

医学教育分野別評価における 評価報告書内容にみる特徴分析

浅田義和（自治医科大学）

1. 本発表の目的と課題

近年、医学教育において、教育の質保証の観点から教学 IR に注目が集まっている。この発端は、ECFMG⁽¹⁾ (Educational Commission For Foreign Medical Graduates) が 2010 年に出した勧告で、2023 年以降、USMLE (United States Medical Licensing Examination、米国における医師国家試験) を受験するにあたっては国際的な基準で認定を受けた大学を卒業することを要件としたことによる。これは通称、2023 年問題とされている。この勧告は、字面通りに受け取るのであれば、あくまで USMLE を受験する際に初めて影響するものであり、国内の医学教育そのものには直接の影響を及ぼすものではない。しかし、一方では国際的な基準での大学の評価という観点に着目した場合、この認定を受けていないことは世界標準からみた教育の質保証という意味で疑問が残ってしまう。このような背景から、国際的な基準を用いた医学教育の評価と質保証を必要とする考えが広がってきた。

現在では、WFME⁽²⁾ (World Federation for Medical Education) の基準を元にした「医学教育分野別評価基準日本版⁽³⁾ (以下、日本版基準)」を用いた医学教育分野別評価が日本医学教育評価機構⁽⁴⁾ (Japan Accreditation Council for Medical Education: JACME) によって行われている。

分野別評価は 2019 年 9 月 1 日時点で 31 大学の評価が終了しており、各大学における自己点検評価および JACME による評価結果が Web サイト上で掲載されている。国内の医学部は 2019 年 9 月 1 日時点で 82 であるため、およそ 4 割の評価が終了したことになる。日本版基準は WFME の基準改定と合わせて適宜アップデートされており、特に 2016 年 6 月 24 日に公開された Ver2.1 とそれ以前とはベースとなっている WFME の基準の版が異なっており、内容に差異が生じている。このため、本論文では Ver2.1 以降の日本版基準を用いた例を対象として扱うこととする。なお、Ver2.1 以降を用いて評価された大学は 31 大学中の 17 である。

日本版基準は 9 つの領域とその下位領域から成り立つ。各領域の内容と下位領域の数を () 内に示す。なお、下位領域は領域 9 を 1 つとカウントして、合計 36 存在している。

表 1 日本版基準の 9 領域

領域 1	(4)	使命と学修成果
領域 2	(8)	教育プログラム
領域 3	(2)	学生の評価
領域 4	(4)	学生
領域 5	(2)	教員

領域 6	(6)	教育資源
領域 7	(4)	プログラム評価
領域 8	(5)	統轄および管理運営
領域 9	(0)	継続的改良

各下位領域には、それぞれ基本的水準（Basic Standards、以下 B 水準）と質的向上のための水準（Quality Development Standards、以下 Q 水準）が設けられており、各々に対して「適合」「部分的適合」「不適合」のいずれかの判定がなされることになる。なお、実際には下位領域の下に小項目が設けられているが、JACME のサイトで公開されている報告書では、下位領域ごとの判定結果のみ公表されている。

9つの領域のうち、特に領域 7 は教学 IR と密接に関連づけて扱われることが多く⁽⁵⁾⁽⁶⁾、国内の医学部における教学 IR の注目が高まっている理由の一つである。また、評価を受けた大学の適合・部分的適合・不適合の割合については、領域ごとに大きく異なっていることが分かる⁽⁷⁾。表 2 に B 水準および Q 水準の件数を示す。なお、領域 1 の NA は防衛医科大学校であり、その特性上評価が実施されていない。また、Q 水準の領域 9 は今後の改良計画に関するものであり、全大学にて評価を実施せずとなっている。

表 2 B 水準および Q 水準における件数

領域	B 領域				Q 領域			
	適合	部分的適合	不適合	NA	適合	部分的適合	不適合	NA
1	5	13	0	1	46	21	0	1
2	84	52	0	0	77	59	0	0
3	3	31	0	0	3	30	1	0
4	53	15	0	0	63	5	0	0
5	20	14	0	0	31	3	0	0
6	75	27	0	0	83	19	0	0
7	11	55	2	0	4	62	2	0
8	85	0	0	0	80	5	0	0
9	16	1	0	0	0	0	0	17

表 2 から分かるように、B 水準・Q 水準ともに領域 3, 7 では部分的適合の件数が多くっており、質保証の観点から改善の余地が残されていることがわかる。また、両領域では、ごくわずかではあるが不適合の結果も存在している。そこで今回、これら 2つの領域に着目し、報告書の記述内容をテキストマイニングすることで、特徴の分析を試みた。

2. 方法

JACME の Web サイトに掲載されている 31 の評価結果のうち、Ver2.1 以降を用いて評価を受けた 17 大学の結果を用いたテキストマイニングを実施した。元データとしては各

大学の評価結果に関して、領域3と領域7の2つの領域に着目し、B水準およびQ水準の評価結果本文を用いた。なお、評価結果は「特記すべき良い点」と「改善のための助言」との2つの要素からなるため、これらについてはカテゴリを分類して解析を行った。また、元データにおいて、下位領域によってはコメントの記載がない部分もあり、マイニングするにあたってデータが不十分となることも考慮し、今回の解析では領域ごとの解析に留めることとした。テキストマイニングを行うにあたってはRMeCabを利用した。

3. 結果

表3として、各領域の文章を形態素解析した結果における総要素数を示す。

表3 各領域の総要素数

領域	カテゴリ	B水準	Q水準
領域3	特記	614	191
	改善	2670	1613
領域7	特記	680	393
	改善	2408	2254

領域3、領域7ともに改善に関するコメント量が非常に多く、全体としてB水準におけるコメントが多く記載されていた。続いて、各領域における語句の頻度に関して、不要語を除いたうえで名詞・形容詞に絞って集計をした結果の上位20語について解析した。まず、表4および表5に領域3 B水準での結果を示す。

表4 領域3 B水準(特記)

語(件数) 1-5	語(件数) 6-10	語(件数) 11-15	語(件数) 16-20
評価(32)	成果(5)	卒業(3)	面(3)
学生(6)	技能(5)	総括(3)	高い(3)
学修(5)	臨床実習(5)	達成(3)	フィードバック(2)
導入(5)	試験(4)	知識(30)	リサーチ(2)
態度(5)	独自(4)	表(3)	参加(2)

表5 領域3 B水準(改善)

語(件数) 1-5	語(件数) 6-10	語(件数) 11-15	語(件数) 16-20
評価(114)	科目(18)	構築(12)	実施(11)
学生(39)	臨床実習(16)	外部(12)	教育(11)
学修(30)	形成(15)	方法(12)	卒業(10)
確実(25)	達成(14)	吟味(11)	促進(9)
成果(19)	導入(12)	学習(11)	仕組み(9)

語数は異なるものの、領域3のB水準では、上位3語はいずれも評価、学生、学修の順

となっていた。また、臨床実習や成果、卒業など、両者に出現している語が複数みられた。続いて、領域3のQ水準の結果を表6・表7で示す。

表6 領域3 Q水準(特記)

語（件数）1-5	語（件数）6-10	語（件数）11-15	語（件数）16-20
評価(13)	検討(2)	臨床実習(2)	判定(1)
導入(4)	方法(2)	次(2)	卒業(1)
試験(4)	科目(2)	法(2)	吟味(1)
新しい(3)	6年(2)	レポート(1)	実施(1)
フィードバック(2)	妥当性(2)	信頼(1)	明示(1)

表7 領域3 Q水準(改善)

語（件数）1-5	語（件数）6-10	語（件数）11-15	語（件数）16-20
評価(65)	臨床実習(15)	外部(12)	回数(9)
試験(22)	活用(14)	科目(11)	性(9)
検証(19)	促進(13)	統合(11)	学習(8)
妥当性(18)	方法(13)	結果(10)	建設(8)
学生(16)	フィードバック(12)	信頼(9)	信頼性(8)

領域3のQ水準においては、上位の語の順番は「評価」以外一致しなかった。しかし、どちらも上位に「試験」が出現していること、および上位語の中に「妥当性」や「信頼性」などの評価結果に関する質保証に起因する内容が含まれていた。

続いて、領域7のB水準の結果を表8・表9で示す。

表8 領域7 B水準(特記)

語（件数）1-5	語（件数）6-10	語（件数）11-15	語（件数）16-20
カリキュラム(15)	プログラム(9)	モニタ(5)	実績(5)
学生(12)	設置(8)	分析(5)	成果(5)
教育(11)	卒業生(8)	収集(5)	委員会(5)
評価(11)	IR(7)	学修(5)	参加(4)
評価委員会(10)	委員(6)	医学部(5)	組織(4)

表9 領域7 B水準(改善)

語（件数）1-5	語（件数）6-10	語（件数）11-15	語（件数）16-20
カリキュラム(49)	教育(35)	改善(16)	フィードバック(13)
分析(43)	学修(21)	卒業生(16)	結果(13)
評価(43)	成果(21)	実績(15)	構築(12)
学生(41)	収集(20)	教員(15)	反映(12)
プログラム(36)	データ(18)	モニタ(14)	対応(11)

領域7のB水準においては、特記カテゴリの中に「IR」の語が出現していた。また、特記・改善両カテゴリに出現した語として「カリキュラム」「学生」「卒業生」など、プログラム評価を行ううえでの調査対象になる語がみられた。

最後に、領域7のQ水準に関する結果を表10・表11に示す。

表10 領域7 Q水準(特記)

語（件数）1-5	語（件数）6-10	語（件数）11-15	語（件数）16-20
評価(19)	学生(4)	病院(3)	実習(2)
卒業生(9)	結果(4)	協議会(3)	導入(2)
フィードバック(7)	分析(3)	臨床研修(3)	担当(2)
実績(5)	カリキュラム(3)	プログラム(2)	活動(2)
教育(4)	教員(3)	収集(2)	関係(2)

表11 領域7 Q水準(改善)

語（件数）1-5	語（件数）6-10	語（件数）11-15	語（件数）16-20
評価(41)	フィードバック(34)	委員会(18)	関係者(13)
カリキュラム(41)	実績(28)	学修(15)	データ(12)
学生(41)	結果(24)	改善(15)	関連(11)
プログラム(39)	分析(23)	成果(15)	他(11)
教育(36)	卒業生(19)	収集(13)	入学(10)

領域7のQ水準では、臨床研修、病院、入学など、プログラム評価を行う時期的な要素に関する語句が出現していた。また、他の結果と同様、特記・改善それぞれに出現する語が複数みられた。

4. 考察

領域3および領域7において、特記・改善の両カテゴリにおいて共通する語が出現していた。分野別評価においては下位領域および小項目によって評価される点が定められており、この部分に関する語が数多く抽出されていることが考えられる。例えば下位領域3.1では、B水準において評価方法や試験の実施回数について、Q水準において試験の信頼性や妥当性についてなどが、それぞれ項目として挙げられている。領域7においても下位領域でプログラムのモニタや評価、学生と卒業生の実績に関する分析が項目として挙げられており、これらがキーワードとして反映されていることが考えられる。

一方、同一の語、例えば領域3の妥当性という語が両カテゴリにおいて出現していたことは、大学によって改善に関する取り組みの進捗が異なっていることを意味する。教育の質保証を目指した改善の取り組みは今が過渡期であり、本研究で行ったような解析を通じて、自大学のみならず他大学の報告書にて記載された内容を踏まえて改善の取り組みを行うための指針を作っていくことも必要であるといえよう。

領域3は学生の評価、領域7はプログラム評価と、異なる対象に関する「評価」を検討する項目となっている。一方、抽出された語をみると両者に共通して存在するものも多数存在している。プログラム評価を行うにあたり、学生の評価すなわち学修の成果を調査することは必要不可欠である。実際、領域7で抽出された「学生」という語の利用状況を確認すると、アンケート等の解答に関するプログラム評価の観点と合わせ、基礎医学や臨床実習などの学内での学習成果、さらには臨床研修病院などにおける卒後の評価といった部分で用いられていることがわかる。教学IRの範囲として、これらのデータを収集し、分析していくことへのニーズもみてとれる。

本研究の限界として、82大学のうちVer2.1以降を用いた17大学の結果解析にとどまっている点、語の頻度解析にとどまっており、共起やトピックモデル分析などを用いた意味の解析に踏み込めていない点などが挙げられる。また、実際は分野別評価の際に各大学が自己点検評価の報告書を作成しており、この内容と比較解析することも必要であろう。

今回は語数の多いものに着目して解析を行ったが、特記カテゴリでは他大学で行われていない、特別かつ意義のある事例に関するコメントが記載されることも多い。このため、共起等の解析を通じて、各大学での特色のある取組みが各下位領域のどのような観点において有意義と判断されたかを整理していくことなども可能であろう。

5. 結論

JACMEによる医学教育分野別評価の結果に関して、領域3および領域7の評価コメントのテキストマイニングを行った。両領域とも、特記・改善の両カテゴリにおいて共通する語句が出現しており、各大学での改善の取り組み進度に差がある点、分野別評価全体において特に着目すべき点などが明らかとなった。

謝辞

本研究はJSPS科研費JP18K02865の助成を受けた。

【参考文献】

- (1) ECFMG <https://www.ecfmg.org/> (アクセス日 2019. 9. 30)
- (2) WFME <https://wfme.org/> (アクセス日 2019. 9. 30)
- (3) 医学教育分野別評価基準日本版 Ver 2.31
https://www.jacme.or.jp/pdf/wfme-jp_ver2.31.pdf (アクセス日 2019. 9. 30)
- (4) JACME <https://www.jacme.or.jp/> (アクセス日 2019. 9. 30)
- (5) Asada, Y. Trends of Institutional Research in Japanese Medical Education: A case study of Jichi Medical University. 7th International Congress on Advanced Applied Informatics. 490-492. (2018)
- (6) 中村真理子, 椎橋実智男, 伊藤彰一, 浅田義和. 医学教育におけるIRの機能と使命. 週刊医学界新聞. 3288. 1-3. (2018)
- (7) Asada, Y. Investigation of the Current Situation of Medical Education in Japan Based on the Results of Accreditation Based on Global Standards. 8th International Congress on Advanced Applied Informatics. 342-346. (2018)