

# IRTの組織サーベイへの応用

持主 弓子\* 今城 志保\*\*

\* University of Illinois at Urbana-Champaign, School of Labor and Employment Relations

\*\* (株) リクルートマネジメントソリューションズ 組織行動研究所

## 1. 背景

サーベイ調査は組織の情報を集めるうえで最も一般的な方法の1つである (Stanton, Sinar, Balzer, & Smith, 2002)。米国における企業のサーベイ調査の歴史は古く、1930年代には従業員のモラル調査のために従業員態度サーベイが幅広く使用されており (Schneider, Ashworth, Higgs, & Carr, 1996)。1950年代までにはサーベイという手法が従業員の意見や職場の状況を知るための有効な手法として多くの企業で広く認知がなされていた (Church & Waclawski, 2001)。1990年代以降、組織サーベイはその測定対象に従業員の意見や制度の機能状況といったものだけでなく、満足感やエンゲージメントといった情緒的な側面も含む心理的構成概念も多く扱うようになり、さらにはサーベイの結果をバランススコアカードなどに盛り込み、経営指標の一つとして用いる (Becker, Huselid, & Ulrich, 2001) といった使い方も多くなされるようになってきた。そのためサーベイには、より複雑な構成概念をより正確に測定することが求められるようになってきたといえる。加えて項目数の多いサーベイは高コスト、低回答率の問題を引き起こすため (Edwards & Thomas, 1993; Paul & Bracken, 1995) むしろ一つ一つの項目の精度をあげるこそが必要といえる。これまではサーベイの精度向上にむけて、回答者に各項目をどのように理解したのかのアンケート調査を行ったり (Edwards, Thomas, Rosenfeld, 1993; Church & Waclawski, 2001)、平均や分散、項目合計相関を利用したりする (Stanton, Sinar, Balzer, & Smith, 2002) といったシンプルな方法が用いられることがほとんどであった。一方、企業で用いられる様々な心理測定ツール、例えば職務態度検査や性格検査では、項目の精査において項目反応理論 (以下 IRT) がしばしば利用され (Donovan, Drasgow, & Probst, 2000; Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001) 精緻な項目検討がなされている。組織サーベイにおいて IRT を活用した例としては、スタークら (2002) が米軍のセクシャルハラスメント調査用サーベイ (SEQ-DOD) の項目削減に使った例や、リバーズら (2009) の米国の公益企業での組織サーベイにおける question effects (項目や選択肢の文言や形式からうける影響) や context effects (項目の並び順等からうける影響) について調査を行った研究があるものの、まだまだ数は少ない。本研究は、組織サーベイに IRT の段階反応モデル (以下、GRM) を使用し、組織サーベイ項目の精査における IRT 利用の有効性を確認する。

## 2. 方法

### <データ>

使用したサーベイは、(株)リクルートマネジメントソリューションズが開発中の営業組織を対象として、組織の文化や制度の機能状況、メンバーの満足感等を確認するためのものである。項目は全部で 93 あり、本研究ではこのうちの会社、職場、仕事についての満足感を確認する 7項目を使用した (項目内容については表 1 参照)。なお、この 7項目は主因子解の推定を最尤法で行い、次元性の確認を行っている。データの収集にあたってはインターネット調査会社を利用し、従業員規模が 1000 名以上の会社で、営業職または販売職についている一般社員から課長クラスまでの 822 名のデータを用いている。

### <IRTモデルとアプリケーション>

使用するサーベイ項目の選択肢は「あてはまらない」「どちらかといえばあてはまらない」「どちらともいえない」「どちらかといえばあてはまる」「あてはまる」の 5 肢であり、項目同士は独立して回答できるようになっている。使用する IRT モデルについては、選択肢が段階づけられた複数のカテゴリである次元モデルが適切である。このタイプのモデルでは、鮫島モデル (1969)、Thissen と Steinberg の部分採点

モデル(1986)、村木の一般化部分採点モデル(1996)があるが、スタークラ(2002)やリバーズら(2009)の先行研究にならない、鮫島モデル(以下 GRM)を採用した。GRM での各種特性値の算出については MultiLQ 7.03 を使用している。

### <IRTの利点の確認>

IRT (GRM) の各特性値を利用する利点を議論するために、GRM の特性値の結果と各項目の平均、SD 修正済み項目合計相関係数の結果の比較を行う。さらに古典テスト理論で項目の精査のためにしばしば用いられる点双列相関係数の結果とも比較を行う。ただし、点双列相関係数については項目の回答が正誤の2値を前提としているため、サーベイの回答の「あてはまらない」「どちらかといえばあてはまらない」「どちらともいえない」までを誤答、「どちらかといえばあてはまる」「あてはまる」を正答として計算を行った。

## 3. 結果

表1に各項目の識別力と困難度を、図1に各項目の項目反応カテゴリ特性曲線(以下、IRCCC)をまとめた。表1の結果から、項目3、6、7の識別力が5.00、4.63、4.71と高く、精度の良い項目であることが分かる。反対に項目1、2については、識別力が2.56、1.77と精度が比較的低い項目であることが分かる。困難度は図1のIRCCCが全体的に左よりであることから、全体的にやさしめ、つまり質問内容に対して、YESと回答しやすい項目が多いことが分かる。ただし項目2については、IRCCCが右よりであり、回答者は質問に対して比較的YESと答えにくい項目であるといえる。

表2は各項目の平均、SD 修正済み項目合計相関係数、点双列相関係数である。修正済み項目合計相関係数、点双列相関係数は各項目の精度を確認することのできる指標である。これらの値とIRTの識別力を比較すると、修正済み項目合計相関係数、点双列相関係数の高い項目ではIRTの識別力が高く、低めの項目では同様にIRTの識別力が低いことが分かり、修正済み項目合計相関係数、点双列相関係数とIRTの識別力での結果に大きな相違はないことが分かる。項目平均は、各項目の難しさ、つまり質問内容に対してYESと答えやすいかNOと答えやすいのかを確認することができる。項目平均では項目2が2.61と低く、項目1や項目4の平均は3.33、3.22と比較的高めである。これらの項目のIRCCCをみると項目2は比較的右より、項目1、4は比較的左よりと項目平均と同じ傾向を示している。

表1>

NO	項目	a	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>
1	この会社に勤めてよかった	2.56 (0.15)	-1.63 (0.09)	-0.98 (0.07)	0.05 (0.05)	1.39 (0.09)
2	会社の決定や方針は従業員に信頼されている	1.77 (0.13)	-1.26 (0.11)	-0.30 (0.07)	1.24 (0.09)	2.69 (0.21)
3	今の仕事にやりがいを感じている	5.00 (0.28)	-1.10 (0.06)	-0.55 (0.04)	0.19 (0.03)	1.38 (0.06)
4	今の仕事を通じて成長できている	3.54 (0.19)	-1.33 (0.08)	-0.68 (0.05)	0.08 (0.04)	1.40 (0.07)
5	今の仕事の中で自分の力を十分発揮できている	3.42 (0.19)	-1.28 (0.07)	-0.60 (0.05)	0.37 (0.04)	1.74 (0.10)
6	今の職場で仕事をすることが楽しい	4.63 (0.24)	-1.09 (0.06)	-0.51 (0.04)	0.38 (0.04)	1.53 (0.07)
7	今の職場で仕事をすることができてよかった	4.71 (0.25)	-1.20 (0.06)	-0.64 (0.04)	0.26 (0.04)	1.43 (0.07)

括弧内は標準誤差

表2>

No	Mean	SD	修正済み項目合計相関	点双列相関
1	3.33	1.10	0.75	0.65
2	2.61	1.03	0.64	0.48
3	3.10	1.18	0.87	0.77
4	3.22	1.14	0.81	0.70
5	3.02	1.09	0.81	0.70
6	2.99	1.14	0.85	0.75
7	3.11	1.12	0.86	0.78

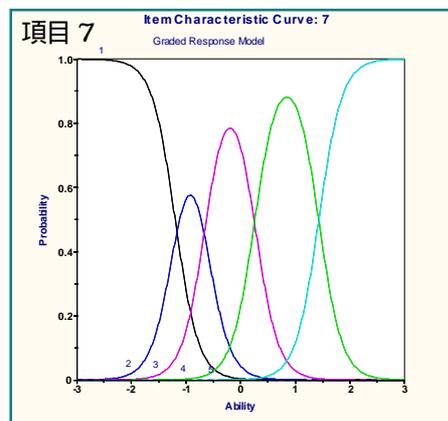
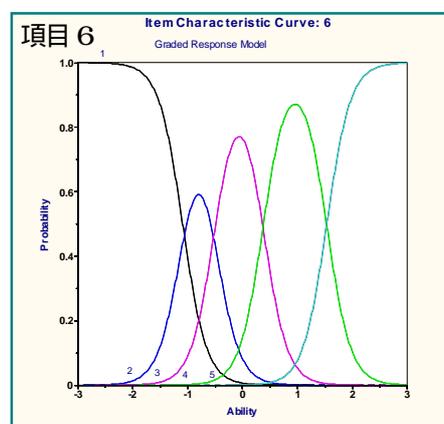
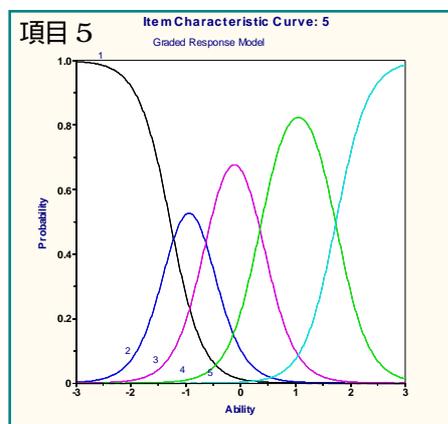
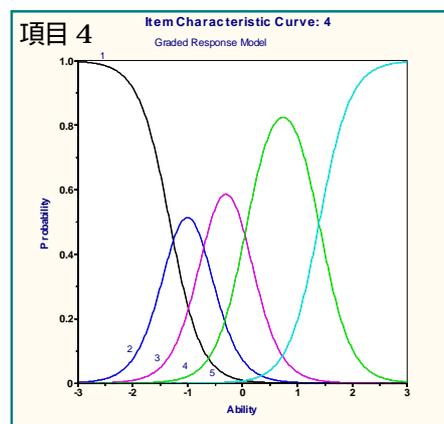
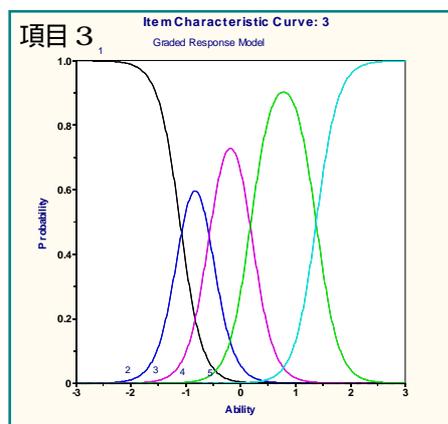
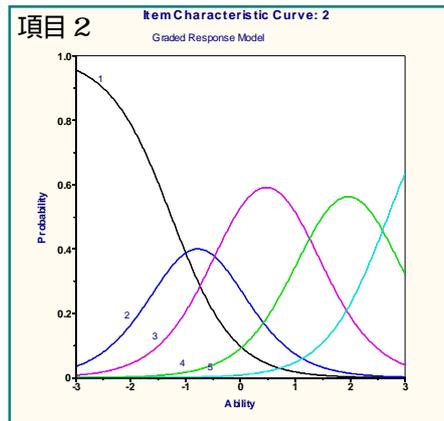
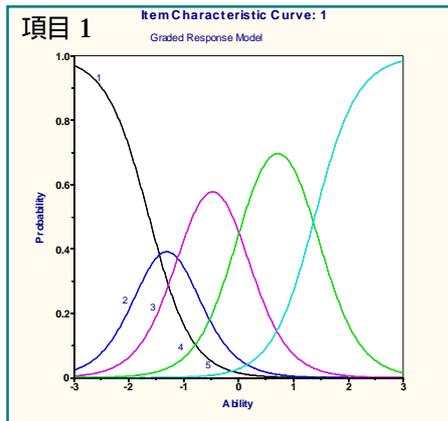
## 4. 考察

どの項目の精度が高いか(または低いか)や、どの項目がYES(またNO)と回答しやすいかについては、項目平均や修正済み項目合計相関、点双列相関でも特定することができる。しかしIRTは従来の手法と異なり、項目や各選択肢の持つ情報量や潜在特性との関係性をより精緻に確認することができる。この点がサーベイ項目の精査にどのように有効であるかを考察する。

図1をみると、今回使用した7項目は全体的に「どちらかといえばあてはまらない」のIRCCCの山が小さい。さらに「どちらかといえば

あてはまらない」と「どちらともいえない」の困難度の差が他の選択肢間の差と比較して小さい。また、項目2以外は選択肢「どちらともいえない」のIRCCCの最大値が潜在特性のネガティブ側にあることが

図 1 >



ら、質問内容に対してネガティブな反応を持つ回答者は、はっきりと「あてはまらない」と回答する、または「どちらかといえばあてはまらない」を選択せずに「どちらともいえない」を選択する傾向にあることが推測される。つまり、選択肢「どちらかといえばあてはまらない」は選択肢としてあまり機能しておらず、回答者は「どちらともいえない」を選択することでネガティブな思いを表現していると考えられる。例えば、項目 1「この会社に勤めてよかった」の選択肢「どちらともいえない」の IRCCC のトップは  $= (-0.98 \div 0.5) / 2 = -0.47$  と比較的大きくネガティブ側にある。つまり、在籍する会社に勤めてよかったかどうかを問われて「どちらともいえない」と回答するのは、ネガティブな気持ちの表現であると考えられる。

項目 4「今の仕事を通じて成長できている」の選択肢「どちらともいえない」の IRCCC の山は他の選択肢の曲線と比較して小さく、この項目で「どちらともいえない」が選ばれる確率は低い。一方、項目 7「今の職場で仕事をすることができてよかった」では「どちらともいえない」の IRCCC の山は高く、「どちらともいえない」が選ばれる確率が高いことが分かる。両項目の項目平均値は 3.22、3.11 と項目の難易度は同程度であり、通常であれば「どちらともいえない」の選択率は同程度となることが予想される。しかし項目 7 と異なり、項目 4 の「どちらともいえない」が選択されにくい結果となったのは、項目の難易度が原因ではなく、項目の文言自体に「どちらともいえない」を選択したいとは思わ

せない何かがある可能性がある。項目の文言をみってみると、項目 4「今の仕事を通じて成長できている」という項目は、現在の職務が自分の能力向上に役に立っているかどうかという限定された点について質問しており、回答者は比較的是っきりと YES か NO かを判断できるのに対し、項目 7「現在の仕事をするのができてよかった」は、問われているものが限定的ではなく、良かったと判断するための要素やエビデンスが様々に想起されることにより、回答者にとって明確に YES NO で判断することが難しい項目である可能性が推測される。したがって、項目 4「今の仕事を通じて成長できている」のような回答者が明確に判断できる項目はたとえ YES NO の 2択項目で設定されていたとしても、回答者が自分の状況を示す適切な選択肢がないとして回答負荷を感じる可能性が低いと考えられる。これに対し、項目 7「現在の仕事をするのができてよかった」のように明確に YES NO を判断しにくい項目は選択肢に「どちらともいえない」がないと回答者は自分の状況を適切に示す選択肢がないとして回答負荷を高く感じるとともに、無理に YES NO の判断を行うため測定の精度が阻害される可能性が考えられる。

IRT は少ない項目で測定の精度をあげるようにサーベイを改善することに役立つだけでなく、項目にとっての適切な選択肢数や回答負荷を考えるうえでの情報としても有効であるといえる。また「どちらともいえない」という一見ニュートラルな回答が、実はネガティブ（またはポジティブ）な反応なのかを示し、結果の得点解釈をより正確に行うために有効な情報を提供することができると考えられる。

## 5. 今後の課題

今後の課題としては以下の 2 点が挙げられる。まず、本研究ではインターネット調査で収集したデータを利用しているが、実際に企業内で実施されたサーベイデータでも同様に IRT 利用の有効性がみられるかを確認する必要がある。企業内で実施されたサーベイはインターネット調査とは異なり、その結果が社内で組織改善等に使用されることを回答者が認識しているため、回答に際して様々なバイアスがかかっている可能性があり、こうしたものが IRT を利用することにどのような影響を与えるのかを確認する必要がある。さらに、本研究では会社、職場、仕事についての満足感という心理的な構成概念を対象としたが、職場の状況や制度の機能状況といった状況認識を問う項目も組織サーベイには多く含まれる。こうした項目に関して IRT が同様に有効であるかを確認する必要がある。

### <References>

- Becker, B. E., Huselid, M. A., Ulrich, D., (2001). *The HR Scorecard: Linking People, Strategy, and Performance*. Boston, MA: Harvard Business School Press.
- Chernyshenko, O. S., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562.
- Church, H. A., Wacławski, J., (2001). *Designing and using organizational surveys: a seven-step process*. San Francisco: Jossey-Bass
- Donovan, M. A., Drasgow, F., & Probst, T. M. (2000). Does computerizing paper and pencil job attitude scales make difference? new IRT analyses offer insight. *Journal of Applied Psychology*, 85(2), 305-313.
- Edwards, J. E., Thomas, M. D., Rosenfeld, P., (eds.) (1993). *Improving organizational surveys: new directions, methods, and applications*. Newbury Park: Sage Publications
- Paul, K. B., & Bracken, D. W. (1995). Everything you always wanted to know about employee surveys. *Training & Development*, 49(1), 45.
- Rivers, D. C., Meade, A. W., & Fuller, W. L. (2009). Examining question and context effects in organization survey data using item response theory. *Organizational Research Methods*, 12(3), 529-553.
- SCHNEIDER, B., ASHWORTH, S. D., HIGGS, A. C., & CARR, L. (1996). Design, validity, and use of strategically focused employee attitude surveys. *Personnel Psychology*, 49(3), 695-705.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55(1), 167-194.
- Stark, S., Chernyshenko, O. S., Lancaster, A. R., Drasgow, F., & Fitzgerald, L. F. (2002). Toward standardized measurement of sexual harassment: Shortening the SEQDoD using item response theory. *Military Psychology*, 14(1), 49-72.