〈報告・資料〉

英語教育における評価のイノベーション

広島大学総合科学部　山　田　　純

キーワード：　正答率　　反応潜時　　事象関連電位

〔要旨〕

本稿では，関連諸学問を念頭に置きながら英語学力の評価のイノベーションを通覧し，まず2つの流れとしての世代を考察する。第1世代は，学習者の言語知識の測定にあたってのイノベーションである。これは，学習事項が限定される　local evaluation　とランダムで全体的な評価を行う　global evaluation　とに分けられる。前者では，言語学の深層構造の概念，後者ではクローズテストなどがイノベーションをもたらした。第2世代は，心理言語学からの知見が端緒である。学習事項の情報処理速度が評価の対象であり，既知のことばの使える度合いを測定しようとしている。その発想自体，イノベーションと呼べる。第1世代と第2世代のつながりは，言語知識と言語運用という大きな理論的問題を内包している。第3世代は，神経言語学に関わるが，まだ始まっていない。事象関連電位をはじめ，多くの手法が開発されつつあり，英語教育学もそれらを早急に導入すべきである。以上の考察は，言語を扱う関連諸学問に依拠しており，英語教育学のいわば縦糸であった。横糸としては，教科教育学がある。そこにどのような共通性と相違性があるか，これからの研究課題である。

はじめに

関連諸学問の概念や技術やイノベーションが当該学問に大きな影響をもたらすことは，これまでしばしば観察されてきた。それは，英語教育学においてもしかりであった。斬新なアイディアやイノベーションは，しばしば関連諸学問から得られている。直接，導入する場合もあったし，英語教育学という学問のフィルターを通し，ひとひねりしなければならない場合も多いように思われる。今後もそのような原料輸入および輸入原料加工の作業は続くように思われる。本稿では，英語教育学における評価を対象として，イノベーションの過去を振り返り，近未来を展望してみたい。

さて，ここに言う評価とは，学習者の言語（英語）能力を的確に把握することに尽きる。それは教育・指導を行うための出発点であり，基盤である。しかし，学習者の言語能力を的確に把握することは，言語学，心理学，教育学，心理言語学の永遠のテーマであると言ってよい。それほど，広範で深淵なテーマであるが，各分野で多くの知見がいろいろなレベルで得られている。ここで，そ

れらを概観してみるとき，２つの大きな流れを捉えることができる。それは，時代の流れにも沿っていて，第１世代と第２世代とに分けられる。ただし，これらの世代は，第１世代が終焉し，第２世代が取って代わるというのではなく，並行して存続，発展する。加えて，両者の関連はどうか，という新たな問題も惹起される。

第１世代

第１世代は，常識的であるが，学習者の言語知識を測定する流れである。これは，ほかの教科の評価と同じで，特定の項目を学習者が知っているか否かを測定するにすぎない。しかし，何をどう測定するかで，イノベーションが見られる。関連諸科学の筆頭は，言語学であった。その言語学における複雑な言語構造の分析の成果を導入することで，これまでにない視点で学習者の言語能力を見るようになった。たとえば，つぎの文は，表層構造は同じである。

(1a) John persuaded Mary to resign.

(1b) John believed Mary to be innocent.

しかし，その深層構造がまったく違うと考えられる(Chomsky, 1981, Radford, 1988)。(1a)と(1b)の深層構造は，それぞれ(2a)と(2b)である。

(2a) John persuaded Mary [$_{S'}$[ce] [$_{S}$PRO to resign]].

(2b) John believed [$_{S}$Mary to be innocent].

ここで，PRO は，空であるが，主語となる代名詞である。このような構造上の違いは，文法文と非文の違いを明確に説明する。たとえば，(3a)が文法文であり，(3b)が非文であることも容易に理解できる。

(3a) John persuaded Mary firmly [$_{S'}$C PRO to resign].

(3b) *John believed [$_{S}$Mary firmly to be innocent].

つまり，firmly という adjunct は，それが修飾する動詞と同じ文の中になければならないという規則がある。(3a)はその規則に従うので文法文であるが，(3b)はそれに違反しているので非文である。つぎのような事例についても同様の説明ができる。

(4a) John persuaded Mary himself [$_{S'}$C PRO to resign].

(4b) *John believed [$_{S}$Mary himself to be innocent].

このような言語学の知見は，当然，英語教育方法と同時に評価の視点にも影響を与える。学習者が(3b)や(4b)を聞き，読むことはないが，話し，書くことはどうかという問題がある。学習者は，ほとんどの場合，(5a)や(5b)という反応をするであろう。

(5a) John firmly persuaded Mary to resign.

(5b) John firmly believed Mary to be innocent.

学習者が(3b)や(4b)のような特殊的な非文を使わないと同時に(3a)や(4a)のような特殊的な文法文を使わないとすれば，過小般化が行われていることになる。それをどのように評価すべきか，新

たな問題が提起されている。Chomsky が登場するまでは，深層構造という概念はなく，persuade のような動詞を含む構文がいかに複雑であるかを教師や評価者は気づかなかった。しかし，今や言語学における深層構造というイノベーションを英語教育学の評価論で検討せざるを得なくなった。それにうまく対応することが英語教育のイノベーションにつながるであろうが，ことは容易ではない。

深層構造についての多くの発見は，確かに画期的である。これまで以上に学習項目の解釈に深みが増し，項目間の関連などの体系化が進んできた。しかし，評価の基本は変わっていない。教科書があり，特定の学習項目が限定され，的をしぼってテストされることに変わりはない。それは，local evaluation と呼ぶことができる。学習者の local linguistic ability が測定される。そういう評価は，教師の経験，さらには英語教育の長い伝統に裏打ちされていて，信頼性が高い。とは言え，そこに偏向や見落としがないとは断定できない。言語項目は無数にある。その無数にある言語項目を限定するとき，無意識的に選択されない項目群が存在したとしても不思議ではない。

では，無限に存在する言語項目をランダムに選択して学習者を評価することはできないか。そのようなタイプのテストは存在している。その代表がクローズテストである。このテストでは，無数に存在するテキストの中から学習者のレベルにふさわしいテキストを選び，そのテキスト中の語を，たとえば7語ごとに削除し，適語を埋めさせる。その例は，以下の通りである。

つぎの文中の ＿＿＿＿ に適当な語を補いなさい。

Suntory Hall

Suntory Hall is located in a very distinguished neighborhood, right in the middle of the Hotel Okura, U. S. Embassy, Akasaka Ark Hills, and TV Asahi in Minato-ku, Tokyo. It was built by Suntory Limited 1.＿＿＿＿ commemorate 60 years of whisky production 2.＿＿＿＿ 20 years of beer sales. It 3.＿＿＿＿ a premier concert hall consisting of 4.＿＿＿＿ stoires, three above ground anf two 5.＿＿＿＿ levels.

Looking at the floor plan 6.＿＿＿＿ the first floor will give you 7.＿＿＿＿ idea of the variety of facilities 8.＿＿＿＿ here. Imagine a long rectangle. The 9.＿＿＿＿, a lobby of mahogany and marble, 10.＿＿＿＿ the Small Hall which seats nearly 11.＿＿＿＿, fill the lower two-fifths of the 12.＿＿＿＿.

解答例は 1. to, 2. and, 3. is, 4. five, 5. basement, 6. of, 7. an, 8. available, 9. entrance, 10. and, 11. 500, 12. plan である。このイノベーションの原理はきわめて単純明快であった。クローズ法は，もともとテキストの読みやすさの測定法のひとつであった。さまざまなテキストを選び，特定の被験者群に空欄を埋めさせる。正答率が高ければ高いほど，そのテキストは読みやすいということになる。そこで，テキストを一定に保つと，被験者の読解力が高ければ高いほど，正答率が高くなる。これは，一見すると，旧来の完成問題であると思われるかもしれない。しかし，そこには大きな違いがある。旧来の完成問題は，local evaluation である。その

ような完成問題の例は，つぎの通りである。

つぎの文中の _____ に適当な語を補いなさい。

(1) I wonder what has become _____ him.

(2) I was forced to drink _____ my will.

(3) The house is _____ construction.

(4) The goods are valued _____ 10,000 yen.

(5) Young _____ he is, he is a man of ability.

(6) It is ten years _____ I saw her last.

(7) _____ in easy English, the book is read by many people.

(8) A: Did you buy a hat?
B: Yes, I bought _____.

解答は 1. of, 2. against, 3. under, 4. at, 5. as, 6. since, 7. Written, 8. one である。解答語を見ると，前置詞や接続詞がほとんどである。動詞の場合は，熟語的表現の中に埋め込まれている。これらは，教師や教科書が教える項目であるので，この種の問題でよい成績を得るためには，その教える項目を効果的に記憶することが必要である。極端な場合，それだけができれば，よい成績が得られる。他方，クローズテストでは，テスト項目がランダムであるので，その対策のための特別な準備を短期的に行うことはできない。その学習者の global linguistic ability が測定される。これは，local evaluation に対して global evaluation である。もっとも，両者を2分法として見るのではなく，連続的に捉えるべきであろう。すべてのテストが2つの極の間のどこかに位置づけられる。しかし，どのように位置づけるかは，これからの研究課題であり，そこにイノベーションが求められる。

いずれにせよ，上記のようなテストでは，学習者の言語知識の量がペーパー上で測定される。それは，0点から100点までのスケールで測定される。たとえば，80点ならば，従来型のテストでは，閉じた領域の中で約80%の知識があると解釈され，overall proficiency test では，ある水準の内の80%の水準に位置すると解釈される。これが第1世代の評価法である。このように言語知識量を測ることは，言語テストの基本であり，それに対して何ら異論を唱える筋はない。このような第1世代言語テストは，絶えずイノベーションを加えつつ，恒久的に存続してゆくであろう。

第2世代

心理言語学では，ひとつのイノベーションが1970年に現れた。それは，Rubenstein による LDT: lexical decision task（語彙判定課題）である。これは，スクリーンに語または無意味語を

提示し，被験者は，語であれば右のボタンを，無意味語であれば左のボタンをできる限り早く押すという課題である。研究者は，それぞれの反応をミリ秒単位で測定し，母語話者のメンタルレキシコンや情報処理のメカニズムを研究した。このイノベーションが出るや否や，プライミング効果や音読反応潜時を測定する研究が爆発的に行われるようになった。そして，このような測定は，まさに母語話者の言語能力を測定していると見ることができる。実際，英語を母語とする児童や大学生の読解力や読語力を測定するために始められた。具体例を図１に示す。

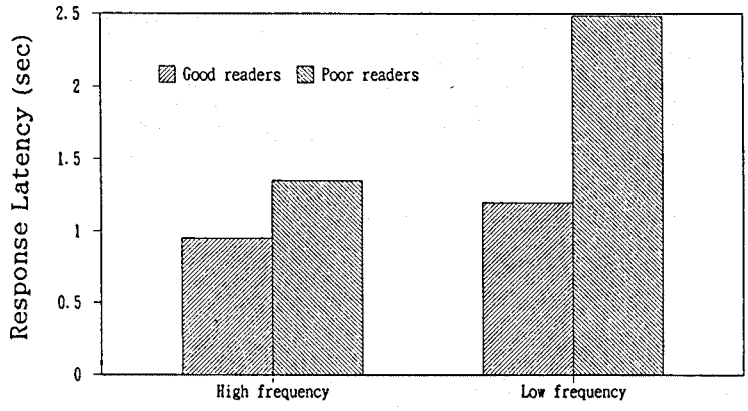

図１　読解力の上位群と下位群の音読反応潜時（Perfetti and Hogaboam, 1976）

被験児はアメリカの児童で，読解テストの上位群と下位群である。スクリーンに英単語を呈示し，被験児はできる限り早くその語を音読する。語の呈示から音読が始まるまでの時間がミリ秒単位で計測される。これらの課題は，既知の言語項目を用い，それによって言語能力を推測しようとする。すなわち，第１世代のように，特定の言語項目を知っているかどうかを測るのではなく，既知の項目をどのくらい早く情報処理できるかを測るのである。これは，今のところ，わが国および海外の英語教育の評価の部門ではほとんど導入されていない。

筆者らは，小規模な実験を少し試みている段階である。図１では，視覚刺激に対する反応時間が測定されている。筆者らは，聴覚刺激の反応時間を測定しようとした。被験者は，広島大学学生で，CELT の聴解力テストの上位群と下位群であった。イヤフォンから英単語（高頻度語と中頻度語）を１語ずつランダムに呈示し，理解できたらすぐにボタンを押すように指示した。そして，反応潜時は，語の呈示終了時点からボタンを押すまでの時間とした。高頻度語と中頻度語の平均値を群ごとにまとめると，図２のようになった。このパターンは，基本的には図１のパターンと同じである。下位群の被験者は，頻度が低くなるにつれて，反応時間に大きな遅れが見られる。したがって，図１から，読解では遅読反応が読解力の低下と結びつき，図２から，聴解では遅延反応が聴解力の低下と結びついている可能性が示唆される。

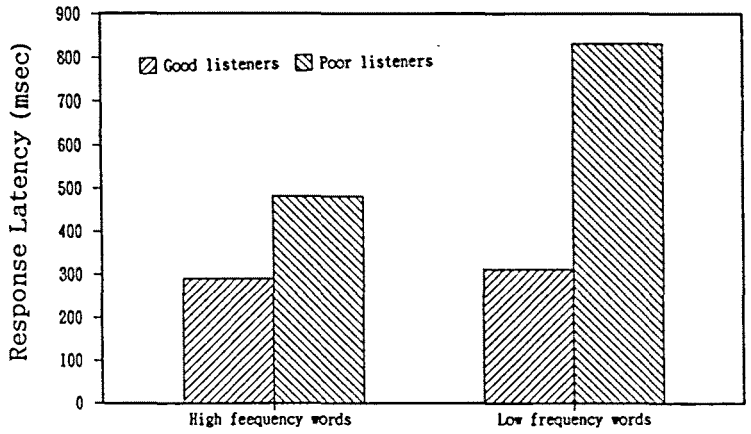

図2　聴解力の上位群と下位群の反応潜時（Sakaki, 1994）

このような既知の言語項目の反応速度の測定は，local evaluation であり，global evaluation を工夫できるかどうかが研究課題のひとつである。ともあれ，当面は，どのような言語項目がなぜ早く情報処理できるのか，といった問題が検討されるであろう。そして，基礎研究がある程度進み，具体的に実践されるようになると，それは，2つの方向に展開してゆくはずである。ひとつは，学習者の言語診断である。学習者のさまざまプロファイルが作成され，教材と教育方法へ貴重な示唆が提供されることになろう。もうひとつは，新たなるテスト形式の導入である。大学入試センター試験は，多くの良問（多くが local evaluation）から成り，全国平均が毎年65点くらいである。たとえば，これをもっとやさしくして，平均が90点くらいになるようにする。そして，制限時間を90分ではなく，25分くらいにする。そうすると，平均点が50点くらいになるだろう。これは，使える英語能力を測定するテストと見なされる。

第1世代と第2世代の結びつき

理論的な問題として，第1世代と第2世代の評価はどのように関連しているかが明らかにされなければならない。図1と図2に関して，2つの世代の結びつきを示唆したが，たとえば，頻度という点で，第1世代の評価においても図1や図2のようなパターンが存在するのであろうか。筆者らは，これについてひとつの答えを得た。広島大学学生を被験者として，英単語（高頻度語と中頻度語）を聴覚呈示し，それを書き取らせる課題を与えた。その正答率の平均値を群別にまとめると，図3のようになった。このパターンは，まさに図1と図2のパターンである。これらの被験者に対して図2の場合のような語を用い，聴覚反応潜時を測定すると，図2のようなパターンが得られるであろうし，図2の被験者に図3のような語を聴覚呈示し，正反応率をまとめると，図3のような

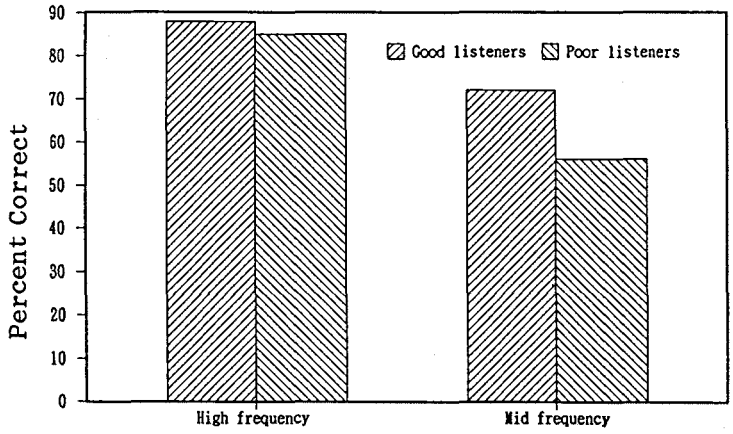

図3　聴解力の上位群と下位群の正反応率（山田・高島，1996）

パターンが得られるであろう。しかし，実際，正答率の高い項目は常に反応時間が短いと断定できるであろうか。両者の相関は，本来，限りなく1に近いものなのであろうか。たとえば，高頻度語は，常に正答率が高く，反応時間が短いのであろうか。否，頻度のみならず，学習時期も重要要因であるというデータが，現在，集積されつつある。では，早期学習項目は，常に正答率が高く，反応時間が短いのであろうか。それに反する事例は存在しないか。あるいは，上述の90分のセンター試験と25分のセンター試験の相関は，どのあたりに落ちつくのであろうか。そしてそれはなぜなのか。このような問題の探求がさらなるイノベーションの創出につながるであろう。

第3世代

さらに第3世代の評価によるイノベーションの可能性を提言したい。イノベーションが新しい道具の導入によることがある。第2世代の勃発は，コンピュータによるミリ秒単位での反応測定が可能になったことも一因である。しかし，第2世代の道具は，物理的に表面に現れた言語行動の測定であった。第1世代も紙と鉛筆という道具を使って overt responses を測定した。

そもそも言語は，大脳皮質に存在する。それならば，じかに大脳皮質の活動を測定できないか。その試みは，神経言語学の長い歴史の中に見られる。しかし，その大半は，失語症患者を対象とする研究であった。ところが，1980年代になって，発達性難読症や発達性失語症の子どもを対象として，さらには，健常児や健常者を対象にして始まった。これは，きわめてエキサイティングな始まりである。近い将来，わが国の英語教育においても大脳皮質の活動を測定する研究が始まるであろう。先陣争いはまだ始まっていない。どこで誰が先鞭をつけるかである。その具体例を音素レベルで見てみよう。周知のごとく，日本人は，/r/と/l/の弁別ができにくい。これを事象関連電位の差

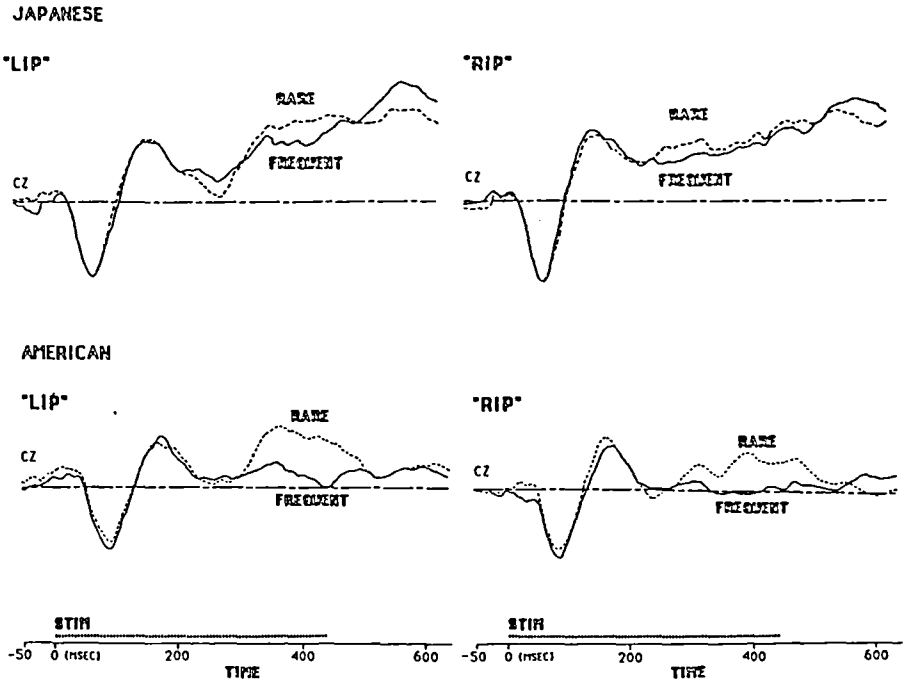

図4　日本人とアメリカ人の "lip" と "rip" 事象関連電位 (Buchwald et al., 1994)

によって明確に示したのが Buchwald et al. (1994) の研究である。アメリカ人の場合は，例えば "lip" を続けて聞いていて，時折 "rip" が呈示されると，図4に示すように，300ミリ秒あたりから大きな電位差が観察されるが，日本人にはそのようなパターンが見られない。このような測定・評価は，知識の測定とか情報処理の反応時間の測定に限定しにくく，第1世代と第2世代とは著しく異なる。もちろん，/r/と/l/の知識を前提としているし，情報処理速度も同時に計測されているが，事象関連電位はより根本的な言語活動を測定していると考えられる。では，第1世代と第2世代で明らかにできないような評価がこの第3世代のイノベーションによって測定できるのであろうか。あるいは，単なるもうひとつの評価方法にすぎず，的確な評価を行う上で必要ではない道具なのであろうか。筆者は，第1世代，第2世代を越える重要な評価法であると考える。第1世代では，知識を問う。それは，特定項目に限定すると，知っているか知らないかの all or none の測定である。しかし，たとえば/l/と/r/の「知識」はそのような2分法の形態をしているのであろうか。拙稿 (1991) では，中間的な状態が存在することを示唆するモデルを示した。一方，図4が示すように，事象関連電位は，連続的である。もし，/l/と/r/の学習が有か無の二者択一であれば，その学習過程をこの事象関連電位によって追跡することにより，その実相が明らかにされるであろう。

第3世代では，音素レベル，語彙レベル，統語レベルにおいてすでにかなりの成果があるように思われるかもしれないが，実際は氷山の一角にすぎない。第3世代も，現時点では，local evaluation の範疇に入る。第2世代の場合と同様に，global evaluation は考えられないか。研究とイ

ノベーションはこれからである。

教科教育における評価

ここまでは，英語教育学の関連諸学問として言語に関わる学問のみに注目した。それは，英語教育学の縦糸を求める，と表現できるかもしれない。一方，もうひとつの密接に関わる学問は，もちろん各教科を対象とした教科教育学である。これは，横糸である。第1世代は，全教科共通の評価法である。そして，それぞれにおいて巧みなイノベーションが存在してきたはずである。ただし，今のところそれらが体系的にはまとめられていないようである。その集大成がなされるとき，ある教科のイノベーションが他の教科のイノベーションのヒントを提供するということも起こってくるであろう。第2世代については，教科の共通性は低いかもしれない。平成8年度広島大学教科教育学会第1回例会(1996年5月25日）において，数学教育学の小山正孝先生より，数学的な思考をする際，情報処理速度は問題とならず，むしろ十分な時間をかけて考えることが望ましい，というような指摘をいただいた。筆者は，概ねそのご指摘は理解でき，この点では英語と数学に基本的な違いがあるように思う。しかし，第2世代は，主として既知の項目の円滑な情報処理を問う。しかるに，算数・数学学習で低次レベルの情報処理の遅延が高次レベルの思考に影響を及ぼさないか。計算障害（dyscalculia）の子どもは，低次の情報処理障害のため高次のステップを踏めないケースが少なくないように思われる（Farnham-Diggory, 1992)。これは，難読症（dyslexia）の場合と類似している。理解力は備わっていても，文字という低次レベルの情報処理の遅延のために本来の能力を発揮できない場合が考えられるのである。このような観点から算数・数学能力と英語力の評価の共通性は探求する余地があるのではなかろうか。

第3世代は，未知の世界である。問題解決の前後の脳の活動状態の解明は，問題の難易，解決の手順，ストラテジーの違いなどの効果をより鮮明に示してくれるものと思われる。

なお，local evaluation と global evaluation という2極的な捉え方が通用しない教科があるかもしれない。もしそうなら，そのような観点からこの2極的な捉え方の妥当性を詳しく検討することが求められよう。

参考文献

Buchwald, J. S., Guthrie, D., and Schwafel, J. (1994). Influence of language structure on brain-behavior development. *Brain and Language*, 46, 607-619

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris Publishers.-

Farnham-Diggory, S. (1992). *The Learning-disabled child.* Cambridge, Mass.: Harvard University Press.

Perfetti, C. A., and Hogaboam, T. Relationships between single word coding and reading

comprehension skill. *Journal of Educational Psychology*, 67, 461-469.

Radford, A. (1988). *Tranformational grammar: A first course.* Cambridge: Cambridge University Press.

Rubenstein, H., Garfiled, L., and Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9, 487-494.

Sakaki, T. (1994). *Relationships between reading comprehension skill and auditory word recognition.* Unpublished BA thesis, Hiroshima University.

Yamada, J. (1991). The discrimination learning of the liquids /r/ and /l/ by Japanese speakers. *Journal of Psycholinguistic Research*, 20, 31-46.

山田　純・高島裕臣（1996）．聴解力と語部分聴覚呈示条件下の反応．第22回全国英語教育学会仙台研究大会口頭発表．1996年8月2日．

（受理　1996年8月30日）

[Abstract]

Innovations in the Testing of English Language Competence

Jun Yamada

Several ideas and innovations in the testing of English language competence are reviewed with reference to the innovations in related disciplines including linguistics and psycholinguistics. Two main streams, or generations, are highlighted. The first pertains to the measurement of linguistic knowledge (e.g., deep structure) and the second involves the measurement of efficiency of language use (e.g., naming latency). While these generations of testing appear to have developed independently for various purposes, the issues concerning relationships between them must be resolved in order to offer a comprehensive framework to testing theories. The possibilities for producing many more innovations are also discussed in the light of recent neurolinguistic techniques, the application of which may lead to the third generation in the future. Finally the implications for testing of the learner's abilities in other academic subjects are briefly discussed.