第1章　英語学力評価論

1. 言語テストにおける構成概念

テストは「個人の行動の特有のサンプルを導き出すために計画された測定のための道具」（池田・大友（監修）, 1997: 24）である。言語テストにおける研究対象は

言語技能や言語運用能力の測定(measurement)
言語能力の査定(assessment)
テストと教育行政の関係を探る教育評価（evaluation）
テストとその結果の省に関する倫理規定の整備・・・など

である。

言語テストとそのデータを用いてそれ自身を評価・判断するために、信頼性(reliability)・妥当性(validity)を確保する必要がある。*Longman Dictionary of Language Teaching & Applied Linguistics* (2002)ではこれらの要素を

Reliability: …a measure of the degree to which a test gives consistent results. A test is said to be reliable if it gives the same results when it is given on different occasions or when it is used by different people. (p. 454)

Validity: …the degree to which a test measures what it is supposed to measure, or can be used successfully for the purposes for which it is intended. A number of different statistical procedures can be applied to a test to estimate its validity. Such procedures generally seek to determine what the test measures, and how well it does so. (p. 575)

と定義づけられている。信頼性はそのテストが測定しているもののスコアがどれだけ安定し、信頼のおけるものであるか示す数値であり非常に重要なものである。同様にテスト作成の際に重要視される指標が妥当性である。妥当性検証によってテスト得点と測定対象能力の関連性が示されるため、あるテストが何の能力を測定しているかを示すために妥当性はテストを論ずるに不可欠な要素である。英語のテストの場合、「英語力」という構成概念(construct)が定義されていなければ何を測定しているか不明瞭なテストになる。そのため、言語能力（e.g. 語彙サイズ、文法力）の範囲を限定し、その構造を明示し、それらの構成要素間の関係付けを理論化させた構成概念の定義(construct definition)を明確にしなければいけない。

言語能力を検証するためには「言語能力」を理論づける必要があるが、その際にはその構成概念を定義づけなければならない。その理論は「個々の事象の背後に潜む規則性を発見し、構成概念を用いて、それらを形式化された法則として提示し、個々の事象をそのような一般的法則の論理的帰

結として説明する」ものである。理論づけのためには検証可能性(verifiability)が満たされることが必要条件であり、それを確保するために精密性(accuracy)、無矛盾性(consistency)、包括性(inclusiveness)、簡潔性(conciseness)を満たす必要がある。テストの実用の際に教員が使えるという条件も加わるため、簡潔性は重要であり、テストの実用性(practicality)に直結する。

測定するためには抽象的な構成概念（言語理論）を具体的に明示しなければならない。コミュニケーションへの関心・意欲・態度というのでは抽象的すぎる上、観察者（採点者）によって評価の手法や基準が異なる（採点者間信頼性の欠如）場合があり、同一の観察者の場合でも一貫した基準で観察することは難しい（採点者内信頼性の欠如）。具体的なタスクを設け、それを学習者（受験者）に行わせることではじめて構築された言語理論によって定義された言語能力を検証することができる。タスクを通じ、受験者の能力を推察(inference)し、それがどの程度適切か確認する作業を妥当性の検証(validation)という。現状では英語学力＝英語を使える能力と認識されているが、測定の観点からは言語に関係する知識（文化等）を排除することはできない。McNamara(1996)は構成概念の定義をする際、1) 言語の知識を習得すること、 2)特定の場面で実際に使えるようになること　としている。

英語をはじめとする言語教育における構成概念（＝言語理論）は、時代・社会の要請を受けて常に形を変えており不変のものではない。そのため、構成概念の定義づけ、テストの妥当性検証は不可欠である。

2. 英語学力論前史

明治初期には数学をはじめとするすべての学科は英語で授業が行われ、教科書も英語で書かれたものであった。そのため学習者は content-base で否が応でも英語を学ばざるを得なかった。しかし、その後翻訳された教科書、日本語で教授する指導者が増えたことから学習者の英語能力が徐々に高水準を保持できなくなってきた。「戦前の高校・専門学校入試英語のレベルは明治・大正期の頻出教材であった英検 1 級レベルに相当する」（小篠・江利川, 2004）とあるよう、明治初期に比べ徐々に英語学習者の平均的能力が低下していることは否めない。そのため、昭和初期には英語廃止論が唱えられるようにもなった。

明治期の入試英語では口述試問、読解、書き取り、英作文、ディクテーション、英会話のようなタスクが設定され、包括的な英語能力の測定が行われていたことがわかる。

	Speaking	Listening	Reading	Writing	Grammar
慶應義塾	○	○	○		
長崎英語学校	○	○	○	○	
商業学校入学試験	○	○	○	○	○
高等商業学校予科	○	○	○	○	○
東京師範学校	○	○			

3. 学習指導要領は英語学力をどのように定義してきたのか

学習指導要領は一般的な目標や内容を書いたものなので、抽象的であいまいである側面があるが、その反面、一貫して内容が通じるものが多い。そのため言語能力測定のための構成概念の定義、テスト作成のための捜査課などに資するところが多い。以下のように、およそ 5~10 年間隔で学習指導要領の改訂が行われているが、徐々に測定可能な構成概念の定義がしやすい学習内容が規定されるようになってきている。

改訂年度	目標・内容・変更点	語彙数（中学）
1947	英語で考える習慣を作り、英語の聞き方、話し方、読み方と書き方を学ぶ 英語を話す国民について知ること、その風俗習慣および日常生活について知ること	設定なし
1951	4 技能の基礎及び基準を身に着ける、4 技能の会議脳として、教師またはほかの生徒が話す簡単な英語の語・句・依頼・文に対して動作・ことばまたは両方で正確に答える能力、例にならって英語の音を発する能力、口頭で自分のものにした事柄に基づいている教材を読む能力	設定なし
1956	英語圏の生活や文化への理解を通して、自らの供用を高め、我が国の文化の向上を図ろうとする態度を養う	1800　語～3000 語
1958		
1969		
1977		
1987		
1998		機能語 100 語 総語数 900 語

1977 年改訂を経て授業時間の削減のため、指導内容、言語材料の大幅な削減が行われた。さらに、「初歩的な」という文言が頻出するようになった。1987 年改訂により、「実践的コミュニケーション能力」の養成が主目標となり、「外国語で積極的にコミュニケーションを図ろうとする態度を育てるとともに、活発な言語活動を促すこと、言語や文化に対する関心を深め、国際理解の基礎を培うこと」とされたが、実際には言語材料の削減の傾向が顕著であった。1998 年度改定では指定語の削減が代表例であるように、指導内容が削減された。

1990 年代には、態度、表現、理解、知識から学習評価を行うようになり、1)コミュニケーションへの関心・意欲・態度　2)表現の能力　3）理解の能力　4）言語や文化についての知識・理解　という観点が設定され具体的に評価の観点が述べられたため、言語使用能力の定義づけに進展が見られた。言語外の知識（4 技能、語彙など）と言語外の知識（文化・風俗など）と構成概念の定義づけに必要な材料はそろっているものの、「英語運用能力」の定義、各要素を関連付けるための理論がないこと

が現状の課題である。

4. 言語評価研究における言語能力

言語能力（構成概念）の定義は数多く行われてきており、近年では Chalhoub-Deville(2005)や Bachman(2007)らが言語能力の枠組みを提案しており、Chalhoub-Deville は 1) ability-based, 2) hierarchical, 3) performance-based 4) interaction-and discourse-based の 4 つを設定しており、Bachman(2007)は 1) skills and elements, 2) direct testing, 3) pragmatic language testing, 4) communicative testing, 5) interaction ability, 6) task-based performance test, 7) interactionism の 7 カテゴリーを設定している。一方、Chapelle & Brindley (2002)は the ability approach と the performance approach の 2 種類を主張している。