

## 新時代における学習評価

(研究会実施日:2010年6月18日)

講師:植野真臣 [電気通信大学 准教授]

### 研究領域

教育工学をベースに eラーニング, eポートフォリオのシステム開発, 実践を通してマネジメント方法の研究に従事。また, テスト理論の研究者でもあり, e テスティングを IEEE で提唱し, 我が国での普及にも貢献。

### 経歴

#### 所属学会

東京工業大学大学院システム科学専攻助手, 千葉大学文学部助手, 長岡技術科学大学助教授を経て,

現在, 電気通信大学大学院情報システム学研究科准教授

日本教育工学学会 (理事, ショート論文編集委員長), 日本行動計量学会 (理事, 副編集長), 日本テスト学会 (監事), 教育システム情報学会 (評議員), IEEE Computer Science(Organizer Committee member)

### 著書

- ・植野真臣・荘島宏二郎 (著) (2010). 学習評価の新潮流 朝倉書店
- ・植野真臣・永岡慶三 (編著) (2009). e テスティング 培風館
- ・植野真臣 (監訳) (2009). 大学生のための学習マニュアル 培風館
- ・植野真臣 (著) (2007). 知識社会における eラーニング 培風館
- ・繁榊算男・植野真臣・本村陽一 (著) (2006) ベイジアンネットワーク概要 培風館

## 要約

近年, 学習理論の主流が構成主義に変化してきたことにより, 評価理論はテストのみによる評価から, より自然で真正な文脈における評価へ移行しつつある。一般の社会では人の能力をペーパーテストのみで評価することは少なく, むしろパフォーマンス (成果) で評価することのほうが多く, 一人の審査者が評価するよりも共同体のメンバーたちによって評価されることも多い。最近の学習評価では, このような自然な学習評価の環境を重視し, 学校化されたテストの仕組みから抜け出し, より真正で実践的な評価が提案されつつある。

そこで, 本報告では, 1. 学習理論の変遷, 2. 新しい学習観・知識観に基づいて提案された評価理論, 3. eポートフォリオや e テスティングなどの最新技術, について紹介する。

## I. はじめに

能力評価は、古くは旧約聖書の中でギレアデびとが「シボレテ」という言葉をうまく発音できるかどうかによって、エフライムの落人をテストしたという記録があるし(士師記12章5・6節)、紀元前413年、不幸にして捕えられた7,000人のアテナイ軍のうち、ユリピデスの詩を暗唱することができるものは釈放されたという話がある(池田,1978)。

東洋で古くからの登用制度の中に、試験を定着させたのは中国であった。6世紀隋の時代からすでに人材発掘の方法として、「科挙」の制度を編み出し、いくたびかの変遷を重ねつつ、20世紀のはじめまで存続させてきたのである。それらが多くの問題点をはらみつつも、実に1300年もの長い間続いたということは、この制度が一方で他に替え難い優れたメリットを持っていたからであろう。

ヨーロッパでは長い間、口述試験が行われてきた。1200年頃のイタリアのボローニャ大学では博士号をとるのに、一週間もの口述試験があったそうである。しかし、居酒屋が試験会場として用いられ、そこでは食事や寝泊りができ、試験が続く間、候補者は試験官である教師やそこにきた学生たちに食事やお酒をごちそうしたそうである。そのため、ほとんどの候補者が合格し、学位も名誉的称号であったそうである(池田,1978)。

1540年に設立されたイエズス会は教育というものの価値を重視して、こうした当時の一般的慣習を破って筆記試験の使用を強調し、1599年には筆記試験実施上の細かい規則を発表している。しかし、実際にヨーロッパで筆記試験が行われたのはそのずっとあとになって1702年にケンブリッジ大学で行われたのが最初であるといわれている(橋本,1976)。筆記試験では全受検者に同一の問題を出し、ほかの答案と比較できたり、採点を見直したりできるために、公平性や客観性が改善された。

我が国の近代学校教育制度は、1872年の学制発布から始まるが、その条文で試験を実施することが明文化され、学業成績は100点満点あるいは甲乙丙丁でつけられた。このとき

の基準は、教師の主観的な絶対評価であったと考えられる。

教育評価が大きく変わり始めたのは1920年代頃の心理学者のThorndikeらによる教育測定運動によるところが大きい。教育測定運動では、「科学性」を重視し、客観テスト・評価の目標・基準の分析・明確化・古典的テスト理論などが導入された。これによって、絶対評価から相対評価への転換が起こったといえる。

その後、科学的なテストを研究する分野としてテスト理論の分野が確立されていった。テスト理論では、テストの良さの指標として、妥当性と信頼性が提案されている。簡単にいうと、妥当性とはテストが測定しようとしている能力を的確に捉えているのかを示す指標であり、信頼性とはテスト項目の持つ統計的一貫性を示す指標である。

その後、認知主義やブルーム(1973)による教育目標分類学の研究に触発されて、現場の中から教育目標の明確化やその目標に基づいた診断的評価が実施されるようになってきた。

このようにして診断的評価への機運が高まり、1980年の文部省小中学校指導要録の改訂では「観点別学習状況」という形で導入され、1991年の改定ではそちらに重点を移し、2001年の改訂ではついに相対評価が廃止された。

以上のように、教育評価の変化の特徴として、能力を一次元尺度により測定する評価から、学習者が何がわかって何がわかっていないのか、どこで行き詰っているか、などを診断できる診断的評価や教育活動自身の改善に用いられる形成的評価への変化、が挙げられる。

## II. 真正な評価へ

従来の評価法は、まだ、「知識は分離可能で普遍的なもの」という前提の下、知識注入主義を基本にしている。前述のように知識や能力は、社会、文化、状況に依存するものである。これらを最も強く主張しているのが、状況論という立場である。状況論では、「能力」とは、人間に元来内在する知識などの資源から表出

されたものではなく、人間の特定の行為を「有能だ」とか「頭が良い」とか呼んだりする承認などの社会的過程を経て構築されてきたものである、と考える(上野,1992)。知識が社会・文化・言語に依存するのは当然のこととして、さらには「能力」が発揮されているその場・そのときの状況にも依存することになる。別の言葉では、人はさまざまな表出を行っているが、それがどのように認識され、どのように評価されているかということは初めから決まっているわけではない。特に何が知識であり、どう行動することが有能とみなされるかということに関して言えば、文化によっても異なるし、状況によっても異なる。したがって、「能力」はそれが発揮される状況と切り離して議論することはできないと考えるのである(加藤,2008)。また、Garfinkelは、人間は日常行動の中で、互いの行動を定式化し、説明し、報告しあうことを通して、互いの行動を意味づけ、理解可能なものにし、それによって、私たちの実践は秩序を達成しているのである、と考えた。これが正しいとすると、出来事や行為に対する評価は、人間が本来行っている行動の定式化の一つのタイプであると考えられる(加藤,2008)。

この立場から評価を捉えなおすと、教室で行われている学習活動がどのようなものであるかということである。それが日常生活の文化的実践とは異質のものであれば、学習の場で可視化されている能力と現実社会における実践能力とは当然異なるものになる。従って、学校での学習活動をより、実社会で意味を持つ「真正の(authentic)」活動へと変えていく必要がある、学校が実社会とリンクせねばならない理論的根拠ともなっているのである。真正な学習の実現のための条件としては、1. 真正な文脈(Authentic Context), 2. 真正な活動(Authentic Activity), 3. 真正な課題(Authentic Tasks), 4. 真正な評価(Authentic Assessment), を満たさなければならない(植野,2007)。

これまでにも現在の学校教育に上のような問題があることは多くの研究者によって指摘されてきた。

例えば、デューイ(1957)は、「第一に学

校は暗記と試験による受動的な学習の場ではなく、その中で子供たちが興味にあふれて活動的な社会生活をいとなむ小社会でなければならない。第二に、この小社会は、単にそこで子供たちの自発的な活動が行われる小社会であるばかりでなく、現代の社会生活の歴史的進歩を代表する社会でなければならない、そのために学校と社会との間に活発な相互作用がなければならない。」と指摘している。

また、Traweek(1988)は、以下のような指摘を行っている。

「高校の物理の生徒が参加し、自らの参加を意味づけているやり方と、専門の物理学者がやっているやり方とはとてつもなく大きな違いがある。すなわち、実際に学習者が学校で習っている物理学は、物理学者が認識しているものではなく、「学校化された人々」のものである。少なくとも大学院に入るまでは、物理学の実践は学べず。絶えず学校化された物理学を強要されるのである。」

さらに、Lave, J. & Wenger, E. (1991)は、上文を引用し、学校という共同体と物理専門家の共同体がまったく異なるものであること、そして現在の学校教育の最も重要な問題は、単純な教授技術上の問題ではなく、このような構造上の問題であると主張している。

Brown, Collins & Duguid (1989)も、状況論的学習の立場から、科学・歴史・文学などの専門家の行っている活動と学校で学ぶ活動がまったく違うことを主張している。

村山(1995)は、学校が学ぶ場として不適切な理由として、そこに実践共同体が存在しないこと、学級は一人の専門家(教師)と多数の同レベルの初心者(児童・生徒)だけで構成されているために実践共同体を構成しにくい性質を持っていること、の二点を指摘している。

近年、注目されているeラーニングは、企業内での知識共有や知識創出などの経営ツールとして脚光を浴びているが、学校教育でも実社会とのリンクを容易にすることができるツールとして注目されつつある(植野, 2007)。

たとえば、高校生の学習と大学での実践的な研究活動をeラーニングを用いて接続した

りする試みなどが実践されているし、真正な評価の一例として、十分にすぐれた成果をあげた高校生が一流学術論文誌に論文が掲載されたことなどが挙げられる(植野,2007)。しかし、このような現実社会との接続は、初等教育では難しい。

学校に閉じた「真正な評価」は、提唱者のWiggins(1998)によると「大人が仕事場・社会生活の場・個人生活の場で『テストされている』その文脈を再現・模擬すること」と定義され、次のような6つの条件を満たさなければならない。

- 課題が現実的で、知識や能力が実社会の状況でテストされるやり方を再現していること
- 課題が体系化されてなく、解決するのに判断と工夫を必要とすること
- 課題を探索的に実行させること
- 大人が実社会でテストされる文脈を再現すること
- 知識や技能を総合的に用いる課題を通して能力を評価すること
- 練習をしたり、下調べをする機会を与え、フィードバックを与え、成果をさらに洗練させること

真正な評価では、このような真正の課題に学習者が取り組み、そこで生み出した成果物や実技などを評価する。すなわち、筆記試験でなく学習活動における学習者のパフォーマンスで評価する。また、必ずしも評価のための特別な時間と評価手段を取らないため、原則として学習場面と評価場面とが一体となっている。

### Ⅲ. 学習理論と評価理論の対応

学習を評価するのは、社会的な要請や学習指導からの観点だけではない。個人の学習を成立させる学習論の視点も必要になる。Table 1に示すように、主に4つの考え方がいままで提示されている。

第1は行動主義に基づくもので、スキナーのプログラム学習が代表とされる(スキナー,1969)。プログラム学習は次の5つの原則を持っている。1つめはスモール・ステップで

あり、学習は小さなステップ毎に分割された簡単な問題形式になっており、2つめはそのステップ毎に学習の反応が求められる。3つめはそれに対して即時フィードバックし、4つめは学習そのものは自分でペース管理できる。そして、最後に学習者の反応結果に基づいて、プログラムの改善が図られるものである。CAIにおいては非常に多用された考え方である。

第2は構成主義であり、これはピアジェに代表される(ピアジェ,1958)。ここでは、学習者たち一人一人が教えられている対象の概念を組み立てていくように教えるという考えである。そこでは学習者自身が能動的に知識を構築していくという考え方があり、その結果学習プロセスの中で質的な変化が学習者自身に起こると考えている。

第3は情報処理論的アプローチと呼ばれるもので(ガニエ,2007)、人をコンピュータのような情報処理モデルと見なし、そのアナロジーを使って効率よく学習を促進させるということが考えられた。そのために学習者をモニタリングし、認知負荷を下げた教材を開発するなどが試みられた。

第4は社会構成主義である。これはヴィゴツキーの「学習は学習者が単独で学習するのではなく、教師や仲間の援助を受ける形で、学習が日常的概念と科学的概念の接続を果たし行われる」という考えに基づいており(ヴィゴツキー,1962)、他者の役割を大きく取り入れている。援助を受け入れることにより学べる領域を最近接領域としている。協調学習の理論的背景になっている。

この他にもレイヴとウェンガーによる状況論(Lave & Wenger,1991)の立場の研究がある。この立場は社会構成主義と同じように学習は個人で行われているものではないとしているが、より学ぶだけでなく生産もしている実践共同体への十全的な正統的周辺参加であるとして、学びそのものを集団への参加プロセスという位置づけにし直した。これを学習実践するものとしてコリンズの認知的徒弟制(Collins,1991)がある。

Table 1 学習理論・評価理論とそれらの支援環境の変遷

	行動主義	情報处理的アプローチ	構成主義	社会的構成主義
主な理論家	スキナー	ガニエ	ピアジェ	ヴィゴツキー
学習	特徴		学校化された学習	
	知識観		真真正正な学習	
	知識は普遍的に真なもの		知識は一人一人が自ら構成するもの	知識は社会的な営みの中で構成するもの
	学習観		学習者の事前知識から事後知識への質的な変化	学習者の事前知識から事後知識への質的な変化(共同体の社会的な営みを通じた内化)
	知識伝達			
	主体		教師中心	
	学習者の態度		学習者中心	
受動的		能動的・自律的		
学習課題		学校化された課題		
真真正正な課題				
学習支援システム		CAI ティーチング・マシン	知的CAI 知的チュータリング・システム エキスパート・システム	LOGO マインドストーム
CSCL eラーニング				
評価	特徴		学校化された評価	
	評価期間		真真正正な評価	
	ある時点		継続的	
	評価形態		学習者のパフォーマンス(学習成果物など)の主観的な評価	
	評価される対象		学習活動のプロセスを通じた学習成果物や記録を重視	
	テストの点数を重視			
	評価の在り方		学習と切り離された評価	
学習に埋め込まれた評価				
評価方法		ポートフォリオ 学習者自身が作成したテスト ダイナミック評価		
テスト				
能力測定		学習プロセス同定と診断的評価	セルフ・アセスメント	ピア・アセスメント(専門家による)他者評価

#### IV. 学習評価の方法

##### (1) テスト理論

従来からテストの理論的根拠として一般にテスト理論(Test theory)が用いられる。テスト理論は、前述のとおり、テストの「妥当性」(validity)と「信頼性」(reliability)を保証するための理論でほとんどのハイ・ステークス・テスト(入学試験や資格試験、昇格試験などテスト結果が受検者にとって重要な影響を与えるテスト)で使われている。テストの妥当性には、①項目内容に誤りがないこと、②項目がテスト全体の測定対象をよく反映していることの二つが重要であり、これらは質的にテスト開発者が吟味しなければならない。信頼性の指標としては、テストを3つ以上の部分テストに分けてそれらの相関係数を求めることによって得られる「アルファ係数」などが一般的であるが、後述の項目反応理論のパラメータなどが最近では信頼性指標として

用いられることが多くなってきている。テスト理論は、a) テストの信頼性と妥当性の概念を規定し、記述統計学の枠組みで信頼性指標などを規定している「古典的テスト理論」(池田,1994) (Classical Test Theory) と b) 項目に対する受検者の正答確率を項目特性パラメータと受検者の能力パラメータによりモデル化した項目反応理論(芝,1991) (Item Response Theory) に大別される。項目反応理論の利点は、1) 異なるテストを受けた受検者を同一尺度上で評価できる、2) テストに依存しない項目特性パラメータ(識別力と難易度)を得ることができ、テスト情報量最大化原理に基づくテスト構成が可能となる、3) 受検者の項目への反応に適応して情報量最大化となる項目を選択・出題する適応型テストを実現できる、などが挙げられ、最近ではeテストの基本原理として用いられる。

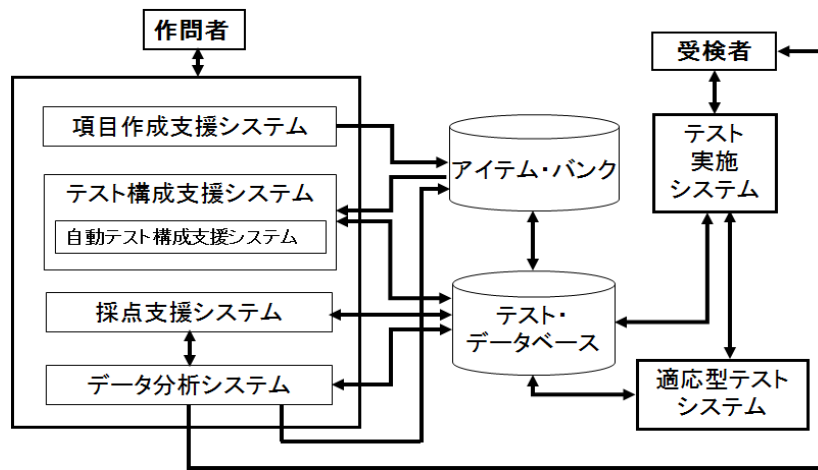


Figure 1. e テスティングの構成

## (2)e テスティング

テストングとは、項目データベースであるアイテム・バンクから複数のテスト開発者がネットワークを介してブラウザ上でテスト構成するのを支援し、ネットワーク上で多くの受検者に対してテスト配信を行い、それらを自動採点し、受検者や管理者にフィードバックを送信するとともにデータ解析した結果よりアイテム・バンクを更新することをシステム上で一貫して行えるシステムをいう(植野・永岡,2009)。

e テスティングの利点として、以下が挙げられる。

1. ペーパー・テストに比較し、テストの印刷、運搬のコストが減少する。
2. 大量のペーパー・テストの束を管理することに比較し、電子ファイル上のテストは管理が容易でセキュリティも高い。
3. テスト実施に際して、テストの配布、回答の回収が自動的にできる。
4. テスト採点が、自動化もしくは半自動化される。小論文方式のテストでさえ、自動採点できる。さらに即時的にテスト結果をフィードバックできる。
5. アイテム・バンクの領域からランダムに項目の抽出が行え、テスト出題バイアスを減少できる。
6. コンピュータによるテスト構成支援機能は、信頼性と妥当性の高いテスト構成を可能にする。
7. 適応型テストなど受検者の能力を随時推定しながらその能力に最適な項目を選択出題

するような動的テストを実現し、テスト時間を短縮し、測定精度も向上させることができる。

8. 様々なテスト目的に応じたテスト出題方略を持つことにより、目的に応じたテスト構成や適応型テストを実現できる。

9. マルチメディア技術を用いることにより、これまで測定することのできなかつた能力を測定できるテスト項目を構築、実施できる。

10. 回答所要時間や回答変容回数、回答見直し回数などペーパー・テストでは測定することができなかつたデータが測定することができるようになり、テスト項目評価の多様性をもたらし、補助的に用いることにより測定精度を向上させることができる。

11. 障害を持つ受検者への対策が容易に行える。例えば、字を大きくしたり、ヘッドホンをつけさせて音を大きくしたり、特別な入力装置(手以外で入力できる入力装置)、出力装置(点字など)を用いることにより障害者の受検を可能にすることができる。

12. 遠隔地にいる受検者の評価を行うことができる。

13. 遠隔地に分散するテスト作成者が、共同でアイテム・バンクを構築し、テスト構成を行うことができる。

e テスティングの一般的な仕組みは Figure 1 に示すとおりである。テスト構成はネットワーク上で一人もしくは複数のテスト開発者がテスト構成支援システムを用いて行う。特に、e テスティングではテスト構成が



Figure 2 テスト得点・所要時間分布予測



Figure 3 自動テスト構成システム

ネットワーク上でブラウザを用いてどこからでも行えることが特徴である。構成されたテストは、テストデータベースにテストID(識別子)を伴い格納され、いつでもWEB上でテスト実施ができる状態となる。テスト構成中のテストの得点分布や所要時間分布を予測したり (Figure 2), 現在のテスト構成状況に適した質問項目を推薦する機能を持つことができ、質問項目 Web 上で複数のテスト構成者が協働してテスト構成を行える環境を提供するシステムも開発されている。最新研究では、同一のテスト得点分布を持つ異なる複数の等質なテストを、数理計画法を用いて自動構成するシステムも実用化段階に入りつつある (Figure 3) (植野・永岡,2009)。e テスティングでは、真正な評価環境が実現できることも特徴である。ビデオや音声を用いた質問項目を配信することができ、より現実的な場面での能力評価が実現できる。例えば、語学テストでは動詞の概念などはビデオを用いて表

現すると非常に明確になるし、科学実験や音楽などもビデオを用いることにより多様な評価が行えるようになってきている。テスト実施システムは、学習者の項目への回答を逐次、テストデータベースに保存するだけでなく、所要時間や見直し回数、回答の書き直しプロセスなど詳細な回答履歴データを蓄積していく。このことも、これまで正誤データのみを情報としてきたペーパー・テストとは異なり、これまで得られなかったような回答履歴データより多様な評価を実現できるのである。

受検者の反応データは、データ解析システムに送信され、まず、自動採点される。当初、コンピュータの得意な自動採点は、多枝選択式項目、真偽式項目のみで行われてきた。しかし、近年、自然言語処理の技術が発達し、単答式項目や論述式項目の自動採点が導入されつつある(石岡,2004)。

各項目への正誤データが生成され、項目所要時間、回答書き直し回数、項目見直し回数などとともにテストデータベースに書き込まれ、データ解析システムへ送られる。

e テスティングの技術で現在最も注目されているものが適応型テスト (computer adaptive testing, CAT) (van der LINDEN, 2000)である。適応型テストとは、テストへのそれまでの回答履歴を用いて逐次、受検者の能力を推定しながら、その能力に最も適応した項目をアイテム・バンクより抽出しながら出題するというものであり、既に多くのテスト業界で実用化されている。適応型テストの利点は、以下が挙げられる。(1)受検者の能力に最適な項目を出題することにより、能力の推定効率を上げる。(2)難しすぎたり、易しすぎる項目が出題されないために、受検者のフラストレーションやテスト不安を減少させる。(3)推定効率が向上するために、ペーパー・テストと同じ精度を保って出題項目数を減少させることができ、受検者の負担を減少させることができる。当初、項目情報量として、フィッシャー情報量が用いられてきたが、近年では、計算機科学の立場から、能力推定値の推定誤差を考慮した完全項目情報量や数理計画法を用いてすべての領域から出題するなどの制約付き出題機能 (van der



LINDEN,2000)などが提案されるとともにそれらの高速計算に関する研究(植野,2009)が盛んに行われている。

また、構成主義における評価観より、学習者自身に項目を作問させて実際にテストや演習問題に用いる方法が注目されている。「問題作り」が問題解決能力や問題分類能力を向上させ、学習者のメタ知識を発達させる働きがあることが示されている(植野,2007)。

真正なテストングを実現するために携帯電話を用いたeテストングも注目されつつある(Figure 4)。携帯電話により、問いたい対象までナビゲーションされ、質問を適応的に生成するしくみなどにより、現実的な場面での能力を測定する仕組みが開発されつつある。



Figure 4 携帯電話によるeテストング

### (3) eポートフォリオ

近年、学習理論の主流が構成主義に変化してきたことにより、評価理論はテストのみによる評価から、より自然で真正な文脈における評価へ移行しつつある。一般の社会では人の能力をテストのみで評価することは少なく、むしろパフォーマンス(成果)で評価することのほうが多く、さらに一人の審査者が評価するよりも共同体のメンバーたちによって評価されることが多い。ここで、パフォーマンスとは、先のテスト結果のみでなく、レポートや製作物などの学習成果などを含む。このような評価を実現するための評価法の一つとして注目されている手法が、ポートフォリオ

評価である。ポートフォリオとは、自身の学習、スキル、業績を実証するためのあらゆる成果(work)を、ある目的のもと、組織化/構造化しまとめた収集物であり、ポートフォリオ作成プロセスと継続的なリフレクション(内省: reflection)が重視される。さらに、電子的な形式で扱われたすべてのポートフォリオをeポートフォリオ(森本,2008)と呼ぶ。eポートフォリオは、1. Multimedia: ビデオや音声データなどのマルチメディア・データを扱うことができる、2. Collaboration: Web上で構成することにより遠隔地にいるメンバーや組織で共有することができる、3. Calculation: コンピュータの推論・計算機能を用いることにより、大量の時系列のポートフォリオ・データから重要知識をマイニングしたり、まとめたり、参考になる他者を推薦したり、と様々な機能を実現することができる、という利点を持ち、その有機的な融合によりその利点は無限に広がる可能性を持つ。

実際の学校教育でのポートフォリオ評価の方法は、1学期・1年という長期に渡って学習者の学習成果物を収集してポートフォリオを作ることによってなされる。学習成果物は、ワークシート、レポート、絵、資料、下書き、プログラム、テストの答案など学習の過程で生み出されるものすべてであるが、実際に何を集めるかは目的に応じて選択する。そして、ときどき、学習者にポートフォリオを系統的に並べ替えさせたり、取捨選択させたり、セルフ・アセスメント(自己評価)させ、自分の学習の過程を振り返る(リフレクションする)機会を設ける。さらに、学習の節目にポートフォリオ検討会を持ち、自分が作ったポートフォリオについて教師や他の学習者と話し合う。検討会では、「どのような作品がよいものなのか」「どのような学習が求められているのか」「どのように自己評価すべきなのか」といった点について教師と学習者とで意見が交わされる。

また、ピア・アセスメントと呼ばれる相互評価により、お互いのポートフォリオを評価しあうことにより、真正な評価が行われる場合もある。ポートフォリオは、eラーニングなどを通じて異なる学校の学習者と共有した



り、ピア・アセスメントが行われる場合もある。

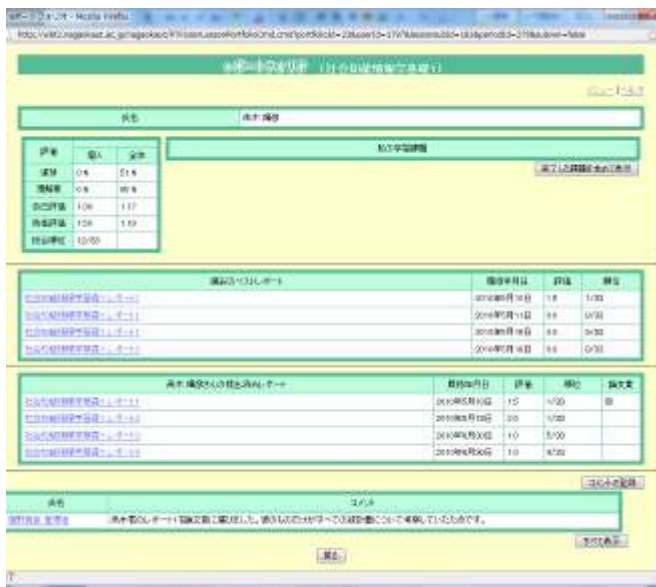


Figure 5 eポートフォリオ Samurai-folio の画面

従来の評価では、教師が学習者を一方的に評価するという形式であったが、ポートフォリオ評価では、評価の当事者が学習者を中心とする学習コミュニティに移行している点の特筆すべきである。ピア・アセスメントは一見、主観的に見えるのであるが、教師一人が評価するよりもより信頼性の高い評価が行えることが示されている。Falchikov (1986)は、

小学生では信頼性の高いピア・アセスメントは実現できていなかったが、中学生ではより高い信頼性の高い評価がされていたことを報告している。また、Amold(1981)は、医学部での講義でピア・アセスメントを導入し、公平で一貫した評価が行われていることを実証している。Orpen(2004)は、大学における学生の学習成果について、教師の評価と学生のピア・アセスメントとを比較し、平均では有意差が見つからなかったこと、信頼性の意味では、教師一人の評価よりもむしろピア・アセスメントのほうが高いことを示している。植野(2008)は、評価者をテスト理論における項目とみなして項目反応理論にあてはめることにより、異なる評価基準を持つ評価者の評価結果を統一的に扱える手法を提案している。これにより、少ない評価者からより信頼性の高い評価を得ることができていることを示している。

Figure 5は著者らが開発しているeポートフォリオシステム「Samurai-folio」の例である(植野,2010)。このシステムでは、自動的にeラーニングでの学習履歴を整理し、学習成果物とともに自動的にポートフォリオを生成し、相互評価、自己評価を管理できるシステムである。

## V. おわりに

本報告では、学習評価理論の推移について解説した。特に理想的な評価理論の実現のために情報通信技術の進展と利用の貢献は大きい。eテスト、eポートフォリオの研究は始まったばかりであり、その可能性も無限にあるといえ、研究マーケットも急速に広がっている。より優秀な研究者が参入し、より発達していく分野であると期待される。

<引用文献>

- Arnold, L. (1981). Use of peer evaluation in the assessment of medical students, *Journal of Medical Education*, 56, pp.35-42
- ブルーム,B.S.(著) 梶田 叡一・渋谷 憲一・藤田 恵壘 (訳) (1973). "教育評価法ハンドブック—教科学習の形成的評価と総合的評価" 第一法規出版
- Brown, J. S., Collins, A. & Duguid, P. (1989) "Situated cognition and the culture of learning", *Educational Researcher*, 18(1), pp.32-42
- Collins, A. (1991). Cognitive apprenticeship and instructional technology, In: Idol L & Jones, B.F. (Eds) *Educational values and cognitive instruction: Implications for reform*, Hillsdale, N.J: L. Erlbaum Associates
- デュローイ,J.(著) 宮原誠一 (訳) (1957). 学校と社会 岩波文庫
- Falchikov, N. (1986). Product comparisons and process benefits of peer group and self assessments, *Assessment and Evaluation in Higher Education*, 11, pp.146-166
- ガニエ,R.M.・ケラー,J.M.・ウェイジャー,W.W.(著) 鈴木 克明・岩崎 信 (監訳)(2007). *インストラクショナルデザインの原理* 北大路書房
- 橋本重治 (1976). *新・教育評価法総説* 金子書房
- 池田 央 (1978). *テストで能力がわかるか* 日本経済新聞社
- 池田 央 (1994). *現代テスト理論* 朝倉書店
- 石岡 恒憲 (2004). 記述式テストにおける自動採点システムの最新動向 *行動計量学* 31 (2), pp.67-87
- 加藤浩 (2008). もうひとつの教育評価 -状況内評価の活用に向けて- *人工知能学会誌*, 23 巻, 2 号, pp.163-173
- Lave, J. & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*, Cambridge: Cambridge University Press.
- 森本康彦 (2008). eポートフォリオの理論と実際 *教育システム情報学会誌* Vol.25, No.2, pp.245-263
- 村山功 (1995). 科学はいかにして学ばれるか 佐伯肝・藤田英典・佐藤学 (編) シリーズ学びと文化3 科学する文化1章 東京大学出版会, pp.1-33
- Orpen, C. (2004). Student versus lecturer assessment of learning, *Higher Education*, 11, pp.567-572
- ピアジェ,J.(著) 波多野完治・滝沢武久 (訳) (1958). *知能の心理学* みすず書房
- 芝 祐順 (編著) (1991). *項目反応理論---基礎と応用---* 東京大学出版会
- スキナー,B.F.(著) 村井 実・沼野 一男 (訳) (1969). *教授工学* 東洋館出版社
- Traweek, S. (1988). *Discovering machines: Nature in the age of its mechanical reproduction, in F. DUBINSKAS (ed.) Making time: Ethnographies of high technology organizations. Philadelphia: Temple University Press.*
- 植野真臣 (2007). 知識社会におけるeラーニング 培風館
- 植野真臣・ソンムアン,ポクボン・岡本敏雄・永岡慶三(2008). ピアアセスメントにおける評価者特性を考慮した項目反応理論 *電子情報通信学会論文誌* Vol. J91-D, No. 2, pp.377-388
- 植野真臣・永岡慶三 (2009). eテストング 培風館
- 植野真臣 (2010). 多機能型eポートフォリオシステム"Samurai-folio"の開発 *日本教育工学会研究報告集* JSET10-3, pp.33-40
- 上野直樹 (1992). 言語ゲームとしての学校文化, 佐伯胖, 汐見稔, 佐藤学 (編) *学校の再生をめざして 1.学校を問う*, 51-81, 東京大学出版
- van der LINDEN, W. J. and Glas, C. A. W. (2000). *Computer Adaptive Testing: Theory and practice*. Dordrecht, The Netherlands: Kluwer Academic Press
- ヴィゴツキー,L.S(著) 柴田義松(訳) (1962). *思考と言語* 上下巻 明治図書
- Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*, Jossey-Bass A Wiley Imprint, San Francisco