

# NTT の実践的利用：

## 2 段階モデルによる英語プレイスメントテストの分析

○木村哲夫

新潟青陵大学看護福祉心理学部

### 1. はじめに

入学生してくる学生の英語の基礎能力の差は非常に大きく、能力別クラス編成を実施したり、リメディアル教育を提供する必要性に迫られている大学は多い。そのための基礎データとしては、入試の英語得点を使う場合もあるが、推薦入学のように英語等の教科試験を受けずに入学してくる学生もいるため、入学時に何らかのプレイスメントテストを全員に実施する場合もある。短時間で実施して短期間に結果を得ることのできるテストはいくつか開発されてきているが、それらはみな連続した間隔尺度上のかなり細かいスコアで示される。スコアをいくつかの階級に区切って、その階級の英語の力を記述した **Can-Do-Chart** が提供されることもあるが、その根拠があいまいであったり、対象者の能力レンジに対して設定されている階級の幅が広すぎて利用しづらい。そもそも、入学生を英語基礎力別にいくつかのクラスに分けることが目的ならば、細かい間隔尺度上のスコアは必要なく、順序尺度上での測定・評価ができれば十分である。そう考えると、テストデータから潜在ランクを分析するニューラルテスト理論(**Neural test Theory: NTT**)によるプレイスメントテスト開発と実施は、まだ事例がほとんどないものの、理にかなったものである。

**NTT** の応用可能性については、項目特徴を示す指標(熊谷, 2007)、入試判定への利用の検討(橋本・荘島, 2008)、等化の精度(荒井・橋本・荘島)、推定値の一貫性の検証と小論文データへの多値型 **NTT** の適用(宇佐美, 2009)など、様々な角度から研究が行われている。

### 2. 目的

本研究の目的は、**NTT** に基づくプレイスメントテストの作成と実施を試み、その信頼性と妥当性を評価した木村(2009)をもとに、異なる潜在能力(たとえば語彙力と読解力と聴解力など)を総合して、求めた英語基礎力を推定する方法を探ることである。語彙・文法問題(**Vg**)、ダイアログを使ったリスニング問題(**Dlg**)、モノログを使ったリスニング問題(**Mlg**)の3種類の問題ごとに、**NTT** により順位数(**Q**)を10として3つの潜在ランク(**R<sub>vg</sub>**, **R<sub>dlg</sub>**, **R<sub>mlg</sub>**)を分析し、3つのランク数の和を求めて簡便的に総合的な英語基礎力とした。本研究では、**NTT** の拡張モデルの一つである段階ニューラルテスト(**graded neural test, GNT**)モデル(荘島, 2008b)によって総合的な英語基礎力を求め、木村(2009)の場合とどのような違いがあるかを検討する。

### 3. プレイスメントテスト

テストに使用された項目は、日本英語検定協会に使用許可を得て、英検 3 級から準 1 級の過去問の一部を利用した。木村(2009)では 222 名の大学生の協力を得て予備テストを行い、CTT の枠組みと、IRT の枠組みの両方の指標を利用するとともに、NTT において各項目のふるまいを見る項目参照プロファイル(item reference profile, IRP)により、misfit 項目と misfit 受験者の除去を行い、Vg32 問、Dlg13 問、Mlg19 問からなるプレイスメントテストを作成した。語彙文法問題とリスニング問題が同数となり、すべてを 60 分程度で終了できるものとなった。

予備テスト参加者のうち、これら 64 項目すべてに解答している者(75 人)のデータをもとに、問題種類ごとに  $Q=10$  の NTT による潜在ランクをもとめ、そのランク数の和をもとに疑似的にクラス分けを試み、良好な結果を得ている。同じデータについて、1PLM による能力推定と単純な正当数との間にも強い相関(.90 ~ .99)が認められた。また、データ数は少ないが、商用英語能力試験(TOEIC Bridge と CASEC)との間にも強い相関係数(.80 ~ .89)が認められ、作成されたテストの妥当性が確かめられた。さらに、このプレイスメントテストは、今年度ある短期大学に入学した学生に対し実施され、実際の英語のクラス分けの資料として十分機能することが確認された。

#### 4. 2 段階モデルによる分析

木村(2009)では、簡単に利用可能なソフトがなかったため、問題の種類ごとに NTT による分析を行ったのち、ランク数の和を総合力とし、それに基づいてクラス分を行ったが、今回は、簡単に利用できるソフト *exametrika* を利用して、問題の種類ごとに NTT による分析を行ったのちに、GNT により総合力の分析するという 2 段階モデルによる分析を試みた。

結果はランク数の和によるクラス分けと順位相関係数で .90 ~ .93 と、おおむね一致していたが、 $Q=10$  の分析では弱順序配置条件が満たされず、 $Q=5$  の分析において 2 ランク以上異なるケースが数例あった。詳細については、大会当日口頭にて発表し、考察を加える予定である。

今後は、ひとつの読解問題に複数項目が存在する問題についても GNT を適用して読解力の潜在ランクを推定するとともに、等化により項目数を増やし、アダプティブなテスト実施を可能にするとともに、潜在ランクの推定値の一貫性についても検討を加えたい。

#### 参考文献

- 荒井清佳・橋本貴充・荘島宏二郎 (2008).ニューラルテストモデルにおけるテスト等化の精度について, 日本テスト学会第 6 回大会発表論文抄録
- 橋本貴充・荘島宏二郎 (2008). 事前分布を伴うニューラルテスト理論を用いた選抜, 日本テスト学会第 6 回大会発表論文抄録
- 木村哲夫 (2009). ニューラルテスト理論による英語プレイスメントテストの作成と評価, 関東甲信越英語教育学会紀, 23, 23-34.
- 熊谷龍一 (2007). ニューラルテスト理論を離散変数型 IRT とみなしたとき項目特徴を示す指標について, 第 1 回ワークショップ「ニューラルテスト理論」(於 大学入試センター).
- Shojima, K. (2007). The graded neural test model: A neural test model for ordered polytomous data. *DNC Research Note*, 07-03
- Shojima, K. (2008). Neural test theory: A latent rank theory for analyzing test data. *DNC Research Note*, 08-01.
- 宇佐美慧 (2009). ニューラルテスト理論の応用可能性—方法論的課題の考察と多値型モデルの適用例—, 日本テスト学会誌, 5, 66-79.