

英語電子メール作成タスクの難易度調査  
—英語ライティング能力に関する Can-do Statements 開発への基礎研究—

**A Study of the Levels of Difficulty of Writing E-mail in English:  
A Fundamental Study on Developing Can-do Statements with  
Regard to Writing Ability in English**

工藤洋路

Yoji KUDO

東京外国語大学大学院博士後期課程/日本女子大学附属高等学校

*Tokyo University of Foreign Studies*

*The Doctor's Program of the Graduate School /*

*The High School Affiliated with Japan Women's University*

**Abstract**

The purpose of this study is to assess the levels of difficulty of writing e-mail in English. During the study, 148 Japanese high school students are asked to work on the six e-mail tasks. Each of the tasks have different kinds of language functions that the subjects are required to fulfill in the reply to each of the given e-mails. The written e-mails are evaluated using two different methods: the first one is “a rating scale for evaluating writing” by Canadian Language Benchmarks. The other one is the method to which “three levels of reaction speech (Tanaka, 2004)” is applied. The scores given these two methods are analyzed using the software, FACETS, by which task difficulty is to be calculated.

From this research, the following findings are made:

- (1) The e-mail task requiring the students to give new information that is not mentioned in the given e-mail is ranked as the most difficult.
- (2) The e-mail task requiring the expression of emotions, e.g., surprise, anger, concern is rather difficult.
- (3) The e-mail tasks where the writers are asked to complete simultaneous use of two language functions are not necessarily evaluated as difficult.
- (4) The range of the levels of difficulty of the six e-mail tasks is not very wide, which implies that replying to given e-mail may not require a wide range of writing ability.

**Keywords**

E-mail, Levels of Difficulty, Writing Ability, Language Function, Can-do Statements

## 1. 研究の背景と目的

言語能力の発達に関して、近年その段階を記述する枠組みの開発が進んでいる。例えば、Common European Framework of Reference for Languages(以下、CEFR)、Canadian Language Benchmarks(以下、CLB)、American Council on the Teaching of Foreign Language(以下、ACTFL)の Proficiency Guidelines などである。それぞれに、設定された言語能力段階ごとに、多くの場合は言語技能別に、Can-do Statements<sup>注1</sup>が付属しており、実際にその言語を使って何ができるかについての記述がある。また、TOEFL<sup>®</sup>や TOEIC<sup>®</sup>、そして英検(実用英語技能検定)にもそれぞれテスト得点ごとに、あるいは能力を示す級ごとに、こちらも多くの場合言語技能別に、Can-do Statements<sup>注2</sup>に該当するものがある。また、英語のコミュニケーション能力を測定するためのテストとして、ベネッセコーポレーションが開発した GTEC for STUDENTS にも、受検者の成績を表すスコアレポートの中に、Can-do Statements の一種である「あなたの英語はこんなに使える」が存在している。テスト得点と実際に英語を使ってどのようなことができるかを関連させ、それをテストフィードバックとして受検者に提示することによって、受検者の学習への動機付けを高めることが Can-do Statements の大きな役割となっている。

しかしながら、現在開発されている Can-do Statements の多くは、その記述の具体性について問題がある。例えば、ベネッセコーポレーションの GTEC for STUDENTS で見られるライティング能力に関する記述例は、「英語の手紙や電子メールで、自分が書きたいことを辞書を引きながら書くことができます。さらに書く練習を積むと、辞書なしでも大体のことは書けるようになるはずです。」というものである。この記述には、「手紙」や「電子メール」といった現実の生活で使用する言語媒体が含まれており、一見具体性に富んだ記述に思われる。しかしながら、「電子メール」一つを取り上げても、「電子メールで書きたいことが書ける」という記述だけでは、電子メールといってもどんなタイプ(こちらからの発信型なのか相手のメールに対する返信型なのか)のものかについては分からない。また、誰に対して送信する電子メールなのか、電子メールを送信する目的は何かなどの点についても具体的とは言えない。これらの点から想定される電子メールの難易度にはかなりの幅が予測されるであろう。例えば、「自分の住所を伝える」という目的で書く電子メールと、「都合の悪い相手を説得する」という目的で書く電子メールでは、求められる英語ライティング能力に大きな差があるのは容易に想像がつく。言い換えると、書くべき電子メールで果たされる言語機能の違いによって、そのタスクの難易度は上下することになる。工藤(2006)によると、特に、驚き、怒り、心配などの感情を表現することと、相手からの勧誘を断ると同時に別の提案を持ちかけるような二段階の言語機能を果たす電子メールは難易度が高いことが判明している。ただし、この研究で用いた電子メールタスクの採点方法が非常に簡易なものであることから、言語機能の達成度を適切に測定する採点基準の構築が今後の課題となっている。そこで、

---

注1 CEFR における Can-do Statements は Self-assessment Grid の中で“I can...”の形で記述され、また CLB では Global Performance として“Can ...”の形で、ACTFL では Proficiency Guidelines として“Able to”という形で記述されている。

注2 TOEFL<sup>®</sup>(詳細には TOEFL<sup>®</sup> iBT) では、Can-do Statements という名ではなく、Language Competence Descriptor として“I can...”の形で、TOEIC<sup>®</sup>では Can-do Tables として“Can ...”の形で、英検では Can-do リストとして「～できる」の形で記述されている。

本研究では、電子メールタスクを2つの新たな採点方法を用いて採点し、求められる言語機能によって、電子メールタスクの難易度にどう差が生じるのかを調査する。リサーチ・クエスションは、「どんな言語機能を達成することが求められる電子メールタスクの難易度が高いか」とする。

## 2. 調査

### 2.1 調査方法の概要

求められる言語機能が異なる複数の電子メールタスクを実際にパソコン上で被験者に実施し、被験者から送信された電子メールを2つの方法を用いて評価し、項目応答理論のラッシュ分析が実行可能なソフトである FACETS を用いて、タスクの難易度を量的に測定する。

### 2.2 電子メールタスク

実際に電子メールを作成してもらうために、パソコンが設置されている教室で、各被験者はパソコン1台を前にして、電子メールを作成する。一般に使用されている電子メールソフトに類似した電子メール作成プログラムを組み、画面上で作業をしてもらう(資料1参照)。また、実際の電子メール作成場面にできる限り近づけるために、返信する人物になりきるためのプロフィール表と週間予定表(資料2参照)が配布され、それに基づいた電子メールの返信を作成する。

電子メールのタスクについては、想定される言語機能を基準に、受信したメールに対する返信を作成する6種類のタスクを設定した。タスクの数を6つとした理由は、被験者として、GTEC for STUDENTS の受検者数の多くを占める高校生を設定するため、高校の通常の授業時間である50分を考慮したからである。パソコンの技術的な説明や実施における注意事項の説明に多少の時間を要することから、この6つのタスクを40分以内で行うこととする。また、送信されたメールに対する返信のメールを作成するという「返信型」の電子メールタスクを採用したのは、こちらからの「発信型」の電子メール作成を被験者に求めた場合、書くべき内容を事前に提示することが必要となってしまう。それが日本語で与えられた場合は結局、和文英訳のタスクになってしまい、また、英語で与えられた場合は、その英語を直接写せばタスクが遂行されてしまう可能性が高くなってしまうことから、「発信型」ではなく「返信型」のタスクを採用する。6つのタスクにおいて求められる言語機能については、高等学校学習指導要領の「言語の働き」の例として挙げられているものを参考に、電子メールの返信にふさわしく、また想定される被験者の能力の幅の範囲で適切だと思われるものを設定した。実際の場面で電子メールを作成する場合は、複数の言語機能を果たすことが同時に求められることが多いと判断し、本研究においても、いくつかのタスクにおいては、複数の言語機能が求められるように設定した。また、電子メールを送る相手に関しては、設定された言語機能から考慮して任意に設定した。その結果、表1に見られるような電子メールタスクが完成した。

表1 電子メールタスク一覧

	主な言語機能	受信メールの要点
タスク1	個人情報の伝達	住所と電話番号を教えてください
タスク2	報告&勧誘応諾	明日放課後テニスをしよう
タスク3	予定の伝達&勧誘拒否	6月16日暇ですか？
タスク4	感情(驚き, 怒り, 心配など)表現&予定伝達	6月15日の約束を忘れてしまった…
タスク5	勧誘拒否&謝礼	6月15日に英語を教える時間とれます
タスク6	勧誘拒否&勧誘	6月16日にミュージカルに行きましょう

## 2.3 被験者

被験者に関しては、3つのスーパー・イングリッシュ・ランゲージ・ハイスクール (SELHi) 指定校より、合計148名を設定する。その内訳は、A校64名、B校21名、C校63名である。

## 2.4 電子メールタスクの評価方法

### 2.4.1 電子メールタスクの評価方法①

電子メールタスクの1つ目の評価方法については、今回のタスクの設定基準である言語機能に焦点を当てることができる評価方法を用いる。本研究で利用したものは、Canadian Language Benchmarks の A rating scale for evaluating writing の観点である、(1) Effectiveness, (2) Appropriateness, (3) Relevance, accuracy and adequacy of content の3点である (Pawlikowska-Smith, 2002)。各観点はそれぞれ4段階で採点される (採点基準については、資料3を参照)。各観点が表す内容は次のものである。

(1) Effectiveness describes whether the global purpose of communication has been achieved according to task requirements. Has the writer demonstrated a global functional ability to perform the writing task according to task requirements? Is the reader able to understand the writer's message? If necessary, would the reader be able to use the writer's text, according to task requirements?

(2) In written texts, appropriateness includes language, suitable format, layout, visual/graphic presentation of text to audience and purpose, appropriate use of formality, register, style, and socio-cultural knowledge and references (e.g., cultural metaphors, quotations, etc.) and naturalness of expression.

(3) Accuracy in the criterion applies to research-based tasks which require factually accurate information.

↓

(3)' Accuracy in the criterion describes whether the written e-mail products have accurate information from “Profile” and “Schedule.”

(3) Relevance, accuracy and adequacy of content の観点に関しては, accuracy についての記述しかなされていない。そこで, relevance や adequacy of content を, タスクの与えられた条件に対して適切であることと判断し, この研究における基準として(3)'を作成した。

#### 2.4.2 電子メールタスクの評価方法②

電子メールタスクの2つ目の評価方法は, 本研究で採用する「返信型」の電子メールタスクの評価に効果的と思われる方法である。採用した採点基準は, 田中(2005)が提示している応答発話の3つのレベルである。このレベルは「応諾と拒否の発話基準」となっており, それぞれ以下のような基準を設定している。表2の発話例は, “Can you help me?”という友人からの依頼に対する応諾である。

表2 応諾の発話基準と発話例

レベル	基準	発話例
1	応諾のみ	O.K.
2	積極的な応諾	Sure. Just a minute.
3	積極的かつ協力的な応諾	Sure. What can I do?

【田中(2005), p.62 より】

表3の発話例は, “Can you lend me your pen for an hour?”という友人からの依頼に対する拒否となっている。

表3 拒否の発話基準と発話例

レベル	基準	発話例
1	拒否のみ	Sorry, I can't.
2	理由を伴う拒否	Sorry, I can't. This is the only one I have.
3	理由と代案を伴う拒否	Sorry, I can't. This is the only one I have. You can use it in a few minutes.

【田中(2005), p.63 より】

この基準では, 「応諾」または「拒否」について, 相手との人間関係を考慮して, 単なる「応諾」や「拒否」のメッセージを伝達することよりも, 「理由」や「積極的な気持ち」を伝えることができる学習者が高いレベルにあることを前提としている。つまり, タスクの条件には明示

されていない「プラス・アルファの感情や情報を伝える」という言語機能の難易度が高いことを意味している。

また、この基準では、発話行為が対象となっているが、「発話における応答」は、「電子メールの返信」と、相手のメッセージに返答するという点で類似していることから、このレベルによる基準を採用する意義が認められる。また、この基準の対象となっている「応諾」と「拒否」の言語機能は、本研究で設定した電子メールタスクの複数のタスクで求められていることから、この基準を採用することが妥当と言える。適用に際しては、電子メールの返信として不適切なものをスコア1点とし、上記のレベル1である「応諾のみ／拒否のみ」をスコア2点、レベル2のものをスコア3点、レベル3のものをスコア4点として、本研究で使用することとする(採点例については、資料4を参照)。

## 2.5 電子メールタスクの評価者

電子メールタスクの評価は、採点方法①については日本人英語教師1名、採点方法②については英語母語話者1名によって、それぞれ実施する。

## 3. 結果

### 3.1 評価方法①の結果

評価方法①を用いた採点結果(資料5を参照)について、FACETSを用いて各観点ごとにタスクの難易度を算出した結果は以下の通りである(FACETSの詳細な分析結果は、資料6～8を参照)。

#### (1) Effectiveness

タスク1	<	タスク2	<	タスク4	<	タスク3	<	タスク5	<	タスク6
(-1.06)		(-0.24)		(0.02)		(0.11)		(0.32)		(0.85)

#### (2) Appropriateness

タスク1	<	タスク3	<	タスク2	<	タスク4	<	タスク5	<	タスク6
(-0.64)		(-0.37)		(-0.28)		(0.04)		(0.16)		(1.09)

#### (3) Relevance, accuracy and adequacy of content

タスク1	<	タスク2	<	タスク4	<	タスク3	<	タスク5	<	タスク6
(-1.13)		(-0.39)		(-0.11)		(0.28)		(0.40)		(0.95)

どの観点においても、タスク1の難易度が最も低い。また、同様に、タスク6の難易度が最も高く、タスク5の難易度が2番目に高い。難易度を表す指標を見ると、タスク1が、どの観点でも群を抜いて易しいタスクと判断され、またタスク6が、同様に、群を抜いて難しいタスクと判断されている。

### 3.2 評価方法②の結果

評価方法②を用いた採点結果は次の表 4 の通りである。

表 4 評価方法②による採点結果

	タスク1	タスク2	タスク3	タスク4	タスク5	タスク6
平均	2.87	2.91	2.99	2.37	3.01	2.52
標準偏差	0.67	0.77	0.85	0.73	1.02	1.1

この結果を、FACETS を用いてタスクの難易度を算出した結果は以下の通りである (FACETS の詳細な分析結果は、資料 9 を参照)。

タスク5	<	タスク3	<	タスク2	<	タスク1	<	タスク6	<	タスク4
(-0.43)		(-0.39)		(-0.24)		(-0.16)		(0.48)		(0.74)

タスク4が最も難易度が高く、タスク5が最も難易度が低いという結果が算出された。

### 3.3 FACETS による分析の評価

#### 3.3.1 信頼性

各採点方法によるテストの信頼性は、資料 6～9 によると、以下の表 5 の通りであり、概ね高い信頼性が得られている。

表 5 テストの信頼性

評価方法		信頼性係数
評価方法①	Effectiveness	0.98
	Appropriateness	0.94
	Relevance, accuracy and adequacy of content	0.96
評価方法②		0.93

#### 3.3.2 モデルの適合度

静(2007)によると、ラッシュモデルを用いた項目分析において、一般的に、Infit の MnSq の値が 0.7～1.3 の範囲から外れる場合は misfit として扱うが、今回の結果においては該当するものはなかった。また、同様に、Infit の ZStd の値が、-2.0～2.0 の範囲を外れる場合は、misfit として扱うが、今回の結果において、評価方法②のタスク1が-3.0 となっており、misfit と判断できる。今回の被験者のレベルから考えて、タスク1が非常に易しかったことが原因と考えられるが、許容範囲(-2.0～2.0)を下回っている項目、つまり、overfit している項目であれば、通常は問題とならない。なぜなら、その項目が、「すでに他の項目が明らかにした能力情報をほとんどなぞっているということなので、テストにとってやや余分な存在(なくてもいい)かもしれませんが、有害な存在(あってはならない)ではない(静、

2007)」項目であるからである。このことから、FACETS を用いた分析結果に見られるモデルの適合度には問題がないと言える。

### 3.3.3 弁別度

項目の弁別度を表す PtBis(point biserial correlation)については、0.25 以上が許容とされている(Henning, 1987)が、今回の分析結果によると、以下のものがその値を下回っている。今回の被験者のレベルから考えて、タスク1～3が易しかったことが原因と考えられる。

表 6 弁別度が低いタスク

評価方法	弁別度が 0.25 以下のタスク
方法① Effectiveness	タスク1 (0.03), タスク2 (0.22), タスク3 (0.22)
方法① Relevance, accuracy and adequacy of content	タスク1 (0.18)
方法②	タスク3 (0.18)

## 4. 考察

### 4.1 全体についての考察

全体的には、タスクの難易度として、1.00 を上回る、または-1.00 を下回るものは、わずかにしか見られなかったことから、難易度に大幅な差があったとは言えない。「言語機能」を基準に据えた評価方法では、コミュニケーションがそれなりに成立している限り、あまりタスク難易度に差が生じないことが理由の一つと推測できる。日本の高校生英語学習者であれば、語彙や文法の誤りの有無を評価基準に入れると、タスク難易度に差が生じることになるのかもしれない。しかしながら、モデルの適合度やテストの信頼性に問題がなかったことから、本調査の分析について、各タスクの結果を考察することは有効である。個別のタスクについて検証してみると、タスクの難易度を規定する条件を引き出すことが可能になるであろう。

### 4.2 各タスクについての考察

2種類の評価方法のいずれの結果より、タスク6の難易度が高いことから、このタスクは根本的に難易度が高いタスクであると言える。複数の言語機能を同時に果たすことが求められている点は、その他のほとんどのタスクに当てはまるため、このタスクの難易度が高い理由は、他のタスクにはない「新たな提案をする」という機能によるものと考えられる。つまり、自ら積極的に新しい話題を提供するような言語機能は、タスクの難易度が上げる条件となりうる。このことから、自らが電子メールを先に送るという「発信型」の電子メールタスクは、一般的に「受信型」よりも難易度が高くなることが予測される。

タスク4については、評価方法①では難易度があまり高くはないが、評価方法②では高くなっている。これは、コミュニケーションを適切に達成すること自体の難易度は低いが、評価方法②の観点である「プラス・アルファの感情や情報を伝える」ことを、このタスクでは達成することが難しいということが分かる。タスクの条件に明示的に感情を表すことが要求されていない場合に、自らの意志で自分の感情を表現することは、難易度が高いと言える。ただし、



ここでの感情とは、驚き、怒り、心配という種類のものであり、例えば、感謝や喜びといった種類の感情を伝えることは求められるライティング能力は異なる可能性もある。

タスク5が評価方法②では難易度が低いが、評価方法①では高くなっている。このタスクは、「こちらが事前をお願いをしていたことに対して、相手から日時の提案がある」というものだが、こちらの元々のお願いについて断りを入れなくてはならないことから、評価方法②の観点である「積極的かつ協力的な応答／理由と代案を伴う拒否」を達成しやすいタスクだったと言える。ただし、二段階の言語機能を達成する必要があることから、評価方法①では、難易度が高いタスクとされたと判断できる。

タスク2と3については、各評価方法において、難易度が近いと思われるタスクである。全体の位置づけの中では、易しめのタスクと判断され、また弁別度においても、低い値が算出されているため、今回の被験者のレベルにとってはやや物足りないタスクであったと思われる。どちらのタスクも、複数の言語機能を果たすことが求められていることから、複数の言語機能を果たすことが必ずしも難易度を上げる条件ではないと言える。

タスク1については、評価方法①において弁別度が低く、また、評価方法②において、overfit していると判断されていることから、今回の被験者にとっては、不要なタスクだったと言える。住所や電話番号を相手に伝えるという返信メールを作成することは、高校生英語学習者にとっては、非常に簡単なものであったということである。

#### 4.3 評価方法①の3つの観点についての考察

評価方法①で利用した3つの観点については、どの観点においても、最も難易度が高いのがタスク6、次がタスク5、そして、一番低いのがタスク1ということについては共通している。特に、タスク6とタスク1は、どの観点においても、難易度だけでなく、度合いを表す数値においても他のタスクを圧倒している。そして、タスク5については、難易度が3番目に高いタスクと数値的にかなり近いことも共通点である。タスク2～4については、観点によって差はあるものの、全体として、この3つの観点が測定している内容がかなり類似していることが考えられる。言語機能の達成度を評価する目的で、**Canadian Language Benchmarks** の **A rating scale for evaluating writing** から採用した観点であるが、それぞれ独立した観点ではなく、かなり依存し合っている観点であると言える。確かに、観点の2つ目である **Appropriateness** で高い評価を得るためには、タスクの目的や条件を考慮した上で適切な言語使用が求められる。その点は、観点の1つ目である **Effectiveness** に共通しており、さらにタスクの条件（返信メール作成のために与えられた「プロフィール」と「スケジュール」）に適しているかどうかを評価する3つ目の観点である **Relevance, accuracy and adequacy of content** とも共通している。

### 5. 研究成果

本研究のリサーチ・クエスチョンである「どんな言語機能を達成することが求められる電子メールタスクの難易度が高いか」に対しては、今回の調査の結果から次の点が研究成果として提示できる。

- (1) 「新たな話題を提供する」という言語機能を達成することが求められる電子メール作成タスクは難易度が高い。
- (2) タスクの条件として明示されていない「驚き、怒り、心配などの感情を伝える」という言語機能を達成することが求められる電子メール作成タスクは難易度が比較的高い。
- (3) 複数の言語機能を達成することが求められる電子メール作成タスクが一概に難易度が高いとは言えない。
- (4) 住所や電話番号を伝えるなどの「個人情報の伝達」という言語機能を達成する電子メール作成タスクの難易度は非常に低い。
- (5) 「返信型」の電子メール作成タスクについては、どんな言語機能を達成することが求められたとしても、難易度に非常に大きな幅があるタスクを設定することが難しい。

## 6. 今後の課題

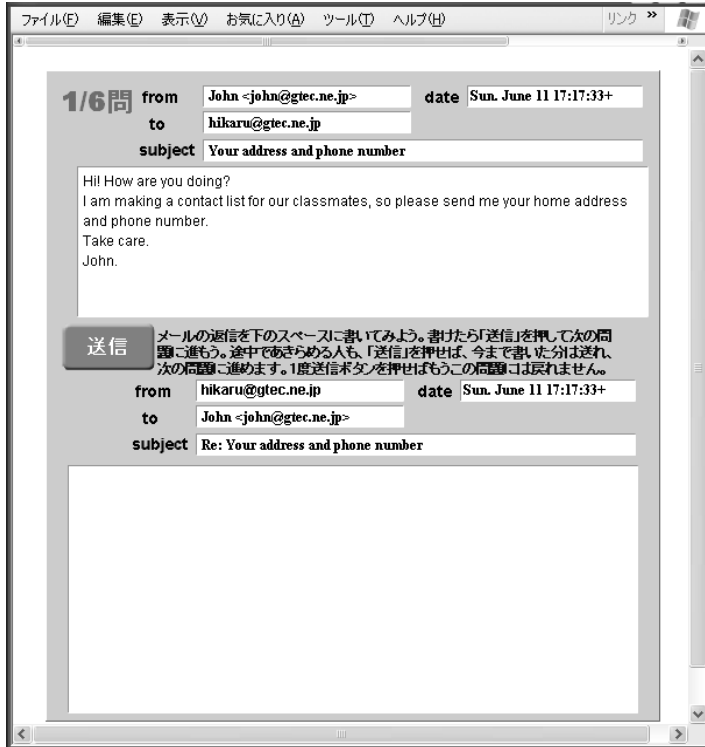
本研究の問題点から今後の課題考えると以下の通りである。また、本研究を応用した研究についても触れてみる。

- (1) 電子メールタスクの採点については、各評価とも1名の採点者によるものだったため、今後は複数の採点者による評価を行う必要がある。
- (2) 今回の被験者にとって、易し過ぎるタスクが複数見られたことから、今後は被験者レベルを考慮したタスク設定を行う必要がある。
- (3) Effectiveness や Appropriateness など言語機能の達成度を測定する複数の観点を用了が、類似した結果になったため、今後は言語機能の達成度を測定する方法の構築も並行して行っていくことが課題であろう。
- (4) コミュニケーションの成否が電子メールタスクの評価の大前提だったため、使用した文法や語彙については、評価観点には組み込まなかった。今後は、文法や語彙などの言語材料にも注目した評価方法を検討していくことも課題である。
- (5) 電子メール作成タスクについては、「返信型」ではなく「発信型」の導入や、目上の人に対するメールを作成するタスクの導入なども検討する意義はあるだろう。「新たな話題を提供する」という言語機能の難易度が高いという結果が出たので、「発信型」の電子メール作成タスクの難易度は高くなることが予測される。
- (6) この研究において電子メール作成タスクの難易度が見え始めてきたが、この結果を応用する手段として、電子メールタスクの結果と、GTEC for STUDENTS のライティングテストの結果を照らし合わせてみるのが考えられる。その結果、GTEC for STUDENTS のライティンググレードごとに、どんな言語機能を持つ電子メールタスクを書くことが可能になっていくかが分かり、テストフィードバックに有益な Can-do Statements の作成が可能となるであろう。

## 参考文献

- Henning, G. (1987). *A Guide to Language Testing: Development, Evaluation, Research*. Boston: Heinle & Heinle.
- Pawlikowska-Smith, G. (2002). *Canadian Language Benchmarks Theoretical Framework*. Centre for Canadian Language Benchmarks.
- 工藤洋路 2006. 「日本人高校生英語電子メール作成能力調査—英語ライティング能力に関する Can-do Statements 開発への基礎研究—」『ARCLE REVIEW』No.1, pp.113-123. Action Research Center for Language Education.
- 静哲人 2007. 『基礎から深く理解するラッシュモデリング—項目応答理論とは似て非なる測定のパラダイム—』関西大学出版部.
- 田中正道 2005. 『これからの英語学力評価のあり方—英語教師支援のために—』教育出版.

## 資料 1 電子メール作成画面



## 資料 2 電子メール作成: 作成者プロフィールと週間予定

Profile	Hikaru
School	Minami High School
Club	Basketball
Likes	Movie, Soccer, Comedy, Musical
Address	1-1-1 Minato-Town, Kanagawa Pref.
Phone Number	045-953-ABCD
Mail Address	<a href="mailto:hikaru@gtec.ne.jp">hikaru@gtec.ne.jp</a>

Schedule	
6/12 [Mon.]	after school: no club (look for someone to spend time with)
6/13 [Tue.]	after school: club (16:00-18:00)
6/14 [Wed.]	after school: no club (look for someone to spend time with)
6/15 [Thu.]	after school: go to Yokohama with Chris to buy shoes & t-shirts, watch soccer game at the stadium together (18:30-21:00)
6/16 [Fri.]	school holiday: study at home (until 12:00), movie with Peter (14:00-17:00), dinner at home with my family (from 18:00)
6/17 [Sat.]	after school: club (16:00-18:00), after club: buy basketball books with Chris
6/18 [Sun.]	want to go to see <u>MUSICAL!</u> (look for someone to go with)

### 資料3 電子メールタスクの評価方法①の採点基準

(1) Effectiveness	
1	Learner is not yet functionally effective in writing; purpose of communication is not achieved according to task requirements
2	Learner is functionally only marginally effective in writing; purpose of communication is only marginally achieved according to task requirements
3	Learner is functionally effective in writing; purpose of communication is achieved according to task requirements
4	Learning is functionally very effective in writing; purpose of communication is achieved with excellence according to task requirements

(2) Appropriateness	
1	Not enough ability to function in writing to demonstrate appropriateness
2	Evidence to suggest insufficient awareness of socio-cultural appropriateness of language and/or format to social contextual factors and purpose
3	Developing sense of appropriateness of language, format / text layout to social contextual factors evident in accomplishing the task. Evidence of some control of formal and informal registers in writing, and of using appropriate idioms, figurative language, and cultural references. Still frequent non-native phrasing of otherwise grammatical text
4	Good appropriateness of language, format, and layout to social contextual factors: control of formal and informal register, knowledge of socio-cultural conventions, evidence of appropriate cultural references, high degree of “naturalness” in expression. Errors likely to cause social misunderstandings are rare

(3) Relevance, accuracy and adequacy of content	
1	Irrelevant, inaccurate and/or inadequate response according to task requirements
2	Response of limited relevance, accuracy and/or adequacy according to task requirements
3	Mostly relevant, accurate, and adequate response according to task requirements; may have some gaps or redundancies and repetitions
4	Relevant, accurate, and adequate response according to task requirements

#### 資料4 電子メールタスクの評価方法②による採点例

スコア	内容	タスク2	タスク3
1	不可	タスク達成していない(間違った情報、相手のメールの誤解、空欄)	
2	応諾のみ／ 拒否のみ	Hi, Mary. Tomorrow I have no club. so I can play tennis with you. See you tomorrow. Hikaru	Mmm... I'm sorry .I can't go to the beach because I don't have free time. So I can't go there. Hikaru
3	積極的な応答／理由を伴う拒否	Hi! I will have no club after school tomorrow. So I can play tennis with you. I am looking forward to playing tennis with you. See you tomorrow. Hikaru.	Hi! I'm sorry, I am not free that day.I'm going to study at home until 12:00, go to the movie with Peter 14:00 for 17:00 and eat dinner at home with my family from 18:00. So I can't go to the beach. Sorry. See you. Hikaru.
4	積極的かつ協力的な応答／理由と代案を伴う拒否	Hi! Mary. What's up? Thank you for your e-mail. I'm fine! I have no club after school tomorrow, so I'm just looking for someone to spend time with me! I'm not so good at playing tennis, so please make me a good player! I'm looking forward to playing tennis. Hikaru	Hi,David! I'm sorry but I can't go to the beach with you. I'm going to study at home until 12:00, in afternoon, I'll go to the movies with Peter, and at night, I'll have a dinner at my home with my famiry that day. I'm looking forward to next chance. Hikaru

#### 資料5 評価方法①による採点結果

	タスク1				タスク2				タスク3			
	EF	AP	RA	TL	EF	AP	RA	TL	EF	AP	RA	TL
平均	3.6	3.3	3.8	10.8	3.3	3.2	3.6	10.1	3.1	3.2	3.4	9.7
標準偏差	0.5	0.6	0.5	1.2	0.7	0.6	0.7	1.7	0.8	0.7	0.9	2.0

	タスク4				タスク5				タスク6			
	EF	AP	RA	TL	EF	AP	RA	TL	EF	AP	RA	TL
平均	3.3	3.1	3.5	9.9	3.0	3.0	3.3	9.4	2.7	2.6	2.9	8.3
標準偏差	0.7	0.8	0.7	1.8	0.9	0.8	1.0	2.4	1.1	1.0	1.2	3.0

EF = Effectiveness / AP = Appropriateness / RA = Relevance, accuracy and adequacy of content / TL = Total

## 資料 6 評価方法①による FACETS の結果 (Effectiveness)

Item Measurement Report (arranged by mN)

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Average	Model Measure	Model S.E.	Infit		Outfit		PtBis	N	Items
						MnSq	ZStd	MnSq	ZStd			
814	292	2.8	2.90	.89	.07	1.0	0	1.0	0	.34	6	6
930	292	3.2	3.32	.34	.07	.9	0	.9	-1	.34	5	5
969	292	3.3	3.44	.11	.08	1.1	0	1.2	2	.20	3	3
983	292	3.4	3.49	.02	.08	.8	-2	.9	-1	.28	4	4
1021	292	3.5	3.60	-.25	.09	1.0	0	1.2	1	.21	2	2
1099	292	3.8	3.82	-1.11	.13	1.1	0	1.1	0	.09	1	1
969.3	292.0	3.3	3.43	.00	.09	1.0	-2	1.0	.2	.24	Mean (Count: 6)	
86.9	.0	.3	.28	.61	.02	.1	1.0	.1	1.2	.09	S.D.	

RMSE (Model) .09 Adj S.D. .60 Separation 6.83 Reliability .98

Fixed (all same) chi-square: 253.2 d.f.: 5 significance: .00

## 資料 7 評価方法①による FACETS の結果 (Appropriateness)

Item Measurement Report (arranged by mN)

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Average	Model Measure	Model S.E.	Infit		Outfit		PtBis	N	Items
						MnSq	ZStd	MnSq	ZStd			
743	296	2.5	2.53	.82	.08	1.2	2	1.2	2	.33	6	6
805	296	2.7	2.75	.38	.09	1.0	0	1.0	0	.33	5	5
882	296	3.0	3.01	-.21	.09	.8	-2	.8	-2	.31	2	2
886	296	3.0	3.03	-.24	.09	.9	0	.9	-1	.30	3	3
899	296	3.0	3.07	-.35	.09	.9	-1	.9	-1	.21	1	1
905	296	3.1	3.08	-.40	.09	1.1	1	1.2	-1	.19	4	4
853.3	296.0	2.9	2.91	.00	.09	1.0	-1	1.0	-2	.28	Mean (Count: 6)	
59.4	.0	.2	.21	.45	.00	.1	1.7	.1	1.6	.06	S.D.	

RMSE (Model) .11 Adj S.D. .43 Separation 3.79 Reliability .93

Fixed (all same) chi-square: 96.2 d.f.: 5 significance: .00

**資料 8 評価方法①による FACETS の結果 (Relevance, accuracy and adequacy of content)**

Item Measurement Report (arranged by mN)

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtBis	N	Items
766	280	2.7	2.86	.85	.06	1.0	0	.9	-1	.34	6	6
907	280	3.2	3.43	.28	.07	.9	-1	.9	0	.37	5	5
940	280	3.4	3.53	.11	.07	1.0	0	1.0	0	.27	3	3
960	280	3.4	3.59	-.01	.08	1.2	2	1.2	1	.24	2	2
982	280	3.5	3.66	-.15	.08	.7	-3	.9	0	.30	4	4
1071	280	3.8	3.88	-1.08	.13	1.2	1	1.0	0	.11	1	1
937.7	280.0	3.3	3.49	.00	.08	1.0	-.3	1.0	-.2	.27	Mean (Count: 6)	
91.9	.0	.3	.31	.58	.02	.2	1.7	.1	.9	.08	S.D.	

RMSE (Model) .09 Adj S.D. .57 Separation 6.62 Reliability .98

Fixed (all same) chi-square: 233.1 d.f.: 5 significance: .00

**資料 9 評価方法②による FACETS の結果**

Item Measurement Report (arranged by mN)

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtBis	N	Items
347	147	2.4	2.37	.74	.11	.7	-2	.8	-2	.28	4	4
369	147	2.5	2.54	.48	.11	1.2	2	1.2	1	.36	6	6
421	147	2.9	2.91	-.16	.11	.7	-3	.7	-2	.33	1	1
427	147	2.9	2.96	-.24	.12	.9	0	.9	-1	.25	2	2
438	147	3	3.03	-.39	.12	1.2	1	1.1	0	.18	3	3
441	147	3	3.05	-.43	.12	1.3	1	1.2	1	.40	5	5
407.2	147.0	2.8	2.81	.00	.11	1.0	-2	1.0	-.4	.30	Mean (Count: 6)	
36	.0	.2	.26	.45	.00	.2	2.3	.2	1.9	.07	S.D.	

RMSE (Model) .11 Adj S.D. .43 Separation 3.79 Reliability .93

Fixed (all same) chi-square: 96.2 d.f.: 5 significance: .00