

CAUAシンポジウム 2021

# 高等教育における 新しい理解度評価の考え方

CAUA会長  
日本オープンオンライン教育推進協議会(JMOOC) 副理事長  
早稲田大学 理工学術院 教授  
深澤良彰

2021-02-16

# (みなさんご存知) 大学の「おざなり」な定期試験

高校までと大違い!

- 大学の教員 (のほとんど) は、教育のための訓練を受けていない
  - 「教員免許」をもっていない
    - そもそも「大学の教員免許」なんか存在しない
  - 自分が学生だった頃受けたのと同じような試験を実施
- 教育よりも研究に力を入れたい
  - 教育の成果よりも、研究の成果の方が評価される
    - 早稲田の任用履歴書でも、10年位前にやっと教育業績の欄ができた
  - いわゆる「研究型大学」において顕著
- 具体的には、
  - 採点済みの答案は学生に返却しない
  - 配点を明確にしない
  - 毎年、同じような問題を出す

## (某W大学で見た) 実例

問1 次の用語について、簡単に説明せよ

- (1) レジスタ
- (2) 仮想記憶
- (3) キャッシュ記憶
- ...

ある大教授の悪口になりますので設問は変えています

- 私が、学部2年の時（20歳）に受けた問題と早稲田に戻ってきて一緒に教えるようになった時（34歳）の問題が同じ
  - きっと私が習う前も、そして、私が一緒に教えなければ、その先生の定年まで
- 採点方法：ある用語を説明するのに必要な内容をチェックし、その数だけ加点
  - 各設問の配点が違う
  - 学生には知らせていない
- もちろん、答えは返却しない

# こんな状況に2つの大波 (1/2)

## ● コロナ禍

### – 授業がオンライン化

- 頑張っ、オンデマンド授業やリアルタイム授業で対応

### – 通常の試験ができない

- レポート課題にすれば問題は少なく、多くの教員がこれを選択
- 最終目的が国家試験である場合には、模擬試験をやりたい
- 実技試験／実習試験をどのようにするか
- リアルタイム授業の場合には、対面の試験と同じように試験を実施できるか
  - 学生のPC環境、ネットワーク環境に左右されないか？
  - うまく入力できない等、コンピュータの利用能力に左右されないか？
  - TV会議システムで監視ができるか？
    - » カンニングの予防は大丈夫か？
    - » 学生個々が有するアプリに左右されないか？

## こんな状況に2つの大波 (2/2)

- MOOCの普及

- 受講生は分散 (世界中、日本中)

- 受講生を一箇所に集めての試験は不可能

- 受講時間もバラバラ

- 試験時間を固定できない

- 場合によっては、数万人以上の受講者

- 選択式とか○×式とかの設問なら機械採点可能だが

- 記述型 (例: 「証明せよ」) は必要不可欠

- 最終試験をパスして、修了証を希望する受講生が存在

- 勤務内容のステップアップ

- 授業の単位

## つまり（今日の内容）

大学を中心とする高等教育機関（の多く）においては、  
学生の理解度の評価方法は重視されてこなかった

コロナ禍によるインター  
ネット授業の実施

MOOCの一般化

新しい理解度評価の出現

理解度評  
価の基礎

オープnbック  
方式の試験

相互評価  
の導入

...

## 評価の意義

単に「成績をつける」だけではありません！

- 教員にとって
  - 学生の理解度の確認・支援
  - 授業の改善
- 学生にとって
  - 到達度の把握
  - 学びの支援
- 組織にとって
  - 質保証
  - 説明責任

評価はゴールではなく、スタート

「卒業とは、出口じゃなく、入口だよ」  
In "GIVE ME FIVE!"  
by AKB48

# 評価指標

- 何を評価対象とするのか？
  - 認知的領域：知識・理解, 思考・判断
  - 情意的領域：関心・意欲・態度
  - 精神活動的領域：技能・表現、スキル



目標に対応

何を評価しようとしているのか？  
何のための評価なのか？

# Bloom の教育目標分類 (タキソノミー)

3つの領域

	認知的領域 (知識)	情意的領域 (態度)	精神運動的領域 (スキル)
高次	評価		
	統合	個性化	自然化
	分析	組織化	分節化
	応用	価値付け	精緻化
	理解	反応	巧妙化
低次	知識	受け入れ	模倣

(Bloom, B.S. 1956)

同種のものも、いろいろあります

例：Finkの「意義ある学習」に関する分類

# 総括的評価と形成的評価

- 総括的評価 (Summative)
  - 達成された学習成果の程度の把握を目的とする
  - 合否判定
  - 学習後に実施
  - 範囲は、広く、発展課題も含む
- 形成的評価 (Formative)
  - 学習プロセスの改善を目的とする
  - 学習活動の逐次修正につながるフィードバック
  - 学習中に実施
  - 範囲は、狭く、学習内容のみ

(Bloom, B. S. 1968)

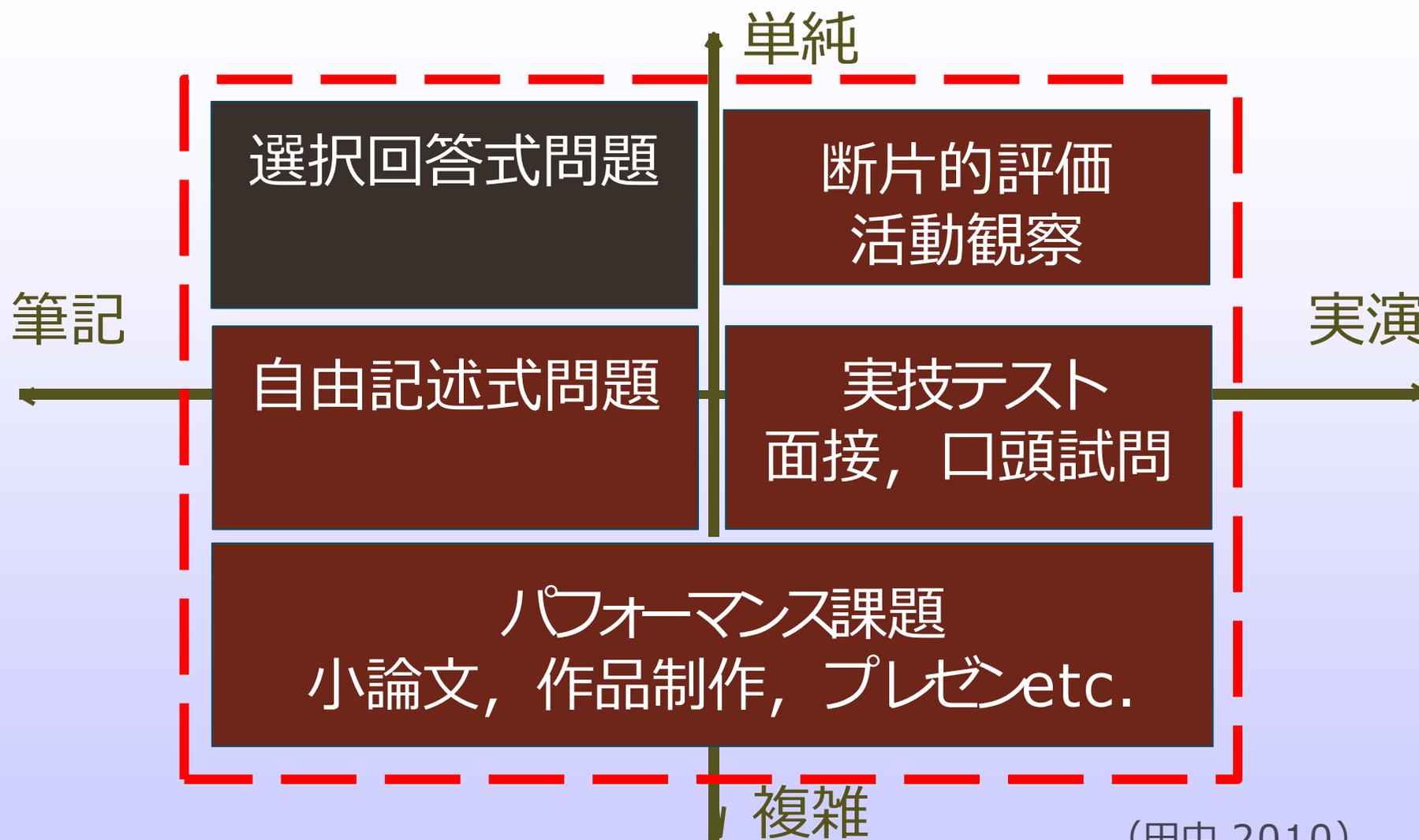
- \* 「どちらか」ではない  
「形成的側面を持った総括的評価」  
その逆もあり

# 「評価」の評価

- 信頼性 Reliability
  - 結果の再現性. テストの精度
  - 同じ集団に同質の試験を何回行っても同じ結果が得られる程度
- 妥当性 Validity
  - 評価方法の適切性
  - 用いる評価方法が測定対象となる能力や行動を測定できているかどうか
- 効率性 Efficiency
  - 評価の時間的経済的な実用性
  - 実施や採点が容易であるかどうか

他にも「〇〇性が重要」という話はたくさんあります

# 評価の方法



(田中 2010)

従来は、これらを使い分けてきている

# 各方法はオンライン授業でも実施可能

問題は厳密にできるか！

単純

録画

Webフォーム

選択回答式問題

断片的評価  
活動観察

Web会議システム

Webフォーム

筆記

実演

自由記述式問題

実技テスト  
面接, 口頭試問

Web会議システム

パフォーマンス課題

小論文, 作品制作, プレゼンetc.

Web会議システム

文書・録画  
・画像提出

複雑

# オンラインで学習を評価するための 10のポイント

- 1. 教室での一斉試験をオンラインで再現することは難しい
- 2. 学問的誠実性について教える
- 3. 学習目標を確認する
- 4. 形成的評価を積極的に取り入れる
- 5. ICTツールを活用して評価する
- 6. 問題を工夫する
- 7. 解答時間を制限する
- 8. 出題パターンを増やす
- 9. ピアレビューを導入する
- 10. 各種リソースを活用することを推奨する

# オンラインで学習を評価するための 10のポイント

- 1. 教室での一斉試験をオンラインで再現することは難しい
- 2. 学問的誠実性について教える
- 3. 学習目標を確認する
- 4. 形成的評価を積極的に取り入れる
- 5. ICTツールを活用して評価する
- 6. 問題を工夫する
- 7. 解答時間を制限する
- 8. 出題パターンを増やす
- 9. ピアレビューを導入する
- 10. 各種リソースを活用することを推奨する

# 問題を工夫する

教科書や授業で扱った多くの知識を暗記し、それを筆記で再現させる試験問題の代わりに、問題を厳選し、深い思考を求める問題にする

たとえば、

- 学習した知識を踏まえて応用問題を解くような問題を課し、あわせてその解答手順や用いた法則・原理も記述させる
- 学習した概念やキーワードの関係性についてコンセプトマップ（概念間の関係を線でつなぐなどして図示化させたもの）で図示化させ、説明を求める
- 学習した知識を使って自ら問題を創り、自ら解答させる



コロナ禍でどのように実現するのか？

# オーブンブック方式

- 試験は、通常、教科書や参考資料を使用せずに行われる  
(クローズドブック評価と呼ばれる)
  - オンライン授業では試験環境を制御することはできない
    - 本人同定：きちんと本人が受験しているか確認が必要
      - カメラで確認しても、デジタル画像が加工されている可能性も
    - 試験中に学生は、インターネットを使用して情報を検索する可能性がある
    - カンニングなどの不正行為の防止
      - カメラで監視したとしても、視野外に、資料やノートPCなど置かれては対応できない
      - LINEなどのコミュニケーションツールの利用もできる
    - 答案の類似度判定をしても100%検出することは難しい
- 
- 試験中に参考資料等何でも使用可能とする  
(オーブンブック評価)

「参考書持ち込み可」の試験のようなもの

# オープンブック方式のメリット・デメリット

## メリット：

- 丸暗記よりは応用力を求められるテストの実施 [1]
- テストに合格するために暗記を減らすことにより、より高い認知レベルの思考を奨励する [2]  
例えば：情報集約の能力とか、批判的な思考力など

## デメリット：

- 教師： 学生が知識を適用することを要求する効果的な試験問題を作成する工夫が必要（知識の暗記を試す試験問題は利用できない）
- 学生： 通常の試験ではないので予習・復習が必要ないと勘違い！して勉強しなくなる

[1] Heijne-Penninga (2008) Open-book tests to complement assessment-programmes: analysis of open and closed-book tests.

[2] Eilertsen (2000) Open-Book Assessment: A Contribution to Improved Learning?

参考：フラナガン・ブレンダン「Open e-Book Assessmentによる成績評価」 第8回  
4月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム2020.5.15

# オープンブック評価を設計する際の考慮

- 1)問題は、単に知識の記憶ではなく、知識の理解や応用力、情報集約力、批判的思考力などを評価するものにするべき
- 2)問題は、シナリオや事例から推測して知識を応用するようにするべき
- 3)問題は、テキストや記憶を利用して回答までに、あまり時間がかかりすぎないように、明確なものにするべき
- 4)問題は、単にテキストやノートの情報を写すだけでなく、情報を集約して理解したことを書くようにするべき
- 5)学習目標を達成したかどうかを問うように問題を設計すべき

参考 : The University of New Castle: Centre for Teaching and Learning  
(2017) A Guide for Academics – Open Book Exams.

参考 : フラナガン・ブレンダン「Open e-Book Assessmentによる成績評価」 第8回  
4月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム2020.5.15

## オープンブック評価の問題の例

1. OOについて、その背景となる情報を書きなさい。
2. OOのデータはどのようなことを示しますか？
3. AとBを比較して、その特徴の違いを述べなさい。
4. OOによって、どのような影響が生じますか？
5. OOに関するいくつかの手法のうち、最も良いものはどれですか？ その理由を書きなさい。

参考 : The University of New Castle: Centre for Teaching and Learning  
(2017) *A Guide for Academics – Open Book Exams*.

参考 : フラナガン・ブレンダン「Open e-Book Assessmentによる成績評価」 第8回  
4月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム2020.5.15

# オープンブック評価のまとめ

オンライン授業では、オープンブック評価がよく行われる

- 教員: 問題を作成する工夫が必要
- 学生: オープンブック試験は十分な予習が必要
- それは本当か？ いくつかの検証結果が存在
  - オープンブックとクローズドブックの成績との間には正の相関がある
  - オープンブック試験で成績上位者は、成績下位者と比較して、多くの予習をしている
  - 検証を行うためには。学習ログが重要



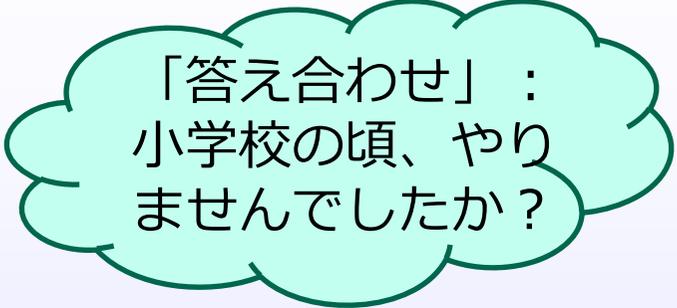
学生はどの問題が理解できていなかったか？ どのページをよくみていたか？等を分析して教材や授業の改善に活用

# オンラインで学習を評価するための 10のポイント

- 1. 教室での一斉試験をオンラインで再現することは難しい
- 2. 学問的誠実性について教える
- 3. 学習目標を確認する
- 4. 形成的評価を積極的に取り入れる
- 5. ICTツールを活用して評価する
- 6. 問題を工夫する
- 7. 解答時間を制限する
- 8. 出題パターンを増やす
- 9. ピアレビューを導入する
- 10. 各種リソースを活用することを推奨する

## 相互評価 (Peer Assessment/Peer Review)

- 学生が相互に評価を行う ● ● ●
- MOOCの大流行で脚光
  - 数万人のレポート、答案の採点
- 明確な基準が示されていれば信頼性あり (= 教員による評価との相関有り)
  - Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of educational research*, 70(3), 287-322.
- 自己調整型学習の手段
  - Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3), 245-281.



「答え合わせ」：  
小学校の頃、やり  
ませんでしたか？

# 相互評価の具体的手続き

- (教員) 課題を提示する
- (学生) 課題に取り組み提出する
- (教員) 学生に3-5名の課題の評価を無作為に割り当てる
- (教員) この評価のための評価基準の提示をし、学生に練習をさせる
- (学生) 割り当てられた課題の評価を行い、結果を提出する
- (教員) 学生の評価結果をとりまとめる
  - 成績に組み入れる場合には、学生の評価結果を採点案として扱い、確認した上で扱う
  - 各学生に評価結果を返却する

# 相互評価参考情報

- マギル大学
  - <https://www.mcgill.ca/tls/instructors/assessment/peer>
  - かなり詳しい解説や課題例が紹介
- アイオワ州立大学
  - <https://www.celt.iastate.edu/teaching/assessment-and-evaluation/peer-assessment/>
  - 手続きも含め、文献を引用しながら具体的に解説
- MOOCプラットフォームedXのスコア計算方法
  - <https://support.edx.org/hc/en-us/articles/360000192027-How-are-peer-assessment-scores-calculated->
  - edXはスコア計算に中央値を利用

# ルーブリックとは？

相互評価：明確な基準が示されていれば信頼性あり

ルーブリック

- ある課題をいくつかの構成要素に分け、その要素ごとに評価基準を満たすレベルについて詳細に説明したもの  
(スティーブンス & レビ, 2014)
- プロジェクトやパフォーマンスを評価する課題における評価軸の可視化
  - レポート課題、演技などの評価しづらいものの評価方法

参考：フィギュアスケートの採点

演技構成点:スケーティング技術、技のつなぎ、演技、構成、音楽の解釈の5項目に分かれ、スケーティング技術なら「流れと滑り」など、各項目について3～6種類の尺度があり、5項目で計23もの基準に照らして採点

# ルーブリックの基本構成

<b>課題</b> ループによるオムニバス講義の模擬授業を評価するためのルーブリック			
グループ ( )		年 月 日	
<b>評価尺度</b>	Excellent	Good	Developing
<b>構成</b>	全体として統一感がしっかりとあり、よい構成であった	全体として統一感をもたせようとした努力がみられた構成であった	全体として統一感に乏しく、各トピックがばらばらな印象を受けた
<b>評価観点</b>			
<b>レベル設定</b>	初学者にとって、全体が「少し手を伸ばせば届くレベル」の適切な教授内容であった。	初学者にとっておおむね「少し手を伸ばせば届くレベル」の適切な教授内容であった。ただし、あるいはやさしすぎる箇所があった。	初学者にとって全体が高度すぎる、もしくは、極端にやさしすぎる内容であった。
<b>学習意欲の喚起</b>	授業内容についてさらに自分で勉強してみたいと興味を	授業内容について授業中はとも興味をもって聞いて満足したが、授業後にさらに勉強し	授業内容について特に興味を

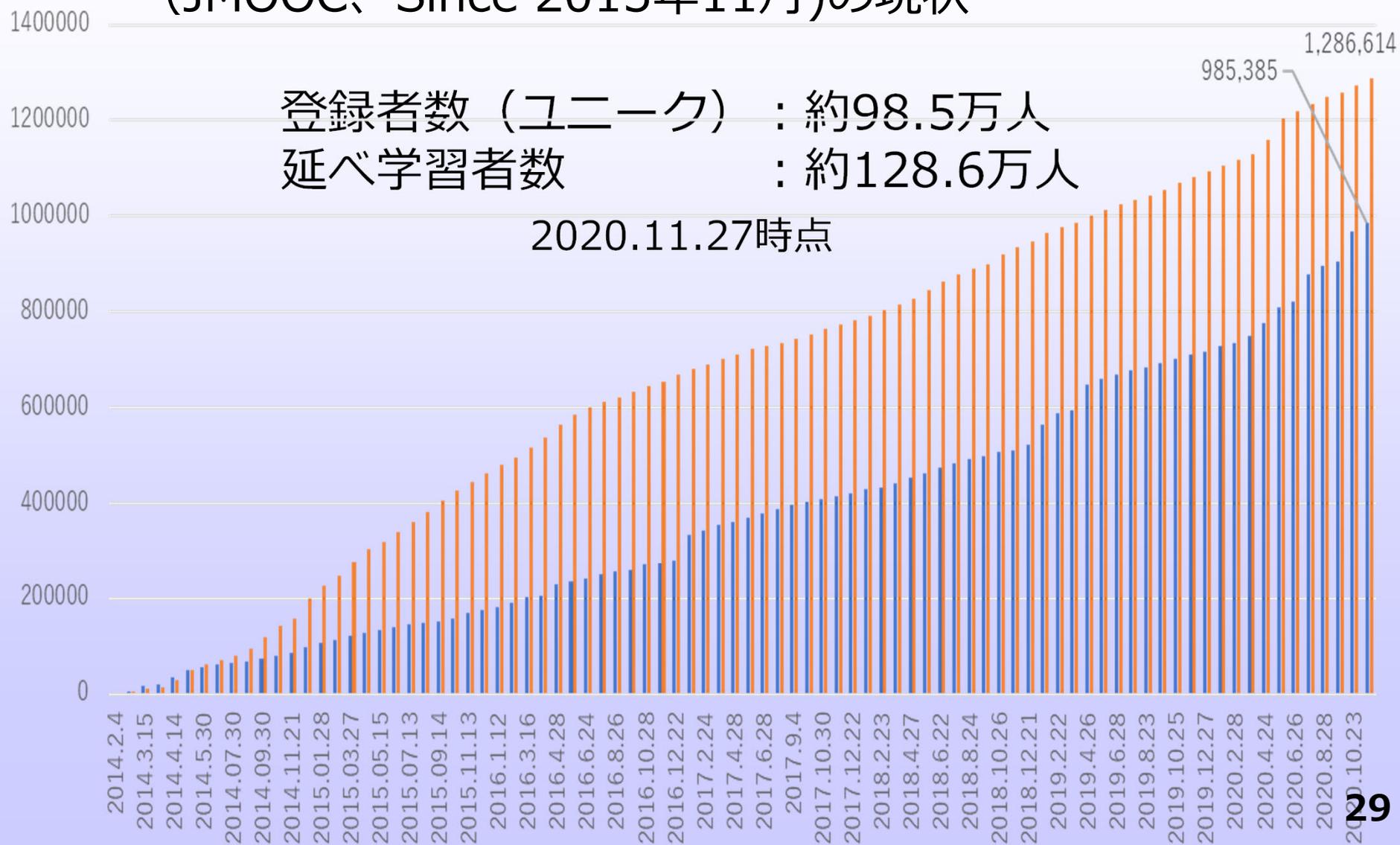
# ルーブリックを積極的に活用した 学生による相互評価例

基本的には「相互評価の具体的手続き」(P24)参照

1. (教員) ルーブリック作成
2. (教員) 学生に課題とルーブリック配布
3. (学生) ルーブリックの理解 (優れた例の配布、評価の練習)
4. (学生) ルーブリックを学習指針として課題に取り組む
5. (学生) ルーブリックを課題に添付し提出 (自己採点をして提出も有)
6. (学生) 相互採点 (場合によっては複数、評価を行う場合は中央値、平均値など)
7. (教員) 相互採点されたものを最終評価/確認
8. (教員) 学生に返却

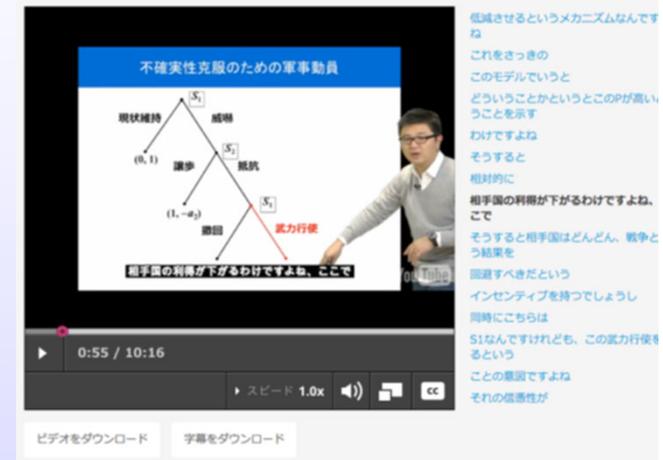
# JMOOCにおける相互評価の評価結果 (1)

一般社団法人 日本オープンオンライン教育推進協議会  
(JMOOC、Since 2013年11月)の現状



## JMOOCにおける相互評価の評価結果（2）

- 対象講義：「国家安全保障論」（栗崎周平）2014年
- JMOOC公認プラットフォーム「Gacco」を利用
- 受講登録者数：12068人
- レポート
  - 頻度：1回／週（合計4回）
  - 文字数：400文字～800文字
  - 期限：出題後2週間以内
- 修了者数：1300人（修了率10.8%）
- 相互評価：5人以上／1レポート
- ルーブリック項目数：  
4項目、4項目、8項目、11項目／各週
- ルーブリックには評価基準の記述だけでなく、レポートのポイントとなる解説も提示
- 有効アンケート回答者：605人



# JMOOCにおける相互評価の評価結果 (3)

## ループリックの例

### 第1回レポート

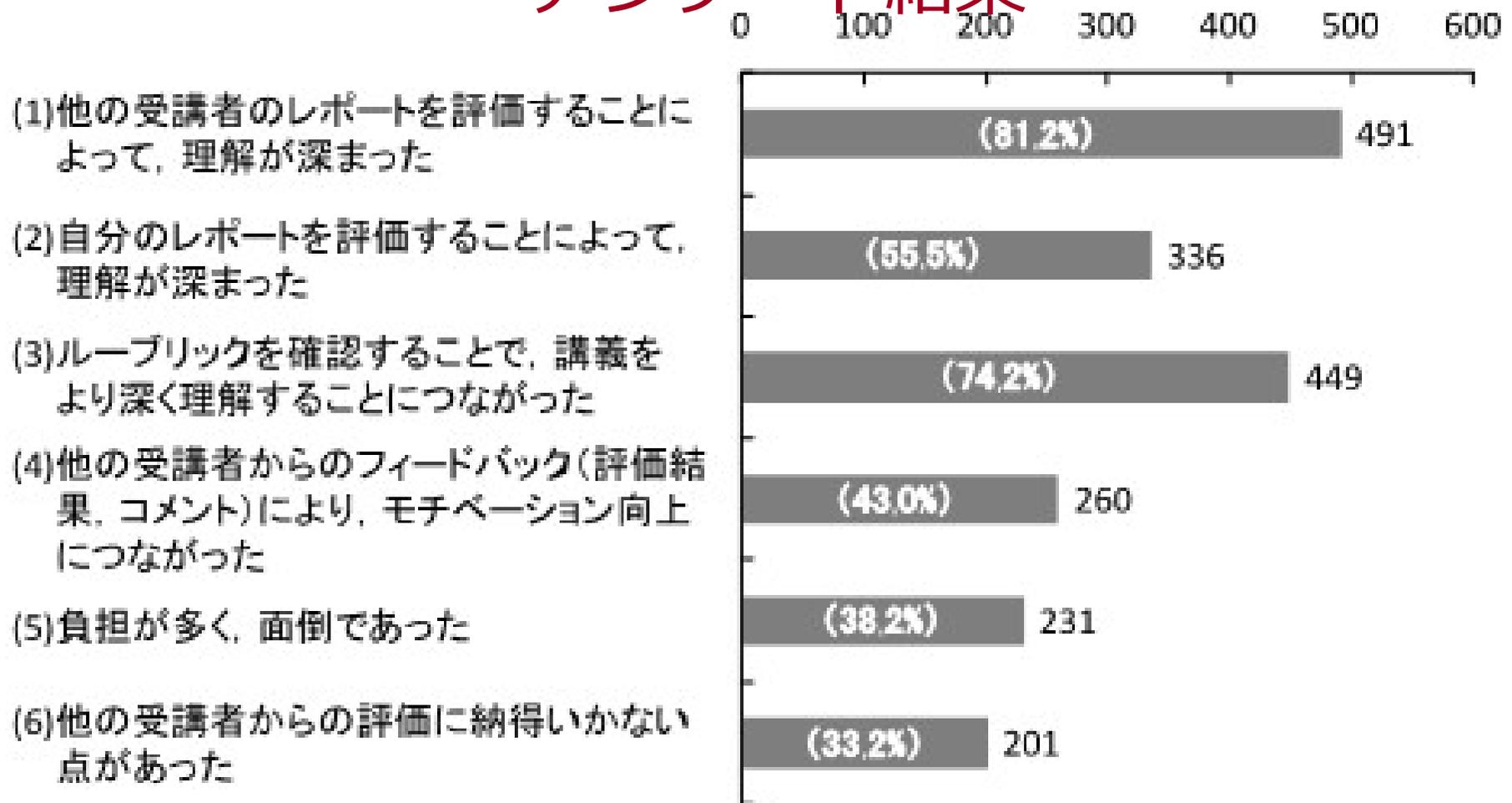
カール・フォン・クラウゼヴィッツは『戦争論』において、戦争を「他の異なる手段を用いた政治過程の延長」、「相手（敵）に当事者の意思を受け入れられるための暴力行為」と定義した。この定義に基づいて戦争が発生することが、なぜ「パズル」であるのか400字程度で説明せよ。

項目1: 1点、項目2: 1点、項目3: 2点、項目4: 1点

- ループリック：項目1：
  - 「通常の政治過程」と「戦争」という二つの手段があることを明示的に言及している。あるいは、レポートの内容からして、このことが前提とされていることが明白：1点
  - 「通常の政治過程」と「戦争」という二つの手段があることが示されていない。あるいは、この点を前提としても踏まえているとは推察できない：0点

# JMOOCにおける相互評価の評価結果 (4)

## アンケート結果



### 相互評価を行った感想 (複数回答可、n = 605)

渡邊文枝, 森裕樹, 向後千春 (2015) JMOOC の講座における相互評価に対する受講者アンケートの分析. 人間科学研究, 28(2): 237-245

## JMOOCにおける相互評価の評価結果（5）

- 相互評価を行うことによって理解が深まると思う学習者は、モチベーションが向上
- 相互評価に対して負担が多く、面倒であると思う学習者は、他の学習者からの評価に納得いかない点がある
- 相互評価を行うことによって理解が深まり、モチベーションが向上すると思う学習者は、負担が多く、面倒であると思わない
- 相互評価に関する肯定意見としては、理解の深化、相互評価の仕組みに対する満足感、新たな視点の獲得など
- 相互評価に関する否定意見としては、負担の大きさ、評価の難しさ、評価人数の多さなど



- 相互評価の導入は、学習者の主観的理解度や満足感を高めることにつながる
- 相互評価を導入する際には、学習者への負荷が高くなりすぎないように考慮して講座を設計する必要あり

# ルーブリックに関して今後やるべきこと (の一つ)

- ルーブリックの項目の信頼性・妥当性の確認
  - 汎用の信頼性・妥当性の評価方法は存在しない
  - 分野を限定すれば、存在する
    - プログラミングの分野[1]
  - ルーブリックの改善に関する研究[2]
    - 適用対象：小学校のプログラミング教育
    - 内容：既開発のルーブリックおよびその改善版を評価・分析
    - 結果：改善後のルーブリックは信頼性と妥当性が上昇



- ルーブリックの改善をしていく体系の必要性

[1] Nathalia da Cruz Alves et al., “A Largescale Evaluation of a Rubric for the Automatic Assessment of Algorithms and Programming Concepts”, 51st ACM Technical Symposium on Computer Science Education, 2020.

[2] Daisuke Saito, Yoshiaki Fukazawa et al., “Assessing Elementary School Students’ Programming Thinking Skills using Rubrics”, IEEE International Conference on Teaching, Assessment, and Learning for Engineering, Dec. 8-11, 2020.

# オンラインで学習を評価するための 10のポイント

- 1. 教室での一斉試験をオンラインで再現することは難しい
- 2. 学問的誠実性について教える
- 3. 学習目標を確認する
- 4. 形成的評価を積極的に取り入れる
- 5. ICTツールを活用して評価する
- 6. 問題を工夫する
- 7. 解答時間を制限する
- 8. 出題パターンを増やす
- 9. ピアレビューを導入する
- 10. 各種リソースを活用することを推奨する

まとめに代えて

## 大学におけるCOVID-19三段跳び

- ホップ（第1期）（2020年度から）  
もに大学に来ることなく、な  
ること

学習の評価に  
焦点を絞って！

- ステップ（第2期）：ネットワークを使った研究・教育  
ならではの「良さ」を、研究・教育の中に活かすこと
- ジャンプ（第3期）：第1期、第2期の結果を用いて、  
大学における研究・教育のDXを引き起こすこと



教育では？



研究では？



Thank you...



なお、本資料は、深澤自身の独断と偏見に基づいている部分が多く、決して早稲田大学およびJMOCの統一的理解ではありません。

ご質問・ご意見等は  
fukazawa@waseda.jp まで

