
プロジェクト報告

ビジネス・コミュニケーション能力を測定する スキルテストの開発

田崎勝也 (青山学院大学国際政治経済学部)
芳賀日登美 (筑波大学大学院)
宮原哲 (西南学院大学文学部)
申知元 (青山学院大学国際政治経済学研究科)

1. 目的

かつて社会心理学者のクルト・レヴィン (Lewin, 1951) は、人間の行動は環境と特性の関数 ($B=f(E, P)$; B =behavior, E =environment, P =personality) によって示されるとした。レヴィンは、人と環境から構成される全体性をその人の生活空間と位置づけ、認知や態度などの心理的要因から、物理的な環境さらには規範や制度などの社会的要因も含む社会的現実のすべてが行動を司る要因となると説いた。コミュニケーション能力をどう定義するかに関しても、特性重視-状況重視の軸にそって様々な見解が存在するが、本プロジェクトで開発を試みる C-Exam では、特性と状況との両側面からコンピテントなビジネス・コミュニケーション能力を考える。田崎・申 (2015) では、企業の取締役や人事担当者などへの面接調査結果に基づき、コミュニケーションの普遍的な側面に着目し、ビジネス・コミュニケーションにおける心理的な側面について、5コンセプト18因子・126項目から構成される心理特性テストを開発した。本論文では、より場面依存的なコミュニケーション能力を射程とし、日本のビジネス・シーンに特化した「スキルテスト」の開発の経緯およびテストの信頼性・妥当性の検証結果を報告する。

さて学力や知能、不安や幸福感、社会階級や中流意識など、教育学・心理学・社会学などで扱う指標の多くは構成概念 (constructs) と呼ばれ、実体を持たないいわば仮説的な構成体である (村井, 2009)。コミュニケーション能力もこうした構成概念の一種と考えることができる。構成概念の測定にはさまざまな方法があるが、紙と鉛筆を前提とした測定法では、段階的な評価基準によって心的傾向測定するリカート法や学力診断

で用いられることの多い多肢選択方式が一般的である。前者はたとえば、質問項目に対して「5 = 当てはまる」から「1 = 当てはまらない」などの基準を基に評価するため、正答がないような態度や適性などの測定に適しており、田崎・申（2015）でも、コミュニケーション能力の心的側面を測定するために使用した。一方後者は、設問の選択肢に正答－誤答があり、たとえば学力テストのように、正答を導きだすことができるような対象の測定に向いている。本プロジェクトで開発をすすめるスキルテストでは後者の方法を用いて、コミュニケーション能力を測定するテストを開発する。日本のビジネス場面や商取引の現場といったように、より詳しく状況設定をすることで、正答や誤答を見極め易くなることが多肢選択方式を採用した背景にある。心理特性テストではビジネス・コミュニケーションの心的な側面に着目し、コミュニケーション能力の一般的な傾向を測定することを目的とした。一方、本調査で開発を試みるスキルテストでは、特定の状況にその能力は限定されるものの、日本のビジネス・シーンにおけるコミュニケーション能力をより詳しくかつ客観的に診断することを目標とする。

2. 方法

設問の作成

スキルテストの設問は、心理特性テスト（田崎・申，2015）の質問項目と同様、主に芳賀・宮原・田崎・申（2015）によるインタビュー調査の結果を基に作成した。芳賀他は、企業人が現代日本のビジネス場面でどのようなコミュニケーション能力を求めているのかについて、企業の取締役もしくは人事担当責任者に面接調査を実施した。調査の結果「コミュニケーションの双方向性」「発信力」「アクティブ・リスニング」「コミュニケーションの基盤」「率直性」「曖昧さの回避」「ビジネスの場での報告の重要性」の7つのテーマが、特に「若手」のビジネス・パーソンに求めるコミュニケーション能力として抽出された。

これら7つのテーマを基盤として、適宜コミュニケーション理論の先行研究を参考にしながら（宮原，2006；末田・福田，2003），実際にコミュニケーションの問題が顕著化するビジネスの場面を想定して125問の設問を作成した。ビジネスの多様な状況を配慮し、たとえば、設問内容は「グループ場面でのイニシアチブ構築力」「2者間での話しの展開力」「複数の相手を対象にした話しの展開力」「ビジネス場面で期待される適応力と判断力」「報告場面でのまとめ力」「ビジネスマナーを含む自己管理能力」などを含み、測定対象および場面設定ともに実際の日本のビジネス・シーンを反映した網羅的かつ実践的な内容とした。質問紙調査を実施するに当たっては、著者間で設問の内容的妥

当性を検討し、最終的には妥当性が認められた100設問のみを質問紙に挿入しデータを収集した。

前述のように、スキルテストの設問は多肢選択方式を採用し、各設問に対して (a) (b) (c) (d) の4つの選択肢から正答をひとつ選ぶ形式を採った。4択からなる選択肢の作成に際しては、コミュニケーション能力が高い受検者がしっかりと正答を導出できるように、設問の識別力を高めるように心がけた。識別力とは項目機能の一種で、項目や設問がどの程度敏感に高い能力をもつ受検者を「高く」、低い能力をもつ受検者を「低く」判別する力があるかを示し (田崎, 2008)、設問ひいてはテスト全体の精度を決める重要な要因となる。具体的には、ビジネス場面における慣習や規範に目を向け、受検者の職種や分野を問わず、正答を導けるように配慮した。また誤答の選択肢には、設問の問いに対して、正答になるコミュニケーション行動には関連するが正答ほど対応が十分でない例や、一般的には正しいがビジネス場面では間違いとされる事例などを積極的に含めるようにした。設問の草稿には、著者だけでなく、さまざまな職域のビジネス・パーソンから意見を求め、正答-誤答の関係性がクリアになるまで修正を繰り返した。

図1は「グループ場面でのイニシアティブ構築力」を問う設問の例である。正答 (ベストアンサー) はbになる。理由としては、ビジネスにおいてグループ力を発揮するためには、限られた時間の中で、最大の効果を上げることが年頭に置く必要があり、メンバー全員の力を上手く結集させることを重要になる。そのため、メンバーの一人ひとりの個性を理解した上で、個々の優れている点を引き出す工夫が求められる。a, c, dは誤答になるが、それぞれ根拠としては、(a) ビジネスでは効率よく結果を導くことが求められるため、時間を考慮せず全員の意見が一致するまで話し合うことは効率を無視した行動となる、(b) 多様性は重要だが、人数が多いと混乱を招く原因となるため、複数で仕事をする場合には適切な人数を考慮する必要がある、(c) 意見の対立を避けることは、良い結果に結びつかない。時間をかけすぎるのは問題だが、決められた時間内で参加メンバーが互いの意見を充分交換することは重要、などが考えられる。

分析手順と調査協力者

在京の日系および外資系企業

Q. グループで仕事をする際、次のうち最も意識しておくべきことはどれですか。

- a. 効率よりも中身が重要なので、時間をかけてでもメンバー全員の意見が一致するまで話し合うことが重要である。
- b. 最少の時間で最大の効果を達成するために、メンバーの一人ひとりの力を最大限に引き出す工夫が重要となる。
- c. 情報量が増えるのでメンバーの数はできるだけ多い方が良い。
- d. 意見が異なり対立が起こりやすいので、できる限りグループ全員での作業は避ける。

図1 設問の例

に勤務する社員342名¹⁾から収集したデータを用いて、設問の最適化および信頼性・妥当性の検討を行った。コミュニケーション能力テスト C-Exam は、測定対象を入社時から入社後10年目までの「若手」を想定したテストである。したがって、職務経歴10年未満のビジネス・パーソンを対象に協力を依頼しデータを収集した。結果、平均勤続年

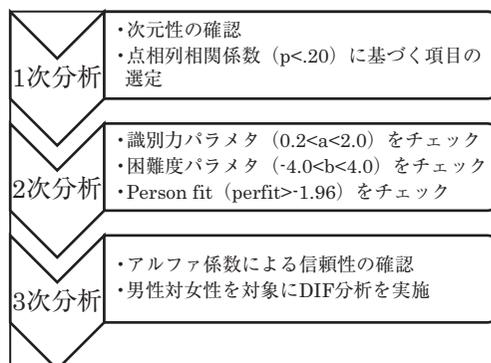


図2 分析のステップ

数は6.09年 (SD=5.66) となり、男女比は男性が264名 (77.2%)、女性が78名 (22.8%) となった。ところで、心理特性テストでは学生データとビジネス・パーソンデータの双方を用いて項目の最適化を行ったが、スキルテストの項目分析にはビジネス・パーソンからのデータのみを使用した。理由として、スキルテストはビジネス場面を射程としたコミュニケーション能力を測定するため、職務経験のない学生データを用いても分析結果に十分な妥当性を得られないと判断したためである。

図2はスキルテスト設問の最適化および信頼性・妥当性の検証についての分析計画である。分析の手順は3ステップから構成され、1次分析では要約統計量を用いた基礎的な項目分析を、2次分析では項目反応理論を用いた項目分析を、そして3次分析では最終項目案に対して信頼性・妥当性の検証を行った。

項目反応理論

冒頭でも述べたように、本調査では多肢選択方式を用いてデータを収集した。こうしたデータは正答-誤答のある2値型データとなるため、本調査では項目反応理論 (Item Response Theory, 以下IRT) を用いて分析を行った。IRTは新しいテスト理論と呼ばれ、近年発展著しいコンピュータを利用した適応型テストでもその理論的枠組みとして用いられている。TOEFLのCBTに代表されるように適応型テストでは設問の解答ごとに受検者の能力を推定し、アップデートされた能力推定値に基づき適した難易度をもつ設問を提示するといった具合にイテレイティブな測定が行われる。受検者は一人ひとり異なる能力に適した難易水準をもつ設問のみに解答することで、テスト全体で精度を考えた古典的テスト理論によるテストと比べ、より少ない設問でより精度の高い測定が行えるといわれている。異なる設問の正否を基に推定した受検者の能力が比較可能性をもつのは、設問の難易度と受検者の能力値が同一尺度上で定義されるため、IRTの

1) 最終的には574名のビジネス・パーソンからデータを収集したが、本分析は2013年8月時点での342名を対象とした。

優れた特性のひとつとして挙げられている（並川・谷・脇田・熊谷・中根・野口，2012）。その他，古典的テスト理論による測定と比較してIRTの特長として議論されるのは（a）テスト特性と被調査者特性を分離することができる，（b）被調査者の標本に依存しないテスト開発が可能，（c）一部の項目だけを利用して信頼性の高いテストが作成できる，（d）潜在特性レベルに応じたテストの精度（信頼性）を考えることが可能，（e）テスト情報量（フィッシャ情報量）から測定誤差を推定することができる，などである（阿久津，2008；豊田，2002）。

IRTでは，受検者の設問への反応（正答率）は潜在的な能力（潜在特性値）の関数として示され，その関数が具体的にどのようなものかは各設問の性質によって定義付けられる（渡部，1993）。正答率と潜在特性値の関係はロジスティック関数を用いた項目特性曲線（Item Characteristic Curve，以下ICC）によって表現されるが，ICCは設問の性質を示す困難度（difficulty），識別力（discrimination），当て数量（guessing）といった項目パラメタによってさまざまな形態をとる。

たとえば，潜在特性を θ で示し，項目 j の困難度を b_j とすると，項目 j に正答する確率 $P_j(\theta)$ は，

$$P_j(\theta) = \frac{1}{1 + \exp(-D(\theta - b_j))} \quad (1)$$

と書くことができる。Dは尺度因子と呼ばれる定数で，通常1.7に設定される。関数にエクスポネンシャル（ \exp ）が入った式(1)はロジスティック・モデルでDの尺度因子を含めると，正規分布と対応関係にある正規累積モデル（normal ogive model）に近似することが知られている（豊田，2002）。式(1)では θ と b_j がゼロになるとき，つまり受検者の能力と設問の難易度が均衡状態のとき，その設問に正答する確率が50%になることを示している。このモデルは

項目の困難度のみがモデル化された1パラメタ・ロジスティック（1PL）モデルと呼ばれる。図3は $b_j = -1.0$ ， $b_j = 0.0$ ， $b_j = 1.0$ とそれぞれ異なる困難度をもつ3つの設問のICCを示したものである。たとえば，50%の正答率を出すために受検

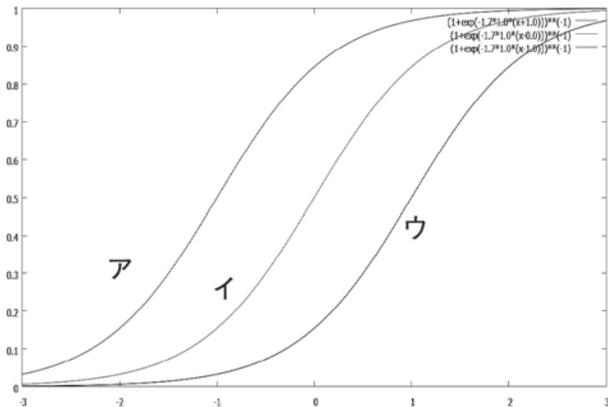


図3 1PL（ア： $b_j = -1.0$ ，イ： $b_j = 0.0$ ，ウ： $b_j = 1.0$ ）

者はアの $b_j = -1.0$ の設問では $\theta = -1.0$ 程度の能力が、イの $b_j = 0.0$ の設問では $\theta = 0.0$ 程度の能力が、ウの $b_j = 1.0$ の設問では $\theta = +1.0$ 程度の能力が必要になる。つまり 1 PL では曲線が左に位置するほど易しく、右に行くほど難しい設問となる。

困難度に加えて、項目の識別力がモデル化された IRT は 2 パラメタ・ロジスティック (2 PL) モデルと呼ばれ、心理測定で最も頻繁に用いられるモデルである。2 PL モデルではある設問に正答する確率は、以下の式で示される。

$$P_j(\theta) = \frac{1}{1 + \exp(-D a_j (\theta - b_j))} \quad (2)$$

右辺の分母をみると、1 PL と比較して θ と b_j の差異が a_j だけ拡張されていることがわかる。図 4 は $a_j = 0.5, a_j = 1.0, a_j = 2.0$ と異なる識別力をもつ 3 つの設問の ICC を示したものである。3 設問の中では $a_j = 2.0$ のウの ICC が $\theta = 0$ 前後で最も急峻に立ち上がっている

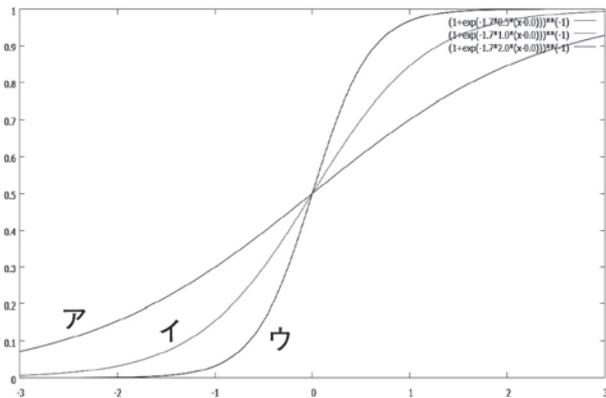


図4 2 PL (ア : $a_j = 0.5$, イ : $a_j = 1.0$, ウ : $a_j = 2.0$)

ことがわかる。識別力に違いによって、正答率にも大きな変化がみられる。たとえば潜在特性値が -1 から $+1$ の範囲での正答率の推移を考えると、アの $a_j = 0.5$ の ICC では、約 30% から約 60% に、イの $a_j = 1.0$ の ICC では約 15% から約 85% に、ウの $a_j = 2.0$ では約 3%、では約 97% まで変化している。つまり、識別力の高い $a_j = 2.0$ では、潜在特性の差が最も顕著に項目反応 (正答率) の差となって現れていることがわかる。

1 PL, 2 PL のほかにも、偶然に正答する確率をモデル化した 3 パラメタ・ロジスティック (3 PL) モデルもある。困難度と識別力の各パラメタに加えて、当て推量パラメタが追加される 3 PL では、潜在特性値が低い受検者でも正答率はゼロにはならない。たとえば、4 択の多肢選択方式のテストでは、当てずっぽうに答えても解答の 25% は偶然に正答することになる。このため 3 PL の ICC は負の方向の漸近線がゼロには近づかず、0.25 付近から立ち上がることになる。

さて、心理テストの開発に IRT を用いる利点の一つに、潜在特性値のそれぞれの推定値に対して、各設問が貢献している情報の大きさを評価できることにある (渡部, 1993)。たとえば、古典的テスト理論ではテスト全体での測定精度を考えたため、クロ

ンバックのアルファ係数などの信頼性係数から測定誤差を推定することはできなかった（阿久津，2008）。一方IRTではテスト情報関数（Test Information Function，以下TIF）という指標から潜在特性値に応じた測定誤差を算出することができる。テスト情報量の平方根の逆数が潜在特性値の測定誤差に対応するので，テスト情報関数を $I(\theta)$ とすると

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (3)$$

となる。項目 j におけるテスト情報量は2PLの場合には，以下のようになる。

$$I_j(\theta) = D^2 a_j^2 P_j(\theta) [1 - P_j(\theta)] \quad (4)$$

TIFは潜在特性値の各点での測定精度を反映したもので，値が大きくなれば，それだけ精度よく測定されていることを示している（藤森，2002）。グラフに描くと，TIFの逆数として示される誤差曲線はTIFを反転した形をとる。

3. 結果と考察

基礎的な項目分析

企業家へのインタビュー調査結果およびコミュニケーション理論の先行研究を参考に作成し，内容的妥当性が認められた100設問に対して，まずは古典的テスト理論をベースとした項目分析を行った。設問の識別力と関連が高い，項目－合計相関係数を求めた。スキルテストの設問は正答－誤答をもつ2値型変数となるため，一方の変数が連続な量的変数，もう一方の変数が2値型の質的変数であるときの相関係数となる点双列相関係数（point-biserial correlation）を算出した。項目削除の基準は，豊田（2002）を参考にし， $r > 0.20$ に設定した。項目－合計相関による分析の結果，10設問で基準に達しない相関係数を示したため，これらを削除することとした。その他，設問の困難度と関係がある平均値や標準偏差などを検討した。平均値に関しては，最小値と最大値を考え，その両端から15%以内を極端に低い（高い）平均値と定義して確認したが（阿久津，2008），極端な値を示す設問は見つからなかった

IRTモデル仮定の確認

項目反応理論を用いた分析には，2つの仮定がある。一つ目は「局所独立の仮定」（assumption of local independence）と呼ばれるもので，項目の正答率は他の項目の正

誤に関係なく互いに独立していることを意味する。たとえば、項目*i*に正答する確率が項目*j*によって影響を受ける場合、この前提が満たされないことになる。もう1つは「一次元性の仮定」(unidimensionality)で、測定対象が一因子から構成されていることを意味す

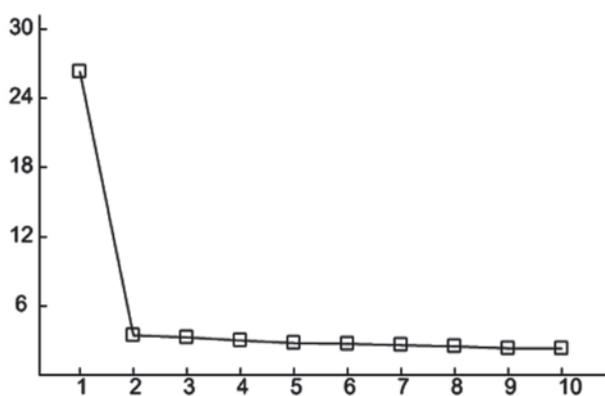


図5 スクリープロットの結果

る。近年開発が進む多次元項目反応理論ではこの限りではないが、通常のIRTでは一次元性を想定する。ただ実際には、前者は後者の仮定に包含されることから (Edelen & Reeve, 2007), 1次元性の仮定のみを確認する研究が多い。

こうしたことから本調査ではまず1次元性の確認を行った。2値型変数間の相関係数を求めたテトラコリック相関行列 (tetrachoric correlation) に基づく因子分析を行った結果、第一因子の固有値は27.227, 第二因子は3.321, 第三因子は3.068と続き、第一因子の固有値が他の因子のそれよりも突出して高いことから一元性が示唆された。スキルテストの1因子構造はスクリープロットの結果にも示されており、図5を見ると明らかかなように、固有値の多寡を示した折れ線グラフは第二因子から大きく落ち込み、その後ほぼフラットな状態で推移していた。

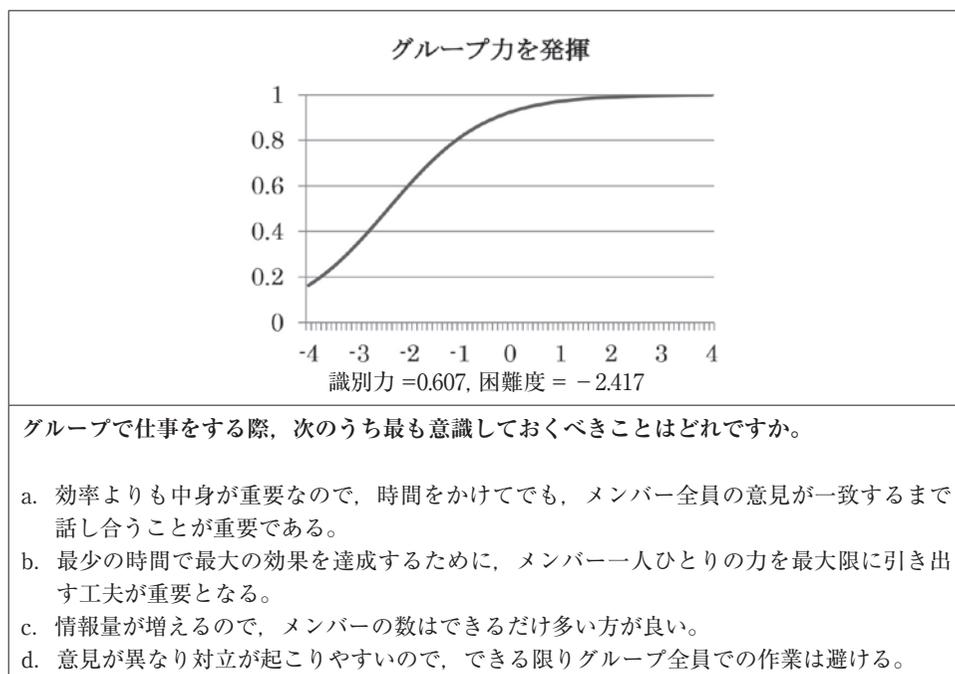
IRT のモデル

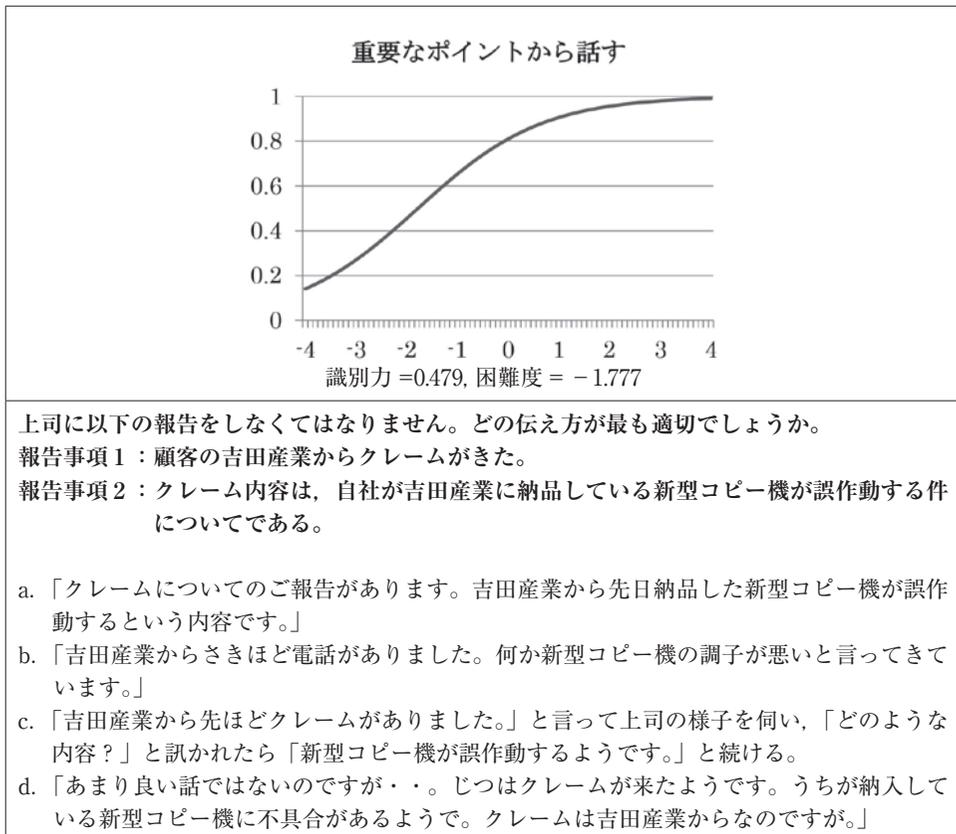
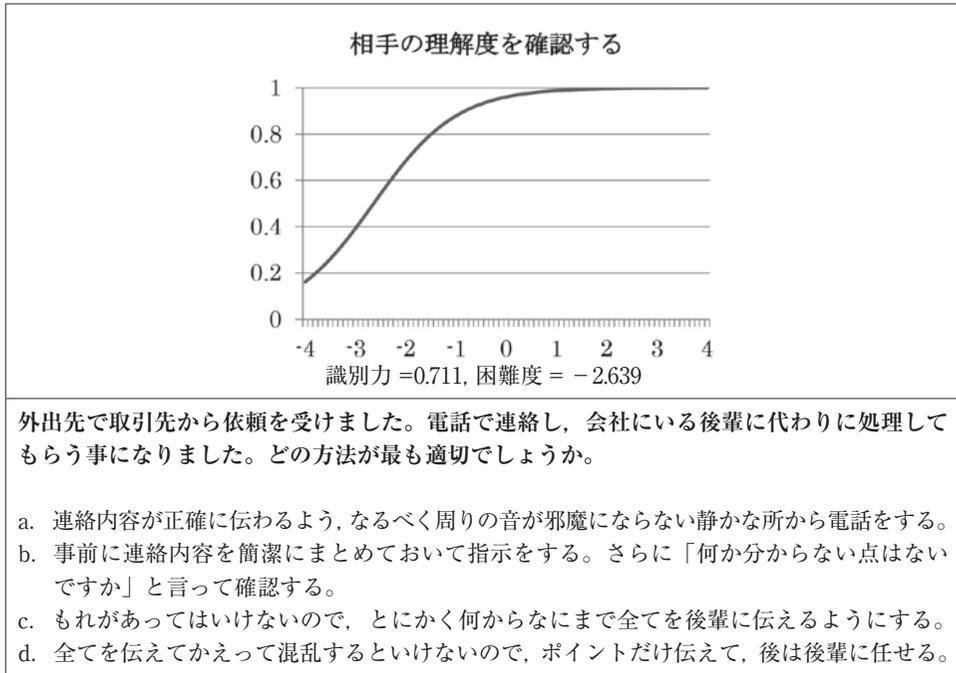
本調査では、1 PL, 2 PL, 3 PL の3つのIRTモデルを想定し、どのモデルがデータとの当てはまりがいいか、尤度比検定 (Likelihood Ratio Test, 以下 LRT) による比較を行った。LRTの結果、 -2LogLikelihood の値は、1 PLでは25,946, 2 PLでは25,645, 3 PLでは25,560となり、各モデルの -2LogLikelihood 値とその下位モデルの -2LogLikelihood 値が自由度の縮減 ($df=90$) に照らして有意に離れていたのは2 PLで²⁾、2 PLと3 PLの間には有意差は見られなかった。また各モデルのAICを確認すると、2 PLのAICは25,825で3つのIRTモデルの中で値が最も小さかった (1 PL: AIC=26,036, 3 PL: AIC=25,911)。これらのことから、2 PLが最もデータ分布の様子を説明していると判断し、2パラメタ・ロジスティック・モデルを最終モデルとして採用することにした。

2) 自由度90に対応するカイ2乗分布の有意水準5%の臨界値 (C.V.) は113.15である。

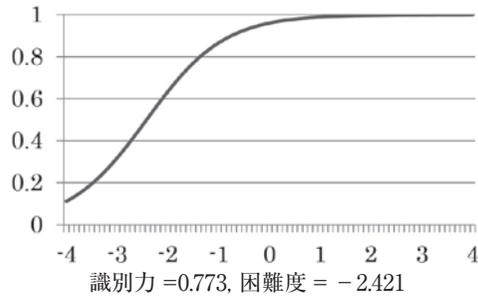
項目パラメタの推定

2PLに基づき、周辺最尤推定法を用いて、識別力 (aパラメタ) と困難度 (bパラメタ) を算出した。推定には統計ソフト EasyEstimation (熊谷, 2009) と BILOG-MG (Zirnowski, Muraki, Mislevy & Bock, 1996) を併用した。識別力パラメタは0.17 から1.38で推移し、推定値の平均は0.60, 標準偏差は0.23となった。一方、困難度パラメタは-3.84から0.73の範囲で値をとり、推定値の平均は-1.74, 標準偏差は0.83であった。識別力パラメタの適性値に関しては研究者によって見解が異なるが、豊田 (2002) によれば、識別力パラメタが0.2以下もしくは2.0以上の値を示した項目に関しては削除を勧めている。また困難度パラメタの適性値についてもさまざまな意見が散見されるが、江口 (2011) は $-4.00 < b < 4.00$ が適当な範囲と主張している。本調査で得た bパラメタの推定値は適切な範囲内に収まっていたが、aパラメタに関しては下限値を下回っていた。くわしく調べたところ、設問24の識別力パラメタが0.17を示し、その他の設問の識別力はすべて $0.2 < a < 2.0$ の範囲内にあった。削除を含めて検討したが、設問24はプレゼンテーション場面での問いであり、当該の場面設定を行っている設問は他にはなかったことから、識別力がやや低いものの、削除は行わずに残すことにした。図6は6つの下位領域における代表的な設問の ICC である。





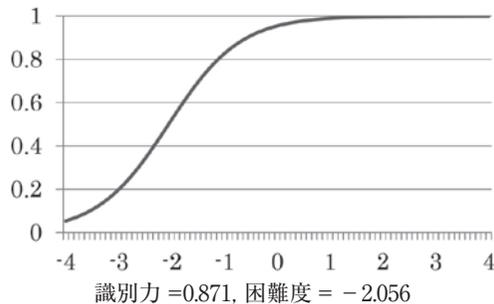
批判を積極的に受け入れる



初対面の顧客から、自分の会社についてかなり痛烈な批判を受けました。その際に、最も効果的だと思われる行動は以下のどれでしょうか。

- 批判は甘んじて受け、「本当に申し訳ありません。」と言って、ひたすら頭を下げ、何とか今後の取引に影響がでないように頑張る。
- まず意見いただいた事に感謝し、今後きちんと改善に結びつけていきたいと述べるとともに、会社の良い点も理解してもらおう機会とするよう努力する。
- 相手の批判に甘んじているのは営業にならないので、相手の批判を上手にかわして、会社の良い点を積極的にアピールする。
- そのような顧客と付き合うのは時間の無駄なので、とりあえず「これからも何卒宜しくお願いします」と言ってすぐに帰る。

ポイントを押さえ要領よく話す



課の上司とメンバーに対して、今回どうしても通したい企画案の説明をする際、次のうちどれに最も注意を払うべきでしょうか。

- メンバー間で意見の衝突などが起きないように、細心の注意を払って自分の発言を組み立て、質問などもなるべく出ない様にする。
- 自分の企画案がいかに自信のあるものかを十分に理解してもらうため、落ち着いて信頼感にあふれる話し方を心がける。
- まずは全体像がわかるような説明をし、大事なポイントはできるだけ早い段階で明らかにし、細かな点はその後順次説明をする。
- 聴いている人たちの様子を見て、あまり関心が無いようなら、準備していたストーリーを変えてでも、興味を持ってもらえる内容にする。

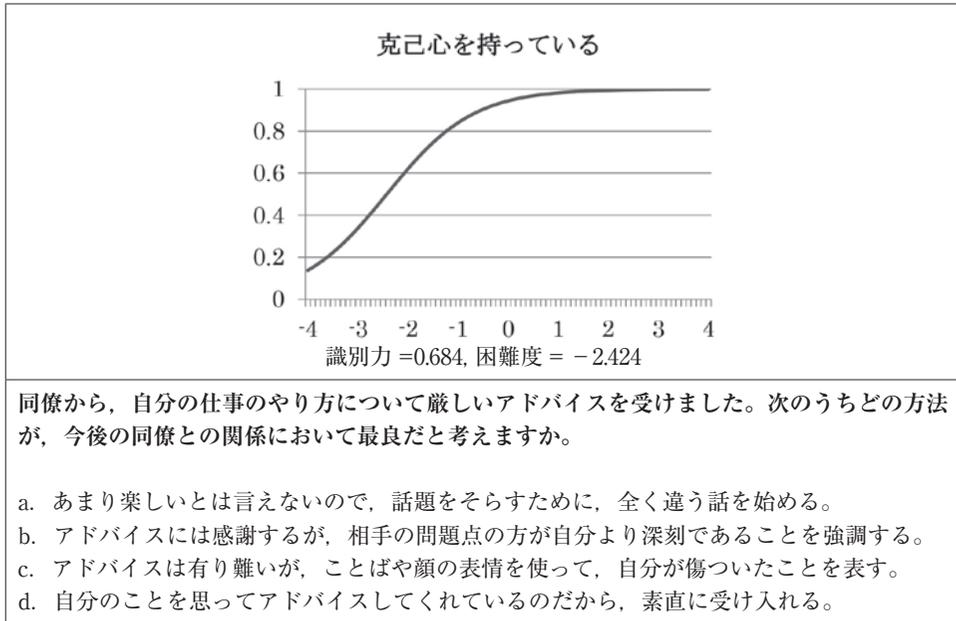


図6 代表的な設問の ICC

潜在特性値の推定

次に、確定した項目パラメタに従い、受検者の潜在特性値 (θ) の推定を行った。全問不正解もしくは全問正解をした受検者はいなかったため、最尤推定法 (MLE) を用いて潜在特性値の推定を行ったところ、 $-3.54 < \theta < 4.65$ となり、その平均値は 0.04、標準偏差は 1.09 となった。潜在特性値の分布は正規分布を仮定して推定をおこなったが、これらの要約統計量によりほぼ想定通りに推定されたことがわかる。IRT では個々の受検者がモデルへの当てはまりのよさを検討することもできる。PersonFit と呼ばれる指標を削除基準 (Embretson & Reise, 2000) に照らして検討した結果、 -2.0 の基準を逸脱した受検者は見つからなかった。また素点は最低点が 14 点、最高点が 89 点となり、平均値は 71.22 点、標準偏差は 12.48 点だった。図 7 は素点の合計点を縦軸に、潜在特性値 (θ) の横軸にとり、両者の関係性を描いたものである。グラフは右肩上がりの曲線となっており、ビジネス・コミュニケーションの能力が上がるほど、スキルテストの素点も高くなることを示している。

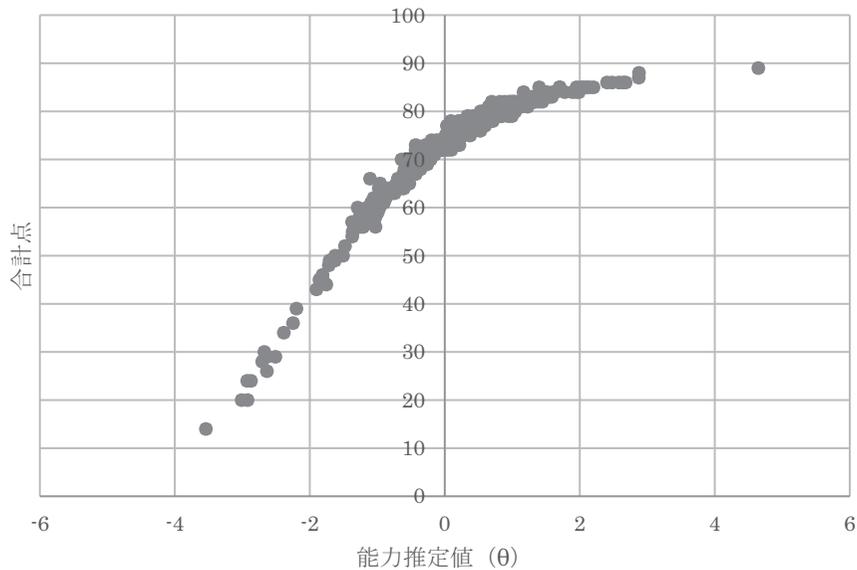


図7 素点と潜在特性値の関係

信頼性と妥当性

妥当性の検証は、心理特性テスト（田崎・申，2015）と同様に、バイアス項目の抽出を目的とした DIF 分析を中心に行った。スキルテストではビジネス・パーソンから収集したデータを用いて分析しているため、男性集団対女性集団を DIF 分析の対象とし、心理特性テストで行ったようなビジネス・パーソン集団対学生集団に対する分析は行わなかった。スキルテストでは項目反応が二値型であることからロジスティック回帰分析（Swaminathan & Rogers, 1990）を援用した DIF 分析を行った。一般線形モデルによる DIF 分析と同じように、合計点、グループ、合計点×グループを説明変数としてモデル化し、グループ変数が有意なら均一 DIF、合計点×グループ変数が有意なら不均一 DIF を疑った。設問ごとに分析を実行した結果、有意な結果は示されず、均一 DIF および不均一 DIF 項目ともに検出されなかった。田崎・申でも議論しているように、DIF 分析は妥当性の中でも結果的妥当性（Messick, 1995）および基準関連妥当性と関連があり、今回の分析でバイアス項目が検出されなかったことはスキルテストがこれらの妥当性を有する証左と考えることができる。

さて信頼性の分析では、まずクロンバックのアルファ係数を算出したところ、0.92 という高い内的整合性を示す値を出力した。既述のように、古典的テスト理論ではテスト全体で測定の精度を考えるのに対して、IRT モデルでは θ の各点における測定誤差をより詳細に検討することができる。図8で描かれている TIF から明らかなように、

潜在特性値が -2 付近で情報量が最大になっている。つまりスキルテストはコミュニケーション能力がやや「低い」受検者が受検した場合に、最も精度よく測定することができる。またテスト情報量の平方根の逆数によって示される測定誤差は -2 付近で最も小さくなり、 θ が $+4$ もしくは -4 に近づくに従って大きくなっていった。

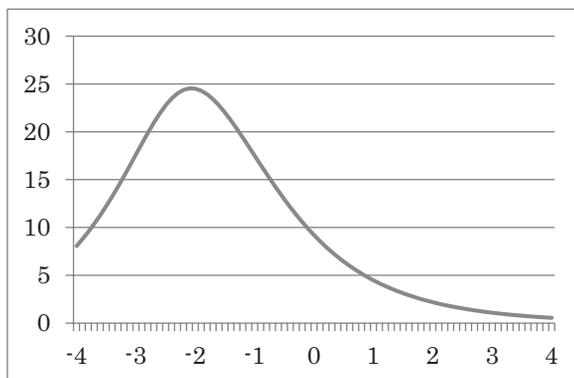


図8 テスト情報曲線 (TIF)

水平テストの作成

IRTでは、同じ概念を測定するテストではあれば、テスト間で得点を比較することが可能である(豊田, 2002)。こうした特性から、テスト内で一部の設問だけを利用して、別のテストを作成することができる。本研究では確定した90設問を基に、ほぼ等しい測定精度を持つ2つの水平テスト

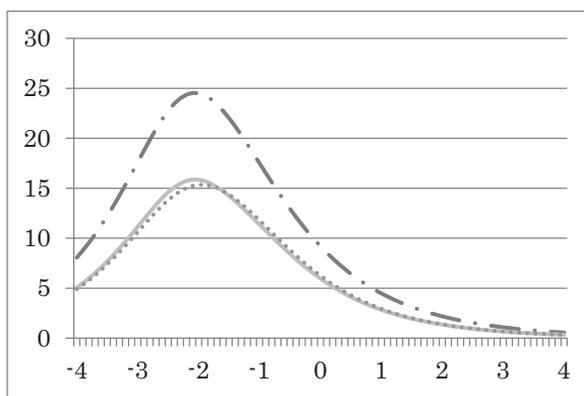


図9 フルテスト、水平A、水平BのTIF

ト(水平Aと水平B)の作成を試みた。まず設問の内容を吟味して、両テストの共通に含まれる24のアンカー設問を設定した。そして、残りの66設問をランダムに振り分け、57設問をもつ水平Aと水平Bを作成した。2PLによる推定を行ったところ、識別力パラメタおよび潜在特性値に関しては表1のようになった。各種パラメタを比較すると、いずれも類似した値が推定されていることがわかる。また図9は3つのバージョンのTIFを描いたものである。フルテストに比べて、項目数少ない水平テストは自ずとテスト情報量は下がるが、いずれのテストも同じような曲線軌道を描き、3バージョンともに潜在特性値が -2 付近で最も精度良く推定が行われている。クロンバックのアルファ係数を算出すると、水平Aでは0.89、水平Bでは0.88となり、フルテスト($\alpha = 0.92$)よりは若干下がるが、双方の水平テストとも高い信頼性を示した。

表1 フルテスト・水平A・水平Bの項目パラメタと潜在特性値

	フルテスト (90項目)	水平 A (57項目)	水平 B (57項目)
1. 識別力 (a)			
平均	0.60	0.61	0.60
標準偏差	0.23	0.24	0.22
範囲	0.17<a<1.38	0.18<a<1.48	0.19<a<1.18
2. 困難度 (b)			
平均	-1.74	-1.73	-1.78
標準偏差	0.83	0.80	0.81
範囲	-3.84<b<0.73	-3.61<b<0.70	-3.73<b<0.26
3. 潜在特性値 (θ)			
平均	0.04	0.07	0.07
標準偏差	1.09	1.18	1.16
範囲	-3.54< θ <4.65	-3.53< θ <5.19	-3.85< θ <4.70
4. 素点 (x)			
平均	71.22	45.16	45.32
標準偏差	12.48	7.97	8.18
範囲	14<x<89	9<x<56	7<x<56

4. 総合考察

本論文では、日本のビジネス場面でのコミュニケーション能力を測定する「スキルテスト」の開発の経緯を説明するとともに、テストの信頼性・妥当性の検証結果を報告した。日本企業のトップ・エグゼクティブへのインタビューを通じて浮かび上がった理想的なコミュニケーション像を基に、ビジネスの現場で実際に起こっている状況を加味して実践型のテストを開発した。各種基準をクリアした90項目は、項目反応理論を駆使した項目分析によって、十分な妥当性と信頼性を有していることが示された。

また水平テストの作成も本プロジェクトの目玉のひとつとして上げられる。ほぼ同様の測定精度をもつ2つのバージョンを持つことで、スキルテストはさまざまなケースの測定に対応できる。たとえば、複数回テストを受ける受検者がいる場合でも、設問への学習効果を抑えてテストを実施することができる。また社内で何らかの教育プログラムを実施する場合、事前-事後で異なるバージョンのテストを用いることで、キャリアオーバー効果などを回避したプログラムの効果を検証することが可能になる。IRT

では設問の難易度と受検者の能力値が同一尺度上で定義されるため（豊田，2002），こうした水平テストの開発が可能になったわけであるが，いずれにしても次元性の仮定や設問の項目パラメタなど，スキルテスト自体の基本性能が十分に満たされていたことが水平テストの作成を可能にした背景にある。

心理特性テストを含めた C-Exam 全体の信頼性と妥当性に関する検討は申・田崎（2015）に譲るとして，最後に「コミュニケーション能力は特性で決まるのか，状況で決まるのか」の議論に戻り，今後の課題も含めて考察してみたい。C-Exam ではコミュニケーション能力の普遍的な側面を測定する心理特性テストと日本のビジネス・シーンという状況下での場面依存的なコミュニケーション能力を問うスキルテストの2部門から構成され，本論文では後者のテストの信頼性と妥当性を検討した。スキルテストは日本のビジネス場面に特化した設問構成となっており，一般的な傾向を問う心理特性テストの項目に比べれば，より場面依存的なコミュニケーション能力を測定していることは自明であるが，一方でスキルテストが職種や業務形態ひいては各々の企業環境といった異なる状況や場面の個別性を十分に拾い上げているのかといった点に関しては疑問も残る。組織行動を測定する諸領域からは，個人や集団，組織の間で差異が大きいといった指摘がされている（渡辺，2003）。こうした観点から考えると，スキルテストもビジネス場面という状況設定はされているものの，程度の差はあるが心理特性テストと同じように，個別のビジネス・シーンの中で共有される通状況的な傾向を測定していることは否めない。

もちろんこうした問題に対処するためには，個別の状況や場面により特化した測定や分析を行えばよいことになるが事情はそうは簡単ではない。状況を詳細に設定し個別の特有性を追求すると，究極的には人それぞれ異なる独自のコミュニケーション行動の諸相が生まれ，測定対象が無限に拡大して収拾がつかなくなってしまう。組織ストレスに関する研究においても同様の問題が発生している（矢富・渡辺，1995）。たとえば，新入社員のストレス，中間管理職のストレス，ホワイトカラーのストレスなど職階による分類に加えて，銀行員のストレスやメーカー社員のストレスなどの業種間での差異，また大企業や中小企業など企業規模によっても異なる状況設定することが可能で，無数の「ストレス」の形態が生み出されている。場面依存的なコミュニケーション能力の追求することは，ストレス研究と同じようなカオスを生む危険性をはらんでいる。コミュニケーション能力測定を初め，コミュニケーション学の目的がコミュニケーションという現象の背後にある本質をできるだけ単純な命題で提示することにあるとするなら，特有性を追及することは本来の目的から離れることを意味するからだ。

これまでのコミュニケーション能力の測定が普遍的な側面に偏重してきたことは明ら

かであり（高井，2011），より状況に特化したコミュニケーション能力を考えることは今後も重要な課題となるが，一方で，コミュニケーションの個別性を追求しすぎるとコミュニケーション能力を測定する本来の目的から離れてしまうといったジレンマがある。結局はコミュニケーション行動の普遍性と個別性のバランスをどう取るか，特に測定に関しては，どの程度詳細に状況を設定してコミュニケーション能力を考えるかが説明力のある診断や測定の成否を決める分水嶺になるだろう。

引用文献

- 阿久津洋巳（2008）. 項目反応理論によるストレス尺度の検討 岩手大学教育学部年報, **67**, 81-94.
- Edelen, M. O. & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, **16**, 5-18.
- 江口圭一（2011）. 労働価値観測定尺度短縮版の項目特性に関する研究：項目反応理論に基づく検討 広島大学マネジメント研究, **11**, 43-53.
- Embretson, S.E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- 藤森進（2002）. テスト得点を統計的枠組みで分析する 渡部洋（編）心理統計の技法 福村出版 pp.113-127.
- 芳賀日登美・宮原哲・田崎勝也・申知元（2015）. 日本において企業が考えるコミュニケーション能力とは一半構造化面接法による探索的研究 *Aoyama Journal of International Studies*, **2**, 81-101.
- 熊谷龍一（2009）. 初学者向けの項目反応理論分析プログラム Easy Estimation シリーズの開発 日本テスト学会誌, **5**, 107-118.
- Lewin, K. (1951). *Frontiers in group dynamics*. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers by Kurt Lewin*. New York: Harper & Row. pp. 188-237.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, **50**, 741-749.
- 宮原哲（2006）. 入門コミュニケーション論 松柏社
- 村井潤一郎（2009）. 尺度構成法 社会心理学事典 丸善出版 pp.518-519.
- 並川努・谷伊織・脇田貴文・熊谷龍一・中根愛・野口裕之（2012）. Big Five 尺度短縮版の開発と信頼性と妥当性の検討 心理学研究, **83**, 91-99.
- 末田清子・福田浩子（2003）. コミュニケーション学 松柏社
- Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, **27**, 361-370.
- 高井次郎（2011）. 対人コミュニケーション能力 日本コミュニケーション学会（編）現代日本のコミュニケーション研究—日本コミュニケーション学の足跡と軌跡 三修社 pp.47-55.
- 田崎勝也（2008）. 社会科学のための文化比較の方法 ナカニシヤ出版
- 田崎勝也・申知元（2015）. ビジネス・コミュニケーションを測定する心理特性テストの開発 *Aoyama Journal of International Studies*, **2**, 103-121.
- 豊田秀樹（2002）. 項目反応理論 [入門編] 朝倉出版
- 渡部洋（1993）. 項目反応理論を学ぶ 渡部洋（編）心理検査法入門 福村出版 pp.76-83.
- 渡辺直登（2003）. 項目反応理論による組織行動の測定とテストング 組織科学, **37**, 56-67.
- 矢富直美・渡辺直登（1995）. 項目反応理論による心理的ストレス反応尺度（PSRS）の分析 経営

行動科学, 10, 23-34.

Zirnowski, M.F., Muraki, E., Mislevy, R.J. & Bock, R.D. (1996). BILOG-MG [computer program].
Chicago, IL: Scientific Software International.