

## 高校現場における「英語テスト」と「観点別評価」の課題

Problems and Challenges with Criterion-Referenced Testing in High School English Courses

村越亮治・江原美明

MURAKOSHI, R. and EHARA, Y.

### 1. 研究の背景

高校現場での教科指導に観点別評価が導入されてから数年がたっている。「児童生徒の学習評価の在り方について(報告)」(文部科学省, 2010a)では、高等学校における観点別評価は小・中学校に比べてあまり定着していないとされていたが、筆者らが現場の教師に聞いてみると、その理解と適用の状況は、その後も学校によりまちまちであると感じられる。平成 25 年度入学生から新学習指導要領が適用され、外国語(英語)については、「授業は英語で行うことを基本とする」という箇所がかなり注目を集めたが、それに先立つ、平成 24 年 7 月に、「評価規準の作成、評価方法等の工夫改善のための参考資料(高等学校外国語)」(国立教育政策研究所, 2012)によって、すでに新指導要領下の科目における観点別評価の具体が示されていた。さらに、都道府県単位で評価に関する研修会が開かれ、現場教師への理解の徹底が図られてきた。しかしながら、指導と一体化した観点別評価は、実効的に十分に行われるようになったのだろうか。

平成 25 年 3 月には、「各中・高等学校の外国語教育における『CAN-DO リスト』の形での学習到達目標設定のための手引き」(文部科学省, 2013)が出され、目標に準拠した 4 技能の指導と評価の必要性が、あらためて高校現場に示された。各高等学校は学校独自の学習到達目標の提出を求められているが、観点別評価と一体化した指導が実現していない場合、さらに 4 技能別に細分化された学習到達目標を実効性のあるものにするには困難だろう。

学習到達目標の達成度を確認し、形成的な指導とつながった観点別評価を可能にするためには、適切なテストを実施する必要がある。しかし、高校現場ではどれくらい妥当性のあるテストが行われているだろうか。神奈川県立国際言語文化アカデミアにおいて実施した、英語テストに関わる研修講座を受講した中学校・高等学校の英語教師のアンケートには、「テストに関して何も知識がなかったのでよかった」「テストについて初めて講義を受け、意識すべきことがわかった」「(妥当性やテストポイントなどを)あまり気にせずテストを作っていたのでとても参考になった」などのコメントが多く見られた。これは、多くの現場の教師が、十分なテストの知識・技能を持たないまま、自己流で、または慣例的にテストを作成し、実施してきたことを表しているかもしれない。そうだとすると、果たしてそのようにして行われたテストが、学習到達目標の達成度を判断し、次の指導につながる観点別評価に寄与するものになっているだろうか。

本研究は、高校現場で実施されている英語の定期テスト(中間テストや期末テスト)の作成プロセスと、テストの妥当性(内容的妥当性および構成概念妥当性)を調べることで、よりよいテストに支えられた、よ

り実効性のある観点別評価と学習到達目標に基づいた指導を実現させるための示唆を得ることを目的にするものである。

## 2. 文献研究

### 2.1 「観点別評価」導入の経緯と現状

「観点別評価」は、学習指導要領の趣旨に沿った指導の充実と評価方法の改善を目的とし、旧学習指導要領時代に本格的に導入された。従来の知識偏重の学力観ではなく、学校教育法第 30 条(同 49 条、62 条)に掲げられた、基本的な知識・技能、思考力・判断力・表現力、自ら学ぶ意欲を含む新しい学力観に基づく指導と評価を目指したものである。それに先立ち、平成 12 年 12 月の教育課程審議会答申「児童生徒の学習と教育課程の実施状況の評価の在り方について」は、目標に準拠した評価(いわゆる絶対評価)のいっそうの重視や、生徒一人ひとりの学習プロセスを評価する個人内評価の工夫などについて、学校が一体となって評価活動の充実に取り組むことを求めた。文部科学省所管の研究機関である国立教育政策研究所が、観点別評価のための参考資料(国立教育政策研究所, 2002, 2004)を公表すると、それに基づき全国の教育委員会や教員の研究組織による研修会が開かれ、学習指導要領の内容に照らした観点別評価の在り方や各観点の定義等について、多くの議論がなされた。

10 年の時を経て、現行学習指導要領(中学校は平成 24 年度から全面実施、高等学校は平成 25 年度から年次進行で実施)でも観点別評価の方向性は維持されることとなり、同時に、旧学習指導要領における観点別評価の現状と課題について調査が行われた。「児童生徒の学習評価の在り方について(報告)」(文部科学省, 2010a)によると、(a) 観点別評価は小・中学校においては定着してきているが、教師の側に負担感があること、(b) 高等学校では小・中学校ほど観点別学習状況の評価が定着していないこと、が課題として挙げられた。この間、神奈川県では、平成 19 年度からすべての県立高等学校で通知表等に観点別評価を記載することを含め、目標に準拠した指導・評価の改善への取組をしてきたことが、同報告でも言及されている。しかし、課題として指摘された教師側の負担感は大きいと考えられ、本来指導の充実のために導入された観点別評価の趣旨が、教師が心から納得のいく形で実践に活かされているか否かについては、現状の把握を踏まえたうえで現実的な議論・提案を行う必要がある。

### 2.2 学習指導要領が求める「観点別評価」

現行学習指導要領における高等学校外国語科の目標は、「外国語を通じて、言語や文化に対する理解を深め、積極的にコミュニケーションを図ろうとする態度の育成を図り、情報や考えなどを的確に理解したり適切に伝えたりするコミュニケーション能力を養う」ことである。外国語科における観点別評価は、この目標に含まれる要素である「コミュニケーションへの関心・意欲・態度」、「外国語表現の能力」、「外国語理解の能力」、「言語や文化についての知識・理解」の4つの観点から、指導内容(「聞くこと」「読むこと」「話すこと」「書くこと」)についての目標達成度を評価し、授業改善に活かそうというものであり、その趣旨および留意事項は、「評価規準の作成、評価方法等の工夫改善のための参考資料(高等学校外国語)」(国立教育政策研究所, 2012)に記載されている。

観点別評価には、テスト等の記録に残す評価だけではなく、授業中の活動の観察や生徒への適時のフィードバックなど、形成的評価の側面も含まれる。今後ますます求められる CAN-DO リスト形式の到達目標に基づく指導に向けては、こうしたプロセス評価と生徒へのフィードバックの充実がより求められるだろう。現在、CAN-DO リストに関し国が求めているのは、観点別評価項目のうち、特に技能面に関わる観点である「外国語表現の能力(話す・書く)」と「外国語理解の能力(聞く・読む)」について、具体的な到達目標を設定し、普通の授業にその目標につながる言語活動を取り入れることであるが(文部科学省, 2013)、そのためには観点別評価の残りの2観点に関わるアプローチ、すなわち「コミュニケーションへの関心・意欲・態度」の観点に関わる情意面の指導と「言語や文化についての知識・理解」に関わる発音・語彙・文法・社会言語的知識に関する指導と評価も重要な要素である。

学習指導要領解説(文部科学省, 2010b)には、改善の基本方針として「4技能の総合的な指導を通して、これらの4技能を統合的に活用できるコミュニケーション能力を育成する」(p. 4)とある。実際の言語使用においてもそうであるように、指導においても複数の技能を駆使してタスクを完成させる活動を行うことが求められている。一方、技能統合型タスク(例えば、講義を聞いてその内容について意見交換をしたり感想を書いたりするなど)を用いて評価をすることについては、「技能統合評価の場合には、定期試験などの到達度テストにおける誤りの原因特定は容易ではない」と根岸(2011)が指摘したように、評価の結果を指導や生徒の学びの改善にフィードバックするという点では扱いにくい面もある。指導過程では統合的に4技能を使う活動を取り入れながら、その成果の評価においては、4技能、情意面、知識面のそれぞれに焦点を当てて分析的に評価するのが観点別評価の考え方である。英語力をタスクの達成度で評価するアプローチも視野に入れながら、英語力ないしは英語によるコミュニケーション能力を分析的に評価することにより、指導にあたる教師自身が指導内容の細目をより明確に意識化することができると思われる。

### 2.3 テスティングに関する文献からの示唆

日本の中学校・高等学校英語教員が入手可能なテスティング関連文献(主として書籍)からは、定期テスト作成に関し多くの示唆を得ることができる。言語テスティングに関する参考書籍には、テスト理論中心のもの、英語教育評価全般に関するもの、テスト作成の方法論に関するものなどがある。

テスト理論に関する基本文献としては、Bachman (1990)がある。Bachman は、コミュニケーション言語能力(Communicative Language Ability)を、「一般知識」「言語知識」「方略的能力」「心理・生理的機能」「コミュニケーションの状況」の5つの要素の相互作用ととらえ、言語使用者は、状況を感じ取ったうえでの判断に基づき、方略的能力を使うことにより一般知識や言語知識を動員してコミュニケーションを行うとした。この考え方は Bachman and Palmer (1996, 2010)にも継承され、言語能力とは言語知識と方略的能力の相互作用であり、言語知識(language knowledge)は、「文法知識」「テキスト(談話)についての知識」「言語機能についての知識」「社会言語学的知識」から構成されるとした。

英語教育評価全般に関する文献に Brown & Abeywickrama (2010) および 松沢(2002)がある。前者はテストを含め、代替評価法(alternative assessment)など幅広い評価法について論じており、評価活動(assessment)を指導(teaching)の重要な一部と位置づけている。後者は、外国語教育にお

いて生徒の学習到達度を判断するための到達度評価の研究が遅れている点を指摘し、具体的改善策として、タスク準拠評価、目標基準準拠評価、継続的評価を提案した。松沢はまた、評価の改善の手法として、「テスト実施報告書」(post-test report)の作成やアクション・リサーチを用いた課題解決法を挙げ、学習者からの聞き取り調査によるテスト問題の改善、記述解答の採点方法の改善(仮採点と採点計画の修正を経て本採点の実施をするなど)による採点者間信頼性の向上などの実践例を紹介した。

テスト作成の方法論に関する文献からは、これまでに議論されたテスト作成に関わる重要考慮事項を確認することができる。Hughes (1989)は、第1章の冒頭でテストが学習者の学習に対し及ぼす影響である波及効果(backwash)について言及し、プラスの波及効果を持ったテスト作成の重要性を強調した。また測定すべき能力を本当に測っているかという妥当性(validity)の問題や、繰り返し行っても安定した測定結果が得られるかという信頼性(reliability)の問題、テストの目的や内容を詳細に記述したテスト明細書(test specifications)作成の重要性について示唆を与えている。Brown (2005)は、客観テスト作成における留意点について、「答えが一つに定まるか」「質問文は生徒のレベルに合った英文で書かれているか」「錯乱肢(distracters)はどれも答えとしてあり得そうなものか(常識的にあり得ない選択肢はないか)」「少なくとも一人の同僚教師に問題を確認してもらったか」などのアドバイスを、表計算ソフトを用いた基本統計量(平均値、最頻値、中央値、標準偏差等)の記述方法、信頼性係数の計算方法等を解説している。

日本の中学校・高等学校における定期テストの問題を指摘した若林・根岸(1993)は、「学校教育におけるテストの第一義は、『教師の授業活動の評価』のための資料収集である」と述べた。教師自身の指導を自己評価するためにテストがあり、指導項目に対応してテストポイントを確認にしたテストづくりをすべきだとした。テストポイントが混在したいわゆる「総合問題」は問題形式として排除することを提案するとともに、聞き取りテストにおいては聞く力以外の力(読んだり書いたり翻訳したりする力)の影響を受けない作問に留意することや、読むことのテストではパラレルテキスト(既習の語句・文法・文体・文章の種類が同じテキスト)を用いることを提案した。こうした主張は、観点別評価に対応した中学校におけるテスト作成の具体例を集録した、根岸・東京都中学校英語教育研究会(2007)に反映されている。

テストポイントを考える視点について、静(2002)は「構成概念」(construct)に焦点を当てて解説した。彼は、構成概念を「『英語力』とか『計算力』とか『文法力』、『語彙力』、『発音の正確さ』等、直接観察はできないが理論上仮定できる人間の持つ特性」(p.5)と説明し、テスト後のフィードバックといった実用的見地、同じ物差しで測定可能かといった次元性(unidimensionality)の見地から、リーディング力、ライティング力、リスニング力、スピーキング力の4技能を構成概念とみなすことで「現実的には必要十分である」(p.15)とした。そのうえで、定期テストにおいては特に、授業内容(ティーチングポイント)がテストの内容(テストポイント)に反映されているかどうかに関わる「内容的妥当性」(content validity)が最も重要であると指摘した。テストは指導の重要な一部であり、テスト形式についても、普段の学習活動を学習者がその問題形式で行うことで英語力が伸びるかどうかを問うことで、その教育的有用性が判断できるとした。さらに、英文和訳・和文要約問題は、英語での表現力育成にとり負の波及効果を及ぼすものとして排除することを提案した。

テストに関する文献からは、このように、コミュニケーション能力の構成要素(構成概念)、波及効果、目標に準拠したテスト作成、教育活動の一環としてのテストの意義について示唆を得ることができる。テスト問題は、指導法と同様、あらゆる状況に対応する万能のものはないのかもしれないが、教えた内容を評価しているかに関わる内容的妥当性や、テスト問題が与える生徒の学習への波及効果は、あらゆる状況に共通して考慮すべき重要点だと考えられる。

#### 2.4 定期テストにおける構成概念妥当性、内容的妥当性、波及効果

観点別評価は、学習指導要領で示された指導内容(「聞くこと」「読むこと」「話すこと」「書くこと」)の指導の成果を、(1) 技能としての「外国語表現の能力(話す・書く)」、・「外国語理解の能力(聞く・読む)」、(2) 技能を活用する原動力となる情意面に関する「コミュニケーションへの関心・意欲・態度」、(3) 技能を支える言語リソースの中味である「言語や文化についての知識・理解」という観点から指導と評価を考えようとするものである。

学習指導要領の示す最終的なゴールは4技能の習得であるから、テスト(紙ベースの定期テストやパフォーマンステスト)では4技能を評価すればよいという考え方には表面的妥当性はあるかもしれない。一方、授業で扱った内容をテストで評価するという内容的妥当性の視点からは、授業で指導した音声、語彙、文法、社会言語知識等についての評価する必要もあるし、コミュニケーション活動に積極的に取り組む姿勢が育成されたかを授業中の言語活動やパフォーマンス評価で見ること必要だろう。また、授業で教えた内容に関する出題をする際(例えば語彙の問題)、その授業内容自体がそもそも4技能の能力を高めることを意識しているかといった点も再考する必要がある。定期テストでの出題形式が実際のコミュニケーションの場面を想定させるようなものであれば、コミュニケーション能力育成にとりプラスの波及効果を期待することもできる。

観点別評価の枠組は、情意面、技能面、知識面という次元に4技能という次元が重なり合い、4技能のそれぞれに情意面と知識面の観点が結びついているため、一次元性を要求する構成概念を考えるうえでは複雑である。しかし、例えば前項(2.3)で言及した Bachman and Palmer (1996, 2010)が、言語能力を言語知識と方略的能力の相互作用と捉えたように、コミュニケーション能力を、言語知識とその活用の工夫(方略的能力)に支えられた4技能の表出ととらえるならば、観点別評価の枠組と言語テスト理論との関連性も見えてくる。言語能力というきわめて複雑な能力の指導と評価にあたっては、その構成要素の理解が教師の側に不可欠であり、定期テストにおける構成概念妥当性、内容的妥当性、波及効果の検証の意義もそこにあるといえる。

### 3. 研究設問

高校現場で実施される英語テストと、それに裏打ちされるべき指導に関する課題を明らかにするために、次の研究設問を立てた。

- ・現場で実施されている英語の定期テスト問題に、どれくらいの妥当性(内容的妥当性、構成概念妥当性)があるか。

- ・より教育的意義のある定期テストの作成、実施、採点にあたり、現場の指導体制に課題があるとすれば、それはどのようなことか？

## 4. 研究方法

### 4.1 アンケート調査

神奈川県における高等学校中核英語教員研修として位置づけられている「英語教育アドヴァンスト研修」の平成 23 年度から平成 26 年度までの受講者および、国際言語文化アカデミアのその他の研修講座受講者で担当者と email によるやりとりを行ったことのある受講者に対し、紙媒体もしくはメールでのアンケート調査を依頼したところ、46 名から回答を得た。質問項目は次の 8 項目であった。

- ① 勤務校の生徒の英語力(回答者の判断で「基礎」「標準」「発展」の 3 レベルから選択)
- ② 定期テストの形式(配点等)の決定方法
- ③ 観点別評価のラベリングの方法
- ④ 読解問題の内容とテキストの出典
- ⑤ リスニング問題の有無
- ⑥ ライティング問題の有無
- ⑦ ライティング問題の評価基準
- ⑧ 定期テストの作成、実施、採点に関わる現場教師の悩み

このうち、①～⑥は多肢選択式(「その他」の選択肢は記述回答欄付き)、⑦⑧は記述回答形式であった。集計後、まず各項目の回答状況を分析し、その後、回答者が判断した勤務校の生徒の英語力別に各回答の傾向を探った。「定期テストの作成、実施、採点に関わる現場教師の悩み」についての記述回答については、コーディングによる分類と解釈を繰り返し、抽出される課題を整理した。

### 4.2 定期テストの分析

実際に高等学校で実施された英語の定期テスト 21 件を集め、2 つの視点で分析を行った。データの内訳は、現行学習指導要領下の科目で、特に 4 技能の総合的・統一的指導が求められている、「コミュニケーション英語 I・II」16 件に加えて、「総合英語」1 件、旧学習指導要領下の「英語 I・II」4 件である。同一校・同一科目のものであっても、出題の諸相が異なれば個々に分析対象としている。

まず、それぞれの観点別の配点を調べ、各テスト内のバランスやテスト間の差異を確認した。次に、大問や各問題に付されている 4 観点のラベリング(「コミュニケーションへの関心・意欲・態度」「外国語表現の能力」「外国語理解の能力」「言語や文化についての知識・理解)」について、問題が測定している知識、技能と合致しているかどうかを調べた。

## 5. 結果と考察

### 5.1 アンケート調査結果

#### 5.1.1 テスト形式および観点別ラベリングの方法

定期テストの形式(配点等)の決定および観点別評価のラベリングの方法について、回答者全体および回答者の判断による勤務校の生徒の英語力レベル(基礎・標準・発展)ごとのばらつき傾向を分析した。定期テストの形式の決定方法(図1)については、全体のおよそ半数(23件)が、「学校として科目ごとに形式が決まっている」もしくは「年度始めに話し合い学年ごとに決めている」と回答した一方、残りの半数(22件)が「年度の最初の作問者の形式にほぼ倣っている」または「その都度作問担当者が決める」と回答した。共通テストであることが前提の質問であったが、「その他」として、定期テストのうち15～20点分のみが共通テストで、残りは各授業担当者が独自で形式を決定しているケースもあった。

学校、学年で定期テストの形式について事前に決められている件数が半数に留まっている現状から、指導内容とテスト形式のすり合わせについて教員間で十分な議論がなされていない(あるいはその余裕がない)状況が推察される。生徒の英語力ごとの回答傾向と、全体の回答傾向との間には大きな特徴の差は見られなかった。

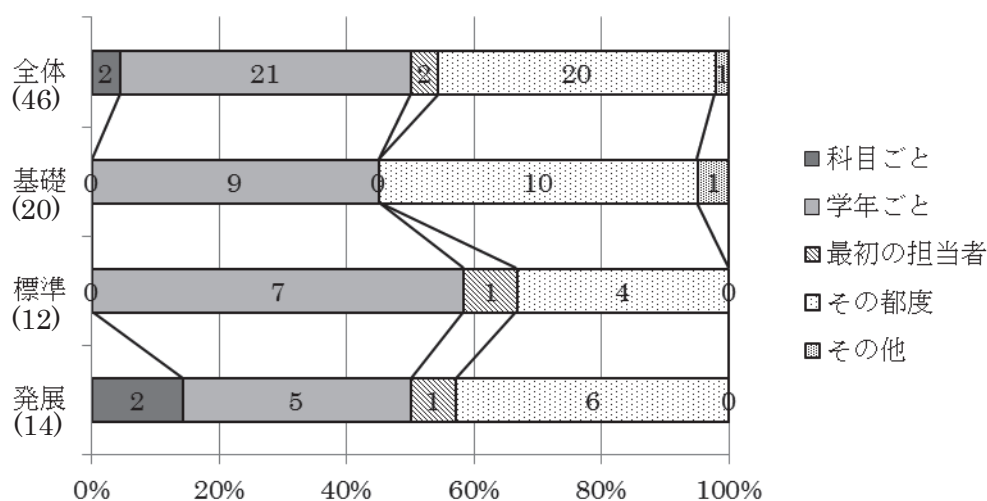


図1 定期テストの形式決定方法 (数字は回答数)

観点別のラベリング方法(図2)について、「あらかじめ問題形式とセットで決まっている」あるいは「作問後担当者間でチェックし、定義の相違があれば話し合っている」と回答したのは全体の37%(17件)で、定期テストの設問の内容や構成概念に基づき観点別のラベリングを行っていると思われるケースが半数を下回った。また、「作問担当者の定義通りにしている(特にチェックしていない)」あるいは「(問題ごとには)特に観点を設定していない」と回答したケースが半数以上にのぼった。「その他」(1件)の内容は、「定期テストでは理解の能力と知識・理解のみを見ている」との回答であり、定期テストとパフォーマンステストで4技能を総合的に評価している学校の例であった。回答結果からは、個々の問題について

の観点別のラベリングについては、構成概念妥当性や内容的妥当性についてあまり精緻に議論されていない様子や、成績処理上の便宜を優先した結果、問題形式や配点を決定した後に観点別のラベリングが行われていると思われる状況が見られた。

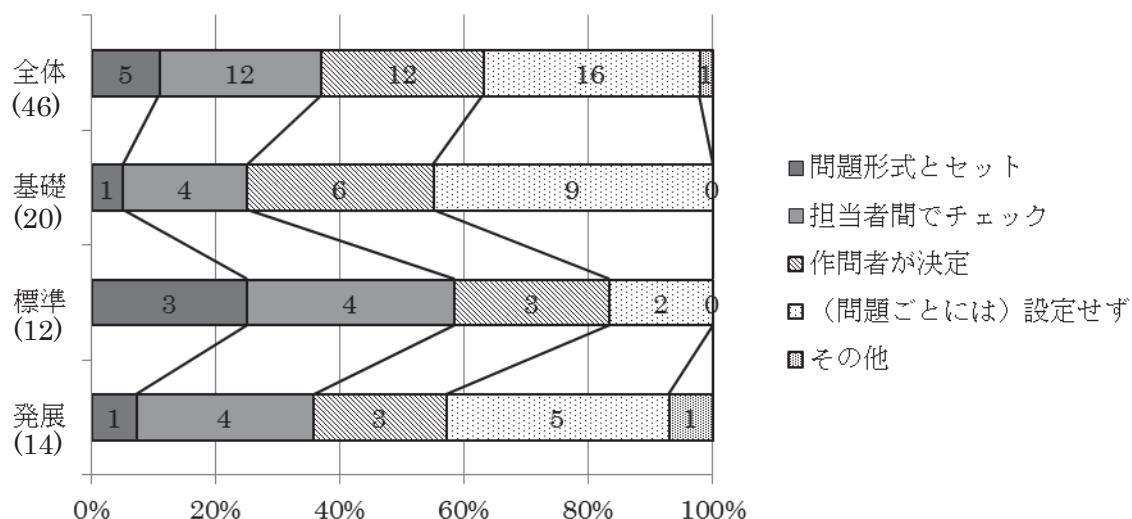


図2 観点別のラベリング方法 (数字は回答数)

定期テストの形式決定方法および観点別のラベリング方法が、回答者の勤務校の生徒の英語力の影響を受けるかどうかについては、データ数が少なく統計的検定は難しいが、今回の調査の回答分布を見る限り、全体の回答状況と生徒の学力別(3レベル)回答状況の分布にそれほど顕著な特異傾向はないようである。つまり、テスト作成や観点別ラベリングについての取組状況は、生徒の英語力ではなく、各学校の文化もしくは教師集団とその同僚性等の要因に依存しているのではないかと推察される。

### 5.1.2 テスト作成に関する現場教師の悩み

定期テストの作成、実施、採点に関する困難点について自由記述回答を求めた項目については、回答者ほぼ全員から記述回答を得た。記述回答の量は回答者により異なるが、各回答者の記述をテーマごとに分類した結果、30種類の意見が抽出された。この30種類の意見に対し、さらにコーディングと分類を繰り返した結果、課題意識の向けられた対象により、(1) 定期テストの妥当性に関する課題 (2) 定期テストの信頼性に関する課題 (3) 教員の相互理解・協調・協力体制に関する課題 の3つのカテゴリーに分類することができた。

表1は、自由記述回答をこれら3つのカテゴリー(主テーマ)、さらにサブテーマとともに示したものである。



表1 定期テストの作成、実施、採点に関する課題についての自由記述回答

主テーマ	サブテーマ	定期テストの作成、実施、採点等について 困っていること・不明点・困難点
定期テストの 妥当性 に関する課題	構成概念	4 技能をバランスよく評価できていない。(「外国語理解の能力」「言語や文化についての知識・理解」が成績の8割を占め定期テストで測られる) 観点別評価の後付け・形骸化・理解不足 観点をテスト問題に明記することに疑問を持っている。 総合問題の一括観点化(負担軽減・ミス回避のため)
	読解問題	教科書の英文を出題することで読解力が測れるのか疑問である。 読解応用問題の内容と割合の妥当性 リライトの質・難易度の設定 読解問題の作成(根強い総合問題) 初見の英文のリソース
	暗記中心	ライティングテストは一度書かせたものを再生させている。初見のトピックを出題する勇気がない。(採点の手間) 生徒の力量を考え、並べ替え、語彙問題などを「表現の能力」として いる。
	内容的 妥当性	授業の活動と共通定期テストの齟齬
定期テストの 信頼性 に関する課題	採点に関する 信頼性	和文回答の採点基準のブレ 採点基準の設定が困難なためライティング問題は出題しない。 (ライティング問題の)採点基準の不徹底(個人差)
	CRT/NRT	平均点の設定(内規の存在)
	配点の問題	テスト形式の信頼性(形式によって点数が異なる)
教員の相互 理解・協調・協 力体制に関す る課題	目標の 共通理解	学習到達目標が形骸化している。 共通理解が困難である。(テストだけでなく活動も) 教材・指導法の共有が図られていないため、共通テストが総合問題 にならざるを得ない。
	協調・協力	共通テストに関して:不参加の教員・作問の遅さ・教員間の指導の相 違、単調なテスト 初見、リライトを認めない教員(暗記テスト化) 教科書ありきの問題は避けたいが、理解が得られない。 出題、配点に関して意思疎通がうまくいかない。(指導力に差がある) 作問者ごとに異なるテスト形式(はじめに話し合われていない) はじめに決めたテスト形式が保持されない。(生徒の混乱を招く) リスニングテストが実施困難である。(設備、協力体制) ライティングの採点基準づくり(困難、話し合いに時間がかかる) テストで使う英文(選択肢等含む)の <b>native check</b> が困難である。
	環境	作成、採点にゆとりがない。(記述+マークシート希望)

定期テストの妥当性については、構成概念妥当性への教師自身の課題意識がうかがえる。4 技能を総合的に評価していないことや、総合問題の妥当性が低いことへの問題意識はある一方で、初見の英文を用いた読解問題作成への負担感があるため、現状を変えられない状況が推察される。また、生徒からの解答の幅を抑え機械的採点を可能にするために、暗記すれば解答できる問題が多くなっている

ことにも触れられている。授業内容と共通テストの内容の齟齬(内容的妥当性の欠如)が課題として挙げられているが、これは後に言及する教師間の共通理解とも深く関わっている。4技能を構成概念とすることに比べ、観点別評価のラベリングについては十分に納得できないままに行っている様子もうかがえる。

定期テストの信頼性については、英作文など記述解答の採点の問題(採点者間信頼性の維持)が挙げられている。この点に関する不安や過去に苦勞した経験が、英作文を定期テストに含めることを阻んでいることが考えられる。テスト問題の内容と配点の整合性についても疑問を感じているとの意見もあった。より大きな問題として、教科間、教員間における評定平均の極端なばらつきを抑えるため学校が内規で評定平均値の範囲を規定することが、目標基準準拠テスト(CRT = Criterion-referenced tests)ではなく、集団基準準拠テスト(NRT = Norm-referenced tests)の考え方であることの矛盾についても指摘された。

教員の相互理解・協調・協力体制に関する課題についてのコメントを分析すると、これらが前述した定期テストの妥当性、信頼性の問題の根底にあることが読み取れる。すなわち、テストの妥当性・信頼性を高める前提条件として、教員間での指導目標の共通理解、評価活動における協調・協力、またそれを可能にする職場環境ないし雰囲気が必要だということである。回答者から得られた意見の分析から、定期テスト作成とその妥当性、信頼性確保に向けての課題は次のようにまとめられるのではないかと考える。

- 指導目標、指導内容、指導技術についての教員間の共通性が希薄なため、共通して教えられる内容が同定しにくく、その結果最大公約数としての「語彙問題」「文法問題」、そして教科書本文を用いた語彙・文法問題ともいえる「総合問題」が定期テストの多くを占めている。
- ライティングをはじめとした英語による発信力の指導と評価については経験とノウハウの蓄積が不十分なため、教師の側には成功体験が少なく、結果として導入が遅れている。
- 英語教育についての考え方や信念は教師により違いがあり、授業目標や内容について話し合いの機会が十分に確保されないなかで共通テストや共通の枠組での評価活動を行うことは、教師の意欲に悪影響を与える。

### 5.1.3 定期テストにおけるリスニング問題の出題状況

定期テストのなかにリスニング能力を問う問題が出題されているかどうかをたずねたところ、46人中の21人が「定期テストのなかにリスニング問題はない」と回答した。そのうちの1件は、授業中にリスニングテストを実施しているということで、どのような形でもリスニング能力を測定していないという学校は、20校(43.5%)であった。4技能を伸ばすという科目の目標を考えれば、リスニング問題がないテストは、構成概念の一つを欠いていると言わざるを得ないだろう。テストに出題されないということは、授業においても、リスニング能力を伸ばすための指導が積極的になされていないということが想像される。

リスニング問題を定期テスト等で生徒に課している56.5%のケースについて、テスト音声の出処をたずねると、大学入試センター試験対策用の市販のリスニング教材(73.1%)、教科書指導用CD(50.0%)、

ALT による録音 (34.6%)、WEB 等からのダウンロード音声 (7.7%) が挙げられた (重複選択あり)。最も多かった市販のリスニング教材については、教科書を使った指導との結びつきが必ずしも強くないため、授業では独立した帯活動のなかで使用され、テストには付属の教師用 CD から類似問題が出題されていることが想像できる。教科書指導用の CD を使う場合は、教科書の各課のまとめとしてのリスニング活動をそのまま使ったり、教科書に出てくる文のディクテーションを出題したりすることが考えられる。教科書を使った指導の一環としてリスニングの指導をし、その内容に関連付けた形でテストを課したいと思えば、オリジナルの音声を用意したり、トピック関連の音声資料を探したりする必要が出てくるだろう。そうすると ALT に録音を依頼したり、教師が自力で WEB 等から入手したりすることになるが、アンケートではその割合は少ない。ALT については、その配置が委託事業になっているため、授業時間外に作業を依頼することが以前よりも困難になっているのかもしれない。WEB 等のダウンロード可能な音声については、テストでの使用に適切なものがなかなか見つからない、またはその存在があまり広く知られていないという可能性もあるだろう。

#### 5.1.4 定期テストにおけるライティング問題の出題状況

「外国語表現の能力」のなかで、定期テストで測定が可能であると思われるライティング問題の出題状況をたずねたところ、46 人中 9 人 (19.6%) が「定期テストのなかにライティング問題はない」と回答した。この結果をそのまま解釈すると、この約 2 割の学校のテストは構成概念としての「書くこと」の測定を欠いているということになるが、定期テスト以外でこの力が測られている可能性もある。その機会もないとすると、ライティングの指導の有無についても懸念される。80.4% という実施率はリスニングのテストよりは高いが、前述のテストに関わる課題についての記述回答を見ると、実施している学校でも、「採点基準を巡って教師の間で意見が異なり話し合いに時間がかかる」、「初見のテーマでは生徒が書けないので、一度授業で書かせたものを再度定期テストで書かせるしかない」、などのコメントが散見されることから、ライティング問題を容易には出題できない実情がうかがえる。

ライティング問題を実施している学校について、その形式を聞いたところ、実施率が高い順に、「部分英作文」(81.1%)、「1 文英作文」(59.5%)、「複数文～1 パラグラフ」(37.8%)、「2 パラグラフ以上」(2.7%) となった (重複回答あり)。採点方法については、「文法やスペリングの誤りを減点する」(97.3%)、「指定語句の使用、内容の了解度に応じて加点する」「複数観点からなるルーブリックによって評価する」(ともに 8.1%) という回答であった (重複選択あり)。試験実施時間の制限もあり、部分的または 1 文の英作文を課し、減点基準にしたがってエラーごとに減点していくという方式が圧倒的に多いことがわかる。このなかには、目標文法や特定の語句の使用を期待した和文英訳も含まれている可能性がある。定期テストには出題されていなくても、他の機会でもまとまった英文を書かせていけば、文法・語彙習得の手段としての作文ではなく、「書くこと」の能力を伸ばすライティングの指導が行われていることが期待できるが、今回の調査からはそのような指導の有無は確認できない。

#### 5.1.5 定期テストにおける読解問題の出題状況

まとまった英文テキストに基づいていくつかの問題が設定されている、いわゆる「読解問題」について

は、5.1.2 で述べたように、一つの英文に異なる観点の問題が設定された「総合問題」が、学校によっては根強く実施されていることが自由記述回答からうかがえる。これについては、定期テスト問題サンプルの観点ラベリングの分析(5.2.2)のところであらためて言及する。ここではアンケート調査で、読解問題で使われている英文の出典と、空所設定等の加工の有無についてたずねることで、読解問題が本当に「読むこと」の能力を測るものになっているかということを調べた。結果は表2のようになった(重複選択あり)。表中の「リライト」とは既習の教科書の英文を書き換えたもの、「関連文」とは、既習の教科書の英文と同一のまたは類似したトピックに関する英文、「任意文」とは、入試問題や資格試験問題など、教科書の内容とは関連しない英文をそれぞれ指している。

表2 読解問題の英文テキストの出典と加工の有無

	教科書 空所等	教科書 そのまま	リライト 空所等	リライト そのまま	関連文 空所等	関連文 そのまま	任意文 空所等	任意文 そのまま
件数	38	21	8	3	9	7	6	7
% (N=46)	82.6	45.7	17.4	6.5	19.6	15.2	13.0	15.2

教科書の英文に空所や並べ替え箇所を設けて、読解問題に使っているという回答が、8割を超えている。空所等はないもの、教科書の英文をそのまま使っているという回答もかなり多かった。さらにくわしく見てみると、24校(52.2%)のテストが、加工の有無を問わず、読解問題には教科書の英文のみを使っていた。一般的にいえば、一度授業のなかで読解した英文については、内容がすでに頭に入っているため、それをもとにしたテストでは、本当の意味での「読むこと」の能力は測れないだろう。記憶力のよい生徒が正答し、またテストに向けた準備も教科書の暗記に終始する、という可能性が高い。例えばそのなかで、目標文法を含む英文和訳の問題などを出題したとしても、日本語訳を何となくでも覚えていれば正答できてしまうかもしれない。

空所、並べ替え等の加工については、段落整序や cloze test などのように、その英文を使って特定のスキルを測ろうとしていけば問題ないかもしれないが、随所に異なる種類の加工がなされていると、まとまった英文として読むことはできないだろう。少なくとも1、2個は、無傷のまとまった英文を読ませる問題があることが望ましいが、教科書そのままでない英文で、そのような問題を出題している学校は15校(32.6%)にとどまっていた。

## 5.2 定期テスト問題サンプルの分析

### 5.2.1 観点別の配点

分析対象とした21件のテストの4つの観点別の配点(%)は下の表3のようであった。付されている観点のラベリングを尊重し、後述するその適切さについてはここでは考慮に入れていない。あくまでも、実際に観点別評価に使われた評価資料としての数値・割合としてそのまま扱っている。表4は、それぞれの観点ごとの基本統計量を示したものである。

表3 観点別の配点(%)

ID	科目名	関・意・態	表現の能力	理解の能力	知識・理解
1	コミュニケーション英語 I	0.0	23.0	27.0	50.0
2	コミュニケーション英語 I	0.0	36.0	34.0	30.0
3	コミュニケーション英語 I	10.0	30.0	30.0	30.0
4	コミュニケーション英語 I	0.0	24.0	44.0	32.0
5	コミュニケーション英語 I	0.0	24.0	46.0	30.0
6	コミュニケーション英語 I	0.0	35.0	35.0	30.0
7	コミュニケーション英語 I	0.0	8.3	50.0	41.7
8	コミュニケーション英語 I	0.0	8.3	50.0	41.7
9	コミュニケーション英語 I	0.0	5.0	37.0	58.0
10	コミュニケーション英語 I	0.0	3.0	44.0	53.0
11	コミュニケーション英語 I	0.0	5.0	35.0	60.0
12	コミュニケーション英語 II	0.0	27.0	28.0	45.0
13	コミュニケーション英語 II	0.0	31.0	27.0	42.0
14	コミュニケーション英語 II	0.0	10.0	15.0	75.0
15	コミュニケーション英語 II	0.0	24.0	50.0	26.0
16	コミュニケーション英語 II	0.0	28.0	28.0	44.0
17	総合英語	0.0	0.0	66.7	33.3
18	英語 I	0.0	30.0	26.0	44.0
19	英語 I	0.0	28.0	42.0	30.0
20	英語 II	0.0	30.0	40.0	30.0
21	英語 II	0.0	10.0	40.0	50.0

表4 観点別配点の基本統計量

	関・意・態	表現の能力	理解の能力	知識・理解
平均値(%)	0.5	20.0	37.8	41.7
最大値(%)	10.0	36.0	66.7	75.0
最小値(%)	0.0	0.0	15.0	26.0
標準偏差	2.1	11.4	11.2	12.4
変動係数	4.5	0.6	0.3	0.3

「コミュニケーションへの関心・意欲・態度」については、一般的にペーパーテストで測定することは簡単ではないと思われるが、1件だけ100点満点の10点分をこの観点に割り振っているものがあつた。具体的には、副教材の単語集からの語彙問題で、授業以外の自律的学習の成果を測るものになっている。

他の3観点を見ると、「表現の能力」の配点平均が他の2観点に比べて低く2割にとどまっている。まったく出題のないテストも1件あった。どの観点でも変動係数が0.1を大きく超えていることから、全体的にばらつきが大きいということが確認できる。これらの観点間の不均衡や、学校間・テスト間の差異については、各学校・各科目で、評点(5段階・10段階)算出の際に各観点別評価がどれくらい反映されるかという「重みづけ」が異なることや、小テストやパフォーマンステストなどの結果が評価資料として加味される割合が違っていることが理由として考えられる。それでも、「コミュニケーションへの関心・意欲・態度」は除いたとしても、3観点の配点バランスがかなり極端なものも見られることから、精査するためには、他の評価資料も合わせて観点別評価にどのような重みづけがなされているかを確認することが必要である。定期テストの「言語や文化についての知識・理解」を測る問題割合が極端に高く、重みづけもそのまま極端に高いとすれば、授業自体がその観点に関わる指導に偏重したものになっている可能性があり、それは望ましい状況ではないだろう。

### 5.2.2 観点のラベリング

テスト問題が本来測るべき能力を測っているかということ、すなわち内容的妥当性を確認するために、すべてのサンプルの問題内容・問題形式と観点のラベリングを調べた。

まず、語彙知識に関わると思われる、さまざまな種類の問題について調べてみると、表5のようなラベリングの齟齬が見られた(重複あり)。もっとも割合が大きいのは、日本語とマッチさせるなどして、空所に入る語(句)を記述させる問題を「外国語表現の能力」とラベリングしているケースであった。生徒が自分の語彙知識から適切なものを引き出してスペリングを書く、という行為が「外国語で表現している」と解釈されていると想像できるが、「外国語表現の能力」で測られるべき力は、あくまでも、言語知識を使いながら自己表現できる力であり、語彙知識そのものではないはずである。そのような力は、答えが一つに決まっているような問題では測れないだろう。

表5 語彙に関わる問題のラベリング齟齬

問題内容(問題形式)	ラベリング		件数	% (N=21)
	実際のもの	(想定されるもの)		
発音・アクセント(多肢選択)	×表現の能力	(知識・理解)	3	14.3
発音・アクセント(多肢選択)	×理解の能力	(知識・理解)	2	9.5
語彙空所補充(記述)	×表現の能力	(知識・理解)	9	42.9
語彙空所補充(記述)	×理解の能力	(知識・理解)	5	23.8
語彙空所補充(多肢選択)	×表現の能力	(知識・理解)	2	9.5
会話表現(多肢選択)	×表現の能力	(知識・理解)	2	9.5

次に、文法知識に関わると判断できる問題のラベリングを確認すると、表6のような齟齬が見られた(重複あり)。もっとも割合が大きかったのは、日本語とマッチさせるなどして、語群を意味の通る英文にしたときの語順を問う問題を「外国語表現の能力」としたケースであった。語彙空所補充問題と同様に、ここ

でも生徒の自己表現の余地はなく、正答も決まっている。文法、特に統語の知識を問う問題に他ならぬだろう。文の書き換えの問題についても同様のことがいえる。和文英訳に関しては、部分的に自己表現要素を入れ込むなど、問題の出し方によっては「外国語表現の能力」を問うことはできる可能性があるが、日本語をそのまま丸ごと英訳させるのは表現の能力を問うてはいえない。まして既習の英文の場合は、暗記力で正答できてしまうかもしれない。

表6 文法に関わる問題のラベリング齟齬

問題内容(問題形式)	ラベリング		件数	% (N=21)
	実際のもの	(想定されるもの)		
文の書き換え(空所記述)	×表現の能力	(知識・理解)	4	19.0
文の書き換え(空所記述)	×理解の能力	(知識・理解)	3	14.3
文の書き換え(全文記述)	×表現の能力	(知識・理解)	1	4.8
文の書き換え(全文記述)	×理解の能力	(知識・理解)	1	4.8
誤りの訂正(記述)	×表現の能力	(知識・理解)	1	4.8
誤りの訂正(記述)	×理解の能力	(知識・理解)	2	9.5
空所補充(記述)	×表現の能力	(知識・理解)	1	4.8
空所補充(多肢選択)	×理解の能力	(知識・理解)	2	9.5
語順整序(多肢選択)	×表現の能力	(知識・理解)	8	38.1
語順整序(多肢選択)	×理解の能力	(知識・理解)	2	9.5
文型判断(多肢選択)	×理解の能力	(知識・理解)	1	4.8
和文英訳(記述)	△表現の能力	(知識・理解)	4	19.0
和文英訳(記述)	×理解の能力	(知識・理解)	1	4.8

語彙の空所補充問題や語順整序の問題が「外国語表現の能力」を問う問題として出題されている背景には、観点別評価が誤って理解されているということもあるかもしれないが、5.1.4 で述べたように、採点の困難さ、担当者間の一致した指導の難しさなどから英作文の問題が出題されにくく、また、他の場面でもその能力を確認する方法がないため、やむなくこれらの問題を「外国語表現の能力」を測る問題と見なしてしまっているということもあるだろう。しかしながら、英語で自己表現ができる生徒を育成するためには、授業のなかにそのための指導、活動がなければならないし、定期テストに限らなくても、その力を確認する手段が必要である。

一つの英文テキストに基づいて、さまざまな知識や能力を問う問題が混在している「総合問題」は、21件中 16 件(76.2%)のテストに出題されていた。問題の内訳を見ると、空所を設けて適切な語(句)を書かせる語彙知識の問題、原形で提示された動詞を適切な時制に書き改めさせたり、文中のバラバラになった語を整序させたりする文法知識の問題、指示語の内容を書かせたり、テキスト内容に合ったステートメントを選ばせたりするような理解に関わる問題、和文英訳、英文和訳など、多岐に渡っている。そ

それぞれの設問に別々に適切な観点のラベリングをしている例も 2 件(9.5%)あったが、複数の観点の問題を含む総合問題を、まとめて「外国語理解の能力」とラベリングしているものが 8 件(38.1%)、「言語や文化についての知識・理解」としたものが 7 件(33.3%)見られた(重複 1 件あり)。総合問題は高校現場で伝統的に出題されてきており、一つの英文でさまざまな問題を作ることができるため、作成する側にとってはある意味重宝する形式である。しかし、どのような知識や力が身についているかということ、教師も生徒も知ることができ、次の指導、学習につながるようなテストにするためには、1つの大問には1つの観点を配し、より正確に知識や能力を測定できるようにすべきであろう。

## 6. 結論と今後の研究課題

本研究では、高校現場で実施されている英語の定期テスト問題の妥当性、および、定期テストの作成、実施、採点にあたっての課題を明らかにすることにより、今後の指導と評価の充実に資する知見を得ることを目的とした。高等学校英語教員からの協力を得て実施したアンケート、および提供を受けた公開済み定期テスト問題の分析から、設定した研究設問について次のことが明らかになった。

第1の研究設問である、英語の定期テスト問題の妥当性については、観点別評価に関する 2 つの次元のそれぞれについて問題があった。まず、4 技能をすべて測定するテストとなっているかという点に関して、特にリスニング、ライティングに注目すると、リスニング問題を含まない定期テストを実施している学校が 43.5%、ライティング問題を出題しない学校が 19.6%で、出題されているライティング問題の内訳もその 81.1%が「部分英作文」という状況であった。筆記テストの性格上スピーキング能力は測れないとしても、こうした状況は構成概念妥当性に問題があるといえよう。また、情意面、技能面、知識面という次元については、情意面は別として、技能(手続き的知識)を測る問題が少なく暗記により蓄えられた知識(宣言的知識)を測る問題が多い傾向が見られた。つまり、「外国語表現の能力」「外国語理解の能力」ではなく「言語や文化についての知識・理解」に関する出題が多く、前者とラベリングされた問題も内容を精査すると後者の範疇であるというケースが多々見られた。例えば、語彙の空所補充問題や語順整序問題は、言語知識を問う問題であって「外国語表現の能力」とはいえないし、一度授業で扱った教科書の英文を用いた読解問題は、暗記力あるいは言語知識を問う問題であって「外国語理解の能力」とは言い難い。作問する側の教師にはさまざまな事情があることが想像されるが、収集したテストサンプルを分析する限り、妥当性に大きな課題があることがわかった。

第2の研究設問である、定期テストの作成・実施・採点にあたっての困難点については、回答者から今後の指導と評価の改善に向けて貴重な知見を得た。それは、多くの現場の教師は定期テストの作成・実施・採点における内容的妥当性・構成概念妥当性・信頼性の重要性は十分認識しつつ、さまざまな要因により自らが理想とする教育実践ができない状況があるということである。その要因として、教師間での指導目標・指導内容・指導技術についての共通理解不足や、表現力の指導・評価に関する経験や情報の不足、さらに教師の対話を促進する学校環境の整備不足などが挙げられる。また、コメントの端々に、同僚との対話を願う教師が、そのような機会を持ってないことに対し不満や課題意識を抱いていることが見え隠れする。時間的余裕のないなかで、観点別評価といった手間のかかる評価活動を行うこ



とに対する疑問の声も散見された。これらの回答の分析から、英語テストと観点別評価の課題は、英語教師集団の同僚性や教師の認知に関わる問題と深く関わっていることが推察される。教育哲学や信念の異なる教師が目標を共有しない限り、指導内容を反映した共通テストや観点別評価の実施は困難といえる。

しかし、筆者らは、現場の教師が作成した定期テストを単に批判したり、教師の側の目標共有の努力不足を指摘したりすることは公平ではないと考えている。多様化する学校業務のなか、さまざまな制約のなかで教育にあたる教師の実践の背後には、それを余儀なくする理由が存在するのである。観点別評価は、英語によるコミュニケーション能力の構成概念をより深く理解し、知識偏重の教育を改め生徒の意欲を引きだす授業実践への示唆を与えるためのものであるにもかかわらず、運用面での煩雑さゆえに、その本来の意義が十分に理解されていない可能性がある。本研究で明確になった課題についてさらに継続して調査するとともに、教員研修の機会をとらえて、指導と評価、テスト作成についてより多くの教師間の対話を広げるとともに、必要なノウハウを共有することが求められる。現在筆者らは、「定期テスト作成マニュアル」の作成に取りかかっているが、こうした資料を教師の研修と対話のための触媒として活用し、その波及効果を検証することが今後の研究課題である。

本研究にあたり、アンケートへの回答、公開済み定期テストの提供にご協力いただいた先生方に深く感謝を申し上げますとともに、本研究が糸口となり日々の英語授業と評価、テスト作成について教師間の対話が促進されることを期待する。

#### 参考文献

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. New York: 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 University Press.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Education.
- Brown, J.D. (2005). *Testing in language programs: A comprehensive guide to English language assessment (New edition)*. New York: McGraw-Hill.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- 国立教育政策研究所 (2002) 「評価規準の作成、評価方法の工夫改善のための参考資料(中学校外国語)」 [http://www.nier.go.jp/kaihatsu/houkoku/index\\_jh.htm](http://www.nier.go.jp/kaihatsu/houkoku/index_jh.htm)
- 国立教育政策研究所 (2004) 「評価規準の作成、評価方法の工夫改善のための参考資料(高等学校外国語)」 [http://www.nier.go.jp/kaihatsu/kou-sankousiryoku/html/index\\_h.htm](http://www.nier.go.jp/kaihatsu/kou-sankousiryoku/html/index_h.htm)

- 国立教育政策研究所 (2012) 「評価規準の作成、評価方法等の工夫改善のための参考資料～新しい学習指導要領を踏まえた生徒一人一人の学習の確実な定着に向けて～(高等学校外国語)」  
[http://www.nier.go.jp/kaihatsu/hyouka/kou/11\\_kou\\_gaikokugo.pdf](http://www.nier.go.jp/kaihatsu/hyouka/kou/11_kou_gaikokugo.pdf)
- 静 哲人 (2002) 『英語テスト作成の達人マニュアル』 東京:大修館書店.
- 根岸雅史 (2011) 「技能統合の評価をどうするか」 『英語教育』Vol. 60, No. 2, pp. 29-31. 東京:大修館書店
- 根岸雅史・東京都中学校英語教育研究会編 (2007) 『コミュニケーション・テストイングへの挑戦』 東京:三省堂
- 松沢伸也(著)・佐野正之・米山朝二(監修) (2002) 『英語教師のための新しい評価法』 東京:大修館書店
- 文部科学省 (2000) 「児童生徒の学習と教育課程の実施状況の評価の在り方について(答申)」  
[http://www.mext.go.jp/b\\_menu/hakusho/nc/t20001204001/t20001204001.html](http://www.mext.go.jp/b_menu/hakusho/nc/t20001204001/t20001204001.html)
- 文部科学省 (2010a) 「児童生徒の学習評価の在り方について(報告)」  
[http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo3/004/gaiyou/1292163.htm](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/004/gaiyou/1292163.htm)
- 文部科学省 (2010b) 『高等学校学習指導要領解説外国語編・英語編』 東京:開隆堂出版
- 文部科学省 (2013) 「各中・高等学校の外国語教育における『CAN-DO リスト』の形での学習到達目標設定のための手引き」  
[http://www.mext.go.jp/a\\_menu/kokusai/gaikokugo/1332306.htm](http://www.mext.go.jp/a_menu/kokusai/gaikokugo/1332306.htm)
- 若林俊輔・根岸雅史 (1993) 『無責任なテストが落ちこぼれを作る』 東京:大修館書店.