

## Assessment & Teaching of 21<sup>st</sup> Century Skills DWP2 Perspectives on methodological issues

「さまざまなステークホルダーにどう結果を伝えるか」が最重要



受検者の将来に影響を与えるので

21 世紀スキルの特徴「creativity and innovation・collaboration (teamwork)・information literacy」として、以下 DWP2 について述べる。

一般論として、アセスメント設計では以下を考える。

- ・ 評価される構成概念の選択：これらの技術はドメイン非依存か、それとも依存か？
- ・ 技術の年齢幅：被験者がK-12・高等教育、それ以上か？
- ・ 分析のレベル：個人・チーム・クラスまたは大きなグループ？
- ・ 文化依存か、非依存か？

### Inferences, evidence and validity (結論、証拠、妥当性)

前提：教育アセスメントのデザインは The evidence-centered approach

21 世紀型スキルのアセスメントには、まず「構成概念の定義」が重要である。これを決めて Figure 1 のサイクル (行動選択・発表・証拠確認・証拠蓄積) を回して洗練させる。



Figure 1: The four-process architecture (Almond, Steinberg & Mislevy, 2003)

その際、“科学的手法”で洗練させて、信頼性を保つ。つまり正例だけ集めてもっともらしいものを作らないことも必要である。

### **Assessment design approaches (アセスメント・デザイン・アプローチ)**

アセスメントシステムの成功のために次の3つの項目が必要である。

- (a) 測定する構成概念の定義とその詳細。
- (b) それらの定義が構成概念のタスクや測定手段を開発・選択するのにどう寄与するか。
- (c) どのように児童・生徒の反応をコード化・分類・定量化・量子化し、構成概念にどう結びつけるか。

### **Defining the constructs (構成概念の定義)**

明確で意義のあるスキルの定義が重要（分析の単位・対象年齢・普遍的なものか・特定の背景に限定されるものなのか・文化的依存性があるか・特定の context や領域に限定されるものか）。

それぞれのスキルの定義によって設問が変わる。

### **Structuring a developmental definition (発展的な定義の構築)**

CCII レポートに基づいて4つの要素に分解。これはスキルの発展的定義を構築するのに必要と考えられる。

- 1) learning Targets (学習目標) : 熟達したということはどのような状況かを明確に定義することが最重要。またそれは指導の結果である必要がある。
- 2) Progress variables : 学習目標の作り上げによってその領域の中心的なテーマが明確になる。そのテーマは中心的な概念構造となり、それぞれのスキルで反映されなければならない。このテーマをどれだけ学んだかを示すものが、1つ以上の progress variable である。
- 3) level of achievement (達成度) : 習熟度には段階がある。
- 4) learning performance (学習パフォーマンス) : 内容標準に到達できた時に、個人が何ができるかの詳細を明らかにしたもの（たとえば、学生が現象を説明できるとか、仮説が検証できるとか）。

### **Assessment of progressions (発達の評価)**

Learning progression (学習の進展) は学習者が領域の取得に向かってたどる道筋を表すもので、その妥当性を追求しなければならないが、(テストの) 実施者に実用的なツールを提供し教授活動を構成するモデルを提供する。従って、学習の進展のための測定は研究と実用の両面にとって重要である。



## **New Tasks for 21<sup>st</sup> century skills (21世紀型スキルの新たなタスク)**

21世紀型スキルの evidence (証拠) にはパフォーマンステストが含まれる。パフォーマンスアセスメントには2つある。

- 1) **Response-centered** 生徒の答えをもとに
- 2) **Simulation-centered** シミュレーションをもとにした定義 (パフォーマンスを観察して学生のモデルをつくる)

technology-mediated (テクノロジーの活用) によるアプローチはどちらにもあり、もっと21世紀型スキルの評価ではもっとつかうべき (エビデンス収集・シミュレーション・仮想世界・センサーなど)。膨大なデータの収集が可能になり、テクノロジーを活用することによって、今まで紙と鉛筆では不可能であった評価が可能になる。

- 1) 総括的評価と形成的評価の組み合わせ：テクノロジーの活用により、総括的評価と形成的評価の結果を合わせての児童・生徒の学びの全体像をつかみ易くなる。教育に使う。
- 2) 集合知：評価の証拠として使えるものが増える。
- 3) タスク分析：テストを学習過程に埋め込み、複雑なタスク分析が可能になる。
- 4) テスト項目の埋め込み：さまざまな埋め込みが可能になる。

## **Valuing the responses (反応・応答の得点化)**

タスクを通して集められた反応とプロダクトを分析するために、生徒のパフォーマンスを分類する質的な異なるカテゴリーが必要になる。

- 1) **Research-based categories** (調査に基づくカテゴリー)：認知と学習の調査に基づくモデルが構想概念の基礎であるべき。
- 2) **Context-specific categories** (背景に特化したカテゴリー)：outcome space は背景に特化すべき。
- 3) **Finite and exhaustive categories** (有限で包括的なカテゴリー)：outcome space はカテゴリー数が限定されていて、可能性のある答えをすべて包括するものでなければならない。
- 4) **Ordered categories** (順序付けされたカテゴリー)：理論と経験に基づくカテゴリーの順番付けが必要である。

以上の4つの基準にあった outcome space の開発は成績基準と評価を教師、児童・生徒、保護者、管理者等にとってもはっきりと曖昧性のないものにする。

## **Delivering the tasks and gathering the responses (試験の配信と答案の収集)**

「どのようなメディアでテストを提供するか」ということは評価にとって大切である。コンピューターで実施するかどうかは設計に影響を及ぼす。また、課題もある。

社会的な側面：「一斉試験を期待する」「コスト」「コンピューターの数の問題」

技術的な側面：「セキュリティ面」「ネット接続環境の有無」「通信の信頼性」

利点は、データが膨大に取れることである。

これから可能性がある試験

- ・ Computer Adaptive test・・・特定問題の消費の問題はあるが、解決策もある。
- ・ Situational Judgment Test・・・状況を判断するもの。
- ・ レスponsがある試験・・・speech actのようなもの。

## **Modeling the responses (反応・応答の分析モデル化)**

モデル化は必要である。分析モデルによって求めたい結果が異なるため、必要に応じて適切なモデルを選ぶ必要がある。

たとえば、グループ別の結果がほしいならば、加重分析やマルチレベルモデル（階層線型モデル）が適切である。

## **Validity evidence (妥当性の証拠)**

他の変数との関係とテストの結果に関する有効性の証明について論じる。

### 1) Relations to other variables (他の変数への関係) :

他の変数との関係で妥当性をチェックしましょう。

### 2) Reporting to users (ユーザーへの報告)

結果はアクションを起こせるものでなくてはならない。関係のある結果はすべてのレベルの関係者に伝えられるべきであり、エンドユーザーはその結果を教育の改善に役立てられるような研修を提供されるべきである。生徒はその報告によって成績向上の環境が作られるべきであり、先生は自分のクラスの傾向を知って改善ができ、社会は教育のシステムの改善が可能になるべきである。

<このレポートでは結果返却に関して気を遣っている>

## Issues in the assessment of 21<sup>st</sup> century skills (21世紀型スキル評価の課題)

- 1) **Generality vs context specificity** (一般かコンテクスト依存か) : 21世紀型スキルの構想概念では、背景による特異性は抹消のこととなり、背景の中で21世紀型スキルを測ることが難しい。また評価の目的も、たとえば教室での形成的評価と全国的な類型的評価によって違うということを考慮すべきである。また初心者と専門家の違いも考慮すべきだ。この件は、学習目標・発達変数・到達レベル・成績に実質的な示唆がある。
- 2) **Large-scale and classroom assessments** (大規模評価と学級内評価) : ユーザーの明確化と分析と報告のレベルの定義が必要だ。教師と政府の両方の懸念を考慮する必要がある。ただテストを追加することによって教師や児童の負担が増えすぎないようにする必要がある。学級で行われるテストに大規模評価で使われる指標を埋め込む方法が考えられるが、その有効性については政府等から疑問がでることも考えられる。また、標準化された大規模評価の結果をどう学級のレベルに使えるかということも課題となる。
- 3) **What can new advances in technology bring to assessments?** (テクノロジーの進化が評価に何をもたらすか) : ダイナミックな視覚や音声やインタラクティブ性・適応性・即答性に加えて、群衆の智・個別化・Adaptive recommender system・Stealth 評価がある。群衆の智に関しては、測定のゴール・観察・スコア付け・測定モデルをさまざまなグループの標準から考えられることである。Adaptive recommender system は、評価プロフィールを情報源と情報探求者を仲介する。Stealth assessment は、反応者に必ずしも評価と意識されないように学習やソーシャルネットワーキングの過程で行われる。個別化はこの Stealth 方法を意思決定に焦点をあてたものといえる。

## Examples of types of measures (測定の例)

- 1) **Assessment of new skills** (新しいスキルの評価) : 先進国では学校のカリキュラムは文化から昇華されたものである。21世紀型スキルにもその方法を適用すると困難である。定義がうまくいったとしても、評価に沢山の項目を要し時間がかかることは避けられない。
- 2) **Self assessment and peer assessment** (自己評価と同僚評価) : 創造性のような特定のスキルはグループ活動で発揮され、自ら学ぶ、互いに学ぶということを啓発できる。いずれにしても適切な構成概念が不可欠である。
- 3) **Creativity/problem solving** (創造性と問題解決) : 創造性は定義も多数あり、評価できるかということは論争の最中である。
- 4) **Group measures** (グループ測定) : 個別化とは個々の児童・生徒が集団とのかかわりと個人の課題に対応できる環境を提供することである。構成概念は21世紀型スキルを集団という観点から捉え、群衆の智を活用して仲間からのフィードバックを活用する。観察は州案の中でさまざまな観点から個人を評価し、予想する。結果空間は個人の成績と集団の成績の2つの面で採点され、集団は方略的に構成される。測定モデルは個人の

成績をさまざまな集団の中で採点し、個人と集団の両面から評価される。

- 5) **Biometrics** (生物学的測定) : 言語外の物理的な行動を測定するもので、反応速度や視線の追跡などがあり、新しい分析が可能となる。

## **Conclusion (結論)**

21 世紀型スキル評価の方法的課題について答えを出すのではなく、問題提起と新しい視点の洗い出しを行った。

6つの部分からそれぞれ課題を選んだ :

1. 構成概念の定義 : 背景 (**context**) の役割を認知的構成概念とどう区別するか、またその必要があるか。 **context** (お釣りが出せる) と、認知的構成概念 (引き算ができる) をどう区別するか、また区別の必要があるか? 今までは、構成概念はテストできるが、**context** はテストの対象外として切り離していた。
2. タスクのデザイン : コンピューターやネットワークによって可能になった新しい試験項目 (**adaptive test**、キーストローク、習熟度ポートフォリオ) は従来の構成概念を変えるか? その場合問題が生じるか (受け入れ等) ?
3. 反応・応答の得点化 : 先生の授業改善のためにコンピューターによる自動採点の良い点 (すぐに結果が得られ、すぐに授業に反映できる) と悪い点 (主観評価ができにくくなる) のバランスは? 可能であっても使用しない方がよい場合があるか?
4. 試験の配信と答案の収集 : 新しい評価方法からデータストリームが多量に入手可能になるが、帰結空間 (**outcome space**、生徒のありうる答えがすべて○×) をうまく分類することも必要か?
5. 反応・応答のモデル化 : 競争状態 (入試、期末試験等順位付けの伴うもの) でのアセスメントと測定 (関連するすべての情報を使う) をどういうときにどちらを選べばよいかどうしたらわかるか? 両方とも (競争状態でアセスメントの結果も測定の結果も) 報告すべきか?
6. 妥当性の証拠 : どうやれば教室に負担をかけずに新しいテクノロジーや新しい評価を活用してより多くの情報を集められるか?  
21 世紀型スキルのために集合知と伝統的な知識観 (知識は個人が持っている) に基づいた情報の好ましい割合は?

行政が行う一斉テスト（日本の学力テストなど）の結果情報を教室活動にどう活用し、教室内のテスト（日々の活動で行われているテスト）のデータを国家の説明責任にどう活用できるか？（相互乗り入れ）

児童・生徒の自らの学びを始動させる 21 世紀型の評価をどう作るか？

以上の点を含んだ今回洗い出された課題を考えなら、いくつかの典型的な 21 世紀型スキルの評価の構築という次の段階に入る。



## 別添

### 証拠中心の設計

評価の設計をすることと、実施された評価をもとに結果を出すことは異なることである。目的を合わせて評価を設計しなければならない。第1段階として評価の目的を展開する、第2段階として目標とするものをサポートするのにどのような証拠が必要か定義する、第3段階としてそれを引き出す方法を列挙することで、タスクモデルは最後に作られる。また設計は基準を元にしていなければならない。達成レベル、証拠モデルとタスクモデルが確定された後、タスク仕様と成績レベル仕様が設計されテスト仕様設計へと続く。実施は予算や時間といった実務的な面に左右される。

### BEAR 評価方法

発達の視点（時間経過と適応性のある評価）・指導と評価の結合（指導に評価を組み込む）・指導者のフィードバック・フィードフォワード・フォローアップによる管理（採点基準を関係者と共有）・高品質の証拠の創出（Wright maps）という考えに基づく。

### 実証フレームワークの使用

証拠を使って科学的理由付を行うことをモデルとしたフレームワーク。指導の最終段階で児童・生徒がどこまで達成しなければならないかの記述を可能とし、発達段階のさまざまな次元での理解を可能とする。生徒の複雑な回答を評価するモデルともなる。

### マイクロソフト・ラーニング

学習の進行具合が達成目標・評価基準と結びついていると、カリキュラム構造と評価をつなぎ合わせるものとなる。

### シスコ・ネットワーキング・アカデミーで使用されているパケット追跡

評価がカリキュラムにうまく組み入れられていて、評価ツールとして特別なものを用意する必要がない。

### チリの国家教師評価

チリの公立学校に勤務する教員が必ず参加しなくてはならない評価システム。プロフィール結果報告の例で、評価段階の結果に基づき形成的な情報を提供する。不十分な教員には教員研修が提供され、評価が高い教員は昇給の対象となる。地域には統合された結果が報告され、必要ならば教員研修の策定などアクションをとる元となる。