

# 日本の大学入試における4技能外部 試験使用の論証に基づく妥当性検証 シミュレーション：早稲田大学を例に

太原 達朗

2019年9月時点において、日本における大学入学試験英語問題の4技能外部試験導入の是非に関して激しく議論が行われている状況にある。4技能外部試験とは、4技能試験と外部試験を合わせたものを指す。4技能試験とはリーディング・リスニング・ライティング・スピーキングの4つの技能を測定する試験を意味し、外部試験は英検、TOEFL iBT<sup>®</sup>、IELTS<sup>™</sup>、TEAP Test<sup>®</sup>、GTEC<sup>®</sup>、ケンブリッジ英検などのテスト機関が作成したテストを指す。各大学が独自に作成する試験（以下、独自試験）に対して外部に委託する試験なので外部試験と呼ぶ。文部科学省がセンター試験に代わる大学入試共通テストとして4技能外部試験を導入しようとするのに対し、公平性や実行性などの観点から反対する研究者が声を上げた（阿部, 2017; 鳥飼, 2018; 南風原(編), 2018)。その結果、文部科学省とその方針に賛成する人々と前述の研究者を含めた反対する人々で対立している構図が出来ている。しかしこの両者の意見の対立による議論の問題点は、言語テストという学問分野が存在するのにも関わらずその知見があまり活用されていない点である。この現状を踏まえて、本稿の目的は言語テストの知見を用いることでどのように既存の議論に新たな視点を得られるかを示すこととする。

本稿の流れは以下の通りである。まず大学入学試験における4技能外部試験導入の経緯と議論を概観する。次に言語テストの概要を説明を行い、特に言語テストの重要概念である妥当性・妥当性検証を説明する。続けて妥当性検証の枠組みの一つである Assessment Use Argument（以下AUA, Bachman & Palmer, 2010）について説明し、妥当性検証においてAUAが有効である点を示す。さらにAUAを用いた妥当性検証が日本における大学入学試験英語問題の4技能外部試験導入の検証の活用例を示す。検証方法をイメージしやすいように、4技能外部試験の活用例として「早稲田大学文化構想学部・文学部の入試においてTEAP Testを用いる」場合を例として設定し、AUAの枠組みを用いて検証シミュレーションを行う。検証シミュレーションの中で公開されている入手可能な情報源から得られるデータと今後必要な情報を提示し、AUAの妥当性検証のプロセスを提示する。

なお、早稲田大学文化構想学部・文学部では一般入試の受験方式の一つに外部試験方式があり、複数の外部試験の中から受験者が選んで出願が可能であり、TEAP Testはその1つである。TEAP Testとは日本英語検定協会と上智大学がベッドフォードシャー大学の言語評価センターが研究協力のもとで共同で開発した試験であり、『『大学教育レベルにふさわしい英語力』を正確に測定する、アカデミック英語能力判定試験』（日本英語検定協会，n.d.a）である。試験は7月・9月・11月に年3回実施される。受験生は高校2年生から受験可能で、高3の11月までに年3回を2年間受験することで高校卒業までに最高6回受験可能である。

## 大学入試英語試験における4技能外部試験導入の経緯

上記の議論が行われるようになった経緯の出発点となる、大学入学試験英語問題の4技能外部試験導入に関して文部科学省が公表してきた方針は以下の通りである。2020年の東京オリンピックを見据えてグローバル化に対応した教育を実施するため、文部科学省は2013年に「グローバル化に対応した英語教育改革実施計画」において「外部検定試験を活用して生徒の英語力を検証するとともに、大学入試においても4技能を測定可能な英検、TOEFL等の資格・検定試験等の活用の普及・拡大」（文部科学省，2013）を行うと述べ、大学入試における4技能外部試験の導入可能性を示唆した。翌年の2014年にも文部科学省は英語教育改革方針の1つとして外部試験の活用を明示した。外部試験を活用する理由としては大学入試において4技能におけるコミュニケーション能力を測定する必要性に加え、「各大学等のアドミッション・ポリシーとの整合性を図ることを前提に、入学者選抜に、4技能を測定する資格・検定試験の更なる活用を促進」する方針を述べた（文部科学省，2014）。さらに文部科学省は2016年に「高大接続システム改革会議『最終報告』の公表について」にて、大学入学者選抜改革の一部として『『大学入学希望者学力評価テスト（仮称）』の英語については、高大接続改革答申や『英語力評価及び入学者選抜における英語の資格・検定試験の活用促進に関する連絡協議会』の議論も踏まえ、今後、『話すこと』『書くこと』『聞くこと』『読むこと』の四技能の評価を推進する。」（文部科学省，2016）と言及した。

4技能外部試験が大学入学試験として導入されることが周知されたのは、文部科学省の方針に加えて国立大学協会と大学入試センターが大学入学共通テストの具体的な実施方針を発表して以降である。2017年に国立大学協会は「大学入学共通テストの枠組みにおける英語認定試験及び記述式問題（国語）の活用に当たっての参考例等について」を公表した。その公表内容において、2020年度以降の大

学入学共通テストとしての英語の試験は、「国立大学としては、新テストの枠組みにおける5教科7科目の位置づけとして認定試験を『一般選抜』の全受験生に課すとともに、2023年度まではセンターの新テストにおいて実施される英語試験を併せて課すこととし、それらの結果を入学者選抜に活用する」（国立大学協会、2017）と方針を述べ、2020-2023年度までは現行のセンター試験の形式を引き継いだ試験（注：試験はリーディング100点、リスニング100点と配点も大きく変わり、試験内容も大きく異なる）と並行して4技能外部試験を用いることを発表した。同時に2024年度以降は大学入学共通テストとして4技能外部試験のみを使用する方針も発表した。英語試験において4技能外部試験を入学試験として扱う方法は、2018年の独立行政法人大学入試センターが発表した「大学入試英語成績提供システム」に述べられ、文部科学省と大学入試センターの協議の上で、資格・検定試験実施主体の中で認定された複数の外部試験の中から受験者が試験を受験し、そのスコアが大学入試センターを通して受験生が応募する大学へと提出されることが決定した（大学入試センター、2018）。また、複数の資格試験を利用するため、異なる試験のスコアを比較するためにはCommon European Framework of Reference for Languages (CEFR, Council of Europe, 2001) と各外部試験の点数換算表を用いることが公表された。CEFRは言語使用者が言語を用いて何ができるかを「〇〇が出来る」という形の例示的能力記述文 (illustrative descriptors) で表し、そのレベルを6段階（低い方から順にA1、A2、B1、B2、C1、C2）で表している。換算表は各資格・検定試験団体がスコアとCEFRの6段階のレベル関係を公表した情報を文部科学省が取りまとめ、この換算表を参考にして異なる試験同士の結果を同一の尺度にまとめている。この換算表は適宜更新され、最新のもののは文部科学省のHP ([http://www.mext.go.jp/b\\_menu/houdou/30/03/\\_icsFiles/afieldfile/2019/01/15/1402610\\_1.pdf](http://www.mext.go.jp/b_menu/houdou/30/03/_icsFiles/afieldfile/2019/01/15/1402610_1.pdf)) に公表されている。

一方、大学入学共通試験における4技能外部試験導入の動きに対して反対する活動が2017年頃から見られるようになった。国立大学協会の発表等を受けて、大学入学共通テストにおける4技能外部試験に反対する大学教員・研究者による書籍が出版された（例、阿部、2017；鳥飼、2018；南風原（編）、2018）。4技能外部試験導入の反対運動は2018年に入って勢いを増し、2018年2月・2019年2月に東京大学において主に4技能外部試験導入に反対する論者たちによるシンポジウムが開催された。反対活動は2019年6月に羽藤由美氏を中心とした研究者による、大学入学共通テストにおける英語民間試験の利用中止のインターネット上における署名運動に発展し、2019年6月18日に国会に請願書類が提出された（羽藤、2019a）。

反対論者が述べる反対の主な理由は以下の通りである（南風原〔編〕、2018；羽

藤, 2019b)。第1に公平性の観点、特に地域差・経済差の観点から公平性が損なわれることが多く指摘されている。具体的には民間の4技能外部試験の受験会場が少なく、地方にいる受験生が受験するのが難しいという地域差がある。また民間の4技能外部試験は複数回受験が可能であり、2種類以上の試験を受験して最も良いスコアを提出することも可能であることであるため経済格差も存在する。このため、地方の生徒や経済的に恵まれない生徒が大都市部の生徒や裕福な家庭の生徒に比べて受験する機会や良いスコアを得る機会が少ないことが問題点として挙げられる。他にも上記のCEFRの基準によって複数の試験の点数のスコアを比較することの是非も問題として挙げられている。各試験は同じ英語試験でも問題構成が異なり、よって測定する能力も微妙に異なるにも関わらず、各試験のスコアをCEFRという6段階の共通の尺度に換算し比較することにも疑問があり、大学入試という人生に大きく影響を与える試験において使用するのに問題があるという考えも根強い。

しかしここで注視すべきなのが、4技能外部試験導入の是非が議論されている中の問題点に言語テストという応用言語学の研究分野で長年行われてきたテスト使用の妥当性検証の手法が参照されていないという点にある。議論が錯綜している理由の一つには、センター試験に代わる大学入学共通テストとして4技能外部試験を使うことと、各大学の独自試験の代わりに4技能外部試験を採用する場合は混同されている点にある(寺沢[2018]の分類を参照)。文部科学省の初期の方針(文部科学省, 2013; 2014)では大学入試において4技能を測定することを推奨し、それに関して外部試験を用いることを推奨してきたのに対して、国立大学協会(2017)の方針の発表以降はセンター試験の代替の大学入試共通テストとして扱う話が中心となり、議論の大半も大学入学共通テストとしての4技能外部試験導入の是非に関するものとなっている。しかし全受験生に影響のある大学入学共通テストと異なり、独自試験の代替として各大学が使用することはまた別の問題である。日本の大学入試試験における4技能外部試験の是非の場合のような、目的に応じた適切なテスト使用に関する検証は言語テストの分野において知見が蓄積されていた点であり、これらを踏まえて議論することが望ましいと考えられる。日本の大学入試の独自試験の代わりに4技能外部試験を「使う・使わない」の2択で考えることは、「テストの結果をどのように使うか」という言語テストという学問分野が積み重ねてきた知見を無視することになり、各大学が望ましい効果を意図して適切に外部試験を使うことを妨げることになる。同時に、なぜ外部試験を用いるかはなぜ4技能を測定する必要があるかにも関連し、どのような英語力を測定するかも考慮する必要がある。これも言語テストにおいては「どのような能力を測定すべきか」に関わる妥当性という概念、そして妥当性をどのよ

うに検証するかという妥当性検証の方法が研究されてきた。この点から、言語テストの知見が4技能外部試験導入の既存の賛成・反対の対立に対して新たな視点を提供出来ると考えられる。

上記の言語テストの必要性を踏まえると、言語テストの知見を活かしてどのように目的に応じた適切なテストを大学入学試験として採用するか、また採用する前にどのような点を参照して選択すべきか、採用後に採用した大学がどのような調査を行いどのようなデータを集めるべきかを知る必要がある。これらを踏まえて、本稿では独自試験の代替としての4技能外部試験を導入する場合に限定して検証を行う。

## 言語テスト

言語テスト (language testing, language assessment) は応用言語学の一学問分野であり、『テストをどのように作り、使ったらよいか』『テストの質をどう評価したらよいか』『言語能力とは何で、テストでは何を測るのか』(小泉, 2018, p. 3) といった問いに答える学問である。言語テストの分野においては *Language Testing, Language Assessment Quarterly* などの国際誌が存在し、また国際言語テスト学会 (International Language Testing Association)、アジア言語テスト学会 (Asian Association for Language Assessment) などの国際学会が設立されている。日本においても1996年に日本言語テスト学会が設立されており、年に1度国内で学会が開催され、年に一度日本言語テスト学会誌 (*JLTA Journal*) が発行される。日本言語テスト学会は大学入試における4技能外部試験の使用に関しても2017年4月1日付で「大学入学希望者学力評価テスト(仮称)における英語テストの扱いに対する提言」([http://jlta2016.sakura.ne.jp/?page\\_id=865](http://jlta2016.sakura.ne.jp/?page_id=865)) を発表し、新たな大学入試における4技能テストの使用に関する方針を発表した(日本言語テスト学会, 2017)。しかしこの声明が現在の議論で引用されることは少なく、文部科学省や連絡協議会などが運営している「英語4技能試験情報サイト」にもこの声明のURLへのリンクは参照されていない。

言語テストの知見に対して初めて知る読者にとって有益なのが、言語テストに対する3つの誤認識、「1. 言語テストへ過度な期待をする」「2. 言語テストの専門家へ過度な依存をする」「3. 言語テストを1つの質の良し悪しだけで判断する」(Bachman & Palmer, 2010, pp. 6-11, 小泉(訳), 2018, pp. 26-29) である。日本の大学入学試験における4技能外部試験導入に関して、誤認識1と3はそれぞれテストに関する単純な考えについて警鐘を鳴らしている。「誤認識1. 言語テストへ過度な期待をする」はテストを変えるだけで良い変化が自動的に生まれ

るわけではないことを示す。例えば4技能外部試験を採用すれば日本人の高校生や大学生が4技能をバランスよく身につけることができるようになるわけでは必ずしもない。「誤認識3. 言語テストを1つ質の良し悪しだけで判断する」は、あるテストが絶対的にいいテストか悪いテストかどうかの2択で判断するのは適切ではない。例えば各大学の目的や文脈を無視して「TEAP Testは良いテストなのか、悪いテストなのか？」と聞かれてもそれだけでは判断することはできない。一方、誤認識2は専門家に全てを委ねてしまうことに対する警告であり、これでは専門家がいなければ4技能外部試験を使用することはできないことになってしまう。専門家でなくても後述する妥当性検証について知り、各大学の目的などに応じて言語テストを使用する態度を持てば、適切に4技能外部試験を使用することができ、この点で専門家以外も言語テストの知識を持つことが4技能外部試験の導入の際に重要となる。

## 妥当性と妥当性検証

言語テストにおける重要概念が妥当性と妥当性検証である。「妥当性」は教育測定における重要な概念であり、言語テストでもこの妥当性の検討を行ってきた。このテスト妥当性についての考え方は時代と共に変遷してきている（日本語による妥当性検証の歴史の文献は澤木（2011）や村山（2012）等を参照）。初期の妥当性は「測ろうとしている能力が測定出来ているかの程度」を表す概念であったが、Messick（1989; 1996）が新たな妥当性検証の枠組みを提唱したことにより、テスト使用の結果まで含めた妥当性の定義が用いられるようになり、「テスト自体が妥当な性質を持つ」という考え方から「テスト得点の解釈と使用の結果がどれだけ妥当か」という考え方にシフトした（Messick, 1989）。ここでいう「解釈」と「使用」というのは専門用語であり、それぞれ「テスト得点がどのような能力を測るものであるか（解釈）」「そのテスト得点をどのように用いるか（使用）」を指す。ここで解釈と使用に関してTEAP Testを大学入試において使用する例をあげると、TEAP Testの内容である「日本における『大学教育レベルにふさわしい英語力』」（解釈）というテスト内容や実施方法は誰がどこで用いても同じだが、TEAP Testの開発者の手を離れたテスト得点の使用方法は各大学によって異なる。

しかし、Messickの考え方は受け入れられこの枠組みを提唱する研究者は他にもいたものの（Bachman & Palmer, 1996; Kunnan, 2004）、実施面においては懸念点もあり、上記の6つの側面の妥当性をどのような順番でどれを優先して検証すべきかがはっきりしないことなどが指摘されてきた（Bachman & Palmer, 2010）。

これらのアプローチでは「どういったデータを」「どういった手続きで」「どの程度集めればいいのか」がわかりにくいことから、Kane (1992; 2013) はツールミン・モデルを用いた論証アプローチに基づく妥当性検証 (argument-based approach to validation) を提唱した。このアプローチによる妥当性検証を行う場合、まずテストの使用と解釈に関してどのような結果が得られるかに関する1つ1つの主張 (claim) に対して、その主張を支える論拠 (warrant) を研究データや文書などの証拠 (evidence) を示すことによって論証していく。2019年現在この論証アプローチに基づく妥当性検証の主流な考え方の1つになっている。Kaneの他にも複数の研究者がこの妥当性検証の論証の枠組みを提唱してきたが (例; Chapelle, Enright, & Jamieson, 2008; Knoch & Chapelle, 2018)、そのベースにあるものはKaneのものである。

言語テストの妥当性検証について日本語で説明を行っている文献は複数存在する (例、飯村, 2017; 熊澤, 2013; 小泉, 2018; 澤木, 2011; 清水, 2004; 水本, 2014)。しかし、テスト開発者側が自らのテストが適切なものであることを提示する研究結果だけではなく、他者が開発したテストをテスト使用者が自らの目的に応じて適切に使用しているかについて検証したものは少ない。今後大学側が独自試験の代替として4技能外部試験を用いる場合に、具体的にどのようなプロセスを経て4技能外部試験の解釈と使用に関してシミュレーションの形で妥当性検証を行うかを理解することは、4技能外部試験を使用しているまたは使用を検討している大学関係者や、大学教員や学生などの実際に4技能外部試験を使用することによって影響を受ける利害関係者や外部から4技能外部試験について批判的に検討する研究者にとっても有益であると考えられる。

## Assessment Use Argument (AUA, Bachman & Palmer, 2010)

上記の論証に基づく妥当性検証の枠組みの1つに Assessment Use Argument (AUA, Bachman & Palmer, 2010) があり、これらが現在の日本の大学入試英語問題における4技能外部試験の検証に適しているといえる。その理由の1つは、各Claimで検証すべき内容の1つ1つの段階がわかりやすいことである (飯村, 2017; 小泉, 2018)。もう1つの理由が言語テストを自ら開発する場合と、他者が開発したテストを使用する場合の両方について検証方法が示されている点である。この点で日本における4技能外部試験における民間のテスト団体が作成したテストを用いるか否かという議論を行うのに適している。大事なのは自ら試験を開発する場合にも、他者が作成した試験を使用する場合のいずれに関しても検証が必要だという点にある。

日本の大学入試における4技能外部試験使用の論証に基づく  
妥当性検証シミュレーション：早稲田大学を例に

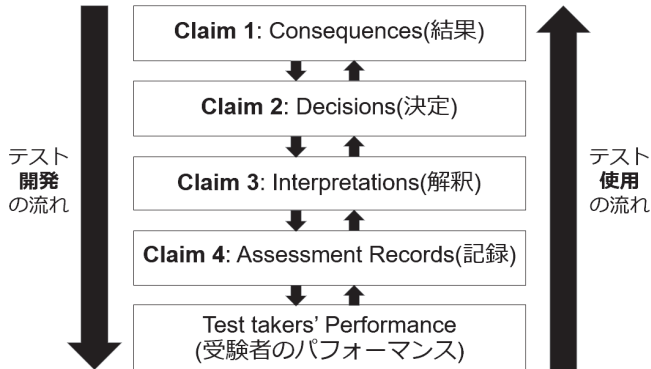


図1. AUAにおけるテスト開発・使用の流れ  
(Bachman & Palmer, 2010, p. 91を元に作成)

AUA (Bachman & Palmer, 2010) におけるテスト開発・使用の流れは図1の通りである。図1の左右の矢印に示す通り、テストを自ら開発する場合と（テスト開発）と他者が作成したテストを使用する場合（テスト使用）では検証方法の順番が異なるが、いずれの場合も最初に図中の各Claimの内容を検討する（各Claimの詳細は後述）。早稲田大学文化構想学部・文学部のTEAP Testを例に取ると、新たなテストを導入することによってどのような結果を期待するか（例：TEAP Testを用いることにより大学入学者がスムーズに大学教育を受けられるようになる）を考え、次にその結果を得るためにどのような決定を可能にするテストが必要かを決め（例：TEAP Testの各技能で65点以上、合計点280点以上取った受験生は英語の試験を免除し、国語と社会の点数で合否を決める）、その決定を下すためにどのような解釈（テストがどのような能力を測定するかという解釈；例：テストは「大学教育にふさわしい英語力」を測定する）を行うか、その解釈がどのような記録（受験者のパフォーマンスを得点化する際に元となる記録された採点結果；例：リーディング・リスニングテストはマークシートによって採点され、スピーキング・ライティングは訓練された採点者によって採点される）を必要とするか、という内容を検討する。これら4つの各Claimの要素を踏まえて、それぞれの基準を満たす既存のテストがあればそれを採用し、右側の矢印の順番のように下から上へと検証を開始する。基準を満たすテストがなく独自で試験を作成する場合は図の左側の矢印の上から下の順番に検証内容を1つ1つ考え、図の一番下の受験者のパフォーマンスが得られたら今度は実証データを元に下から上へと順番に検証していく手順を取る。

AUAの枠組みを用いて早稲田大学文化構想学部・文学部の大学入試英語問題



において4技能外部試験を用いる場合、図1の右側の上向きの矢印の手順を取る。TEAP Testを採用することが決定し実施に使用するところから議論が始まるため、「受験者のパフォーマンス」から「結果」まで下から上の順番に検証を行っていく。図2は4技能外部試験の中のTEAP Testを日本の大学入試英語問題の独自試験の代わりに用いる場合のAUAに基づく検証内容の例である。このように、各段階のClaimでどのような内容を検証すべきかを決め、そこからデータや文献などを用いて検証を行っていくのがAUAにおける妥当性検証のプロセスである。

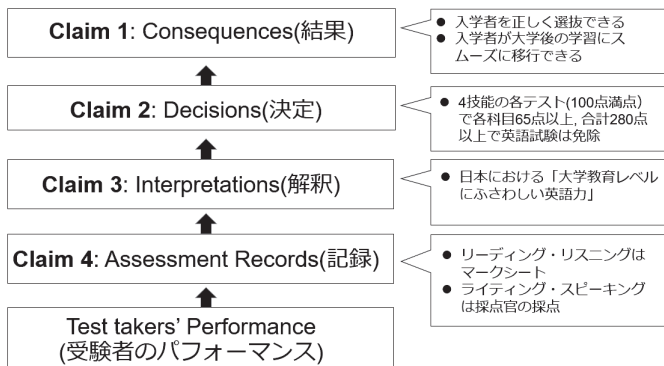


図2. AUAを用いたTEAPの検証内容とその例  
(Bachman & Palmer, 2010, p. 104を元に作成)

「AUA」「早稲田大学文化構想学部・文学部」「TEAP Test」をそれぞれ本稿で扱う理由は以下の通りである。AUAの枠組みを用いる理由は、上述の通り日本の大学入学試験において議論されている「自らテストを開発するか、それとも他者が作成したテストを使用するか」、という2択のいずれの場合にも実行可能で手順がわかりやすく、またテスト使用者の外部からの検証も比較的容易であるためである。早稲田大学文化構想学部・文学部を具体例として設定したのは、妥当性検証はテストの使用目的や文脈を考慮して行うものであるためである。前述の通り、各大学の目的や文脈を無視して「TEAP Testは良いテストなのか、悪いテストなのか？」と議論することは不適切であり、実際の使用の場を考慮する必要がある、また読者にとって妥当性検証のイメージがわかりやすいと考えられる。早稲田大学文化構想学部・文学部は早い段階で大学入学試験の一般入試で4技能外部試験型の受験方式を導入した学部であり、同時にTEAP連絡協議会において積極的にTEAP Test (<https://www.eiken.or.jp/teap/>)を始めとした4技能外部試験の使用を推進することを表明してきたことも理由である。

## 早稲田大学文化構想学部・文学部における TEAP Test使用の検証シミュレーション

ここから実際にAUAの枠組みを用いて、早稲田大学文化構想学部・文学部におけるTEAP Testのテスト使用の検証シミュレーションを行う。テスト使用の検証の順番通りに(図1と図2参照)AUAのClaim 4からClaim 1にかけて各Claimで検証すべき内容を具体的に述べる。各Claimにおける用語はBachman and Palmer (2010)を元に筆者が日本語訳し、必要に応じて追加説明を入れている。検証を行うための証拠、つまり情報は早稲田大学文化構想学部・文学部やその教職員がTEAP Testに関してHP上で公開している情報やアドミッション・ポリシー、カリキュラム・シラバスなどの文書、TEAP Testを運営している日本英語検定協会が公開しているTEAP研究レポート(<https://www.eiken.or.jp/teap/group/report.html>)、そして必要に応じて言語テストの研究成果などを用いた。なお、本稿では公開されている情報のみに基づいた検証シミュレーションであり、このため学内で現在進行形で実際に検証をしている情報など外部からアクセスできない情報には言及を行っておらず、同時に検証シミュレーションが実際の学内における議論とは異なる可能性があることに注意したい。

### Claim 4 記録 (Assessment Records) に関する妥当性検証

Claim 4: 評価の記録 (assessment records) は、異なる言語使用タスク・評価手順の異なる側面、そして異なるテスト受験者の集団を通して一貫している (consistent)

一貫している (consistent) ここでは、各受験者のテスト結果が何度受けても安定した点数が取れるか、また異なる採点者同士でその一貫性が保たれているかどうか等を示す。この点は外部からはほとんど検証できない点であるが、最低限テスト機関が公開しているテストの説明による「採点の質を確保するための方策」(<https://www.eiken.or.jp/teap/merit/>)と、「TEAP研究レポート」におけるライティングとスピーキングの開発段階において、採点の一貫性の検証が行われ、大学受験における十分な評定者間信頼性と評定者内信頼性があることが検証されていることが確認できる (Nakatsuhara, 2014; Weir, 2014)。後者のTEAP研究レポートでは、トライアル版のテストデータを用いて検証を行い、スピーキングにおける5つの評価基準「発音」「文法の範囲と正確性」「語彙の範囲と正確性」「流暢さ」「やり取りの適切さ」、ライティングの言語面における3つの評価基準「語彙」「文法」「文章と文章のつながり」がうまく機能しているかを検証を行い、ま

た受験後の受験者や採点後の試験官にインタビューを行い、面接時の応答や採点の評定者トレーニングについての情報を収集していたことがわかる。

### Claim 3 解釈 (Interpretations) に関する妥当性検証

Claim 3: 評価する能力の解釈 (つまり、テスト得点がどのような能力の指標となるかという解釈) (interpretations) は

- 特定の学習シラバス、目標言語使用領域 (Target Language Use [TLU] Domain\*) で言語を実行するのに必要な能力、言語能力とは何かを表す一般的な理論に照らし合わせて、またこれらの組み合わせに照らし合わせて意味がある (meaningful)
- \* 目標言語使用領域 = テスト受験者が実生活で言語活動を行う場面や状況 (Bachman & Palmer, 2010, p. 60)。今回の例で言うと「早稲田大学文化構想学部・文学部における英語使用場面」。
- すべてのテスト受験者にとってテスト問題の偏りがない (impartial)
- 目標言語使用領域へ一般化できる (generalizable)
- 決定を下すのに関連している (relevant)
- 決定を下すのに十分である (sufficient)

意味がある (meaningful)、関連している (relevant)、十分である (sufficient) ことでの「意味がある」はそれぞれテストで測定している内容が測定したい言語能力と合致しているか、またそういった意味で開発・使用したテストが意味のあるものであるかを指す。また、「関連している」「十分である」はテストの結果を用いて判断を下すために、それぞれテストの各問題が測定したい言語能力と関連しているか、その問題の数は十分であるかに関することである。つまり、期待する結果を起こすために必要な決定を行うために必要十分なテスト内容を含んでいるかどうかである。これらの点については、早稲田大学文化構想学部・文学部がどのような能力を測定しようとして TEAP Test を採用したかという基準と、TEAP Test の測定している能力やその元となったテスト開発者側の検証結果を照らし合わせて検証した。その結果、前者に関しては公開されている情報は確認できなかった。もし大学入試英語問題をできる限り大学入学後に使用する英語使用と関連づけてテストを使用したい場合はニーズ分析 (Bachman & Palmer, 2010) を行い、大学生活や授業内においてどのような能力が求められるかの調査を行うのが望ましい。日本の大学入試と大学英語使用に関するニーズ分析の調査例は Sawaki (2017) や Tahara (2018) などが存在する。後者に関しては TEAP Test の HP で『『大学教育レベルにふさわしい英語力』を正確に測定する、アカデミック英語能

力判定試験」(日本英語検定協会, n.d.a)だと確認できる。また日本英語検定協会(n.d.b)「問題構成・見本問題」によって大問と問題形式、ねらいが詳細に述べられている。TEAPの問題内容は高校生と大学生、高校教員を対象にしたニーズ分析を元に作成されており、日本の文脈を踏まえて測定内容を決定している(Green, 2014)。TEAPで測定する内容はライティングとスピーキングに関しては評価基準や解答例、CEFRにおける採点例が書かれており、こういった能力が求められているかが詳細に述べられている。同時に測定している能力が問題内容によって変わるので、各問題の測定内容が使用目的に応じて関連しているか、十分であるかも確認する必要がある。例えばリーディングテストでは語彙力や図を理解する能力、短文を理解する能力、長文を理解する能力などが問われており、ライティング試験では文の要約能力や与えられた情報を用いて議論文を書く能力が問われていることがわかる。重要な点は、同じ4技能を測定している外部試験同士でもその問題構成は異なるため、各試験が具体的にこういった能力を測定しているか、またそれが各大学が測定したい内容かどうか一致しているかどうかを確認する必要があることである。TEAP Testにおける測定内容やテスト細目(テストの設計図)のより詳細な情報は「TEAP研究レポート」に各技能ごとに記述されている(例, Nakatsuhara, 2014; Taylor, 2014; Weir, 2014)が、基本的には4技能とも大学生活に必要な能力、アカデミックな英語力を測定していることがわかる。

また、どのようなレベル(難易度)の能力を測定しているかどうかは文部科学省(2018a)の「各資格・検定試験とCEFRとの対照表」が参考になる。このCEFRの換算表とTEAP Testの4技能の総点との関連を見ると、TEAPはCEFR A2 からC1 までのレベルを評価できる。またTEAP TestとCEFRの対応付けの方法は調査報告書から確認できる(日本英語検定協会, 2018)。日本の高校生の実態調査(文部科学省, 2018b)では日本の高校生はA1 からB1 の受験者が多く、その意味ではTEAP Testはその範囲をカバーしていると言える。一方で実態調査において用いられたテスト問題とTEAP Testでは測定している能力が異なり、異なる測定方法から得られたCEFRのレベルが一致しているとは限らない。同時に一般の高校生と実態調査と早稲田大学における受験者の層も異なるので、その点も含めてTEAP Testが大学入試英語問題として幅広く受験者層を測定できているかどうか適切さをさらに検証していく必要がある。

**偏りがない (impartial)、一般化できる (generalizable)** この偏りがないというのは、問題内容や実施方法に受験者がコンピュータを用いたテストの実施状況などで結果が変わったりせず、またテストの受験機会やテスト対策における勉強の機会が受験生によって偏りがなく、この偏りに関しては英検のHP

(<https://www.eiken.or.jp/exam-univ/info/>) において経済的に困難な受験生の配慮、障害等のある受験生の合理的配慮、試験監督及び採点の公平性の方策について公表しているものが確認できる（日本英語検定協会，n.d.c）。また、ここでいう一般化は、TEAP Test で良い点を取れることはテストの問題を解けるだけでなく、より一般化した目標言語使用領域（ここでは大学における英語使用場面）でも英語を使えることを表しているかどうかに関するものである。この2点はテストの設計段階やテスト団体の実施方法で決まってしまうため外部からの検証が不可能であるが、英検のHPやTEAP研究レポートを確認することで英検側が行っていることを確かめることができる。

## Claim 2 決定 (Decisions) に関する妥当性検証

Claim 2: テスト得点の解釈を基に行った決定 (**decisions**) は

- テストを使用するコミュニティの価値観 (**value-sensitive**) や関連する法的要件を考慮に入れている
- 決定によって影響を受けるすべての利害関係者にとって公平である (**equitable**)

**価値観に合う (Value-sensitive)** ここでいう価値観は、テストが使われる国や地域、教育機関などにおけるテストに対する社会・文化的な価値観と合致しているかを表す。早稲田大学文化構想学部・文学部のHPを見ると「アドミッション・ポリシー」が公開されている。この中で具体的な能力に関して言及している箇所では「文化構想学部[文学部]の基礎から専門に亘るカリキュラムを理解して、入学後の修学に必要な基礎学力を有する。また入学以前に幅広い学習と経験を積んでいる。」(早稲田大学文学部, n.d.a; 早稲田大学文化構想学部, n.d.a) と述べられている。この文書は一般的な学力に関する情報で求められる英語力に関するものではない。そのため4技能外部試験としてTEAP Testに求められているテスト使用が、後述する熟達度テストとしての用途か達成度テストとしての用途なのかはここからは判断できない。しかしTEAP Testは日本の高校生を意識して作成された試験であることから、少なくとも大卒でTEAP Testの使用がアドミッション・ポリシーの趣旨に沿うものであることがわかる。同時に学内における大学入学試験における価値観は外部からは観察できないが、民間の試験を使うことや、下記に述べる通り外部試験において一定のスコアを満たすと英語の試験を免除するような運用が学内関係者に受け入れられているかに関しても検証が必要である。

**公平である (equitable)** ここでいう公平であるとは、どの受験者もテストを含む公表されている選抜基準以外において合否が決められたりすることなく、どの

ような手続きを用いて合否が決められるかを十分に周知されているかことを示す。注意すべきなのは、ここでいう公平は決定を下す判断における公平であり現在4技能外部試験導入の議論において言及される公平性の意味とは必ずしも一致しない。Claim 3のimpartialもテスト問題における公平性に関するものであり、同じ公平性に関するものでも各claimによって検証する内容が異なる。TEAPの最高点はReading, Listening, Writing, Speakingの各技能においてそれぞれ100点、合計点で400点になる。TEAP Testを文化構想学部・文学部の外部試験型の試験方式において使用した場合の選抜基準は、各技能別ではそれぞれ65点、総合点で280点の基準点が設定されており、これらの基準を満たした場合は受験が可能であり、これをクリアした受験者は国語と社会の独自試験を受験しその点数によって合否が決められる(早稲田大学文学部, n.d.b; 早稲田大学文化構想学部, n.d.b)。この点に関して重要なのは、基準スコアを設定するということは、入学前までにある一定の英語力を求めていると考えられる点である。どのような英語力が求められているかはここでは明確になっておらず、その理由として大学入試の独自試験においては大学側がテスト細目(テストの設計図)を公開していないためにテストの目的が不明であり、外部からは入学試験が選抜目的である他に達成度テスト(特定のカリキュラム、日本の場合だと学習指導要領に基づいて一定の能力基準を満たしたかどうか)なのか熟達度テスト(学習の経過に関わらず、どれだけの能力を持っているか)がはっきりしないという点がある(尾崎, 2008)。4技能外部試験は測定する能力と対応した基準スコアを設定することで入学前にどのような能力を受験者が保持しているかを示すことにより、独自試験の点数とは異なる試験結果の使用が可能になる。今回基準スコア方式を採用する経緯は公開されている情報は確認できなかったが、この点が明らかになることで4技能外部試験を用いることのメリットが明確になる。

TEAP Testを採用する際に公平を検討する場合には2つの観点がある。一つは他の4技能外部試験との比較である。早稲田大学文化構想学部・文学部の入試情報のHPにはTEAP TestだけではなくTEAP CBT(コンピュータ型)やIELTS、実用英語技能検定(英検)、TOEFL iBT、ケンブリッジ英検、GTEC CBTが使用可能である(早稲田大学文学部, n.d.b; 早稲田大学文化構想学部, n.d.b)。早稲田大学文化構想学部・文学部におけるこれらのテストの選択の選定基準や各テストの基準スコアの設定方法は公開されていなかったが、複数のテストを用いた結果を検証していくことが重要であり、例えば使用したテストによって入学直後のプレースメントテストのスコアや入学後の学業成績(例:GPA)が変化するかなどの比較検討などが可能であろう。

もう一つの観点は一般入試において、英語試験を4技能外部試験で受験する方

式（外部試験型）と独自試験を受験する型（独自試験型）の2種類が存在する点である。これらの方式を2つに分けたことで特定の受験者層が有利または不利になったかどうかの検証が必要だが、各試験方式の倍率などの情報を除いてこれらの情報は公開されていない。今後の検証方法としては外部試験型でも独自試験型でも共通で受験する国語と社会の試験の成績の比較や、外部試験型と独自試験型で入学した学生の入学直後に異なる方法で英語力を測定する方法が考えられる。同時に2020年度受験の外部試験側の場合、2018年2月1日以降に受験した外部試験のテスト結果が使用可能であるが、これらの複数回数受験した受験生に関する情報は見つけることができなかった。入試のシステム上、実際の高校生が早稲田大学文化構想学部・文学部を受験するまでにどれぐらいの外部試験を受験したかどうかは検証できないが、アクセスのしやすい入学者に関して受験1-2年前にどのような外部試験を受験したか、スコアはどれぐらいだったかなどが調査可能だと考えられる。

### Claim 1 結果 (Consequences) に関する妥当性検証

Claim 1: 評価を使用した結果 (consequences)、そしてその使用の決定の結果 (consequences) はすべての利害関係者にとって有益である (beneficial)。

**有益である (beneficial)** ここでいう有益であるとは、テストを使用する教育機関において関わるすべての関係者（受験生・教職員・大学当局など）に対してテスト導入によって有益な結果を得られるか、また有害な結果を避けられるのかという意味である。外部試験型の試験方式の導入後にどのような結果をもたらしたかに関しては導入から時間があまり経っていないこともありその結果があまり公開されていないが、英語4技能外部試験サイトにおいて早稲田大学文化構想学部の安藤文人教授と外部試験方式で入学した2名の学生がインタビューに答えている。外部試験導入のメリットとして(1) 2月の試験前に外部試験で点数をクリアした場合に受験勉強で他の科目の学習に集中できる(2) 入学後に自信を持って英語を使えると述べられている(英語4技能試験情報サイト, n.d.)。しかしこれらは組織的に行われた実証研究ではなく個人の体験談を紹介したものであり、文化構想学部・文学部全体でどのような影響があったかに関してデータに基づいて得られた結果で公開されているものは見つけられなかった。また、第6回TEAP連絡協議会において、安藤氏は英語のカリキュラム改革として入学後に4技能を用いて英語を使用することを目指していることを述べている(TEAP連絡協議会, 2016)。よって、どのような結果を期待して4技能外部試験を採用したかについては確認できるが、これらも実際に外部試験型で入学してきた学生が英語科目や

英語を用いる専門科目においてより英語を用いてスムーズに英語を使用できるようになったかどうかに関する検証結果は2019年9月現在では公開されていない。文化構想学部・文学部における英語4技能テスト利用型の一般入試が開始したのは2017年度入試からであり、その導入の結果の検証には複数の年数が必要であると考えられる。今後外部試験の使用前後における変化を継続的に調査し検証することで長期的な外部試験導入の効果が明らかになるだろう。また受験生や大学生だけでなく、教員・職員を含む大学全体における導入の結果を検証することも必要になると考えられる。

## 本稿の限界と今後の課題

本研究では最初に大学入試英語問題における4技能外部試験導入の是非の議論に言語テストの知見が必要であることを述べ、言語テストの重要概念である妥当性・妥当性検証の説明を行った。次に妥当性検証の枠組みの1つである Bachman and Palmer (2010) の Assessment Use Argument (AUA) の枠組みが日本の大学入試における4技能外部試験に適していることを述べた。さらにAUAの枠組みを用いて早稲田大学文化構想学部・文学部におけるTEAP Testの解釈と使用に関する妥当性検証のシミュレーションを行った。その結果、4種類の各Claimのうち検証として使用しうるデータが公開されている情報から一部得られたことを示し、同時により良い4技能外部試験の使用に向けて検証が望まれる点も挙げられた。特に外部試験導入によって期待する結果や(Claim 1)、期待する結果を得るためにどのような決定を可能にするテストが必要か(Claim 2)に関してはより透明性のあるテスト使用の結果の検証とその結果報告が待たれる。同時に、4技能外部試験の導入の是非を議論する際にはAUAの各Claimのどのデータが妥当性検証の観点からプラス・マイナスかというように具体的な点について考慮出来ることを示した。特に、「公平性・公正性」と一言で言われていた点が地域・経済差(Claim 2)だけではなく、問題の内容(Claim 3)や採点方法(Claim 4)に関するものであることに注意して議論が行われるのが今後望ましいと考えられる。

本研究における限界を述べる。まず本稿で行ったことはあくまで検証シミュレーションであり、早稲田大学文化構想学部・文学部におけるテストの使用目的や背景は一般に公開され入手できる情報源だけで検証を行った点である。現実には学内で議論されているテストの使用目的や選択した理由など公開されている情報以外の要素も複雑に絡み合っていると予想されるが、それらを考慮に入れることはできなかった。本来ならば大学内の各利害関係者や学内外の研究者が共同で各Claimにおけるデータを包括的に収集して実施するのが本来の妥当性検証



である。AUAで本格的な実証研究を行う際には本稿で紹介した各Claimの各要素もより詳細に研究する必要がある。また、今回はTEAP Testのみに焦点をおいて検証を行ったため、他の選択肢として存在する複数の外部試験（TOEFL iBT, IELTS, GTEC等）の併用の適切性までを十分に検証するまでには至らなかった。

しかし、今回の検証シミュレーションはAUAの妥当性検証の枠組みを用いることによって、現時点で公開されているデータが各Claimのどの論拠になりうるのか、どのようなデータが不足しているかを示すことが出来た。これにより、今後独自試験において4技能外部試験の導入を検討している大学が各4技能外部試験の使用の適切性を検討するためにどのように調査を行えばいいか指針を得ることができた。これらの方法によって使用方法について検証を行い、また外部からもこのように検証を行っていくことが今後必要である。

日本の大学入試入学試験における4技能外部試験の使用と解釈に関して今後の課題を述べる。まず4技能外部試験の大学入学共通テストでの使用については2019年9月現在も情報が錯綜し多くの議論が巻き起こっている状態であり、2024年以降から4技能外部試験のみを使用する段階に入る前に多くの問題を解決することが喫緊の課題である。次に、4技能外部試験を各大学の独自試験の代替として使う場合は本稿の検証シミュレーションで例示した通り、各大学の目的や文脈に応じて4技能外部試験を用いるのは問題ないが、それでも社会に対する影響は大きいことを自覚し、各大学が継続して妥当性検証を行っていくことが必要である。

また、既に4技能外部試験を使用している大学も積極的に現在の検証を行っていくことが必要である。もし妥当性検証の観点からマイナスの点が見つかったとしても、それだけでそのテストを使ってはいけないということになるわけではない（小泉, 2018, p. 146）。同時に妥当性検証は一度行えば終わりという性質のものではなく、最初に良い妥当性の証拠が見つかったとしても適切なテスト使用が継続しているかどうかを検証すべく、社会状況やカリキュラムの変化などの各大学の背景や文脈に応じて検証を続けていくことが大事であり、これが将来の効果的な4技能外部試験の使用に貢献すると考えられる。

## 備 考

本稿は2018年10月27日に開催された2018年度多元文化学会秋期大会研究発表会における発表「大学入試英語問題における4技能外部試験の使用に関する3つの疑問に答える—言語テスト研究の知見から—」の内容を元に、構成を大幅に変更して加筆修正を行ったものである。

## 参考文献

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Test score interpretation and use. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language™* (pp. 1-26). New York: Routledge.
- Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.
- Green, A. (2014). The Test of English for Academic Purposes (TEAP) impact study: Report 1—Preliminary questionnaires to Japanese high school students and teachers. *Tokyo: Eiken Foundation of Japan*.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement: Issues and Practices*, 50, 1-73.
- Knoch, U., & Chapelle, C. A. (2018). Validation of rating processes within an argument-based framework. *Language Testing*, 35(4), 477-499.
- Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27-48). Cambridge, UK: Cambridge University Press.
- Messick, S. (1989). Validity. In R. u. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Nakatsuhara, F. (2014). A research report on the development of the Test of English for Academic Purposes (TEAP) speaking test for Japanese university entrants—Study 1 & Study 2. *Tokyo, Japan: Eiken Foundation*.
- Sawaki, Y. (2017). University faculty members' perspectives on English language demands in content courses and a reform of university entrance examinations in Japan: a needs analysis. *Language Testing in Asia*, 7, 1-16.
- Tahara, T. (2018). Japanese university students' perspectives on English language needs in secondary school and university education. *The Bulletin of the Graduate School of Education of Waseda University*, 26, 153-169.
- Taylor, L. (2014). A report on the review of test specifications for the reading and listening papers of the Test of English for Academic Purposes (TEAP) for Japanese university entrants. *Tokyo: Eiken Foundation of Japan*.
- Weir, C. (2014). A research report on the development of the Test of English for Academic Purposes (TEAP) writing test for Japanese university entrants.
- 阿部公彦 (2017). 史上最悪の英語政策—ウソだらけの「4技能」看板. ひつじ書房
- 飯村英樹 (2017). 測定と評価—妥当性と信頼性. 平井明代 (編) 教育・心理系研究のためのデータ分析入門 第2版 (pp. 1-19). 東京書籍

英語4技能試験情報サイト (n.d.). 英語4技能テスト利用入試がもたらした効用と課題.

Retrieved from [http://4skills.jp/selection/advanced/waseda\\_univ.html](http://4skills.jp/selection/advanced/waseda_univ.html)

尾崎茂 (2008). 言語テスト学入門：テスト作成の基本理念と研究法. 大学教育出版

熊澤孝昭 (2013). 学内開発プレースメントテスト得点解釈と使用の妥当性の評価について.

*JALT Journal*, 35(1), 73-100.

小泉利恵 (2018). 英語4技能テストの選び方と使い方—妥当性の観点から—. アルク

国立大学協会 (2017). 2020年度以降の国立大学の入学者選抜制度—国立大学協会の基本方

針—Retrieved from <https://www.janu.jp/news/files/20171110-wnew-nyushi1.pdf>

澤木泰代 (2011). 大規模言語テストの妥当性・有用性検討に関する近年の動向. *言語教育  
評価*, 2, 54-63.

清水裕子 (2004). 測定における妥当性の理解のために. *立命館言語文化研究*, 16(4), 241-254.

大学入試センター (2018). 参加要件の確認方法と確認結果の概要等. Retrieved from [https://www.dnc.ac.jp/albums/abm.php?f=abm00033007.pdf&cn=01\\_%E5%8F%82%E5%8A%A0%E8%A6%81%E4%BB%B6%E3%81%AE%E7%A2%BA%E8%AA%8D%E6%96%B9%E6%B3%95%E3%81%A8%E7%A2%BA%E8%AA%8D%E7%B5%90%E6%9E%9C%E3%81%AE%E6%A6%82%E8%A6%81%E7%AD%89.pdf](https://www.dnc.ac.jp/albums/abm.php?f=abm00033007.pdf&cn=01_%E5%8F%82%E5%8A%A0%E8%A6%81%E4%BB%B6%E3%81%AE%E7%A2%BA%E8%AA%8D%E6%96%B9%E6%B3%95%E3%81%A8%E7%A2%BA%E8%AA%8D%E7%B5%90%E6%9E%9C%E3%81%AE%E6%A6%82%E8%A6%81%E7%AD%89.pdf)

TEAP連絡協議会 (2016). TEAP連絡協議会レポート 第6回TEAP連絡協議会開催!.

Retrieved from [https://www.eiken.or.jp/teap/group/pdf/teap\\_council\\_report\\_20160607.pdf](https://www.eiken.or.jp/teap/group/pdf/teap_council_report_20160607.pdf)

寺沢拓敬 (2018). 政策研究の観点から見た「四技能入試」論議. 外国語教育メディア学会

(LET) 全国大会 (2018年8月9日). Retrieved from <https://www.slideshare.net/tterasawa/let2018>

鳥飼玖美子 (2018). 英語教育の危機. 筑摩書房

日本英語検定協会 (n.d.a). TEAP. Retrieved from <https://www.eiken.or.jp/teap/>

日本英語検定協会 (n.d.b). TEAPの特徴とメリット. Retrieved from <https://www.eiken.or.jp/teap/merit/index.html>

日本英語検定協会 (n.d.c). 大学入学英語成績提供システムに関する基本資料. Retrieved from <https://www.eiken.or.jp/exam-univ/info/>

日本英語検定協会 (2018). TEAP 4技能 (リーディング・リスニング・スピーキング・ライティング) スコアとCEFRレベル対応付け. Retrieved from <https://www.eiken.or.jp/teap/group/pdf/teap-score-cefr.pdf>

日本言語テスト学会 (2017). 大学入学希望者学力評価テスト (仮称) における英語テストの扱いに対する提言. Retrieved from [http://jlta2016.sakura.ne.jp/wp-content/uploads/2017/04/JLTA\\_proposal2017J.pdf](http://jlta2016.sakura.ne.jp/wp-content/uploads/2017/04/JLTA_proposal2017J.pdf)

南風原朝和 (編) (2018). 検証 迷走する英語入試—スピーキング導入と民間委託. 岩波書店

羽藤由美 (2019a). 2021年度 (2020年度実施) の大学入学共通テストにおける英語民間試験の利用中止を求めます. Retrieved from <https://nominkaninkyotsu.com/>

羽藤由美 (2019b). 新制度の問題点. Retrieved from <https://nominkaninkyotsu.com/problem/>

水本篤 (2014). 測定の妥当性と信頼性—よいデータの必須条件とは. 水本篤, 竹内理 (編). 外国語教育研究ハンドブック改訂版 (pp. 17-31). 松柏社

村山航 (2012). 妥当性概念の歴史的変遷と心理測定学的観点からの考察. *教育心理学年報*, 57, 118-130.

日本の大学入試における4技能外部試験使用の論証に基づく  
妥当性検証シミュレーション：早稲田大学を例に

- 文部科学省 (2013). グローバル化に対応した英語教育改革実施計画. Retrieved from [http://www.mext.go.jp/a\\_menu/kokusai/gaikokugo/\\_icsFiles/afieldfile/2014/01/31/1343704\\_01.pdf](http://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afieldfile/2014/01/31/1343704_01.pdf)
- 文部科学省 (2014). 今後の英語教育の改善・充実方策について 報告(概要)～グローバル化に対応した英語教育改革の五つの提言～. Retrieved from [http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/102/houkoku/attach/1352463.htm](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/houkoku/attach/1352463.htm)
- 文部科学省 (2016). 高大接続システム改革会議「最終報告」【概要】. Retrieved from [http://www.mext.go.jp/component/b\\_menu/shingi/toushin/\\_icsFiles/afieldfile/2016/06/02/1369232\\_02\\_2.pdf](http://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/afieldfile/2016/06/02/1369232_02_2.pdf)
- 文部科学省 (2017a). 「大学入学共通テスト」について. Retrieved from [http://www.mext.go.jp/component/a\\_menu/education/micro\\_detail/\\_icsFiles/afieldfile/2017/10/24/1397731\\_001.pdf](http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2017/10/24/1397731_001.pdf)
- 文部科学省 (2017b). 大学入学共通テスト実施方針. Retrieved from [http://www.mext.go.jp/component/a\\_menu/education/micro\\_detail/\\_icsFiles/afieldfile/2017/10/24/1397731\\_001.pdf](http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2017/10/24/1397731_001.pdf)
- 文部科学省 (2018a). 平成30年度「英語教育実施状況調査」の結果について. Retrieved from [http://www.mext.go.jp/a\\_menu/kokusai/gaikokugo/1415042.htm](http://www.mext.go.jp/a_menu/kokusai/gaikokugo/1415042.htm)
- 文部科学省 (2018b). (大学入試改革) 民間の英語4技能試験の結果の提供について (平成30年8月10日). Retrieved from [http://www.mext.go.jp/a\\_menu/koutou/koudai/detail/1408090.htm](http://www.mext.go.jp/a_menu/koutou/koudai/detail/1408090.htm)
- 早稲田大学文学部 (n.d.a). 教育理念. Retrieved from <https://www.waseda.jp/flas/hss/about/policy/>
- 早稲田大学文学部 (n.d.b). 入学試験情報. Retrieved from <https://www.waseda.jp/flas/hss/applicants/admission/>
- 早稲田大学文化構想学部 (n.d.a). 教育理念. Retrieved from <https://www.waseda.jp/flas/cms/about/policy/>
- 早稲田大学文化構想学部 (n.d.b). 入学試験情報. Retrieved from <https://www.waseda.jp/flas/cms/applicants/admission/>

# Simulation of an Argument-Based Approach to Validate the Use of External Four-Skill Tests as University Admission Tests in Japan: The Case of Waseda University

TAHARA Tatsuro

There has been a heated discussion regarding the introduction of four-skill tests for use as university admission tests in Japan starting in 2020 as an alternative to the English test in the National Center Test for University Admissions. However, this discussion seems to overlook that whether universities develop in-house tests themselves or use commercial tests, it is necessary to conduct test validation according to the purposes of test uses within the context of each university. Language testing (language assessment), an academic field of applied linguistics, can offer a new perspective on this discussion. Specifically, argument-based approaches to test validation, which have been investigated and used in language testing, can provide a clear view of how to conduct research on choosing appropriate external tests as new university admission tests in Japan. To better understand how to use the framework of test validation, a simulation of test validation was performed for the use of the TEAP® Test for admission to the Faculty of Letters, Arts and Sciences at Waseda University using an Assessment Use Argument (AUA, Bachman & Palmer, 2010) — a framework of the argument-based approach to test validation. Although published information on the use of the TEAP® Test at Waseda University is limited, AUA can provide an example of how to conduct test validation for the use of external tests as university admission tests and can indicate what kind of research is needed to obtain more information on improving the use of tests at the university. The test validation procedure using AUA will be helpful for testing users, including university stakeholders who consider the use of external tests, and for researchers conducting test validation research on the use of new tests in specific contexts.