

**外科領域におけるノンテクニカルスキルの教育訓練プログラム開発と
その評価システムの構築に関する研究
ノンテクニカルスキル評価を標準化するため評価点数の乖離を可視化する試み**

研究代表者

相馬 孝博

榊原記念病院

副院長

研究要旨

スコットランド外科学会のノンテクニカルスキル（NOTSS）マスタークラスでは、評価点数の突き合わせをディスカッションと共に行って、評価基準の標準化を進めていた。ノンテクニカルスキル評価に新規に参加する場合、自身の評価が評価者全体のどこにあるかが明確になれば、その後に行う評価は標準化に向かうと思われる。まずノンテクニカルスキルが一定水準以上にある評価者群がお手本となる評価点数を蓄積する。その後の新規の評価者が、同じサンプルビデオを評価し、自らの評価点数と前者との乖離をレーダーチャートにて確認することにより、評価すべき振る舞いのポイントやその善し悪しが標準化されることを目指した。

A．研究目的

ノンテクニカルスキルの評価者は、必ずしも指導的地位にある人間である必要はないが、業務内容（テクニカルスキル）を熟知した上での客観的評価が求められる。スコットランド外科学会のNOTSS（Yule, et al.2008）マスタークラス（講習会）では、サンプルビデオを視聴して、評価点数の突き合わせをディスカッションと共に行い、参加者全体の中での各人の評価の位置づけを確認することにより、評価基準の標準化を進めていた。

本研究では、NOTSS システムをカスタマイズし、日本語翻訳したjNOTSS のパイロットシステム構築を行うため、こうしたマスタークラスに参加しなくても、一定水準の評価ができるように、自身の評価を振り返るツールを開発する。すなわちお手本となる評価を、ノンテクニカルスキルが

一定水準以上にあると思われる評価者群（指導医クラス）がサンプルビデオを視聴して評価し、評価点数を蓄積し、参照評価とする。次にノンテクニカルスキル評価に新規に参加する利用者が、自身の評価が評価者全体のどこにあるかが明確になれば（＝参照評価との乖離が可視化されれば）その後に行う評価は標準化に向かうと思われる。この乖離をレーダーチャートを用いて可視化するツールの試案を提示する。

B．研究方法

Yule らは、多くの外科専門医とともにいくつかの分析手法を用いて、NOTSS の分類および行動評価システムを開発した（Yule, et al.2008）。当初このスキルは5つのカテゴリー（状況認識、意思決定、タスクマネジメント、コミュニケーション

とチームワーク、リーダーシップ)で構成され、さらにそれらは14の要素に分けられていたが、その後のNOTSSシステムは、タスクマネジメントが除かれて4つのカテゴリー(12要素)に整理された。カテゴリーおよび要素はともに「poor」から「good」の4段階による評価がなされ、本NOTSSシステムの内容の妥当性は、感度(sensitivity)、評価者間信頼性(inter-rater reliability)および内部構造(internal structure)による検証により確認されている。簡単にまとめると、感度(sensitivity)とは「(専門家による)参照評価(reference rating)と(調査の)回答者の正確さのレベルを測る指標」、評価者間信頼性(IRR; internal-rater)とは「複数の回答者がどれだけ正確に同じ結果を出せるかを表す指標」、内部構造(internal structure)とは「カテゴリーとそれを構成するサブカテゴリー(要素)の評価の関係」である。これらをチェックすることで、NOTSSシステムの有効性を測っている。本研究では、Yuleたちの開発したNOTSSのシステムをカスタマイズし、日本語翻訳したjNOTSSの開発を行っており、その最初のステップとして、専門家による参照評価の作成等を試みている。なお、一般的なNOTSS、jNOTSSの手順(サンプルビデオのシナリオの内容などについては相馬(2013)等を参照されたい。以下、簡単に本研究における調査(評価)結果の分析手順等を説明する。

サンプルビデオは、マスタークラスで使用された英語オリジナル版を用い、日本語字幕をつけることにより言語的な障壁を除いた。複数のシナリオのビデオを、指導医クラスの外科専門医(16名)に視聴させ、jNOTSS評価票(図1)を渡し、その内容を4つのカテゴリー(状況認識、意思決定、コミュニケーションとチームワーク、リーダーシ

ップ)から評価してもらった。それぞれのカテゴリーは3つのサブカテゴリー(要素)から構成されており、Yuleたちと完全に同じものである。評価スケールも同様に、「poor」(患者の安全を脅かす、あるいは潜在的に危険な要素あり、非常な改善を要する)から「Good」(手術の遂行は、一貫して高い標準を維持患者の安心も促進、良い見本足りうる)の4段階とN/A(このケースではスキルは不要)の選択肢を用意し、カテゴリー評価および要素評価を行ってもらった。さらにこの評価表を提出後に、参加者全員でディスカッションを行い、その後改めて同じサンプルビデオを視聴して再評価を行った。この手続きによりマスタークラスと同様に、外科専門医の評価者が、評価すべき振る舞いのポイントやその善し悪しを共有化することを目指した。ただし個々の評価者の評価が変わる可能性もあれば、変わらない可能性もある。それゆえに、各シナリオに関して、評価者ごとにディスカッションの前後の評価が存在しており、その対応がわかるようになっている(事前/事後)。

これらによって収集されたデータをもとに、統計分析を試みている。まず、図1の各カテゴリーに含まれる個々の要素が内的整合性を持つかどうか(目的とする特性を測定する質問項目群であるか)を判定するために、クロンバックの α 信頼性係数を調べる。次に、カテゴリー毎にPRIDIT分析を行い、得点化を行った。PRIDIT分析とはノンパラメトリックな手法で累積相対度数(リジット)を計算して、それにて重み付けを行った後、主成分分析を行うものである(Bross,1958; Lieberthal, 2007)。主成分分析を行う前にリジットによる重み付けを行う点が特徴である。これにより、一般的な質問紙調査結果に対して主成分分析を行うことの統計的問題が少しは解消される。

PRIDIT 分析において計算された固有ベクトル (重み付け) を用いて、回答者の得点を計算し、それをレーダーチャートにて図示することができ

る。この固有ベクトルを用いて、専門家の平均得点との乖離度合いについての情報を得ることもできる。

図 1: NOTSS 評価表 (例)

病院: KCC 指導者: AT 日時: 20-- / -- / -- AM, PM
 NOTSS 評価表(例) 研修医: YT 手術: Total +PS

カテゴリー	カテゴリー評価	要素	要素評価	振り返り還元事項
状況認識		<ul style="list-style-type: none"> □ 情報を集める □ 情報を理解する □ 先を見通し行動する 	<p>3</p> <p>3</p> <p>4</p>	<p>麻酔プランの相談</p> <p>臓器浸潤の可能性・術式の相談</p>
意思決定		<ul style="list-style-type: none"> □ 選択肢を検討する □ 選択を行いチームに伝える □ 選択を実行し経過を確認する 	<p>3</p> <p>3</p> <p>4</p>	<p>PSについての認識不足</p> <p>他メンバーに周知不足</p>
コミュニケーションとチームワーク		<ul style="list-style-type: none"> □ メンバー間で情報を交換する □ 相互的な理解をつくりあげる □ チームの活動を調整する 	<p>3</p> <p>3</p> <p>4</p>	<p>振り返り(What's good/not?)</p>
リーダーシップ		<ul style="list-style-type: none"> □ パフォーマンスの水準を設定しそれを維持する □ メンバーをサポートする □ チームプレッシャーに対処する 	<p>3</p> <p>4</p> <p>3</p>	<p>適宜委任した</p>

評価スケール

1 Poor
 2 Marginal
 3 Acceptable
 4 Good
 N/A

患者の安全を脅かす、あるいは潜在的に危険な要素あり、非常な改善を要する
 心配な要素あり、かなり向上する必要あり
 標準的で満足行くレベルだが、向上の余地あり
 手術の遂行は、一環して高い標準を維持患者の安全も促進、良い見本足りうる
 このケースでは“Skill”は不要

自由記載: T4P1 でR1切除可能, アプローチ選択困難, 術中の意思決定スムーズ

さらに、シナリオ毎に事前と事後で各評価者の評価が変わるか、変わらないかについても Wann-Whitney 検定によって検証を試みる。

C. 研究結果

本研究においては、シナリオ 4 に対して 16 人の外科専門医に NOTSS 評価を行ってもらった。しかしながら、カテゴリーの要素によっては「N/A」(このケースではスキルは不要)とした評価者も存在したため、カテゴリーによってサンプル数が異なることを最初に断っておく。PRIDIT の事前

および事後の結果は以下の通りである。

クロンバックの α の値は、事前の状況認識が 0.5117(事後は 0.6885) 事前の意思決定は 0.4829(事後は 0.7699) 事前のコミュニケーションとチームワークは 0.7491(事後は 0.5311) 事前のリーダーシップは 0.7438(事後は 0.6035)となり、いずれも必ずしも高い値とはなっていない。また、特徴として、前者 2 つは事後の方が事前よりもクロンバックの α の値が高いのに対して、後者 2 つは事前の方が事後よりも値が高くなっていることがわかる。表 1 で求められた固有ベクトルを用いて、例えば、カテゴリー毎に得点を計算できる。

例えば、事前の「状況認識」であれば、 $-0.4158 \times$ 「情報を集める」の評価点 $+0.6013 \times$ 「情報を理解する」の評価点 $+0.6823 \times$ 「先を見通し行動する」の評価点によって計算される。なお、「N/A」と評価した場合、その評価者の該当カテゴリーの得点化は行われない。表 2 には、各カテゴリーの基本統計量を示している。なお、リジットによる重み付けを行っているために、平均値はゼロに近く、分散もそれほど大きくなっていない。

例えば、3人がシナリオ 4 を見て、(事前)表 3 のような評価を行ったとすると、表 2 の結果を用いて、各カテゴリーの得点を計算してやると、図 2 のようになる。

「16名の外科専門医の平均値」による参照評価は、いずれのカテゴリーもゼロにかなり近いものとなり、これから乖離するほど、彼らと異なることがわかる。例えば、ID1 と ID2 は状況認識において、2つの要素で「poor」の評価をつけているため得点は -1.91 となり、0よりも小さくなっている。逆に ID3 はいずれの要素も「acceptable」以上の評価をつけているため得点は 2.22 となっている。状況認識においては参照評価と乖離していることがわかる。意思決定においては若干の差異はあるが、ID1 から ID3 は大きく参照評価から乖離はしていないことがわかる。しかしながら、コミュニケーションとチームワーク、リーダーシップについても、参照評価と大きな乖離があることがわかる。

さらに、シナリオ 4 に関して、事前と事後で各評価者の得点に差異があるか、ないかを調べるために、(リジットによる重み付けする前の評価において) Wann-Whiteney 検定を行ったところ、カテゴリー「コミュニケーションとチームワーク」における要素「相互的な理解をつくり上げる」に

おいてのみ 10%で統計的に差異が確認され、それ以外の要素については統計的な差異は確認されなかった。

表 1 : PRIDIT 結果

シナリオ 4 (事前)

			# of Obs.	14
			Rho	0.5833
状況認識				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.75	0.841626	0.5833	0.5833
Comp2	0.90837	0.566737	0.3028	0.8861
Comp3	0.341633		0.1139	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq1_10	-0.4158	0.6974		
wsq1_20	0.6013	0.3673		
wsq1_30	0.6823	0.1853		
			# of Obs.	13
			Rho	0.497
意思決定				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.49096	0.424403	0.497	0.497
Comp2	1.06655	0.624067	0.3555	0.8525
Comp3	0.442487		0.1475	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq2_10	0.1196	0.9787		
wsq2_20	0.7193	0.2287		
wsq2_30	0.6844	0.3017		
			# of Obs.	13
			Rho	0.6578
コミュニケーションと チームワーク				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.97338	1.26282	0.6578	0.6578
Comp2	0.710561	0.3945	0.2369	0.8946
Comp3	0.316061		0.1054	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq3_10	0.5764	0.3443		
wsq3_20	0.5082	0.4903		
wsq3_30	0.6399	0.192		
			# of Obs.	12
			Rho	0.6353
リーダーシップ				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.90593	1.07645	0.6353	0.6353
Comp2	0.829478	0.564881	0.2765	0.9118
Comp3	0.264597		0.0882	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq4_10	0.398	0.6981		
wsq4_20	0.6486	0.1983		
wsq4_30	0.6488	0.1977		

シナリオ 4 (事後)

			# of Obs.	14
			Rho	0.5512
状況認識				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.65348	0.869972	0.5512	0.5512
Comp2	0.783503	0.220481	0.2612	0.8123
Comp3	0.563022		0.1877	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq1_11	0.5831	0.4377		
wsq1_21	0.6288	0.3463		
wsq1_31	0.5144	0.5625		
			# of Obs.	12
			Rho	0.7278
意思決定				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.18325	1.49516	0.7278	0.7278
Comp2	0.688091	0.559433	0.2294	0.9571
Comp3	0.128658		0.0429	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq2_11	0.6015	0.2101		
wsq2_21	0.6442	0.09409		
wsq2_31	0.4725	0.5126		
			# of Obs.	13
			Rho	0.5308
コミュニケーションと チームワーク				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.59227	0.740342	0.5308	0.5308
Comp2	0.851928	0.296125	0.284	0.8147
Comp3	0.555803		0.1853	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq3_11	0.454	0.6717		
wsq3_21	0.6166	0.3947		
wsq3_31	0.6432	0.3413		
			# of Obs.	12
			Rho	0.5487
リーダーシップ				
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.64603	0.841662	0.5487	0.5487
Comp2	0.804364	0.254754	0.2681	0.8168
Comp3	0.54961		0.1832	1
Principal components (eigenvectors)				
Variable	Comp1	Unexplained		
wsq4_11	0.4826	0.6166		
wsq4_21	0.6232	0.3607		
wsq4_31	0.6154	0.3767		

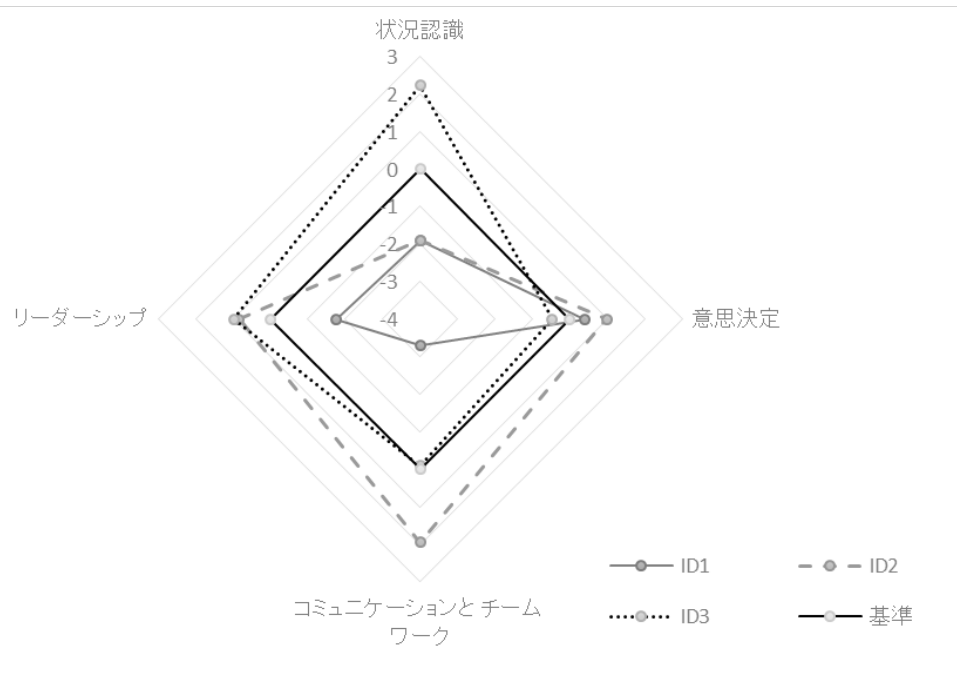
表 2：基本統計量

基本統計量	シナリオ4(事前)				シナリオ4(事後)			
	状況認識	意思決定	コミュニケーションとチームワーク	リーダーシップ	状況認識	意思決定	コミュニケーションとチームワーク	リーダーシップ
平均	-6.43E-08	2.31E-08	-8.46E-08	1.17E-07	1.43E-08	-4.17E-08	-5.38E-08	-8.33E-09
中央値	0.400	0.476	-0.103	-0.276	-0.256	-0.024	0.749	-0.182
最頻値	0.400	0.476	-0.103	-0.574	-0.256	1.395	0.749	-0.182
分散	1.750	1.491	1.973	1.906	1.653	2.183	1.592	1.646
尖度	-0.640	-0.250	1.387	-0.548	1.130	-1.958	1.460	-0.692
歪度	-0.063	0.126	-0.538	0.179	0.436	-0.070	-1.623	-0.057
範囲	4.135	4.197	5.416	4.708	4.940	3.714	3.521	3.702
最小	-2.073	-1.821	-3.196	-2.163	-2.038	-1.808	-2.633	-1.981
最大	2.062	2.376	2.220	2.545	2.902	1.905	0.888	1.721
標本数	14	13	13	12	14	12	13	12

表 3：一例

	状況認識			意思決定			コミュニケーションとチームワーク			リーダーシップ		
	sq1 1	sq1 2	sq1 3	sq2 1	sq2 2	sq2 3	sq3 1	sq3 2	sq3 3	sq4 1	sq4 2	sq4 3
ID1	3	1	1	2	2	2	1	1	1	1	2	1
ID2	3	1	1	1	2	3	2	3	3	3	3	2
ID3	4	4	3	2	2	1	2	2	2	2	3	3

図 2：カテゴリー毎の得点



D．考察

本研究では、スコットランド外科学会と同じサンプルビデオを 16 人の日本人外科専門医に視聴してもらい、同様の NOTSS 評価方法にて、参照評価を作成し、jNOTSS 評価のためのパイロットシステムの構築を検討した。評価をしてもらった 16 人の専門家の間でも、図 1 の各カテゴリーの評価にばらつきがあった。その評価表を提出後に、

参加者全員で議論して、再びシナリオに対して評価したとしてもそのばらつきは必ずしも小さくならなかった。ノンテクニカルスキルについての観察ポイントには相当の個人差が存在すると思われる。また項目の評価のばらつきをコントロールするために、リジットによる重み付けを行い、それをもとに主成分分析を行う方法 (PRIDIT) を採用した結果、jNOTSS の各カテゴリーの評価をゼロ

で標準化する 1 つの参照評価を得ることができ、誰でも図 2 のようなレーダーチャートを描くことが可能となった。このように可視化することで、新たな jNOTSS 評価者による評価が、専門家（参照評価）とどれくらい乖離しているのかを簡単に見ることができ、自身の評価のどのカテゴリーが専門家と違うかについて知ることができることになった。

E . 結論

本研究では、Yule たちが行ったように評価者間信頼性を行うことも検討したが、項目の評価のばらつきをコントロールするために、リジットによる重み付けを行い、それをもとに主成分分析を行う方法（PRIDIT）を採用した。そうすることで、jNOTSS の各カテゴリーの評価をゼロで標準化する 1 つの参照評価を得ることができた。また、これを計算するために得られた専門家の評価によるリジット、また主成分分析結果から得られた各カテゴリーの固有ベクトルなどを用いて、誰でも図 2 のようなレーダーチャートを描くことが可能となる。しかしながら、現時点では、1 つでも要素に「N/A」がある（もしくは欠損値がある）場合、そのカテゴリーの評価点が計算できない。そのため、今後は「N/A」などをどのように扱うか、また専門家の数（サンプル数）を増やし、Yule たちと同様に、評価者間信頼性等に関する統計分析を試みて、jNOTSS の参照評価をより頑健なものにしていきたいと考えている。

参考文献：

・ Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., Paterson-Brown, S. (2008)

Surgeons' Non-technical Skills in the Operating Room: Reliability Testing of the NOTSS Behavior Rating System, *World J Surg*, 32: 548-556

・ 相馬孝博 (2013) 「外科領域におけるノンテクニカルスキルの教育訓練プログラム開発とその氷塊システムの構築に関する研究」(厚生労働科学研究費補助金 地域医療基盤開発推進研究事業・平成 24 年度 総括・分担研究報告書)

・ Bross, I.D.J. (1958) How to Use Ridit Analysis, *Biometrics*, 14(1): 18-38

・ Lieberthal, R.D. (2007) Hospital Quality: A PRIDIT Approach, *Health Research and Educational Trust*, 43(3): 988-1005

F . 健康危険情報

なし

G . 研究発表

1 . 論文発表

・ 青木貴哉, 浦松雅史, 相馬孝博: The Joint Commission の警鐘事象情報に学ぶ, *病院* 72(1): 50-55, 2013

・ 相馬孝博: 医療事故を防ぐには, *心臓* 45(9)1197-1198, 2013

・ 相馬孝博: 医療安全からみたノンテクニカルスキル オーストラリア・ニュージーランドの外科医養成プログラムからみた具体的な問題行動, *臨床外科* 68(7)764-772, 2013

・ Kaneko T, Nakatsuka A, Hasegawa T, Fujita M, Souma T, Sakuma H, Tomimoto H: Postmortem Computed Tomography is an Informative Approach to Determining Inpatient Cause of Death but Two Factors Require Noting from

the Viewpoint of Patient Safety. JHTM1:1-9, 2013.

- ・竹村敏彦，浦松雅史，相馬孝博：東京医科大における医療安全意識の経年比較分析.東医大誌 71 (4)：363-375, 2013

2. 学会発表

- ・相馬孝博：呼吸器外科医のノンテクニカルスキル第 30 回日本呼吸器外科学会 安全教育セミナー，2013 年 5 月 9 日，名古屋（特別講演）
- ・相馬孝博：WHO 患者安全カリキュラムガイド多職種版について，日本薬学協議会，2013 年 6 月 28 日，東京(特別講演)
- ・相馬孝博：世界標準の患者安全教育 - WHO 患者安全カリキュラムガイド多職種版から学ぶ.第 32 回日本歯科医学教育学会，2013 年 7 月 13 日，札幌(特別講演)
- ・相馬孝博：世界標準の患者安全教育 - WHO 患者安全カリキュラムガイド多職種版から学ぶ.第 45 回日本医学教育学会，2013 年 7 月 26 日，千葉（モーニングセミナー）
- ・相馬孝博：医療安全の基礎，医療・病院管理研究協会，2013 年 8 月 23 日，(特別講演)
- ・相馬孝博：世界標準の患者安全教育 - WHO 患者安全カリキュラムガイド多職種版から学ぶ.第 36 回日本高血圧学会総会医療倫理・医療安全講習会，2013 年 10 月 24 日，大阪（特別講演）
- ・相馬孝博：WHO カリキュラムガイドに学ぶノンテクニカルスキルの重要性.第 8 回医療の質・安全学会学術集会，2013 年 11 月 23 日，東京(共催セミナー)
- ・相馬孝博：安全対策と感染対策の連携の必要性.第 8 回医療の質・安全学会学術集会，2013 年 11 月 23 日，東京（シンポジウム）

- ・相馬孝博：WHO カリキュラムガイドの医療専門職の基礎教育への活用，第 8 回医療の質・安全学会学術集会，2013 年 11 月 23 日，東京（ワークショップ）

H. 知的財産権の出願・登録状況（予定を含む）

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

