

第37回 研究助成

A 研究部門・報告Ⅱ・英語能力テストに関する研究

高等学校の観点別学習状況の評価における 教室内英語テストの妥当性検証と改善

研究者: 大鋸 雄介 神奈川県／神奈川県立七里ガ浜高等学校 教諭
東京都／東京学芸大学連合大学院 在籍

《研究助言者: 小泉 利恵》

概要

本研究では, Kane(1992)が提唱し, Chapelle et al. (2008), Chapelle and Voss(2021)が発展させた論証に基づく妥当性検証アプローチ(argument-based approach to validation)を用いて, 教室内英語テストの妥当性を, 七つの推論のうちの領域定義推論(Domain Definition Inference)及び波及効果推論(Consequence Implication Inference)の観点から検証することを目的とした。具体的には, 筆者の勤務校において, 英語科教師への質問紙調査, 生徒への質問紙調査, 言語テスト分野の専門家による文書レビューを実施した。その結果, 教師は授業や教科書との整合性を重視してテストを設計している一方で, 単元計画やCan-doリストとの接続には課題が見られること, 構成概念への理解にはらつきがあることが明らかとなった。また, 生徒質問紙調査により, 定期試験がリーディングや語彙・文法の領域に対する学習動機づけに一定の影響を及ぼしていることが示された一方で, リスニングについては波及効果が限定的であり, 配点や出題の有無が学習行動に影響している可能性が示唆された。これらの知見は, 論証に基づく妥当性検証における領域定義推論及び波及効果推論の前提を支持する証拠の一部を提示しつつも, 教室内テストに内在する課題を提示するものであり, 高等学校における教室内英語テストの妥当性論証の構築に寄与するものである。一方で, その他の推論に関する証拠は本研究の範囲外であり, それらを今後の研究で補完することで, より包括的な妥当性検証が達成されると考えられる。

1 はじめに

平成29年・30年告示学習指導要領では, 各教科等の目標及び内容を, 「知識及び技能」「思考力, 判断力, 表現力等」「学びに向かう力, 人間性等」の三つの柱で再整理した(文部科学省, 2017a, 2017b, 2018)。観点別学習状況の評価(以下, 観点別評価)とは, 目標に準拠した評価を実施するために, 「知識・技能」「思考・判断・表現」「主体的に学習に取り組む態度」の三つの観点で学習状況を分析的に捉えることであり, これらを総括して評定を定めることとなった(国立教育政策研究所, 2020a, 2020b, 2021)。外国語科においては, 小・中・高等学校で一貫した目標を実現するために, ヨーロッパ言語共通参照枠(CEFR)を参考に, 五つの領域で外国語科の各科目の目標を設定し, その目標を実現するために行う言語活動の内容もCEFRを参照するという方針が示された(文部科学省, 2017b, 2018)。この方針に基づいて, 外国語科では五つの領域それぞれにおいて設定された目標の実現の状況を三つの観点ごとに適切に評価するために, 五つの領域についてそれぞれ三つの観点で評価規準を設定することになった(国立教育政策研究所, 2020a,

2020b, 2021)。さらに、約10年ごとに行われる学習指導要領の改訂に向けた中央教育審議会特別部会の議論では、現行の3観点のうち「主体的に学習に取り組む態度」について、評価の妥当性や現場負担の観点から評定に直接反映させない方向が検討されている一方、「知識・技能」及び「思考・判断・表現」は次期学習指導要領においても引き続き評価観点として用いられる見込みである(中央教育審議会, 2025)。

しかし、外国語科における観点別評価には、以前からさまざまな課題があると指摘してきた。村越・江原(2015)は、ペーパーテスト作成時における各設問の観点分類に問題があることや、教師間で分類に関する認識にずれがあることを明らかにしている。この調査は旧学習指導要領に基づく評価を扱ったものであるが、平成29年・30年告示学習指導要領下においても課題が見られる。今井(2023)や馬場(2023)は、国立教育政策研究所の『「指導と評価の一体化」のための学習評価に関する参考資料』において、「知識・技能」と「思考・判断・表現」の評価規準作成時の留意点が類似している点を挙げ、「思考・判断・表現」の観点の評価の難しさを指摘している。

その結果、教師は何に基づいて評価をすればよいのか分からず、同様に生徒もどのように学習すればよい評価を得られるのかが分からず、その結果、生徒の学習に負の波及効果を及ぼすことが懸念される。そのような中で、ペーパーテストのみならず、幅広い方法で生徒の学習状況を上記の観点別に評価するよう求められる学校現場は非常に混乱しているように感じられる。勤務校でもこの方針に則り、定期試験及びパフォーマンステストを通じた評価を行ってきたが、特に受容技能を中心とする定期試験において、どの構成概念を測定しているのかが曖昧であるという課題がある。

しかし、外国語科の観点別評価が、生徒の外国語学習の到達度を測る評価として妥当性が高いかを実証的に検証した研究は調べた限り見つからない。高等学校の評価は大学入試にも幅広く活用され、生徒の人生を左右する重要な評価であるといえる。そのため、外国語科における観点別評価の妥当性を検証することは喫緊の課題である。本研究は、言語テストにおける妥当性を検証するための「論証に基づく妥当性検証アプローチ(argument-based approach to validation)」を用いて、高等学校における教室内英語テストの妥当性を検証することを目的とする。

2 先行研究

妥当性は構成概念だけでなく、外部基準との関連性や波及効果、信頼性など、さまざまな要素で構成される概念である(Messick, 1996)。そのため、言語テストの妥当性を高めるには、これら複数の要素をどの程度満たしているかを多角的な手法を用いて検証する必要がある。このような妥当性の検証手法として、「論証に基づくアプローチ(argument-based approach)」が提唱された(Kane, 1992)。このアプローチでは、テストスコアに基づいて妥当性を主張するために必要な証拠を体系的に提示することで、妥当性の各要素を支持することが目指される(小泉, 2018)。特に、Kane(1992)のアプローチを修正・発展したChapelle et al.(2008)の「論証に基づく妥当性検証の枠組み(validation framework of argument-based approach)」は妥当性を論じた多くの文献で引用され、言語評価の妥当性検証にも使われている。

この枠組みにおいては、以下の六つの推論を通じてテストの妥当性を検証する:(1)領域定義推論、(2)得点化推論、(3)一般化推論、(4)説明推論、(5)外挿推論、(6)利用推論。Chapelle and Voss(2021)では上記六つの推論にテストの波及効果についての要素を含む(7)波及効果推論を加えて七つの推論とし、それらを裏付ける前提や証拠を示しながら推論の正当性を立証していく(図1)。

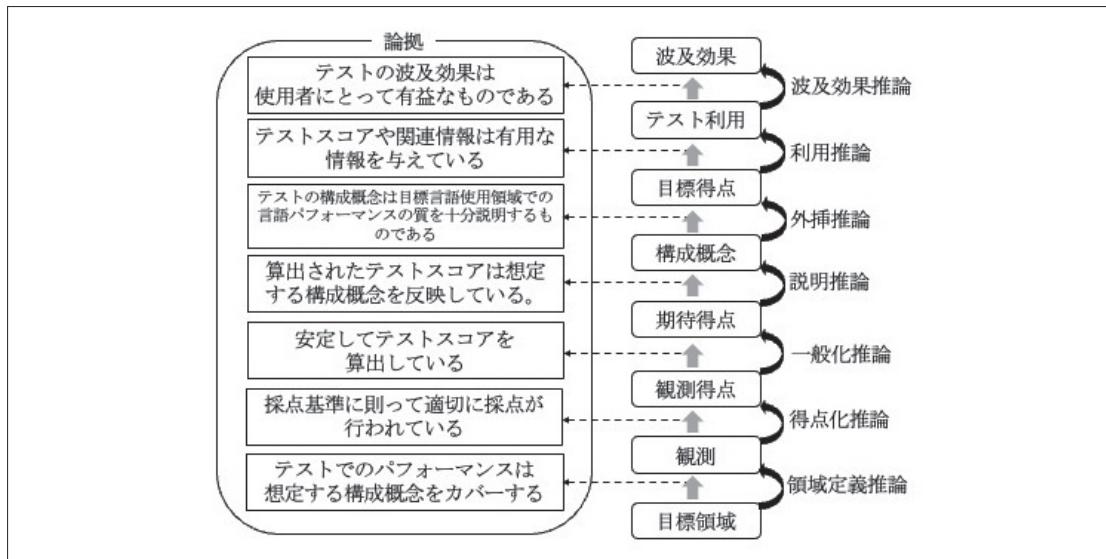


図1: Chapelle and Voss (2021)の論証に基づく妥当性検証の枠組み

注:小泉(2018), 久保(2022)に基づいて著者が作成

論証に基づく妥当性検証では、まず「学習到達度の測定や指導改善への活用といった目的で実施される高等学校の定期試験が、意図した能力を適切に測定し、その使用が有益である」というような主張を立てた上で、領域定義から波及効果に至る各推論に対応する前提や反論、必要な証拠を含んだ論証を構築し、証拠の収集と論証の見直しを繰り返しながら妥当性を検証していく(小泉, 2025)。Chapelle and Voss (2021)による論証に基づく妥当性検証の枠組みでは、七つの推論を通じて言語テストの妥当性を多角的に検討する。

(1) 領域定義推論では、テストで測定すべき言語使用の文脈(目標言語使用領域: Target Language Use [TLU] domain)が適切に定められているか、またテストがそのTLUを適切に反映しているかを検証する。TLU領域とは、テストの場面とは異なる現実の場面において、受験者が実際に英語を使って何らかのやりとりやタスクを行うことが求められるような、具体的な状況や環境のことを指す(Bachman & Palmer, 2010)。例えば、高等学校の教室内テストが学習指導要領に示された言語活動や学習目標に基づいて設計されているかどうかが問われる。

(2) 得点化推論では、テストによって観測された受験者のパフォーマンスが適切に採点されているかどうかが論点となる。例えば、多肢選択式のテストに対しては、項目困難度や弁別力の観点から、スピーキングやライティングなどのパフォーマンステストに対しては、ルーブリックの明確性や採点の一貫性などの観点から検証する。

(3) 一般化推論では、観測得点(採点されたテストスコア)が異なる実施日や異なる採点者でも安定して得られるかを分析し、信頼性について検証する。

(4) 説明推論では、期待得点(目標領域で得られた信頼性の高いテストスコア)が、想定する構成概念をどの程度反映しているかを検証する。

(5) 外挿推論では、テスト得点がTLU領域における実際のパフォーマンスをどの程度予測・反映するかを検証する。

(6) 利用推論では、テストスコアや関連情報が意思決定のために有益に活用されているかを検証する。大規模テストにおいては、受験者を適切に分類することやカットスコアの正当化が焦点となるが、本研究の対象である教室内テストでは、教師がスコアレポートを用いて学習指導の改善や生徒へのフィードバックを行うといった教育現場での活用が主な検討対象となる。

(7) 波及効果推論では、テストが学習や指導にどのような影響を及ぼしているか、例えば「定期試験が文法暗記偏重の学習を促してしまっていないか」や「スピーチングテストが生徒の発話量を増やすきっかけとなっているか」といった教育的影響を考察する。

これらの推論を統合的に検討することで、テストの設計・実施・運用に関する妥当性の全体像を論証的に構築することが可能となる。以下の表1は、上述した妥当性論証における主要な推論・前提・証拠の例をまとめたものである。

■表1: 言語テストの妥当性論証における主要な推論・前提・証拠の例 (Dai et al., 2024に基づく)

推論	前提の例	証拠の例
領域定義推論 主張: テストタスクの選定・設計・実施が、関連する目標領域(TLU領域)を考慮している	・テストタスクが目標領域を反映している ・テストタスクが目標領域を十分に捉えている	・領域分析、領域関係者へのインタビュー及び質問紙調査
得点化推論 主張: 受験者のパフォーマンスが、意図された特性をもつ得点として適切に評価されている	・実施手順が適切である ・項目の統計的特性は適切である ・テストタスクは受験者を適切なレベルに分類できる ・テストタスクは受験者が能力を発揮する十分な機会を提供している ・採点には信頼性がある	・テスト条件の検討、受験者へのインタビュー ・統計分析(項目分析、Rasch分析) ・関係者へのインタビュー ・応答特性を含むテスト問題のレビュー
一般化推論 主張: 観測された得点は、他の類似課題や形式、採点者にわたって再現可能なものである	・タスク・項目・採点者数は信頼性の高い得点を得るのに十分である ・テスト得点の等化処理が適切である ・構成概念とは無関係な分散が、実施条件や受験者の母語、性別といった要因によって生じない ・採点はテストの実施ごとに一貫している	・一般化可能性理論研究 ・テスト開発文書のレビュー ・DIF(特異項目機能)分析、統計的モニタリング
説明推論 主張: 期待得点は、テストが測定しようとする構成概念に起因している	・必要とされる知識・プロセス・方略が理論と一致している ・テストスコアの内部構造が理論と整合的である ・タスクにおけるパフォーマンスは、関連する他のタスクにおけるパフォーマンスと関連している	・受験者の言語使用の発話プロトコル ・因子分析 ・相関分析
外挿推論 主張: テストで測定された構成概念は、現実の目標領域におけるパフォーマンスを説明できる	・テストでのパフォーマンスが TLU のパフォーマンスと関連している ・受験者がテストタスクで行うやりとりが TLU のパフォーマンスと似ている	・領域関係者へのインタビュー ・テストと現実場面の発話分析
利用推論 主張: テストスコアや関連情報は適切な意思決定のために有益な情報を与えている	・評価は受験者を意思決定に適切なレベルに分類できる ・カットスコアが正当化できる	・統計分析 ・基準設定方法のレビュー ・利用者へのインタビュー
波及効果推論 主張: テストの波及効果は使用者にとって有益なものである	・合格者は当該領域で十分な能力を有する ・テスト形式が当該領域の言語使用に合致している ・テスト得点やフィードバックが学習や指導に良い影響を与える	・受験者、雇用者、教師、利害関係者へのインタビュー

Dursun and Li (2021) は、2000年から2018年にかけて発表された言語テスト分野の論証に基づく妥当性検証研究を対象としたレビュー研究であり、その使用傾向や地域的分布を明らかにしている。彼らは論証に基づく妥当性検証の枠組みが近年増加傾向にあり、これが主流な枠組みであること、また、対象となった言語テストは大規模テストが最も多く、教室内テストを対象とした研究が最も少ないことを指摘している。そして、同枠組みにおいて証拠が示された推論のうち、領域定義推論が最も少なく、次いで利用推論及び波及効果推論が少ないことも明らかとなった。また、論証に基づく妥当性検証の枠組みを用いた研究の半数は北米(特に米国)で実施されており、70の対象研究のうち56(80%)が北米の研究者によって発表されている。一方、アジア地域の研究者によるものはわずか6件であり、その中に日本の事例も含まれているものの、全体に占める割合はごく少数である。このことは、論証に基づく妥当性検証が国際的に広がりを見せているものの、日本における実証的な適用例は依然として限られていることを示しており、今後の国内研究においてこの枠組みを積極的に活用していく必要性を示唆している。

日本の言語テストを対象に、論証に基づく妥当性検証の枠組みが用いられた数少ない例として以下の二つの研究が挙げられる。Koizumi et al. (2011) は、日本人英語学習者を対象とした文法診断テスト (EDiT Grammar) に対して、論証に基づく枠組みを用いて妥当性検証を行い、Chapelle et al. (2008) の六つの推論に沿って解釈的主張を構築し、ラッシュ分析や発話プロトコル分析を通じて、得点化推論及び説明推論を支持する証拠を示している。Kumazawa et al. (2016) は日本の大学生向けに開発された英語プレースメントテストである VELC Test® を対象に、Kane (2006) に基づいて妥当性検証を行った。Kane の提唱する論証に基づく妥当性検証では、採点 (scoring)、一般化可能性 (generalizability)、外挿 (extrapolation)、決定 (decision) という四つの推論に焦点が当てられ、VELC Test® に対するそれぞれの推論に対応する証拠が統計的に収集された。例えば、ラッシュモデルに基づく項目分析や一般化可能性理論 (G理論) を用いた信頼性の検証、受験者の Can-do に基づくスコアマッピングなどを通じて、テストスコアが構成概念を妥当に反映していることが示された。また、スコアレポートが受験者や教師にとって学習支援のツールとして有用であることも検討されており、Chapelle and Voss (2021) における利用推論や波及効果推論につながる視点も含まれている。このように論証に基づく妥当性検証の枠組みを用いた研究において日本の言語テストを対象とした研究は少なく、特に教室内英語テストの妥当性を検証した研究は見当たらない。よって本研究では、日本の教室内英語テストの妥当性を論証に基づく枠組みを用いて検証し、その中でも領域定義推論及び波及効果推論に係る前提についての証拠を提示し、より良い教室内英語テスト開発のために寄与することを目的とする。

領域定義推論は、上述したようにテストで観察されるパフォーマンスが、想定する構成概念を十分にカバーしていることを前提とする推論である。その検証は主に二つの視点から行われる。第一に、評価が行われる文脈における TLU 領域で測定すべき能力が明確に定義されているか、第二に、設定されたテストタスクが、その TLU 領域における能力をどの程度反映しているかである。例えば、高等学校の教室内英語テストにおいては、学習指導要領に示された言語活動や学習目標に基づき、実際の言語使用状況を反映したテストタスクが設計されているかどうかが検討対象となる。領域定義推論で用いられる研究手法として主にカリキュラム分析、授業観察、ニーズ分析、コーパス開発と分析が挙げられる。例えば、Park (2021) は、軍用航空管制官を対象とした航空英語評価の開発において、文書分析・専門家へのインタビュー・ニーズ分析を通じて TLU 領域を明確にし、その TLU を反映した評価タスクを仮想環境で設計・実施することで、領域定義推論に基づく妥当性の検証を行っている。

波及効果推論は、テストが指導や学習に与える望ましい、あるいは望ましくない影響を明らかにする推論である。波及効果推論で用いられる研究手法としては、テスト使用者へのインタビュー、分類精度と一貫性の分析、ラッシュ分離指数、縦断的データ分析が挙げられる (Dursun & Li, 2021)。Allen et al. (2024) は論証に基づく妥当性検証の枠組みは用いていないものの、生徒質問紙調査・授業観察・インタビュー等の混合的手法を通じて日本の高等学校における英語4技能外部試験の波及効果を調査し、明確な

指導・学習への影響は確認されなかったものの、当該試験がカリキュラムや授業と整合していること、そして大学入試による負の波及効果が依然として存在することを明らかにした。

本研究の妥当性論証においては、領域定義推論の前提を「高等学校における教室内英語テストにおいて、TLU領域は明確に定義されており、テストのタスクがTLU領域の重要なスキルを適切にカバーしている」とし、波及効果推論の前提を「高等学校における教室内英語テストは生徒に有益な波及効果を与えている」と設定した。これらの前提を検証するために、以下のリサーチクエスチョンを設定した。

RQ1 高等学校における教室内英語テストにおいて、目標言語使用領域は明確に定義されているか。

RQ2 高等学校における教室内英語テストのタスクが、目標言語使用領域の重要なスキルを適切にカバーしているか。

RQ3 高等学校における教室内英語テストは、生徒にどのような波及効果を与えているか。

3 方法

3.1 対象となる教室内英語テストの文脈

本研究の対象校は、神奈川県内に所在する公立高等学校であり、全日制・共学の普通科高校である。生徒数は約1,100名で、英語科には11名の専任教師が勤務している。定期試験は年間5回実施されており、パフォーマンステストは学年によって異なるものの年間1~2回実施されている。定期試験では主にリーディング及びリスニングの受容技能を、パフォーマンステストではスピーキング及びライティングの産出技能を測定している。定期試験の作成は、原則として学年担当教師間で分担されており、事前に出題範囲や試験形式に関する調整が行われる。また、パフォーマンステストについても、共通の課題とループリックを用いながら、各教師が担当クラスでの実施・評価を行っている。いずれのテストも、実施の1~2週間前を目安に、生徒に対して出題形式や評価基準等の概要を事前に周知することが慣例となっている。本研究では、高等学校における教室内英語テストの妥当性検証の一部として、領域定義推論及び波及効果推論における前提を支持する証拠を収集することを目的とし、教師質問紙調査、専門家による文書レビュー、生徒質問紙調査の三つの調査を行った。各調査の概要は表2の通りである。

■表2: 本研究で実施した調査の概要

調査名	対象	調査内容	実施時期	備考
教師質問紙調査	筆者の勤務校の英語科教師10名	教室内英語テスト全般についての認識・実践を調査	2025年3月	特定のテストに限定せず、全般的な傾向を把握
専門家による文書レビュー	大学教授(専門家)	・3年生対象の定期試験(1回) ・3年生対象のライティングテスト(1回) ・2年生対象のスピーキングテスト(1回)	2025年3月	各テストに関する関連文書(ループリック、単元計画等)も併せて送付
生徒質問紙調査	高校1~3年生	各学年の1学期中間試験(定期試験)についての受け止めを調査	2025年5月	英語コミュニケーションの試験実施後に調査

3.2 領域定義推論に関する調査(RQ1及びRQ2)

3.2.1 教師質問紙調査

本調査には、筆者の勤務校である神奈川県内の公立高等学校に勤務する英語科教師10名が参加した(筆者は除く)。調査対象者の英語指導歴は、0~5年が3名、6~10年が4名、11~20年が2名、21年以上が1名であり、若手からベテランまで幅広い経験年数の教師が含まれていた。質問紙調査は2025年3月にGoogle Formsを用いてオンラインで実施され、事前に調査の目的、回答内容の匿名性、研究への任意参加である旨を説明し、文書による同意を得た(資料1参照)。質問項目は Jun(2014)を参考に筆者が作成した(資料2参照)。質問項目は英語指導歴と担当学年を問う項目から始まり、Q9を除くQ3~Q19及びQ22、23は6段階リッカート尺度(6:「とてもよく当てはまる」、1:「全く当てはまらない」)を用いた項目であり、中間の5、4、3、2についてはラベルを付与せず数値のみを提示した。Q20及びQ21は、定期試験やパフォーマンステストを作成する際に参考する文書について尋ねる複数回答式の項目であった。Q24~Q27は定期試験やパフォーマンステストを作成・実施・採点する際の困難点や工夫している点について自由記述回答を求める項目である。Q9は「ご自身の勤務校ではCan-doリストを設定していますか?」という質問に対して「設定している」「設定していない」「分からない」の三つから選択する項目であった。自由記述を含む設問においては、回答者個人が特定されないように十分な配慮を行った。なお、本調査は特定の定期試験やパフォーマンステストに対して行われたものではなく、昨年度各教師が実施した定期試験やパフォーマンステストを作成・実施・採点する際の意識を全般的に調査したものであることや、本研究における結果の解釈及び一般化に際しては、同一校内の教師のみを対象とした調査であることに留意する必要がある。

3.2.2 専門家によるフィードバック

2025年3月、言語テストを専門とする大学教授に対し、勤務校で実施された1回分の定期試験、スピーキングを測るパフォーマンステスト及びライティングを測るパフォーマンステストをそれぞれ1回分、並びに関連文書(単元の指導と評価の計画、ルーブリックやCan-doリスト、教材のコピー等)を添付し、評価の妥当性に関するフィードバックをメールにて依頼した(送付した文書の一部は資料3、4参照)。送付した文書の一覧は以下の通りである(表3)。フィードバックを行う際には、以下の三つの観点でフィードバックをいただくよう依頼した:①テスト問題が、TLU領域を適切に反映しているかについて、②評価基準(ルーブリック)の適切さと明確さについて、③試験全体の妥当性や改善すべき点について。翌月の2025年4月に、当該教授より文書によるフィードバックを受領した。フィードバックの対象となった定期試験の概要は表4の通りである。

表3: 専門家に送付した文書一覧

分類	文書名及び内容の詳細	テスト実施時期	対象学年	作成者
リスニング・リーディングテスト関連	英語コミュニケーションⅢ 1学期期末試験	2024年7月	高校3年生	筆者
スピーキングテスト関連	英語コミュニケーションⅡ Lesson 4 ゴールタスク(スピーキングタスク)	2024年7月	高校2年生	勤務校英語科教師 (経験年数6年~10年)
ライティングテスト関連	論理・表現Ⅲ Lesson 9 ゴールタスク(ライティングタスク)	2024年9月	高校3年生	筆者
その他参考資料	・勤務校のCan-doリスト ・各科目の「単元の指導と評価の計画」 ・該当単元の教科書及びワークシートのPDF	—	—	—(該当科目全体の資料として添付)

■表4: 対象となった定期試験の概要

大問	小問	問題概要	形式	項目数	配点	評価の観点
大問1	A	副教材からのリスニング問題(ディクテーション)	記述式	8	8	知識・技能
	B	副教材からのリスニング問題(内容理解)	多枝選択式	3	6	思考・判断・表現
	C	副教材からのリスニング問題(内容理解)	多枝選択式	3	6	思考・判断・表現
大問2		副教材からの文法・語法問題	多枝選択式	20	20	知識・技能
大問3		副教材からの語彙問題	多枝選択式	20	20	知識・技能
大問4	A	教科書からのリーディング問題 (教科書をパラフレーズした問題)	多枝選択式	9	10	思考・判断・表現
	B	教科書からのリーディング問題 (教科書をパラフレーズした問題)	多枝選択式	5	10	思考・判断・表現
大問5	A	大学入試からのリーディング問題	多枝選択式	5	10	思考・判断・表現
	B	大学入試からのリーディング問題	多枝選択式	5	10	思考・判断・表現

今回フィードバックの対象としたスピーキングテストは、高等学校2年生の英語コミュニケーションⅡの授業において実施された。当該単元は動物の生態をテーマとし、スピーキングテストのタスクは、各生徒が自ら選択した動物について調査し、まとめた内容を発表する発表形式のタスクであった。タスクは1種類であり、場面設定は、近隣の中学校にゲストスピーカーとして招かれ、理科の授業で動物に関するスピーチを行うというものであった。発表は授業内で行われ、一人ずつポスターを用いてスピーチを行い、その場で担当教師が事前に提示されたルーブリックに基づいて採点した。教科書の単元は四つのパートから構成されており、動物に関する説明文の構成、関連語彙・文法はこのタスクに関連しており、単元全体が発表タスクの準備プロセスとして構成されていた。図2に対象となったスピーキングテストの指示文を示す。

The Goal Task

「Animal 100 選 大図鑑！これがガ浜の生徒の総力！」

In the end of this lesson, you are going to go to KOSHIGOE junior high school as a guest speaker. You will introduce an animal in science class. You are going to choose one animal and become a master of it.

■図2: 対象となったスピーキングテストの指示文

今回フィードバックの対象としたライティングテストは、高等学校3年生の論理・表現Ⅲの授業において実施されたものである。タスクは一つで、ソーシャルメディアの利点と欠点に関する英文を読んだ上で、“Do the benefits of social media outweigh the drawbacks? Agree or disagree?”というテーマについて、80~100語程度のエッセイを作成するものであった。生徒は授業時間内に20分間でライティングを行い、終了後に提出した。採点は事前に生徒へ提示されたルーブリックに基づいて授業時間外に教師が行った。授業内容との関連として、当該単元では意見文の構成や主張・理由の展開方法について事前に指導を行っており、本タスクはそれらの学習内容の到達度を評価する目的として位置づけられていた。図3は対象となったライティングテストの指示文である。

STEP 3 Read **Input A and B** and write an essay of approximately 80-100 words about “**Do the**

benefits of social media outweigh the drawbacks? Agree or disagree?” in which you clearly state your position on the issue.

■図3: 対象となったライティングテストの指示文

3.3 波及効果推論に関する調査(RQ3)

波及効果推論の前提に関する証拠収集の一環として、2025年5月に実施した1学期中間試験の直後に、生徒を対象とした質問紙調査を実施した。対象は筆者の勤務校に在籍する高等学校1~3年生全体($N = 1,103$)であり、調査はGoogle Formsを用いてオンラインで実施され、事前に調査の目的、回答内容の匿名性、研究への任意参加である旨を説明し、文書による同意を得た(資料5参照)。質問項目は、定期試験が生徒の学習行動や学習意欲に与えた影響を把握することを目的とし、Allen et al.(2024)を参考にして筆者が作成した。質問項目は、主に6段階リッカート尺度(6:「とてもよく当てはまる」, 1:「全く当てはまらない」)を用い、中間の5, 4, 3, 2についてはラベルを付けず数値のみを提示した。また、選択式及び自由記述式の項目も含まれていた(資料6参照)。任意回答とした結果、924名から有効回答が得られた(回収率83.8%)。なお、調査実施にあたり、回答結果は評価や成績には一切影響を及ぼさないことを生徒に周知し、匿名性とプライバシーの保護に十分配慮した。また、本研究における結果の解釈及び一般化に際しては、同一校内の生徒のみを対象とした調査であることに留意する必要がある。

4 結果

4.1 教師質問紙調査による領域定義推論の検証

質問紙調査の質問項目、及び自由記述回答の項目を除く結果を資料2に示す。資料2より、テストで測る英語力は学習指導要領や単元の指導と評価の計画等に明示されているかという質問に対して、Q3(定期試験)及びQ4(パフォーマンステスト)の平均値はそれぞれ4.3及び4.1となり、6段階尺度(6:「とてもよく当てはまる」, 1:「全く当てはまらない」)の中ではやや肯定的寄りの傾向が見られた。なお、本調査では2~5の中間値にはラベルを付与していないため、これらの数値は端点との相対的位置に基づく傾向として解釈している。実際に明示されているかどうかは、別途文書分析を通じて検証する必要があるが、概ね参加教師はTLU領域で測定すべき能力が明確に定義されていると認識していることが分かった。英語指導経験年数別に見ると、経験年数が長いほど明示されているとする傾向があった。ただし、標準偏差はいずれも約1.5と大きく、教師間で認識にばらつきがあることが示唆された。

観点別評価の評価観点の違いについての理解は、技能ごとに平均値に若干の差が見られた。Q5(読むこと: $M = 3.8$)、Q7(書くこと: $M = 3.7$)に対し、Q6(聞くこと: $M = 3.3$)及びQ8(話すこと: $M = 3.1$)はやや低い値を示した。特に「話すこと」では標準偏差が1.1と大きく、教師間で理解のばらつきが相対的に大きいことがうかがえた。

さらに、詳しく回答結果を見てみると、Q5からQ8について6段階の尺度の最高値である「6」及び最低値「1」を選択した参加者はいなかった。また、「読むこと」「書くこと」「聞くこと」は6段階中の「4」を選択した参加者が最も多かったのに対し、「話すこと」については6段階中「2」を選択した参加者が最も多かった。本調査では実際にどのように評価観点を理解しているかを調査していないため、正しく理解しているかどうかは判断できない点について留意すべきではあるが、評価の根拠となる目標や観点が曖昧であり、理解が教師によって異なる可能性が示唆された。

なお、定期試験及びパフォーマンステストの設計において、教科書との整合性については高い認識が共有されていた(Q15 = 4.6, Q19 = 4.7)。また、Q22及びQ23ではいずれも平均値が5.0であり、授業内容とテストとの整合性についても高い認識が得られた。これにより、参加者は教室内テストを授業内容や教科書を反映させるものとして設計していることが示唆された。一方で、「学習指導要領」、「Can-do リスト」や「単元の指導と評価の計画」との整合性については、Q12($M = 3.7$), Q13($M = 3.5$), Q14($M = 3.6$), Q16($M = 4.3$), Q17($M = 3.7$), Q18($M = 3.9$)に見られるように、パフォーマンステストと学習指導要領の整合性への意識は一定程度見られるものの、教科書や授業内容との整合性と比べて相対的にやや意識が低い傾向も確認された。

さらに、Q20「定期試験の設計プロセスにおいて、どのような資料を参照していますか？」及びQ21「パフォーマンステストの設計プロセスにおいて、どのような資料を参照していますか？」の回答結果を表5に示す。いずれのテストにおいても「教科書」(定期試験:90%, パフォーマンステスト:90%)が最も多く参照されており、次いで「過去の定期試験」(同90%, 70%)や「英検などの外部試験」(同80%, 70%)といった既存の問題例を参考にする傾向が強く見られた。一方で、「学習指導要領」(同10%, 0%)や「単元の指導と評価の計画」(同30%, 30%), 「Can-do リスト」(同10%, 30%)など、制度的な評価枠組みの参照は相対的に少なかった。これらの結果は、テスト設計が教科書や過去問題、外部試験といった資料に強く依拠している一方で、学習指導要領やCan-do リストなどの制度的文書との接続が十分でない可能性を示唆している。

また、Q9「ご自身の勤務校ではCan-do リストを設定していますか？」の質問に対して、「設定している」が6名、「設定していない」が2名、「分からない」が2名であった。Can-do リストについては各校で作成し、活用することが推奨されており、勤務校でも独自のCan-do リストは作成してあったが、その認知度には個人差があることが分かった。このことから、一部の教師にとってはCan-do リストが形骸化している可能性が示唆された。

■表5: Q20及びQ21への回答結果(複数回答可)

参考とする資料	Q20:定期試験	Q21:PT
	回答件数 (%)	回答件数 (%)
学習指導要領	1 (10%)	0 (0%)
単元の指導と評価の計画	3 (30%)	3 (30%)
Can-do リスト	1 (10%)	3 (30%)
教科書	9 (90%)	9 (90%)
過去の定期試験	9 (90%)	7 (70%)
英検などの外部試験	8 (80%)	7 (70%)
大学入試過去問	1 (10%)	—
その他の問題集	1 (10%)	—
参考となる写真、動画	—	1 (10%)

注. PT = パフォーマンステスト

自由記述回答内容については質的に分析し、意味のまとまりごとにコード化した上で、類似の内容を基にカテゴリに整理した。まず、Q24の「定期試験を作成する際に難しいと感じる点」に対する回答について、教師が定期試験を作成する際に直面している主な課題として、観点別評価における「思考・判断・表現」の観点に対応した問題作成の難しさが指摘された。また、単元の指導と評価の計画に即した問題を作成する時間的余裕の不足も挙げられた。このことは、上記の単元の指導と評価の計画との整合性の低さを示した結果の理由を説明する回答であった。さらに、テストの平均点の調整の難しさ、複数教師による担当のため試験方針にばらつきが生じやすいこと、マーク式の導入による記述問題の制限、教科書の単なる暗記に偏らない試験作りに苦慮している実態も明らかとなった。これらは、村越・江原(2015)の教師質問調査の結果と類似している点が多く、旧学習指導要領下における定期試験作成に関する困難点は、平成29年・30年告示学習指導要領下でも解消されていない可能性が示唆されている。

次に、Q25の「パフォーマンステストを作成・実施する際に難しいと感じる点」に対して、大きく分けて時間的制約と評価の信頼性に関する回答を得た。時間的制約に関しては、40名規模のクラスで実施する場合、授業回数を複数回費やす必要があり、他の内容を扱う時間が確保しづらいという意見が複数見られた。本調査は特定のパフォーマンステストについて調査しておらず、実際の手順は不明であるが、筆者の勤務校では、スピーキングのパフォーマンステストを行う際、しばしば生徒一人ずつを廊下に呼び、パフォーマンスを行わせ、授業者が評価・フィードバックを行うことが慣例となっている。そのため、50分で1時間分の授業とすると、一度のパフォーマンステストで2~3時間費やすことになる。また、トピックや形式の検討に時間がかかることや、記述式やスピーキング活動において準備・実施・採点にかかる負担の大き

さも指摘された。小泉(2022)は、授業中に発話を録音して提出し、後で採点するなど1回の授業内で実施が可能であるスピーキングテストの実施形態を紹介しており、妥当性、信頼性、実現可能性の三つの要素から学校環境に応じた実施形態を検討することを推奨している。したがって、時間的制約の課題を軽減するためには、こうした代替的な実施形態の導入を検討することが有効であると考えられる。評価の信頼性に関しては、共通のルーブリックを用いたとしても担当者によって評価にばらつきが生じることや、フィードバックの在り方、評価者の主觀を避ける難しさなど、評価基準の共通理解や指導と評価の一体化に関する課題が多く挙げられた。その他、生徒によってモチベーションが異なる中での到達目標の設定や、テストのモデルを事前に示すことにより回答が似たものになってしまう点も挙げられた。

Q26の「定期試験やパフォーマンステストを作成・実施する際に工夫している点」に対しての回答は、大きく授業との整合性と学習者支援・動機づけの二つに分類できる。まず、授業との整合性に関しては、多くの教師が、教科書や副教材、授業で扱った内容に即した出題を意識しており、生徒が授業で得た知識や技能を試験で活用できるように設計している。例えば、語彙や文法をそのまま出題するのではなく、言い換えや応用的な文脈で問うことで、表層的な暗記では対応できないような問題に工夫している。また、試験範囲や出題形式、評価項目を事前に明示し、学習の見通しを持たせる工夫も見られた。

一方で、学習者支援と動機づけの観点では、大学入試とは直接関係のない活動であっても、英語がコミュニケーションの手段であることを意識させ、学ぶ意味や目的を丁寧に伝えることで、生徒の内発的動機づけを促す工夫が見られた。特に話す活動に対して不安を感じやすい生徒に向けては、安心して取り組めるような説明やモデル提示、練習機会の提供を通して、生徒の不安軽減を図る工夫が共有された。また、大学入学共通テストなどの大学入試に似た形式の問題を導入することで、生徒の関心を引き、学習の意欲につなげている事例もあった。この取り組みは、廣森・小金丸(2024)が指摘するように、生徒が有している学習動機を意識的に捉え、それを授業や評価の設計に反映することでエンゲージメントを促進する有効な事例であると位置づけられる。

最後に、Q27の評価を実施する際の疑問点に関する質問に対しては、A～Cの観点別評価から5段階の評定への変換の妥当性や、主体的学習に取り組む態度の評価方法に対する疑念が見られた。また、ライティング課題における生徒のAI使用やスピーキング評価の基準に関する悩みも複数報告された。こうした声は、学習評価を取り巻く環境の変化やツールの進展に対する対応の難しさを反映しており、今後の制度設計や研修の在り方に関わる重要な示唆といえる。

4.2 専門家フィードバック

4.2.1 目標言語使用領域の適切性について

専門家は、定期試験及びパフォーマンステストがTLU領域を適切に反映しているかを検討した。はじめに、テストにおいて何をTLU領域と定義しているかによって適切かどうかが変わるという指摘があった。勤務校ではテストごとにテスト仕様書(Test Specifications)を作成しているわけではなく、TLU領域がテストごとに明示されているわけではない。テスト仕様書とは、テストの目的、測定したい能力、問題の形式や内容、項目数や配点、難易度や出題範囲などを明確に定めたテストの設計図のような文書であり、評価の信頼性や妥当性の向上に寄与する(Jin, 2021; 小泉ら, 2017)。そのため、領域定義推論の観点から、改善点としてテストの実施の際はテスト仕様書を作成し、TLU領域を明示することが考えられる。

学習指導要領に基づいて「高校生にとっての日常的、社会的な英語」のようにTLU領域を定義したと仮定すれば、筆記の定期試験、パフォーマンステストともねらいとするTLU領域と、テストで測ろうとしている言語使用の領域は概ね合致しているとのフィードバックを受けた。しかし、送付した定期試験の大問5A(大学入試過去問)については、場面や目的の設定が乏しく、実際の言語使用状況との対応が弱いとの指摘があった(図4参照)。この大問5Aにおいては、問題の指示文がRead the passage below and answer the following questions.であり、目的や場面、状況の設定が欠如している。国立教育政策研究所(2021, p. 65)の参考資料が示す「読むこと」の評価事例では、You are asked to read the following

article about the Universal Design Font (UD font) to prepare for a presentation in English class. という状況設定が記載されている。さらに, 大学入学共通テスト英語(リーディング)においては, Your host family in the US is planning to buy a small aquarium. You found this helpful pamphlet. などという状況設定が提示されている(大学入試センター, 2025a, p. 4)。これらを参考に, 今回フィードバックを依頼した大問5の例においては, You are asked to read the following article about Irene Sendler to prepare for a presentation in English class. などのような記述を冒頭に加えることが考えられる。目的や場面, 状況の明示における解釈についてはさまざまな議論があるが(渡邊ら, 2025), テストを実施する際に, TLU領域を明確に提示することだけでなく, その言語使用領域との対応を明示するために, 問題項目に言語使用の目的や場面, 状況を明記することの重要性が示唆された。

5 【応用問題】〈マークシート問題〉思考・判断・表現[20]

[A] Read the passage below and answer the following questions.

- 1 In 1942, as Jews throughout Europe were being rounded up and transported altogether to Nazi *concentration camps to face an unthinkable fate, one woman took courageous action and risked her own life to save thousands of others.
- 2 Her name was Irene Sendler. Though she herself was not Jewish, she feared for the lives of Jews around her, particularly the children. As head of the children's section in the Polish underground movement known as Zegota, she was unable to sit by and not do anything. So, she went into the Warsaw Ghetto and persuaded Jewish parents and grandparents to place their children in her care, saying that they were certain to die in the Ghetto or in the Nazi death camps unless they could be spirited away to safety.
- 3 *Smuggling the children past Nazi guards through a variety of means — hiding them in body bags or under loads of goods — Ms. Sendler took them into the homes of Polish families, where they were adopted and raised with false identities. Ms. Sendler made lists of these children and placed the lists in a jar that she buried in a garden, hoping she could someday dig up the jar, locate the children and inform them of their past.
- 4 From 1942 to 1943, Ms. Sendler managed to smuggle 400 children out of the Warsaw Ghetto before she was captured by the Nazis and severely injuring for her actions. Even under extreme torture, she refused to reveal where the lists of the smuggled Jewish children were hidden. Eventually, a member of the Polish underground bribed a guard to release her, and she entered into hiding. Even then, she continued to work with Zegota to rescue another 2,100 children.
- 5 It's a remarkable story and considering all the remarkable stories from the Holocaust that have surfaced over the past several years, it's hard to believe that this one went largely unnoticed. And it might have stayed that way, were it not for four high school students from Uniontown, Kansas.

(上智大学 改題)

*concentration camps 強制収容所 smuggle ～を密輸する

(1) Choose the appropriate summary for Paragraph 2~5. (2 点×4)

Paragraph 2	61
Paragraph 3	62
Paragraph 4	63
Paragraph 5	64

① Irene Sendler's story remained largely unnoticed until four high school students from Uniontown, Kansas brought it to light.

② She smuggled the children out using various means and placed them with Polish families, recording their names in a jar for future reference.

③ She was captured and tortured by the Nazis, refusing to reveal the hidden lists. A member of the Polish underground bribed a guard to release her, and she continued to rescue another 2,100 children while in hiding.

④ Irene Sendler persuaded Jewish parents to place their children in her care to save them from the Warsaw Ghetto.

(2) Which of the following title would be the most appropriate for this passage? 65 (2 点)

- ① "The Warsaw Ghetto: Life and Struggles During the Holocaust"
② "The Role of Zegota in World War II"
③ "High School Students Discover Historical Secrets"
④ "Irene Sendler: The Heroine Who Saved Thousands of Jewish Children During the Holocaust"

■図4: 送付した定期試験問題の大問5A(上智大学, 2010を参考に改変)

4.2.2 評価基準(ループリック)の適切さと明確さ

4.2.2.1 スピーキング

スピーキングの評価基準(ループリック)については、いくつかの観点から改善の余地が指摘された。フィードバックの対象となったスピーキングテストのループリックを表6に示す。対象となったループリックは言語的側面や表現内容に基づく五つの分析的評価観点「Accuracy」「Delivery」「Eye-contact」「Structure」「Content」に、「Accuracy」「Content」の観点は4レベル(S～C)、「Delivery」「Eye-contact」「Structure」の観点は3レベル(A～C)の設計になっている。

まず、「Accuracy」のS評価にはイントネーションや発音の正確さも含まれているが、これはA～Cで評価している語彙・文法の誤りとは異なる観点であるため、A～Cの記述にもイントネーション等の評価観点を含める必要があるのではないかとの指摘があった。また、各ランクにおける記述文の表現と評価内容の一貫性にも注意が必要であり、例えば「語彙表現や文法の誤り」と「語彙や文法の誤り」など表記の揺れが見られるほか、誤りの頻度に加えて意味の伝わりやすさが評価に含まれている点も整理が求められる。さらに、「Eye-contact」のA評価における「プロフェッショナルに見える」という表現は曖昧であり、スピーキングの場面設定(高校生が中学生に動物について説明する場面)との整合性にも疑問が残るとの指摘があった。この点については、「プロフェッショナルに見える」といった抽象的な表現ではなく、「聞き手と適切にアイコンタクトをとり、自然で自信のある印象を与える」など、観察可能で具体的な行動を基準とした表現に修正することが考えられる。

これらの評価観点の記述文の内容に加えて、ループリック全体として「Accuracy」「Delivery」「Eye-contact」「Structure」「Content」の5観点に問題はないものの、筆記テストと同様に「知識・技能」「思考・判断・表現」といった学習指導要領に基づく評価観点と対応づける必要があるのではないかという指摘もあった。国立教育政策研究所(2021)の参考資料は、「書くこと」の評価において、「知識・技能」を問う問題は主に英語使用の適切さ、「思考・判断・表現」を問う問題は主に表現内容の適切さで評価すると整理しており、「話すこと」の指導においても、「表現内容の適切さ」という内容面と、英語使用の適切さという言語面からの指導を行うとの記述から、同様の解釈が可能であることが考えられる。また、山本ら(2021)は、開発したスピーキングテストの評価基準における「正確さ」は「知識・技能」に、「内容」「応答」「質問」「提案」は「思考・判断・表現」「主体的に学習に取り組む態度」に対応づけた。これらを踏まえると、本調査における上記5観点の分析的な評価観点については、英語使用の適切さに関連する「Accuracy」「Delivery」「Eye-contact」を「知識・技能」へ、表現内容の適切さに関連する「Structure」「Content」を「思考・判断・表現」「主体的に

表6: スピーキングテストのループリック

Accuracy		Delivery	Eye-contact	Structure	Content
S	(Aに加えて) 語や句、文におけるイントネーションなどの発音の特徴をとらえ、正しく発音している。				(Aに加えて) 具体例を効果的に示していて、情報量が多い。
A	・語彙表現や文法の誤りがほとんどない。 ・聞き手に負荷を与えない。	以下の条件を全て満たしている。 ・相手に十分伝わる声で話している。 ・制限時間内で話し続けている。	充分なアイコンタクトがあり、プロフェッショナルに見える。	「生態」→「特徴」→「リスクや危機」→「解決策、メッセージ」という構成で話を伝えている。	・重要な情報を漏れなく伝えられている。 ・話の流れが明確で一貫性がある。
B	・語彙や文法の誤りが時々あるが、意味は伝わる。 ・聞き手に負荷を与えることがある。	上記条件を一つ満たしている。	スピーチの半分以上でアイコンタクトが取れる。	上記の構成に多少のずれがある。	・重要な情報に多少の漏れがある。 ・話の順序に多少の前後がある。
C	Bに満たない。	上記条件を一つも満たしていない。	Bに満たない。	上記の構成にずれが多く、何をまとまりで話しているか伝わらない。	・重要な情報に多くの漏れがある。 ・話の順序に一貫性がない。

「学習に取り組む態度」へ対応づけることが考えられる。その結果、言語的側面や表現内容に基づく分析的評価観点を設定したルーブリックを活用しつつも、ルーブリックの評価観点と観点別評価の評価観点の対応づけを明示することで、学習指導要領の目標に基づいて測定したい能力を測るテストとしての妥当性がより高まることが考えられる。一方で、観点別評価との対応を前提とした「知識・技能」「思考・判断・表現」「主体的に学習に取り組む態度」の3観点をそのまま用いたルーブリック設計に疑義を呈する見解もあり、今後の検討においてこうした批判的見解も考慮する必要がある(駒井, 2024)。

4.2.2.2 ライティング

ライティングテストのルーブリックについては、評価観点の表記に一貫性があり、基準も明確である点は高く評価された。表7に示すとおり、対象となったルーブリックは「内容」「構成」「正確さ」の言語的側面や表現内容に基づく三つの分析的評価観点から成り立っており、記述は簡潔かつ分かりやすいものであると評価された。一方で、スピーキングのルーブリックと同様に、設定された評価観点それ自体には問題はないものの、筆記テストにおける他の技能(リーディングやリスニング)では「知識・技能」「思考・判断・表現」など、学習指導要領に基づく観点別評価が行われていることを踏まえると、ライティングにおいてもそれらの観点と対応づけたルーブリック設計が望ましいのではないかという指摘があった。

例えば、英語使用の適切さに関する「正確さ」の観点は「知識・技能」に対応し、表現内容の適切さに関する「構成」「内容」の観点は「思考・判断・表現」に対応しうると考えられる。こうした言語的側面や表現内容に基づく分析的評価観点と学習指導要領における観点別評価の3観点との関係性を明示することで、教室内でのパフォーマンス評価と制度的な評価の枠組みとの整合性がより高まることが期待される。

さらに、複数パラグラフの英文を書く課題において、「内容伝達に支障をきたす誤り」が複数出現する可能性が高いという懸念も示された。例えば、ある生徒が序論一本論一結論の3パラグラフを構成した場合、一つのパラグラフ内で主語・動詞の欠落や語順の誤りなどが複数含まれると、それだけでC評価に分類される可能性が高くなる。その結果、多くの受験者がC評価になってしまふとパフォーマンス評価の妥当性を下げる要因にもなり得るため、ルーブリックの記述内容やA～Cの基準設定において慎重な検討が求められる。

表7: ライティングテストのルーブリック

内容		構成	正確さ
A	内容に一貫性があり、例・理由などに説得力がある。	序論一本論一結論の構成が確立し、さらに結論に工夫が見られる。	内容伝達に支障をきたす誤りがない。
B	内容に一貫性がある。	序論一本論一結論の構成が確立している。	内容伝達に支障をきたす誤りが一つある。
C	内容に一貫性がない。	序論一本論一結論の構成が確立していない。	内容伝達に支障をきたす誤りが二つ以上ある。
備考	内容の面白さではなく、与えられたトピックに関して首尾一貫しているか、付加情報があるなどを判断する。	「結論に工夫」とは、単に序論を繰り返すのではなく、言い換えたり、判断・感想などを加えたりしていることとする。	S+V構造の欠如、語順の誤り、時制の誤りを含むものなどに加えて、意味が通じないものも含む。

4.2.3 試験全体の妥当性や改善すべき点について

本調査においては、学習内容の定着度を測定する到達度テストとして、①Can-doリスト・指導と評価の計画とテスト内容が一致しているか、②学習指導要領に基づいた観点別評価(知識・技能、思考・判断・表現)が適切に行われているか、の観点で検討された。全体としては、定期試験はリーディング及びリスニ

ングを中心に構成されており、文法・語法・語彙といった言語知識も含まれている点から、ある程度は授業内外の学習内容に即した出題であったとの評価が得られた一方で、学習内容との関連性や観点別評価観点との整合性について複数の課題が指摘された。

大問1では、小問Aはディクテーションの形式を取っているが、単元の指導と評価の計画においてディクテーション活動が明記されておらず、学習内容との関連性がやや不明瞭であるとされた。したがって、ディクテーションを評価に用いる場合には、計画段階でその活動を明記し、授業内外での学習を経た上で試験に組み込むことが妥当であると考えられる。大問1の小問B及びCについては、「待ち合わせの場所について話しています」といった簡潔な場面が提示されてはいるものの、場面や目的、聞き手の立場などの情報が不十分であり、「思考・判断・表現」の観点からの評価には適していない可能性があるとの指摘があった。図5に大問1の設問内容を示す。国立教育政策研究所(2021, p. 64)による「聞くこと」の「思考・判断・表現」を評価する問題例においては、例えば Kasumi and her classmate Michael are talking in their classroom. Listen to the dialogue and choose the best option from among the four choices. You will be able to listen to the dialogue only once.といった場面設定が導入されている。しかし、渡邊ら(2025)はこのような例について、場面のみが記述されており、目的や状況が明示されていない点を問題視している。さらに、大学入学共通テスト英語(リスニング)においても本調査の定期試験と同程度の場面設定が多く受けられる他、場面設定の提示のない問題項目もある(大学入試センター、2025b)。そのため、目的や状況をどの程度まで場面設定として明示すべきかについては、今後さらなる検討が必要であるが、学習指導要領に基づいて生徒の思考力・判断力・表現力を測る問題を作成するためには、より明確な場面設定を導入することは改善点の一つとして挙げられる。

大問4のA(1)に関しては、内容理解ではなく英文のつながりに焦点を当てた設問となっており、「思考・判断・表現」の観点で評価するには不適切ではないかとの指摘があった。大問4のAの設問内容は図6に示す通りであり、(1)では英文中に設けられた空所に対し、前後の文脈を手がかりに適切な語句を選択させる形式である。国立教育政策研究所(2021, p. 38)は、「読むこと」における「思考・判断・表現」の評価規準として、「コミュニケーションを行う目的や場面、状況などに応じて、日常的な話題や社会的な話題について、必要な情報を読み取り、書き手の意図や概要、要点を捉えている」としている。この規準に照らすと、空所補充のように語彙や接続表現の知識に依存する設問は、書き手の意図や概要・要点の把握には直接関与せず、「知識・技能」の観点に分類するのが適切であると考えられる。一方、渡邊ら(2025)は、「読むこと」の「知識・技能」と「思考・判断・表現」を区別する際に、「内容を捉える技能」と「概要・要点を捉えている状況」を明確に分けた問題作りの難しさを指摘している。実際にこのようなテスト項目と評価観点の分類との齟齬は、村越・江原(2015)においても指摘されており、平成29年・30年告示学習指導要領下においても同様の問題が依然として存在していることが示唆される。

次に、技能ごとの配点の割合については、リスニング20%、リーディング40%、文法・語法20%、語彙20%という構成であった。この配分に対しては、英語4技能の育成を目的とする英語コミュニケーションⅢの評価としては、文法・語法・語彙に対する比重がやや高く、一方でリスニングが相対的に少ないという課題が指摘された。この配分がCan-doリストや単元の指導と評価の計画と整合的であるかについては、改めて検証する必要がある。実際に、勤務校のCan-doリスト及び単元の指導と評価の計画においては、英語4技能をバランスよく指導・評価することが示されており、リスニングも他技能と同等の比重で配点されることが理想である。なお、本研究では、定期試験及びパフォーマンステストが年間成績に占める割合や、両者を合わせた全体の評価比率を正確に算出することは困難であった。そのため、ここで示す配分は定期試験内での技能比率に限られる点に留意する必要がある。

また、観点別評価における配点の割合は、「知識・技能」48%、「思考・判断・表現」52%と一見バランスが取れているように見えるが、前述のように観点のラベリングに齟齬が含まれている可能性がある。そのため、各設問がどの評価観点に適切に分類されているかを再確認した上で、実際にバランスの取れた評価が行われているかを見直す必要があるとの指摘があった。

1【リスニング】知識・技能[8] 思考・判断・表現[12]

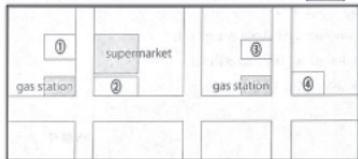
【A】〈記述問題〉以下の英文はこれから読まれる英文のスクリプトです。英文を聞き、空所(あへく)に読まれた語または数字を書き入れ、解答用紙にその答えを書きなさい。1回読れます。(知技1点×8)

(※解答欄は解答用紙の裏面にあります。)

You might think that the traditional French ham and butter baguette sandwich would be the (あ) popular fast food in France. But think again. A recent study shows that sales of American-style hamburgers have (い) the sales of these sandwiches. According to a French restaurant marketing company, about 1.2 billion ham and butter baguette sandwiches were sold in France in 2017, while 200 (う) more burgers were sold than sandwiches during the same period. Sales of both baguette sandwiches and hamburgers are on the (え), the study says. It notes that baguette sandwiches saw 1.3 percent growth in 2017, while burger sales recorded a 9 percent (お). Curiously, in France, the number of hamburgers sold by a major fast food restaurant chain has not varied for years. What has changed, however, is the (か) number of restaurants putting burgers on their menus. In fact, (き) percent of restaurants surveyed in France included burgers on their menus in 2017. The rise of the American-style burger was noted in a newspaper story which reported that many famous restaurants in Paris (く) burgers to their menus for the first time 10 years ago.

【B】〈マークシート問題〉対話とそれについての問い合わせを聞き、答えとして最も適切なものを一つずつ選びなさい。1回流します。(思判表2点×3)

問1 特ち合わせ場所について話をしています。 1



問2 昨晩のバーディについて話をしています。 2

- ① Bored.
- ② Concerned.
- ③ Funny.
- ④ Surprised.

問3 夫婦がお店で話をしています。 3

- ① \$100
- ② \$120
- ③ \$140
- ④ \$160

【C】〈マークシート問題〉対話を聞き、問い合わせとして最も適切なものを一つずつ選びなさい。1回流します。(思判表2点×3)

問1 男性が女性に今朝の状況を説明しています。 4

- How did the man arrive?
- ① By airplane.
 - ② By car.
 - ③ On the 9:30 train.
 - ④ On the 10:00 train.

問2 ホームパーティについて話をしています。 5

- Where is the man's house?
- ① Across from the high school.
 - ② Around the corner from the women's house.
 - ③ In the neighborhood of the city office.
 - ④ Next to the convenience store.

問3 映画について話をしています。 6

- Which of the following is true about the women?
- ① They didn't watch the movie in the afternoon.
 - ② They each spent \$10 to enter the movie.
 - ③ They enjoyed drinking instead of watching the movie.
 - ④ They paid as much for drinks for watching the movie.

■図5: 対象となった定期試験問題の大問1(A~C)

4【FLEX】知識・技能〈マークシート問題〉思考・判断・表現 [20]

【A】 You are taking a lecture about optical illusions and reading the following text. Read the text and answer the questions below.

Take a look at the object in the illustration below. What does it look like to you? To some people this might appear to be a rabbit facing to the right. (47), it may seem like a duck looking in the opposite direction to others. This illustration is an example of what is called an “optical illusion”, (48) it shows how our brains can interpret the same image in different ways. Illusions such as these have a variety of uses and are often seen around us in our daily lives.

To start, these types of illusions have traditionally been incorporated into various structural and architectural designs. For example, the *karesansui* (dry landscape) garden at the Zen temple Ryōanji in Kyoto illustrates this concept well. At first glance, its entire surface may appear perfectly flat, but if we look closer, we can see that it actually inclines towards the garden’s south-east corner. (49), the earthen wall on the west-side is slightly higher on its northern end. These features combine to create an optical illusion and to our eyes the surface actually appears to be deeper than it truly is. This same effect can also be seen in the Ponzo Illusion. In the Ponzo Illusion, the uppermost horizontal line appears to be longer the line just below it, but (50), the two lines are exactly the same length.

Second, if we look at the image below, we can see another example of an optical illusion. The interior of the house appears to have an extremely high ceiling though it is actually much lower than it looks. Our brains interpret the image in this way as a result of the Fick Illusion (Figure 2). In the Fick Illusion, the vertical line seems longer than the horizontal line. (51), they are exactly the same length. This effect can be very useful in architecture, as it can cause us to think a room is much more spacious than it actually is.

Third, another valuable technique which has been used successfully in recent times is the creation of what appears to be a three-dimensional object on a two-dimensional surface. (52), in one particular train station people are often unable to find an elevator because it is located in a blind spot, so a sign that appears three-dimensional has been painted on the floor in order to help them find it more easily. The sign quickly catches people’s attention and has been extremely successful in helping reduce the number of people who have become lost while trying to locate the elevator.

Last, they have been found to be useful in the field of transportation, in particular at intersections which don’t have traffic lights. Crosswalk markers which appear to be three-dimensional have been painted on some roads in an effort to prevent traffic accidents from occurring. From the viewpoint of approaching drivers, it looks like they are about to hit large concrete blocks which are in the road. (53), these were created in hopes that the drivers will see them and slow down as they get nearer the crosswalk.

Optical illusions are just illusions. They cannot magically make a garden larger, nor can they truly make a ceiling higher. Over time, even if a crosswalk is painted to appear three-dimensional, it will eventually lose its effectiveness in preventing accidents as drivers who regularly see these markers will get used to them and come to understand that what they are seeing is just an illusion and they won’t slow down. (54), even if we recognize such common limitations of these illusions, we should be able to find numerous other useful applications for optical illusions in society.

(1) Put an appropriate word or phrase in each blank in the text. (1点×8)

47	①On the other hand	②In fact	③As a result	④Besides
48	①conversely	②therefore	③for example	④in other words
49	①However	②Besides	③On the other hand	④For instance
50	①In conclusion	②In short	③in fact	④meanwhile
51	①In general	②Consequently	③Moreover	④However
52	①Additionally	②For instance	③Therefore	④As a result
53	①On the contrary	②In fact	③Thus	④Furthermore
54	①To sum up	②Nevertheless	③In addition	④In other words

(2) Which is the main point of the text? (2点) 55

- ① Optical illusions can create the illusion of depth in gardens and structural designs.
- ② Optical illusions are adaptable tools used in various aspects of daily life, including architecture, navigation, and transportation.
- ③ The effectiveness of optical illusions diminishes over time as people get accustomed to them.
- ④ The illustration of the rabbit and duck is a classic example of how optical illusions work.

図6: 対象となった定期試験問題の大問4のA

パフォーマンステストについては、全体として授業内容との関連性は高く評価された。例えば、スピーキングテスト(図2参照)では、教科書に登場する動物に関するトピックを基に「中学生に向けて動物について説明する」といったタスクが設定されており、学習した語彙や表現を活用できる設計となっていた。このように学習内容との整合性は見られた一方で、タスクと単元の指導と評価の計画との整合性には課題が見られた。具体的には、評価観点に「Eye-contact」や「Delivery」が含まれていたが、これらの技能については指導と評価の計画に記載がなかったため、計画上の学習活動との対応が不十分であるとの指摘があった。

ライティングテスト(図3参照)については、リーディング活動と組み合わせた技能統合型のタスクとして構成されており、与えられた英文資料を読み取った上で、自分の意見を英語で記述するという形式が取られていた。この点は、実際の言語使用場面に近いタスク設計であるとして肯定的に評価された。しかしながら、評価の中心が「思考・判断・表現」に置かれているにもかかわらず、「誰に向けて、どのような状況で書くのか」といったコミュニケーションの目的や場面、状況設定が明示されていなかったことから、評価規準の前提条件が不明瞭であるとの指摘があった。国立教育政策研究所(2021, p. 76)の参考資料の例においては、「下記は海外交流校の生徒が送ってきた質問です。質問を読み、どちらがより効果的だと思うか英語で書きなさい。」という指示文が提示されており、ライティングに際して誰に向けて書くべきかが明確になっている。本調査で対象となった「ソーシャルメディアの利点や欠点についての意見を書く」というタスクであれば、「海外交流校の高校生との議論に向けて」など、ライティングの目的や想定読者を明示することで、学習者の表現の方向性が明確になり、評価の妥当性も高まると考えられる。

これらのことから、パフォーマンステストの設計においては、学習活動との一貫性を確保しながら、課題文の条件設定をより明確化するとともに、Can-doリストとの対応関係やループリックの評価観点の構造を可視化することが、妥当性の向上につながるといえる。また、Can-doリストに基づく学習到達目標が、実際の指導内容や評価活動と一貫性をもって連動しているかを検証することも求められる。最後に、本フィードバックを踏まえた今後の改善点として、これまでに指摘された事項に加え、テスト仕様書の作成が重要であるとの指摘があった。これにより、各テスト項目が測定しようとする能力(construct)を明確に定義することができ、テスト全体の妥当性の向上に資することができる。

4.3 波及効果推論に係る生徒質問紙調査

本調査では、定期試験後に勤務校の高校1年生から3年生までの生徒計1,103名を対象に、英語コミュニケーションの定期試験に関する質問紙調査を実施し、924名から回答を得た(回収率83.8%)。学年別の回答者数は、1年生370名、2年生300名、3年生254名であり、全学年から概ね均等に回答が集まった。各学年の英語コミュニケーションの定期試験は、概ね教科書を題材としたリーディング問題、文法・語法問題、さらには大学入試問題や外部英語試験を基にしたリーディング問題で構成されている。なお、リスニング問題は2年生の試験にのみ含まれており、1年生及び3年生については、定期試験とは別に、授業内でリスニング試験が実施されているため、本調査で扱った定期試験には含まれていない。また、生徒が回答した定期試験の内容は、本研究において専門家レビューの対象となった定期試験問題とは異なるものである。

Allen et al.(2024)の質問紙調査では、英語圏での滞在経験を有する生徒は全体の25.9%、将来的に英語圏での生活や留学を希望している生徒は68.5%であった。これに対し、本調査では滞在経験を有する生徒は11.6%、留学等を希望している生徒は50.2%にとどまった。これらの結果から、本研究の対象生徒は一定の関心を示してはいるものの、英語圏での経験及び将来的な関心のいずれにおいても、先行研究の対象生徒に比べて相対的に低い関心を示していることが分かった。生徒質問紙調査の質問項目及び調査結果並びに選択式回答項目の選択肢及び回答結果を資料6,7に示す。

本調査における授業に関する設問では、Q5「先生が話し、生徒が聞く時間が多い」($M = 3.5$)、Q6「先生は主に英語を話す」($M = 3.9$)といった項目には6段階の中間値(3.5)と同程度及びやや高めの値を示した。本調査では3と4のラベルが設定されていないため厳密な解釈は難しいが、平均値としては中程度及びや

や高めといえる。一方で、Q7「生徒が先生と話す際に英語を使う」($M = 2.9$)、Q8「生徒同士のやりとりで英語を使う」($M = 3.2$)といった生徒のアウトプットに関する項目は中間値よりも低い傾向を示した。ただし、Q9「即興で英語を話すことが求められることが多い」という項目については($M = 3.7$)と中間値よりもやや高めだった。全体としては教師によるインプットは比較的豊富である一方で、生徒の英語によるアウトプットの機会は限定的であることがうかがえる。これらの結果は、今後の授業改善において、生徒の発話を促進する授業設計の必要性を示唆するものである。一方、Allen et al. (2024) の同様の設問においては、Q5「先生が話し、生徒が聞く時間が長い」($M = 2.4$)と低く評価される一方、Q6「先生は主に英語を話す」($M = 5.3$)、Q7「生徒が先生と英語で話す」($M = 4.2$)、Q8「生徒同士で英語を使う」($M = 4.2$)、Q9「即興で英語を話すことが求められる」($M = 5.1$)と、アウトプット重視の授業実践がより積極的に行われている様子が示されている。この比較から、本研究の対象校では教師主導の授業が中心であるのに対し、Allen et al. の対象校では生徒の主体的な英語使用がより重視されていることがうかがえる。今後は、生徒の発話機会を体系的に増やすための授業デザインの検討が、よりバランスの取れた言語活動の促進に寄与すると考えられる。

Q10「英語コミュニケーションの授業の活動は主に何に焦点を当てていますか? (複数回答可)」という設問に対しては、「スピーキング」(58.9%)、「リスニング」(58.6%)、「リーディング」(56.9%)といった技能項目が語彙や文法などの知識項目より高い割合で選ばれた(表8)。学年別に見ると、リスニングは3年生で最も高い割合(63.4%)を示し、定期試験に含まれていないにもかかわらず、生徒は授業で重視されていると認識していた。一方で2年生は定期試験にリスニング問題が含まれているにもかかわらず、その割合は55.7%と低く、授業と試験設計の方針が必ずしも一致していないことが示唆された。特に3年生では、大学入学共通テストにおけるリスニング配点の増加や英検などの外部検定試験対策が授業内容に影響している可能性があるが、本調査はその因果関係を直接検証するものではなく、教師・生徒へのインタビューなど他の方法による裏付けが必要である。

一方、「スピーキング」に焦点を当てていると回答した割合は、1年生71.4%、2年生53.7%、3年生46.9%と学年が上がるにつれて低下した。これはAllen et al. (2024) が指摘するように、大学入試におけるスピーキング技能の軽視が授業内容や学習活動の変化に負の波及効果(negative washback)を与えている可能性がある。

表8: Q10「英語コミュニケーションの授業の活動は主に何に焦点を当てていますか?」(複数回答可)

	語彙	文法	発音	リスニング	リーディング	スピーキング	ライティング
全体	回答件数	474	414	405	541	526	544
	回答率	51.4%	44.8%	43.8%	58.6%	56.9%	58.9%
1年	回答件数	200	152	211	213	214	264
	回答率	54.1%	41.1%	57.0%	57.6%	57.8%	71.4%
2年	回答件数	148	159	105	167	144	161
	回答率	49.3%	53.0%	35.0%	55.7%	48.0%	53.7%
3年	回答件数	127	103	89	161	168	119
	回答率	50.0%	40.6%	35.0%	63.4%	66.1%	46.9%

注. N=924. 複数回答を可としたため、回答率の合計は100を超える。

Q24「今回の定期試験のために学校外では主に何を中心に学習しましたか?」(複数回答可)への回答では、多くの生徒が語彙(59.4%)、リーディング(49.7%)、文法(41.6%)を重点的に学習しており、それ以外の領域は低い割合にとどまった(表9)。特にリスニングは、1年生(2.7%)や3年生(2.4%)で極めて低く、2年生では21.7%と相対的に高かったものの、依然として他の領域に比べて低水準であった。この結果は、2年生で

リスニング問題が定期試験に含まれていることが学習行動に一定の影響を与える可能性を示す一方で、その波及効果は限定的であることも示唆している。その要因の一つとして、定期試験におけるリスニングの配点が比較的低く、学習者にとって優先度が下がってしまうことが考えられる。一方で、ライティングやスピーキングは定期試験ではなくパフォーマンステストで評価されるため、定期試験における波及効果はほとんど見られない。今後は、パフォーマンステストにおける波及効果についても検証する必要がある。

■表9: Q24「今回の定期試験のために学校外では主に何を中心に学習しましたか?」(複数回答可)

		語彙	文法	発音	リスニング	リーディング	スピーキング	ライティング	ない
全体	回答件数	549	384	40	81	459	56	102	74
	回答率	59.4%	41.6%	4.3%	8.8%	49.7%	6.1%	11.0%	8.0%
1年	回答件数	228	153	17	10	192	26	61	25
	回答率	61.6%	41.4%	4.6%	2.7%	51.9%	7.0%	16.5%	6.8%
2年	回答件数	157	140	18	65	154	19	33	21
	回答率	52.3%	46.7%	6.0%	21.7%	51.3%	6.3%	11.0%	7.0%
3年	回答件数	164	91	5	6	113	11	8	28
	回答率	64.6%	35.8%	2.0%	2.4%	44.5%	4.3%	3.1%	11.0%

注. N=924. 複数回答を可としたため、回答率の合計は100を超える。

スコアレポート(得点個表)に関する設問では、Q30「自身の強みや弱点を理解するのに役立つ」、Q31「内容が明確である」、Q32「解釈しやすい」(いずれも $M = 4.6$) といった肯定的な評価が得られた。学年によって多少の違いはあるものの、勤務校で用いられているスコアレポートは生徒の自己評価や学習の振り返りを促す有効な資料として機能していることが示唆される(図7参照)。一方で、小泉ら(2023)の研究では、スコアレポートを有効に活用するためには、単なる情報提供にとどまらず、解説動画や自己分析活動と組み合わせることで、学習への結び付きが強化されることが指摘されている。こうした知見を踏まえると、勤務校におけるスコアレポートの活用においても、どのように活用されているかを検討し、生徒のより自律的な学習支援へつなげていく余地があると考えられる。

11組 1番 氏名 模範解答																																																																															
番号	正答	マーク	判定	観点	得点	正答率	番号	正答	マーク																																																																						
(1)	2	2	○	思判	2	54.2	(3)	3	3																																																																						
(2)	1	1	○	思判	1	53.3	(29)	1	1																																																																						
(3)	4	4	○	思判	2	47.6	(30)	4	4																																																																						
(4)	1	1	○	思判	2	9.1	(31)	1	1																																																																						
(5)	3	3	○	思判	2	40.3	(32)	5	5																																																																						
(6)	3	3	○	思判	2	51.9	(33)	4	4																																																																						
(7)	3	3	○	思判	2	37.5	(34)	1	1																																																																						
(8)	4	4	○	思判	2	62	(35)	2	2																																																																						
(9)	1	1	○	思判	2	32	(36)	3	3																																																																						
(10)	1	1	○	思判	2	54.2	(37)	5	5																																																																						
(11)	3	3	○	知接	2	71	(38)	3	3																																																																						
(12)	2	2	○	知接	2	90.2	(39)	5	5																																																																						
(13)	4	4	○	知接	2	56.7	(40)	6	6																																																																						
(14)	4	4	○	知接	1	64	(41)	4	4																																																																						
(15)	4	4	○	知接	1	40.3	(42)	1	1																																																																						
(16)	2	2	○	知接	1	67.8	(43)	3	3																																																																						
(17)	2	2	○	知接	1	84.4	(44)	2	2																																																																						
(18)	2	2	○	知接	1	85.9	(45)	2	2																																																																						
(19)	1	1	○	知接	1	85.9	(46)	4	4																																																																						
(20)	2	2	○	知接	1	78.8	(47)	5	5																																																																						
(21)	3	3	○	知接	1	82.6	(48)	2	2																																																																						
(22)	3	3	○	知接	1	79.3	(49)	3	3																																																																						
(23)	2	2	○	知接	1	61.2	(50)	1	1																																																																						
(24)	4	4	○	知接	2	22.4	(51)	2	2																																																																						
(25)	3	3	○	知接	2	56.5	(52)	3	3																																																																						
(26)	2	2	○	知接	2	68.5	(53)	1	1																																																																						
(27)	4	4	○	知接	2	32.5																																																																									
<table border="1"> <tr> <td>マーク</td> <td>46</td> <td>/</td> <td>46</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>思判</td> <td>36</td> <td>/</td> <td>36</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>記述</td> <td>10</td> <td>/</td> <td>10</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>思判</td> <td>8</td> <td>/</td> <td>8</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>小計</td> <td>56</td> <td>/</td> <td>56</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>思判</td> <td>44</td> <td>/</td> <td>44</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>合計</td> <td>100</td> <td>/</td> <td>100</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>										マーク	46	/	46							思判	36	/	36							記述	10	/	10							思判	8	/	8							小計	56	/	56							思判	44	/	44							合計	100	/	100						
マーク	46	/	46																																																																												
思判	36	/	36																																																																												
記述	10	/	10																																																																												
思判	8	/	8																																																																												
小計	56	/	56																																																																												
思判	44	/	44																																																																												
合計	100	/	100																																																																												

■図7: 勤務校で用いられるスコアレポートのサンプル

注:正答率は学年の正答率を示す。

5

考察

5.1 教師質問紙調査と専門家レビューを通じた総合的考察(RQ1及びRQ2)

本研究では、教室内英語テストの妥当性を論証に基づく妥当性検証の領域定義推論の観点から検討するため、現職教師に対する質問紙調査と、言語テストの専門家による文書レビューを実施した。これら二つの調査は、それぞれ異なる視点から現行のテスト実践における妥当性の課題と改善の方向性を浮き彫りにしている。

まず、教師質問紙調査の結果から、多くの教師が教科書との整合性や授業内容との連動を強く意識して定期試験やパフォーマンステストを設計している一方で、学習指導要領や単元の指導と評価の計画、Can-doリストとの接続に関しては、教師間での理解の不十分さやばらつきが見られた。しかしながら、本来教室内での評価は、学習指導要領に示された目標に基づいて、Can-doリストや単元ごとの指導と評価の計画が作成され、それらと一貫性をもった形で設計・実施されるべきである。したがって、教師は教科書や授業内容のみならず、TLU領域を定義する制度的文書との整合性を含めた、全体の連動性を意識する必要がある。このような視点に基づく教師間の認識の共有と自治体などの支援が、教室内評価の妥当性を高める上で不可欠であると考える。さらに、観点別評価における評価規準の理解に教師間でばらつきが見られた。これは、観点別評価を運用する際の測定対象(構成概念)や評価観点に関する理論的理解と共通認識が十分に共有されていないことを示唆している。このように、評価の設計が表面的な整合性にとどまり、構成概念や能力記述に基づいた評価規準の理解が曖昧なまま運用される場合、評価の信頼性及び妥当性を損なうリスクがある。

これらの傾向は、専門家レビューの結果とも一致する部分がある。専門家からは、定期試験やパフォーマンステストにおいて、学習内容との整合性は概ね良好であるとの評価がなされた一方で、ルーブリックの言語的侧面や表現内容に基づく分析的評価観点と学習指導要領に基づく評価観点(知識・技能、思考・判断・表現)との対応関係が不明確であること、TLU領域における測定すべき能力の明示が一部不十分であることが課題として指摘された。このことは、現場教師が観点別評価における構成概念を統一的に捉えることの難しさを示しており、自治体などによる研修を含めた実践・理論の両面からの支援が必要であることを示唆している。

以上の結果から、現行の教室内英語テストは、授業実践に基づいた設計という点では一定の妥当性を有するものの、TLU領域が定義されているとされる学習指導要領やCan-doリスト、単元の指導と評価の計画などの制度的な文書との整合性について、さらに観点別評価における測定すべき構成概念についての教師間の共通理解について課題が見られることが明らかとなった。したがって、本結果は領域定義推論における前提である「高等学校における教室内英語テストにおいてTLU領域は明確に定義されており、テストのタスクがTLU領域における重要なスキルを適切にカバーしている」を一部支持するものの、上記のように改善すべき課題が見出された。これらの課題は、今後の教室内評価の改善に向けた重要な論点であると考えられる。

5.2 生徒質問紙調査の考察(RQ3)

本調査の結果から、定期試験が生徒の学習に対して一定の波及効果を及ぼしている可能性が示唆された。まず、試験準備では、多くの生徒が語彙(59.4%)、リーディング(49.7%)、文法(41.6%)といった言語知識やリーディングに重点を置いており、試験後に「自信がついた」項目でもリーディング(29.1%)、語彙(25.0%)、文法(24.2%)が上位を占めた。さらに、試験後に「さらに練習したい」と回答した項目もリーディング(58.0%)、語彙(57.7%)が高く、定期試験を契機とした継続的学習意欲の喚起が示唆される。一方で、2年生にのみ定

期試験にリスニング問題が含まれていたものの、他の領域と比べて低水準であった。対象が異なる試験であるが、専門家によるフィードバックにおいて定期試験のリスニングの配点が小さいという指摘も踏まえると、リスニングは授業で一定程度扱われているにもかかわらず、定期試験での出題の有無や配点が他の領域よりも少ないことが、学習者の優先度を下げ、プラスの波及効果を制限している可能性も考えられる。大学入学共通テストにおいてリーディングとリスニングの配点が同じであることも踏まえると、今後は教室内評価におけるリスニングの比率を高める必要性があるかもしれない。また、ライティングやスピーチングは定期試験ではなくパフォーマンステストで評価されることを前提としているため、定期試験における波及効果はほとんど見られない。今後は、パフォーマンステストの波及効果についても検証する必要がある。

また、「試験の内容が授業と似ていた」「授業が試験準備に役立った」といった項目は平均4前後と中間値よりもやや高い評価を得ており、授業と試験内容の整合性がある程度確保されていることが分かる。これは、授業で学んだ内容が試験準備へと直接つながるというプラスの波及効果がある程度現れていると考えることができる。しかし、6段階評価において満点に近い値ではなくことから、その効果は十分に最大化されているとは言い難く、改善の余地がある。さらに、授業の内容がそのまま試験に出題され、試験準備が単なる暗記になってしまふ可能性があり、この点については別途インタビュー調査などで検証が必要である。また、専門家によるフィードバック調査における、定期試験の問題項目について目的・場面・状況の設定が不十分であるという指摘を踏まえると、日頃の授業から目的・場面・状況を明確に設定した言語活動を行い、それを評価タスクに反映させることができ、授業と試験の一貫性をより高め、プラスの波及効果をより引き出すことにつながる可能性がある。

スコアレポート(得点個表)については、「自身の強みや弱点を理解するのに役立つ」($M = 4.3$)、「内容が明確である」「解釈しやすい」とともに($M = 4.6$)といった肯定的評価が得られ、生徒の自己評価や学習の振り返りを促す有効な資料として機能していることが示唆された。しかし、6段階評価において満点に近い値ではなく、依然として改善の余地がある。改善の方向性としては、小泉ら(2023)が指摘するように、単なる情報提供にとどまらず、解説動画や自己分析ワークシートとの組み合わせ、授業内での振り返り活動や教師による具体的なフィードバックの導入などが考えられる。これにより、スコアレポートの活用がより能動的な学習行動につながりやすくなると期待される。

以上より、本研究の対象となった定期試験は、特にリーディングや語彙・文法といった領域でプラスの波及効果を示す兆候がある一方で、リスニングへの波及効果は限定的であり、スコアレポート活用にも改善の余地がある。これらは波及効果推論における前提「高等学校における教室内英語テストは生徒に有益な波及効果を与えている」を一部支持しつつも、一部反論する結果となっていることが考えられるため、定期試験やスコアレポートの活用についてさらなる改善が求められる。しかしながら、パフォーマンステストの波及効果の検証など、本研究では検討できていない要素が多数存在することに留意が必要である。また、本研究で得られた結果は、生徒質問紙調査という自己報告データのみに基づくものであり、実際の授業行動や学習成果を直接観察したものではない。そのため、定期試験やパフォーマンステストが学習に及ぼす波及効果の全体像を把握するには限界がある。今後は、授業観察や生徒・教師へのインタビュー、学習成果の総合的測定など、多角的な方法を組み合わせることで、より多面的で信頼性の高い検証を行う必要がある。

6

結論

本研究では、高等学校における教室内英語テストの妥当性を、Chapelle and Voss(2021)の論証に基づく妥当性検証の枠組みにおける「領域定義推論(Domain Definition Inference)」及び「波及効果推論(Consequence Implication Inference)」の観点から検討した。現職教師に対する質問紙調査、言語テストの専門家による文書レビュー、生徒に対する質問紙調査を通して、教室内テストの設計・運用に関する

現状と課題を明らかにした。

まず、RQ1「教室内英語テストにおいて目標言語使用領域は明確に定義されているか」については、多くの教師が教科書や授業内容との整合性を意識して評価を設計している一方で、学習指導要領、Can-doリスト、単元の指導と評価の計画といった制度的な文書との接続に対する理解や実践にはばらつきが見られ、TLU領域の定義が必ずしも明確に共有・運用されているとは言えない点が確認され、領域定義推論の前提を一部支持するものの、改善が必要であることが分かった。

次に、RQ2「テストのタスクが目標言語使用領域の重要なスキルを適切にカバーしているか」については、学習内容との整合性は一定程度確認されたものの、タスクの設計において測定すべき構成概念の明示が十分でない場合や、教師によって分類された評価観点のラベリングに齟齬があること、さらにはループリックの評価観点と学習指導要領に基づく評価観点との対応関係が不明確であるといった課題が指摘された。これらは、領域定義推論の前提を一部支持するものの、評価設計における理論的な裏付けの不足や教師間の共通認識について今後の改善の必要性を示唆している。

また、RQ3「教室内英語テストは生徒にどのような波及効果を与えているか」については、特にリーディングや語彙・文法といった領域においてはプラスの波及効果の兆候が確認された一方で、リスニングについての波及効果は限定的であり、プラスの波及効果を与えていない兆候も見られた。これらは、波及効果推論の前提を一部支持しつつも、一部反論する結果となっていることが考えられるため、定期試験における配点の割合のバランスなどの改善が求められる。また、スピーキングやライティングといったアウトプット技能については、別途実施されるパフォーマンステストの波及効果の検討が求められる。

本研究で扱った教師質問紙調査・生徒質問紙調査・専門家レビューはいずれも筆者の勤務校における実施に限定されており、結果の一般化には限界がある。また、本研究では質問紙調査と文書レビューを主な方法として採用したが、教師や生徒へのインタビュー、授業観察、テスト結果の得点分析など、より多角的なアプローチを取り入れることで、より深い洞察や裏付けが得られる。したがって、今後は複数校を対象とした調査の実施や混合研究法による検証を通じて、教室内英語テストの妥当性に関する理解をさらに深めていく必要がある。

本研究は、高等学校における教室内英語テストを対象に、論証に基づく妥当性検証の枠組みを適用した、国内では先駆的な試みである。これまでの妥当性検証は主に大規模標準化テストに焦点を当てたものが多く、日常的に実施される教室内テストにこの枠組みを応用した研究は極めて限られていた。こうした背景のもと、本研究は、実際の教育現場で行われる定期試験やパフォーマンステストについて、領域定義推論及び波及効果推論という二つの視点から妥当性を検討し、評価実践に内在する課題を可視化とともに、今後の改善に向けた示唆を提供するものである。今後は本研究を皮切りにして、さらに領域定義推論と波及効果推論の証拠を補強するとともに、得点化推論をはじめ、一般化推論、説明推論、外挿推論、利用推論の他の推論についても継続的に検討することが求められる。約10年ごとに改訂される学習指導要領に基づく学習評価については、本研究で採用した論証に基づく妥当性検証の枠組みなどの系統的な理論的枠組みを用いて、その妥当性を検討することが重要である。本研究はその端緒を示すものであり、今後、この枠組みを活用した検証が進むことで、より適切な指導と評価の実現につながることが期待される。

謝辞

はじめに、本研究を実施する機会を与えてくださった公益財団法人 日本英語検定協会の皆様、及び選考委員の皆様に心より感謝申し上げます。特に、研究助言者である筑波大学の小泉利恵先生には研究実施から執筆にあたり貴重なご指導を頂戴いたしました。深く感謝申し上げます。また、本研究にあたって貴重なご意見をくださった琉球大学の深澤真先生をはじめ、英検研究助成についてご紹介ください、日々の研究活動の中で多くのご助言をいただいている横浜国立大学の斎田智里先生、及び同大学院ゼミの皆様にも深く感謝申し上げます。最後に、本研究にご協力してくださいました高校生並びに先生方に感謝申し上げます。なお、本報告書に関して、開示すべき利益相反関連事項はありません。

引用文献

- Allen, D., Yamamura, K., Meguro, S., & Nakamura, T. (2024). Washback from four-skills English exams in the Japanese secondary school context. *Language Assessment Quarterly*, 21, 473-505. <https://doi.org/10.1080/15434303.2024.2440895>
- 馬場哲生 (2023). 日本の中・高等学校の英語教育における観点別学習状況の評価: 歴史的変遷と今後の課題. 東京学芸大学紀要, 74, 467-481. <https://u-gakugei.repo.nii.ac.jp/records/55991>
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Chapelle, C., Enright, M., & Jamieson, J. (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.
- Chapelle, C., & Voss, E. (Eds.). (2021). *Validity argument in language testing: Case studies of validation research*. Cambridge University Press.
- 中央教育審議会 (2025). 豊かな学びに繋がる学習評価の在り方～過度な負担を生じさせない在り方との両立～. https://www.mext.go.jp/content/20250711-mxt_kyoiku01-000043568_03.pdf
- Dai, D. W., Vu, T., Knoch, U., Lim, A. S., Malone, D. T., & Mak, V. (2024). Expanding Kane's argument-based validity framework: What can validation practices in language assessment offer health professions education? *Medical Education*, 58(12), 1462-1468. <https://doi.org/10.1111/medu.15452>
- 大学入試センター. (2025a). 令和7年度大学入学共通テスト本試験の問題英語(リーディング). https://www.dnc.ac.jp/albums/abm.php?d=771&f=abm00005680.pdf&n=2025_op_20_reading.pdf
- 大学入試センター. (2025b). 令和7年度大学入学共通テスト本試験の問題英語(リスニング). https://www.dnc.ac.jp/albums/abm.php?d=771&f=abm00005676.pdf&n=2025_op_25_listening.pdf
- Dursun, A., & Li, Z. (2021). A systematic review of argument-based validation studies in the field of language testing (2000-2018). In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 45-70). Cambridge University Press.
- 廣森友人・小金丸倫隆 (編). (2024). エンゲージメント×英語授業「やる気」と「意欲」を引き出す授業のつくり方. 明治図書.
- 今井裕之 (2023). 外国語科における三観点評価の課題. KELES ジャーナル, 8, 82-86. https://doi.org/10.18989/keles.8.0_82
- Jin, Y. (2021). Test specifications. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (pp. 271-288). Routledge. <https://doi.org/10.4324/9781003220756-22>
- Jun, H. S. (2014). *A validity argument for the use of scores from a web-search-permitted and web-source-based integrated writing test*. Unpublished doctoral dissertation, Iowa State University.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Koizumi, R., Sakai, H., Ido, T., Ota, H., Hayama, M., Sato, M., & Nemoto, A. (2011). Toward validity argument for test interpretation and use based on scores of a diagnostic grammar test for Japanese learners of English. *Japanese Journal for Research on Testing*, 7, 99-119. https://doi.org/10.24690/jart.7.1_99
- 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. アルク選書.
- 小泉利恵 (編). (2022). 実例でわかる英語スピーチングテスト作成ガイド. 大修館書店.
- 小泉利恵 (2025). 言語テスト開発理論の現在. 全国英語教育学会50回研究大会記念特別誌, 302-305.
- 小泉利恵・阿川敏恵・高橋闇 (2023). 英語4技能テストのスコアレポートを用いた学習の有用性—大学生の反応から. 清泉女子大学 言語教育研究所言語教育研究, 15, 21-43.
- 小泉利恵・印南洋・深澤真 (編). (2017). 実例でわかる英語テスト作成ガイド. 大修館書店.
- 国立教育政策研究所 (2020a). 「指導と評価の一体化」のための学習評価に関する参考資料(小学校編). https://www.nier.go.jp/kaihatsu/pdf/hyouka/r020326_pri_gaikokg.pdf
- 国立教育政策研究所 (2020b). 「指導と評価の一体化」のための学習評価に関する参考資料(中学校編). https://www.nier.go.jp/kaihatsu/pdf/hyouka/r020326_mid_gaikokg.pdf
- 国立教育政策研究所 (2021). 「指導と評価の一体化」のための学習評価に関する参考資料(高等学校編). https://www.nier.go.jp/kaihatsu/pdf/hyouka/r030820_hig_gaikokugo.pdf
- 駒井健吾 (2024). 観点別評価は言語活動の評価デザインとして適切か. 全国英語教育学会 第49回福岡大会. 2024年8月25日.
- 久保佑輔 (2022). 妥当性検証の枠組みに基づく批判的思考テストの基準関連妥当性の調査—批判的思考力と態度の関係性—. 外国語教育メディア学会中部支部外国語教育基礎研究部会 2022年度報告論集.
- Kumazawa, T., Shizuka, T., & Mochizuki, M. (2016). Validity argument for the VELC Test® score interpretations and uses. *Language Testing in Asia*, 6(2). <https://doi.org/10.1186/s40468-015-0023-3>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 243-56. <https://doi.org/10.1177/026553229601300302>
- 文部科学省 (2017a). 小学校学習指導要領(平成29年告示)解説(外国語活動・外国語編). https://www.mext.go.jp/content/20220614-mxt_kyoiku02-100002607_11.pdf
- 文部科学省 (2017b). 中学校学習指導要領(平成29年告示)解説(外国語編). https://www.mext.go.jp/content/20210531-mxt_kyoiku01-100002608_010.pdf

引用文献

- 文部科学省 (2018). 高等学校学習指導要領(平成30年告示)解説(外国語編). https://www.mext.go.jp/content/1407073_09_1_2.pdf
- 村越亮治・江原美明 (2015). 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, 4, 1-18. https://doi.org/10.20686/academiakyou.4.0_1
- Park, M. (2021). Domain Definition Inference for a Virtual Interactive Aviation English Test (VIAET) for Military Air Traffic Controllers. In C. A. Chapelle & E. Voss (Eds.), *Validity Argument in Language Testing: Case Studies of Validation Research* (pp. 73-95). Cambridge University Press.
- 渡邊大志・大鋸雄介・佐藤扶・鶴篠憲一・斎田智里 (2025). 神奈川県公立高等学校入学試験英語問題の分析—観点別学習状況評価の観点から—. 教育デザイン研究, 16, 146-155. <https://ynu.repo.nii.ac.jp/record/2001651/files/16-1-17.pdf>
- 山本大貴・近藤暁子・鳴海智之・多田ウェンディ・吉田達弘 (2021). 日本の高等学校での活用を想定したペア型スピーキングテストの開発. 外国語教育メディア学会中部支部紀要, 58, 99-128. https://doi.org/10.24539/let.58.0_99

資料1：教師質問紙調査の指示文

高等学校英語科における定期試験・パフォーマンステストの評価に関するアンケート

B I U ↳ X

本研究は、高等学校英語科における定期試験およびパフォーマンステストに対する教師の認識を調査し、より妥当性のある学習評価を実施するための改善策を検証することを目的としています。この研究で得られた情報は研究者のみが所持し、あなたの個人情報が漏洩されることはありません。また、あなたの個人情報や個人のデータ等の開示を希望される際はそれに応じます。

以下に自身のメールアドレスを入力する欄があります。メールアドレスの回答は自身の回答の確認用です。で、「同意する」と回答していただいた方でもメールアドレスの情報は使用いたしません。

ご協力のほど、よろしくお願ひいたします。

資料2：教師質問紙調査の質問項目と調査結果

		指導経験(年)			M	SD
		0~5	6~10	11~		
Q1	高等学校での英語指導歴	—	—	—	—	—
Q2	令和6年度にご担当された学年(該当するものを選んでください:複数回答可)	—	—	—	—	—
Q3	「定期試験で測る英語力」について、学校や学年として、「単元の指導と評価の計画」等に明示されていると思いますか?	3.0	4.5	5.3	4.3	1.6
Q4	「パフォーマンステストで測る英語力」について、学校や学年として、「単元の指導と評価の計画」等に明示されていると思いますか?	3.0	4.0	5.3	4.1	1.5
Q5	あなたは観点別評価における「読むこと」の「知識・技能」・「思考・判断・表現」の評価規準の違いについて、どの程度理解していると感じますか?	4.0	4.0	3.3	3.8	0.9
Q6	あなたは観点別評価における「聞くこと」の「知識・技能」・「思考・判断・表現」の評価規準の違いについて、どの程度理解していると感じますか?	3.3	3.3	3.3	3.3	0.8
Q7	あなたは観点別評価における「書くこと」の「知識・技能」・「思考・判断・表現」の評価規準の違いについて、どの程度理解していると感じますか?	4.0	3.5	3.7	3.7	1.0
Q8	あなたは観点別評価における「話すこと」の「知識・技能」・「思考・判断・表現」の評価規準の違いについて、どの程度理解していると感じますか?	2.7	3.0	3.7	3.1	1.1

(資料2 続き).....

		指導経験(年)			M	SD
		0~5	6~10	11~		
Q9	ご自身の勤務校ではCan-doリストを設定していますか?	—	—	—	—	—
Q10	令和6年度にご自身がご担当された科目の「定期試験」は、学習指導要領の目標、Can-doリスト、教科書の内容、単元の指導と評価の計画等に基づいて適切に設計されていると思いますか?	3.3	3.8	4.0	3.7	1.5
Q11	令和6年度にご自身がご担当された科目の「パフォーマンステスト」は、学習指導要領の目標、Can-doリスト、教科書の内容、単元の指導と評価の計画等に基づいて適切に設計されていると思いますか?	3.0	4.3	4.3	3.9	1.5
Q12	令和6年度にご自身がご担当された科目の「定期試験」が、学習指導要領の目標を適切に反映していると思いますか?	3.7	3.5	4.0	3.7	1.1
Q13	(Can-doリストを設定していると回答した場合)令和6年度にご自身がご担当された科目の「定期試験」が、Can-doリストを適切に反映していると思いますか?	2.5	3.5	5.0	3.5	1.5
Q14	令和6年度にご自身がご担当された科目の「定期試験」が、単元の指導と評価の計画を適切に反映していると思いますか?	3.3	3.5	4.0	3.6	1.0
Q15	令和6年度にご自身がご担当された科目の「定期試験」が、教科書の内容を適切に反映していると思いますか?	4.7	4.5	4.7	4.6	0.7
Q16	令和6年度にご自身がご担当された科目の「パフォーマンステスト」が、学習指導要領の目標を適切に反映していると思いますか?	4.3	4.0	4.7	4.3	1.0
Q17	(Can-doリストを設定していると回答した場合)令和6年度にご自身がご担当された科目の「パフォーマンステスト」が、Can-doリストを適切に反映していると思いますか?	3.5	3.8	5.0	3.7	1.7
Q18	令和6年度にご自身がご担当された科目の「パフォーマンステスト」が、単元の指導と評価の計画を適切に反映していると思いますか?	3.3	3.8	4.7	3.9	1.2
Q19	令和6年度にご自身がご担当された科目の「パフォーマンステスト」が、教科書の内容を適切に反映していると思いますか?	4.7	4.5	5.0	4.7	1.2
Q20	定期試験の設計プロセスにおいて、どのような資料を参照していますか?	—	—	—	—	—
Q21	パフォーマンステストの設計プロセスにおいて、どのような資料を参照していますか?	—	—	—	—	—
Q22	定期試験の問題は、授業で扱う内容を反映していると思いますか?	5.0	5.0	5.0	5.0	1.1
Q23	パフォーマンステストの内容は、授業で行う活動と一致していると思いますか?	5.3	4.5	4.3	5.0	1.2
Q24	定期試験を作成・実施する際に難しいと感じる点について、自由にご意見をお聞かせください。	—	—	—	—	—
Q25	パフォーマンステストを作成する際に難しいと感じる点について、自由にご意見をお聞かせください。	—	—	—	—	—
Q26	定期試験やパフォーマンステストを作成・実施する際に工夫している点(例:生徒の学習につながるように意図している点)について、ご記入ください。	—	—	—	—	—
Q27	評価を実施する際に、「この点は適切にできているのか?」と疑問を感じることがあればご記入ください。	—	—	—	—	—

資料3：送付した単元の指導と評価の計画(一部)

教科	外国語科	科目	英語コミュニケーションⅢ	単位数	4
Lesson 1	[REDACTED]				
	内容のまとめごとの評価規準			主な評価方法	
	知識・技能	思考・判断・表現	主体的に学習に取り組む態度		
・読みこと	[知識] 前置詞を伴う関係代名詞、動名詞の意味上の主語、受け身の完了形、倒置、付帯状況の with を理解している。 [技能] 教科書本文の音声を聞いたり、本文を読んだりして、本文中の前置詞を伴う関係代名詞、動名詞の意味上の主語、受け身の完了形、倒置、付帯状況の with が使われた英文の意味を把握する技能を身につけている	錯視が日常生活に使われている方法について理解できる。	錯視が日常生活に使われている方法について自ら理解しようとしている。	ペーパーテスト 振り返りの記述 活動の観察	
(やり取り)	[知識] [技能]	理解した内容を自分の言葉で表現し、相手に伝えることができる。また、錯視が日常生活に使われている方法について自分なりに意見を持ち、それを表現することができる。	本文から学んだ錯視が日常生活に使われている方法について自分の意見を持ち、積極的に発信しようとしている。		
(発表)	[知識] [技能] 前置詞を伴う関係代名詞、動名詞の意味上の主語、受け身の完了形、倒置、付帯状況の with を用いて一つのテーマについて書き、それを発表する技能を身につけている。	理解した内容を自分の言葉で表現し、相手に伝えることができる。また、錯視が日常生活に使われている方法について自分なりに意見を持ち、それを表現することができる。	本文から学んだ錯視が日常生活に使われている方法について自分の意見を持ち、積極的に発信しようとしている。	個人発表 振り返りの記述 活動の観察	
書くこと	[知識] [技能] 前置詞を伴う関係代名詞、動名詞の意味上の主語、受け身の完了形、倒置、付帯状況の with を用いて一つのテーマについて書き、それを発表する技能を身につけている。	理解した内容を自分の言葉で表現し、相手に伝えることができる。また、錯視が日常生活に使われている方法について自分なりに意見を持ち、それを表現することができる。	本文から学んだ錯視が日常生活に使われている方法について自分の意見を持ち、積極的に発信しようとしている。	ワークシート 振り返りの記述 活動の観察	

記録に残す評価 (○)

時間	主な学習活動	知	思	主
1 6	<ul style="list-style-type: none"> 単元の目標や進め方を確認し、学習の見通しをもつ。 提示された質問を基に自分について考える。 錯視が日常生活に使われている方法について聞いたり、読んだりし、理解する。 教科書の内容や自分の考えを自分の言葉で他に伝えたり、書いたりする。 本文で使われている前置詞を伴う関係代名詞、動名詞の意味上の主語、受け身の完了形、倒置、付帯状況の with の使い方を理解し、身につける。 教科書の内容を応用し、自分の考えを積極的に自分の言葉で他に伝えたり、書いたりしようとする。 本文で使われている前置詞を伴う関係代名詞、動名詞の意味上の主語、受け身の完了形、倒置、付帯状況の with の使い方を理解し、積極的に応用場面で使おうとする。 上記の振り返りを Booklet に記入する。 	○	○	○
後日	<ul style="list-style-type: none"> ペーパーテスト 初めて聞いたり読んだりする英文からも出題されます。 	○	○	

活動状況を見取り、学習や指導に生かすための評価は毎時間行います。

資料4：送付したCan-doリスト

目標		1年	2年	3年
READING	英語で書かれた様々な文章に触れ、多様な考え方・ものの見方を理解し、自分の考えを持つことができる。 (教科書でも多様なジャンルを扱う。共創研究で論文にも触れさせたい)	ジャンルを意識しリーディングスキルを使い分けることができる。パラグラフ構成について理解して筆者の主張を掴むことができる。WPMは100を目指す。(共通テストで余裕があるレベル)	ディスコースマーカーや結束性(cohesion)、一貫性(coherence)などを意識して、パラグラフのつながりや文章全体を理解できる。WPMは130を目指す。(英語テストで余裕があるレベル)	論文なども含め、Authenticな英文を読むことができる。WPMは150を目指す。(TOEICですかべて読み終わるレベル)
WRITING	コミュニケーションの相手やジャンルに応じた英文を書くことに抵抗を感じない。 (Fluency, Global/Local Error) 基本的な型(理由の述べ方、パラグラフライティング)を理解しそれに従って英文を書くことができる。ジャンルや読み手を意識してライティングをすることができる 100語程度の文量を書くことができる。	英文の細かい部分まで意識できる。 (Accuracy, Local Errorにも注意) 内容の深化を図る。(論理的に考えて文章が構成されているか) 200語程度書くことができる。	genre(手紙やメールなど)や receiver(友達や上司など)に応じて適切な表現や幅広い表現(無生物主義、分詞構文など)を用いて英文を書くことができる。 20分で200語程度(分量最大 IELTS の 250語)の英文を書くことができる。	genre(手紙やメールなど)や receiver(友達や上司など)に応じて適切な表現や幅広い表現(無生物主義、分詞構文など)を用いて英文を書くことができる。 20分で200語程度(分量最大 IELTS の 250語)の英文を書くことができる。
LISTENING	様々な英語に触れ、要点などを押さえることができる。 (海外の研究発表を開けるように)	身近なテーマ(自己紹介など)や最新の話題について聞き取り、理解ができる。 話される内容をメモしながら聞くことができる。 実際の英語を聞き取ることができる。	社会的なテーマ(英語での専門的な授業など)や最新の話題について聞き取り、理解ができる。1分程度の会話を聞き取ることができる。 実際の英語を聞き取ることができる。	社会的なテーマ(英語での専門的な授業など)や最新の話題について聞き取り、理解ができる。また、複数の話者や雑音が入っているものも聞き取ることができる。
SPEAKING (INTERACTION)	英語を話すことに対して抵抗感を感じず、積極的に他者とコミュニケーションを図ることができる。聞き手に応じた適切な表現を用いて、自分の考えを伝えたり、質問したりできる。 (海外の研究発表の質疑応答)	英語を話すことに抵抗を感じない。(Fluency) Filler や適切な表現を使いながら会話をつなげることができる。 自分の周りのことについて1分間話すことができる。(CEFR A2 レベル) 質問文を作ることができる。(CEFR B1 レベル)	発話の細かい部分まで意識できる。Errorを減らしていく。(Accuracy) 聞き手からの質問など(+1)があつても自然な回答や会話を続けることができる。 自分の周りの社会のことについて1分間話すことができる。(CEFR B1 レベル)	register や genre, context を意識して会話を続けることができる。(Accuracy) 自然な流れで会話の目的を達成することができます。様々な内容について1分間話すことができる。(CEFR B2 レベル)
SPEAKING (PRESENTATION)	聞き手に応じた適切な表現やアイコンタクトなどの工夫を生かして、自分の考えを伝えることができる。 (ノンハーバルとハーバルを育成する)	聞き手を意識して発表をすることができる。ノンハーバルの観点の定着。	聞き手が理解しやすい発表をすることができる。ノン定型句などを用いての発表ができる。ハーバルの視点の定着。	聞き手を意識して理解しやすい発表をすることができる。グラフの説明やデータを用いた研究を英語で行うことができる。
意識している内容・活動	・アカデミックな点での資質・能力の育成 →海外での発表に耐えうる資質・能力の育成 例) 論文の講解、発表や質問を聞く 論文を書く、発表や質問に対する応答	R: 多様なジャンルの説明 W: パラグラフ、ジャンル・読み手の意識 L: メモの取り方、音の特徴(脱落など)に慣れる S: 聞き手の意識、ストラテジー、ノンハーバル	R: まとまった分量・論文(一部)の説明 W: エッセイライティング L: 様々な状況のリスニングに慣れる S: 聞き手の意識、ストラテジー、ハーバル	R: 多様なジャンルを大量に読む、論文の読み取り W: エッセイライティング L: まとった分量の聞き取り S: グラフなども含めてハーバル、質問を作る

資料5：生徒質問紙調査の指示文

高等学校英語科における定期試験に関するアンケート

B I U ↵ X

このアンケートは、みなさんが英語科の定期試験についてどのように感じているかを調査するものです。みなさんの声を集めて、これからの英語の学習やテストが、さらに良くなるように役立てたいと考えています。

このアンケートで集めた情報は、研究目的でのみ使用します。公表に際しては情報提供者のプライバシーが侵害されないように最大限の注意を払い、個人名が特定されないように配慮いたします。

また、このアンケートに答えたことで成績などに影響が出ることもありません。安心して、自分の考えを書いてください。

アンケートの内容について、知りたいことや確認したいことがあれば、英語科：大鶴（おおが）まで声をかけてください。

ご協力のほど、よろしくお願ひいたします。

資料6：生徒質問紙調査の質問項目及び調査結果

		1年		2年		3年		全体	
		M	SD	M	SD	M	SD	M	SD
Q1	英語の学習を始めたのはいつですか？	—	—	—	—	—	—	—	—
Q2	英語圏の国に留学または滞在したことがありますか？	—	—	—	—	—	—	—	—
Q3	(「はい」の場合)どのくらいの期間滞在しましたか？	—	—	—	—	—	—	—	—
Q4	英語コミュニケーションの授業では、主にどのように活動していますか？(複数回答可)	—	—	—	—	—	—	—	—
Q5	英語コミュニケーションの授業では、先生が話し、生徒が聞く時間がが多い。	3.1	1.1	3.8	1.3	3.8	1.2	3.5	1.2
Q6	英語コミュニケーションの授業では、先生は主に英語を話す。	3.9	1.2	3.9	1.1	4.1	1.1	3.9	1.2
Q7	英語コミュニケーションの授業では、先生と話すとき、生徒はほとんど英語を使う。	2.9	1.1	3.0	1.2	2.7	1.1	2.9	1.1
Q8	英語コミュニケーションの授業中、生徒同士で話すとき、ほとんど英語を使う。	3.4	1.2	3.3	1.2	2.8	1.1	3.2	1.2
Q9	英語コミュニケーションの授業のスピーキング活動では、即興で英語を話すことが求められることが多い。	3.9	1.2	3.6	1.2	3.6	1.2	3.7	1.2
Q10	英語コミュニケーションの授業の活動は主に何に焦点を当てていますか？(複数回答可)	—	—	—	—	—	—	—	—
Q11	英語コミュニケーションの授業では何に多くの時間を使いますか？(複数回答可)	—	—	—	—	—	—	—	—
Q12	英語コミュニケーションの授業では定期試験対策の学習を多く行っている。	3.7	1.2	3.3	1.1	3.3	1.3	3.5	1.2
Q13	1週間にどのくらいの時間、英語の宿題をしますか？	—	—	—	—	—	—	—	—
Q14	学校の宿題の内容は主に何ですか？(複数回答可)	—	—	—	—	—	—	—	—
Q15	高校の英語コミュニケーションの授業に満足している。	4.5	1.1	3.8	1.3	3.8	1.3	4.1	1.3
Q16	高校の英語コミュニケーションの授業によって、どのスキルが向上したと思いますか？(複数回答可)	—	—	—	—	—	—	—	—
Q17	高校の英語コミュニケーションの授業について、他にコメントがありますか？	—	—	—	—	—	—	—	—
Q18	今回の定期試験は楽しかった。	2.8	1.5	2.6	1.3	2.8	1.4	2.8	1.4
Q19	今回の定期試験は難しかった。	3.9	1.3	4.3	1.1	4.1	1.1	4.1	1.2
Q20	今回の定期試験の指示は明確だった。	4.5	1.2	4.5	1.1	4.3	1.3	4.4	1.2
Q21	今回の定期試験の試験時間は十分だった。	4.0	1.5	3.5	1.4	3.1	1.4	3.6	1.5
Q22	今回の試験時間のために学校外でのテスト準備に何時間費やしましたか？	—	—	—	—	—	—	—	—
Q23	今回の定期試験の詳細について知ったとき、どのように試験の準備をしましたか？	—	—	—	—	—	—	—	—
Q24	今回の定期試験のために学校外では主に何を中心に学習しましたか？(複数回答可)	—	—	—	—	—	—	—	—
Q25	学校の授業は今回の定期試験の準備に役立った。	4.4	1.1	4.0	1.1	3.8	1.3	4.1	1.2
Q26	今回の定期試験の問題は学校の授業で行った内容と似ていた。	4.6	1.1	4.2	1.1	3.9	1.2	4.3	1.2
Q27	今回のテストは、定期試験として適している。	4.8	1.0	4.4	1.1	4.4	1.1	4.6	1.1
Q28	今回の定期試験後、自信がついたものはどれですか？(複数回答可)	—	—	—	—	—	—	—	—
Q29	今回の定期試験後、さらに練習したいと思ったものはどれですか？(複数回答可)	—	—	—	—	—	—	—	—
Q30	スコアレポート(得点個表)は自身の強みや弱点を理解するのに役立つと思う。	5.0	1.0	4.4	1.2	4.4	1.3	4.6	1.2
Q31	スコアレポート(得点個表)は明確である。	5.0	1.0	4.5	1.0	4.4	1.2	4.6	1.1
Q32	スコアレポート(得点個表)は解釈しやすい。	4.9	1.1	4.5	1.0	4.3	1.2	4.6	1.1
Q33	授業で受けた指導は、今回の定期試験を受けるのに十分だったと思う。	4.7	1.1	4.2	1.1	4.0	1.2	4.4	1.2
Q34	そう思う理由を書いてください。	—	—	—	—	—	—	—	—

資料7：生徒質問紙調査の選択式回答項目の選択肢及び回答結果

			回答人数 (%)
Q1	英語の学習を始めたのはいつですか？	0～6歳	132(14.3%)
		7～12歳	567(61.4%)
		13～16歳	225(24.4%)
Q2	英語圏の国に留学または滞在したことがありますか？	はい	107(11.6%)
		いいえ	817(88.4%)
Q3	（「はい」の場合）どのくらいの期間滞在しましたか？	1ヶ月未満	95(10.3%)
		1～12ヶ月	12(1.3%)
		12ヶ月以上	5(0.5%)
Q4	英語コミュニケーションの授業では主にどのように活動していますか？ (複数回答可)	個人	611(66.1%)
		グループ	380(41.1%)
		ペア	787(85.2%)
		クラス全体	207(22.4%)
Q10	英語コミュニケーションの授業の活動は主に何に焦点を当てていますか？ (複数回答可)	語彙	475(51.4%)
		文法	414(44.8%)
		発音	405(43.8%)
		リスニング	541(58.6%)
		リーディング	526(56.9%)
		スピーチング	544(58.9%)
		ライティング	234(25.3%)
Q11	英語コミュニケーションの授業では何に多くの時間を使いますか？ (複数回答可)	語彙	361(39.1%)
		文法	346(37.4%)
		発音	311(33.7%)
		テキストを読んで質問に答える	592(64.1%)
		モノologueを聞いて質問に答える	75(8.1%)
		ダイアログを聞いて質問に答える	64(6.9%)
		メールを書く	12(1.3%)
		記事を書く	38(4.1%)
		物語を書く	12(1.3%)
		エッセイを書く	45(4.9%)
		レビューを書く	20(2.2%)
		ペアで身近な話題について質問・回答する	457(49.5%)
		一人で絵を説明する	42(4.5%)
		ペア・グループで絵について話し合う	128(13.9%)
		ペア・グループでアイデアやトピックを議論する	259(28.0%)
		ペア・グループで意思決定を行う	194(21.0%)
Q13	1週間にどのくらいの時間、英語の宿題をしますか？	1時間未満	475(51.4%)
		1～2時間	271(29.3%)
		2～3時間	104(11.3%)
		4時間以上	74(8.0%)

(資料7 続き).....

			回答人数 (%)
Q14	学校の宿題の内容は主に何ですか?(複数回答可)	語彙	417(45.1%)
		文法	273(29.5%)
		発音	110(11.9%)
		リスニング	182(19.7%)
		リーディング	278(30.1%)
		スピーキング	239(25.9%)
		ライティング	226(24.5%)
Q16	高校の英語コミュニケーションの授業によって、どのスキルが向上したと思いますか? (複数回答可)	リスニング	284(45.6%)
		リーディング	421(42.9%)
		ライティング	190(20.6%)
		スピーキング	396(30.7%)
Q22	今回の試験時間のために学校外でのテスト準備に何時間費やしましたか?	0時間	135(14.6%)
		1~5時間	636(68.8%)
		6~10時間	120(13.0%)
		11時間以上	33(3.6%)
Q24	今回の定期試験のために学校外では主に何を中心に行いましたか? (複数回答可)	語彙	549(59.4%)
		文法	384(41.6%)
		発音	40(4.3%)
		リスニング	81(8.8%)
		リーディング	459(49.7%)
		スピーキング	56(6.1%)
		ライティング	102(11.0%)
		ない	74(8.0%)
Q28	今回の定期試験後、自信がついたものはどれですか? (複数回答可)	語彙	231(25.0%)
		文法	224(24.2%)
		発音	32(3.5%)
		リスニング	71(7.7%)
		リーディング	269(29.1%)
		スピーキング	27(2.9%)
		ライティング	37(4.0%)
		ない	291(31.5%)
Q29	今回の定期試験後、さらに練習したいと思ったものはどれですか? (複数回答可)	語彙	533(57.7%)
		文法	417(45.1%)
		発音	108(11.7%)
		リスニング	244(26.4%)
		リーディング	536(58.0%)
		スピーキング	128(13.9%)
		ライティング	167(18.1%)
		ない	39(4.2%)