

令和 5 年 6 月 14 日現在

機関番号：12701

研究種目：基盤研究(C)（一般）

研究期間：2017～2022

課題番号：17K03005

研究課題名（和文）英語学習到達度評価の垂直尺度化に関する研究

研究課題名（英文）Study on vertical scaling of English achievement

研究代表者

齊田 智里（Saida, Chisato）

横浜国立大学・教育学部・教授

研究者番号：50400594

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：英語教育ではパフォーマンステストや学習到達度評価において、正答・誤答の2値型採点に加えて部分点や観点別の段階評定といった多値型採点の場面が多い。本研究はまず「書くことの調査」（H22）と「話すことの調査」（H17）（国立教育政策研究所実施）の実データに多値型データ分析用の段階反応IRTモデルを適用し、部分点が有効に機能するのは正確さでなく適切さを求める問題であることを明らかにした。次に英語学習到達度目標を段階反応モデルで分析し、中学生までの基礎的運用力の尺度と高校生以上の高度な運用力の尺度とに分けて学習到達目標をIRTスコアに対応付け、英語力の発達を可視化する垂直尺度化を実現することができた。

研究成果の学術的意義や社会的意義

英語教育で扱うデータは、正答・誤答の2値型のみならず、部分点や観点別段階評価による多値型が多い。本研究の学術的意義は、英語によるパフォーマンステストの多値型データの分析に段階反応IRTモデルが活用できる点を示せたことと、英語の学習到達度の発達段階を可視化するために、Can-do形式による自己評価の多値型データを段階反応モデルで分析し、能力記述文の垂直尺度化が実現できた点である。パフォーマンステストの部分点があまり機能していない場合もあることや、学習到達度とIRTスコアとを中学生までの尺度とそれ以降の尺度に分けて対応づけし、英語力の発達段階を可視化することができた点に社会的意義があると考えられる。

研究成果の概要（英文）：Performance assessments and self-assessment questionnaires are often scored polytomously rather than dichotomously. First, this study analyzed the polytomous data sets from the writing test (H17) and the speaking test (H22) conducted by the National Institute of Educational Policy Research using the graded response IRT model. The results indicated that the partial scoring system did not work effectively on approximately 75% of the performance items. The partial scoring system worked better for the items to assess the appropriateness of language use rather than the items to assess the accuracy of language use. Second, the self-assessment data set of CEFR-J was analyzed using the graded response IRT model to create vertical scaling. The analysis made it possible to link IRT scores to Can-do descriptors on different scales, one for the basic English ability for students up to junior high school and the other for the advanced English ability for senior high school and university students.

研究分野：外国語教育

キーワード：英語力 学習到達度 垂直尺度化 項目応答理論 Can-doリスト 段階反応モデル CEFR-J

1. 研究開始当初の背景

Can-do 形式により学習到達目標を設定し指導・評価方法を改善していくことが英語教育で求められている。達成可能な目標を設定し、その達成状況を妥当な方法で評価し、学習者の英語力の発達状況を可視化できる垂直尺度化の研究は、小中高大一貫した英語教育実現のために重要である。英語教育では学習到達目標に対する自己評価やパフォーマンス評価などの分析で、「できる」「できない」や「正答」「誤答」の2値データではなく、段階評価や部分点による多値データが用いられることが多い。2値評価より段階評価のほうが学習者の状況をより詳細に把握できると考えられる。そこで、本研究ではまず、パフォーマンス評価の多値データに段階反応 IRT モデルを適用して段階評価の有効性について検討する。次に、学習到達目標を英語教育の現場で有効に活用していくために、学習到達度評価とタスクの項目応答理論 (IRT) モデルに基づく垂直尺度化に取り組む。学習到達目標を多値型 IRT モデルによる分析を行い、IRT スコアの示す意味を Can-do 表示により明らかにし、英語力の発達状況を可視化することを目指す。

2. 研究の目的

(1) パフォーマンス評価では、部分点や観点別段階評定の多値による採点が多く行われる。多値型採点は、手間はかかるがよりきめ細かく学習状況を評価することが可能になるという利点がある。一方で、多値型採点が評価の目的に照らして有効に機能しているかという検証は必要である。国立教育政策研究所実施の「書くことの調査」(H22)及び「話すことの調査」(H17)の実データをもとに、正答、準正答、誤答の多値型採点の有効性の検討を行うこととする。

(2) 日本の学校教育で広く活用されている CEFR-J は、英語初級者から上級者までの英語を使ってできることを CEFR のレベルに基づき発達段階ごとに Can-do 表示で示している。Can-do リスト検証の過程で、5,468 名の日本人英語学習者に自己評価調査を実施した(投野(編)2013)。全 110 項目に対して4段階の自己評価が行われ、2値データに変換した上で2パラメタ・ロジスティックモデルを用いて項目困難度を推定し記述文の並べ替えを行っている。このデータは CEFR-J 付属 CD-Rom で公開されている。この自己評価データを多値データとして段階反応モデルを用いて分析することで、Can-do 形式による学習到達目標の垂直尺度化を行い、IRT スコアの示す意味を Can-do 表示により明らかにし、英語力の発達段階を可視化する。

3. 研究の方法

(1) 研究 1

国立教育政策研究所が「教育課程実施状況調査」(抽出方式、H13-H25)を実施し、その枠組みでは把握が難しい内容について、「特定の課題に関する調査」を実施していた(<https://www.nier.go.jp/kaihatsu/tokuteikadai.html>)。英語では、「話すこと」については平成 17 年 11 月に、「書くこと」については平成 22 年 11 月に、全国の中学校 3 年生を対象に調査が実施された。調査データは国立教育政策研究所から教育研究用に CD-ROM として提供されており、許可を得てそれらを手に入して分析を行った。

「話すことの調査」

対象者は、全国の国公立中学校から無作為抽出された学校 33 校に通う中学 3 年生 1,090 人である。「話すこと」の、イラストが表す名詞を正しく発話する能力(10項目)、英文を正しく再構成して、意味を正確に伝達する能力(6項目)、話しかけ(質問)の内容を聞いて理解し、それにあった内容を聞き手に正しく伝える力(5項目)、自分の考えや気持ちなどが聞き手に伝わるように話す力(1項目)を評価するため、ヘッドセットに付属したマイクによるコンピュータを用いて調査が実施された(http://www.nier.go.jp/kaihatsu/tokutei_eigo_2/index.html)。全 22 項目の採点に、正答、準正答、誤答の3段階の採点基準が適用された。正答 2、準正答 1、それ以外を誤答とみなし 0 として全項目の分析を行った。

「書くことの調査」

対象者は、全国の国公立中学校から無作為抽出された学校 101 校に通う中学 3 年生 3,225 人である。「書くこと」の基礎的・基本的な知識・技能と、まとまりのある文章を書く力を評価するためにペーパーテストが実施された。ペーパーテストは、共通項目のある 2 つの調査票 A、B から構成され、調査票 A は 51 校(1,644 人)、調査票 B は 50 校(1,581 人)に実施された(http://www.nier.go.jp/kaihatsu/tokutei_eigo_2/index.html)。テストには 2 値型採点と 3 段階の多値型採点項目が含まれ、前者については正答 1、それ以外を誤答とみなし 0、後者については、正答 2、準正答 1、それ以外を誤答とみなし 0 として分析を行った。調査票 A 28 項目、調査票 B 28 項目のうち、共通に 9 項目があり、合計 47 項目を分析対象とした。

段階反応モデル (graded response model) は、受検者の回答が段階を持った順序尺度の場合に適用できる多値型 IRT モデルである(野口・大隅, 2014; Samejima, 1969)。各項目の識別力と境界特性曲線の位置母数(カテゴリー数 - 1)を推定した。分析には、Easy Estimation(熊谷,

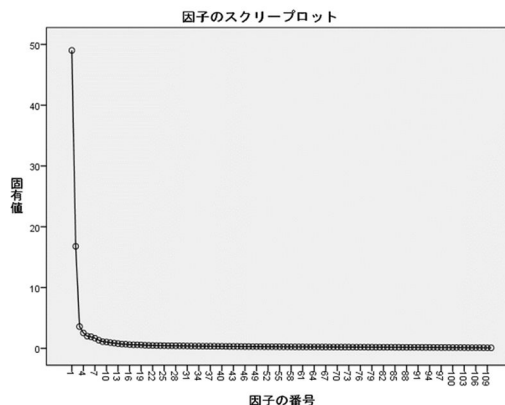
2009)を用いた。

(2) 研究2

CEFR-J の Can-do 表示の学生自己評価データを分析に用いた。全項目無回答を除く 5,458 名 (中学生 1,683 名、高校生 2,537 名、大学生以上 1,238 名) の回答を使用した。領域ごとに分けず、全データをまとめて IRT 適用条件である一次元性の仮定が満たされているかを検討したところ、B1.1 か B1.2 の前後で 2 つの因子に分けて考えたほうがよいことがわかった。

第一因子: 基礎的な運用能力 (聞くこと: B1.2 まで 16 項目、読むこと: B1.2 まで 6 項目、話すこと (やりとり): B1.1 前半まで 13 項目、話すこと (発表): B1.1 まで 14 項目、書くこと: B1.1 まで 14 項目 合計 73 項目 (黄色の部分)

第二因子: 高度な運用能力 (合計 37 項目) (緑色の部分)



聞くこと	Pre1	Pre1	A1.1	A1.1	A1.2	A1.2	A1.3	A1.3	A2.1	A2.1	A2.2	A2.2	B1.1	B1.1	B1.2	B1.2	B2.1	B2.1	B2.2	B2.2	C1	C2
読むこと	Pre1	Pre1	A1.1	A1.1	A1.2	A1.2	A1.3	A1.3	A2.1	A2.1	A2.2	A2.2	B1.1	B1.1	B1.2	B1.2	B2.1	B2.1	B2.2	B2.2	C1	C2
話すこと (やりとり)	Pre1	Pre1	A1.1	A1.1	A1.2	A1.2	A1.3	A1.3	A2.1	A2.1	A2.2	A2.2	B1.1	B1.1	B1.2	B1.2	B2.1	B2.1	B2.2	B2.2	C1	C2
話すこと (発表)	Pre1	Pre1	A1.1	A1.1	A1.2	A1.2	A1.3	A1.3	A2.1	A2.1	A2.2	A2.2	B1.1	B1.1	B1.2	B1.2	B2.1	B2.1	B2.2	B2.2	C1	C2
書くこと	Pre1	Pre1	A1.1	A1.1	A1.2	A1.2	A1.3	A1.3	A2.1	A2.1	A2.2	A2.2	B1.1	B1.1	B1.2	B1.2	B2.1	B2.1	B2.2	B2.2	C1	C2

そこで、データを 2 つの因子 (基礎力、応用力) で分けて、段階反応モデルを適用して分析をした。PreA1 から B1.1 (一部 B1.2) の基礎力の評価に関する回答を扱い、各項目の識別力と位置母数、英語で「できること」の能力値を推定した。項目番号と CEFR-J のレベルを対応させて解釈を行った。コンピュータプログラムは、熊谷 (2009) の *EasyEstimation* を使用した。項目困難度にあたる位置母数 (location) 3 つ (「ほとんどできない」と回答する確率が 0.5 となる能力値、「ある程度はできる」と「ほぼできる」と回答する確率が 0.5 となる能力値、「ほぼできる」と回答する確率が 0.5 となる能力値)、識別力 (slope)、能力値 (EAP 値) を推定した。

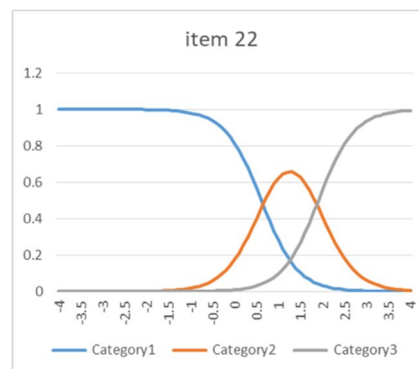
4. 研究成果

(1) 研究1

話すことの調査

全 22 項目について因子分析を行ったところ、IRT 適用前提の一次元性の仮定を満たしていると判断できた。そこで、段階反応モデルを適用して項目分析を行った。部分点が与えられる全 22 項目のうち 15 項目 (72%) で、部分点が能力値推定に全く機能しておらず、正誤の採点で十分であることが示唆された。

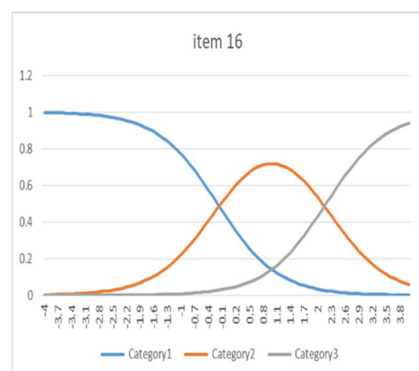
準正答とする解答類型が、発音に難がある、複数形や冠詞、前置詞の欠如などで、意味は通じるが形式における正確さの観点から部分点とした場合はあまり機能しておらず、正誤の 2 値での分析で十分能力推定が可能といえる。部分点が有効に機能していたのは、好きな季節を一つ選び、選んだ理由やその季節でしたいことを 1 分間で話す問題で、単一の文のみ話す場合を準正答とした場合 (項目 22) である。



書くことの調査

全 47 項目について因子分析を行ったところ、IRT 適用前提の一次元性の仮定を満たしていると判断できた。そこで、段階反応モデルを適用して項目分析を行った。47 項目中、部分点が与えられるのは 29 項目で、そのうちの 16 項目 (55%) で、部分点が能力値推定に全く機能しておらず、正誤の採点で十分であることが示唆された。

文字や符号の形、つづり字や大文字・小文字の書き分けなど、形式の正確さに欠くといった「意味」に直接関係のない場合の部分点はあまり機能しておらず、特に困難度が中程度の場合はこの傾向が強い。逆に文法・語法の誤りや、文のつながりの工夫の仕方、手紙の表現として不十分であるなど「意味」に関わってくる部分点の場合は、有効に機能している傾向がみられた (例、項目 16)。



以上の分析結果から導かれる知見を下記にまとめた。

- ・「話すこと」や「書くこと」のパフォーマンスを 3 段階の全体的評価で行う場合、部分点の与

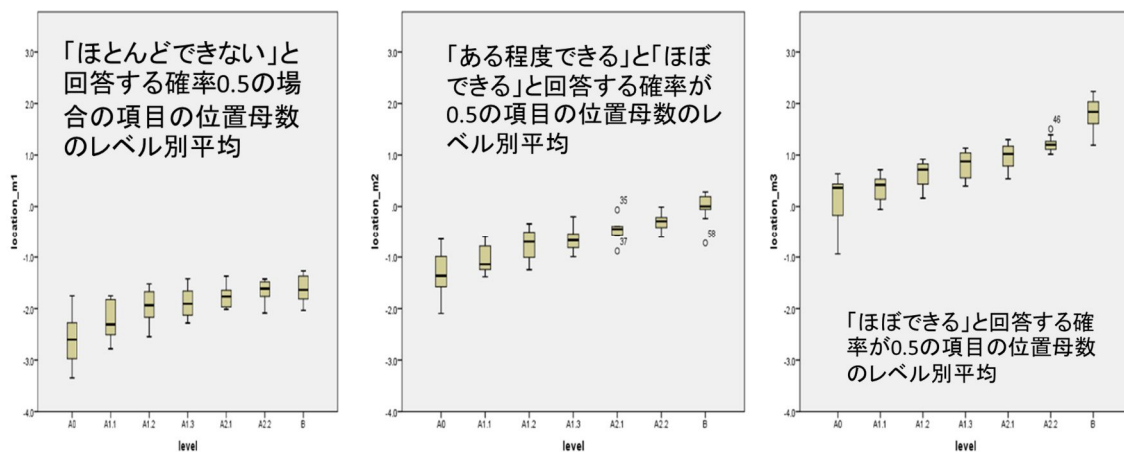
え方には注意が必要である。

- ・ 正確さを評価する場合には、部分点が有効に機能していない場合がほとんどであった（判断基準に「正しく」といった表現がある場合）。
- ・ 内容の適切さや一貫性を評価する場合は、部分点が有効に機能していた。
- ・ パフォーマンスを評価する場合、正確さについては、2段階でもよく能力を推定できる。適切さについては、3段階による評価が能力推定に有効である。
- ・ 全体的評価ではなく、分析的評価を行い、正確さと適切さに分けて分析をすることも段階評価の有効活用に効果的である。

中学生のパフォーマンス評価の場合、形式の誤りをどの程度厳格に採点すべきかについては、目的に照らして検討が必要である。

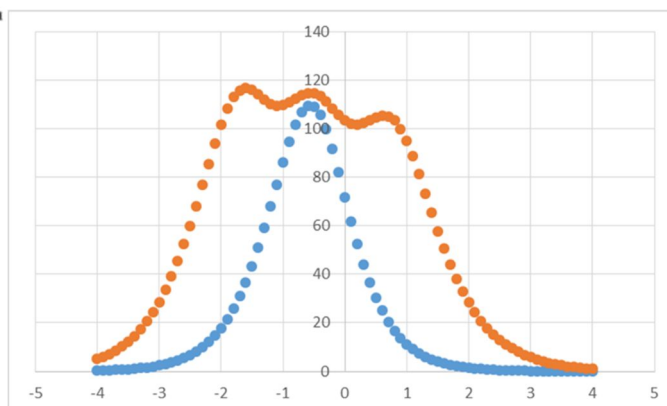
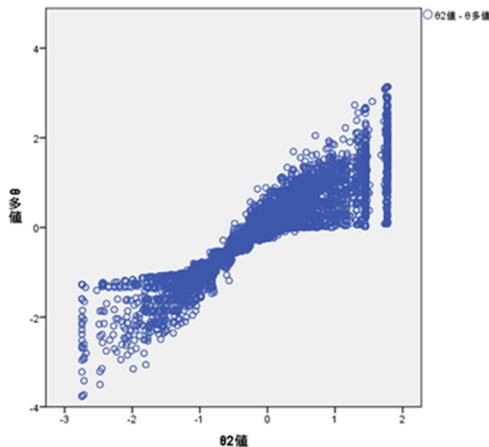
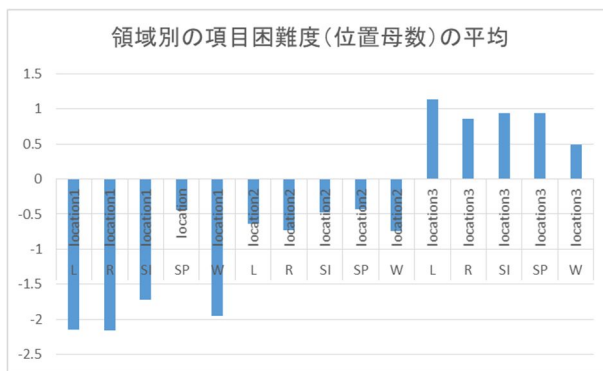
(2) 研究2

CEFR-JのCan-do記述文の段階評価は、どの段階も回答者の能力値推定によく機能していた。「ある程度はできる」の回答確率が高くなるのは、この自己評価尺度における平均的な能力値をもつ中学生はA1.1からA1.2、高校生はA1.2からA1.3、大学生以上はA2.1で、平均より1標準偏差以上高い能力値をもつ回答者で「ほぼできる」と回答する確率が高くなるのは、中学生でA1.3、高校生でA2.1、大学生以上でA2.2であった。



領域別の自己評価の特徴として、「話すこと(やりとり)」と「話すこと(発表)」は能力値の低い層で他の領域より低く自己評価をする傾向がみられた。また、能力値の高い層で「書くこと」を高く評価する傾向がみられた。

2値型分析による能力値と多値型分析による能力値の散布図では、多値型分析による能力値推定のほうが、2値型分析によるものより、能力値が高い層と低い層の双方で、より幅がある推定値となっていた(相関係数は.915)。推定の標準誤差は、多値型分析のほうが2値型分析より小さかった。情報量の比較においても、2値(青線)より多値(赤線)のほうが大きいことがわかった。



以上の結果から得られた知見を下記にまとめた。

・CEFR-J の Can-do 学生自己評価アンケートの回答結果を因子分析したところ、基礎的運用能力と高度な運用能力との2因子がみられたため、2つの尺度に分けてIRT分析を行った。英語力の発達段階の垂直尺度化を行う際には、基礎的運用能力の尺度と高度な運用能力の尺度を分けて尺度化したほうがよいことがこのデータ分析から示唆された。

・段階評価のまま2値評価による能力記述文の項目困難度の順番は、5領域全体ではCEFR-Jレベル順になっていたが、領域別でみると順番の入れ替えをしてもよい項目があった。各レベルの項目数が少ないことも影響しているだろう。

・CEFR-Jの各Can-doディスクリプタの4段階評価は回答者の自己評価による自信度を適切に推定していた。段階評価の自己評価データを多値データとして分析することで、より精緻化されたきめ細かい学習者の情報が得られた。

・「あまりできない」から「ある程度はできる」への回答確率が高くなるのは、この自己評価尺度における平均的な能力値をもつ中学生はA1.1からA1.2、高校生はA1.2からA1.3、大学生以上はA2.1である。また、平均より1標準偏差以上高い能力値をもつ回答者で「ある程度はできる」から「ほぼできる」と回答する確率が高くなるのは、中学生でA1.3、高校生でA2.1、大学生以上でA2.2であった。

・領域別に、段階評価の違いは見られる。話すこと(やりとり)は、能力値の低い学習者には難しく感じられ、書くことは、能力値の高い学習者には易しく感じられている。

・段階評価のほうが2値評価より多くの情報量が得られている。能力値推定においては、能力値のより低い層やより高い層で、より正確に能力値を推定することができるのは、段階評価による分析のほうである。

・学習到達目標としてのCan-do descriptorsとIRT尺度値とを基礎的運用能力の尺度と高度な運用能力の尺度とにそれぞれ対応させることができ、段階反応モデルによる垂直尺度化を行うことができた。その結果、英語の発達段階を共通の尺度でトレースできるようになる。

以上2つの研究から得られる共通の知見と今後の課題を下記にまとめた。

英語教育では学習到達目標に対する自己評価やパフォーマンス評価など場合、「できる」「できない」や「正答」「誤答」の2値型データではなく、段階評価や部分点による多値型データが用いられることが多い。多値型データを2値型データに変換して分析を行うより、多値型データのまま分析をしたほうが、能力推定値と項目特性値の推定においていずれもより詳細で正確な情報が得られることが実証され、その結果は英語教育に有益な知見をもたらすことがわかった。パフォーマンス評価にとって重要な採点基準の有効性の検討を多値型データ分析で行うことができる。学習到達目標の自己評価による英語力の発達段階をIRT尺度値と学習到達目標のCan-doディスクリプタの内容とを結びつけて共通尺度上で可視化することができる。

今後の課題は、パフォーマンス課題を含む大規模英語学力テストの多値型データのIRT分析を行い、英語教育の現状と課題を明らかにすること、及び、Can-doディスクリプタに結び付けられるテスト問題やタスクの困難度を推定し生徒の英語力を推定していくことである。

<引用文献>

熊谷龍一(2009)。「初学者向けの項目反応理論分析プログラムEasyEstimationシリーズの開発」『日本テスト学会誌』5, 107-118.

国立教育政策研究所「特定の課題に関する調査」

<https://www.nier.go.jp/kaihatsu/tokuteikadai.html>

英語:「話すこと」の調査(平成17年11月実施)全国の中学生3年生(抽出)

英語:「書くこと」の調査(平成22年11月実施)全国の中学生3年生(抽出)

野口裕之・大隅敦子(2014)。「テストニングの基礎理論」東京:研究社.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph, 17. Psychometric Society.

投野由紀夫(編著).(2013)。「英語到達度指標CEFR-Jガイドブック」東京:大修館書店.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件 / うち国際共著 1件 / うちオープンアクセス 2件）

1. 著者名 Chisato Saida	4. 巻 33
2. 論文標題 Analysis of the Junior High School English Test Items of the National Assessment of Academic Ability Using Item Response Theory	5. 発行年 2022年
3. 雑誌名 Annual Review of English Language Education (ARELE)	6. 最初と最後の頁 33,48
掲載論文のDOI（デジタルオブジェクト識別子） 10.20581/arele/33/0_33	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Chisato Saida	4. 巻 14(3)
2. 論文標題 Creating a Common Scale by Post Hoc IRT Equating to Investigate the Effects of the New National Educational Policy in Japan	5. 発行年 2017年
3. 雑誌名 Language Assessment Quarterly	6. 最初と最後の頁 257、273
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/15434303.2017.1368518	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 斉田 智里・亀山 弘二郎・吉田 早希・和田 雅広・Stebbins Antoine	4. 巻
2. 論文標題 横浜国立大学附属鎌倉中学校英語科における目標に準拠した評価の実践	5. 発行年 2018年
3. 雑誌名 横浜国立大学教育学部紀要. 1, 教育科学 = Journal of the College of Education, Yokohama National University. The educational sciences,1	6. 最初と最後の頁 58、75
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計8件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 斉田智里
2. 発表標題 パフォーマンス評価における多値型採点の有効性の検討
3. 学会等名 全国英語教育学会第45回弘前研究大会
4. 発表年 2019年

1. 発表者名 斉田智里
2. 発表標題 段階反応モデルを用いたCan-do自己評価項目の垂直尺度化と活用
3. 学会等名 関東甲信越英語教育学会第43回神奈川研究大会
4. 発表年 2019年

1. 発表者名 斉田智里、関口友子
2. 発表標題 センター試験英語はそれほど悪いテストか？－「論証に基づく妥当性検証」の試み－
3. 学会等名 関東甲信越英語教育学会第42回栃木研究大会
4. 発表年 2018年

1. 発表者名 亀山弘二郎、斉田智里
2. 発表標題 学習到達目標の改善と評価方法の工夫 - 評価ポートフォリオの活用 -
3. 学会等名 関東甲信越英語教育学会第42回栃木研究大会
4. 発表年 2018年

1. 発表者名 斉田智里、田中秀毅
2. 発表標題 中高英語科教育の諸課題：文法指導と学習評価
3. 学会等名 平成30年度横浜国立大学教員免許状更新講習
4. 発表年 2018年

1. 発表者名 斉田智里
2. 発表標題 「書くことの調査」データの特異項目機能（DIF）分析 学校所在地域の規模・校種の観点から
3. 学会等名 全国英語教育学会 第43回 島根研究大会
4. 発表年 2017年

1. 発表者名 亀山弘二郎・斉田智里
2. 発表標題 学習到達目標・年間指導計画・学習評価三位一体の英語指導体制の構築 CAN-DOリスト形式での学習到達目標の設定と指導・評価改善のための活用
3. 学会等名 関東甲信越英語教育学会 第41回新潟研究大会
4. 発表年 2017年

1. 発表者名 斉田智里
2. 発表標題 英語免許状所有の有無と小学校英語教科化への意識との関係
3. 学会等名 関東甲信越英語教育学会 第41回新潟研究大会
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 酒井英樹・廣森友人・吉田達弘（編著）	4. 発行年 2018年
2. 出版社 大修館書店	5. 総ページ数 319
3. 書名 「学ぶ・教える・考える」ための実践的英語科教育法 第9章「学習評価のあり方・進め方：目標に基づく指導と評価の一体化の実現のために」	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------