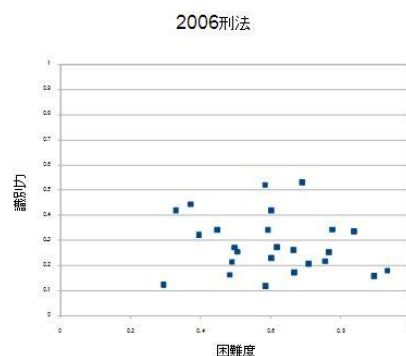


図1 2006年刑法問題項目の困難度と識別力

(2) 信頼性についてはクロンバックの α を信頼性係数 ρ の推定値として用いて検討した。その結果、主要7科目(憲法・民法・刑法・民訴法・刑訴法・商法・行政法)のうち α が0.6を越えるのは憲法と民法だけであることがわかった【表1】。全般に各科目の問題セットの一貫性という点では信頼性が低いということである。他方で、科目内の知識構造が多面的になっているともいえ、法知識の構造を、その法体系との関連で検討するという新しい課題が明らかとなった。

【表1】既修者試験年度ごとの科目別信頼性係数

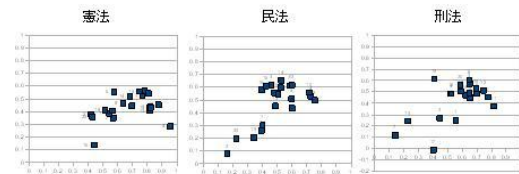
	全体	憲法	民法	刑法	民事訴訟法	刑事訴訟法	商法	行政法
2006	0.86	0.57	0.70	0.62	0.41	0.46	0.32	0.38
2007	0.83	0.60	0.66	0.39	0.47	0.39	0.46	0.32
2008	0.83	0.60	0.62	0.41	0.49	0.35	0.35	0.50

信頼性係数=複数の項目が、全体として、同一の能力を測定しているかを測る指標の一つ。項目数が大きければ値が大きくなる傾向がある。正解したかどうかで信頼性係数を計算している。「全体」は125項目、各科目は25項目となる。

(2) 法学検定試験委員会事務局より提供を受けた法学検定試験2級、3級、4級の実試験データを用い、古典的テスト理論にもとづき、信頼性と妥当性を検討した。その結果、問題セット全体としての信頼性は一定程度の高さにあった【表2】。4級から2級へと難度をあげるにつれ、信頼性係数はやや低下する傾向にある。項目別にみると、識別力が0.20未満と低い項目が科目ごとに少数ながら含まれていた【図2】。これらの項目について、そこで問われている法体系上の分野や判例知識の有無について整理を行った。識別力が低い問題には、肢別に異なる判例知識を問う問題であったり、学説の微細な違いを問うたりするような問題が目立っている。

【表2】法学検定三級の科目別信頼性係数

Chronbach's α	項目数	2006	2007	2008
憲法	20	0.759	0.757	0.739
民法	20	0.820	0.798	0.738
刑法	20	0.763	0.794	0.788
三科目	60	0.906	0.906	0.887
N				
憲法		4032	3479	3192
民法		5383	4692	4094
刑法		3199	2800	2397
三科目		2323	2035	1848



【図2】2006年法学検定3級項目別通過率（横軸）と識別力（縦軸）

(3) ウェブ調査では、実試験とWEB調査の受験者集団の能力の違い、解答姿勢の違い等により、民法・刑法についてはWEB調査において能力の識別がほとんどできていなかった。すなわち、民法・刑法分野では、調査対象者の法律学習程度に比して難しい問題であったためか、通過率平均が期待値程度にとどまり、十分な分散を得ることができな

かったため、以後の分析は行えなかった。通過率について検討した結果では、憲法ではオリジナル出題の結果と強く相関する分布となっている。憲法については項目群（ここではオリジナル出題年ごと）の通過率の違いがあっても、能力（ここでは全項目による正答数得点ベース）が一定程度識別された【表3】。これは実試験項目から識別力の高い項目を選ぶことで（古典的テスト理論の範囲内でも）より質の高い試験を作成することができる可能性を示している。

【表3】調査-能力は問題セットの通過率にかかわらず判定されているか？

		憲法正答数(三層)					
		<=5		6-7		8+	
		度数	行のN%	度数	行のN%	度数	行のN%
憲法06正答数(三層)	<=2	56	81.2%	11	15.9%	2	2.9%
	3-3	17	37.8%	20	44.4%	8	17.8%
	4+	3	6.5%	8	17.4%	35	76.1%
	合計	66	100%	69	100%	45	100%
憲法07正答数(三層)	<=1	54	77.1%	12	17.1%	4	5.7%
	2-2	20	42.6%	18	38.3%	9	19.1%
	3+	2	4.7%	9	20.9%	32	74.4%
	合計	66	100%	69	100%	45	100%
憲法08正答数(三層)	<=1	67	72.0%	19	20.4%	7	7.5%
	2-2	9	25.0%	12	33.3%	15	41.7%
	3+	0	0%	8	25.8%	23	74.2%
	合計	76	100%	39	100%	45	100%

(4) さらに、問題項目セットを模擬的に複数作ることで、IRTによる等化を行い、能力値の検討を行った。その結果、法律試験のような複雑な知識の測定においても、等化によるテストの質の向上が可能であることが示唆された。具体的には、ウェブ調査参加者を共通受験者として、その受験者能力尺度値を素得点により推定し、年度間等化を行い、非実試験項目（実試験では用いられたが、ウェブ調査には含まれていない項目）の項目特性値を共通尺度化した。この結果、実試験結果の年度間比較が可能となり、また受験者集団の能力分布の違いも観察することができた。

したがって、十分な項目プールが存在すればではあるが、法律学テストにおいても、現代的テスト理論（項目反応理論）を応用した運用ができ、「競争・選抜」型の一回限りのテストから、「資格・能力測定」型の比較可能なテストとして実施することができることがあきらかとなった。これは、法学検定試験の結果を「〇級合格/不合格」の二値ではなくて、たとえば「法学能力値=〇〇〇点」という形で示すことができることを意味する。

しかし、このような項目プールを前提とした資格試験の運用には、一回限りで問題を完全公開する競争試験を重用する「日本のテスト文化」が障害となっていることも明らかとなった。それでもなお運用方法に工夫を凝らし法学検定試験を能力値を示すことができるテストとして実施していく意義は大きい。

(4) またウェブ調査の分析において、解答時間を測定した結果を分析したところ、不誠実な回答を析出することができた。全データ

(250 ケース)、解答に要した総時間が短いケースを除いたサンプル (183 ケース)、ウェブ画面が切り替えとなる 5 問単位で回答時間が短いケースを除いたサンプル (159 ケース) を比較した。画面単位の回答時間で不誠実回答を析出した場合に、平均正答率、識別力ともにやや上昇することが観察された。回答時間測定は煩雑に思えるが、ウェブ調査であれば項目枚であっても容易に可能である。今後社会調査やマーケティングで多用されていくと思われるウェブ調査の精度を上げるためのみならず、項目の質を厳密に測定するためにも解答 (回答) 時間を個別項目ごとに調査することの意義が明確となった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

野口裕之・倉元直樹 (2009) 「外国語としての日本語能力を支えるテスト理論」クオリティ・エデュケーション 2, 93-118 頁。

藤田政博 (2009) 「裁判員時代の法廷用語」言語 38 巻 9 号、16-19 頁。

[学会発表] (計 4 件)

野口裕之 「テストの分析理論－項目応答理論 (Item Response Theory)」 「テスト理論による法学試験の質の向上」研究会、2009 年 5 月 18 日、静岡大学。

藤本亮 「法学択一試験の質の検討－テスト理論の応用可能性」日本法社会学会 2010 年度学術大会、2010 年 5 月 8 日、同志社大学。

Akira Fujimoto "The Crisis of the Bar Exam and Legal Education in Japan" Law and Society Association 2010 Annual Conference, May 29, 2010, Chicago, Illinois, USA.

藤本亮 「ウェブ調査による法知識測定の試み」情報ネットワーク法学会 2010 年度研究大会、2010 年 12 月 11 日、成城大学。

[図書] (計 2 件)

藤田政博・藤本亮 (2008) 「法学既修者試験についての分析」(第 9 章) 適性試験委員会編『法科大学院統一適性試験テクニカル・レポート 2007』(JLF 叢書 Vol. 15)、97-119 頁。

藤本亮 (2010) 「法知識とその測定の課題～

拡散性と系統性」(第 2 章) 松村良之・村山眞維編『法意識と紛争行動』東京大学出版会、23-50 頁。

6. 研究組織

(1) 研究代表者

藤本 亮 (FUJIMOTO AKIRA)
静岡大学・法務研究科・教授
研究者番号: 80300474

(2) 研究分担者

吉川 真理 (YOSHIKAWA MARI)
静岡大学・法務研究科・教授
研究者番号: 70289750

宮下 修一 (MIYASHITA SYUICHI)
静岡大学・法務研究科・准教授
研究者番号: 803777121

小谷 順子 (KOTANI JUNKO)
静岡大学・人文学部・准教授
研究者番号: 403599721

野口 裕之 (NOGUCHI HIROYUKI)
名古屋大学・教育発達科学研究科・教授
研究者番号: 601148151

藤田 政博 (FUJITA MASAHIRO)
関西大学・文学部・准教授
研究者番号: 603771401