



東北大学歯学部FD 兼 大学間連携共同教育推進事業FD

2014.2.13

＝学習成果の基礎と最近の動向＝ 学習成果をどう評価するか？

松下 佳代

京都大学・高等教育研究開発推進センター
matsushita.kayo.7r@kyoto-u.ac.jp



大学での学習成果とは？



CONTENTS

- 大学での学習成果とは？
- 学習成果の評価のタイプ
- パフォーマンス評価をめぐる動向
- 新潟大学歯学部でのパフォーマンス評価の開発
- 学習としての評価



学習成果への注目

- Outcome-based education
 - 〈教員が何を教えたか〉よりも、〈学生が何を学んだか〉に焦点
- 日本では、学士課程答申から
 - 中央教育審議会「学士課程教育の構築に向けて(答申)」(2008)
 - 「学士力」
 - ＝学士課程修了段階での学習成果 (learning outcomes)

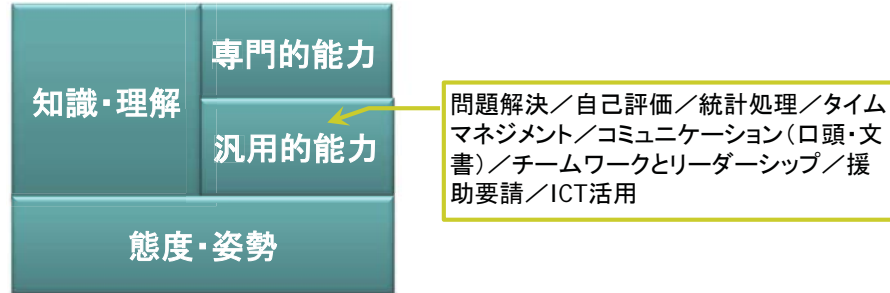
知識・理解	多文化・異文化に関する知識の理解／人類の文化、社会と自然に関する知識の理解
汎用的技能	コミュニケーション・スキル／数量的スキル／情報リテラシー／論理的思考力／問題解決力
態度・志向性	自己管理能力／チームワーク／リーダーシップ／倫理観／市民としての社会的責任／生涯学習力
統合的な学習経験と創造的思考力	

新潟大学歯学部の場合



● 卒業時に獲得が期待される学習成果

= “新潟大学歯学部版学士力” (小野・大内・前田, 2011)



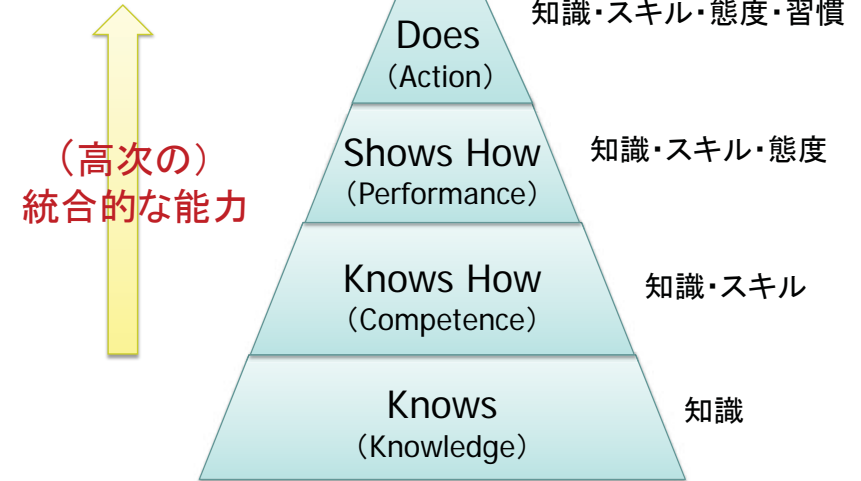
⇒ **高次の統合的な能力**にみあった評価が必要
= 能力目標と評価の連動 (alignment)

5

学習成果にみあった評価①



「臨床能力評価の枠組み」
(Miller, 1990)



6

学習成果の評価のタイプ



● <高次の統合的な能力にみあった評価>とはどんな評価なのか？

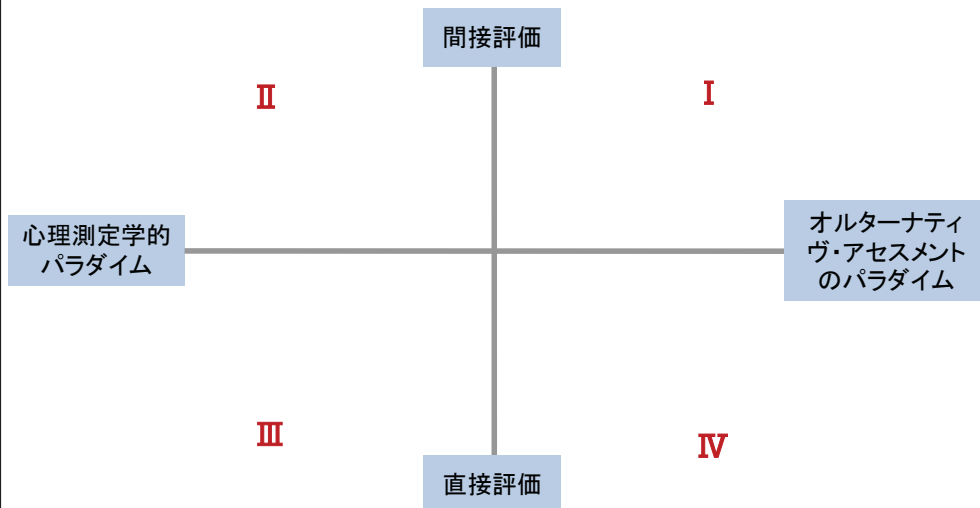
- 現在行われている学習成果の評価は、
 - 直接評価 vs. 間接評価
 - 心理測定的パラダイム vs. オルターナティブ・アセスメントのパラダイム

という2つの軸で捉えられる

7

8

学習評価の構図①



9

間接評価と直接評価

(Palomba & Banta, 1999)

● 間接評価

- 学生の学習行動や学習についての自己認識を通じて、
学習成果を間接的に評価 **何ができているか？**
- 質問紙による「学生調査」
 - 【例】NSSE (National Survey of Student Engagement)
JCIRP (日本版大学生研究調査プログラム) (山田, 2012)

● 直接評価

- 学生の知識や技能などの表出を通じて、
学習成果を直接的に評価 **何ができるか？**
- 【参考】「臨床能力評価の枠組み」

10

評価の2つのパラダイム

	心理測定的パラダイム	オルタナティブ・アセスメントのパラダイム
学問的基盤	心理測定学、知能理論	構成主義、状況論
評価目的	アカウンタビリティ、能力判別	教育改善・指導
評価項目	分割可能性	複合性
評価機能	総括的評価	形成的評価
評価場面	脱文脈性 統制された条件	文脈性 シミュレーション、真正の文脈
評価基準	客観性	間主観性
評価データ	量的データ	質的データ
評価主体	評価専門家、政策担当者	実践者自身(教員・学生)
評価法	客観テスト(標準テスト)、 学生調査 など	真正の評価、ポートフォリオ評価、 パフォーマンス評価 など

11

オルタナティブ・アセスメントの評価法

- **ポートフォリオ評価** (portfolio assessment)
 - ポートフォリオに収められた資料にもとづいて、学習者の成長のプロセスを評価する方法
- **真正の評価** (authentic assessment)
 - 学習者に、仕事場や市民生活など現実世界の課題と類似した、本物らしさ(真正性)をもった課題に取り組みさせる評価の方法
- **パフォーマンス評価** (performance assessment)
 - ある特定の文脈のもとで、さまざまな知識や技能などを用いながら行われる、学習者自身の作品や実演(パフォーマンス)を直接に評価する方法

12

● 3つの評価法の関係

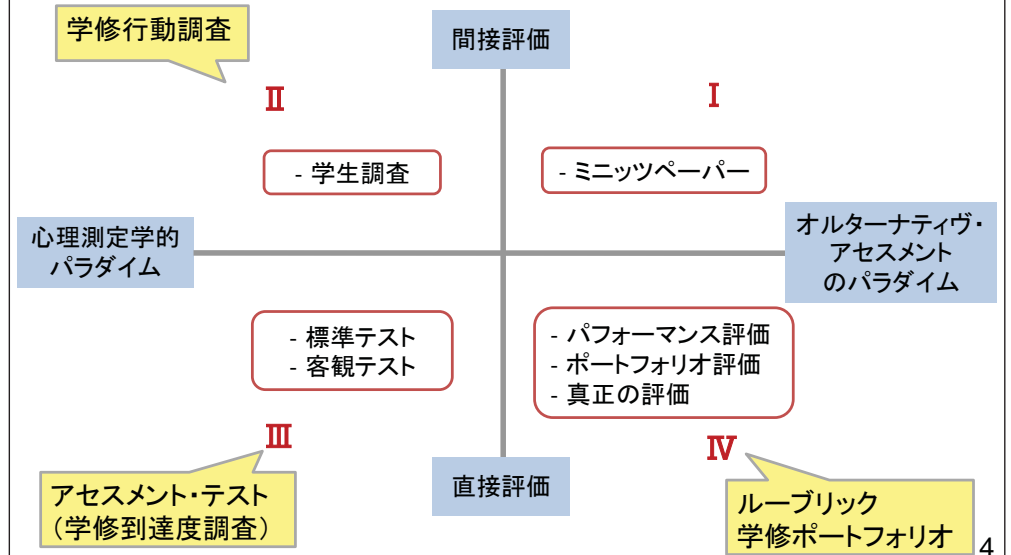
- 概念的区別…どこに焦点があるかの違い
 - ポートフォリオ評価: ポートフォリオに収められたエビデンス (一定期間の成長プロセス)
 - 真正の評価: 文脈や課題の「真正性」
 - パフォーマンス評価: ある文脈での「遂行」やその結果

【参考】質的転換答申

- 中教審「新たな未来を築くための大学教育の質的転換に向けて(答申)」(2012)

「特に、成果の評価に当たっては、学修時間の把握といった学修行動調査やアセスメント・テスト(学修到達度調査)、ルーブリック、学修ポートフォリオ等、どのような具体的な測定手法を用いたかを併せて明確にする。」

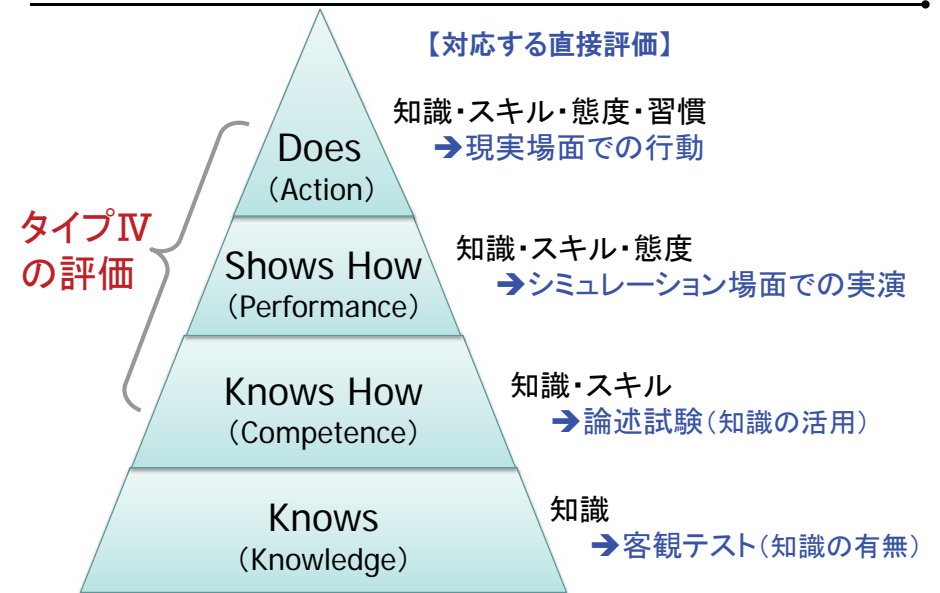
学習評価の構図②



それぞれのメリット・デメリット

- **タイプ II** 【例】新潟大学歯学部卒業生調査
 - メリット: 集団のデータが容易に収集でき量的に分析可能、アカウントビリティに使いやすい
 - デメリット: 構成概念以外の要因が入り込むおそれ(いい加減な回答態度、社会的望ましさや要求水準の反映など)
- **タイプ III** 【例】各授業科目での客観テスト
 - メリット: 客観性が高い、評価負担が小さい
 - デメリット: 高次の統合的な能力が評価できない
- **タイプ IV** 【例】改良版トリプルジャンプ(小野報告)
 - メリット: 高次の統合的な能力が評価できる、教育改善や個々の学生の学習改善に使いやすい
 - デメリット: 信頼性を担保しにくい、評価負担が大きい

学習成果にみあった評価②





パフォーマンス評価をめぐる 動向

17



パフォーマンス評価の現状

- 大学教育における現状
 - 学生の作品や実演の直接評価は多い
 - 実技(医療・教員養成など)、演奏や美術作品(芸術分野)、製作物(PBLなど)、レポート・卒業論文、プレゼン・口頭試問など
 - しかし、教員個々人の主観にゆだねられていて、妥当で信頼性のある評価方法になっていない
- ↓
- パフォーマンス評価に必要な道具立て
 - 評価課題： パフォーマンス課題
 - 評価基準： ルーブリック など

18

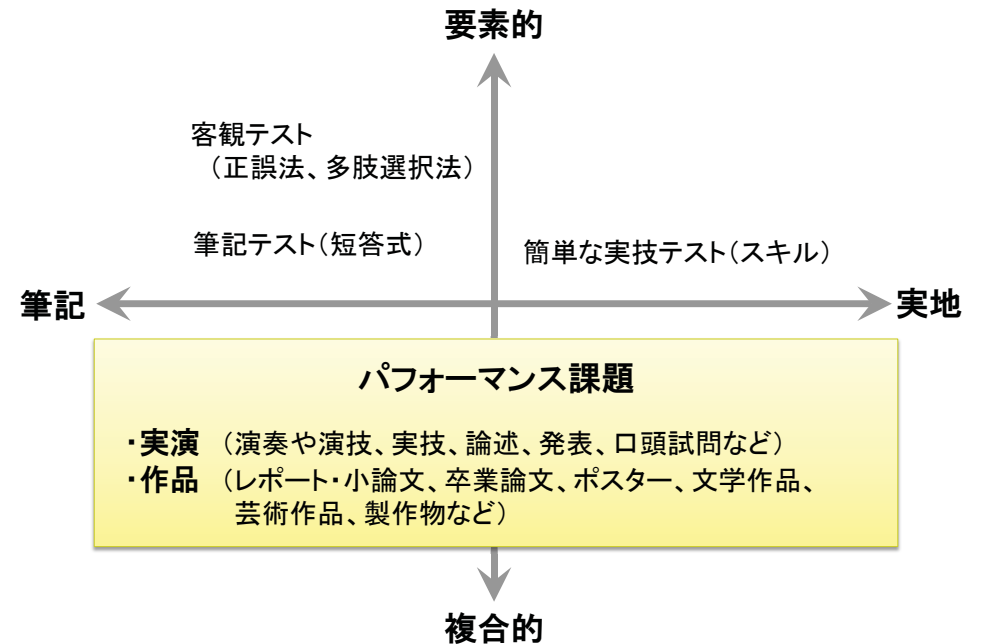


パフォーマンス課題

- パフォーマンス課題 (performance task)
 - 学習者のパフォーマンスを評価するためにデザインされた課題
 - リアルな状況で、様々な知識や技能を総合して使いこなすことを求めるような課題
- パフォーマンス課題の要素 (Wiggins & McTighe, 2005)
 - 1. ゴール／2. 役割／3. 相手／4. 状況／5. 作品・実演／6. 満たすべき規準

19

【評価課題の分類】



ルーブリック



● ルーブリックとは

=パフォーマンスの質を段階的・多面的に評価するための評価基準表

- パフォーマンスの質 →ルーブリック→ 量的表現

● ルーブリックへの注目

- 中教審「質的転換答申」で言及
 - ← 厳格な成績評価、アカウントビリティの要請

	レベル4	レベル3	レベル2	レベル1
観点1
観点2
観点3
観点4

観点(規準) → (左側) レベル (上側) 記述語 (右下)

● ルーブリックのタイプ

- (a) 構造
 - 基準を複数設定して分析的に評価する「分析的ルーブリック」
 - 基準を複数に分けずに全体的に評価する「全体的ルーブリック」
- (b) スコープ
 - ある領域で一般的に適用できる「一般的ルーブリック」
 - 当該課題だけに適用される「課題特殊のルーブリック」
- (c) スパン
 - 複数年にまたがって使われる「長期的ルーブリック」
 - 短期間、あるいはスナップショット的に使われる採点用ルーブリック

パフォーマンス評価の2つの方向性



● 標準化志向 (心理測定的パラダイム)

- 大規模な標準テスト(同一の評価課題)
- 同一の評価基準

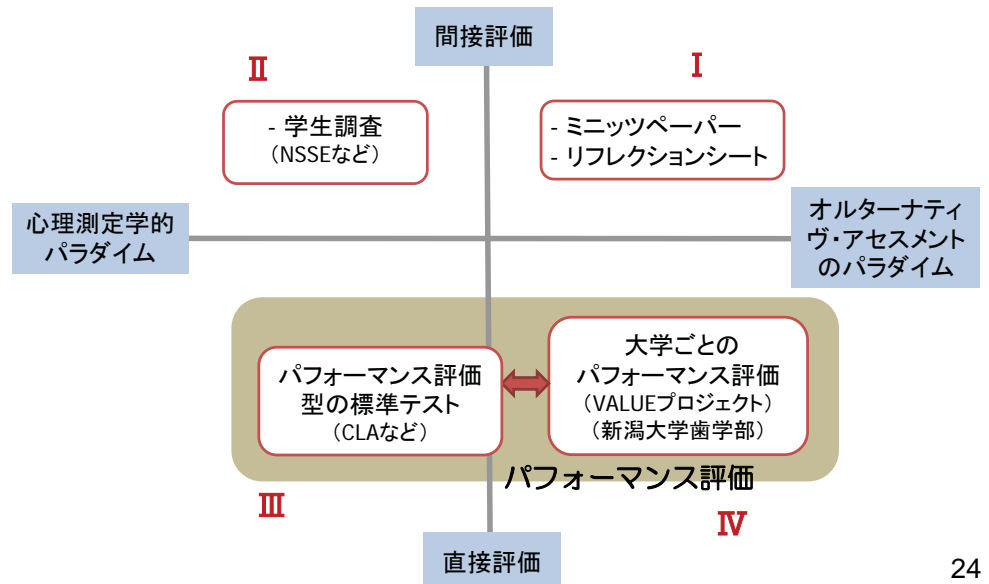
【例】OECD-AHELOにおけるジェネリックスキルの評価(CLA)

● 非・標準化志向 (オルターナティブ・アセスメントのパラダイム)

- 大学や学科独自の評価課題
- 評価基準の共通性と多様性

【例】AAC&U(アメリカ大学・カレッジ協会)のVALUEプロジェクト、新潟大学歯学部

学習評価の構図③





パフォーマンス評価型の標準テスト

- OECD-AHELO (Assessment of Higher Education Learning Outcomes)
 - 高等教育修了段階で実施 = “大学版PISA”
 - 大学ごとに評価
 - 4つの柱
 - 工学、経済学
 - ジェネリックスキル (批判的思考、分析的推論、問題解決、コミュニケーションなどを用いながら、様々なソースからの証拠資料を集め、現実的な状況の中のオープン・エンドな問いに答えられるか)
 - 付加価値 (入学時と卒業時でどれだけ伸びたか) ×
 - 2010~12年の間に試行調査(フィージビリティ・スタディ)
 - 17カ国参加 (日本は工学分野で参加)

25



CLA (Collegiate Learning Assessment) の評価法

- OECD-AHELOの汎用的技能の評価で採用 ← 北米の大学で普及
- パフォーマンス課題とルーブリック型の評価基準を使用
 - 現実場面を模したシナリオ (例: 「犯罪削減」)

《あなたはジェファーソン市の市長選挙における現職市長のコンサルタントです。市の犯罪削減についての対立候補の主張の長所と問題点について、メモを作成するように依頼されました。…》
 - 文書ライブラリを使って問いに答える(90分)
 - 評価基準にそって評価
- 大学教育によって得た「付加価値」を分析
 - 入学時—(2年生修了時)—卒業時 の変化
 - 学生の学習成果から、大学の教育力を測定

27

AHELO: 4 strands of work

Discipline strand in Economics

Initial work on defining expected learning outcomes through 'Tuning' approach.

+ contextual data

Discipline strand in Engineering

Initial work on defining expected learning outcomes through 'Tuning' approach.

+ contextual data

Generic skills strand

International pilot test of the US Collegiate Learning Assessment (CLA), to assess the extent to which problem-solving or critical thinking can be validly measured across different cultural, linguistic and institutional contexts.

+ contextual data

Research-based “Value-added” or “Learning gain” measurement strand

Several perspectives to explore the issue of value-added (conceptually, psychometrics), building on recent OECD work at school level.



26



Collegiate Learning Assessment

「犯罪削減」課題 問2

[cla] [cwra]

2 of 3 1 hr 27 min 32 sec

文書ライブラリ

Question 2

Do you agree or disagree with Dr. Eager's statement that "we should take the money that would have gone to hiring more police officers and spend it on the XYZ drug treatment program?" Use the box below to explain why you reached this conclusion. In other words, why do you believe Dr. Eager's position regarding this matter does or does not make sense? Is there a better solution? Be specific as to the strengths and limitations of these options. Be sure to cite the information in the documents and any other factors you considered (such as the quality of the research conducted on various treatment programs) that led you to this conclusion.

Copy Paste

Select document: Document Library Contents

© 2007 CAE. All rights reserved.

Document Library Contents

- Document 1 Instructions
- Document 2 Investigator's memo
- Document 3 Newspaper Story
- Document 4 Police Tables
- Document 5 Report on XYZ
- Document 6 Crime Statistics
- Document 7 Dr. Eager's Chart
- Document 8 Research Abstracts

28



- AHELOの試行調査の結果 (OECD, 2013)
 - 汎用的技能のテスト (CLA) の問題点を指摘
 - 文脈があまりにアメリカ的
- ↓
- 汎用的技能は専門分野の文脈の中で測定するのが適切

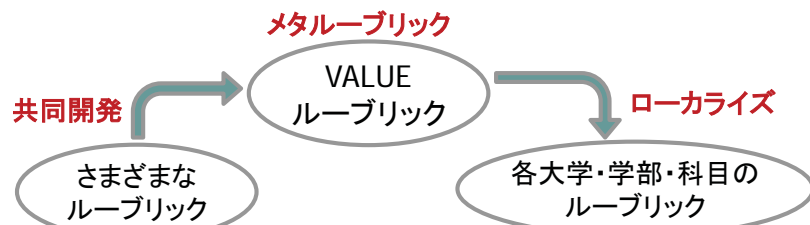


標準化によらないパフォーマンス評価

- AAC&UのVALUEプロジェクト
 - *Valid Assessment of Learning in Undergraduate Education
- eポートフォリオとルーブリックの組み合わせによる評価
 - ← 標準テストへの批判
 - (サンプルの学生のみ / 一回的 / 教育・学習の改善に役立たない)
- 教養教育における汎用的能力の評価基準として、ルーブリックを共同開発
 - 探究と分析 / 批判的思考 / 創造的思考 / 文章コミュニケーション / 口頭コミュニケーション / 読解 / 量的リテラシー / 情報リテラシー / チームワーク / 問題解決 / 市民参加 / 異文化知識・能力 / 倫理的推論 / 生涯学習の基礎とスキル / 統合的学習 / グローバルな学習



- VALUEルーブリック
 - 観点別で、領域一般的な、長期的ルーブリック
 - 領域分野横断的な16領域
 - 学士課程4年間の長期的ルーブリック
 - 「メタルーブリック」としての性格
 - … 共通性と多様性の両立を図る



【例】文章コミュニケーション VALUEルーブリック

	キャップストーン		マイルストーン		ベンチマーク	
	4	3	2	1	1	1
文章作成の文脈と目的	文脈・読者・目的について完璧な理解を示し、それによって、与えられた課題に対応し、作品のあらゆる要素に焦点をあてている。	……	……	……	……	……
内容の展開	適切に関連性があり説得力に富む内容を用いることによって、科目の習得ぶりを示すとともに、書き手の理解したことを伝え、作品全体を形づけている。	……	……	……	……	……
ジャンルと学問分野の約束事	特定の学問分野や文章作成課題に関連する広範な約束事(構成、内容、提示、書式、文体選択を含む)に対し、細かい注意を向けうまく遂行している。	……	……	……	……	……
資料と根拠	当該の学問分野やジャンルにふさわしいアイデアを展開するために、質が高く、信頼でき、関連性のある資料をうまく使いこなしている。	……	……	……	……	……
構文と技法を操ること	読み手に明確かつ流暢に意味を伝えることができる格調ある言葉遣いをしてい。ほとんど全く誤りがない。	……	……	……	……	……

(注) レベル1を満たさないものはレベル0とされるので、実質的には5段階のレベルからなる。



【例】ローカライズされたVALUEルーブリック

	2	1	0	得点
文章作成の目的 (文章コミュニケーション)	・読者を考慮していること、および、研究の焦点が明確であることが、タイトルに示されていた。	
問題の特定 (文脈的能力)	・研究の主要問題が一般的な読者に関心のあるものになっていた。 ・	
エビデンス (批判的思考)	・ソース(情報源)からの情報を評価し、自分自身の解釈を生み出していた。 ・	
問題の説明 (批判的思考)	・すべての問題が記述され、明確化されていた。 ・	
構文と技法を操ること (文章コミュニケーション)	・使用された言葉は、読者に意味を明確に伝えていた。 ・	

33



● メリット

- 組織的利用を想定
 - ←大学・学部ごとにローカライズ
- 分析的で、一般的な、長期的ルーブリック
 - 1年次と卒業時の比較によって、「付加価値」を評価

● デメリット

- 教養教育における汎用的能力のみを対象
- 評価課題は、検討対象外

34



新潟大学歯学部での パフォーマンス評価の開発

35

パフォーマンス評価開発の経緯

- 2011年度
 - 初年次教育科目「大学学習法」のレポート評価のためのルーブリックを開発 (松下・小野・高橋, 2013)
- 2012年度
 - PBLテュートリアルの評価のためのパフォーマンス課題とルーブリックを開発(=改良版トリプルジャンプ)、トライアル実施
- 2013年度
 - 改良版トリプルジャンプを、口腔生命福祉学科2年生対象に実施(前・後期)

36

● PBLテュートリアルの評価

- 新潟大学歯学部の求める学習成果に対応したパフォーマンス課題
 - 「専門的能力」と「汎用的能力」の両方を含む
 - 「知識」の獲得も求める
 - 一部、「態度・姿勢」も含む
- 専門分野のルーブリック
 - 分析的 or 全体的 → 分析的
 - 一般的 or 課題特殊的 → 一般的 + 課題特殊的
 - 長期的 or 採点用 → 長期的

37

トリプルジャンプの検討: 先進校の訪問調査

- 2012年5月
 - 東京医科歯科大学(TJ実施中→改訂予定)
 - 昭和大学(eポートフォリオ)
- 2012年9月
 - アデレード大学歯学部
→TJは評価負担が大きく、信頼性に問題あり
- 2013年9月
 - ハワイ大学医学部
→TJを実施中。新潟大学の取り組みに対して好意的評価

38

評価の一般的要件

- 妥当性
 - 当初意図した内容や構成概念を評価しているか
- **信頼性(または比較可能性)**
 - 正確に安定して評価しているか
- カリキュラム適合性
 - 構成概念、領域、またはカリキュラムが明確に規定され、評価がカリキュラム全体をカバーしているか
- 公正性
 - 公正か(平等性、教育への影響、社会的承認など)
- **実行可能性**
 - 入手可能な資源と時間の限度内で評価できるか

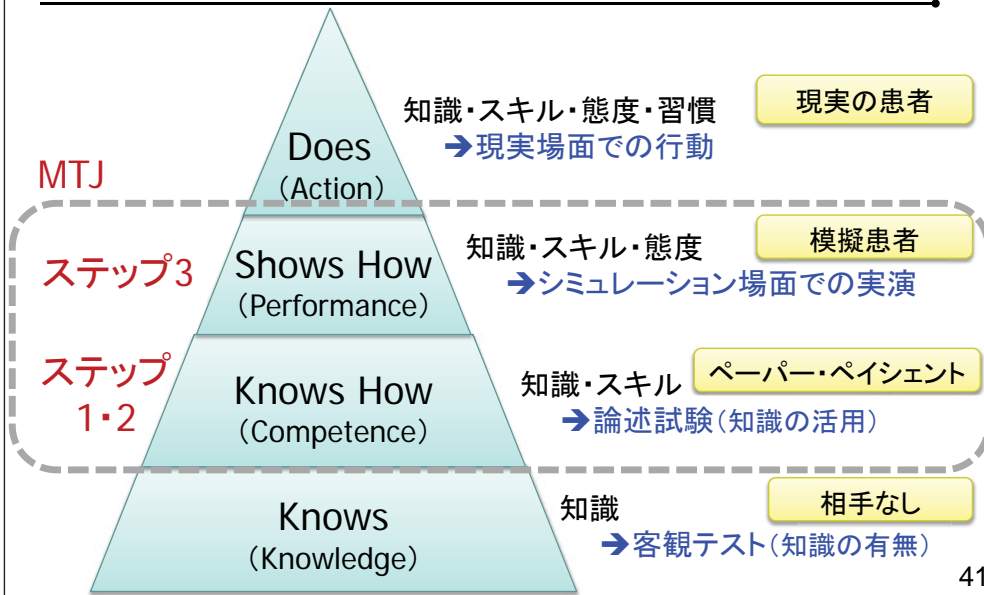
39

アデレード大学訪問調査からの宿題

- トリプルジャンプ(TJ)の改訂
 - 信頼性と実行可能性を高める
 - 総括的評価より、形成的評価の機能を高める
 - ↓
 - **改良版トリプルジャンプ**(Modified Triple Jump: MTJ)の**開発**
 - ステップ1・2とステップ3のルーブリックを作成し、信頼性を検定
 - ステップ1・2でのワークシートの活用により、評価負担を軽減
 - ステップ3で、パフォーマンス評価(ロールプレイによる実演)と教員からのフィードバック
- cf. 小野報告

40

学習成果にみあった評価③



学習としての評価



42

「学習としての評価」という考え方



- 3つの評価 (Earl, 2003; Alverno, 1994)
 - 学習の評価 (Assessment of learning) …総括的評価
 - 成績の決定／教師が評価主体
 - 学習のための評価 (Assessment for learning) } 形成的評価
 - 授業改善／教師が評価主体
 - 学習としての評価 (Assessment as learning) }
 - 学生自身の学習の自己調整／学生が評価主体
- 「学習としての評価」という考え方の拡張
 - 学習の自己調整 + 学習経験としての意味

43

「学習としての評価」としてのMTJ



- MTJ(トライアル)での学生の反応
 - 「やったらやった分だけ実になるような気がします。また、自分がどれくらい成長できるのかが分かるので、ただ学習しているだけよりも、もしかしたら手ごたえがあるかもしれません。実際一人ずつ評価コメントしなくてはならない先生方には大きな負担になるかもしれないのですが。」
- 「学習としての評価」へ
 - パフォーマンス評価は評価負担が大きい(=実行可能性が高くない)が、「学習としての評価」になりうるのであれば、実施する価値があるのではないだろうか？

44

ご清聴ありがとうございました



松下 佳代

matsushita.kayo.7r@kyoto-u.ac.jp

45

- 松下佳代・小野和宏・高橋雄介 (2013). 「レポート評価におけるルーブリックの開発とその信頼性の検討」『大学教育学会誌』第35巻第1号, 107-115.
- Miller, G. E. (1990). The assessment of clinical skills/ competence/ performance. *Academic Medicine*, 65 (9, Suppl.), S63-S67.
- 西岡加名恵 (2005). 「教育評価の方法原理」田中耕治 (編)『よくわかる教育評価』ミネルヴァ書房, 66-75 .
- OECD (2013). *AHELO feasibility study report, Vol.2*. OECD.
- 小野和宏 (2012). 「大学学習法へのパフォーマンス評価の導入—新潟大学歯学部の初年次教育の場合—」大学教育学会第34回大会ラウンドテーブル「学習成果の直接評価に向けて—パフォーマンス評価の可能性—」.
- 小野和宏・大内章嗣・前田健康 (2011). 「学習者主体PBLカリキュラムの構築—新潟大学歯学部口腔生命福祉学科7年のあゆみ—」『新潟歯学会誌』41(1), 1-12.
- Palomba, C., & Banta, T. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey Bass.
- Rhodes, T. (Ed.) (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: AAC&U.
- Rhodes, T. L., & Finley, A. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment*. Washington, DC: AAC&U.

47

文献

- Alverno College Faculty (1994). *Student assessment-as-learning at Alverno College*. Alverno College Institute.
- Benjamin, R. et al. (2009). *Returning to learning in an age of assessment: Introducing the Rationale of the Collegiate Learning Assessment*. (<http://www.collegiatelearningassessment.org/>)
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Gipps, C. V. (1994). *Beyond testing: Toward a theory of educational assessment*. Falmer Press. ギップス, C. V. (2001) 『新しい評価を求めて—テスト教育の終焉』(鈴木秀幸訳) 論創社.
- 平井洋子 (2006). 「測定の妥当性からみた尺度構成—得点の解釈を保證できますか—」吉田寿夫 (編)『心理学研究法の新しいかたち』誠信書房, 21-49.
- 松下佳代 (2007). 『パフォーマンス評価』日本標準.
- 松下佳代 (2010). 「学びの評価」佐伯胖監修『「学び」の認知科学事典』大修館書店.
- 松下佳代 (2012). 「パフォーマンス評価による学習の質の評価—学習評価の構図の分析にもとづいて—」『京都大学高等教育研究』第18号, 75-114.

46

- Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. San Francisco: Stanford University Press.
- Wiggins, G. & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development. / ウィギンズ, G. & マクタイ, J. (2012). 『理解をもたらずカリキュラム設計—「逆向き設計」の理論と方法—』(西岡加名恵訳) 日本標準.
- 山田礼子 (2012). 『学士課程教育の質保証へむけて—学生調査と初年次教育からみえてきたもの—』東信堂.

48