

大学入試における共通テストの 複数回実施は実現可能か

—日本のテスト文化やこれまで
見送られてきた理由などからの検討—

石井 秀宗

—<要 旨>—

大学入試における共通テストの複数回実施は、検討が始められてから30年以上経過してもまだ実現されていない。共通テストの複数回化は受検者にとって利益があるとされながら、日程上の問題や技術的な問題等を理由に見送られ続けている。日本のテスト文化も共通テストの複数回実施には適していない。しかし、近年実施されている、もしくは実施予定のいくつかのテストは、共通テストの複数回実施を可能にしうる要素を含んでいるように見受けられる。そこで本稿では、日本のテスト文化、これまで複数回実施が見送られてきた理由や課題、近年のテストの状況等の観点から、複数回実施の実現可能性について検討した。その結果、(1) 選抜試験にするのか資格試験にするのかというテストの活用法の問題、(2) テスト得点の過度な重視とそれに起因する問題、(3) テストを合理的に設計・開発するための知識基盤が脆弱であるという問題、という3つの課題が依然残る要因として指摘された。

1. はじめに

2016 (H28) 年3月に出された高大接続システム改革会議最終報告書において、大学入試における共通テストの複数回実施は、またも見送られることとなった(高大接続システム会議 2016)。共通テスト(大学入試センター試験)を同一年度内に複数回実施することは、木村ほか(2017)によれば、1985 (S60) 年に出された臨時教育審議会第1次答申で言及され、

また、1997（H9）年に出された中央教育審議会第二次答申などにおいても論じられており（中央教育審議会 1997）、30年以上の歳月を経ても実現には至らなかったことになる。

共通テストの複数回化は、受検機会が増え一回勝負の状況を改善するので受検者にとって利益があるとされながら、日程の問題や技術的な問題等を理由に見送られ続けている。

確かに、日本のテスト文化は、大学入試など人生を左右するようなハイ・ステークスなテストの複数回実施には不向きな面がある。しかし、近年実施されている、もしくは実施予定のいくつかのテストは、共通テストの複数回実施を可能にしうる要素を含んでいるように見受けられる。そこで本稿では、日本のテスト文化や、これまで複数回実施が見送られてきた理由及び課題、近年のテストの状況等の観点から、大学入試における共通テストの複数回実施の可能性について改めて検討することにする。

まず次節において、直近の大きな入試改革であった英語リスニングテストの導入、記述式問題の導入の経緯について概観する。次に、日本のテスト文化、続いて、複数回実施が見送られてきた理由や課題について整理する。そして、近年実施されている、もしくは実施予定のテストが、共通テストの複数回実施実現にあたってのどのような問題を突破しているかを検討し、依然残る要因は何かを明らかにする。

なお本稿では、テスト・スタンダード（日本テスト学会 2007）に則り、「受験」ではなく「受検」、「多肢」ではなく「多枝」という表記を用いることにする。

2. 英語リスニングテスト及び記述式問題導入の経緯

2006（H18）年度入試から、大学入試センター試験に英語リスニングテストが導入された。内田・大津（2013）によれば、共通テストへのリスニングテストの導入は、国立大学協会によって1970年代はじめから検討され、大規模な実施調査まで行われていた。しかし、1977年に始まった共通第1次学力試験では結局、リスニングテストは導入されなかった。その後も、大学入試センターにおいて、リスニングテストの導入に向けた研究や検討が続けられたが、1997（H9）年の時点では、設備面、技術面、実施面などの問題があるとして、当分の間は導入困難であるとされていた。

ところが、2年後の1999（H11）年に出された中央教育審議会答申（中

中央教育審議会 1999) や、翌年の大学審議会答申(大学審議会 2000) では、大学入試センター試験において「リスニングテストの実施に向けて検討を進めることが必要」「早急にリスニングテストの導入を図ることが必要」と書かれ、平成 14 年度の文部科学白書では、「リスニングテストの導入など様々な改善方策が提言されており、文部科学省ではこれらの提言について、実現可能なものから随時施策化していく」というように、共通テストにおけるリスニングテストに関する記述が載った(文部科学省 2003)。そして、翌平成 15 年度版の文部科学白書(文部科学省 2004) で、「『外国語』教科の『英語』科目にリスニングテストを導入することとしました」と明言されるに至り、瞬く間にリスニングテストが導入されることとなった。その後、数回の試行テストを経て、2006(H18) 年度入試から大学入試センター試験に英語リスニングテストが導入されたのである。このようにリスニングテストは、急転直下の導入が決まり、瞬く間に実施に至った感はあるものの(内田・大津 2013)、当初の構想から考えれば、数々の研究、検討が重ねられ、30 年以上かかってようやく導入決定に至っている。

これに対し記述式問題は、異例とも言える速さで導入が決まっている。記述式問題についての記述を遡ると、2012(H24) 年に設置された中央教育審議会高大接続特別部会の第 2 回会議で言及された記録がある(中央教育審議会高大接続特別部会 2012)。そして、2014(H26) 年の中央教育審議会答申において、共通テストに記述式問題を導入する方向性が示され(中央教育審議会 2014)、2016(H28) 年の高大接続システム改革会議最終報告書で、2020(H32) 年度入試から共通テストにおいて記述式問題を導入するとされている。

記述式問題の導入は、構想から導入決定まで長く見積もっても 4 年であり、30 年かかったリスニングテスト、30 年以上経っても実現しない複数回実施に比べ、非常に速い速度で導入が決定されている。これには、情報技術や通信技術が既に発展しているということも寄与しているが、多枝選択式テストでは学力は正しく測れず、記述式テストなら「真の学力」が測れるとの考えが広く受け入れられているという、日本のテスト文化も大きく影響していると考えられる。

3. 日本のテスト文化の特徴

日本のテスト文化については、Arai and Mayekawa(2005)、前川(2015)、

光永 (2017)、柴山 (2008) などの研究でその特徴が示されている。ここでは、これらの文献を参考にしながら、大学入試における共通テストを中心に、日本のテスト文化について整理する。本稿では、日本のテスト文化を「テストの実施」「テスト得点」「テスト問題」「テスト仕様」「テストについての理解」という5つの視点から整理する。結果を表1に示す。

表1 日本のテスト文化の特徴

テストの実施に関して

- (1) 年に1回、一斉に実施される
- (2) 各科目、問題冊子は1つであり、同一科目受検者は、全員同じ問題冊子に解答する
- (3) 共通テスト実施から出願までの期間が極めて短い

テスト得点に関して

- (4) テスト得点として素点が用いられる
- (5) 満点 (100点) や0点に特別な意味を見出す
- (6) テスト得点が受検者に通知されるのは入試が終わった後である
- (7) 共通テストの自己採点結果を収集して、合否予想が大々的に行われる
- (8) テスト得点の有効期限が示されていない

テスト問題に関して

- (9) テスト問題は実施後公開される
- (10) テスト問題を非公開にしても、問題を記憶するアルバイトを雇うなどして、組織的に復元される
- (11) 問題は毎回新しく作成される
- (12) 問題は予備調査されることなしに、実際の試験で用いられる
- (13) 一度出題された問題は使い捨てられ、繰り返しテスト問題として利用されることはない

テスト仕様に関して

- (14) 各科目のテストの設計、問題作成、編集等を、別に職を持つ当該学問領域の専門家が、兼業として行う
- (15) 大問形式が多く利用される
- (16) 多枝選択式問題1問あたりの解答時間が2~4分と長めに設定されている
- (17) 大規模テストとクラスルームテストの違いが明確でない
- (18) 選抜テストと資格テストの違いが明確でない

テストについての理解に関して

- (19) テスト得点を過度に重視する
 - (20) テスト問題の領域代表性を考えず、結果を過度に一般化する
 - (21) 多枝選択式テストでは学力は正しく測れず、記述式テストなら「真の学力」が測れると信じている
 - (22) テスト理論が普及していない
-

注：Arai and Mayekawa 2005、前川 2015、光永 2017、柴山 2008などを参考に作成

まず、テストの実施に関しては、「年に1回、一斉に実施される」「各科目、問題冊子は1つであり、同一科目受検者は、全員同じ問題冊子に解答する」という特徴がある。テストは公平・公正であることが求められるが(西郡・倉元 2007)、日本においては全員が同じ問題を同時に受検するこ

とが公平・公正だと考えられているのである。また、「共通テスト実施から出願までの期間が極めて短い」のも特徴である。例えば、フランスのパカロレアやドイツのアビトゥーアなどは資格型の試験であり、取得した大学入学資格は生涯有効である（大桃ほか編 2007 など）。

テスト得点に関しては、「テスト得点として素点が用いられる」「満点(100点)や0点に特別な意味を見出す」などの特徴がある。TOEFL等のテストでは、標準化された得点と違って、素点ではなく、解答パターンから推定された能力値をテストの得点とする。その値は受検者に早期に通知され、各受検者はその結果を出願先大学に提出する。これに対し日本は、「テスト得点を受検者に通知されるのは入試が終わった後である」ので、「共通テストの自己採点結果を収集して、可否予想が大々的に行われる」。また、基本的にテスト得点は当該入試でしか有効でないため、「テスト得点の有効期限が示されていない」。

次に、テスト問題に関してであるが、「テスト問題は実施後公開される」。センター試験や個別試験の問題と解答が、翌日には新聞に掲載されるし、一般書店で過去問題集が販売されている。また、もし「テスト問題を非公開にしても、問題を記憶するアルバイトを雇うなどして、組織的に復元される」ので、問題を非公開にすることは非常に難しい。問題がすべて公開され、過去問として丹念に学習されてしまうので、「問題は毎回新しく作成される」ことが必要になり、「問題は予備調査されることなしに、実際の試験で用いられる」「一度出題された問題は使い捨てられ、繰り返しテスト問題として利用されることはない」という事態になる。実は、この慣習がテストの複数回実施を妨げる要因の1つになっている。

4 つめはテストの仕様に関してである。日本では「各科目のテストの設計、問題作成、編集等を、別に職を持つ当該学問領域の専門家が、兼業として行う」ことが多く、「大問形式が多く利用される」。大問形式とは、1つの出題テーマがあり、それに関連していくつかの問題が連なる問題形式を言う。1つ1つの問題が独立した小問形式のテストは品がないなどと忌避される傾向もあり、たとえ選択式テストであっても、「多枝選択式問題1問あたりの解答時間が2~4分と長めに設定されている」。また、「大規模テストとクラスルームテストの違いが明確でない」「選抜テストと資格テストの違いが明確でない」など、対象や目的、用途の違いを考慮せず、どの場合も同じようなテストを使っている。例えば、授業である単元を学んだ後に児童生徒の理解状況を把握するのが目的であれば、その単元に焦点をあ

てたクラスルームテストを実施し、指導に当たった教師が採点することによって、各児童生徒の理解状況を把握することができる。一方、国全体の教育施策を考えるために各単元の修得状況を把握するためには、単元数が多く、とて1つのテストで全単元をカバーすることはできないので、問題のある程度ずつ入れ替えた複数の問題冊子を作成してテストを実施し、受検者全体の解答を分析することで、その目的を果たすことが可能となる（日本テスト学会編 2010）。しかしながら、全国学力・学習状況調査は全員に同じ問題を解かせており、その冊子に含まれるごく狭い範囲の議論だけしかできないのが現状である。

最後は、テストについての理解に関してである。日本では「テスト得点を過度に重視する」傾向が強いが、これは、日本のテスト制度は中国の科挙に近く（宮崎 1963）、立身出世のためにはテストで高い点を取ることが必要だったからだと考えられる。そういう状況では、とにかくテストの点が高ければ良いという風潮が生まれ、「テスト問題の領域代表性を考えず、結果を過度に一般化する」傾向が生じる。また、「多枝選択式テストでは学力は正しく測れず、記述式テストなら『真の学力』が測れると信じている」人はかなり多く、これは共通テストにおける記述式問題導入の追い風にもなったと考えられる。確かに記述式問題だと、受検者は自分の能力をいかに発揮できるという満足感を味わえるかもしれないが、採点するのは他の人間であり、受検者自身の意図を正しく読み取ってくれる保証はない。採点者間で採点基準は異なりうるし、採点者の好みも影響しうる。採点者内であっても採点基準がだんだんズレてくることは良くある（宇佐美 2013 など）。記述式問題で「真の学力」を発揮できたとしても、それが正しく「測られる」には、解決すべき課題は大きいのである。このように、テストについて必ずしも合理的でない理解が蔓延しているのは、日本で「テスト理論が普及していない」ためである。テスト理論とは、服部（2011）によれば、テストのあるべき姿、作成手順、実施と採点手順などに関する知識体系である。戦後日本では、教育測定学や計量心理学などテスト理論を扱う専門家の養成をほとんど行わなくなったため（木村 2010）、さまざまな場面で、これだけテストが利用されているにもかかわらず、現在でもテスト理論は普及していないのである。

4. 共通テストの複数回実施が見送られてきた理由及び検討すべき課題

本節では、共通テストの複数回実施が見送られてきた際に付された理由や、検討すべき課題について整理する。結果を表2に示す。

表2 共通テストの複数回実施が見送られてきた理由及び検討すべき課題

中央教育審議会第二次答申（平成9年6月1日）
高等学校教育への影響
受験競争の早期化
受験者の負担の増大
実施会場となる大学等の負担の増大
大学審議会答申（平成12年11月22日）
高等学校教育への影響
受験準備の早期化
受験期間の長期化
実施会場となる大学等の負担の増大
テスト得点の早期通知の困難性
難易度調整の困難性
資格試験化の検討の必要性
高大接続システム改革会議（平成28年3月31日）
IRTやCBTの導入検討の必要性
問題を非公開にすることの必要性
多量の問題を集積した問題プールの必要性
日程の問題
資格試験化の検討の必要性
テスト作成に関する課題
問題冊子作成業務の増大と複雑化
予備調査データの分析
多数の小問からなるテストへの形式の変更
問題プールのメンテナンス
同一科目のテストの内容の同質性の確保
問題作成・修正体制の整備
テスト得点に関する課題
標準化されたスコアへのテスト得点の変更
異なる回のテスト得点間の等化作業
テスト得点の有効期間の検討
その他検討すべき課題
実施、採点、運営等にかかる業務の増大
受験料等、受験にかかる費用の負担増

1997 (H9) 年の中央教育審議会第二次答申では、複数回実施を見送る理由として、高等学校教育への影響、受検競争の早期化、受検者の負担の増大、実施会場となる大学等の負担の増大などが挙げられている。

2000 (H12) に出された大学審議会答申 (大学審議会 2000) では、これらの理由に加え、受検準備の早期化、受検期間の長期化、テスト得点の早期通知の困難性、難易度調整の困難性が挙げられている。またこの答申では、共通テストの資格試験化の検討の必要性も触れられているが、これは、資格試験型のテストであるバカロレアやアビトゥーア等のテストが複数回実施されていることを意識したものと考えられる。

これらとは性質を異にするのが、冒頭で挙げた高大接続システム改革会議最終報告書である。この報告書では、高等学校や大学、受検者等の負担について、日程上の問題と一言ある以外はまったく触れられておらず、IRT (Item Response Theory : 項目反応理論。項目応答理論とも言う) や CBT (Computer Based Testing) の導入検討の必要性、問題を非公開にすることの必要性、多量の問題を集積した問題プールの必要性、資格試験化の検討の必要性など、TOEFL 等アメリカ型のテストの複数回実施を念頭に置いた理由付けがなされている。これは、高大接続システム改革会議等において、テスト理論の1つである項目反応理論 (IRT) を基礎としたコンピュータ利用型テスト (CBT) の導入が議論されていたことによると考えられる。

ではなぜ、IRT や CBT の導入は困難であると判断されたのか。以下では主に IRT について、日本のテスト文化を考えながら検討を進める。

IRT を利用するには、テストが相応の仕様に則って作成・実施される必要がある。センター試験のような仕様では IRT は利用できない。まず問題を非公開にしなければならない。テストを複数回実施して、その得点を比較可能にするためには、最初に予備調査を行って問題の特性を把握し、次に実際のテスト問題として使用するというように、同じ問題を複数回 (最低でも 2 回) のテストで提示することが必要である。ところが、過去問が丹念に学習されるような状況下で問題を公開してしまえば、せっかく把握した問題の特性が変化してしまい、次に実施したときには得点の比較ができなくなってしまうのである。

それぞれの問題は、大問の中の1つの問題ではなく、独立した小問であることも必要である。そのような問題を多量に備蓄したデータベース (問題プール) の構築も必要で、そのメンテナンス、恒常的に問題作成や修正を行う体制の整備も行わなければならない。アメリカでは Item Writer や

Test Developer などの専門の職業がある。

毎回のテストが同質（一次的）であるようにすることも必要である。1つのテストの中には、予備調査段階の問題と、能力推定に用いられる問題が含まれる。また、1回のテストにおいても、構成の異なる複数の問題冊子を用いなければ予備調査が追いつかない。それらの問題冊子をどのように編集するかも毎回考えなければならない。

このことは、受検者は全員同じ問題に解答するという状況を破ることになる。さらに、CBT を利用した場合は、一斉実施ということでもなくなり、テストにおける公平・公正の考え方を変える必要姓にも迫られる。

問題編集の作業が増大かつ複雑になるだけでなく、実施、採点、運営等にかかるすべての業務や費用も増大する。受検者について言えば、受検料や受検準備の負担が増える。

テスト得点も変更しなければならぬ。異なる回のテスト得点を比較するには、素点ではなく、解答パターンから推定される能力値をテスト得点とする必要がある。もちろん、そのための得点等化や、有効期間の検討という作業も必要になってくる。

以上のように、IRT を利用した共通テストの複数回実施には、日本のテスト文化と整合しない部分があるのだからである。

5. 近年のいくつかのテストについて

日本のテスト文化は共通テストの複数回実施には合わないが、近年実施されている、もしくは実施予定のいくつかのテストを見ると、複数回実施を見送る理由や日本のテスト文化の問題を突破しているように見えるものが存在する。本節ではそれらのテストについて、どのような部分がこれまでと異なっているかを整理する。整理した結果を表3に示す。

まず挙げられるのは、2020（H32）年度からの大学入学共通テストで導入される英語の認定試験である。これは、英語力の向上、英語教育の強化、英語4技能の評価の名の下に、民間業者が行う英語4技能試験を共通テストの一部として活用するというものである。本稿執筆時、まだどの民間試験が認定されるかは採用数も含め分かっていないが、申請された試験を見ると、それらは、既に年複数回実施されており、受検者は受検回を選択でき、少なくとも回ごとに問題が異なり、テスト得点として素点でなく標準化されたスコアが用いられている。受検者は、高3の4月から12月までの

間に実施されるテストのうちの2回分までの成績を、共通テストの成績として利用することができる」とされている。このように、同一問題を一齐に解かず、素点でもないテストの成績を入試で利用するというのは、これまでの日本のテスト文化とは異なるものである。

表3 日本のテスト文化や複数回実施が見送られてきた理由
及び課題とは整合的でない事例

認定試験の導入（大学入学共通テスト）
年複数回実施
受検回を選択可
複数の問題冊子
テスト得点として、素点でなく、標準化されたスコアが用いられる
段階別評価が利用される
受検のたびに受検料が必要
試行調査（プレテスト）の実施（大学入学共通テスト）
10万人規模の本番さながらのテストを、本試験の数ヶ月前に実施
各大学等に協力要請
記述式問題の導入（大学入学共通テスト）
膨大なコストがかかる
評定者間、評定者内の評定のズレの問題
テスト得点に占める割合の問題
医療系大学間共用試験
臨床実習の可否に影響するハイ・ステークスな試験
問題は非公開
テスト得点として、素点でなく、標準化されたスコアが用いられる
埼玉県学力・学習状況調査
指導と学力の関係の分析
問題は非公開
テスト得点として、素点でなく、標準化されたスコアが用いられる
就職試験における適性検査等
就職に影響するハイ・ステークスな試験
問題は非公開
テスト得点として、素点でなく、標準化されたスコアが用いられる

次に挙げられるのは、やはり大学入学共通テストにおける、記述式問題導入のための試行調査（プレテスト）である。全国10万人規模の本番さながらのテストを本試験の数ヶ月前に実施するにあたり、全国の大学等に協力要請がなされ、試行調査が実施されるのである。規模にして5分の1程度とはいえ、全国の大学等を試験会場とし、試験実施本部を設置したりして本番さながらに実施するのであるから、複数回実施の予行演習にもなり

うるものである。

その記述式問題は採点等に膨大なコストがかかる、評定者間、評定者内で基準にズレが生じるなどの問題を孕みながらも、最終報告書から4年後の導入が決められている。採点は民間業者に委託し、評定基準についてはズレが生じないような、条件への適合性を問うような記述式問題を出题することで対応するとされている。

共通テストから離れたものでは、医歯学部教育で用いられている医療系大学間共用試験がある。この試験は、医歯学生が臨床実習に入る前に修得すべき到達目標に達しているかを評価するハイ・ステークスな試験であり、社団法人医療系大学間共用試験実施評価機構により実施されている。この試験の一部ではIRTが用いられており、TOEFL等と同様に、問題は非公開で、同じ問題を複数回利用し、テスト得点として標準化されたスコアが用いられている。

埼玉県教育委員会が実施している埼玉県学力・学習状況調査は、指導と学力の関係の分析を行うことが主目的のテストであるが、ここでもIRTが利用されている。やはり問題は非公開で、同じ問題を複数回利用し、テスト得点は標準化されたスコアである。

就職試験における適性検査等においても、就職に影響するハイ・ステークスな試験であるにもかかわらず、IRTを使い、問題は非公開とされ、同じ問題を複数回利用し、標準化されたスコアを使っているものがある。

6. 共通テストの複数回実施実現を考えたときに残る要因の検討

本節では、近年のテストが、共通テストの複数回実施を見送ってきた理由や課題、日本のテスト文化の問題のどの部分を突破しているか、しかしまだどのような要因が残っているかを検討する。結果を表4に示す。以下、表4に沿って考察する。

高大接続システム改革会議で挙げられたような、問題の非公開の必要性、標準化スコアへの得点の変更、小問形式への移行、IRTやCBTの導入、また、受検料等受検者にかかる負担については、決定事項として実施主体側がテスト仕様をそのように設定することで解決されている。もちろん、そう決定できたのは、そのような仕様のテストの導入が可能だと判断されたからであるが、この手法を用いるならば、年1回一斉実施、問題の公開、大問形式、素点主義などの文化的要因は、導入困難の理由にならなくなる。

表 4 共通テストの複数回実施を考えたときに残る要因の検討

方法	理由・課題	テスト文化
決定事項として設定 (認定試験) (医療系試験) (埼玉県試験) (就職試験)	問題を非公開にする必要性 標準化されたスコアへのテスト得点の変更 多数の小問からなるテストへの形式の変更 IRT や CBT の導入検討の必要性 受検料等、受検にかかる費用の負担増	年に1回、一斉実施 1つの問題冊子に全員解答 テスト問題は実施後公開 大問形式 1問あたりの解答時間が長い テスト得点は素点 大規模/クラスルームの違いが不明確
民間委託 (記述式問題) (認定試験)	実施、採点、運営等にかかる業務の増大 問題冊子作成業務の増大と複雑化 同一科目のテストの内容の同質性の確保 予備調査データの分析 難易度調整の困難性 異なる回のテスト得点間の等化作業 多量の問題を集積した問題プールの必要性 問題作成・修正体制の整備 問題プールのメンテナンス テスト得点の有効期間の検討 テスト得点の早期通知の困難性	毎回新作問題 予備調査なし 問題は使い捨て 得点の有効期限不明 出願までの期間短い 得点通知は入試後
必要性を強調し 問題視せず (認定試験)	日程の問題 高等学校教育への影響 受検準備の早期化 受検競争の早期化 受検期間の長期化 受検者の負担の増大	多枝選択式テストでは学力を測れない 満点や0点に特別な意味
協力要請 (試行調査)	実施会場となる大学等の負担の増大	
残る要因	資格試験化の検討の必要性	選抜/資格の違いが不明確 テスト得点を過度に重視 テスト結果の過度な一般化 自己採点、合否予想 非公開問題の復元 学問領域の専門家がテストを設計 テスト理論が普及していない

実施、採点、運営、問題冊子作成等にかかるコストの増大及び複雑化、問題作成にかかる一連の業務体制の整備、また得点の等化、通知等については、それらの業務を民間委託することによって解決されている。毎回新しい問題を作っては捨てるということもなくなり、問題プールを構築したりメンテナンスするために必要な Item Writer や Test Developer などの新たな業種が生まれ、IT 産業や教育産業など、テストに関連する産業の成長

も見込まれる。

高等学校教育への影響や受検準備の早期化、受検期間の長期化などの日程の問題、受検者の負担の増大という問題については、英語認定試験の活用のように、そのテストの必要性を強く主張し、問題視しないか、やむを得ないという対応が取られる。民間の英語試験が主として多枝選択式のテストであるということも、特に議論にはなっていない。

実施会場となる大学等の負担の増大については、共通テストの試行調査のように、協力要請というかたちで突破されている。協力要請に対して応じないという選択枝はまずないであろうから、実質は強制である。ただし、実施会場がどこになるかについては未確定の要素がある。CBTが採用されれば受検者が一斉に集まる必要はなく、逆にシステム容量などの問題から、受検会場、受検日は分散したほうが良いからである。

以上のように考えると、共通テストの複数回実施実現を考えたときに、これまで言われてきた理由や課題の中で残る要因は、資格試験化の検討の必要性のみとなる。確かに、フランスのバカロレア、ドイツのアビトゥーア等の試験は、複数回実施されている。これらの試験ではIRTは用いられていないので、大問形式、記述式等、日本のテスト文化にあった仕様のテストで複数回実施を実現することが可能である。解決しなければならない問題は、共通テストを資格型にするということである。フランスやドイツでは大学の定員管理が厳しくないので、選抜性の高い大学・学部でない限り、有資格者は希望する大学への入学が許可される。そのような入試制度にするのか、それとも、有資格者は各大学が実施する個別試験を受検する資格を得るとするのか。いずれにしても、大きな入試制度の変更が必要となる。

資格化の問題にも関連する文化的要因として、テスト得点を過度に重視することが、残る要因として挙げられる。何にせよとにかく点が高いほうが良いという文化においては、あらゆるテストが選抜型の様相を帯び、同資格者であっても得点が高い者のほうが優遇されるなど、選抜試験と資格試験の違いが不明確で、そもそも共通テストの資格化自体が困難である。また、今般の入試改革においても、調査書や推薦書の見直しや、志願者本人が記載する資料の利用等、多面的・総合的評価を行うことが言われているが、「原則として学力検査を免除し」という表現が削除され、共通テストを用いるなど何らかの方法で「学力」を評価することが求められている（文部科学省 2017）。結局、共通テストの得点が合否に影響するとなれば、テ

スト得点の早期通知により自己採点はなくなるかもしれないが、大規模な合否予想は依然として残るであろう。IRTを利用した場合は、少しでも高い点を取るために、組織的に非公開の問題を復元し販売しようとする業種が生まれてくることは想像に難くない。そうなれば問題の非公開性が崩れ、テストの複数回実施そのものができなくなる。

もう1つ残る大きな文化的要因は、テストの設計、問題作成、編集等を、別に職を持つ当該学問領域の専門家が兼業として行い、テスト理論が普及していないということである。複数回実施のためのテストを設計しようとしても、それに必要な知識が十分には普及しておらず、またテスト理論に精通した専門家も極端に少ないことから、複数回実施に適したテストの設計や開発の議論が進み難いという問題がある。複数回実施が検討され始めてから30年も経って、ようやくテスト理論に基づいた困難理由が表に出てきたことも、このことをよく象徴している。テスト得点を過度に重視しているが、日本ではそのテストを合理的な理論に則って設計・開発するということができていないのである。

以上を簡潔にまとめると、共通テストの複数回実施を実現するために、決定事項として問題非公開などのテストの仕様を定め、民間業者に膨大複雑な実施関連業務を委託して経済を潤し、必要性を強調して受検者の負担は問題視せず、大学等に負担を強制したとしても、選抜試験にするのかそれとも資格試験にするのかというテストの活用法の問題、テスト得点の過度な重視とそれに起因する問題、テストを合理的に設計・開発するための知識基盤が脆弱であるという問題が、依然残る問題として挙げられる。

7. おわりに

本稿では、大学入試における共通テストの複数回実施の可能性について、日本のテスト文化、これまで複数回実施が見送られてきた理由や課題、近年のテストの状況等の観点から検討を行った。その結果、依然残る要因として、(1) 選抜試験にするのか資格試験にするのかというテストの活用法の問題、(2) テスト得点の過度な重視とそれに起因する問題、(3) テストを合理的に設計・開発するための知識基盤が脆弱であるという問題の3つがあることが指摘された。

共通テストの複数回化に向けこれらの問題が解決されるかどうかは分からないが、今後、記述式問題の時と同じように、異例の速さで共通テスト

の複数回実施が導入される可能性はある。民間業者、受検生、高校、大学等はそれに備えておく必要に迫られているのかもしれない。

参考文献

- Arai, S., and Mayekawa, S., 2005, “The Characteristics of Large-scale Examinations Administered by Public Institutions in Japan: From the Viewpoint of Standardization”, *The Japanese Journal for Research on Testing* (日本テスト学会雑誌), 1: 81-92.
- 中央教育審議会、1997、『21世紀を展望した我が国の教育の在り方について(第二次答申)』。
- 中央教育審議会、1999、『初等中等教育と高等教育との接続の改善について(答申)』。
- 中央教育審議会、2014、『新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について：すべての若者が夢や目標を芽吹かせ、未来に花開かせるために(答申)』。
- 中央教育審議会高大接続特別部会、2012、『第2回議事録』。
- 大学審議会、2000、『大学入試の改善について(答申)』。
- 木村拓也、2010、「日本における『テストの専門家』を巡る人材養成状況の量的把握」『日本テスト学会雑誌』6: 29-49。
- 木村拓也・安野史子・荒井克弘、2017、「大学入学者選抜制度における「複数回受験」の研究－能研テスト昭和39年・40年連続受験者の得点分析－」『データ分析の理論と応用』6: 3-19。
- 高大接続システム改革会議、2016、『高大接続システム改革会議「最終報告」』。
- 服部環、2011、「古典的テスト理論」松原望他編『統計応用の百科事典』丸善出版、434-5。
- 前川眞一、2015、「日本の試験風土」『大学入試の日本的風土は変えられるか』大学入試センター入学者選抜研究に関する調査室、251-60。
- 光永悠彦、2017、『テストは何を測るのか：項目反応理論の考え方』ナカニシヤ出版。
- 宮崎市定、1963、『科挙－中国の試験地獄』中公新書。
- 文部科学省、2003、『平成14年度文部科学白書』。
- 文部科学省、2004、『平成15年度文部科学白書』。
- 文部科学省、2017、『平成33年度大学入学者選抜実施要項の見直しに係る予告』。
- 日本テスト学会編、2007、『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。

- 日本テスト学会編、2010、『見直そう、テストを支える基本の技術と教育』金子書房。
- 西郡大・倉元直樹、2007、「日本の大学入試をめぐる社会心理学的構成研究の試み」『日本テスト学会誌』3: 147-60。
- 大桃敏行・上杉孝實・井ノ口淳三・植田健男編、2007、『教育改革の国際比較』ミネルヴァ書房。
- 柴山直、2008、「日本のテスト文化について」『人事試験研究』208: 2-13。
- 内田照久・大津起夫、2013、「大学入試センター試験への英語リスニングテストの導入に至る歴史的検討とその評価」『日本テスト学会誌』9: 77-84。
- 宇佐美慧、2013、「論述式テストの運用における測定論定問題とその対処」『日本テスト学会誌』9: 145-64。

謝辞

本稿を執筆するにあたり、東北大学大学院教育学研究科教授の柴山直先生及び九州大学人間環境学研究院の木村拓也先生から多数の有益な示唆を頂きました。ここに記して感謝申し上げます。