

Socio-cognitive Framework を用いた英文和訳課題の妥当性検証 — 高校の定期テストを中心に —

黒川 智史¹

¹ 東京大学大学院 総合文化研究科 言語情報科学専攻 〒153-0041 東京都目黒区駒場 3-8-1

E-mail: ¹ skurokawa0521ut@gmail.com

あらまし 本稿では、Socio-cognitive Framework で高校の定期テストに出題されている英文和訳の妥当性を検証した。出題に値する利点としては、(1) 教師側では英文和訳の問題作成が容易、(2) 実施環境に左右されにくい、(3) 生徒のメタ言語能力の向上に貢献する可能性があるという 3 点が示された。妥当性を損なう可能性のある点としては、(1) 出題意図が不明確だと授業内容とテスト内容の間にズレが生じる、(2) テキストの文化的な背景知識が必要な問題は訳すことが困難な場合がある、(3) 指示文が曖昧な場合、テスト受験者がどのように回答すべきか不明瞭、(4) 産出スキルとして適切な採点方法が実施されていない可能性がある、(5) 適切な外的基準と英文和訳のスコアの相関分析ができていない、(6) 授業で扱ったテキストをそのまま出題すると負の波及効果が発生するということが示された。その結果、検証した英文和訳は高校の定期テストの出題に値するほど妥当性がないという結論に至った。

キーワード Socio-cognitive Framework, 英文和訳, 定期テスト,

Study on English to Japanese Translation Tests using Socio-cognitive Framework

Satoshi KUROKAWA¹

¹ Department of Language and Information Sciences, Graduate School of Arts and Sciences, the University of Tokyo

E-mail: ¹ skurokawa0521ut@gmail.com

Abstract This validity study investigated English to Japanese translation tests in high school's term-end examinations using Socio-cognitive Framework (Weir, 2005). The study revealed that merits of English to Japanese translation tests (1) are easy to make the test from teachers' point of view, (2) are not influenced by the test environment, (3) evoke test takers' meta-linguistic awareness. However, the translation tasks (1) have the gap between the teaching task and teaching task because few teachers did not clarify the object of this task (2) find it difficult to translate cultural context, (3) have ambiguity of the directive statement, (4) can be a lack of scoring validity, (5) can be a lack of criterion-related validity, (6) may have negative washback effect on the test takers' learning. Given that the large number of lacks validities, English to Japanese translation tasks should not be on the test unless the tasks remedy lack of validities.

Keywords Socio-cognitive Framework, English to Japanese Translation Tests, Term-end Examinations

1. はじめに

2020 年度より、従来のセンター試験を廃止し、記述式問題などを取り入れた「大学入学共通テスト」が導入されることに注目が集まっている。しかし、もう 1 つ重要な改革が行われる。その改革とは、2020 年度から一般入試において、受験者の各高校が提出する調査書を、その大学のアドミッションポリシーに基づき、可否判定に用いることができるようになる、というこ

とである (文部科学省, 2017)。そのため、調査書に記入される学内の中間テストや期末テストなどの総称であるいわゆる「定期テスト」のスコアも、生徒の志望校次第では、大学入試で評価対象になる。それゆえ、高校生にとって定期テストの重要度が上がるだけでなく、テストを作成する高校教師は、より一層、公正公平な定期テストを実施し、妥当性・信頼性の高い定期テストの作成に努めなくてはならない。つまり、これを契機に高校の定期テストの問題形式の妥当性と信頼

黒川智史, “Socio-cognitive Framework を用いた英文和訳課題の妥当性検証—高校の定期テストを中心に—”,
言語学習と教育言語学 2019 年度版, pp. 17-25,

日本英語教育学会・日本教育言語学会合同編集委員会編集, 早稲田大学情報教育研究所発行, 2020 年 3 月 31 日.

Copyright © 2019-2020 by Kurokawa, S. All rights reserved.

性を検証し、高めていかなければならないだろう。

高校の英語教育に焦点を当てると、妥当性と信頼性に疑念を抱かせるような問題形式が今もなお習慣的に全国の高校の定期テストで用いられている(若林 & 根岸, 1993; 村越 & 江原, 2015)。とりわけ英文和訳に関しては、解答するのに日本語能力が必要であることから、英語のテスト問題として出題すべきかについて議論されてきた(斎藤, 2006; 馬場, 2006; 静, 2006)。しかしながら、これまでの議論は、研究者自身の学生時代の経験や、教育者としての指導方針に基づいて議論しているものが多く、あまり言語テストの観点から議論されてこなかった。また、そもそも研究者ごとに議論の対象にしている英文和訳自体が異なっていることも多かった。例えば、雑誌『英語青年』において、斎藤(2006)は、英文和訳を文化的な背景をうまく日本語に落とし込むような高度な翻訳能力を測るものであるとみなして議論している。一方で、静(2006)は、文化的な背景をあまり絡めずに表層的な英語と日本語を訳し、授業で用いられる文構造や語彙知識などの能力を測る課題として、英文和訳とみなして議論していた。そのため、英文和訳自体のコンセンサスが取れていない状態で議論していたといえる。したがって、まず今回検証する英文和訳を提示してから、言語テストの観点から検証し、出題に値する利点や妥当性が損なわれている可能性がある点について様々な観点から検証し、高校の定期テストで英文和訳を出題すべきかの議論を進めることに貢献できるだろう。

近年、言語テスト研究において、「社会的認知の枠組み(Socio-cognitive Framework)」がテストの開発に用いられている(Weir, 2005; Green, 2014; Bannur and Abidin, 2015)。社会的認知の枠組みは、Weir(2005)が提唱したテストの妥当性検証アプローチであり、Green(2014)によると、Test of English for Academic Purpose (TEAP)の開発の際に社会的認知の枠組みが用いられたという。また、Bannur and Abidin(2015)では、リビアの大学内で実施された特定の学習者向けの英語テストの妥当性の検証にも社会的認知の枠組みが用いられた。したがって、日本独自の問題形式である英文和訳の検証にも応用可能であると考えられる。

そこで、本研究では、社会的認知の枠組みを用いて、高校の定期テストで英文和訳を運用することによって、どのような出題に値する利点と妥当性を損なわれている可能性があるのかを検証し、高校の定期テストで英文和訳は出題するほど妥当性があるのかを考察する。

2. 社会的認知の枠組み

2.1 社会的認知の枠組みとは

社会的認知の枠組みでは、5つの妥当性の構成要素

に分かれており、それぞれ異なる検証を行っている。文脈妥当性(Context Validity)と理論基盤妥当性(Theory-based Validity)は、テスト実施前に検証する構成要素とされており、採点妥当性(Scoring Validity)、基準準拠妥当性(Criterion Related Validity)、そして結果妥当性(Consequential Validity)は、テスト実施後に検証する構成要素となっている(Weir, 2005)。また、社会的認知の枠組みでは、妥当性と信頼性は不可分であるとされているため、構成要素の中には、妥当性と信頼性に関する項目が共存しているのが特徴である。したがって、以降本稿では特に但し書きがない限り、「妥当性」という言葉は社会的認知の枠組みにおける、信頼性を指すことにする。

社会的認知の枠組みは、本来リーディング、リスニング、ライティング、スピーキングの課題を検証するものであり、それらの4技能に属さない日本独自の課題である英文和訳を対象にするため、4技能の枠組みをそのまま用いずに、英文和訳の特性を考慮しながら検証する。なお、妥当性の検証の判断材料とするものは、3.1で述べる問題形式の英文和訳とする。次のページにある図1はWeir(2005, p.44 - 47)を基に作成した社会的認知の枠組みに関するものである。2.では社会的認知の枠組みの構成要素の詳細を説明し、第3章では、定期テストの英文和訳に関する、出題に値する利点と妥当性が損なわれている可能性がある点があるのかを考察する。

2.2 文脈妥当性

文脈妥当性は、従来、内容妥当性(Content Validity)と言われていたものである(Weir, 2005)。Weirは、テストを社会的な文脈において論じるために、内容妥当性という用語ではなく、文脈妥当性としている。文脈妥当性は、要求タスク、実施環境、タスク設定の3種類に分かれている。

要求タスクとは、その課題によって、どのような能力を要求されるかを指している。例えば英文和訳であったら、課題で使用されている語彙レベルや、和訳するテキストの長さなどが挙げられる。

実施環境とは、その課題を行う場所が、受験者が安心して集中できる環境なのかを指している。安心な環境とは、テスト中に不当に妨害されることがない場所であることなどが挙げられる。集中できる環境とは、例えば、リスニング課題では、音声テスト受験者の座席にかかわらず、全員聞こえているかどうかなどが挙げられる。

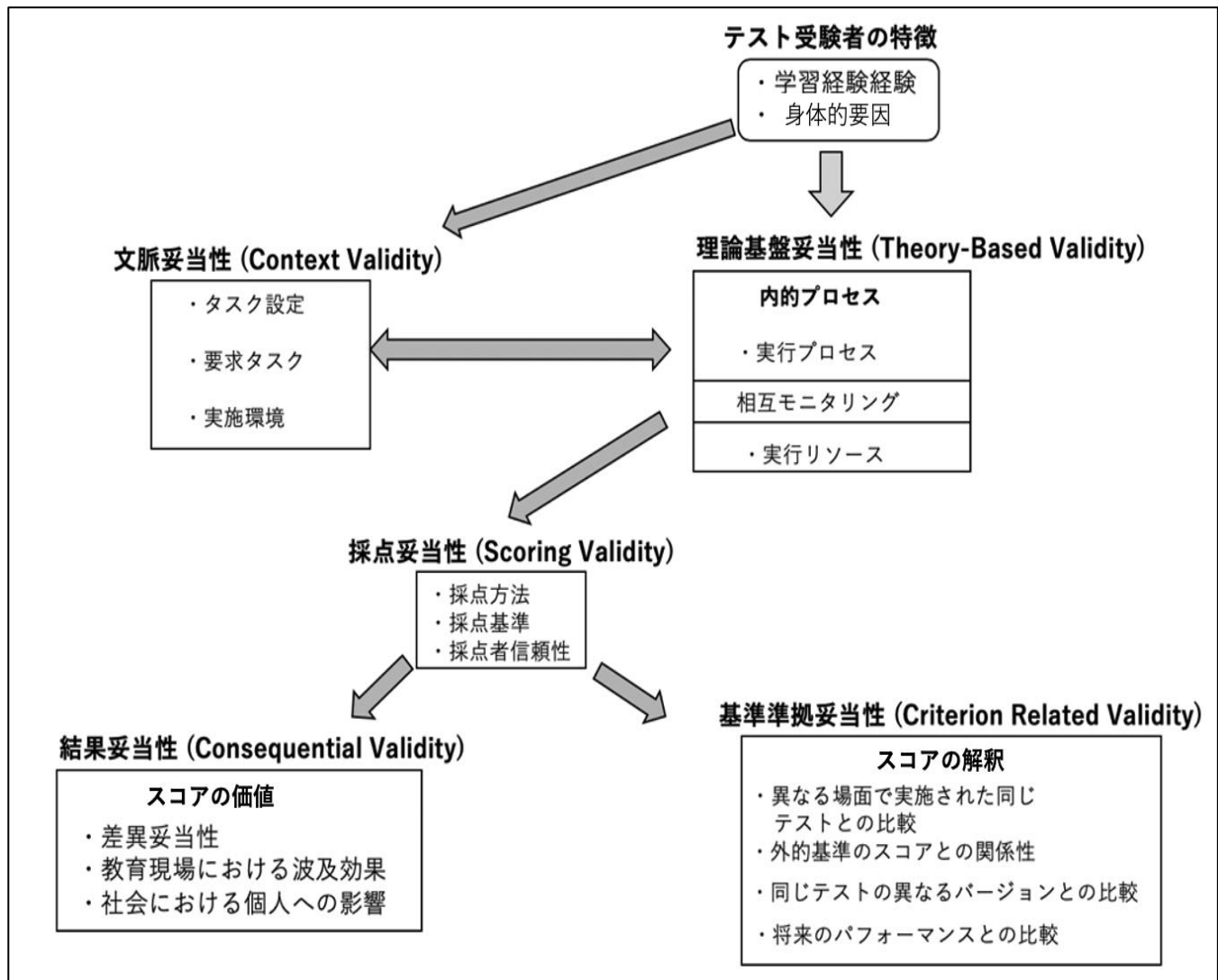


図 1 Weir (2005) の社会的認知の枠組みの日本語訳 (pp.4 4-47)

タスク設定とは、そのタスクを設定した目的や、解答方式、課題の数や、予想される解答時間など、課題のフォーマットや解答形式に関するものを指している。このように文脈妥当性は、従来の内容妥当性の検証に加えて、テスト受験者の実施環境にまで配慮していることが特徴といえる。

2.3 理論基盤妥当性

文脈妥当性が、テストを実施する環境やテスト課題自体など「受験者の外的なプロセス」に関するものに対し、理論基盤妥当性は、実際にテスト受験者がその課題を解いているときにどのように解答しているのかという「受験者の内面的なプロセス」に関するものである。理論基盤妥当性には、実行リソースと実行プロセスに分かれている。実行リソースは、その課題に解答するために、どのようなスキルが必要であるかを指している。例えば、長文読解のような課題では、語彙力や文法能力が不可欠であるため、それらの知識が無い場合は解答できないだろう。他にも、受験者が課題

の内容に関する前提知識 (Content Knowledge) がどの程度必要なのかを考慮することである。実行プロセスでは、テスト受験者の課題目標設定、課題内容の認識などが挙げられる。実行プロセスと実行リソースは課題に解答し終わるまで、テスト受験者の頭の中で相互にモニタリングが行われていると考えている。つまり、理論基盤妥当性の考え方では、実行プロセスが実行リソースに適しているかを受験者自身が調整して課題を遂行しているということになる。このように、課題を解く際の「受験者の内面的なプロセス」がテスト作成者の意図したものであるかを検証するものが理論基盤妥当性であるといえる。

2.4 採点妥当性

テスト受験者が課題を解答後、テストを回収し、採点する。その際の妥当性が採点妥当性である。採点妥当性はリーディングやリスニングのような受容スキル (Receptive Skills) か、ライティングやスピーキングのような産出スキル (Productive Skills) によって大きく

異なる。前者の場合は、選択式問題が多数を占めているため、テスト項目が意図した能力を測定できているのか、または、テスト項目が一貫して同じ能力を測定しているのかを検証することが主な調査対象である。後者の場合は、採点者間の信頼性があるのか、採点者内の採点の一貫性があるのか、または、採点項目、採点基準として妥当なものが設定されているかなどが検証の対象となっている。

2.5 基準準拠妥当性

採点が終わった後、「そのスコアの価値」を知ることができれば、課題に相応の価値をつけることができるだろう。基準準拠妥当性は、同じようなテスト課題とどの程度相関があるのか、または、その課題のスコアが、他の英語テストのスコアとどの程度相関があるのかなどを調査することで、「その課題のスコアがどのような意味があるのか」を検証することを指している。

2.6 結果妥当性

基準準拠妥当性が「そのスコアの価値」に関する検証であるのに対し、結果妥当性は、「そのスコアの解釈」に関する検証である。具体的には、その課題のスコアが、どのような社会的価値があるのか、その課題のスコアが、テスト受験者の将来のパフォーマンスをどの程度約束するものなのか、または、その課題を用いることの波及効果はどのようなものかを検証することである。

3. 定期テストの英文和訳の妥当性検証

本項では、実際に高校の定期テストなどで運用された英文和訳を取り上げ、2.で言及した、社会的認知の枠組みで、先行研究に基づいて、定期テストの妥当性を検証する。

3.1 今回検証する英文和訳

第1章で述べたように、研究者が議論している英文和訳にも様々な種類がある。例えば、2つの似た意味の英文が提示されて「文章の違いが分かるように訳せ」という指示がされる問題などもある（若林&根岸, 1993）。したがって、本研究で検証する英文和訳とはどのような問題形式なのかをあらかじめ定義する必要があるだろう。実際に定期テストで使用されている英文和訳を提示するために、若林と根岸（1993）が収集したある高校の「総合問題」形式のテストで、実際に使われたテストの中から英文和訳を抜粋した。

設問 2 下線部 (イ) の意味を日本語で書きなさい。

I liked her very much, in fact, that (イ) I would go to the library and read the cards in the back of the books to find the ones that she had borrowed.

解答例: 私は図書館に行き、彼女が借りた本を見つけるために、本の裏にある閲覧カードを読もうとした。

(若林 & 根岸, 1993, pp. 32)

このように、英文を「日本語で書きなさい」と指示されるような問題形式が英文和訳であるといえる。また、中沢(2014)によると、進学校の高校生が中心に受験しているとされている「進研模試」では、以下のような英文和訳が出題されているという。

例 1: 下線部(A)を日本語になおせ。

(A) No one knows for sure when Harrison's interest in clocks began.

解答例: いつハリソンが時計に興味を持つようになったのかは、誰にもはっきりとはわからない。

(中沢, 2014, pp. 92)

模擬試験は定期テストではないため、調査書の内容には影響がないと考えられるが、進学校においては、模擬試験を参考にして定期テストを作成している可能性が高い。また、(1)文中の下線部の英文を訳させている、(2)「日本語で書きなさい」や「日本語になおせ」といった指示文などは、若林と根岸（1993）が提示した英文和訳と一致している点が多い。

今回提示した1つ目の課題の訳の解答例において、若林と根岸（1993）は、「図書館に行って、後ろにある（閲覧）カードを読んだ」という訳では、点数がもらえず、ポイントは助動詞の *would* の訳であると述べている(pp.35)。また、2つ目の課題の訳のポイントは、中沢によれば文全体が *SVO* の文型を取っていること、*no one* と *for sure* の意味を理解しているか、目的語の部分に間接疑問 (*When*~)があること、および、*When* の内部の文型を理解しているか、ということであるという(pp.93)。

このように英文和訳が採点されるポイントを1.で言及した英文和訳が測定している能力に関する議論と照らし合わせると、文化的背景を交えた高度な翻訳能力を測定している齋藤（2006）よりも、どちらかというと文化的背景を絡めず逐語訳で文法、英語の構文知識を問うような能力を測定している静（2006）に近いと考えられる。したがって本研究においては、逐語訳

で文法、英語の構文知識を問う問題として英文和訳について議論することにする。

3.2 社会的認知の枠組みを用いた英文和訳の検証

3.2.1 英文和訳の文脈妥当性

英文和訳の文脈妥当性が該当する、要求タスク、実施環境、タスク設定を考察する。要求タスクは、テスト受験者、つまり、その高校の生徒の英語力や、授業で実際に用いた語彙レベルやテキストの長さなどを調整すれば妥当な問題になるだろう。また、テスト作成者側から見れば、和訳をさせたい部分に下線を引けば問題が完成するため、テスト作成が容易であることも出題に値する利点として考えられる (馬場, 2006)。

実施環境に関して英文和訳は、リスニング課題のように、テスト受験者の座席によって音源の聞こえやすさが異なる可能性などを配慮する必要はない。そのため、通常の定期テストを実施する環境が整っていれば妥当性を確保できるだろう。それゆえ、英文和訳は実施環境に左右されにくい問題形式であり、その部分は出題に値する利点であると考えられる。

その一方、タスク設定に関して考察すると、若林と根岸 (1993) によると、特に「総合問題」型のテストにおいて英文和訳は、目的意識がなく出題されている場合があると指摘している。村越と江原 (2015)は、普段の授業内容 (ティーチング・ポイント) と、テストの内容 (テストティング・ポイント) が一致させることが重要であると述べている。英文和訳は、ティーチング・ポイントとテストティング・ポイントが一致させるには、授業内で和訳を提示しなくてはならないだろう。その際に、英文和訳を出題する目的が不明確だと、授業内容とテスト内容の間にズレが生じてしまうだろう。さらに、現行の学習指導要領におけるコミュニケーション重視の英語教育、そして 2020 年度の改定により、4 技能をバランス良く伸ばす英語教育が施行され、和訳を授業のティーチング・ポイントとして設けることはますます困難である可能性がある。今後、教師は、授業のティーチング・ポイントに和訳をあまり組み込んでいない場合、安易に前例を踏襲して英文和訳を出題すると、定期テストの妥当性を損なってしまう可能性がある。そのことに留意し、教師が多肢選択など他の方法を用いて、英文和訳で測りたい能力を測定できないかを検討していく必要があるだろう。

3.2.2 英文和訳の理論基盤妥当性

理論基盤妥当性の実行リソースと実行プロセスを英文和訳に適応して考察する。実行リソースに関しては、その高校の授業で扱っている文法事項や、語彙を使えば過度に負担のかかる課題とはならず、妥当性は確保できるだろう。これは文脈妥当性で述べた言語要

求タスクが生徒にとって妥当なものであるならば、大きな問題になりにくいと同様である。しかしながら、課題の前提知識に関してはテスト作成時に考慮する必要がある。例えば、自国の文化と大きく異なる文化が関連する出題に関しては、たとえその語彙を知っていても、和訳をすることが困難である可能性がある。これは、訳すという行為の中に、異なる文化的背景を有する二言語間を行き来しなければならないためである。英文和訳も例外ではないため、もし、前提知識を必要とする問題を出題する場合は、補足するような注意書きなどを提示する方法が考えられる。

実行プロセスに関しては、英文和訳は実行プロセス上の妥当性が損なわれている可能性がある点が存在する。Weir (2005) は、参加者全員が同じようなプロセスで課題を行うことが望ましいため、指示文は誰でも理解できるようにすべきであると主張している。その一方、若林と根岸 (1993) のように、英文和訳は「この文を日本語になおしなさい」という指示文であると、原文の文構造をできるだけそのまま反映させた訳にするのか、それとも前後の文脈から自然な日本語になるように訳すのかが、分からないという指摘している。この点に関しては、ただ単に「この文を和訳しなさい」という曖昧な指示ではなく、テスト受験者が誤解しないような指示文にすることが求められるだろう。先行研究においてもしばしば英文和訳の指示文に関して述べられていることがある。例えば、Ushiro et al (2005) では、英文和訳の指示文を明瞭にすべきであると主張している。しかしながら、両研究とも、実際にどのような指示文にすれば明瞭になるかまではあまり述べられていなかった。それゆえ、今後は、どのような指示文ならば授業のティーチング・ポイントを反映させられるかを考察する必要があるだろう。

3.2.3 英文和訳の採点妥当性

英文和訳の採点妥当性を検証する。英文和訳は、記述式問題であるため、ライティング課題やスピーキング課題のような産出スキルとみなすことができる。その場合、採点方法は、ライティング課題のような採点方法で行う必要があるだろう。つまり、英文和訳も採点項目や採点基準を設ける必要があるということである。さらに英文和訳の場合は、採点の一致度を高めるためには、採点例など複数を設け、採点妥当性を高める方法といえるだろう。

しかしながら、英文和訳の採点の信頼性に関する麻生 (2012) などの先行研究においては、具体的な採点項目が用いられていないものが多く、採点者の裁量に任されている場合が多く、採点者のトレーニングの機械や、採点項目を設けていない研究が多い (Buck, 1992; 麻生, 2012)。英文和訳 2020 年度の大学入試より

調査書が非常に重要になってくることを考慮すると、高校の定期テストにおいて採点の妥当性が確保されていない可能性がある状態は、他の妥当性と比べても、とりわけ大きく妥当性が損なわれている可能性がある。

英文和訳の採点の信頼性を調査するために、黒川(2019)では、ライティング課題と同様に、採点項目を設置し、採点者のトレーニングの機会を設け。その結果、採点者間信頼性は、非常に高かったという($r = .95$)。黒川の研究では、採点者数が少ない(2名のみ)ことや、採点項目に挙げていた4項目同士の一貫性に懸念がある(i.e., 日本語能力も採点項目に入っている点など)など多くの課題が残っている。また、採点者内信頼性などの分析も行われていない。それゆえ、黒川(2019)の研究に不足している上記の点を今後検証していくことで、英文和訳の採点妥当性を検証できるだろう。

3.2.4 英文和訳の基準準拠妥当性

「英文和訳のスコアの価値」は、主に「英文和訳のスコアと英語力と関係性」という文脈において検証されてきた(Buck, 1992; 青木, 2000)。代表的なものでいえば、青木(2000)がある。青木は、アメリカのミシガン大学が留学生の英語レベルを分けるために使用している、Michigan Test of English Placement Test の Reading Comprehension に該当する20問と、英文和訳の20問を参加者に解答させ、ピアソンの相関分析を行った。その結果、両者は有意に弱い相関であったという($r = .36, p < 0.01$)。しかしながら、金谷(1995)が指摘しているように、英文和訳は読解能力のみを測定しているのではないため、青木のように、英文和訳と読解力のテストの間が弱い相関になるのは当然であり、その2つの相関係数を求めることの根拠が乏しい。

また、青木(2000)の場合は、採点基準は各採点者に委任しており、また、採点妥当性の部分でも述べたように、トレーニングの機会を設けておらず、採点妥当性が保たれているか懸念があった。高校の定期テストにおいて、より正確に基準準拠妥当性を検証するには、高校英語の到達度を測定できる課題と英文和訳の相関を見る必要があり、その際には、採点妥当性も考慮して検証しなくてはならないだろう。

また、英文和訳は図1の右下の基準準拠妥当性の中にある、「異なる場面で実施された同じテストの比較」や、「同じテストの異なるバージョンとの比較」、「従来のパフォーマンスとの比較」に関しても、今後検証していく必要があるだろう。

例えば、「異なる場面で実施された同じテストの比較」に関していえば、英文和訳は記述式問題であるものの、あまり日本語に触れる機会がなく、外国語に多く触れる海外の日本人学校の生徒と、日本在住の公立校高校の生徒では、異なる結果になる可能性も高く、

和訳の違いを記述的に分析することができるだろう。

「同じテストの異なるバージョンとの比較」、「従来のパフォーマンスとの比較」に関していえば、同じ高校の定期テストにおいて、毎回出題していき、どのようにテストのバージョンが変化することで、生徒のパフォーマンスが変化するのか、また、従来のパフォーマンスから、徐々に向上していくのかを比較することができるだろう。

3.2.5 英文和訳の結果妥当性

英文和訳の結果妥当性は、多くの議論がある。最も代表的なものは、波及効果に関するものである(静, 2006)。静は、大阪大学や京都大学が二次試験において英文和訳が出題しているため、学校の定期テストにおいても出題されていると主張している。しかしながら、2022年度より新学習指導要領施行に伴い、高校の授業は、高校の授業にリーディング、リスニング、ライティング、スピーキングの4技能をバランス良く指導されるため、高校の定期テストも4技能を測定することを取り入れる可能性がある。それに伴い、大学入試において入学試験内容の見直しが行われ、英文和訳の出題数も減少する可能性がある。そのため、今後英文和訳が定期テストに出題されることによる波及効果は現在よりも弱くなると考えられる。

また、教師が英文和訳を授業で扱った既習テキストをそのまま出題してしまうことも懸念材料である。テスト受験者である生徒は、次のテストに備えて、授業で扱ったテキストの和訳を暗記する可能性が出てくる。既習テキストの和訳を暗記することは、英語力と関係がなく、実用性も低いいため、負の波及効果が生じる可能性があると考えられる。また、生徒が既習テキストを暗記することは英語学習であるという錯覚を起こしてしまう可能性も否定できない。そのため、卯城(2012)が述べるように、生徒の「授業の理解度」や「取り組み状況」を測るなどの目的がない限り、授業で扱ったテキストをそのまま使うことは避ける必要があるだろう。定期のテストでは、生徒の語彙、文法事項などの学習状況を加味しながら、新たなテキストをテストに用いる、もしくは、授業で扱った内容を教科書以外の語彙や表現を用いて言い換えるなどの工夫がある(卯城, 2012)。そうすることで、テストの結果妥当性を高めることに貢献するだろう。

また、今回は静(2006)の定義により近いものが英文和訳の測定したいものであるとしていたが、仮により翻訳能力を測定するような斎藤(2006)の定義で測定しようとする場合、江利川(2011)によると、英文和訳は、日英の二言語間を行き来するプロセスが含まれており、そのプロセスが知的訓練に繋がると主張している。仮に英文和訳が生徒の知的訓練に役立つならば、

定期テストに出題する意味があるともいえる。鳥海、大津、斎藤、江利川と林 (2017) は英語教育の目的の1つとして、メタ言語能力を身につけることを主張している。メタ言語能力とは、言語についての文法や構文など言語のメタ的な機能を理解し、正しく運用できる能力のことを指す (松崎, 1991)。確かに、英文和訳が生徒の知的訓練やメタ言語能力の育成に貢献するならば、定期テストに出題する意味があるだろう。しかしながら、英文和訳を解答する際、生徒の内面では実際にどのようなプロセスがあるのかは解明されておらず、また、英文和訳によってメタ言語能力が向上するかは定かではない。今後は、英文和訳がメタ言語能力向上させるのかを実証的に検証する必要があるだろう。例えば、受験者に英文和訳解答中してもらい、解答の最後に発話プロトコル分析することで、英文和訳を解いている間にどのようなメタ認知が働いているのかを解明する手がかりになるだろう (長谷川, 2016)。

4. 議論:英文和訳は高校の定期テストに必要なか

第3章で指摘した高校の定期テストにおける英文和訳の出題に値する利点と妥当性が損なわれている可能性がある点を整理すると以下ようになった。

- ・ 出題に値する利点

- (1)教師側は英文和訳の問題作成が容易
 - (2)実施環境に左右されにくい
 - (3)生徒のメタ言語能力の向上に貢献する可能性がある
- ・ 妥当性が損なわれている可能性がある点
- (1)出題意図が不明確だと授業内容とテスト内容の間にズレが生じる
 - (2)テキストの文化的な背景知識が必要な問題は正確な和訳が困難な場合がある
 - (3)問題の指示文が曖昧な場合、テスト受験者がどのように回答すべきか不明瞭
 - (4)産出スキルとして適切な採点方法が実施されていない可能性がある
 - (5)適切な外的基準と英文和訳のスコアの相関分析ができていない

- (6)授業で扱ったテキストをそのまま出題すると負の波及効果が発生しうる

このように、社会的認知の枠組みを用いて、定期テストの英文和訳を運用する際の出題に値する利点と妥当性が損なわれている可能性がある点を挙げた。本項では、各妥当性の検証を統合し、英文和訳に出題すべきかについて考察する。

英文和訳には、教師が容易に作成でき、実施環境に左右されにくいことが示された。このような出題に値する利点があるからこそ、英文和訳はこれまで長きにわたり高校の定期テストに出題されてきたと考えられる。他にも、英文和訳の出題に値する利点として考えられることとして、生徒のメタ言語能力を向上させる可能性があるということが示唆された。英文和訳とメタ言語能力の関係は未だ解明されておらず、今後英文和訳によってメタ言語能力が向上するのかを調査する必要もあるだろう。

しかしながら、英文和訳には多くの妥当性を損なう妥当性が損なわれている可能性がある点も社会的認知の枠組みにより示された。英文和訳の採点妥当性が確保されていることは、とりわけ今後の高校の定期テストにおいて重要になる。採点の妥当性が確保されていない最も大きな理由として、出題意図が不明確であることが挙げられる。出題意図が不明確であると、英文和訳のどのような部分を評価し、また、評価しないと決めることができない。また、英文和訳の評価方法が決まっていないことが、問題の指示文が曖昧になってしまうこと、そして、適切な外的基準と英文和訳のスコアの相関分析ができないことにつながってくると考えられる。それゆえ、英文和訳の妥当性は、社会的認知の枠組みにおける5つの妥当性のうち一つも完全に満たしていない。したがって、今回社会的認知の枠組みで検証したように、英文和訳は妥当性を損なうような部分が多く散見されたため、現状、高校の定期テストに出題に値するほどの妥当性がないという結論に至った。

しかし、これは飽くまで今回検証した英文和訳の妥当性検証を行った結果であり、教師が上記の妥当性が損なわれている可能性がある点を十分に理解し、授業のティーチング・ポイントに組み込んでいけば、英文和訳も出題可能だろう。今回の検証した妥当性が損なわれている可能性がある点を改善できれば、高校の定期テストに出題するに足りうる妥当性を有する英文和訳を出題することも可能であろう。その意味で、言語テストの観点から、英文和訳の妥当性が損なわれている可能性がある点を指摘できたといえる。今後

は、どのような英文和訳が社会的認知の枠組みで検証してきたような妥当性を確保できるのかを調査する必要があるだろう。

5. おわりに

本稿で検証した課題は、長きにわたり、全国の定期テストにおいて出題されているにもかかわらず、その課題の妥当性に関して、教師や研究者が疑念を抱いている問題形式である英文和訳とした。まず、社会的認知の枠組みで定期テストにおける英文和訳の妥当性を考察し、多くの出題に値する利点と妥当性が損なわれている可能性がある点を指摘することができた。具体的には、英文和訳を出題する出題に値する利点としては、(1)教師側は英文和訳の問題作成が容易、(2)実施環境に左右されにくい、(3)生徒のメタ言語能力の向上に貢献する可能性があるという3点が示された。また、妥当性が損なわれている可能性がある点としては、(1)出題意図が不明確な場合、がある、(2)テキストの文化的な背景知識が必要な問題は訳すことが困難な場合がある、(3)指示文が曖昧な場合、テスト受験者がどのように回答すべきか不明瞭、(4)産出スキルとして適切な採点方法が実施されていない可能性がある、(5)適切な外的基準と英文和訳のスコアの相関分析ができていない、(6)授業で扱ったテキストをそのまま出題すると負の波及効果が発生するという6点が示された。

上記の妥当性検証を整理し、定期テストにおける英文和訳の出題に値する利点と妥当性が損なわれている可能性がある点を整理し、本稿で議論したような問題形式の英文和訳を出題すべきではないという結論に至った。その一方で、英文和訳は実施環境に左右されにくいということが示唆も得られた。

本来、社会的認知の枠組みを用いたテスト課題の妥当性検証は、4技能の課題に妥当性があるかを検証するものであり、定期テストの4技能に属する課題の妥当性を検証する際に用いることが可能である。このように、社会的認知の枠組みを用いて定期テストの他の課題の妥当性の検討を行うことも視野に入れる必要があるだろう。問題形式に関しては、今回検証した英文和訳だけではなく、選択式問題などを検証することで、より高校の定期テストの改善することに繋がるだろう。

なお、本研究の限界は、Weir (2005)のみで検証していることが挙げられる。例えば、O'Sullivan and Weir (2011)などは、Weir (2005)の社会的認知の枠組みに、これまでの妥当性検証の要素を加えたものを提案している。それゆえ、今後は、他の妥当性検証の枠組みで英文和訳を検証することで新たな知見を見つけることができるだろう。

参考文献

- [1] 文部科学省, “平成31年度大学入学者選抜実施要項” Retrieved May 28, 2019, from http://www.mext.go.jp/component/a_menu/education/detail/___icsFiles/afieldfile/2018/06/07/1282953_02_1.pdf
- [2] 若林俊輔, 根岸雅史, “無責任なテストが「落ちこぼれ」を作る: 正しい問題作成への英語授業学的アプローチ”, 大修館, 東京, March, 1993.
- [3] 村越亮治, 江原美明, “高校現場における「英語テスト」と「観点別評価」の課題” 神奈川県立国際言語文化アカデミア紀要, vol.4, pp.1-18, May 2015.
- [4] 斎藤兆史, “英文和訳、大いに結構” vol.152, No.1, pp.2-5, April, 2006.
- [5] 馬場哲生, “英文和訳テストの功罪,” 英語青年, vol.152, no.7, pp.408-410, October 2006.
- [6] 静哲人, “これでいいのか、大学入学英語問題-英語教育およびテスト理論の立場から,” 英語青年, vol.152, No.1, pp.2-5, April, 2006.
- [7] C. J. Weir, *Language Testing Validation: An Evidence-Based Approach*, Palgrave Macmillan, Hampshire, November 2005.
- [8] A. Green, The Test of English for Academic Purposes (TEAP) impact study: Report 1—Preliminary questionnaires to Japanese high school students and teachers, pp.1-47, Tokyo: Eiken Foundation of Japan, 2014.
- [9] F. Bannur, S. A. Z Abidin, A. Jamil “A Validation Process of ESP Testing Using Weir's Socio-cognitive Framework (2005). *Procedia-Social and Behavioral Sciences*, 202, pp.199-208 August 2015.
- [10] 中沢敏浩, “模擬試験再考: 和訳がなくなるとはかもしれないもう一つの理由”, 中国地区英語教育学会研究紀要, vol.44, pp.91-96, March 2014.
- [11] 卯城祐司(編). “英語リーディングテストの考え方と作り方”, 研究社, 東京, September 2012.
- [12] Y. Ushiro, Y. Hijikata, M. Shimizu, Y. In'nami, K. Kasahara, A. Shimoda, H. Mizoshita, & R. Sato, Reliability and Validity of Translation Test as a Measurement of Reading Comprehension, *Annual Review of English Language Education in Japan*, vol.16, pp.71-80, Gunma May 2005.
- [13] G. Buck, Translation as a language testing procedure: does it work?, *Language Testing*, vol.9, no.2, pp.123-148, December 1992.
- [14] 青木優子, “英文和訳テストの妥当性調査,” 関東甲信越英語教育学会研究紀要, vol.14, pp.35-42, 東京, July 2000.
- [15] 麻生雄治, “英文読解力評価のための英文和訳テストの信頼性と妥当性,” EIKEN BULLETIN, vol.24, pp.189-197, November 2012.
- [16] 黒川智史, “一般化可能性理論を用いた英文和訳問題の採点の信頼性に関する検証,” 言語学習と教育言語学 2018 年度版, pp.1-9, March, 2019.
- [17] 金谷憲, 英語リーディング論, 河原社, 東京, September 1995.
- [18] 江利川春雄, 受験英語と日本人: 入試問題と参考書からみる英語学習史, 研究社, 東京, March 2011.
- [19] 鳥飼玖美子, 大津由紀雄, 江利川春雄, 斎藤兆史, 林徹, 英語だけの外国語教育は失敗する: 複言語主義のすすめ. ひつじ書房, May 2017.
- [20] 松崎正治, “<<メタ言語能力>>を育てる教材の開

発(<特集>音声言語の教育をどうするか”, 国語科教育, vol.38, pp.27-34, March 1991.

- [21] 長谷川佑介. “質的方法” 渡部良典・小泉利恵・飯村英樹・高波幸代 (編) 日本言語テスト学会誌 20周年記念特別号 (pp. 244-247) 日本言語テスト学会.2016.
- [22] O’Sullivan, B., & Weir, C. J., Test development and validation. In B. O’Sullivan (Ed.), *Language testing: Theories and practices* (pp. 13-32). Palgrave Macmillan. New York, 2011.