大学入試センターシンポジウム 2016

高大接続における 学力評価の最新動向

平成 29 年 (2017年) 3 月

独立行政法人大学入試センター 入学者選抜研究に関する調査室

MCE EXAM



入学者選抜研究に関する調査室報告書3

大学入試センターシンポジウム 2016 高大接続における学力評価の最新動向

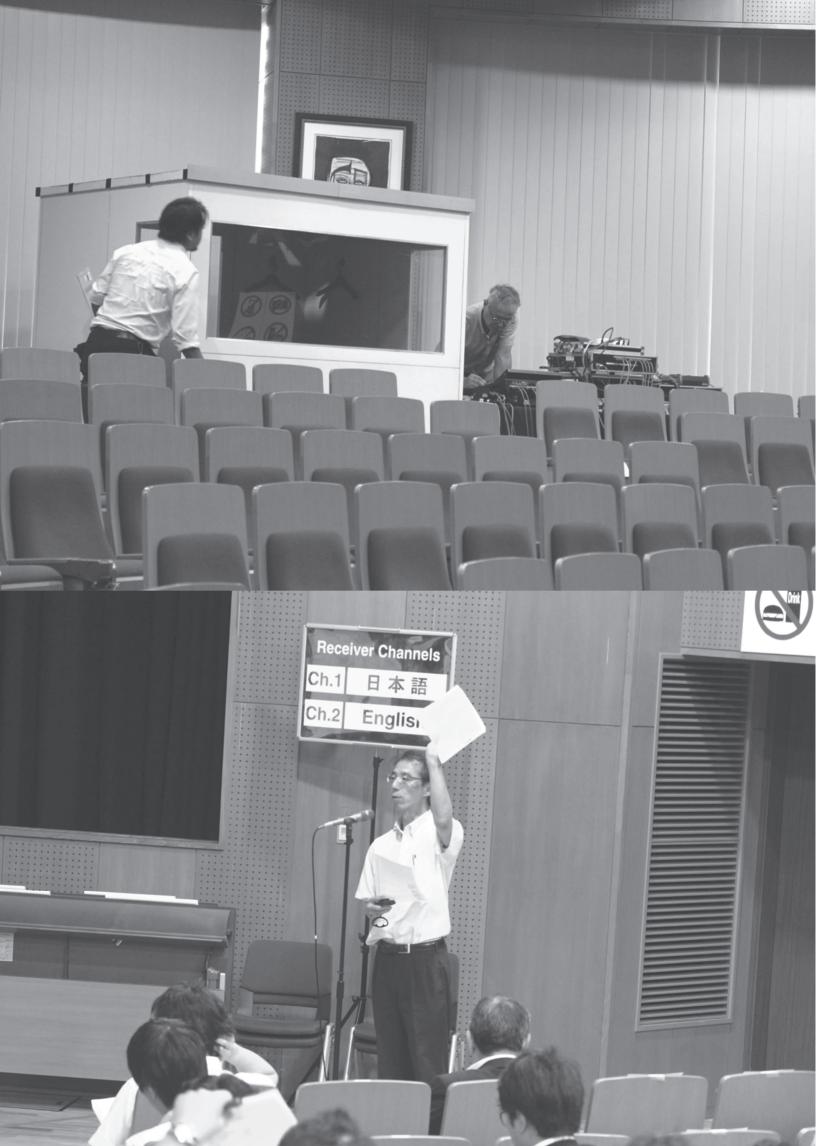
独立行政法人大学入試センターは、平成 28 年 9 月 10 日(土)、東京工業大学大岡山キャンパス(東京都目黒区)において「大学入試センターシンポジウム 2016」を開催した。本報告書は、同シンポジウムにおける講演、報告及び討論等を収録したものである。

上

- 開会挨拶 山本廣基(大学入試センター理事長)-5
- 趣旨説明 大塚雄作(大学入試センター試験・研究統括官)-7
- 基調講演 認知科学にもとづいたアセスメント 思考力・判断力・表現力の評価・改善をめざすアセスメントシステムPaul Deane(米国 Educational Testing Service 首席研究員) 10
- 報告 1 パフォーマンス評価をどう活かすか-京都大学教育学部における 特色入試の取り組み 西岡加名恵(京都大学大学院教育学研究科准教授)-104
- 報告 2 学習科学からみた資質・能力の育成と評価 白水 始(東京大学大学総合教育研究センター教授) – 128
- 報告 3 コンピュータで実施するテストの採点をめぐる課題 石岡恒憲(大学入試センター研究開発部教授) - 147
- 指定討論 清水和弘(福岡大学附属大濠中学校·高等学校副校長)-159
- 全体討論 Paul Deane/西岡加名恵/白水 始/石岡恒憲/清水和弘-163
- 閉会挨拶 大津起夫(大学入試センター研究開発部長) 183

〔総合司会〕山地弘起(大学入試センター試験・研究副統括官)





〔開会挨拶〕

山本 廣基(大学入試センター理事長)



本日は、全国より多くの方々が大学入試センターシンポジウムにご参加くださいました ことを、心より御礼申し上げます。

当センター主催のシンポジウムは2年ぶりの開催となります。これまで大学入学者選抜に係るその時々の時宜にかなったテーマで開催してまいりましたが、今回は「高大接続における学力評価の最新動向」というテーマを設定いたしました。高大接続改革につきましては、皆様方もご承知のとおり、教育再生実行会議の提言や中央教育審議会の答申を経て、今年3月末に、文部科学省の「高大接続システム改革会議」において「最終報告」が取りまとめられたところです。

この最終報告を受け、現在、この改革の実現に向けた検討・準備グループが文部科学省に設置され、改革の推進に向けた具体的な検討がなされているところです。現行の大学入試センター試験に代わって平成32年度からの導入が予定されている「大学入学希望者学力評価テスト(仮称)」についても、十分な知識・技能を基盤にしつつ、思考力・判断力・表現力をよりよく評価できるものとすること、高等学校教育の指導改善や入学後の大学教育にも好影響を与えることのできるものとすること、個別大学の入学者選抜において適切に組み合わせながら効果的に活用することができるものとすること、こういったことが「最終報告」でうたわれています。この新テストの問題作りに当たっては、思考力・判断力や、

思考の過程や結果を表現する力などを評価することができるよう,自ら文章を書いたり図 やグラフ等を描いたり式を立てたりすることを求める記述式問題を導入することや従前からのマーク式問題の一層の改善を図ることが提言されており、大規模共通の入学者選抜テストとしては、これまでに前例のない新たな取組が求められています。

私ども大学入試センターにおきましても、「新テスト実施企画本部」を立ち上げ、記述式問題の採点方法・体制・結果表示の在り方など、またマーク式問題の改善、英語 4 技能の評価、問題作成体制の見直しなど、種々の専門的・実務的な検討を開始したところです。今後、さらに専門的知見を得ながら、文部科学省と一体となって、様々な課題を乗り越え、新テストの円滑な実施のための準備を整えてまいりたいと考えています。

言うまでもなく、入学者選抜は、大学教育への適応力であるとか、それに必要な学力などについて測定し、その測定結果に基づいて選抜が行われるということになりますが、特に思考力・判断力・表現力や学ぶ意欲等を客観的に評価するための方法の開発も必要であると考えています。そこで、改めて、本日は、認知科学や教育評価を専門としておられる先生方にお集まりいただき、先ほど申し上げた視点からの学力評価の最新動向について情報提供していただくことにしました。

基調講演をお願いしています Paul Deane 先生は、アメリカのプリンストンにあります ETS (Educational Testing Service) よりはるばるお越しいただきました。先生には、モデルに基づいたアメリカでの評価方法の事例を紹介していただきます。現在中心となって取り組んでおられる読解力・論述力を評価する問題作成と自動採点の高度化等について貴重なお話を伺う機会となるでしょう。

続いて,事例報告をお願いいたしました西岡加名恵先生,白水始先生,石岡恒憲先生は, それぞれ,教育方法学,認知科学,教育工学がご専門の先生方であり,最先端の視点から の課題をお伝えいただけるものと思います。

指定討論といたしましては、清水和弘先生に高等学校教育の現場から、本日の基調講演 や各報告へのご意見・ご感想を頂戴しつつ、全体討論への橋渡しをお願いしたいと思いま す。

午後いっぱいの長丁場のシンポジウムになりますが、実りあるシンポジウムになりますよう、皆さまのご協力をお願いして挨拶といたします。

〔趣旨説明〕

大塚 雄作(大学入試センター 試験・研究統括官)



皆さん、こんにちは。今日はお忙しい中、大学入試センターシンポジウムにお集まりいただきまして、本当にありがとうございます。理事長から登壇者の簡単なご紹介がありましたので、私からはこのシンポジウムの背景となるようなところを皆さんと共有したいと思います。

このシンポジウムは、入試センターの中に置かれている「入学者選抜研究に関する調査室」という組織で企画させていただきました。「調査室」の室長を、私、試験・研究統括官の大塚が担当しており、室長補佐を、今日、司会をしている山地副統括官、それから研究開発部長の大津副統括官に担当していただいております。

入試センターには、入試研究に関しては研究開発部がありまして、入学者選抜に関わる 恒常的な研究課題に取り組んでおりますが、入試は今まさに、高大接続改革のさなかにあ り、タイムリーにいろいろな調査研究課題にも応じていく必要が出てきます。そのような アドホックな調査研究課題に対応するための体制として、「調査室」が位置付けられている ということです。

ここのところ、センター主催のシンポジウムは、調査室が企画してきておりますが、昨年度は開催いたしませんでした。高大接続改革答申が出されましたのが一昨年暮れで、その後、高大接続システム改革会議が設置されて、入試改革の具体的な方策について議論されておりまして、昨年は、入試改革がどういう方向で進められていくのかがまだはっきりとしていなかったということがありました。このようなセンター主催のシンポジウムなど

がありますと、往々にしてそこで議論されたことが実際の新テストにつながるのではといった誤解を招くことにもなりかねないということもありまして、そういった微妙な時期という判断に基づいて昨年度は自重したということです。

来年 4 月ぐらいには新テストの具体的な実施方針なども出されると聞いておりますが、 先日、文科省から公表されました高大接続改革の進捗状況などを見ましても、実は今年も まだ微妙な時期が続いている感があります。したがいまして、本来ならば、新テストに直 接結び付くようなテーマ設定が皆さんの関心にはよりマッチするということがあるのだろ うと思いますが、入試改革の議論をより建設的に進めていくためには、その基盤となる評 価あるいは測定に関わる基本的な事項を共有していくことが肝要であり、そういう場を設 定してみようという趣旨で、私どもは本日のシンポジウムのテーマを選んだ次第です。

そういうことで、いろいろな学習評価の手法や実践が紹介されると思いますが、くれぐれも、それがすぐ新テストに結び付くという誤解はなさらないでいただきたいと、まずはお断りしておかなければいけないと思います。もちろん、認知科学、学習科学、教育評価、パフォーマンス評価などに関わる一線級の研究者の先生方にご登壇いただき、その理論的な基礎の部分は何にしても違うわけではないということもありますので、そのベースのところで、センター試験や新テストの方に生かしていくことのできるアイデアやヒントは十分に得られると思いますし、企画の立場からしますと、そのような形で今後、このシンポジウムが生きていければ本望という部分もないわけではありません。

一方、現在の高大接続改革の流れは、高校教育、大学教育、入学者選抜の三つが一体となって改革していくというコンセプトに立っておりますので、教育のプロセスにおける評価のあり方を考えていくということも重要な課題となっております。その点で、今日のそれぞれの話題は、実際の教育現場でどういう評価が有効であるかという点でも、いろいろな示唆が得られると思います。

大学入試につきましては、共通試験のみならず、各大学の個別試験も合わせて一つのシステムをなすものでありまして、個別試験の改革も求められているところであります。今日は、特色入試を始められた京都大学の個別入試の経験も紹介されると思います。そういった独特の入試の方法は、どの程度のサイズで実行可能なのか、そして、具体的にどのような評価が行われているかといった点で、それぞれの大学でこれから個別試験を改革していこうというときのヒントにもなるかと思っております。

今日は、様々な立場におられる方々がお集まりかと思いますが、それぞれの立場で行われている測定や評価の基本となる部分を共有することを通して、それらの総合力として、より豊かな教育、より実りある入試改革と結び付く礎となればと思っているところです。

いずれにしましても、「高大接続における学力評価の最新動向」というテーマは、大きなテーマでありまして、ある意味で、今の入試改革の言葉を借りれば、答えが一つに定まらない、非常にチャレンジングなテーマだと思っております。ですから、往々にして、シンポジウム後のアンケートに、パネルディスカッションなどのまとめがどうもはっきりしないというような感想も聞かれることがあるのですが、それこそ今日のテーマも、拡散することはあれ、一つに収束するということは考えにくいテーマでもありますので、私からは、

ご参加の皆さんお一人お一人の立場で、この入試改革に当たって、あるいは、自らの教育 実践における評価に対して、こうしたらどうなんだろうといった問題意識を、この場を通 じて新たに作っていっていただければとお願いしておきたいと思います。

その点については、アンケート用紙を本日も皆様にお配りしていますので、そこにご意見などを書いていただければ、概要という形になるかとは思いますが報告書に掲載させていただければと思います。報告書につきましては、例えば、2年前に実施しましたシンポジウムにつきましても、大学入試センターのホームページに掲載されていますので、またご覧いただければと思いますが、それと同じような形で、皆さんお一人お一人の考え方が、ウェブや報告書を通じて共有できるというチャンネルも用意しております。また、パネルディスカッションでは、フロアの皆様方からのご質問・ご意見を伺う機会もあると思います。是非、お一人お一人、主体的にこのシンポジウムに参加していただきまして、より有意義なシンポジウムに盛り立てていただくことをお願いいたしまして、趣旨説明とさせていただきます。長時間になりますが、どうぞよろしくお願いいたします。

〔基調講演〕

認知科学にもとづいたアセスメント―― 思考力・判断力・表現力の評価・改善をめざすアセスメント システム

Paul Deane(米国 Educational Testing Service 首席研究員)

(ポール・ディーン)

米国 Educational Testing Service (ETS) 首席研究員。シカゴ大学で言語学 Ph.D.取得。ETS の研究開発部において過去 10 年ほど CBAL (Cognitively Based Assessment of, for and as Learning) プロジェクトをリードしてきた。認知科学・学習科学の知見を踏まえて、おもに K-12 段階の英語 (国語) の読解力・論述力の評価問題作成と自動採点の高度化に取り組み、学習改善・教員研修までをセットにした次世代型アセスメントシステムを開発している。



【山地】では、基調講演に入りたいと思います。まず、Paul Deane 先生の簡単なご紹介を申し上げます。これは皆さんのお手元にあるプロフィールのとおりなのですが、アメリカの ETS の首席研究員をなさっています。お聞きすると、2007 年ぐらいからこの CBAL のなかの英語(国語)を担当しておられるとのことです。

この CBAL は、もちろん詳しくは後で内容が出てくるのですが、"Assessment of, for and as Learning"ということで、最近の構成主義的、あるいはオーセンティックなタスクをもとにした評価というところでずっとやっておられるので、of だけではないということですね。Assessment of Learning だけではなく、for と as もくっつ

いてくる新たな評価方法について様々な開発を進めておられますので、本日はそれ らを具体的にご紹介いただくことになっています。

おもには K-12, つまり高校生段階までなのですが、お聞きすると K-16, 大学生までを込みにした形での、いろいろな認知機能をチェックする問題も作っておられるし、そのフィードバックの仕方についてもパイロット研究を行っておられるようですので、今の大学の学修成果の可視化などにも関連してきます。そこまで可能性を広げてお聞きいただければ、非常に大きな示唆があるのではないかと思っております。

それでは、Deane 先生、よろしくお願いいたします。

ありがとうございます。このシンポジウムに参加できて大変光栄に思っております。日本は美しい国であるとともに、皆さんが一生懸命に教育システムにおいて尽力されていることに感銘を受けました。日々、いろいろと勉強させていただいています。

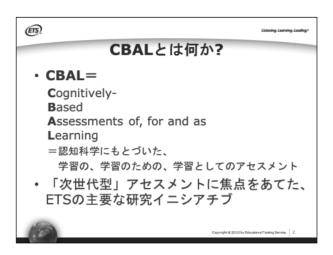


●CBAL の概要

では、まず背景からご説明します。私が説明するのは、運用システム自体ではありません。そもそも研究が始まったのは、2007年のことでした。取組として革新的な形で、アセスメントをいかに構造化するかということを考えたのです。

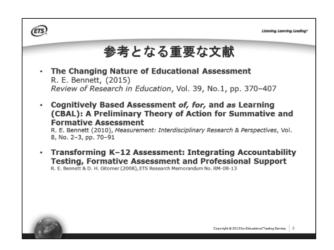
背景として考えのもとになったのは、アセスメントと指導の関係です。場合によってはマイナスの関係が出てくることもあると考えました。ハイ・ステイクスな試験が将来控えている場合、教師としてもいろいろな妥協をして、生徒を試験に備えさせなければいけないこともあるということです。

したがって、慎重に考えて、指導自体にプラスの影響があるようなアセスメントはどう あるべきかということを考えたのです。これが背景にありますので、ぜひそれを念頭に置 いていただければと思います。

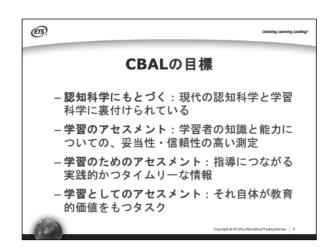


CBAL というのは、Cognitively-Based Assessments of, for and as Learning(認知科学に基づいた、学習の、学習のための、学習としてのアセスメント)です。これらの一語一語が意味を持ちます。これは私が在籍している ETS の主要な取組であり、アセスメントシステムを概念化するということです。単一のアセスメントだけではなく、アセスメントシステムというと、いろいろな要素が含まれます。進捗をモニタリングするという評価もあるでしょうし、形成的な評価、ベンチマーク的評価、それから教室における評価もあると思います。これらがどうつながるかを考えるということです。

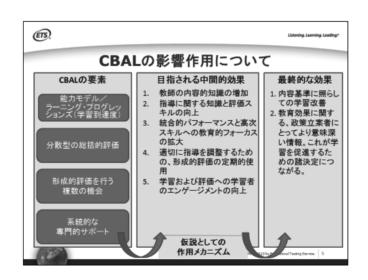
CBALという略称ですが、認知ベースという主要な概念が含まれています。認知科学に基づいて、学習者がどのように学び考えるのかを理解する。そして、学習でなければいけない、学習のためでもなければいけないということです。将来の指導や、あるいは教えている最中に使えるようなものでなければいけません。アセスメント自体が教育活動として意味のあるものでなければいけないということで、学習者にとっても学習経験になるようなアセスメントにしたいということです。



途中、このように文献を紹介いたします。CBALの取組を報告している文献です。この 講演では詳しくお話しできませんので、詳細を知りたいという場合には、CBAL 関連の文 献、出版物に詳細が記述されていますので、ご参考までに示します。



今まで申し上げたことの繰り返しになりますが、CBALの目標は、認知ベースの評価であるということで、現代の認知科学と学習科学に裏付けられていて、学習者の知識と能力についての妥当性・信頼性の高い測定、そして指導につながる実践的かつタイムリーな情報を提供するということです。



では、こうした目標にいかに向かっていくのか、どのような方法論を使って達成してい くのか、ということを明確にしなければいけません。

単純化したバージョンではありますが、CBALの影響セオリー、すなわちアセスメントがいかに指導にインパクトを及ぼすのかを示します。目標は、教育において成果を出すことですので、CBALの理論としては、いろいろな構成要素がモデルの中に入っています。まず、「能力モデル/ラーニング・プログレッションズ(学習到達度)」によって、どのようなスキルを重要視しているのか、どうやって学ぶのかということが枠組みとして提供されています。

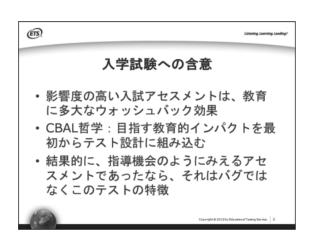
「総括的評価」についてその革新的なところは、分散型になっているという点です。一つの年度末の総括評価を1回限り大掛かりにやるのではなく、より小規模な評価を分散させて年度を通じて蓄積していきます。年度末になる頃には、学習者の到達度の全体像が分かるようにするということです。こうした方法をとる理由は、カリキュラムに直接関連の

ある形で指導ができる、また学習経験としても豊かになるからです。1 回限りの試験を年度末だけに行うと、それだけで学習者は圧倒されてしまいます。

それから,「形成的評価を行う複数の機会」です。いろいろなチャンスを通じて,教師が 到達度を見て調整できるようにするということです。「系統的な専門的サポート」も提供し ます。

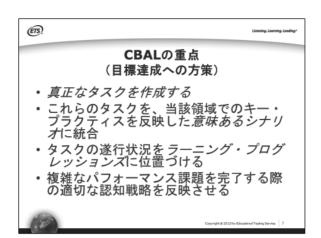
これらの構成要素は、アセスメントシステムの設計概念に入っているものです。CBAL の一環となっています。

そして「期待される中間効果」としては、教師の内容的知識の増加、指導に関する知識と評価スキルの向上、また学習者のエンゲージメントの向上も図っていきたいということです。つまり、学習者の学習と、教師・教育者として得る情報にプラスの効果をもたらしたいということです。



本日のシンポジウムのテーマは入試ですが、先ほども申しましたが、やはり影響度の高い入試アセスメントというのは、教師や学習者にとって大きなインパクトがあります。哲学・理念としては、最初からそのインパクトを考えておこうということです。アセスメントシステムの設計上、最新の研究成果に基づいて、教育的インパクトをプラスにしようということで、テストの準備だけに資源を割いてそれが学習者の役に立たないということがないようにしたいのです。

後でアセスメントの実際をご覧に入れますが、教室の活動のように見られると思います。 意味のある教育ということで、いわゆる紙ベースだけの伝統的なテストにとどまらないと いうことです。学習指導のように見えるアセスメントであったら、それはバグではなく、 テストの特徴だと思ってください。



目標達成に向けて、幾つかの方策を考えています。一つは、真正なタスクモデルの作成です。どのようなやり方を学習者に覚えてほしいのか、どのようなスキルをどのような文脈で使えてほしいのか、そのような文脈に合ったタスク、つまり真正なタスク作りをするということです。

もう一つは、意味のあるシナリオ作りをするということです。タスクを意味のある形で シナリオ上に並べ、実生活でスキルを適用するに当たって直面するような場面を再現した いということです。そして、キー・プラクティス(鍵となる活動)と言っていますが、そ うしたものを考えて、例えば、実際のスキルとして職業生活上必要になるものを学んでほ しいということです。

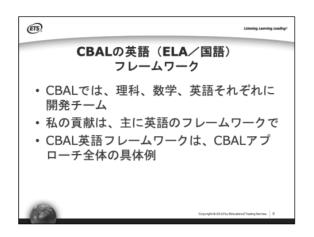
これはヴィゴツキーの活動システムの考え方なのですが、実践することによりスキルも 学んでいく、その認知的な発達があるということで、この考えを説明した文献はたくさん あります。なぜパフォーマンス課題が重要かというと、パフォーマンス課題により、達成 してほしい重要な取組の経験をしてもらうことで、複雑なタスクでも実際に遂行できる能 力を身に付けていけるからです。

作成するシナリオは、いわゆる伝統的なパフォーマンス課題と必ずしも同じではありません。段階ごと、ステップごとに評価をしていきます。パフォーマンス課題を分解して、ステップ・バイ・ステップで、そのタスクにいかにアプローチしたらいいかということが分かるようになっています。タスク全体をいきなり投げて、何のサポートもせずに測定するのではないということです。

それから、ラーニング・プログレッションズを重視しています。通常どのようなステップを経て学習が進んでいくのかを分析し、未熟な状態から熟達のレベルが高くなっていくためにはどのような質的変化が求められるか、ということを理解しておく必要があります。 タスクを作成するに当たっても、ラーニング・プログレッションズのなかのどの段階にいるのか、この段階には達しているか、あるいは次の段階に達するにはどうしたらいいか、といったことが分かるようにします。

最後の点は、適切な方略をモデル化して、複雑なパフォーマンス課題を達成できるようにするということです。やり方が分かっている学習者がいる一方で、できると分かっていても失敗する学習者はどうするのか、複雑なタスクは遂行できない学習者はどうするのか、

ということがあります。そこで、アセスメントにおいてシグナルを出してもらって、このようなステップを経ていかなければ達成できませんよ、ということが分かるようにしたいのです。

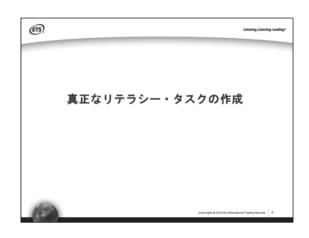


私自身が専門としているのは、国語としての英語です。Randy Bennett が CBAL 全体の リードをしていまして、数学や科学分野を担当している他の同僚もいます。私自身が寄与 しているのは英語です。

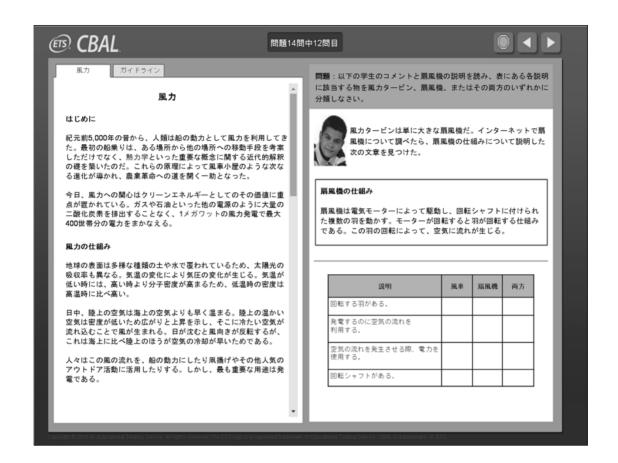
したがって、ここからはその話が中心になります。ただし、数学や科学でも CBAL の研究は盛んです。ETS のウェブサイトをご覧いただきますと、いろいろな文献が出ております。参考文献もご覧に入れていきます。

今回の話は、CBAL 全体のアプローチの一例だと思ってください。いろいろな領域があるので、私が中心となっている英語だけではない、ということを念頭に置いておいてください。

●真正なタスクの作成



では、幾つかの方略を見ていきたいと思います。どういうことをするのか、もう少し具体的に見ていきましょう。まず最初は、真正な(authentic)リテラシー・タスクのモデル化をすることです。



これは読解のタスクです。文章を読んで、情報の比較をさせます。例えば風力発電と扇風機を比較するのです。共通点と相違点があるものを比較させるわけです。伝統的なペーパーテストでは、必ずしもこういうテスト形式にならないと思います。

ここで問うているのは何かというと、表形式になっていて、どの表現が扇風機に当てはまるか、風力発電に当てはまるか、を答えるようになっています。グラフィック・オーガナイザーというのですが、読んで情報を整理するに当たって、グラフィック・オーガナイザーを使えるのは重要なことで、意味のある方略のモデルです。

私たちのタスクではこういうものを多用しています。それぞれの領域で情報を表現する に当たって、どのような方法であればその文脈で真正となり得るのか考え、真正な活動を モデル化するのです。心理測定レベルにおいても真正であるかどうかを見ています。

もちろん同じような問題を多肢選択問題でも出せるのですが、このような情報を整理するという活動は意味があって、実生活でもやっていることを反映しています。



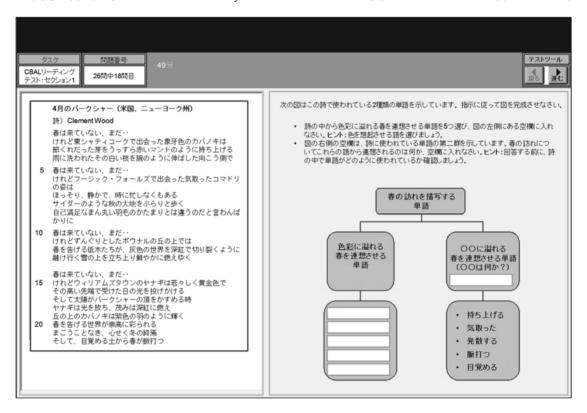
これはもう一つの例です。CBALのタスクで、より真正なものにと考えて作ったものです。テキストを読んで概略をまとめなさいという内容なのですが、まとめるに当たって有効なのは、例えば下線を引く、ハイライトを入れる、といったことです。

客観テストで多肢選択式でしたら、どの情報が重要か選びなさいとすることもできるのですが、そうではなく、大事な箇所に下線やハイライトを入れなさい、としてはどうでしょうか。デジタルで入れさせるのです。パラグラフを読んでいく中で、重要なところをハイライトさせ、自動的に評価できます。ハイライトするというタスクを与えることによって、これは重要なスキルなのですよというメッセージを出すことになるのです。重要なところをハイライトするというのは、実生活で文章を読むに当たっても使っています。



こちらはまた別の例なのですが、この文章を読んで、客観的・主観的なところを見分けなさいと。意見を表している用語を選びなさい、客観的な文章にしなさい、それで意見を表明している部分を削除しなさい、そしてこの文章をより客観的なものにしなさいということです。

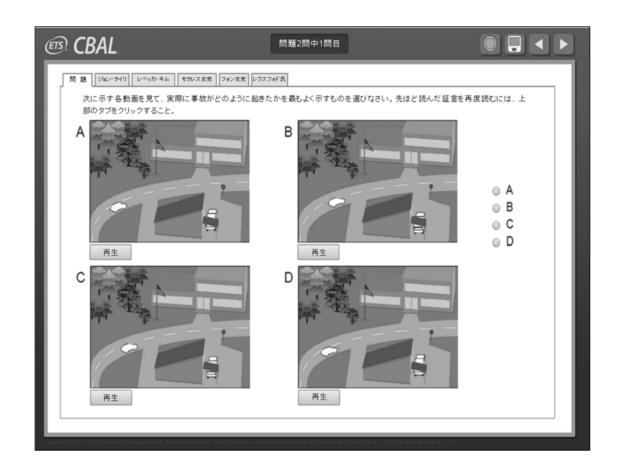
例えば、あるものを読む、あるいは書く際に編集や推敲(すいこう)をしますが、主観的でないようにするために、例えばこういう作業をやると思います。この中の文章で"unfortunately(残念なことに)"というのは主観的な用語です。したがって、事実の陳述から意見の陳述に、この unfortunately があることによって変わっていることが分かります。



これも、グラフィック・オーガナイザーという表を使っています。もう少し複雑なものです。英語では、このような表を多々いろいろな形で使っています。

ここでは英語の詩を読ませて、それを分析させます。そして、情報がかなり埋まっていますが、オーガナイザーをもとに空欄を埋めさせます。空白を与えて何もサポートしないのではなく、例えば色彩を表す用語を抜き出しなさい、それから右側は、春の表現として共通しているものをピックアップしなさい、としています。

このような表,グラフィック・オーガナイザーを使うことが助けになるのです。決して全部ゼロからやれということではありません。これは高校の1年生ぐらいでしょうか。これはそれぐらいのスキルレベルです。大学ではありません。大学であれば,もうグラフィック・オーガナイザーは,ゼロから文言を埋めなさいとするかもしれません。ただ,そこまでスキルがないかもしれないということになると,どれぐらいの習熟度なのか分かるように,一部埋めて,その上で出すということです。



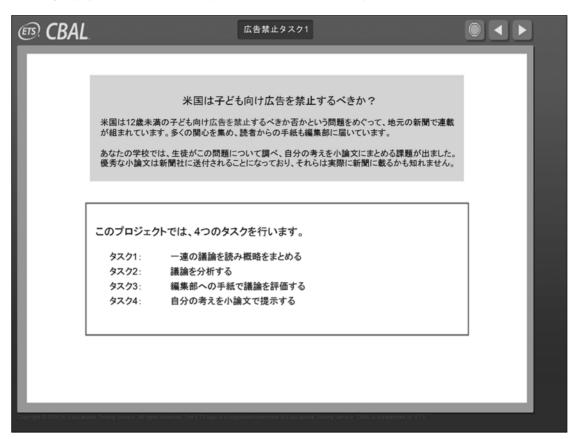
もう一つ、こちらは多肢選択式の客観テストです。ビデオを選ばせます。A・B・C・D とあります。アセスメントの一環として、過去に起きたことの推論を行うのです。事故が起きたとします。学校の外で事故が起き、いろいろな証人がどのように事故が起きたのか証言をします。それを聞いた上で、本当は何が起きたのかということを判断します。

したがって、最終的にいろいろな証拠が、証言としていろいろな証人から与えられます。 アニメを四つ見て、実際に証言を聞いた上で、事故として何が起きたか、どれが正しいと 思いますかということを選ばせるのです。情報を抽出して、自分でそれをまとめて判断し ます。これは実際何が起きたのかということの判断をさせるという真正なタスクです。言 葉だけではなく、動画も使っています。

●意味のあるシナリオ



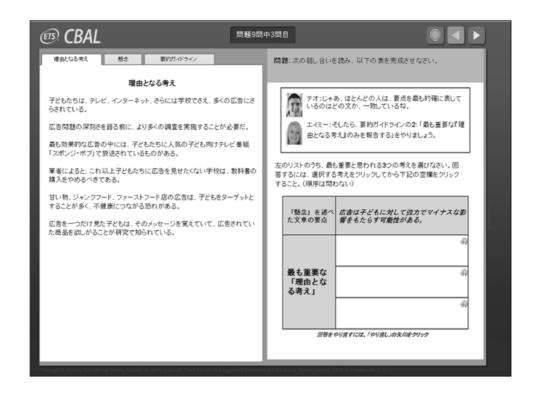
次に、意味のあるシナリオを活用するということです。



意味のあるシナリオとはどういうものなのかというと、考え方として、例えば英語のスキルを教える際には、いろいろな連続したことをやって、最終的な目的を達成できるようにしてほしいということです。例えば討論をするのであれば、そのテーマ・論点に関する情報を収集して、それを整理して、集めた情報からその議論の、自分は賛否両論のどちら側なのかを決めて、それを発表して、反論にも備えます。

これは活動としての討論のいろいろなステップです。そうしたステップを経て、議論の構築、討論の構築をするのです。もちろん、学習者はそれら全てが整理できているわけではないかもしれません。成功裏に全てのスキルが提供できるわけではないかもしれません。あるいは、段階的にそのステップを経ることができないかもしれません。

そこで、シナリオを作るのです。全ての設問がシナリオに関わっている形で、シナリオをたどらせるのです。だからこそ、複数回テストをやる意味があるのです。議論のシナリオ、研究のシナリオ、そして文学作品の読解のシナリオをそれぞれやるのですが、各シナリオで段階的に意味のあるステップを経て、スキルの構築をしていける形になっています。



では、どういう形で進むのかお見せしたいと思います。まず、アセスメントの最初の段階です。議論、討論をするに当たり、これを読みなさいと与えて、そしてそれについて誰かがメモした内容を読んで、何がメインのアイデアで何が補助的なアイデアなのかをまとめさせます。グラフィック・オーガナイザーの表を使っています。タイプするのではなく、ドラッグ・アンド・ドロップするなどして、テキストからグラフィック・オーガナイザーに入れて表を埋める形にしています。これは自動採点ができます。



次は、サマリーが書けるかどうかというタスクです。読んだ文章についてまとめなさい、 ということです。こちらのスクリーンですが、子供向けの広告は禁止すべきか否かという ことで、 賛否それぞれの意見が設問ごとに表されます。ここでは、一方の議論が書かれて いて、それをまとめなさいというタスクです。

幾つかの設問を経て、スキルが構築されるようになっています。うまくサマリーが書けたかどうかの基準を示して、そして欠点があるようなサマリーも事例として見せて、それでメインのアイデアは何か、補助的なアイデアは何か、と選ばせるタスクも与えて、一連の作業をする、そして最終的に自分の意見を陳述できるようにするのです。

® CBAL	問題7問中1問目	
意見文 コマーシャルは大体が覚えやすく忘れにくいようになっている。このため、自分のために良くない商品であっても、コマーシャルによって購入してしまう人がたくさんいる。健康な大人の多くは、子どもの頃にたくさんのコマーシャルを見て育った。 広告の効果について十分な知識を持ち、子どもを守りたいと考える親たちでも、子どもの見る広告についてあまり心配しない。 ジャンクフードの広告主は、子ども向けに広告することを好む。これは、子どもは大変影響されやすいこと、また幼少時についた習慣は大人になっても変わりにくいことを	意見文を吟味しなさ 問題:それぞれの意見文を読み、子ども向け広告について しなさい。回答するには、選択する意見をクリッグしてから、 禁止	て、禁止または容認のどちらの理由になっているかを判断
知っているためである。 子どもが広告を見てその商品を親にねだり続けると、その商品の種類に関わらず、親は大概聞き入れてしまう。 多くの人は、目にする広告のほとんどを無視する。 広告が何かを買わせようとしているとは、子どもは理解しないかもしれない。 広告に影響されて子どもが商品を欲しかったなら、親はその機会を活用して、賢い選択の仕方を子どもに教えることができる。		
大概において、親は子どもがテレビで何を見ているか確認 する手間をかけないため、子どもが見ている広告について 無知である。 子ども向け広告に賛成を強く唱える人々の中には、広告業 界から大きな利益を得ている人がいる。したがって、そう した人の主張を鵜呑みにしてはいけない。		

今度は、議論の分析をする段階です。こちらのタスクは、もちろん多肢選択式でもできますが、T チャートと呼んでいるものを私たちはよく使います。グラフィック・オーガナイザーで、横棒と縦棒を T 字型に引いて、その上にトピックを書きます。T 字の例えば左側には賛成派、右側には反対派を入れるのです。

10 のコメントがここに書かれていて、これは賛否のどちらに入るのかを分けさせるのです。左側が禁止すべき、右側が容認すべきという立場の理由です。実際、自分の思考をある論点に関してまとめる際には、こういうことをやると思います。それぞれの文を読んだ際に、賛成か反対か、必ずしも明示されていません。しかし読んで、これは賛成だ、反対だ、ということを判断させるのです。重要なスキルです。

このタスクにおいてこれができないと、2~3 文、書くだけに終わってしまいます。例えば、読解や思考することに問題がある学習者には、これは難しいタスクです。したがって、まずはこういうことができるようにスキルを構築しないと、最終的に小論文などは書けないわけです。

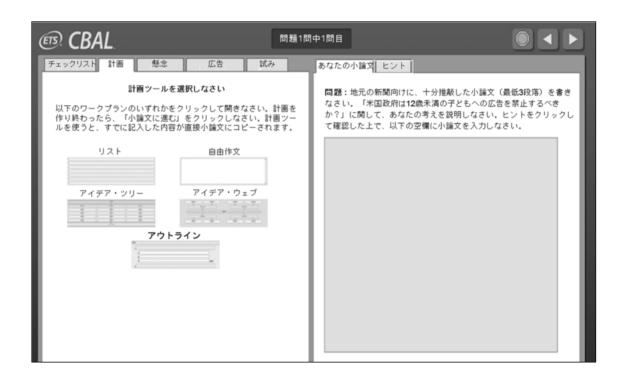


今度は、議論とそれに伴う裏付けの証拠を示して、その証拠が強いかどうかという評価をさせます。



さらに複雑なタスクがこちらです。議論・討論を与えるのですが、かなりまずい議論を読ませて、それを批評させます。これは8~9年生レベルのタスクなのですが、これはアメリカの8~9年生にはかなり難しいタスクだと思います。上位の学習者しかうまくできないだろうという想定です。そもそも批評や反論が分かっていなければいけません。それから、メッセージの発信として、批評すること、反論することは重要で、議論においては本質的であるということです。それを教えるものです。

このような、意味のある一連のタスクをやってもらって、分析をさせます。最終的に、パフォーマンス課題として、小論文を書きなさい、となります。読んだ文章、広告の賛否両論に関してもアクセスできる、見られるようにしています。スコアの基準がどうなっているかということも示します。



また、計画ツールとして、考えをまとめるようなツールもちゃんと提供します。ステップを経てこの段階まで来ると、かなり記述の評価ができると予想されます。最後の論述のタスクまで来た頃には、トピックに関してかなり豊かな知識が積み上がっているはずです。したがって、トピックについて何が分かっているか、何を知っているか、それについてどう思考しているか、ここまでタスクをやってきたのですから相当評価ができるということです。

以上、CBALの評価とはこういう形で進める、ということをお話ししました。

●CBAL の実際(ビデオ)



では、ここでビデオをご覧に入れたいと思います。ETSのウェブサイトにも掲載されているものです。CBALの実際について感触をつかんでいただくために、アメリカの学校の教室でどういう形で実践されているか、見ていただきます。

…………ビデオ開始(カギ括弧部分は教師が生徒に向かって話している箇所)…………

Rafano (英語教師)》「学校でこういうジャンクフードを売っていいかどうか、ということを議論してください。賛否両論の議論を実際に読んでみてください。最終的に、小論文を書いてもらいます。」

今日は、生徒に対して情報源はどこなのか、あるトピックについて、この情報がいい、悪い、の判断をする前に、情報源の検証をします。

- Nick (男子生徒 (9 年生))》 CBAL では、自分の論述がうまくいくために、どのような構成要素が必要か教えてくれるのです。
- Deane》なぜこうした異なるスキルが重要なのか、ということが分かるようになっていま す。ちゃんと目的に結び付いているからです。
- Marquez(Assessment Developer, ETS)》そしてステップごとに、足場かけと呼んでいますが、節目節目があり、教師側としても立ち止まって、ちゃんと理解が進んでいるか、見分けることができます。発達のレベル、理解のレベルを確認することができます。
- Rafano (英語教師)》CBAL の情報により、実際に先に進めて構わない生徒は誰なのかということも分かりますし、タスクを達成するに当たって、何が必要か分かっていない生徒を見分けることもできます。最終的なプロジェクトに到達するに当たって、どの生徒がどの段階にいるかということが見分けられます。

「では、ペアになって、次の五つのタスクをパートナーとやってください。後で、全体で発表してもらいます。」

小グループで作業してもらいます。例えば同じようなタスクやトピックで、同じレベルで苦労している生徒同士をペアで組ませます。それからピア・ティーチングといって、生徒同士で教えさせます。実際に聞いたり見たりすることによって、いろいろな学習モードを通じて CBAL では学習させます。

- Payal (女子生徒 (9 年生))》 CBAL はすごく好きです。なぜ CBAL がいいかというと、家庭でもアクセスできるし、いろいろな図解やビジュアルが載っていて、そちらの方が分かりやすいからです。
- Marquez(Assessment Developer, ETS)》シナリオベースなのです。新鮮な文脈の中で、シナリオが提示されます。そして設問は相互に関連していて、積み上げられるようになっています。直観的なレベルから、より高次の公式の理解が構築できるように進められるところがいいのです。

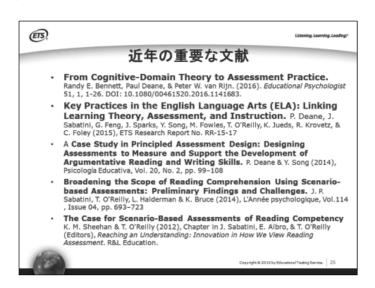
評価はコンピュータ・シミュレーションが伴うことが多いです。コンピュータ・シミュレーションを使うことによって理解が進み、重要な概念を発見することができます。

- Perry(数学教師)》こうしたタスクで、シミュレーションを通じて実世界の状況として、自分たちがなじみがあるものに結び付けて理解することができます。

 子供はビジュアル化して見ることが好きなのです。ただ板書で、私が言葉で説明するだけではなく、シミュレーションすることにより、例えば比や割合がどう違うのかが、シナリオごとに分かるようになっています。
- Tiffany(女子生徒(9 年生))》CBAL により、学んでいることとビジュアルを結び付けることができるのが非常にいいです。テストはつまらないと思います。
- Marquez (Assessment Developer, ETS)》 CBAL のアセスメントの特徴は、なぜそう思ったのか説明させることによって、その生徒の思考が分かることです。
- Perry (数学教師)》子供たちはすごくタイピングが速いのです。コンピュータも携帯電話も 使い慣れているから、すごく解答も速いです。
- Gabrielle(女子生徒(9年生))》比と割合、それから分数をやったのですが、違う見方ができる、違う形で教えてくれるのがいいと思います。
- Perry(数学教師)》「では、比と割合として、これに合うのはどれでしょうか。」
- Deane》このアプローチの要となっているのは、テストの中のタスクを、教師がその概念を教えるに当たって違和感のないものにしたいということです。この研究はまだ継続中で、いろいろな洞察、方法論、多くの質問も出てきています。まだまだ研究途上です。

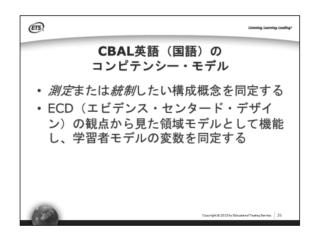


CBAL の素材が実際にどう使われているか、ご覧いただきました。感触をつかんでいただけたでしょうか。

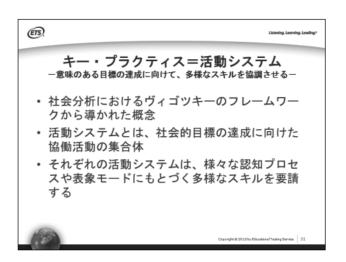


時間が限られていますので、認知科学的な背景については、何枚かスライドを飛ばすことをお許しください。さらに詳しくお知りになりたい方は、我々は幾つかの文献を出しておりますので、それらをご参照ください。

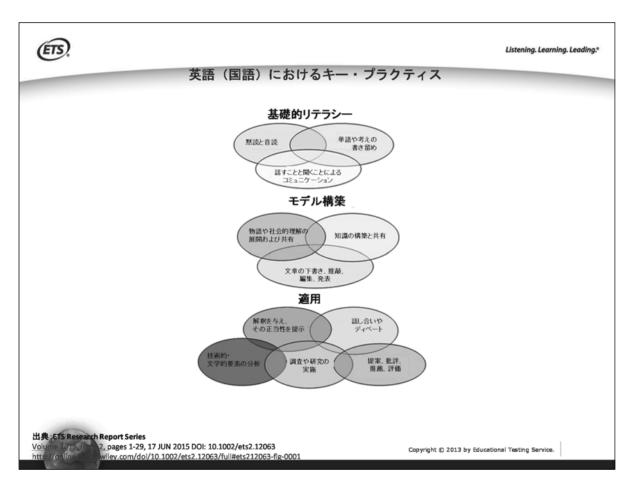
●キー・プラクティスとタスク分析



アセスメント設計の枠組として、一つのコンピテンシー・モデルがあります、測定したい構成概念や統制したい構成概念を考えた上で、アセスメントの世界で言うところの領域モデル、つまり一体何をテストしようとしているのかというモデルを立てます。明示的なモデルとして、これはテストの設計を最初から規定するものなのです。



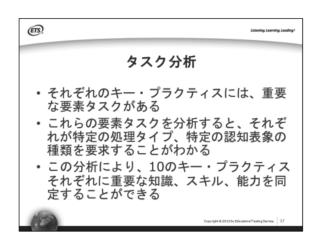
大事なのが「キー・プラクティス」で、主要な活動、鍵となる活動、と言われているものです。意味のある社会的な相互作用を通して人々が携わる活動で、教育上も重要なものです。



図のなかに「適用」とありますが、大学に行くと、そうした英語のスキルはもちろん重要です。例えば解釈、文学作品を理解する、メインのアイデアを議論する、リサーチする、問題解決をする、提言や評価をする、という一連のいろいろな活動があり、人々はそうしたことを大学で、あるいは職業上やっているのです。

職業上の活動では、例えばビジネスにおける問題を特定しなければいけないし、特定したら、次は解決策を出せと言われます。そしてそれが正しい解決策なのか評価し、そして計画を立てて実施し、その後、成果を評価せよと。これは活動として1か所で起きるのではなく、何度も反復して、いろいろなコンテクストの中で繰り返し実践される、またいろいろなバリエーションがあり得る、複雑なスキルです。研究であれ、議論や討論であれ、文学作品の解釈であれ、いろいろな活動の段階があり、それを経ていかなければいけません。

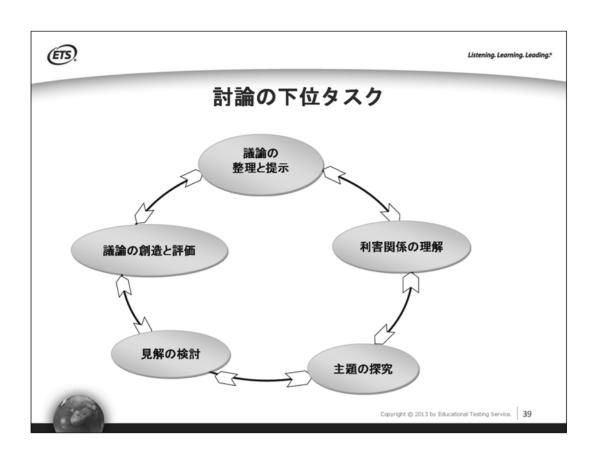
各段階で必要なスキルがあります。その段階の活動を行うには、特定のスキルが必要です。したがって、それぞれの活動を評価し、それがどうつながっているのか、意味のある続き方になっているのか、またそれを支えるスキルとして何が必要か、といったことを示さなければなりません。



それがタスク分析です。すなわち、どんな認知的な処理を行っているのか、またどのように情報を表象しているのか、それらに関する分析を行うのです。

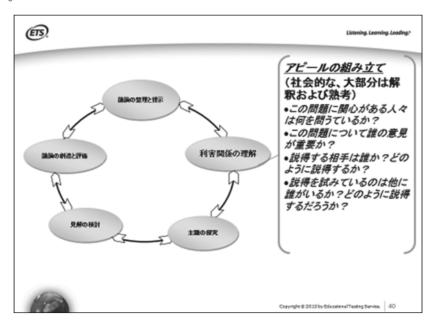


ここで例として紹介するのは、討論あるいはディベートに参加するときに、どんなタス クがあるかということです。

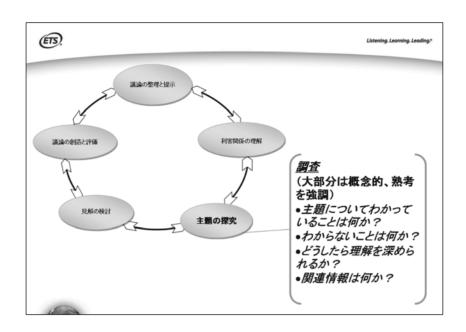


下位タスクに分けて考えていきます。すなわち、議論に参加することがうまい人であるならば、恐らくこの五つのことをやっていると考えられます。

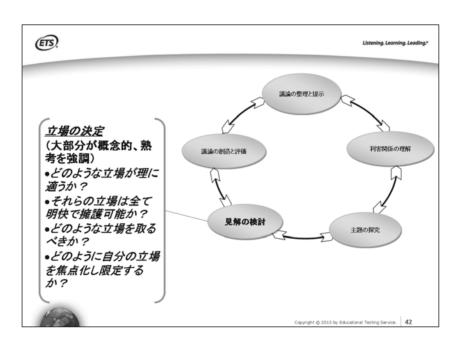
そして、文書ではなく、実際に対面して行う討議の場合は、これが対話として何度も繰り返されながら使われているのです。例えば口頭でうまく主張することができる人は、書くのもうまい、作文もうまい、ということがしばしばあります。また、お互いにこのようにして討論しながら対話をするのですが、熟達者の場合はかなりのスキルを持っていますので、いろいろ後戻りをしながらやる必要がなく、それでも成功裏に意見をまとめることができます。



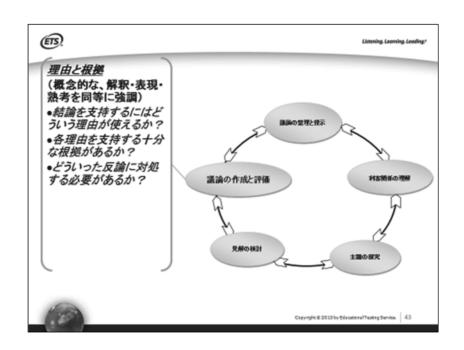
そのためには、この討論で誰がどんな主張をしているのか、またその理由は何なのか、 またどうすれば説得することができるのか、といったことを知っている必要があります。



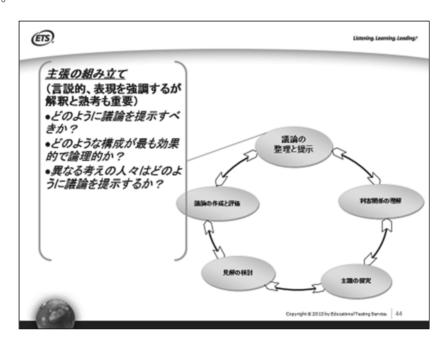
主題についての知識も重要です。ですから、調査や研究のスキルにも関係しています。



ロジカルな主張をしていくためには、その立場として複数のことを考える必要があります。どの主張が合理的であるのか、どんなものであれば支持できるのか、また、どの立場を少し変えることで、擁護可能なものにできるのか、といったことを考えます。討論を実際に行う、あるいは立論する前に、このようなスキルを身に付けていくのです。



また、その評価をするために、証拠や反論、様々なことをあらかじめ考えるのですが、 最終的には、ここまで行った思考を全てまとめた上で、他の人に発表していかなくてはな りません。

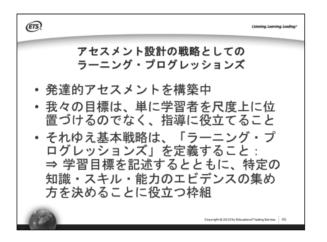


主張をどのように構成するのか、テキストとして効果的に発表するにはどうしたらいいのか、ということを考えなければなりません。そこでは、様々なスキルが必要になります。 それぞれのスキルの中にさらにサブスキルがあり、それらを積み上げていくことで、初めてうまくできるようになるのです。

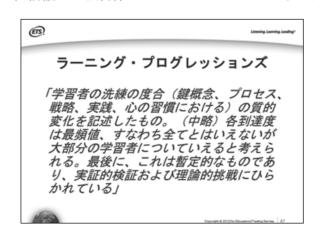
●ラーニング・プログレッションズとアセスメント設計



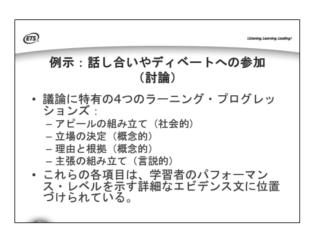
それぞれのスキルのラーニング・プログレッションズ、すなわち、学習者が今どの段階 にいるのかをどのようにして評価していくのか、お話しします。



ラーニング・プログレッションズは、発達を見るためのアセスメント枠組です。それぞれのスキルの要素をモデル化しており、それぞれの段階の間が関連しています。学習者がどのゴールを達成しているのか、発達的に見て力を伸ばしていく上で重要な要素を、どの学習者は既に達成しているのか。例えば、賛成・反対の部分だけを理解しているのか、もっと深く理解しているのか、またその主張についてどう正当化するかまで理解しているのか。理解が深まるほど、討論や主張自体が豊かなものになっていきます。



ラーニング・プログレッションズは、その言葉自体はまだまだ曖昧さの多い用語ですが、 CBAL における定義がこちらに出ています。すなわち、鍵となる概念、処理過程、戦略、 実践、心の習慣に関して、熟達レベルの質的変化を記述したものということです。つまり、 知識であれ、スキルであれ、理解がどのようなレベルに到達しているのかを考える際の用 語です。



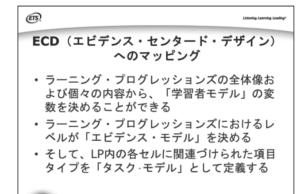
例えば、討論を進めていくためのスキルを学ぶときにも使います。

全体像一理由と根拠	
見解は、聴衆が納得する理由で裏付ける必要があるという 考えを理解する	準備
根拠の必要性をある程度わかった上で、理由を認識し作り 出し詳述する。また自分と馴染みがあり関係がある文脈で あれば、自分自身の議論で他者の議論に対抗する	基礎
根拠の活用について理解する。さらに、主張に直接関連し 論理的に問題なくそれを支持する根拠や理由を提示する必 要性を明確に把握する	初級
批評や反論の役割を理解し、反証や批判的質問に対し議論 したり応答したりすることができる	中級
ディベートというものについて体系的なメンタルモデルを 構築し、そのモデルを使って自分の知識構成の試みを方向 付ける	上級

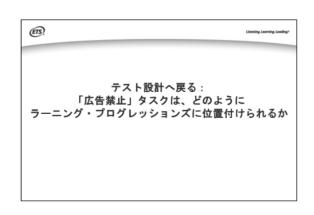
概観しますと、この五つのレベルがあります。例えば9年生であったならば、レベル4にいる人たちが多いでしょう。でも、レベル2や3にいる人たちもまだいるのではないか。そして、大学になると大体レベル3~4の学習者が多いでしょう。しかし、小学校であれば、恐らくレベル1~2の生徒がほとんどだと分かります。ですから、どの辺に多くの学習者が入ってくるかも、一つ理解しておくべきことです。

,,,,,,	解釈	表現	グレッションズの一部 熟考
準備	特定の主張を支持す る根拠を見つける	特定の主張を支持する理 由を少なくとも一つ考え 出し、文章にまとめる	議論形成の方略のテンプレー を適用できる(リストを作っ り賛否図を埋めたり等)
基礎	裏づけとなる理由や 根拠を文章中でみつ け、それらが支持す る主張と関連づける	主張を裏づける複数の理 由を考え出し、自分と関 係があり馴染みのある文 脈であれば、それらを他 者に反論するために使う	議論の一般的な枠組に関する 黙の理解を反映するものとし 分析的な方略を適用して主張 裏づけるのに必要な情報をみ けることができる
初級	主張と裏付けの言明 の関係を認識し説明 できる。どの根拠が どの主張の裏づけと なっているか把握し ている	主な主張・補助的な主張 を支持する理由や根拠を 選び整理することで、論 理的かつ階層的に議論を 組み立てる	一般的な三段論法と誤謬を認 することで、根拠の有効性を 価し、正当な議論とそうでな 議論を識別することができる
中級	文章中で、異論や反 論に弱い主張がわか る	他者の議論の要約とそれ への意見を批判的にまと めた簡単な批評や反論を 書く	討論の一般的枠組における掛 的質問方略を適用して、反論 考え、主張を補強する方法を めることができる

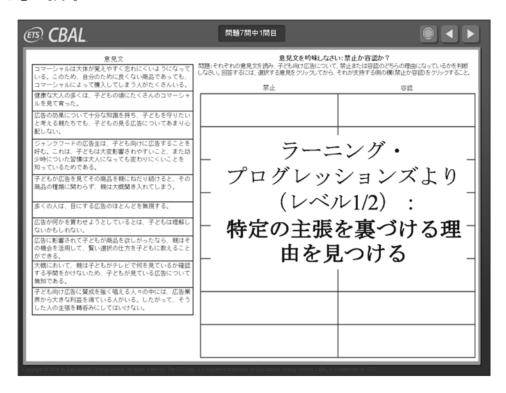
より具体的な例が上のものです。



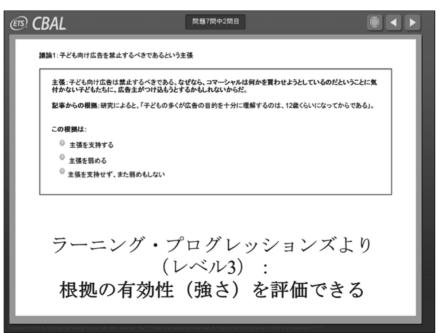
そして、やるべきタスクを理解し、各セルの記述に基づいてそれをマッピングしていきます。今、学習者は何をすることができるのか、プログレッションズのレベルを見ていきます。各セルの記述は、一部は文献から引用したものですし、また実際の教授法、教育現場から取ってきたものもあります。また、仮説も含まれています。このようなものを設計していく上で、エビデンスを使いながらまとめたものです。



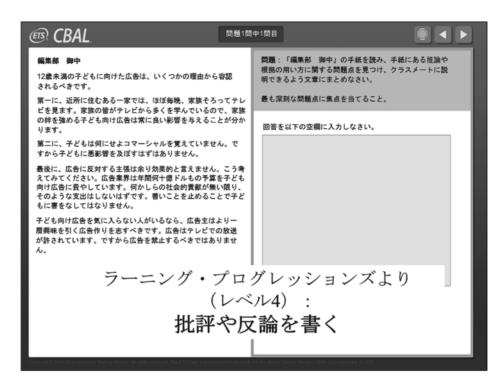
ではここで、これらがシナリオベースの評価にどのように結び付いていくかをお話しし たいと思います。



先ほどお見せした T チャートですが、これはプログレッションズのレベル 1 及び 2 のものです。「特定の主張を裏づける理由を見つける」というタスクで、討論の非常に基本的な部分です。ここができないと、学習者は深刻な問題を抱えていることになります。



このタスクは「根拠の有効性を評価できる」かどうかを見るもので、レベル3のタスクです。すなわち、単に理由やエビデンスを理解するだけではなく、どの理由、あるいはどのエビデンスが適切かを評価できなければなりません。



この「批評や反論を書く」タスクはレベル4です。これを成功裏に行うためには、何が よい議論かを理解しているだけでなく、より深く分析ができなければなりません。

こうした一連のステップは、教室での実際のプロジェクトに対応しています。それぞれのステップで、学習者がラーニング・プログレッションズのどの段階にいるかを知ることができるのです。そして、全体のステップが終わる頃には、学習者のパフォーマンスのパターンが把握できるのです。

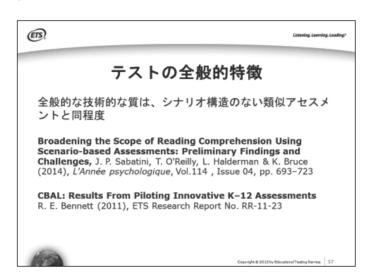
非常に一般的なパターンとして、T チャートはとてもうまくできるが、根拠の評価はまあまあ、そして批評については全くできないということがよくあります。8 年生によく見られるパターンです。大学進学前の場合には、批評のタスクが半分くらいはできてほしいところですが、実際はかなり難しいでしょうね。GRE という、ETS が大学院向けに行っているテストがありますが、その中のライティングのタスクとかなり近いものなので。

●CBAL の評価



では、現在集まっているエビデンスで、CBALのアプローチがどの程度うまく機能しているかを示したいと思います。というのは、このようなアプローチが従来型の評価とかなり違うので、批判する人たちもいるからです。

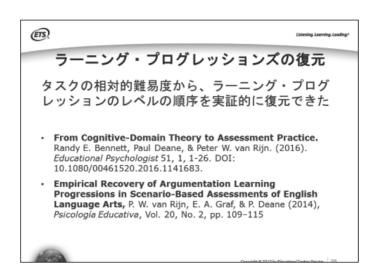
例えば、このような設問を全て一つのトピックに関してやるのですか、それは危険ではありませんか、つまり、全部同じトピックだと情報が失われませんか、と言う人もいます。あるいは、設問がだんだん複雑になっていきますね、それは足場かけになっているようですが、それでこれまでよりもよい評価が可能ですか、などと言う人もいます。これまでと違って、まるで教えるのと同様な感じでのテストを創るアプローチですから、いろいろな批判が出てきます。



シナリオベースのテストですが、このテストの技術的側面は、従来の多肢選択式や記述 式テストとそれほど異なっているわけではありません。シナリオ構造になっていることに ついて、最初にできる主張としては、もしこれをやっても測定という点で害がないのであ れば、またテストがより直観的に自然に取り組むことができ、教師にとっても何を教える べきなのかがはっきりするものであれば、この形態でのテストをやってみるべきだという ことです。



一方、トピック情報について、設問間に関連があるために情報が少なくなり、テストが偏った結果をもたらさないかという疑義があるかもしれません。しかし、次元性の分析結果の示すところでは、読み書きの違いはトピックの違いよりも大きかったのです。また、タスクの相対的な難易度からラーニング・プログレッションズの順序性を復元することもできたのです。



出版予定の成果

(ETS)

論述タスクの前に準備タスクを行うと、 論述タスクの後に他のタスクを行う場合 と比較して、より信頼性が高い全体スコ アが得られた

(Mo Zhang他による出版予定の研究)

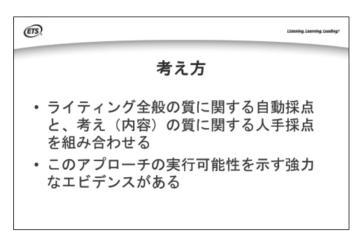
最近行われた研究として、小論文の前に準備タスクを行った場合と、準備なく小論文を書かせ、その後でいろいろな質問をする場合とを比較したものがあります。一つのトピックに関して小論文を書かせて、その後で質問をするというやり方よりも、同じトピックについてまず読ませて、質問をして、それから小論文を書かせた方が、評価という意味では信頼性のある結果が得られました。逆の順番にすると、結果はそれほど悪かったわけではないのですが、少し信頼性が劣りました。

しかも、この小論文を最初に書かせる、あるいは準備をしてから書かせる、どちらも評 点自体はそれほど変わらなかったのですが、最初に準備をしてから小論文を書いた学習者 の方が、たとえスコアが同じだったとしても、時間の使い方がずっと効率が良かったので す。書き始めたときに、既に内容について学習者はよく知った段階で書いているからです。

●ライティングの採点と報告



ではここで幾つか、CBAL プロジェクトの他の側面についても説明したいと思います。 特にここではライティングの採点と報告について申し上げたいと思います。



まず、いろいろなライティングの側面をどう採点するのか、また何に関しては自動採点が使えるのか、どういう場合は使わない方がいいのかをお話ししたいと思います。自動採点の大きな問題は、内容や主張を理解できないという点です。言語の要素などは自動採点で行い、内容や主張は人の採点で行うやり方であれば、非常に有望です。



また、重要な文献もあります。ここに挙げている文献などは、そのようなやり方をした ときに何が起こったか、結果について教えてくれます。

(ETS)

Listenina, Learnina, Leadina,

議論の質を評価するルーブリック

レベル4 (明確に要求を満たす)

議論の構成の仕方

- 問題に関して明確に見解を述べている
- 見解を支持するために理由と根拠を用いている
- 反論を考慮して対処しているところもある

議論の質と展開

- 問題に関して理に適った主張をしている
- 関連した理由や例を引用して説明を加え、主張を裏付けている
- 根拠の用い方が大体において正確である

読み手への意識

- 対象となる読み手に適した内容に主眼をおいている
- 読み手に適切な、かつライティングの目的に適う語調で考えを表現している

ここではルーブリックがどんな形になっているか例示したいと思います。それぞれのライティングの種類ごとにルーブリックがあります。例えば議論の質について、ルーブリックには5段階あるのですが、このレベル4というのはかなり能力が高いことを示しています。例えば、問題に関して明確に意見を述べることができる、見解を支持するための理由と根拠を用いている、あるいはエビデンスの使い方もうまい、そして聴衆のことも念頭に置いた上で議論をしているなど。小論文の採点をする上で、このようなルーブリックを使えば成功裏に行うことができます。



Listening. Learning. Leading.*

全体的なライティングの質を評価する ルーブリック

レベル4 (明確に要求を満たす)

適切に構成されている

全般として、タスクに対する回答が明確かつ適切に構成されている。関連した考えが適切にグループ分けされ、必要に応じて節や段落に分割されている。考えのグループの間の移行が、適切に合図されている

首尾一貫している

大半の新しい考えが適切に導入されている。文の連なりにおいて、一つの考えから次の考えに移る際に読み手を混乱させたり焦点が変わったりすることがほとんどない。文中での、また文と文との間でのつながりが、代名詞、接続詞、従属関係等を使用することで、必要に応じて明確に示されている

適切に表現されている

考えが明確かつ簡潔に表現されている。単語選択が十分多様な語彙力を表している。文の長さや 構造に適切な多様性があり、焦点や強調点をコントロールしている

標準的な英語の書き言葉を適切に運用している

標準的な英語の書き言葉のルールに従って文法と用法を用いているが、軽微な誤りが見られることがある。同様に、標準的な英語の書き言葉のルールに従って綴り、句読点、大文字を用いているが、軽微な誤りが見られることがある



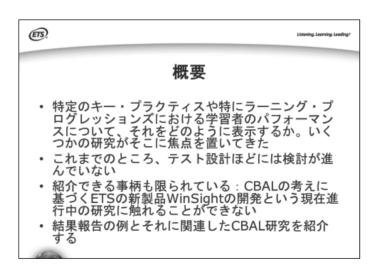
Copyright $\mathop{\circledcirc}$ 2013 by Educational Testing Service.

全く違うところにフォーカスしたのがこちらです。全体的なライティングの力に焦点を当てています。すなわち、全体として構成がどれだけ適切であるか、首尾一貫しているか、表現が適切であるかということを焦点としています。主旨に関してはあまり取り上げていません。またこのような内容であれば、自動採点であっても大体同じような結果になります。

一般的に、全般的なライティングの質と議論の質との間には強い相関があるものですが、 ミスマッチがあった場合、例えば議論の質は高くとも言語的な問題を抱えているような学 習者の場合は、両者の情報を得ることが助けになります。しかし、非常に流暢(りゅうちょう)な書き手であるけれども必ずしも意味ある主張を構成することができないという場 合もあります。



次に得点の報告についてですが、どんな研究、実験が行われているのかをここで紹介します。

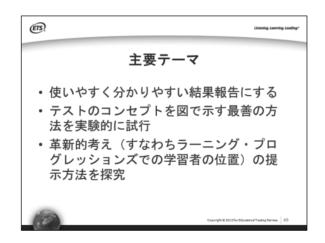


私たちは、まだ非常に予備的な段階なのです。すなわち概念実証という段階であり、まだまだ現場で使う段階ではないということです。いろいろな問題を解決するために、こんなことをやってみました、こんなことを考えてみました、という段階です。

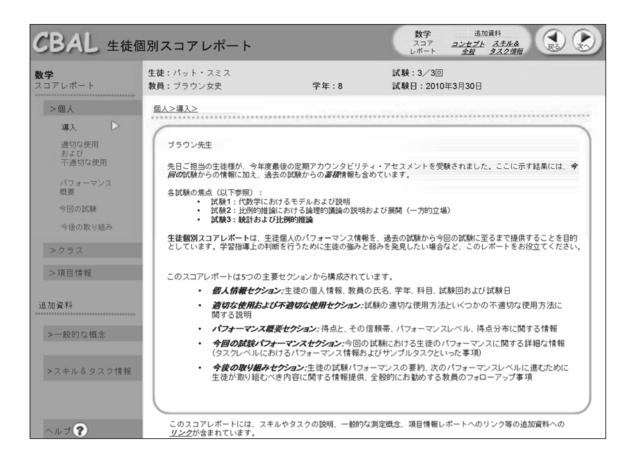
このような情報を報告する際、考えなければいけないのは、次に挙げることです。スコ アが何を表しているかを皆が理解しているかどうか、あるいはスコアをどう有効に使った らいいかが分かるか、また、そのスコアを学習にどのように結び付けたらいいか分かるか。 もし、学習の改善のためにどうスコアを使ったらいいのかが分からないとすると、そうい うスコアにはほとんど意味がないのです。



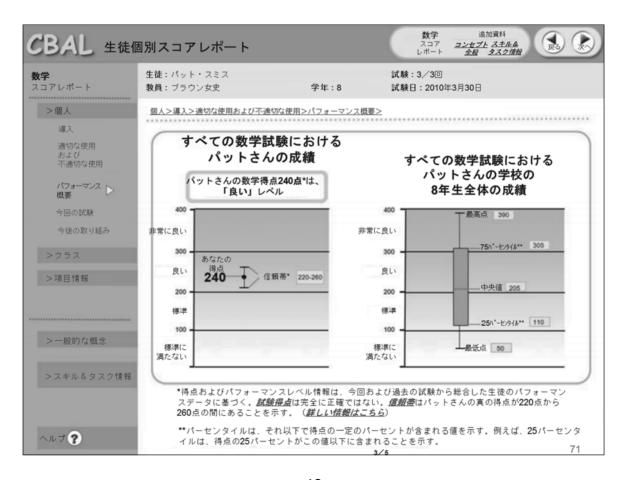
ここでも幾つかの重要な文献を挙げています。



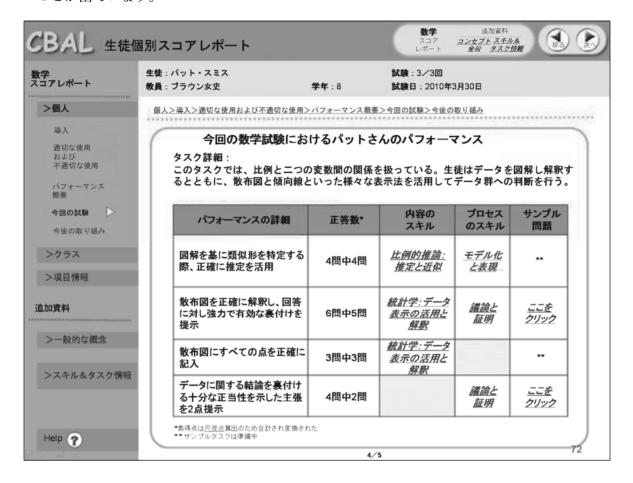
また主要テーマとしては、どうやったらこのスコアレポートを使いやすく分かりやすいものにできるのか、例えば、図やその他のビジュアルを使って分かりやすくするにはどうしたらいいのか、また特にグラフを使った場合、ラーニング・プログレッションズでの学習者の位置と次の学習の焦点をどのように示したらいいのか、といった検討も行っています。



実際にどういうことをやっているか、これは Diego Zapata という私の同僚が論文にしている内容です。

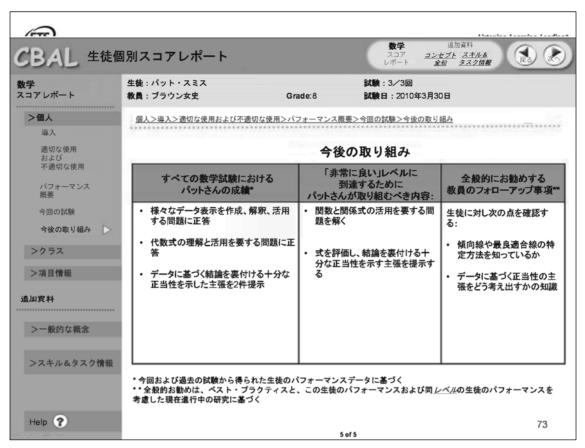


テストのレポート,これは数学のテストですが、例えば割合についてのものです。その構成の仕方としてどんな要素が入っているのか、非常にクリアに示しています。ここには、全体のパフォーマンスがどうであったのか、またスコアをどう解釈したらいいのかということが出ています。

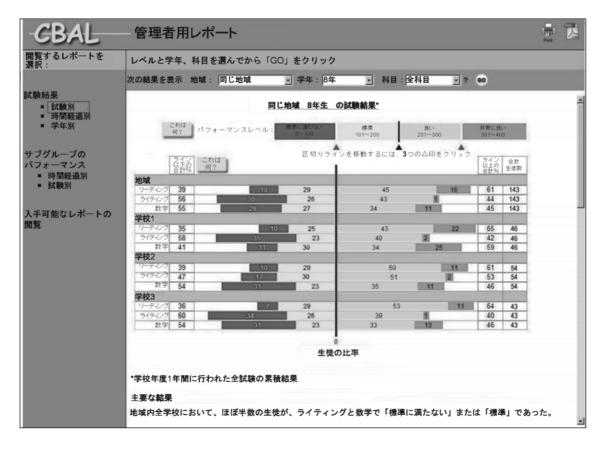


ここでは、そのパフォーマンスを分解して、実際に評価の対象になっていたものに分けて見ています。例えば一連のテストがあり、それぞれが意味のある関連性を持っていたのであれば、このようなものに分けて考えると非常に豊かな情報を引き出すことができます。 以上は割合に関しての例ですが、もし同じことを討論・議論に関してやるとすると、例

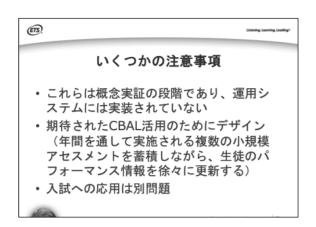
以上は割合に関しての例ですか、もし同じことを討論・議論に関してやるとすると、例 えば討論で賛成か反対かに情報を分類するのは得意だが、その主張を強めたり弱めたりす るエビデンスの評価が十分でないとか、そうした情報もラーニング・プログレッションズ の枠組の中で示すことができます。



以上に示したように、スコアレポートは、試験の結果に基づいて次に何をなすべきかに つながるものでなければなりません。この点で、私は 640 点だった、私は 700 点だった、という単なるスコアカードとは違っているのです。



これが最後のものですが、様々な分野におけるパフォーマンスレベルをどのようにして 捉えるか、管理者向けのレポーティングとしてまとめたものです。



レポーティングについて様々な検討をしていますが、先ほども言いましたように、これらはまだ実用段階ではなく概念実証の段階です。運用システムについても、このように実装されることを考えてはいます。しかし、重要な点は、ここで示したものは CBAL を活用する場合ということであり、そこで焦点を当てているのはアメリカの K-12 の教育です。毎年の試験をやるときに、入試と違って、実際に学習が十分進捗しているかどうかに焦点を当てています。そして、その結果をもとにどのような改善が必要かに焦点を当てているのです。

以上の試みは、期待される CBAL 活用の事例向けにデザインされており、そこでは年間を通してアセスメントを蓄積していきます。多くの観点から大変興味深く思えるのは、もし全ての試験を 1 か所で行う必要がなく、かつ試験が教育上有用な場合には、試験の機会が増えることで学習者のパフォーマンスの情報をより多く得ることができるという点です。しかし、これが入試の場合にはどういう意味合いを持つか、そこはかなり違ってきます。

●入試への含意



そこで最後に、入試との関係についてお話しします。



ning, Learning, Leading

文脈化の価値

- 試験への従来のアプローチでは、しばしばアセスメント・タスクの脱文脈化が図られる
- より多くの真正なタスクを活用する試みは、 測定として不十分と批判されることが多い
- CBAL研究では、現実的で真正なタスクと思われるアセスメントにおいても、望ましい測定の特性は維持できることが示唆されている

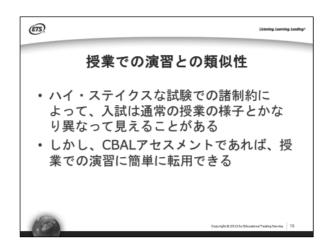
まず文脈化について取り上げます。従来の多くのペーパーテストでは、しばしば脱文脈 化がなされています。学校でやっていることや実際の生活と関係がないようなことが試験 に出るわけで、非常に抽象的になりがちです。その背景には、もっと試験を客観的にしよ う、もっと良い測定をしよう、などの様々な理由があります。しかし、CBALのこれまで の研究によれば、従来の測定方法を用いたままでも有用な文脈化が可能であるということ が示唆されています。

測定の二つの違ったアプローチの間に分断があることが問題です。パフォーマンス課題を行う場合,従来のパフォーマンス課題では多くの資料を与えてタスクに取り組ませます。そして 30 分,あるいは 1 時間,あるいは 4 時間ぐらい使って,小論文のような何らかの成果物を出させます。そしてそれを採点しますが,得られる情報はその得点だけなのです。その他は分からないのです。例えば 1 という全く駄目な出来ばえであっても,なぜうまく成果が作れなかったのか分かりません。

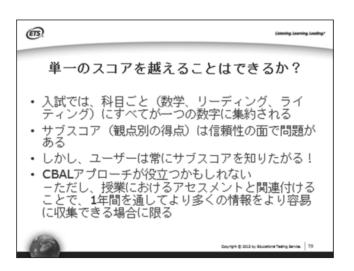
一方,従来の評価の仕方,例えば多肢選択式を使う場合,同じような時間をかけて,非常に詳細な得点情報が得られます。サブスコアの情報も得られるでしょう。例えば,ライティングよりもリーディングに問題がある,とか,他の単元よりも割合に関して出来が悪い,とか。けれども,そこまでで終わってしまいます。脱文脈化されているために,どうやって授業につなげたらいいのか明確ではありません。

CBAL アプローチの長所は、そうした構造的なアセスメントを、実生活でもありそうなパフォーマンス課題を用いて行うことができるという点です。その結果、一つの総括的な得点ではなく、例えばよい小論文を書くのに必要なクリティカルなタスクそれぞれの得点情報を得ることができるのです。

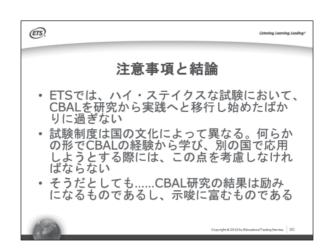
ただ、入試をする場合には、その目的が何かを考える必要があります。純粋に、ベストの学習者を選ぶことなのか、それとも学校が適切な指導をして、学習者が大学に向けて十分な準備をできるようにするのかということです。もし純粋に選抜だけを考えて学校への影響は考えないのであれば、CBAL アプローチはあまり関連がありません。しかし、入試で学校にも影響を与えたいということであれば、CBAL がやろうとしていることは皆さんにとって大変興味深いかもしれません。



次のポイントです。CBALのデザインは授業での演習のように見えるのですが、これは意図的です。ハイ・ステイクスの試験では、テストのやり方が限られてきます。CBALでは、他の革新的なアセスメントのようなことはやっていません。例えばゲームベースのアセスメントなどはクラスルームでは素晴らしいですが、入試で行うことは難しいものです。したがって、CBALのシナリオベースのアセスメントにおいては、妥協として、純粋なパフォーマンステスト、あるいはゲーム、シミュレーションを含めながら、同時に、多くの従来的な測定の方法も用いています。一方、文脈化によって CBAL で行っているような構造化を行えば、授業にうまくつながることが明らかになります。そのことが入試でも役立つかもしれません。



入試ではスコアの問題もあります。皆がサブスコアを知りたがりますが、同時に、1時間や3時間のテストでは信頼性の高いサブスコアが得られないということもよく知られています。しかし、CBALのように、カリキュラムと統合されたテストを蓄積する分散型のアプローチをすれば、より多くの情報によってサブスコアの報告をサポートすることができるでしょう。



注意事項もあります。まだ実際にこのシステムが運用に至っているわけではありません。 これらの考えは将来の成果につなげたいとは思いますが、まだまだ研究途上です。

また、今日お見せしたのはアメリカの文脈でやっていることです。日本では文化が違います。文化的な違いを考慮することが重要で、何に価値があり、何が適切な文脈やトピック、デザインなのかを考えなければいけません。しかし、そうだとしても、CBAL 研究の成果は非常に励みになるものであり示唆に富むものと思います。

どうもご清聴ありがとうございました。



●質疑応答

【山地】ありがとうございました。実は打ち合わせのときに、かなりスライドが多いから端折りながらどんどんやってくださいとお願いしたら、15分も早く終わってしまいました。Deane 先生にはご協力をありがとうございます。

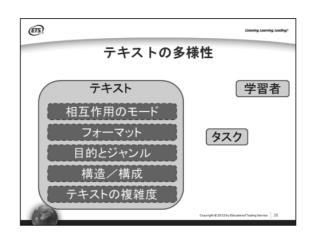
それではこの時間を使って、お手元の資料ではかなりスキップしたところもありますし、いろいろな片仮名の用語もありますので、分かりにくいことがあるかもしれませんから、確認をしていただきたいと思います。いかがでしょうか。はい、お願いします。

【西岡】京都大学の西岡です。大変興味深いお話,ありがとうございました。課題を作る際の,High-Level View, Varieties of Text, Individual Variation など,その辺は多分省略されたと思うので,そこをもう少し補足していただけたらと思ったのと,自由記述で3段落程度文章を書くという問題があったと思いますが,その採点でルーブリックが使われているということは分かったのですが,採点は担当している学校の先生方がなさっているのか,ETSでやっておられるのか,そのあたりを教えてください。

【Deane】幾つかご質問いただいたので、一つずつお答えしたいと思います。



キー・プラクティス(鍵となる活動)という話をいきなりしましたが、鍵となる活動という概念は、認知的・社会的な文脈を背景にしています。テキストとタスクをどう結び付けるかということなのですが、学習者が何ができて、何を知っているかということを把握するに当たって、どのようなタスクをさせるか、そしてタスクをさせた際に学習者について何を知り得るか、ということを考えなければいけません。



まず、テストを設計する上で考えなければいけないのは、どのようなテキストを 用意するかということです。そこには、相互作用のモードがあります。何か黙読し ている普通の状況なのか、あるいは音読をさせるかもしれない、あるいは対話型に したいかもしれない、ないしはマルチメディアを使用したいかもしれない。他の人 とのやりとりの方式がいろいろあるのです。テキスト、対話、あるいは他の言語形 式もあるかもしれません。スキルを評価するに当たって、いろいろな条件があり得 るということです。どのようなテキストを与えるかによって、何を測定するかも変 わってくると思います。

フォーマットも重要です。紙のテキストですと、それなりの特徴があります。構造化された文章で、最初から順次読んで、しかもまた読み直せるという点です。しかし、ページを繰って読み直すことしかできない、そういう意味ではある程度限界があります。デジタルでの情報提示になると、また違う可能性があります。あちこち飛んだりできるからです。

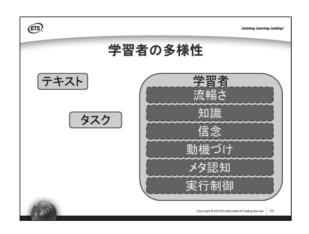
それから、目的とジャンルも重要です。情報型の文章、議論型のもの、それ以外のタスクなど違いがあるので、ジャンルによって目的も違ってきます。したがって、どのようなテキストに取り組ませるかを決めなければなりません。例えば大学入試であれば、恐らく情報テキストをきちんと読み込むことができるか、情報を抽出できる能力があるか、きちんとそれを検討できるか、などを試したいと思うのではないでしょうか。

また、提示されている議論が理解できているかどうかを評価したいかもしれないですし、他にもいろいろと見たいところがあるでしょう。したがって、教室の中で何を読ませるのか、そして読んだ後にそのテキストで何をさせるのか、ということが大事であるわけです。英語でも日本語でも、どんな国語であったとしても、どのようなテキストを使うかは重要です。

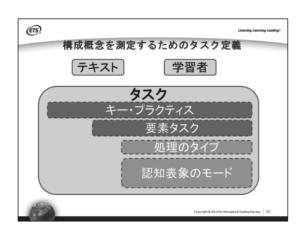
そのテキストの構造・構成も重要です。いろいろなテキストが、同じ目的であっても構成が変わる場合があるかもしれません。古典的な例として、例えばアングロ・アメリカンですと、メインポイントがあって、サブポイントがあって、そしてそれ

を裏付ける理由があって、というように構造化されています。フランス語だったら、 テーゼがあって、アンチテーゼがあって、そしてジンテーゼ、という形になり、構成・構造が違ってくるのです。言語によっても違ってくるということです。

また、テキストの複雑さも多様です。ボキャブラリー(語彙)の面だけでなく、 表現のスタイルも複雑になっている場合があります。それらも検討しなければいけ ないということです。

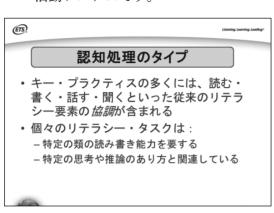


それから、学習者にもばらつきがあります。表現の流暢さ、それからトピックに 関する知識、また信念はどういうものか。信念次第でパフォーマンスも違ってきま す。動機付け、やる気があるのかどうか。メタ認知や実行制御はどう働いているか。 複雑なタスクを、やっていることの意味が分からないままやっているということに なっていないか。テストを設計するに当たって、これらを測定しコントロールでき ているかどうかの視点が必要です。



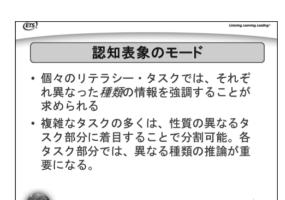
テキストと学習者が明確になったら、では実際に何をやらせるか、タスクをどうするかということになります。どのような活動が重要か。議論、討論、解釈、それからストーリーテリング、どれができることが重要なのかということです。大学入試の場合は、ストーリーテリングは必ずしも重要でないかもしれませんが、これは、価値判断としてそれぞれの機関で決定しなければいけません。それに合わせたタスクが必要なのです。キー・プラクティスというのは、これらを自然な形でつなげる

活動システムです。



構成概念	認知処理
聞く、読む	解釈
話す、書く	表現
戦略的・批判的思考	熟考

さらに進めますと、キー・プラクティスを横断的に見た場合、どのような認知処理が必要なのか。聴解し、読解し、解釈する、それから、会話、論述をして表現する。 これらは自動化しているところもありますが、より戦略的なクリティカル・シンキングが熟考の際には必要になってくるということで、どんなタスクであったとしても、解釈、表現、熟考をどう要求しているかが問われます。



認知表象の5つのモード				
特徴				
人々とその伝達目的				
世界のメンタルモデル				
テキストや言説の構造				
言語によって伝達される構造と意味				
言語やテキストが物理的な形に表される、 正書法などのモード				

それからもう一つ、認知的な表象のモードをどうするかという問題もあります。 CBAL では五つあると考えています。まず、タスクの社会的なレベルがあります。 人とのコミュニケーションの目的です。それから、概念的なレベル、すなわち世界のメンタルモデルがありますし、またテキストの構造(ディスクール)のレベル、さらに、使っている言葉のレベル、そして活字体にどういう形で表現されているのか、正書法、音声など。これらは複雑なタスクの中では様々に綯(な)い交ぜになっています。

典型的な論述のルーブリックの場合、オーディエンスに向けてのスタイルはといえば、社会的なレベルです。よく練られた議論ということであれば、概念レベルに当てはまります。明確なテーゼを導入する、それはテキストの構造レベルに関係します。言葉が明確で、正確で、言葉数が多過ぎない、混乱しない使い方になっているか、それが言葉のレベル。そして書き方が正しいかどうか、それが活字のレベルということになります。

複雑なパフォーマンス課題というのは、同時に幾つかのこうした表象をきちんと

組み合わせて使えるかどうかを見ているのです。自動的にできる人はできるのですが、より高次のスキルを獲得するに当たっては、タスクの複雑さを検討することが重要だということです。それがタスクの分析につながるのです。

一つの例示として、講演の中では討論の例を挙げました。討論の活動における各段階が、認知表象のどのモードに対応しているのかということです。例えば、そこでの問題に関して、誰がどういうことに着目して賛否どちらを主張しているのか、これは社会的なレベルですね。トピックについてどれだけの知識があるのか、あるいはないのか、どの情報が重要とされているのか、これらは概念的なレベルです。また、どのスタンスを取るのか、そのスタンスをいかに抗弁するのか、これらも概念的レベルです。小論文を書く段階では、どういう構成にするのか、どう表現するのか、などを考えて構成や言語をきちんと調整することになるので、これはディスクール、テキストの構造レベルです。活動のどのフェーズにいるのかによって、異なるスキルや表象が求められるということになります。以上が、スキップしたスライドの部分のお答えです。

次に、誰がスコアを付けているかということです。結局質問として、どうやってスコアを付けるのかということになると思うのですが、その状況の制約次第で変わってくると思います。例えば入学試験だと、大学の教授を採点者とするのであれば、それはとても適切だと思います。特段、利益相反もないと思います。教授たちには、不正確に採点をするインセンティブは働かないからです。できるだけ優秀な人たちを入れたいと思っています。他方、学校でテストをして教師に採点させると、実は利益相反が生まれることがあるのです。

アメリカでは、きちんと採点者が訓練されているかどうか、そして利益相反のない人に採点させているかどうかに気を遣います。州の中には、自分が教えている生徒の採点はしない、あるいは他の学校の生徒の採点をランダムに行うという州もあります。

CBAL では、そのような負担を自動採点でいかに軽減できるかを考えています。 人による採点のいいところを残しつつ、ということです。人の採点であれば、本当 におかしいと思ったらそこで立ち止まって、見直す必要があるとすぐに判断するこ とができます。人中心の採点であれば、言語や文法、全体的な構成などをコンピュ ータで自動評価させても結果は悪くない、ということが研究でも出ています。機械 が検出できないものは、人に見させるということです。

伝統的なルーブリックを当てはめるのは、幾つかのことを同時に評価しなければいけませんので、これは難しい場合があるのです。評価を正確にするためには、やはり人を訓練し、同じような形で重み付けができるように、そろえなければいけません。

ETS では、1950 年代に人による採点を検討していたのですが、評価者によって やり方が全く異なり、結果に大きなばらつきがありました。例えば、文法的な正確 さにばかり目がいく人、スタイルばかりを見る人、中身だけを重視する人、ちゃん と 5 パラグラフの構成になっているかということばかり気にする人など、様々でした。しかし、採点でそんなにばらつきがあっては困るのです。

人による採点の方法は、いかに信頼性の高いものにするかが課題でした。そのためには、厳格な訓練をし、ほぼ同じ方法で同じ基準を当てはめる一貫性が成り立つよう、プロセスもかなり細かく決めておかなければなりません。そこから乖離(かいり)しない、それないようにしなければいけないのです。

自動採点のメリットは、同じ小論文を毎回一律に同じように評価できるということです。しかし、何か変なもの、例えば全く意味不明な言葉が並んでいて、一見まともに見えるがよく読むと実はそうではないという場合、自動評価ではうまくいかないことがあります。したがって、GREの場合は、人と自動評価の両方を組み合わせるという方法です。一方に人間がいて、もう一方にマシン・スコアラーがいて、意見が一致したらそれでオーケー、もし意見が食い違ったらもう一人加えて、二人プラス機械で評価するということにしています。人だけを評価者にすると相当コストがかかりますから、この方法でコストを下げることができます。

もし教師を評価者にする場合には、どれだけの負担を強いることになるかの考慮が必要です。教師に採点してもらうのは、専門職の能力開発という点では意味があるかもしれませんが、相当多数の評価をしなければいけないとなると教師側も疲弊します。採点の体制をどうするのかということは、よくよく考えておかなければいけないのです。



質の高い「人による採点」を生かしつつ、教育目標上、重要なものを犠牲にしないようにすることが大切です。目標は何で、資源がどれだけあるのか、人的財政的コストをどこまでかけて目標達成に向かうのか、ということを慎重に検討しなければならないのです。

【山地】ありがとうございます。西岡先生から、どなたが実際に評価なさるのかというご質問もあったかと思うのですが、それも含めて、現在 CBAL はパイロット段階でやっているのか、それとも K-12 で実際に使っているのか、そのあたりもご紹介くださいませんか。

【Deane】CBALは、現在は実際に使われているものではなく、ETSのイニシアチブという段階です。実践上の課題を検討しており、例えば一つの設問があったとき、人による採点と自動化された採点を、あまりコストをかけ過ぎずに組み合わせるにはどうしたらよいか、といったことなどを実験しています。また、採点者として教師にも入ってもらって、同じ質の採点ができているかどうか、つまり、様々な学校から選んだ教師による採点と、ETSのプロフェッショナルによる採点とを比べることもしています。実際に正確な採点を行う教師もいれば、そうでない教師もいます。そのようなところを洗い出していくのも一つの目的なのです。



CBALではいろいろなやり方を試みています。教師による採点、ETSのプロフェッショナルによる採点、自動化された採点、そしてそれらを組み合わせた採点。例えば小論文を評価する場合、どこで自動採点を使うべきなのか、どこは人間による採点にした方がよいのか、ということでいろいろとトレードオフの関係があるわけですので、最適の組合せを探ります。

最も現場に近いプログラムとしては、ETSは、通常は採点者を使ったプログラムをもっていて、彼らはいろいろな修士号を持っている人であったり、他の資格を持った教師であったりしますが、資格を持っているからといって採点がすぐにできるということではないのです。つまり、トレーニングが必要です。そして、正しく採点できることが担保されなければいけないのです。それが証明されて初めて、実際に採点をすることができます。

また、目的をよく考えた上での設計が必要です。例えば入学試験であったならば、 採点はかなり精度の高いものでなくてはいけません。しかし、例えば教師に対して フィードバックをする程度であれば、また個々の生徒には適用されないものである ならば、教師同士で採点をするのがよいかもしれません。ケースバイケースでいろ いろトレードオフがあるわけです。私たちは、ETS やその他の評価を行っている団 体などから、多くの意見をもらいました。自動採点についても有意義な意見を得て いるので、どういうときに、どんな目的の場合に自動採点を使った方がよいのかが 明らかになりつつあります。

【山地】ありがとうございます。この自動採点と人による採点については、この後の石岡 先生の、報告3のところでも扱われると思います。

再びフロアの方で、ここは確認しておきたい、あるいは聞いておきたいということはございませんか。はい、お願いします。

- 【大塚】先生、ありがとうございます。真正な文脈を入れるという問題例がたくさん示されましたが、これは私も教育における真正性の重要性を学んできておりますし、これから重要になっていくと思っております。ただ、ETSでも、SATなど共通テストをやられていますが、私どもも共通テストを担当しておりまして、文脈を組み込んだ問題はそれへの馴染みの個人差が生じてしまうということで慎重さが求められるということを経験しております。かつてはアメリカでもカルチャーフリーテストと言われるような文化差や背景に依存しないテストの開発などが随分やられてきたと思いますが、共通テストなどの場合に文脈を導入するということが、受験生の公平性をある意味で乱す要因になるのではないかという可能性について、先生自身はどのようなお考えをお持ちですか。
- 【Deane】この問題は本質的に、まさにより真正な試験を行うことに関わってきます。より真正な試験をやるには、やはり文脈がなければいけません。文脈には文化的な要

素が入ってきます。それについて懸念しなければいけません。問題は、試験から文化の要素を省くということは、人々の背景知識を除外することになるということです。

たとえかなり一般的な、文化的にそれほど固有ではないトピックを選ぶことができたとしても、最終的には一部の人たちがより馴染んでいて他の人より知識があるということがあるわけです。それが誰かは分かりません。それをコントロールはできないのです。これまでSATやその他の試験で、試験があまりにも文化固有になることを避ける手立てを取ってきましたが、文化的なバイアスはなくすことができないので、どんなことをやっても批判を受けます。

私たちが分かったことは、試験全体としては、トピックの違いはそれほど大きな意味を持っていないということです。総和としては相対的なパフォーマンスが現れてきます。様々に異なる設問があることで、各学習者の達成度の違いが反映されます。

例えば、三つのトピックを考えることができます。一つ目は、12歳に対する広告を禁止すべきかどうか。二つ目は、生徒とソーシャルネットワーキングについて、例えばスマホ等を学校から排除するかどうか。三つ目は、生徒の耳の不自由を扱うようなトピックです。トピックを選ぶ際、三つ目のものは様々なバイアスがかかってきますから、慎重を期する必要があります。

意味のある文脈という点では、必ずしも試験によって重要な意味のある情報が提供されるとは限りません。感情的になったり前提知識がなかったりすれば、試験への解答の仕方が変わってきます。それは深刻な問題ですが、全体の設問構成を幅広いものに工夫することによって、そのような問題は緩和されます。

【山地】ありがとうございます。他に何かご質問はありますでしょうか。

一つ確認ですが、先ほど追加で示された内容は、我々にとっては例の思考力・判断力・表現力を評価する問題を作るときに、どういう認知要素、どういう認知機能を前提にして問題を作っていくか、その枠組として参考になると思うのですが、CBALでは実際の作問体制はどうなっているのでしょうか。ラーニング・プログレッションズに沿って、様々な要素の組合せで問題を大量に作らなければいけないということになると思うのですが、このあたりは一体どうなさっているのですか。

【Deane】前倒しの投資をします。思考力を反映する豊かなアセスメントをするのであれば、最初に相当練って作らなければいけません。私たちのものも含めて、文献は豊富にあります。ウェブサイトでは、私たちの報告の最新版が出ています。いろいろなリソースや文献があり、どのようなタスクがどのような場面で適切かを教えてくれます。既に成果として挙がっているものを活用することができます。

もう一つ重要なのは、CBALの理念の大きな部分として、進化するモデルを作っていくということです。タスクを並べる際、スキルをテストするに当たってどうい

うモデルが適切かということを考えなければいけません。例えば議論・討論のスキルではいろいろなモデルがあります。既にあるモデルをベースにしてタスクを作っているのです。これらを土台にして、より明示的な内容にして、より多くの側面を従来以上にテストしようと考えて細分化しています。

- 【山地】今日示されたような具体的な問題例はどなたが作っておられるのか、つまりどういう体制でそれぞれの問題を作っておられるのかという点はいかがでしょうか。
- 【Deane】チームとして、経験のあるアセスメント開発者や研究者、またリーディング、 ライティング、クリティカル・シンキングなどの専門家がいます。学校と連携し、 教師たちと一緒に共同開発をしました。プロセスとしては、まずアセスメントの考 え方をまとめて、研究者やアセスメントの専門家と連携し、そして教師にも協力を 仰いだということです。モデルの開発をそういう形で行いました。

それで、シナリオベースのアセスメントでいろいろなモデルができることになったわけです。そうしたモデルをアセスメントの開発者に渡して、スペックに沿った形でテスト問題を作ってもらいます。

【山地】ありがとうございます。学校の先生方、それからリサーチャー、専門家がチーム を組んでやっておられるわけですね。

最後にもう一つくらい質問の時間があると思いますが、もし何かありましたら。 はい、どうぞ。

- 【大塚】もう一つ質問したい点は、「ラーニング・プログレッションズ」ということについてです。その言葉は、私にとっては新しい言葉として聞いているのですが、これは学習者にもいろいろな個人差があり、いろいろなレベルがあるということだと認識しております。そして、そのラーニング・プログレッションズの段階によって、同じトピックであっても、問い方などが違ってくるということをお話しくださったと思います。その際に、一つのクラスであっても、小さなクラスであっても、生徒たちの間に個人差があって、プログレッションズのレベルが違ってくると思うのですが、その子供に適した問い方を、コンピュータを利用することで、例えば、アダプティブ・テスティング(適応型テスト)という方法などもあるわけですが、それと似たような方式で子供のレベルに合った問い方が自動的に提供されるようなシステムをお考えなのでしょうか。その辺の動向などを教えていただければと思います。
- 【Deane】ラーニング・プログレッションズが、学習の理論の中でどこに位置付くかからお話ししたいと思います。これはピアジェの理論やそうした発達モデルとは違うものです。それらは、人はどんな場合でもある一定の段階順序を経て発達していくという考え方で、学習の際も、一つのスキルを自分のものにしていくためには通常一

定のステップを経ていくと考えます。

それに対してラーニング・プログレッションズというのは、各段階でどんな介入をするのが適切であろうかということを理解するためのものです。例えばある議論をしていて、賛成・反対のそれぞれの主張を強めるものや弱めるものが何であるか理解していなければ、議論はできないわけです。しかし、特定の意見が果たして賛成なのか反対なのか分からない、というようなレベルもあるかと思います。その場合は介入が必要になります。この学習者の場合は読解に問題がある、あるいは思考の仕方に問題がある、など、ラーニング・プログレッションズを踏まえれば分かってくることがあるのです。

このようなシナリオベースの評価では、あるレベルの学習者がいるとき、そのレベルであればどういうフォローアップが適切なのかということが分かります。これはアダプティブ・テスティングとも違います。もちろんそれも可能だと思いますが、私たちはその点についてはフォーカスしていません。アダプティブ・テスティングは、学習者の答え方をもとに、それがラーニング・プログレッションズのどこにあるかを特定し、その次の設問をそれに基づいて作っていくというやり方です。

アメリカでは指導の個人化(パーソナライゼーション)ということが教育制度の中で重視されているため、テストから得られる情報に基づいて、それぞれの学習者が次のレベルに行くためにはどんなヘルプをしたらよいのか、ということを読み取ることが求められます。ですから、これは入試と言うより、日々の授業・指導の中での話です。テストからの情報を用いて、教師がどうやったらより良く指導できるかを考えるのです。多くの学習者が、例えば小論文が全く書けないわけではないのだけれど、X、Y、Zというような問題を抱えているという場合、学校における活動として、それらを解決するためにはどうしたらいいのかということを明らかにするのです。

【山地】ありがとうございました。では、時間になりましたので、ここで 10 分間の休憩を入れさせていただきます。Deane 先生にお礼の拍手をお願いします。どうもありがとうございました。

(*64~103ページに,英文スライドを掲載)

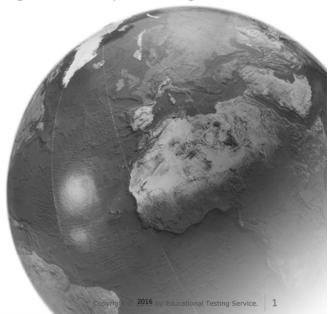


Cognitively Based Assessment

An assessment system that measures and develops competencies in thinking, judging, and expressing

Paul Deane **Educational Testing Service**

Presentation for NCUEE





Listenina, Learnina, Leadina,

What is CBAL?

· CBAL =

Cognitively-**B**ased Assessments of, for and as Learning

 A major ETS research initiative focusing on "next generation" assessment





Some Key Publications

- The Changing Nature of Educational Assessment R. E. Bennett, (2015) Review of Research in Education, Vol. 39, No.1, pp. 370-407
- Cognitively Based Assessment of, for, and as Learning (CBAL): A Preliminary Theory of Action for Summative and **Formative Assessment**

R. E. Bennett (2010), Measurement: Interdisciplinary Research & Perspectives, Vol. 8, No. 2-3, pp. 70-91

Transforming K-12 Assessment: Integrating Accountability Testing, Formative Assessment and Professional Support R. E. Bennett & D. H. Gitomer (2008), ETS Research Memorandum No. RM-08-13



Copyright © 2013 by Educational Testing Service. 3



Listenina, Learnina, Leadina,

CBAL Goals

- Cognitively based: grounded thoroughly in the modern cognitive and learning sciences
- Assessment of learning: valid, reliable measurement of what students know and can do
- Assessment for learning: actionable, timely information that can guide instruction
- Assessment as learning: tasks that are worthwhile educational experiences in their own right





CBAL Theory of Action

CBAL Components

Competency Models/ **Learning Progressions**

Distributed Summative Assessments

Multiple Opportunities for Formative Assessment

Systematic Professional Support

Intended Intermediate Effects

- Increased teacher content knowledge
- 2. Increased pedagogical knowledge and assessment skill
- 3. Greater instructional focus on integrated performance and higher-order skills
- Roiutine use of formative 4. assessment to make appropriate instructional adjustments
- 5. Improved student engagement in learning and assessment

Hypothesized Action Mechanisms

Ultimate Effects

- 1. Improved student learning with respect to content standards
- 2. More meaningful information for policymakers regarding the effectiveness of education, leading to decisions that facilitate learning





Listenina, Learnina, Leadina,

Implications for Admissions Testing

- A high-stakes admissions assessment has powerful washback on instruction.
- CBAL philosophy: Build the intended instructional impact into the test design from the start.
- · If the result is an assessment that looks more like instruction - that is a feature, not a bug.





CBAL Emphases (how we achieve our goals)

- Model authentic tasks
- Combine these tasks into meaningful scenarios that reflect key practices in the domain.
- Link task performance to learning progressions
- Emphasize designs that model appropriate strategies with which to complete complex performance tasks.





Listenina, Learnina, Leadina,

The CBAL English Language Arts (ELA) Framework

- · Different CBAL teams have focused on science, mathematics, and the English Language Arts.
- My contribution has primarily been to the English Language Arts framework
- The CBAL English Language Arts framework provides a concrete illustration of the larger CBAL approach



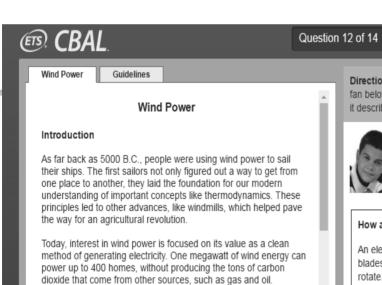


Modeling Authentic Literacy Tasks



Copyright © 2013 by Educational Testing Service.

0



How an Electric Fan Works

An electric fan is run by an electric motor that powers a set of fan blades attached to a rotating shaft. As the motor spins, the fan blades rotate. The result of this rotation is a moving airstream.

Directions: Read Dexter's comment and the description of an electric fan below. Then classify each statement in the table according to whether

A wind turbine is just a big electric fan. I looked up electric

fans on the Internet and came across the following

description of how an electric fan works.

it describes a wind turbine, an electric fan, or both.

Statement Wind Turbine Both Has blades that spin Uses a moving airstream to generate electricity Uses electricity to generate a moving airstream Has a rotating shaft

How Wind Power Works

Since Earth's surface is made up of very different types of land and water, it absorbs the Sun's heat at different rates. Changes in temperature cause changes in air pressure. In cold air, the molecules are more closely packed together than in warm air, so cold air is more dense than warm air.

During the day, the air above the land heats up more quickly than the air over the water. Since it is less dense, the warm air over the land expands and rises, and heavier, cooler air rushes in to take its place, creating winds. When the Sun goes down, the winds are reversed because the air cools more rapidly over land than over water.

People use this wind flow for sailing boats, flying kites, and other popular outdoor activities. But the most important use is generating

68











©James Wilson/iStockphoto # 7293761

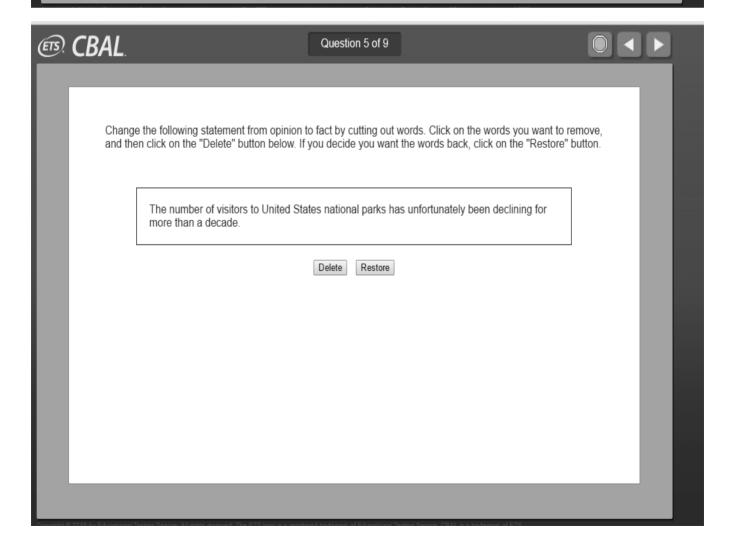
Summarizing "Krypton"

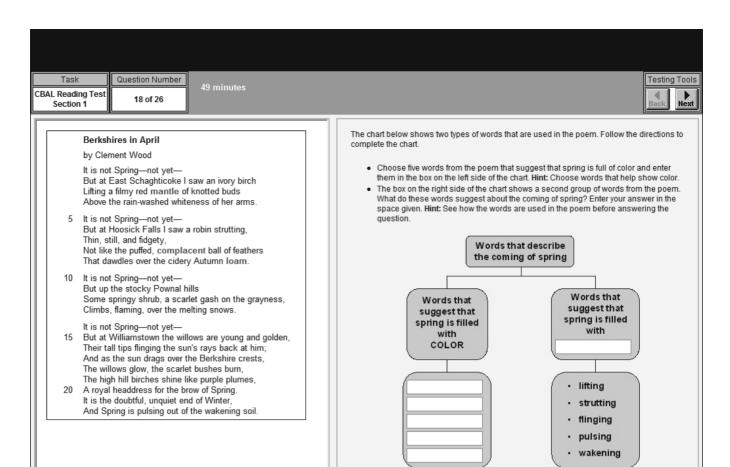
While doing research for a science report on the gas krypton, you find the following paragraph and decide to summarize it in one or two sentences. Highlight the parts of the paragraph that are most important to include in your summary. Click-drag your mouse to highlight text. Click "Start Over" if you change your mind.

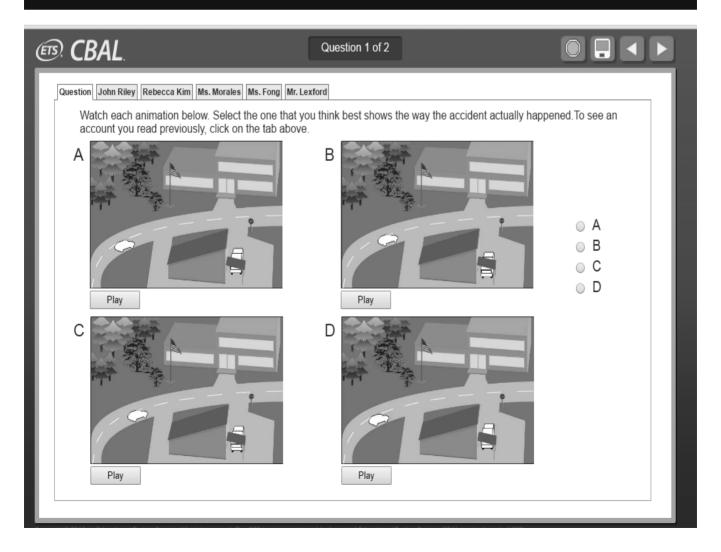
The famous cartoon character Superman has many super powers. Everybody knows that. He's the Man of Steel, has X-ray vision and hearing so good he can tune in on one voice in a crowded city. And, of course: He's faster than a speeding bullet! More powerful than a locomotive! Able to leap tall buildings in a single bound! But there's one substance that weakens Superman: kryptonite. If exposed to kryptonite, Superman experiences pain and loses his super powers. He can even die. Kryptonite, of course, is purely fictional. Despite the similarity in names, krypton (element 36) is a very different matter. Kryptonite is make-believe: a green glowing rock that can cause great harm — to Superman, anyway. Krypton is real: an inert gas that has no effect on anything.

Start Over

Adapted from "Krypton (revised)." Chemical Elements: From Carbon to Krypton. Ed. Lawrence W. Baker and David E. Newton. Vol. 2. Detroit: UXL, 2006.







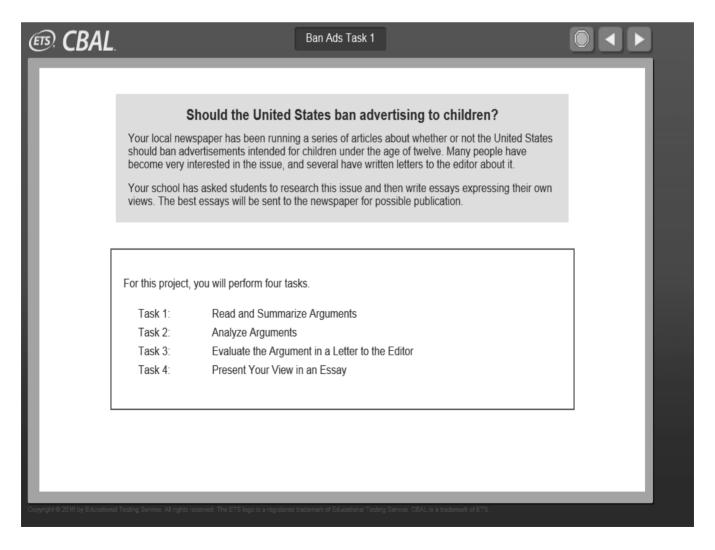


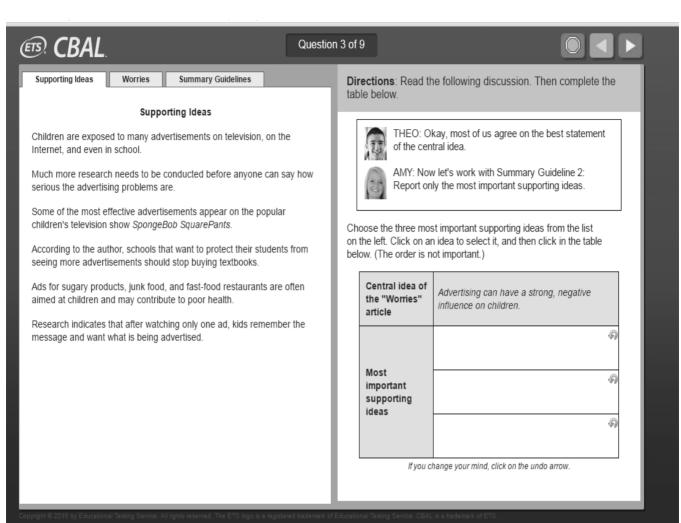
Meaningful Scenarios

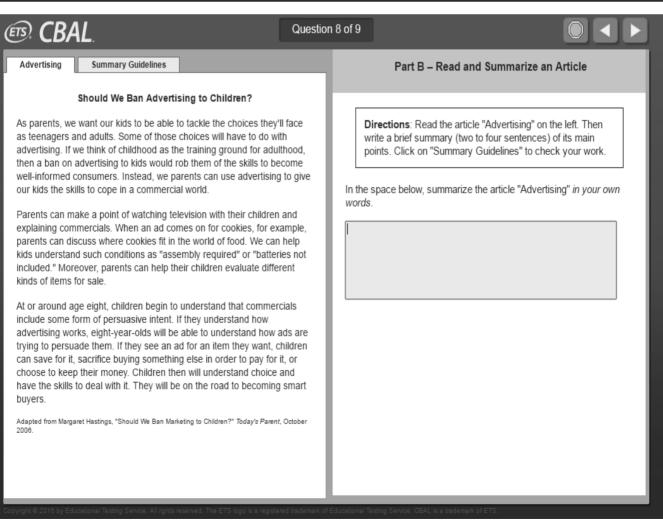


Copyright © 2013 by Educational Testing Service.

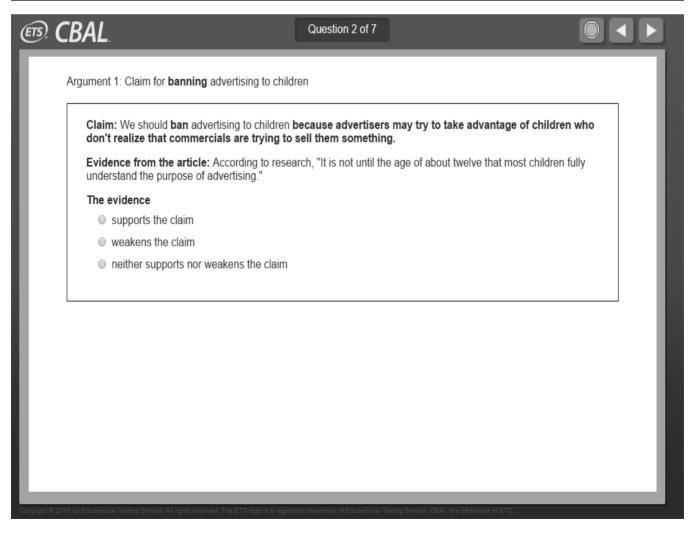
10

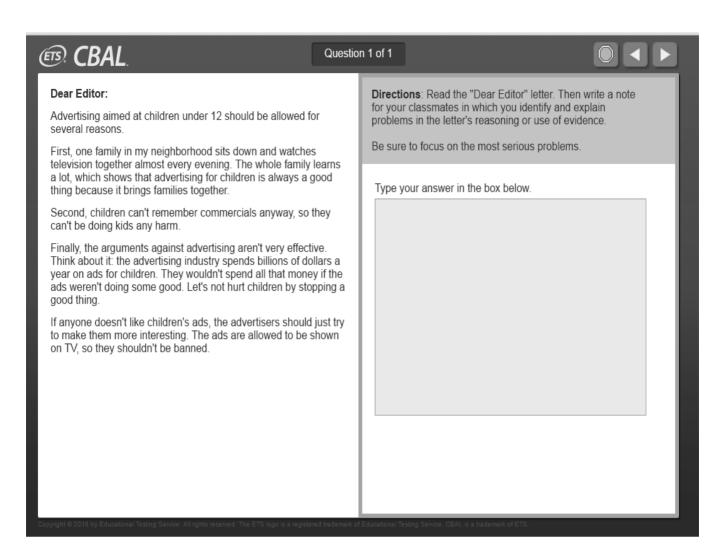


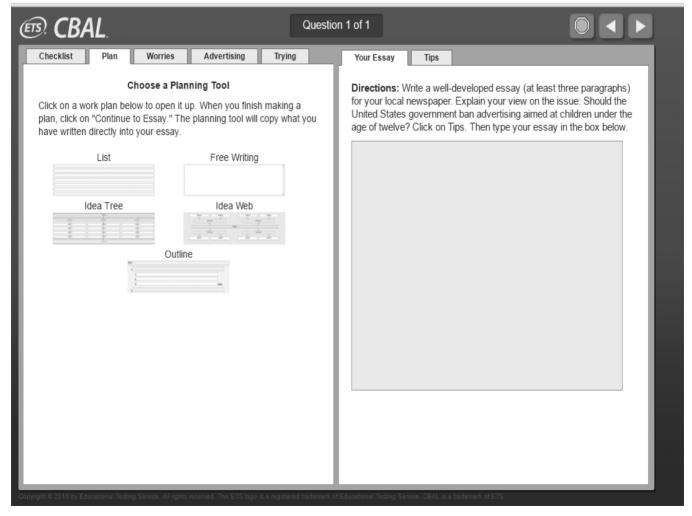




ES. CBAL.	Question 1 of 7	
STATEMENTS		nts: Ban or Allow?
Commercials are usually easy to remember and hard to forget, so an advertisement can get a lot of people to buy something, even if it is not good for them.	Directions: Read each statement and decide whether it children. Click on the statement to select it. Then click in issue the statement supports. BAN	gives a reason to ban or allow ads intended for a box under Ban or Allow to indicate which side of the ALLOW
Most healthy adults watched a lot of commercials when they were young.		
Parents who are well-informed about the effects of advertising and are concerned about protecting their children rarely worry about the advertisements their children watch.		
Junk food advertisers like advertising to children because they know that children are very easy to influence and that the habits people form as children will often last the rest of their lives.		
When a child sees an advertisement and keeps asking the parents to buy an advertised product, the parents usually give in, no matter what the child is asking for.		
Most people ignore most of the advertisements they see.		
Children may not understand that advertisements are trying to trick them into buying something.		
If an advertisement makes a child want a product, parents can use that as an opportunity to teach the child how to make smart choices.		
Parents don't usually take the time to check what their children are watching on TV, so they don't know what advertisements their children are seeing.		
Some of the people who argue the hardest in favor of ads for children earn big profits from the advertising industry, so you can't trust what they say.		









What CBAL Looks like in Practice

https://www.ets.org/s/research/videos /18879 cbal.html





Listening. Learning. Leading.®

Domain Analysis: Key Practices and Learning Progressions





Some Key Recent Publications

- From Cognitive-Domain Theory to Assessment Practice. Randv E. Bennett, Paul Deane, & Peter W. van Rijn. (2016). Educational Psychologist 51, 1, 1-26. DOI: 10.1080/00461520.2016.1141683.
- **Key Practices in the English Language Arts (ELA): Linking** Learning Theory, Assessment, and Instruction. P. Deane, J. Sabatini, G. Feng, J. Sparks, Y. Song, M. Fowles, T. O'Reilly, K. Jueds, R. Krovetz, & C. Foley (2015), ETS Research Report No. RR-15-17
- A Case Study in Principled Assessment Design: Designing Assessments to Measure and Support the Development of Argumentative Reading and Writing Skills. P. Deane & Y. Song (2014), Psicología Educativa, Vol. 20, No. 2, pp. 99-108
- Broadening the Scope of Reading Comprehension Using Scenariobased Assessments: Preliminary Findings and Challenges. J. P. Sabatini, T. O'Reilly, L. Halderman & K. Bruce (2014), L'Année psychologique, Vol.114 , Issue 04, pp. 693-723
- The Case for Scenario-Based Assessments of Reading Competency K. M. Sheehan & T. O'Reilly (2012), Chapter in J. Sabatini, E. Albro, & T. O'Reilly (Editors), Reaching an Understanding: Innovation in How We View Reading Assessment, R&L Education.





Listenina, Learnina, Leadina,

The CBAL ELA Competency Model

- Identifies constructs we wish either to measure or to control
- Functions as the domain model from an ECD perspective, and identifies likely student model variables





A High-Level View

Text

Student

Task





Listening. Learning. Leading.®

Varieties of Text

Text

Student

Mode of Interaction

Format

Purpose and Genre

Structure / Organization

Textual Complexity

Task



Individual Variation

Text

Task

Student Fluency

Knowledge

Beliefs

Motivation

Metacognition

Executive Control





Listening. Learning. Leading.®

Defining Tasks to Measure the Construct

Text

Student

Task

Key Practices

Component Tasks

Types of Processing

Modes of Cognitive Representation





Key Practices = Activity Systems -

Coordinating skills to achieve meaningful goals

- Concept derives from Vygotskian frameworks for social analysis
- An activity system is a collection of activities in which people work together to achieve social goals
- Each activity system recruits many different types of skills, drawing on different cognitive processes and modes of representation



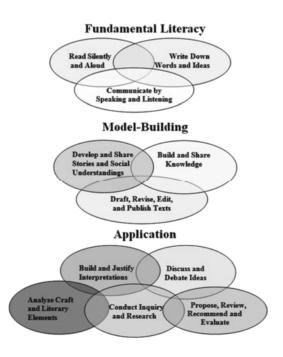
Copyright $\ensuremath{\text{@}}$ 2013 by Educational Testing Service.

21



Listening. Learning. Leading.®

Key Practices in the English Language Arts



From £15 Research Report Series

Volume 2015, ISS 2, pages 1-29, 17 JUN 2015 DOI: 10.1002/ets2.12063

http://o.kina.series.org/doi/10.1002/ets2.12063/full#ets212063-fig-0001



Types of Cognitive Processing

- Most of the key practices involve coordination of traditional literacy constructs such as reading, writing, speaking and listening
- Individual literacy tasks
 - require specific kinds of reading and writing
 - are connected to specific forms of thinking and reasoning



Copyright © 2013 by Educational Testing Service.

- -



Listening. Learning. Leading.

Three Types of Cognitive Processing

Construct	Cognitive Process
Listening, Reading	Interpretation
Speaking, Writing	Expression
Strategic & Critical Thinking	Deliberation

Each key practice requires different specific forms of interpretation, expression and deliberation



34

Copyright © 2013 by Educational Testing Service.

Confidential and Proprietary, Copyright © 2009 Educational Testing Service. All rights



Modes of Cognitive Representation

- Specific literacy tasks often require an emphasis on different kinds of information
- Many complex tasks can be broken down by focusing different parts of the task where different kinds of reasoning are critical



Copyright © 2013 by Educational Testing Service.

0 =



Listening. Learning. Leading.*

Five Modes of Cognitive Representation

Level	Description	
Social	People and their communicative purposes	
Conceptual	Mental models of the world	
Discourse (Text Structure)	The structure of texts and discourses	
Verbal	The structure and meaning communicated by language	
Print	Orthography and other modes by which language and text are embodied in physical form	

36

Copyright © 2013 by Educational Testing Service.

Confidential and Proprietary, Copyright © 2009 Educational Testing Service, All rights

reserved



Task Analysis

- Each key practice has critical component tasks
- Each of these tasks can be analyzed to identify the specific types of processing and the kinds of cognitive representations it requires
- This analysis helps us identify the knowledge, skills and abilities that matter for each of the ten key practices





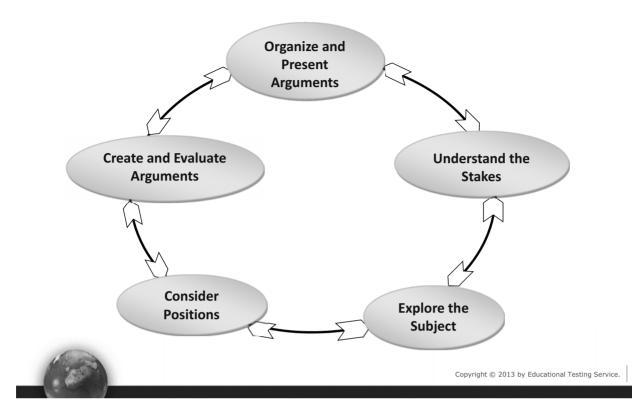
Listenina, Learnina, Leadina,

An Example: **Analyzing Task Requirements** for the Key Practice, 'Participate in **Argumentation and Debate'**

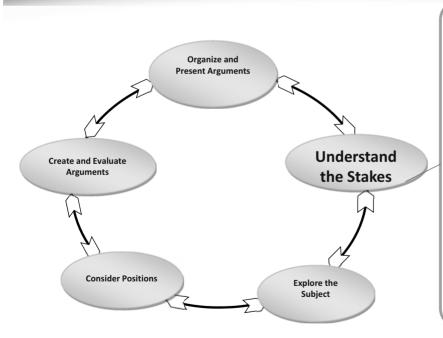




Argumentation Subtasks



Listening. Learning. Leading.*



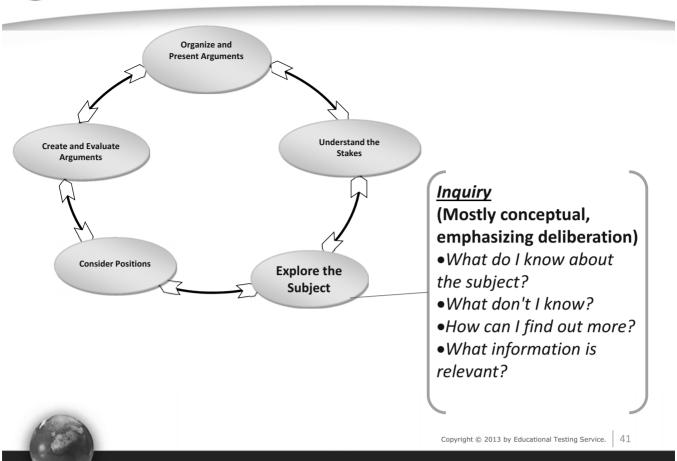
Appeal-Building (Social, mostly

interpretation and deliberation)

- •What do people who are interested in this issue care about?
- •Whose opinions about this issue matter?
- •Who am I trying to convince? How will I convince them?
- •Who are others trying to convince? How might they convince them?





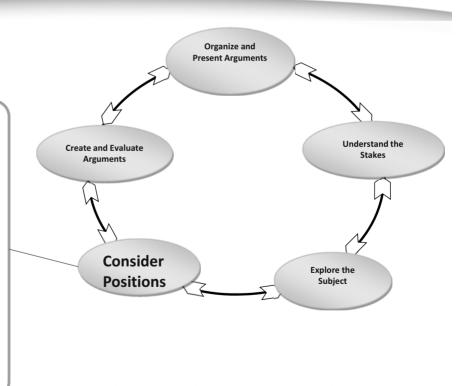




Listening. Learning. Leading.®

Taking a Position (Mostly conceptual, emphasizing deliberation)

- •What positions are reasonable?
- •Are they all clear and defensible?
- •What position should I take?
- •How should I focus and limit my position?

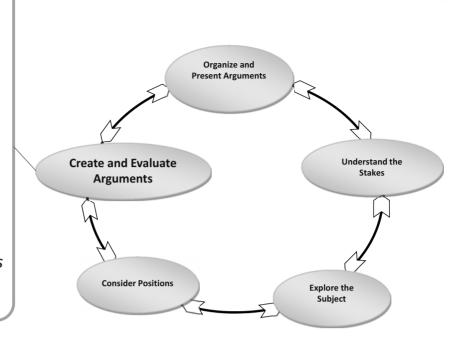




Reasons and Evidence

(Conceptual, equal emphasis on interpretation, expression, and deliberation)

- •What reasons can I use to support my conclusions?
- •Do I have enough evidence to support each reason?
- •What counterarguments do I need to address?





Copyright © 2013 by Educational Testing Service.

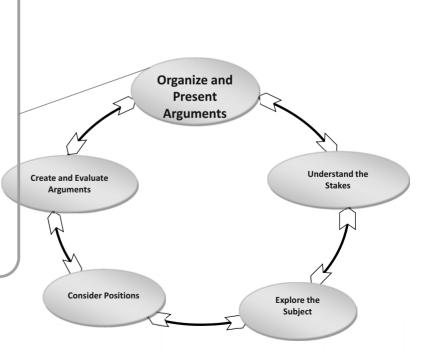


Listening. Learning. Leading.®

Framing a Case

(Discourse, emphasis on expression, though interpretation and deliberation are also important)

- •How should I present my arguments?
- •What structure is most effective and logical?
- •How will people with different perspectives present their arguments?







Learning Progressions and Assessment Design





Listenina, Learnina, Leadina,

Learning Progressions as a Strategy for Assessment Design

- · We are building a developmental assessment
- Our goal is to inform instruction, not simply place students on a scale
- So our basic strategy is to define "learning progressions" that function both to describe learning goals and to define how we will collect evidence for specific knowledge, skills and abilities.





Learning Progressions

a description of qualitative change in a student's level of sophistication for a key concept, process, strategy, practice, or habit of mind. Each progression is presumed to be modal, i.e., to hold for most, but not all, students. Finally, it is provisional, subject to empirical verification and theoretical challenge."





Listenina, Learnina, Leadina,

Illustration: Participating in Discussion and Debate (Argumentation)

- Four learning progressions specific to argument:
 - Appeal-building (social)
 - Taking a Position (conceptual)
 - Reasons and Evidence (conceptual)
 - Framing a Case (discourse)
- Each of these is mapped to detailed evidence statements targeting levels of student performance





High Level Overview – Reasons and Evidence 1. Understands the idea that positions may need to be supported **Preliminary** with reasons that will be convincing to the audience 2. Recognizes, generates and elaborates on reasons in writing, with **Foundational** some awareness of the need for evidence, and uses one's own argument to counter others' argument in an engaging, familiar context 3. Understands use of evidence and clearly grasps the need to **Basic** provide evidence and reasons that are directly relevant to and support the main point and which are logically sound 4. Understands the role of critique and rebuttal and is able to reason Intermediate about and respond to counterevidence and critical questions 5. Builds systematic mental models of entire debates, and use that **Advanced** model to frame one's own attempts to build knowledge.



Æ	TC)	
Ŀ	13/	

Listening. Learning. Leading.®

Detail: Part of the fine-grained progression for Reasons and Evidence					
	Interpretation	Expression	Deliberation		
Prelim-	Identifies reasons	Generates at least one	Can apply template-based		
inary	people give to support	reason to support a specific	argument-generation strategies		
	a specific point	point, in sentence form	(such as making lists or filling in a pro-con chart)		
Founda-	Identifies supporting	Generates multiple reasons	Can apply analytical strategies to		
tional	reasons or evidence in	to support a point, and uses	identify information needed to		
	a written text and	these reasons to counter	support a point, reflecting implicit		
	relates them to the	others' argument in an	understanding of common		
	point they support	engaging, familiar context	argument schemes		
Basic	Recognizes and	Builds logical,	Can evaluate the strength of		
	explains the	hierarchically structured	evidence and distinguish sound		
	relationship between	arguments by selecting and	and unsound arguments by		
	main and supporting	arranging reasons and	recognizing common syllogisms		
	points and keeps track	evidence to support main	and fallacies		
	of which evidence	and subsidiary points			
	supports which point				
Interme-	Identifies specific	Writes simple critiques or	Can apply critical-question		
diate	points in a text that are	rebuttals that critically	strategies for commonly used		
	vulnerable to	provide both summaries of	argumentation schemes to generate		
	objections and	and responses to other	counterarguments and determine		
	counterarguments	people's arguments	how to reinforce specific points		



Mapping to ECD (evidence-centered design)

- The overview and specific learning progressions define possible student model variables.
- The levels in the learning progressions define the evidence model.
- · We then define item types linked to cells in the LP, as our task models.

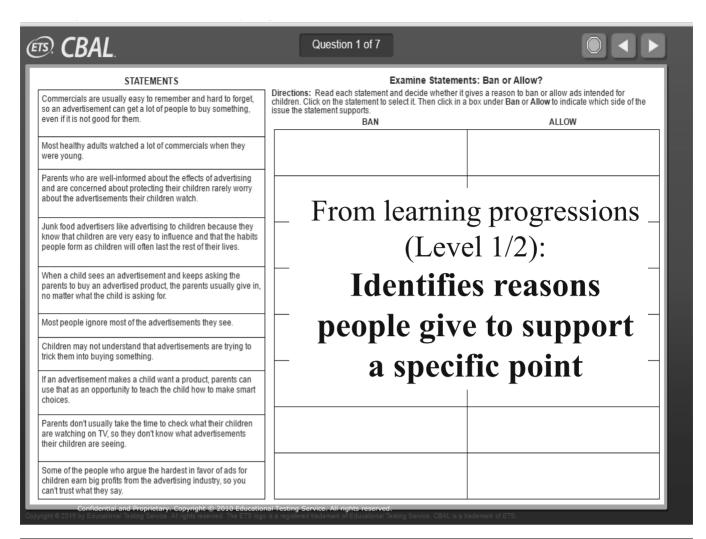


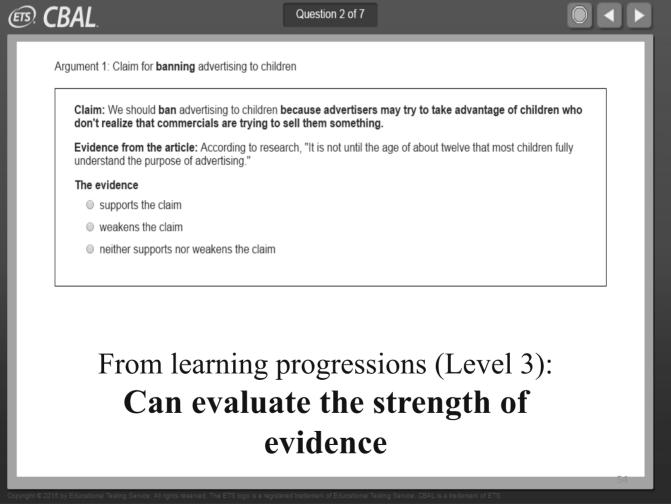


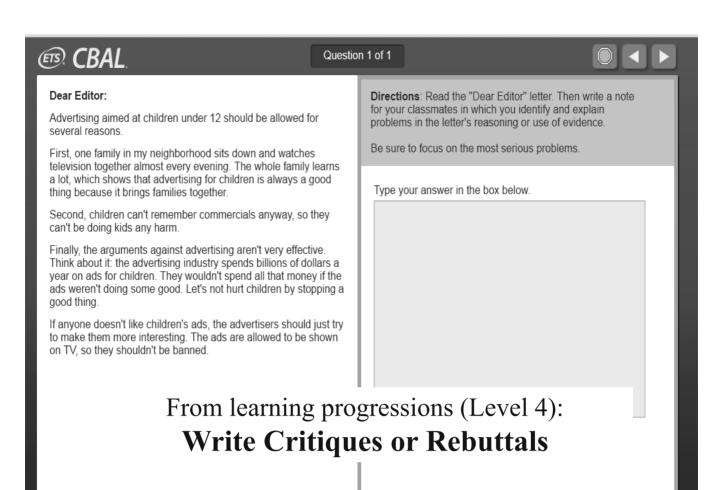
Listenina, Learnina, Leadina,

Back to the Test Design: How some of the 'Ban Ads' tasks map back to the Learning Progressions











Listening. Learning. Leading.

So how well does this approach work?





Overall Test Characteristics

Overall technical quality comparable to similar assessments without scenario structure

Broadening the Scope of Reading Comprehension Using Scenario-based Assessments: Preliminary Findings and Challenges, J. P. Sabatini, T. O'Reilly, L. Halderman & K. Bruce (2014), L'Année psychologique, Vol.114, Issue 04, pp. 693-723

CBAL: Results From Piloting Innovative K-12 Assessments R. E. Bennett (2011), ETS Research Report No. RR-11-23





Listenina, Learnina, Leadina,

Dimensionality

Best fitting IRT model had distinct dimensions for writing (constructed response) vs. reading (selected response) but not for choice of topic

- From Cognitive-Domain Theory to Assessment **Practice.** Randy E. Bennett, Paul Deane, & Peter W. van Rijn. (2016). Educational Psychologist 51, 1, 1-26. DOI: 10.1080/00461520.2016.1141683.
- **Empirical Recovery of Argumentation Learning Progressions in Scenario-Based Assessments of** English Language Arts, P. W. van Rijn, E. A. Graf, & P. Deane (2014), Psicología Educativa, Vol. 20, No. 2, pp. 109 - 115





Recovery of Learning Progressions

Empirically recovered the order of the levels of the learning progression from relative difficulty of tasks

- From Cognitive-Domain Theory to Assessment Practice. Randy E. Bennett, Paul Deane, & Peter W. van Rijn. (2016). Educational Psychologist 51, 1, 1-26. DOI: 10.1080/00461520.2016.1141683.
- **Empirical Recovery of Argumentation Learning Progressions in Scenario-Based Assessments of English** Language Arts, P. W. van Rijn, E. A. Graf, & P. Deane (2014), Psicología Educativa, Vol. 20, No. 2, pp. 109-115





Listenina, Learnina, Leadina,

Some Forthcoming Results

Presenting the preparatory tasks before the essay writing task produced more reliable overall scores than presenting the essay writing task first, followed by the other tasks.

(forthcoming study by Mo Zhang and others)





Scoring Writing





Listenina, Learnina, Leadina,

Leading Ideas

- Seek to combine automated scoring for general writing quality with human scoring for the quality of the ideas
- Strong evidence for feasibility of the approach





Key Publications

- On the Relation Between Automated Essay Scoring and Modern Views of the Writing Construct. P. Deane (2013). Assessing Writing, Vol. 18, No. 1, pp. 7-24
- **Automated Essay Scoring in Innovative Assessments of Writing** from Sources. P. Deane, F. Williams, V. Weng, & C. S. Trapani (2013) Journal of Writing Assessment, Vol. 6, No. 1
- Covering the construct: An approach to automated essay scoring motivated by a socio-cognitive framework for defining literacy skills. P. Deane (2013). Chapter in M. D. Shermis & J. Burstein (Editors), The Handbook of Automated Essay Evaluation: Current Applications and New Directions. Routledge.
- **Automated Scoring Within a Developmental, Cognitive Model of** Writing Proficiency. P. Deane, T. Quinlan, & I. Kostin (2011) ETS Research Report No. RR-11-16.





Listenina, Learnina, Leadina,

Rubric for Quality of Argument

Level 4 (Clearly Competent)

Command of Argument Structure

- States a clear position on the issue
- Uses reasons and evidence to build a case in support of the position
- May also consider and address counterarguments

Ouality and Development of Argument

- Makes reasonable claims about the issue
- Supports claims by citing and explaining relevant reasons and/or examples
- Is generally accurate in its use of evidence

Awareness of Audience

- Focuses primarily on content that is appropriate for the target audience
- Expresses ideas in a tone that is appropriate for the audience and purpose for writing





Rubric for General Writing Quality

Level 4 (Clearly Competent)

It is adequately structured

Overall, the response is clearly and appropriately organized for the task. Clusters of related ideas are grouped appropriately and divided into sections and paragraphs as needed. Transitions between groups of ideas are signaled appropriately

It is coherent

Most new ideas are introduced appropriately. The sequence of sentences leads the reader from one idea to the next with few disorienting gaps or shifts in focus. Connections within and across sentences are made clear where needed by the use of pronouns, conjunctions, subordination, etc.

It is adequately phrased

Ideas are expressed clearly and concisely. Word choice demonstrates command of an adequate range of vocabulary. Sentences are varied appropriately in length and structure to control focus and emphasis

It displays adequate control of Standard Written English

Grammar and usage follow Standard Written English conventions, but there may be minor errors. Spelling, punctuation, and capitalization follow Standard Wirtten English conventions, but there may be minor errors.



Listenina, Learnina, Leadina,

Score Reporting





Overview

- Some research, focusing on methods of presenting how well students are performing on specific key practices and especially learning progressions
- Not as well developed as the test design piece thus far
- Limits to what I can say without touching on work currently being done to develop new ETS product, WinSight, based on CBAL thinking.
- Will present some examples of reporting ideas and CBAL research related to it



Copyright © 2013 by Educational Testing Service. 67



Listenina, Learnina, Leadina,

Some Key Publications

- Keeping Your Audience in Mind: Applying Audience Analysis to the Design of Interactive Score Reports. D. Zapata-Rivera & R.I. Katz, (2014)/ Assessment in Education: Principles, Policy & Practice, Vol. 21, No. 4, pp. 442–
- Comparing Graphical and Verbal Representations of Measurement Error in Test Score Reports. R. Zwick, D. Zapata-Rivera, & M. Hegarty (2014). Educational Assessment, Vol. 19, No. 2, pp. 116–138
- **Exploring Teachers' Understanding of Graphical Representations of Group Performance.** D. Zapata-Rivera, & M. Vezzu, & W. VanWinkle. ETS Research Memorandum No. RM-13-04
- **Applying Score Design Principles in the Design of Score Reports for CBAL™Teachers.** D. Zapata-Rivera, W. VanWinkle, & R. Zwick (2012). ETS Research Memorandum No. RM-12-20.
- **Designing and Evaluating an Interactive Score Report for Students.** M. Vezzu, W. VanWinkle, & D. Zapata-Rivera (2012). ETS Research Memorandum No. RM-12-01
- **Designing and Evaluating Score Reports for Particular Audiences** D. Zapata-Rivera (2011). ETS Research Report No. RR-11-45
- An Evidence-Centered Approach to Using Assessment Data for Policymakers. J. S. Underwood, D. Zapata-Rivera, & W. VanWinkle (2010). ETS Research Report No. RR-10-03





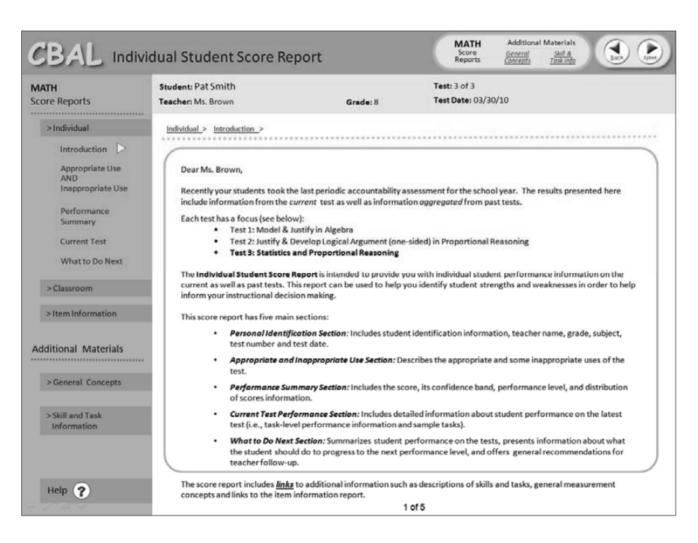
Major Themes

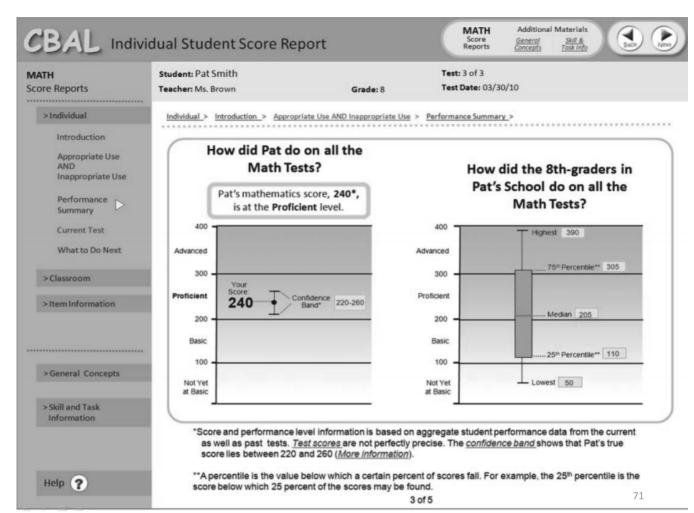
- Focus on how to make score reports easier to use and understand
- Experimentation with best ways to present testing concepts graphically
- Exploration of how to present innovative ideas (i.e., student status on learning progressions)

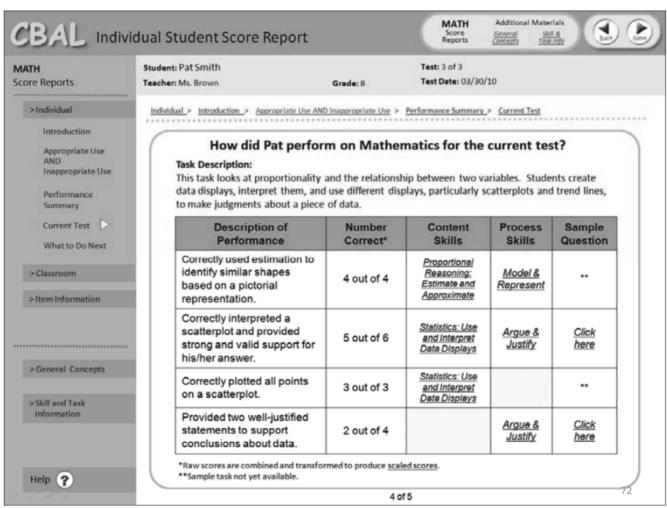


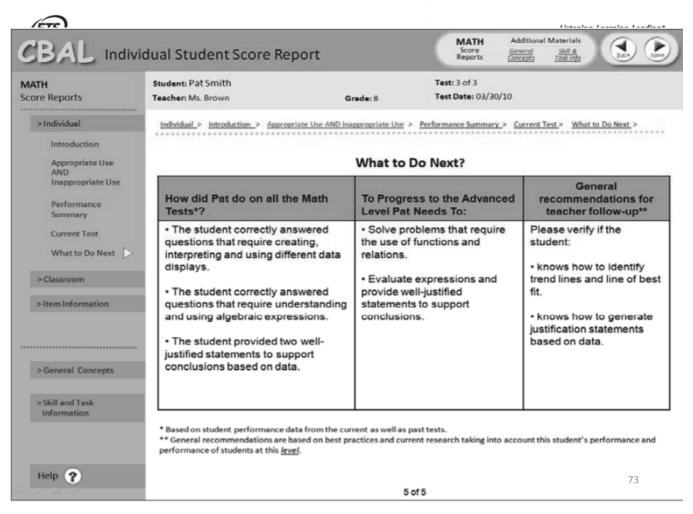
Copyright © 2013 by Educational Testing Service.

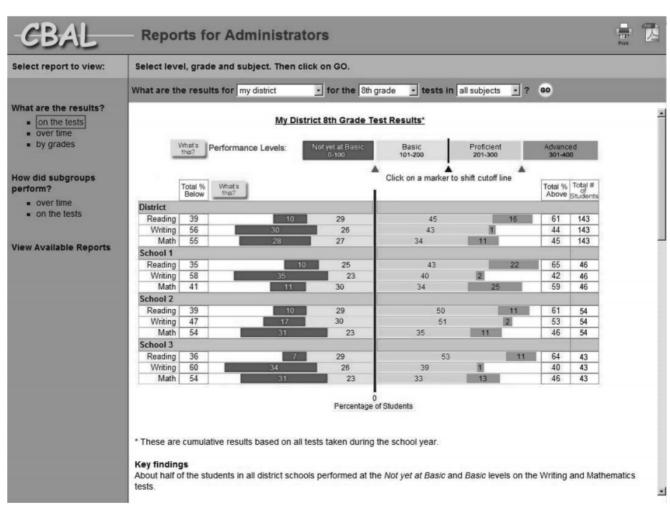
-













Some Caveats

- These are proof of concept ideas, not part of an operational system
- They are designed for the intended CBAL use case (gradually updating information about student performance from multiple smaller assessments given over the course of the year)
- Application to admissions testing would be different



Copyright © 2013 by Educational Testing Service. 75



Listenina, Learnina, Leadina,

Implications for Admissions Testing





The Value of Contextualization

- Traditional approaches to testing often try to decontextualize assessment tasks
- Attempts to use more authentic tasks are often criticized on the grounds that they damage measurement.
- CBAL research suggests that we can maintain desirable measurement properties in an assessment that looks much more like a realistic, authentic task





Listenina, Learnina, Leadina,

Similarity to Classroom Exercises

- The constraints of high-stakes testing can make admissions tests look very different from classroom practice
- But CBAL assessments are easily repurposed as classroom exercises.





Can we go beyond a single score?

- In admissions testing, everything reduces to a single number for a subject area (math, reading, writing).
- Subscores are limited in reliability.
- But users always want subscores!
- CBAL approach might help but only if link to classroom assessment makes it easier to gather more information over the course of the year





Listenina, Learnina, Leadina,

Caveats and Conclusions

- At ETS we are only just beginning to move this work from research into practice in high-stakes tests
- Testing regimes differ with national culture. This must be taken into account in any attempt to learn from CBAL experience and apply it in another country.
- But even so ... the results of CBAL research are encouraging, and suggestive.



〔報告 1〕

パフォーマンス評価をどう活かすか—— 京都大学教育学部における特色入試の取り組み

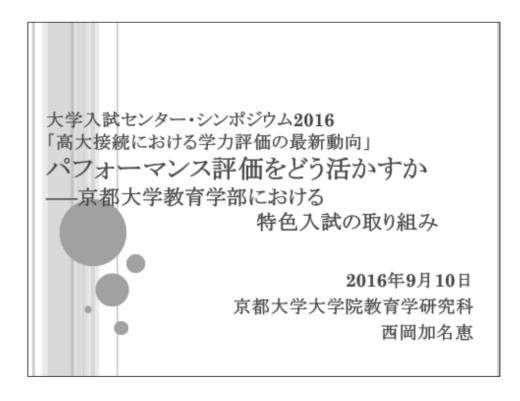
西岡 加名恵(京都大学大学院教育学研究科准教授)

京都大学大学院教育学研究科准教授。京都大学大学院教育学研究科修士課程修了。英国バーミンガム大学で Ph.D. (Ed.) 取得。鳴門教育大学講師を経て、2004 年 4 月より現職。専門は教育方法学(カリキュラム論、教育評価論)。日本教育方法学会常任理事、日本カリキュラム学会理事。主著に、『教科と総合学習のカリキュラム設計』(単著、図書文化、2016 年)、『新しい教育評価入門』(共編著、有斐閣、2015 年)など。



【山地】時間になりましたので、これから三つの報告をいただきたいと思います。最初に、 西岡加名恵先生です。西岡先生は、文科省関係の委員会でも教育方法、カリキュラ ム等についていろいろ助言をなさっており、教育方法学、特にパフォーマンス評価、 パフォーマンス課題の作り方等については日本での第一人者と申し上げてよいかと 思います。本日は、京都大学で始まった特色入試、その教育学部での取組について お話をいただきます。

では、西岡先生、よろしくお願いいたします。



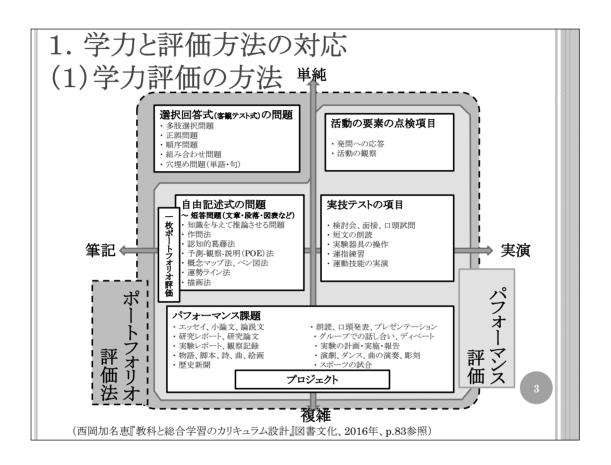
皆さん、こんにちは。京都大学の西岡と申します。本日はこのような機会を頂きまして、 どうもありがとうございます。

■自己紹介

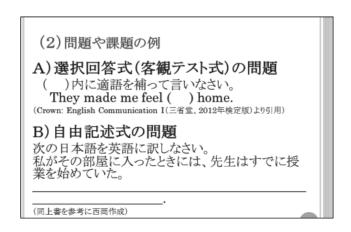
- 教育方法学(カリキュラム論、教育評価論)
- パフォーマンス評価(ポートフォリオ評価法、パフォーマンス 課題、ルーブリック)
- アメリカやイギリスの調査
- 日本の学校の先生方と共同研究開発
- 大学では教員養成の仕事も担当(教育実習指導など)
- 京都大学大学院教育学研究科E.FORUMの講師と運営を 担当 http://www.educ.kyoto-u.ac.jp/e-forum/
- ・中央教育審議会 初等中等教育分科会 教育課程部会 児童生徒の学習評価の在り方に関するワーキンググループ(2009年6月~2010年3月)、育成すべき資質・能力を踏まえた教育目標・内容と評価の在り方に関する検討会委員(2012年12月~2014年3月)、高等学校における「多様な学習成果の評価手法に関する調査研究」評価・推進委員会委員(2013年6月~2016年3月)

ただ今のご紹介にありましたように、私は教育方法学を専門としておりまして、主には 小・中・高等学校との共同研究において、パフォーマンス評価を取り入れた実践改善の共 同研究を進めております。 さて、今回は「パフォーマンス評価をどう活かすか」というテーマで、特に京都大学教育学部の特色入試についてご紹介したいと思います。

●学力と評価方法の対応



まず、高大接続に関わる改革動向を検討する上では、学力と評価方法との対応関係を確認しておくことが有効と思います。この図では、上に行くほど単純な評価方法、下に行くほど複雑な評価方法を示しています。左側には筆記による評価、右側には実演による評価の方法を並べています。



具体例をお示しした方が分かりやすいかと思いますので、問題や課題の例を幾つか並べてみました。英語科の筆記による評価の場合です。

- A) の問題,「() 内に適語を補って言いなさい」, これは単純に"at home"というフレーズを覚えていて, 再生できるかどうかを試す問題です。〇×で採点できるという点で, 選択解答式の問題と言えます。
- B)の和文英訳問題,「私がその部屋に入ったときには,先生はすでに授業を始めていた」。 これは, when 節や時制などの文法事項を使いこなせるかどうかを見る自由記述式の問題 です。

C) パフォーマンス課題

あなたは園部高校を訪れている外国からのお客さんに日本のことや暮らしを紹介することになりました。下にあるような内容を適宜選んで自由に英語で書いてください。その際,あなた自身の経験や感じたことなどを書いて,できるだけ具体的になるようにしてください。

わたしからのお奨め観光スポットやもの・私の住む町・市・私の家族・趣味・好きなこと・学校生活・週末の過ごしかた・など

(京都府立園部高等学校における実践。ただし、課題文を一部修正した。)

C) のパフォーマンス課題は、さらに複雑な課題です。「あなたは園部高校を訪れている外国からのお客さんに日本のことや暮らしを紹介することになりました。下にあるような内容を適宜選んで自由に英語で書いてください。その際、あなた自身の経験や感じたことなどを書いて、できるだけ具体的になるようにしてください」。

ここでは、特定の状況において自分の持っている知識やスキルを総動員して作品を作る ことが求められているということがお分かりいただけるかと思います。

(3)パフォーマンス課題とは・・・

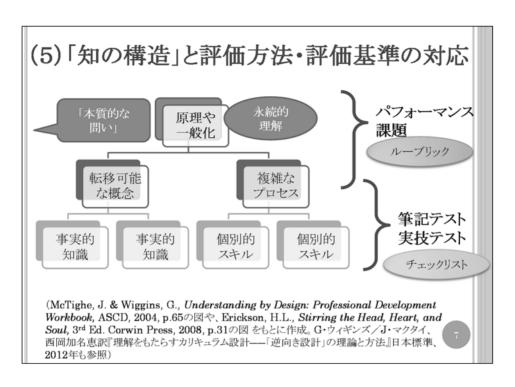
- ○様々な知識やスキルを総合して使いこなすことを求めるような、複雑な課題。
- 具体的には、論説文やレポート、展示物といった 完成作品(プロダクト)や、スピーチやプレゼン テーション、実験の実施といった実演(狭義のパ フォーマンス)を評価する課題。

(4)パフォーマンス評価とは・・・

- の知識やスキルを使いこなす(活用・応用・総合する)ことを求めるような評価方法(問題や課題)
 - ←学力観の転換

このように、パフォーマンス課題(performance tasks)というのは、様々な知識やスキルを総合して使いこなすことを求めるような、複雑な課題を指します。なお、パフォーマ

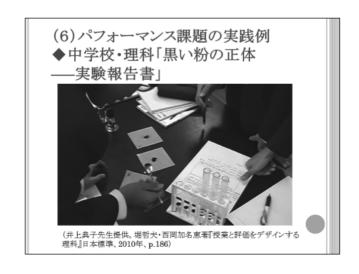
ンス評価 (performance assessment) というのは、知識やスキルを使いこなすことを求めるような評価方法の総称として捉えております。



ただ今ご紹介した事例から、評価方法と、評価対象となる学力の種類とが、ある程度対応していることが実感していただけたのではないかと思います。私自身は、アメリカの、

「真正の評価」に関する研究者、ウィギンズ(Wiggins, G.)とマクタイ(McTighe, J.)が提唱している『理解をもたらすカリキュラム設計』の理論、通称「逆向き設計」論に依拠して、学校との共同研究を進めています。「逆向き設計」論では、この図のように「知の構造」と評価方法の対応が整理されています。

パフォーマンス課題については、教科の中核に位置するような「原理や一般化」についての「本質的な問い」や「永続的理解」に対応させて用いることが提案されています。



パフォーマンス課題は、2008年改訂の現行の学習指導要領で、知識やスキルを活用する 思考力・判断力・表現力を育成し評価する方法として注目されるようになりました。学校 現場でも、かなりの実践例が開発されてきています。例えばこれは中学校の理科ですが、 理科室から正体不明の黒い粉が入った3本の瓶が出てきた。これらの瓶に入っている粉の 正体を解明するような実験を自分たちで計画して、実施して、報告しようという課題です。

◆高等学校·英語科

※動画: 京都大学オープンコースウェアに掲載
 (西岡「教育課程論 II 」2014年10月23日)
 http://ocw.kyoto-u.ac.jp/ja/03-faculty-of-education-jp/13-9234001
 国際文化コースの1人の生徒の変化

→普通科旧 I 類の2年次よりの選択コース。



(京都府立東舞鶴高等学校 大槻裕代先生提供)

与えられたテーマについてプレゼンテーションを行うというような課題もあります。学校紹介の課題だけではなくて、自分の好きな絵について語る、観光名所について語るなどですね。もちろん、主張に関して何らかの論説を展開するというような課題も取り組まれています。

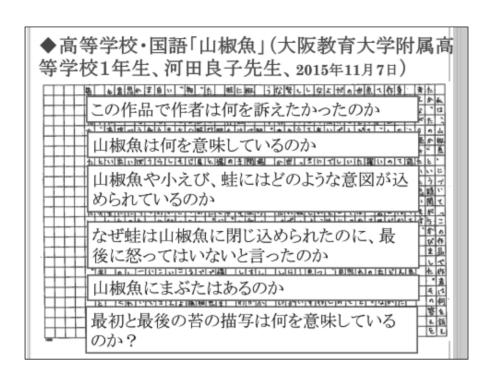
◆中学校・社会科 「民主的な国家を 提案しよう!」

(三藤あさみ「検討会で関連 づけて思考する力を育成す る」西岡加名恵・田中耕治編 著『「活用する力」を育てる授 業と評価・中学校』学事出版、 2009年



何らかのテーマについて、まとまったレポートを書くという課題もあります。これは中

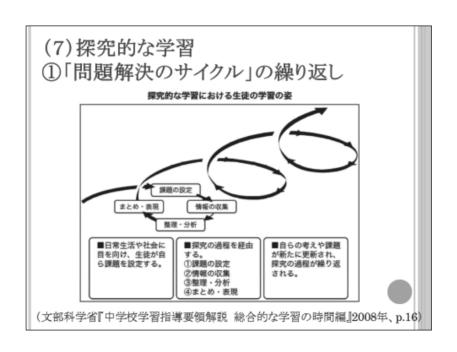
学校の社会科で、独裁国家が崩壊したときに、どうすれば民主的な国家が作れるのかということを提案しようということで、授業の中で学んだ知識やスキルを総合して使いこなすということが求められています。



これは高等学校の国語の例ですが、「山椒魚」という物語について、自分なりの視点を定めて批評文を書こうというものです。非常にユニークな様々な視点が出てきたのですが、 これはもちろん学校の先生が、そのような指導をされているからこそです。



なお、各教科のパフォーマンス課題の例については、京都大学大学院教育学研究科で E.FORUM というネットワークを作って、教員研修の場を提供しているのですけれども、 そちらのウェブサイトでもご紹介していますので、またご参照いただければと思います。



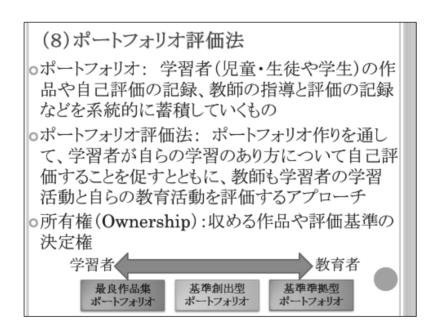
一方で、いわゆる探究的な学習、つまり生徒自身が課題を設定し、情報収集、整理・分析、課題の再設定というサイクルを繰り返しながら学ぶ学習も広がっています。小学校・中学校においては、総合的な学習の時間が既にあるわけですけれども、最近ではスーパーサイエンスハイスクールやスーパーグローバルハイスクールなどを中心に、高等学校でも広がってきているのを感じます。

②合科・探究型の課題

D) 皆さんに、地域や人々を幸せにする方法を提案してもらいます。地域の特色について3~5人のグループで、フィールドワークにより調査したうえで、企画を考え、提案してください。また、その企画を地域に住む外国人にも紹介するための、英文リーフレットを作成してください。(金沢大学人間社会学域学校教育学類附属高等学校の実践をふまえて、筆者作成。Cf. 山本哲次「SGHと探究学習の実践」E.FORUM教師カアップ研修「探究力をどう育成するか」2015年3月28日、http://ocw.kyoto-u.ac.jp/ja/opencourse/92/videos)

これは、スーパーグローバルハイスクールの一つである金沢大学附属高校の実践を下敷きに作ってみた課題文です。「皆さんに、地域の人々を幸せにする方法を提案してもらいます。地域の特色について3~5人のグループで、フィールドワークにより調査した上で、企画を考え、提案してください。また、その企画を地域に住む外国人にも紹介するための、英文リーフレットを作成してください」というような課題です。

先ほど紹介した A) ~C) の問題や課題と比べ、教科の枠にとどまらない、さらに総合的な課題となっています。地域の特色を捉える地理の力、英文リーフレットを作る英語力、インタビューをする国語力、そういう力に加え、チームワーク、企画力など、汎用的なスキルが総合的に試される課題となっています。

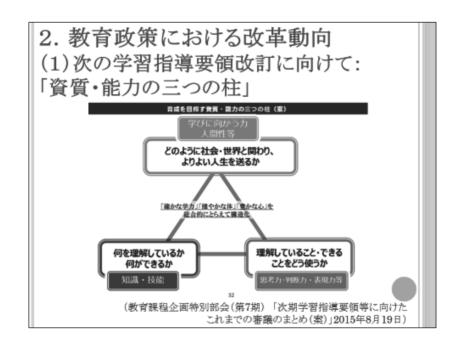


こういった学習がどのように進められているかということを見るためには、ポートフォリオ評価法を活用することが有効だと考えられます。



ポートフォリオ評価法とは、学習の足跡を示す資料をファイルなどに蓄積し、学習者の学習成果を幅広く評価するものです。これもパフォーマンス評価の方法の一種と言えます。

●教育政策における改革動向



以上、ご紹介してきた A)~D)の問題や課題を念頭に置いていただくと、近年の教育改革の動向も理解しやすいように思います。2017年に予定されている次の学習指導要領改訂に向けては、「育成を目指す資質・能力の三つの柱」が提案されています。「何を理解しているか、何ができるか(個別の知識・技能)」に当たるものが、先ほどの問題 A)や問題 B)、「理解していること・できることをどう使うか(思考力・判断力・表現力等)」が課題 C)、「どのように社会・世界と関わり、よりよい人生を送るか(学びに向かう力、人間性等)」に当たるのが課題 D)というふうに理解するとよいのではないかと私は考えております。



ちなみに、高大接続システム改革会議の「最終報告」でも、類似の3要素を、幅広い評価方法を用いて評価するという方向性が目指されています。

●京都大学の特色入試



(京都大学特色入試: http://www.nyusi.gakusei.kyoto-u.ac.jp/tokushoku/message/)

さて、京都大学の特色入試は、「高校3年間+大学4年間」で「一人一人を丁寧に育てて いこう」という理念の下に、2016年度入試より導入されました。

学部•学科		選抜方法	試験実施方式 提出書類	
総合人間学部		書類審査、能力測定考査、 大学入試センター試験の成績	学力型AO	調査書、学業活動報告書、 学びの設計書
文学部		書類審査、論文試験、論述試験 大学入試センター試験の成績	学力型AO	調査書、学業活動報告書、 学びの設計書
教育学部		書類審査、課題、口頭試問、 大学入試センター試験の成績	学力型AO	調査書、学びの報告書、 学びの設計書
法学部		書類審査、小論文試験、 大学入試センター試験の成績	後期日程	調査書
経済学部		書類審査、論文試験 大学入試センター試験の成績	学力型AO	調査書、学業活動報告書、 学びの設計書
理学部		書類審査、能力測定考査、口頭試問 大学入試センター試験の成績	学力型AO	調査書、学業活動報告書、 学びの報告書
医学部	医学科	書類審査、小論文試験、面接試験	推薦	調査書、推薦書、学びの設計書、 TOEFL-iBTの受験者成績書、 特色事項に関する資料
	人間健康科学科	書類審査、論文試験、面接試験、 大学入試センター試験の成績	学力型AO	調査書、学業活動報告書、 学びの設計書
薬学部	薬科学科	書類審査、論文試験、面接試験、 大学入試センター試験の成績	学力型AO	調査書、学業活動報告書、学びの記 計書、TOEFL-iBTの受験者成績書
	地球工学科	書類審査、口頭試験、面接試験、 大学入試センター試験の成績		調査書、推薦書、学びの設計書、顕著な活動実績の概要
工学部	電気電子工学科	書類審査、 大学入試センター試験の成績	TT EE	
	情報学科	書類審査、口頭試問、 大学入試センター試験の成績		
	工業化学科	書類審査、大学入試センター試験の成績		20
農学部	食料·環境経済学 科	書類審査、小論文試験、 大学入試センター試験の成績	学力型AO	20

主として学力型 AO 入試が採用されておりますが、具体的な設計は学部により異なりま す。今日お話ししますのは、あくまで教育学部の例だということをご承知置きください。

4. 京都大学教育学部の特色入試:導入過程 (1) 導入に向けての議論や調査

- o 2012(H24)年度 教育学部:多次元入試研究会
- 2012(H24)年度 教育学部:入試改革ワーキング・グループ →学部の全教員で検討
- o 2013(H25)年4月 全学:特色入試実施準備委員会発足
- o 2014(H26)年4月 全学:特色入試実施委員会発足
- 2013(H25)年5月 教育学部:特色入試実施準備委員会 発足
 - 総長裁量経費を得て、他大学への訪問調査など
 - 求める人物像、定員、出願書類、試験方法の検討
 - →要項やサンプル問題の作成
 - 高校対象アンケートの実施(約120校)
 - オープンキャンパスにおける高校生アンケートの実施(400名X2年)
 - 特色入試説明会(2015年3月、学部独自)
- o 2015(H27)年4月 教育学部:特色入試実施委員会発足

特色入試の導入に当たっては、2012年度より、様々な議論や調査を行ってきました。教

育学部については、センター試験と個別試験の組み合わせで行われる一般入試がうまく機能しているという実感がありますので、正直申しまして、当初、大半の教員は特色入試の導入には消極的でした。しかしながら、全ての学部で特色入試を導入するということが全学の方針として決まってしまいました。

いざ導入するとなれば、教育学的な知見の真価が問われる改革ですので、学部の全教員が参加して慎重な議論や調査を重ねました。教員の間では、「やるからには、私たち教育学部で蓄積している知見を高校生や社会に発信する機会にしたい」「高校現場の教育改善のための提案をするような入試にしたい」「一般入試で入ってくる学生たちのリーダーにもなってくれるような学生が入ってくれればうれしい」といったことを話し合いました。

(2)主として参考にした先行事例

- 。筑波大学AC入試 ... 自己推薦書
- ◦九州大学「21世紀プログラム」AO入試
 - ・・・第1次選抜: 活動歴報告書など
 - ・・・第2次選抜

1日目: 講義・レポート(120分)×3 2日目: グループ討論(150分)

小論文(270分)うち15分が個人面接

○京都工芸繊維大学 ダビンチ(AO)入試

o東北大学AO入試

Cf. 京都大学大学院教育学研究科 大学院入試

具体的な入試方法を考えるに当たっては、他大学の事例も調査いたしました。特に参考になったのは、筑波大学のAC入試、九州大学21世紀プログラム、東北大学や京都工芸繊維大学などで取り入れられているAO入試などです。また、私たちがこれまでに実施してきた大学院入試のイメージも下敷きになりました。

ここには書いておりませんが、海外の事例もかなりいろいろ調べました。



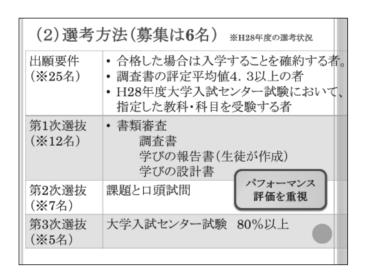
また、E.FORUMの研修の機会などを使って、高校の先生方と共に高大接続を考える研修・セミナーなども行いました。

5. 京都大学教育学部の特色入試:設計(1) 求める人物像

- ○教科の学習及び総合的な学習の時間などにおいて 学習を深め、テーマを設定して探究活動を行い、卓 越した学力を身につけ、成果をあげた者、あるいは、 学校内外の活動で豊かな経験を積み、創造的な熟 達を通して、深い洞察を得ている者
- ○人間と社会、教育や心理について関心を持ち、論理的・批判的に思考し、問題を解決する能力とコミュニケーション能力を持つ者
- の将来、教育や心理にかかわる専門的識見を発揮して、社会に貢献する志を持つ者

(京大·特色入試: http://www.nyusi.gakusei.kyoto-u.ac.jp/tokushoku/)

さて、具体的な制度設計に当たっては、まず「求める人物像」を明確にしました。ここに示した 3 項目は、それぞれ、過去の実績、現在の能力、将来の展望に対応しています。特色入試の実施に当たっては、選抜の基準がこの「求める人物像」になります。計画・実施の過程では、何度もこの人物像に立ち戻りました。「求める人物像」を明確にしておくことは、このような入試を行う上でとても重要だと実感した次第です。

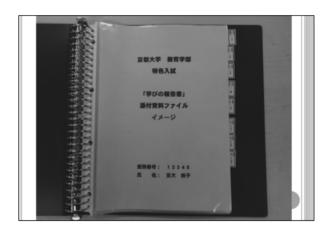


具体的な選抜方法としては、第 1 次選抜が「調査書」「学びの報告書」「学びの設計書」による「書類審査」、第 2 次選抜が「課題と口頭試問」、第 3 次選抜が大学入試センター試験で 80%以上です。特に第 1 次選抜・第 2 次選抜でパフォーマンス評価を重視している点に特徴があります。

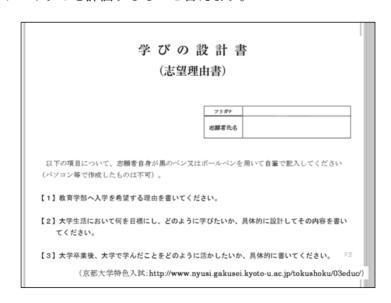
2016年度の場合,25名の出願があり,第1次選抜で12名が合格,第2次選抜で7名が合格,第3次選抜で5名が最終的な合格となりました。

	(3)第1次選考: 書類選考 学びの報告書(H28年度)				
間、読 過に沿 (1) 「 て (2) 「 資 (3) 「	【1】 中学時代から現在までに取り組んだ「学び」の活動(各教科での学習や総合的な学習の時間、読書、課外活動、学校行事での活動、ボランティア活動等)のうち、主なものを時間の経過に沿って記述してください。 (1) 「時期」欄には活動を行った時期(西暦で○年○月、○年○月~○年○月など)を記入してください。 (2) 「活動内容」欄には活動の名称と簡単な説明を書いてください。その際、その成果を示す資料を添付してください。 (3) 「資料番号」欄には対応する資料の番号を記入してください。 ※取得した資格や各種の検定については、上記とは分けて【2】の欄に記載してください。				
時期	活 動 内 容	資料番号			
	(京都大学特色入試:http://www.nyusi.gakusei.kyoto-u.ac.jp/tokusho	≈∉ ku/03educ/)			

教育学部の特色入試の特徴として、受験者本人が作成する「学びの報告書」があります。 これは「中学時代から現在までに取り組んだ『学び』の活動のうち、主なものを時間の経 過に沿って記述」することを求めるものです。



さらに、その成果を示す資料をファイル 1 冊分添付することを勧めています。これは、 すなわちポートフォリオを評価するものと言えます。



「学びの設計書」については、志望理由、大学生活において行いたいこと、卒業後の展望を書いてもらうものとなっています。

(4)第2次選考: 課題と口頭試問

○課題(H28年度)→次スライドへ

近年の日本においては、青少年の規範意識やモラルの低下を理由に、道徳教育の強化を求める声が聞かれる。青少年に関する資料(資料集掲載の資料1~資料6)を用いながら、以下の問いに答えなさい。

なお、用意された資料は、問いで指定されたものを 除いて必ずしもすべてを使う必要はない。また、解答 にあたって資料に言及するときには、「資料1によれ ば」「表2-2-Aによれば」といったように表記すること。 (計100点)

(http://www.nyusi.gakusei.kyoto-u.ac.jp/tokushoku/past_issues/)

- ◎口頭試問
- 提出書類についての質疑など

第2次選考は、1日目が3時間かけて取り組む「課題」です。2016年度の場合は、次のような問題を出しました。「近年の日本においては、青少年の規範意識やモラルの低下を理由に、道徳教育の強化を求める声が聞かれる。青少年に関する資料(資料集掲載の資料1~資料6)を用いながら、以下の問いに答えなさい」。

●課題:資料集

資料 1. Collective Delinquency

(出典) B. S. Griffin & C. T. Griffin, *Juvenile Delinquency in Perspective*, Harper & Row, 1978, pp.77-79

資料2. シンナー等有機溶剤濫用(乱用)に関する資料

2-1. シンナー等有機溶剤濫用事犯

(出典) 法務省『昭和62年版 犯罪白書』1987年,第1編第2章第1節4

2-2. シンナー等有機溶剤乱用少年の心理

(出典) 土井敏彦「シンナー等有機溶剤乱用少年の心理的側面 ―― 頗回乱用者特性を中心

として」『科学警察研究所報告防犯少年編』第 16 号, 1975 年, pp.10-18

資料3. 青少年に関する各種の統計資料

<中略>

資料4. 谷川俊太郎の詩

(出典) 河合隼雄『子どもと悪 --- 今ここに生きる子ども』岩波書店,1997年,裏表紙

資料5. リミナリティとコムニタス (出典) V. W. ターナー (富倉光雄訳)『儀礼の過程』 思索社, 1976年, pp.125-129 3.3 3

資料6. 逸脱行動としての非行

(出典) 大村英昭『非行のリアリティ』世界思想社, 2002年, pp.106-108

この課題では、テーマに関連する複数の資料を掲載した資料集を与えます。英語や日本 語の文章のほか、統計資料や詩も掲載しました。

●課題

問1. **資料1**に示した英文は、青少年の集団非行に関するものである。この文章について、次の問いに答えなさい。(30点)

- (1)下線部を和訳しなさい。(10点)
- (2)この文章で著者が述べていることを600字以内で要約しなさい。(20点)

問2. 次のグラフ(図A)は、薬物乱用で送致された少年(20歳未満の者)の人数の変遷を示したものである。送致人数が急速に減少している事実の解釈について、資料集に示した複数の資料を踏まえて推測されることを述べなさい。その際、どのような根拠に基づいて推測したのかを含めて説明しなさい。(40点)

問3. 子どもから大人に移行する過程で、自律的な規範意識が培われるためには、どのようなことを体験し、どのように成長していくことが求められるのだろうか。 資料集に示した複数の資料に言及しながら論じなさい。 (30点)

問題は、問1が英文の要約と下線部訳、問2が特定の資料を他の資料を用いつつ検証する問題、問3が資料を用いつつ自分の主張を提案する問題となりました。

また、2日目の第2次選抜では、口頭試問を行いました。そこでは、提出してもらった書類についての質疑応答などが行われました。

6. 京都大学教育学部の特色入試: H28年度合格者への聞き取り調査

(2016年5月10日、http://www.nyusi.gakusei.kyoto-u.ac.jp/tokushoku/03educ/)

- (1)なぜ特色入試を受験しましたか?
- ○京大教育学部を志望 「チャンスが1回増える」 「直接京大への想いを伝えられる絶好のチャンス」
- 人物像に適合していそう
- ○高校時代の探究活動などが活かせそう
- ○書類の形式に魅力を感じた
 - →自分の活動を振り返りたい

この特色入試が成功しているかどうかは、この入試による入学者が卒業する時点にならないと分からないと思います。しかしながら、5月に合格者への聞き取り調査をしましたので、その様子を一部、ご紹介したいと思います。

まず、受験理由については、「チャンスが増えるため」「自分が人物像に適合していそうだから」という意見が多く語られました。変わったところでは、「学びの報告書」の書式に魅力を感じた、と言った学生もいました。

(2)どんな準備をしましたか?

- o「私は, 興味のあることや好きなことに欲張って取り組んだ結果, それらが準備に役立ちました。 具体的には, 生徒会活動やボランティア活動です。」
- 「学校から与えられたチャンスを最大限に活かしてきたことだと感じています。私の高校は、とにかく国際交流に熱心でした。かつ部活もしっかりやる学校だったので、全国大会に出場したりもしながら、自分のやりたい海外派遣の参加などにも取り組みました。」
- 「私は高校3年間かけて一貫したテーマで探究活動を 進め、最終的にはそれを論文としてまとめたため、それ を強みにして挑もうと思いました。」

役に立った準備について尋ねたところ, 高校での様々な活動に積極的に取り組んだことが生きたという学生が3名おりました。生徒会活動やボランティア活動,国際交流や部活,探究活動などが挙げられております。

- 「私は中学校まではほとんど特別支援学校で、そこから高校へ行きたいと言って、ただ純粋に普通学校で学びたいと思ってしてきた活動について、書類に書きました。」
- 「教育学部生としてふさわしい高校活動としてこれといった のが思いつかなかったので、自分が教育学部でやりたいことと、これまでやってきたことが何とかつなげられないかと 思って、そういった視点から自分の過去の活動を意味付け するといったことを2週間だけですがやりました。」

◎後輩へのアドバイス・メッセージ

- 「自分の経験を発信する力を十分に持ってほしいので、その練習もどこかでしておけば、特色入試の課題でも口頭試問でも自分をアピールできると思います。」
- ○「私は様々な活動をする上で、それらの活動に一貫性をもたせることを意識しました。自分の中で軸となるものをしっかりと確立していたことで、高校での活動、大学でやりたいこと社会に出てやりたいこと、の3つを書く時に一貫性をもって書くことができたと思っています。」

また、一方では、個人的に取り組んできたことが生かせたと語った学生も2名おりました。学校が提供した活動だけでなく、自分自身が独自に取り組んだ活動を生かして受験にチャレンジしてくれたことがお分かりいただけるかと思います。

(3)後輩へのアドバイス・メッセージ

①高校生活の充実を

- o 「日常の中で,自分がちょっと無理かなと思うことに背伸びして, やりたいけどやれるかなという躊躇を取り払ってやってみるのが 良いと思います。自分で自分の限界を作らないとよく言いますが。 完璧にできる必要はなくて,上手くいかないこととか反省点とか はありますが,......結果的に上手くいかなくても成長できるし, 糧には確実になる.....。」
- 「中学高校のうちに興味のあるところに手を出してみたら、今まで自分が見ていた世界ではないところが見えてくるので、そのような活動には積極的に参加することをお勧めしたいです。」
- 「この入試は受験を選択するまでに多くの経験をしていなければスタート地点にも立つことができないというものなので、中学生・高校生の時間があるときに多くの活動に貪欲に参加することが必要だと思います。しかし、それらの活動は受験のためにするというようなものではありません。......その順序が間違っていれば意味はないと思います。」

後輩へのアドバイス・メッセージを尋ねたところ、高校生活の充実がまず挙げられました。ただし、これは単に受験のためにというのではなくて、人生として豊かに生きるために、それが大事なんだということを言っておりました。

②自分の経験を振り返る良い機会

- ○「特色入試を通してこれまでを振り返ることで、改めて 自分を知ることができるのは本当に良い機会でした。」
- ○「私は突っ込んだ、結構ギリギリのことも書いて出したのですが、それを評価してもらえたり、面接で話し合えたりというのは、限定的な話ですがすごく楽しかったし、やって良かったと思います。書類を書く時に自分の経験をいかに分析するか、どういう切り口で見るかを考えて、自分が意義を与えるという経験をしたのは良かったです。一つの経験でも捉え方によって違いますし、それを一冊の提出書類に作るという経験は高校ではなかなかないので、一度はやっていいと思います。」

また,「学びの報告書」や「学びの設計書」が,自分の経験を振り返る良い機会になった, という声も複数の学生からありました。

自分がなぜこの学部を受験するのかということをしっかり考える,あるいは、今まで自 分が今まで取り組んできたことの意味は何なのか、自分なりに整理する非常に良い機会に なったと言ってくれています。

③将来を考える良い機会

- o 「特色入試は、将来どう進むかを深く考える機会にもなりま す。」
- 「私は,特色入試のおかげで今,これからの大学生活4年間をどうするか,当時作った資料を見ながら考えています。これからの人生の指針になりますし,本当に挑戦して良かったなと思います。」
- o「私は特色入試を受けるまでは自分が大学へ行く意味も, 入ってからどうするか,自分のこれまでやってきたことが何の 役に立っているかも分からなかったけど,書類を書いている うちに自分の活動がどういうふうに今の自分を形作っている か,自分がそれを経てどんなビジョンを持っているかなどに ついて,漠然と頭で考えていたことが型にはまって一つの筋 道が見えました。だからちょっと無理かなと思っても受験する ことが糧になるので受けてみたらどうかなと思います。」

そのことが、すなわち大学へ行く意味、将来を考えるという点でも良い機会になった、 という声も寄せられています。

例えば「私は特色入試を受けるまでは自分が大学へ行く意味も、入ってからどうするか、 自分のこれまでやってきたことが何の役に立っているかも分からなかったけれど、書類を 書いているうちに自分の活動がどういうふうに今の自分を形作っているか、自分がそれを 経てどんなビジョンを持っているかなどについて、漠然と考えていたことが、一つの筋道として見えてきました」というようなことを語ってくれています。

④独特の大変さ、対策は困難

- 「教育学部は他の学部よりたくさんの書類を書かないといけない し、センター試験のプレッシャーも含めて本当にしんどいと思う」
- 「準備は言葉で表せないほど大変です。周りの人とは違う不安を 抱えながら受験に立ち向かうことになります。対策して合格できる 入試ではないということも実感しました。」
- 「受かった今だから言えることがたくさんあります。リスクがかなり 大きくて、失うことや犠牲にすることが多いので、私たちが良い意味で実験台になれたらいいなと思います。相当覚悟がないと出願から合格発表までの4ヶ月の入試なんてやっていられないです.....。」
- o 「率直に言ってこの受験は対策のしようがないものです。ですから、いかなる状況に立たされた時にでも、それに対応する能力が 試されます。逆境に立たされた時にこそ強さを発揮できるような 人がこの試験に向いているのだと思います。」
- 「4人が言ってくれたように、対策は私もしていないし、どうしていいかも分からないし、そもそもないと思うので、培った経験で勝負できる面白い入試だと思います。」

ただし一方では、たくさんの書類を書いたり、センター試験の結果を待ったりする点では、かなり「しんどい」という声もありました。なお、この入試に関しては、「対策のしようがない」ものだと感じた受験生もいたようです。

●意義と課題

7. 意義と課題(西岡の私見)

- (1)意義
- 入学者の幅を広げる。
- 特色入試そのものが学習機会となる。
- ○高校現場への発信。
- 学部教育の見直しを迫られる。
 - →1回生向けの科目「教育研究入門」の改革

(2016年度~)

(2)課題

- 教員の負担が大きい。
- 受験生の心理的負担が大きい。
- o評価できる学力にやや偏り?
- ⇔一般入試の二次試験(4教科)
- →高校の調査書の比較可能性を高めてほしい

最後に、あくまで私個人の私見ではありますが、今回の教育学部の特色入試の意義と課題について考えます。まず、入学者の幅を広げるという点では、一定の意義があると感じました。また、特色入試そのものが意義深い学習機会となっているという指摘も、合格者から聞かれたことに喜んでいます。「探究的な学習」などの意義を高校現場へ発信するといった点で、高校教育の改善にも資することを願っています。

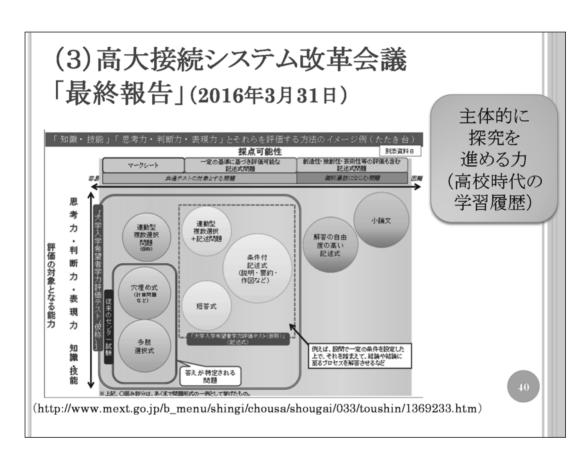
また、実は学部教育の側でも、1回生向けの必修の講義科目を演習型の科目にするという改革の契機となりました。大学で何をやりたいかということを散々尋ねておいて、入っ

てきたら、「さあ、学部の授業は座って聞いておきなさい」というのは、あんまりかなとい う声が大学教員から出てきたということです。

ただし、一方で最大の問題としては、実施する教員の負担が非常に大きいことが挙げられます。特色入試には、課題の作成、書類審査、口頭試問などで膨大な時間と労力がかかります。因果関係は証明できませんが、特色入試を導入した初年度、教育学部では病気になる教員が続出いたしました。

第2に、受験生の負担が危惧されます。一般入試の二次試験でも、受験生は7時間半の 筆記試験を受けますので、物理的な負担は同じぐらいかなと思いますが、特にセンター試 験の結果を待つ期間がつらいという声を複数の入学者から聞きました。

第3に、特色入試の場合、できるだけ多角的な評価を行う設計にはしてありますが、や はり評価できる学力には限界がある点が気になっています。



ここに示したのは、文部科学省の高大接続システム改革会議で、センター試験の後継である「大学入学希望者学力評価テスト(仮称)」の検討の際に示された図です。教育学部の特色入試の場合、この枠の外に位置付けたような「主体的に探究を進める力(高等学校時代の学習履歴)」を評価するものとなっています。また、一番左側の部分、センター試験で5教科の知識・技能を評価するという点もカバーされています。

しかし、現状ではこの真ん中のあたりに大きなギャップがあるわけですね。もちろん、 今後、大学入試センターの方で良いテストを作っていただくと、多少幅が広がってくるわ けなのですけれども、これまでの一般入試で評価されてきたような、各教科での「自由度 の高い記述式の問題」を出題することは、英語以外はできていません。

また、この図において「小論文」とある部分については、本来は、複合テーマ型の小論文と、各教科の内容知識を生かしつつ取り組むようなパフォーマンス課題とを区別しておく必要があると思います。教育学部の2次選抜で出している課題は複合テーマ型の小論文と言えるかと思うのですが、しかし、現行の方法では、学校で取り組まれているような各教科の、教科内容に即したパフォーマンス課題に対応する力を評価することは困難です。

今後,「大学入学希望者学力評価テスト (仮称)」で短い記述式の問題などが開発・出題されたとしても、これらの限界を乗り越えることができない、つまりこの黒い丸の部分が評価できないという点は注意しておく必要があると思います。

			資質・能力の要素(目標の柱)			
	能力・学習活動の階 層レベル(カリキュ ラムの構造)		知識	スキル		情意(関心・意欲・態
				認知的スキル	社会的スキル	度・人格特性)
教科等の枠づけの中	教科等のお	知識の獲得 と定着 (知ってい る・でき る)	事実的知識、技能(個別的スキル)	記憶と再生、機械的実行 と自動化	学び合い、知識の共同構 築	達成による自己効力履
	らいりの中	知識の意味 理解と洗練 (わかる)	概念的知識、方 略(複合的プロ セス)	解釈、関連付け、構造化、 比較・分類、帰納的・演 繹的推論		内容の価値に即した内 発的動機、教科への関 心・意欲
	での学習	知識の有意 味な使用と 創造(使え る)	見方・考え方 (原理と一般化、 方法論) を軸と した領域固有の 知識の複合体	知的問題解決、意思決定、 仮説的推論を含む証明・ 実験・調査、知やモノの 創発(批判的思考や創造 的思考が深く関わる)	プロジェクトベースの対 話 (コミュニケーショ ン) と協働	活動の社会的レリバンスに即した内発的動機教科観・教科学習観 (知的性向・態度)
	学習の枠づけ ちが決定・再	自律的な課 題設定と探 究 (メタ認 知システ ム)	思想・見識、世 界観と自己像	自律的な課題設定、持続 的な探究、情報収集・処 理、自己評価		自己の思い・生活意 (切実性)に根差した 内発的動機、志やキャ リア意識の形成、
$\left\{ \right\}$	†自体を学習者た 再構成する学習	社会関係の 自治的組織 化と再構成 (行為シス テム)	人と人との関わりや所・文化によっての選出でいての意識、や明文化に共同体の運営方法に関する方法	生活問題の解決、イベント・企画の立案、社会問題の解決への関与・参画	人間関係と交わり (チームワーク)、ルールと分業、リーダーシップとマネジメント、争いの処理・合意形成、学びの場や共同体の自主的組織化と再構成	社会的責任や倫理意識 に根差した社会的動榜 道徳的価値観・立場性 の確立

そこで、私個人として一番希望しているのは、高等学校の調査書の比較可能性を高めるような工夫を考えていただきたいということです。国際バカロレアやイギリスの GCSE、GCE などであれば、各学校で先生方が取り組ませておられるパフォーマンス課題の評価の比較可能性を確保する仕組みがあります。例えば先生方が、学校で生徒たちが取り組んだ作品について、採点をする。作品と共に点数を試験実施機関に提出をして、抽出調査を行い、採点が適切に行われているかということをチェックするといった仕組みです。

日本においても、そういう事例を参照しつつ、制度設計を行っていただけるとありがたいと思っています。例えば、生徒たちが生み出したレポートなどを県単位などで集めて、シャッフルして先生方が採点していくというような取組も考えられるのではないでしょうか。そういう取組というのは、単に入試だけではなく、教員研修としての意味もかなり大

きいものがあると期待しています。

以上です。後半の討論の中で、いろいろとご批正いただければと思っております。どう もありがとうございました。

●質疑応答

- 【山地】どうもありがとうございました。それでは、西岡先生へのご質問を受けたいと思います。どなたかございますか。
- 【フロア】特色入試の募集要項の内容を見ますと、一般入試よりも優秀な人を求めたいという気持ちがありありと出ていて、入学者のアンケートの回答を見ても、結構「しんどい」と。来年度、再来年度と、出願者が継続的にたくさん来てくれることを期待したいわけですけれども、その辺の見込みはどうなのでしょう。ちょっと厳しいのではないかなという気がするのですが。
- 【西岡】どの角度から答えればいいでしょうか。私たちはいろいろな大学の AO 入試の先進事例を調査しましたが、やはり AO が成功していると評価している大学は、入った学生が劣等感を持っていない。逆に、AO が実はうまくいっていないと判断される大学の場合は、学生たちが大学に入ってから「僕たちは AO でしか入れなかったから」というスティグマを持っているということを聞きました。私たちはそういうAO にはしたくない、入った学生たちが「俺たち AO 組だから」と思ってしまうのは嫌だなと思っていました。

一般入試と比べてレベルが高いか低いかというと、私たちとしては一般入試の二次試験とは、ちょっと違う種類の力を試していると考えております。ですから、実際に入った学生同士の間で、周りの一般入試で入った子たちは「やはりこの特色入試で入った学生たちは、自分たちにはないような部分があるな」と、ちょっと尊敬の念を持って見ている反面、特色入試で入った学生たちの方は、「でも私たちは一般入試で入れたかどうか分からない」というような、不安とまではいきませんが、「頑張らなくちゃ」というような意識を感じさせるような対話も聞いております。そういう点では、学生が持っている力の多様化といいますか、相互に影響し合いながら高めてくれるようなコミュニティを作っていけたらというようなイメージで作っています。

継続的に受験に応募してくれるかということに関しては、まだ2年目が行われていませんので。この入試は、志願者に殺到されても無理なのですね。だから、25人ぐらい来てくれると一番いいのですが。そういう印象です。

【フロア】教員が協力して全学でやるというのが、すごくうらやましいというか、普通は 執行部が言っても、なかなか言うことを聞いてくれない先生が多い中で、さすが京 都大学だなと思いました。

【西岡】ただ、定員を見ていただきますと、学部によって実はかなり温度差があります。 全学の方針は1割でしたけれども、定員1割を採っていない学部もありますし。四 百何十人教員がいる工学部と、三十何人しかいない小規模教育学部では、導入する ときの教員にかかる負担が全く違うのですよね。

【山地】ありがとうございます。病気で倒れた先生方も結構いるというお話でした。なかなか覚悟がないとできないなと思いました(笑)。西岡先生,ありがとうございました。



〔報告 2〕

学習科学からみた資質・能力の育成と評価

白水 始(東京大学大学総合教育研究センター教授)

東京大学 大学総合教育研究センター教授,同大学大学発教育支援コンソーシアム推進機構機構長。 名古屋大学大学院文学研究科 修士課程修了。中京大学情報科学部助手,同大学情報理工学部准教 授,国立教育政策研究所 初等中等教育研究部総括研究官を経て,2016年4月より現職。博士(認 知科学)。専門は学習科学,認知科学,教育工学。主著に『資質・能力 理論編』(共編著,東洋館 出版社,2016年),「新たな学びと評価は日本で可能か」『21世紀型スキル』(共著,北大路書房, 2014年)など。



【山地】それでは、白水先生のご報告に移りたいと思います。白水先生は3月まで国立教育政策研究所におられ、4月から東京大学の大学総合教育研究センターに移っておられます。といっても、国立教育政策研究所のフェローとしてそちらにもオフィスをお持ちのようなので、実質的には両方を兼務しておられる形かと思います。日本での学習科学研究の草分けと言ってもよい方で、今後、東京大学の入試にも関わっていかれるとのことです。本日は、学習科学と認知科学、さらには入試とのオーバーラップしたあたりをお話しいただけるかと思います。

では、白水先生、よろしくお願いいたします。

Consortium for Renovating Education of the Future

2016年9月10日(土) 大学入試センターシンポジウム2016 「高大接続における学力評価の最新動向」 15:25-15:45 @東京工業大学

学習科学からみた 資質・能力の育成と評価

しろうず はじめ

東京大学大学総合教育研究センター教授 大学発教育支援コンソーシアム推進機構(CoREF)機構長 国立教育政策研究所フェロー

白水です。よろしくお願いします。「学習科学からみた資質・能力の育成と評価」という話なのですが、簡単に言いますと、子供たちが授業の中で対話を通して Deane 先生がおっしゃるような読み書きのスキルを身に付けていく授業を本当にやっていこうというものです。その中で、一つ大きな違いは、それがどんなスキルの発達を見せるかという点についての考え方です。ラーニング・プログレッションズというのがどこか天の上にあって、それに従ってルーブリックを決めて、だから今日の授業はこうすべきだ、評価はこうすべきだというようにシステマチックに決まるわけではなくて、一つ一つの授業で、先生がルーブリックを手作りして、それを積み重ねてラーニング・プログレッションズを作っていく。そんなことをやっていくと、作問の世界と、授業で子供たちの学力を伸ばすということが融合してくるのではないかという話をしたいと思います。

●学習科学における評価の考え方

出典:高等学校における「多様な学習成果の評価子法に関する調査研究」 平成25年度 - 平成27年度(東京大学CoREF)

Consortium for Renovating Education of the Future

学習科学からみた資質・能力の育成と評価

- 評価をどう考えるか:理論的前提
 - 1. 評価とは、児童生徒の認知過程を観察可能なデータから解釈すること ⇒「評価は常に不完全。合わせ技と手法自体の継続的改善が必須」
 - 評価は、個人間の優劣の評定より、まずは潜在的な学ぶ力を引き出し、 伸ばしながら見取ることに使うべき⇒「評価が変化を捉え、変化を生む」
 - 授業と評価と資質・能力の向上は一体的に進む⇒「入試センターだけでなく、小中高大の教員も子どもも、一人ひとりが評価の主体であるべき」
- 従来の評価観及び評価手法の刷新が必要

従来:

刷新:

総括的評価-形成的評価

絶対評価 一相対評価

客観評価 一主観評価

個人評定一授業評価-教育政策評価

一人ひとりを伸ばしながら 見取るために学習過程を 捉える評価へと刷新する

(あらゆる手法はこの目的のために総合利用)

学習科学から見て資質・能力の育成と評価をどのように考えているのか。学力の3要素を一体的に育成・評価してこそ、初めて意味があるのだろうと思います。そういう意味で、学力は非常に総合的で複雑なものになってきます。ですから、その実態を知るためには、多様な文脈で、多様な手法を用いて、継続的に観察して、たくさんのデータを手に入れて、解釈を深めていく。そういう評価の合わせ技ができるといい。そう考えると、多分、西岡先生がおっしゃるケースのように倒れてしまう先生方だけにお任せするのではなくて、小中高大、いろいろな人が関わって評価の主体になっていく、一人一人が必要なレベルの評価のリテラシーを身に付けるというのが、今回の改革の成否を左右する大きなポイントではないかと思います。

Consortium for Renovating Education of the Future

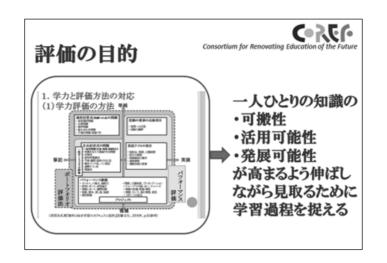
新しい教育と評価へのニーズ

- ●社会の情勢の変化
 - ●新規な課題に対して新しい解を生み出したい
 - ●体験や知識の断片を組み合わせて解を作る
- ●認知科学・学習科学の進歩
 - 個人の経験則が「対話」を通して科学的概念へと変化する仕組みの解明
 - ●一人ひとりの知識が個別で多様であるからこそ、 対話が各自の知識の適用範囲を拡げ、社会全体 の知的レベルを向上させる

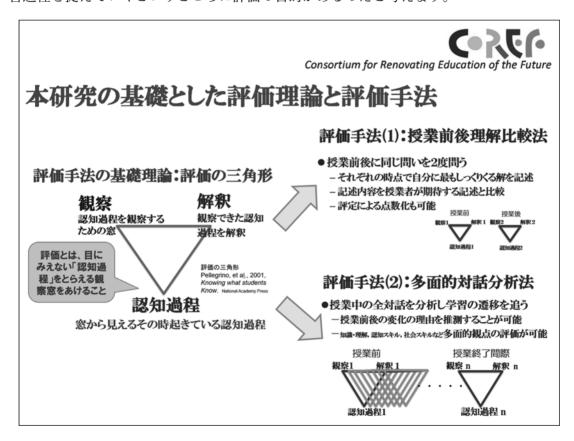
それでは、なぜこんなことが必要になっているのか。社会が変わってくるからなのだろうと思います。社会が変わってくるので、新規な課題に対して新しい解を生み出せるような人材を育てたい。そのためには、今までに分かってきている知識・技能のようなものでは解けない課題に対して、体験や知識の断片しかないのだけれども、目の前の情報と組み合わせて解を作るような人材が求められている。

では、そんなことが本当にできるのかというところに、Deane 先生に紹介していただいた認知科学、学習科学の進歩があります。一人一人が持っている経験則は多様なのだけれども、対話しているうちにだんだん科学的な概念に変わっていく。それを信じていいのだ、そこにメカニズムがあるのだということが、認知科学・学習科学から分かってきています。

なおかつ、その過程をもっと詳細に見てみると、本当に一人一人の知識というのが個別で多様だということが見えてきています。ラーニング・プログレッションズという、一本道で決まらないほど、本当に多様な道筋を通って子供たちは学んでいきます。だからこそ、対話してみると、いろいろ考え方が違うので、それぞれの知識の適用範囲が広がっていく。その多様性を大切にしながら評価をやっていきたい。

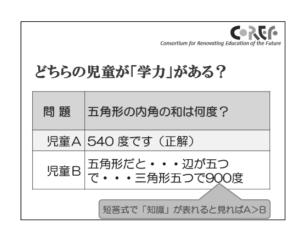


そのために、西岡先生が紹介してくださったようないろいろな評価方法を、目的としては右に書いたように使いたい。つまり、一人一人の知識が、可搬性(portability)、学んだ場から持ち出せるか。いろいろなところで使えるような dependability を持っているか。いろいろなところで作り変えて、新しいことを学んで発展させていけるか。サステイナブルなものになっているか。こうした知識の質が高まるように伸ばしながら、見取るために学習過程を捉えていくというところに評価の目的があるのだと考えます。

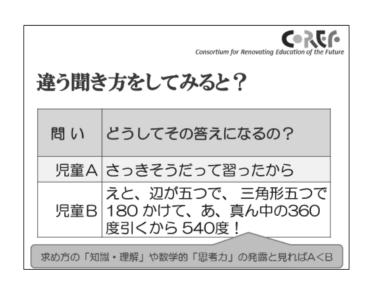


では、資質・能力以外の評価も含め、学習科学から見て評価というのはそもそも何をすることなのか、簡単にお話しします。簡単に言うと、子供の頭の中で起きている認知過程は見えません。何とか知りたいものを見るための観察の窓を開ける。窓から得られたたく

さんのデータを解釈して、こんなふうになっているのではないかという妥当な解釈へにじ り寄っていく。これが評価だろうと思います。



簡単な例でいきますと、次のどちらの児童が「学力」があるかと聞かれたら、先生方、どうでしょうか。「五角形の内角の和は何度ですか」と聞かれた小学生 A は即座に「540度です」と、元気に正解を言ってくれる。児童 B は「五角形だと……、辺が五つで……三角形五つで900度?」みたいなことを答えてくる。



もし、短答式で「知識」が表れると見ると、A 君の方が B 君よりできると見られるのですが、先生がもう一歩進んで、「どうしてその答えになるの?」と聞いてみると、A は「さっきそうだって習ったから」と答える。それに対して B は、「えっと、辺が五つで、三角形が五つで、180 を掛けて 900 度。あ、真ん中の 360 度を引くから、540 度」と、恐らく五角形をピザのように切って、「180 度が五つ分。その中の真ん中の 360 度は要らないから、540 度!」というように考えて答えてくれた。こんな風に答えてくれると、どうも「知識・理解」、数学的な「思考力」、その場で考える力のようなもので言うと、B の方がありそうだということになります。

Consortium for Renovating Education of the Future

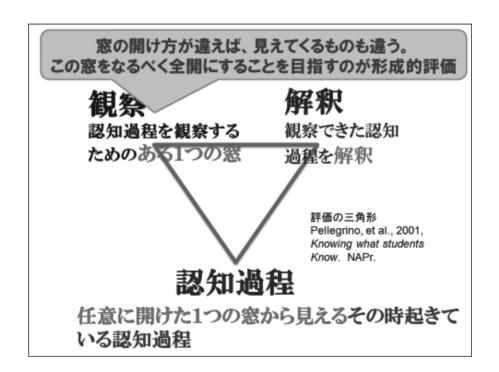
知識・理解の「評価」と言ったときに

- ●答えを教えた直後にYes/Noクエスチョンや穴 埋め問題で高い点がとれるか
- ●しばらく経ってから、抜き打ち的に発展問題や 原理的な理解を問う問いを出されたときに答 えを考えられるか

穴の開け方が違えば、見えてくる「理解」も違う

→伸ばしたいのはどんな理解か

そう考えると、知識・理解の「評価」と言ったときに、答えを教えた直後に Yes/No を問う問題や穴埋め問題で高い点が取れることを指すのか、それとも、しばらくたってから、抜き打ち的に原理的な理解や活用を問う問いに答えを出せるのか、そのどちらを「知識・理解」と言っているのかによって、観察窓の開け方も違ってくるわけです。



ただ、窓の開け方が違えば、そこで見えてくる理解も違うのだとしても、例えば記述式で理由を聞けば、もちろん短答式よりはベターかもしれないが、それがベストかどうか分からない。その意味で、ベストの窓を一つ開けることを目指すのではなく、窓をなるべく全開にすることを目指そう。それが私にとっての「形成的評価」というものです。

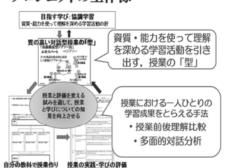


本研究の基本的デザイン

授業づくりプロジェクトとしての評価研究

- 1. 学習科学に基づいて, 資質・能力を「使いながら育てる」授業の型と, その特徴を活かした評価手法をセットで提案
- 2. 実践者が手法を使ってみて,知 見を出し合い,授業と学びにつ いての知見を深める研修を教委 と連携して実施※
- ⇒一人ひとりの授業者が子ども一人 ひとりの学びを見取りながら伸ばす 「形成的評価」を日常化

プロジェクトの全体像

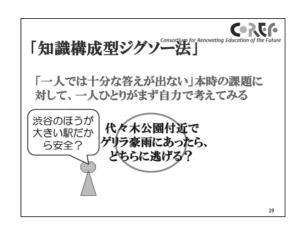


※研究規模:20以上の自治体、 1000名上の小中高教員と連携、 70,000名以上の児童生徒に実施

そのために、学習科学から思考力や協調問題解決能力の育成・評価に向けて、こんなことをやっております。

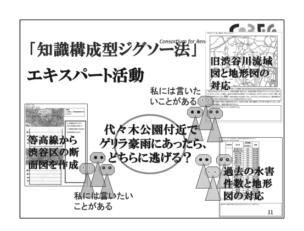
まず、学習者は思考力の資質・能力を潜在的に持つのではないかと仮定します。しかし、 その発揮は、まさに文脈に依存しますので、その実態をできるだけ的確に見取るために、 資質・能力を使う必要性のある環境を整えて、スキルを発現させていく必要があるだろう。 そのために回しているサイクルが、少し図が小さいのですが、こういうものです。

資質・能力を使って理解を深めるような授業の型というのを先生方にお渡しする。けれども、問いを用意して、資料を用意するのは先生方。それでやってみた後に、一人一人の児童生徒がどれぐらい学んだか。授業前後でどれぐらい理解を深めたか。そして、その間にどういう対話をしているかというのを全部分析してみると、実際に一人一人の子供がどんなふうに知識の質を伸ばしていくかを先生方が見て取る「形成的評価」ができるのではないかと考え、このようなサイクルを、年間7万人ぐらいの児童生徒に実施してデータを取っております。

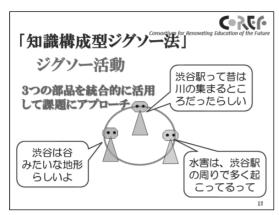


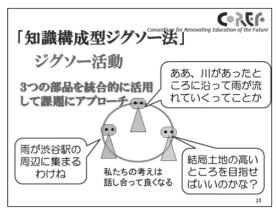
では、どういう授業の型かというのがこれです。最初に、「一人では十分な答えが出ない」 課題を先生が考えてみる。例えば高校生の地理の時間に「代々木公園付近でゲリラ豪雨に 遭ったら、どちらに逃げる?」というような問題を出してみます。

対象は埼玉の高校生なので、何となく「渋谷駅の方が大きいから安全かな」みたいなと ころから始まるのですが、これを書き留めた後で、「いや実は、ヒントがあるんだ」と、三 つの資料をクラスの中で分担して読んでもらいます。まさに、Deane 先生の文章の読み解 きのようなものに当たります。

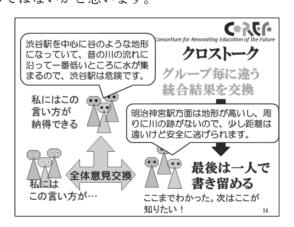


まず、等高線から渋谷区の断面図を見てみると、「結構、渋谷って低いのね」ということになります。地下鉄銀座線が渋谷で地上に出てくるように、渋谷は断面図で見たら結構低いのです。それから、渋谷川というのが昔ありまして、現在も地下を流れています。それと今の地形図を対応させてみる。また、過去に渋谷で起きた水害の件数。地下の方で、内水氾濫がたくさん起きているというような事実もある。このような三つの資料をそれぞれで読み込むと、一人一人言いたいことができてきます。これが、例えば資質・能力の一つであるコミュニケーション能力を賦活します。





これを読み込んだあとで、席替えをして、違う資料を読んだ3人で、まさにジグソーパズルのピースを合わせるように情報を合わせてみる。「渋谷は『谷』っていう漢字が入ってるんだけどさ、地形が谷みたいなところだったらしいよ」「渋谷駅って昔は川の集まるところだったらしい」「水害は渋谷駅の周りで結構多く起こってるって」と、それぞれの資料から大事なポイントを話してみると、お互いに、「ああ、じゃあ、雨が渋谷駅の周辺に集まるのね」「川があったところに沿って雨が流れていくのか」「結局、土地の高いところを目指さなきゃいけないんじゃないかな」みたいなことを言い、だんだん、自分たちの考えを話し合いでよくしながら、答えににじり寄っていくのです。これがまさに協働、コラボレーションの基礎になるのではないかと思います。



では、こんなふうにできた答えは、同じ問いに同じ3資料を使って考えた3人で一致するのでしょうか。グループでクロストークをやってみますと、「渋谷駅を中心に谷のような地形になっていて、昔の川の流れに沿って一番低いところに水が集まるので、渋谷駅は危険です」と言う班もあれば、「明治神宮駅方面の方が、地形が高くて、周りに川の跡がないので、少し距離は遠いんだけど安全に逃げられそうです」と、それぞれ同じような肝はつかみつつも、表現が違います。その表現を聞いているうちに、一人一人で「ああ、こういう言い方もあるんだ」「こんな言い方もあるんだ」と、情報を聞きとめて、最後にもう1回書き留めてみると、「最初はここしか書けなかったけど、今はこんなことが書けるようになった。次はここが知りたいぞ」というような、大人から見たら当たり前でも子供から見たら新しいことを作り出したというような、イノベーションの基礎体験ができます。

資質・能力を使う機会 = 見取る機会 - 知識構成型ジグソー法の授業の場合 -

	ワークシートへの記入 <書く>	対話/発表 <話す>	
はじめに	記入	一多面的	
エキスパート活動	記入 将数	対話 対話 対話分析法	
ジグソー活動		解比較法 対話	
クロストーク		発表	
終わりに	記入	_	
データの記録	・ 児童・生徒が紙に書く 将来的には:児童・生徒がタブ レット/PCへ手書き、文字認識 上、テキスト化	10111201210 27 222000711	

授業の型に組み込まれた機会を活用し、資質・能力がどのように使われているかについて、真正なデータを得られる

15

こんなふうに、協調の中で、読んだり書いたりするアクティビティを通して、子供たちはまず同じ紙の上に授業の最初の解答を書いて、最後に、下の方に授業の後の解答を書きます。自分たちでやはり「量が増える」「質が上がる」ということを見た後で、同じグループで3人交換したりすると、同じグループで同じ結論にたどり着いたのだけれど、微妙に表現が違って、ポイントの置きどころと、次の疑問点、見えてくるものが違うことが感じ取れるのです。まさに、アセスメントをその場でラーニングとしてやっていくことができます。

いま,ワークシートの手書きをどのぐらい自動認識できるかといったことに挑戦したり,全部のグループにマイクを置いて,会話を全部記録して,先生方の評価を支援しております。

研究成果1:授業前後理解比較

授業前後理解比較の例

●高1 地理 の事例データ 「代々木公園でゲリラ豪雨にあったら、どう逃げる?」

	生徒A 神宮→神宮	生徒B 淡谷→明治神宮	生徒€ 流谷→流谷
前	明治神宮駅方面が 高いから。洗谷方面 は水たまりやすそう。	JR駅の線路は上に はしっているから開 で止まることはない と思うから。	渋谷駅の方が低いと 思うけど水の流れに さからって歩くのはき つい。
後	最短ルート、地形が高い(明)。渋谷駅を中心にして谷になっている!川の近くで水害が多く起こっているから、川が近くにない明治神宮駅の方が安全。	明治新宮駅の方が 流谷に比べて標高 が高く最短ルートに 比べればすこしかか るかもしれないが安 全に歩いていくこと ができるから。	流谷のほうが内水氾濫の被害が少ない。 流谷は明治神宮駅と 同じく低いけど最低 のところではないか ら。



評価規準の設定

授業のねらいに基づき,期待する解答の要素(教えた い教科内容の本質),その他使われそうな情報を同定

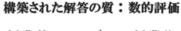
経験		経験や感覚	
	地理1(expA)	土地の高低差	
	地理2(expB)	河川の存在	
地理	地理3(expC)	過去の水害発生件数	
	地理4	現在地からの距離	
	地理5	施設、環境	
ほか		経路、金銭など、経験と地理的 情報以外の要素	

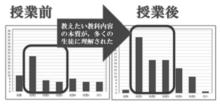


€756

使ってみた手広え: どの実践でも評価できそうなもの

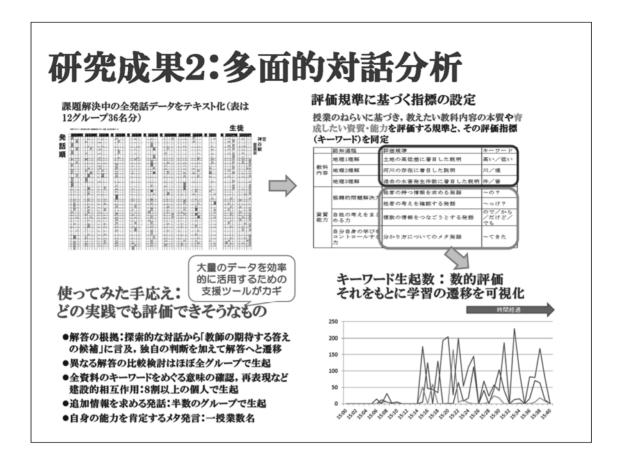
- ●生徒の学力 数的/質的 評価規準の設定が カギをにぎる
 - ●構築される解答の質
- ●解答構築過程で使われた資質・能力の想定
- ●自分なりの問いや着眼点(ポテンシャル)
- ●授業の成果や課題
 - ●生徒の既習度
 - ●生徒の学習過程の多様性、時間のかかり方





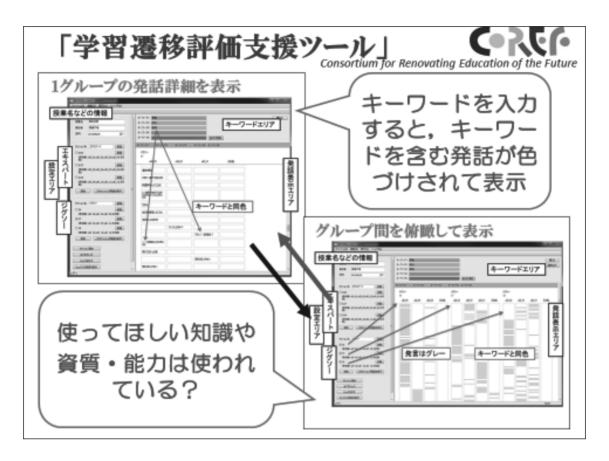
これが簡単な授業前後の3人の生徒の理解比較です。そうすると、Xは、神宮に逃げる という答えは授業前後で変わっていないのだけれども、書くものの質が変わってきます。 Y は逃げる先が渋谷から神宮に変わって、Z はなぜか渋谷から渋谷という、先生からする と期待しない答えにスティックしている。では、この解答を、先生が自分の教室の中で理 解を見取る基準を設定して、どうやって評価していけるか。

例えば資料には、土地の高低差、河川の存在、過去の水害発生件数があるので、やはり それを入れてほしいということで、事前と事後の解答をポイントに分けて分析してみます と、授業の中でねらいが達成されたかどうかという「構築された解答の質」で、確かな評 価が一つできるようになります。



それでは、途中、子供たちはどんな解答を作っていたのかを知りたくなりますので、その間をこんなふうに全部発話を分析して、今のところは手起こしをしているのですが、今後は自動化しながら、その手起こしを加えていったときにどんなに役立つものなのかというのをやっております。

これが一人1行で、これが1グループ。この縦で1分ぐらいしかないのです。もう、見る気が失せるくらいデータ量が多いのですが、これはそのままだと使いものになりませんので、こんなふうにキーワードを入れたら、該当箇所が明るく光るようになっております。例えば「神宮」と言ったか、「谷」と言ったか、「駅」と言ったかと入れてみると、そこが見えるようになって、先生方の評価の力も上がっていくのではないかと思います。キーワードを変えながら、授業が進むにつれて子供たちがどんなふうに表現を変えていくか、ということが日々の授業の中で評価できるようになると、先生の評価しやすさがだいぶ上がってくるのではないでしょうか。



これをやってみますと、先ほどの渋谷の例ですと、例えば土地の高低について、発話の中でどんな言葉で言っていそうか。最初は「高い」「低い」と言っていそうだと検索するのですが、実際に調べていくと、子供たちは、「上っていく」「下りていく」という、まさに自分から見たような表現で高低を表現していたりします。

あるいは、協調問題解決能力で、他者の考えを確認するというコンポーネントがあるのだったら、例えば実際の発話で、「つまり」という言葉が出るのではないかと大人は考えて、「つまり」と打ち込む。しかし、「つまり」と言いながら話し合う高校生はいないのですね。「で」「じゃあ」「だったら」などという高校生の口語表現の中で、「じゃあ、何々ってこと?」という言い方で、彼らは仲間の考えを確認していることが見えてきます。

では、先ほどの生徒 Z、「渋谷から渋谷」で変わらなかった子供は、事前事後記述だけ見ていると考えていないように見えるのですが、実際にそうなのか。そこで、この子の発話に「?(はてな)」というキーワードを入れてみて、その「?(はてな)」の周りで彼がどういう発話をしているかを見に行くと、他の担当の生徒が、「いやあ、富ヶ谷の 2 丁目は近くに川が通っているようには見えなかったのに、なぜこんなに浸かったかと思ったら……」と言ったら、この Z 君は、「ここら辺に川があるよね。宇田川があるよね」と言いながら、「だから危ないのかな」というようなことを言っていろいろ考えていました。ですが、先生が「じゃあ、あと 1 分でどちらに入れるか発表してください」と言ったことで、「じゃあ渋谷一択で行こうか」というようにいきなり考えが変わって結論を出していました。そう考えると、事前・事後に書いているものだけから読み取れないくらいに複雑な思考過程を通じて子供たちは考えているということが分かりました。

●形成的評価の意義

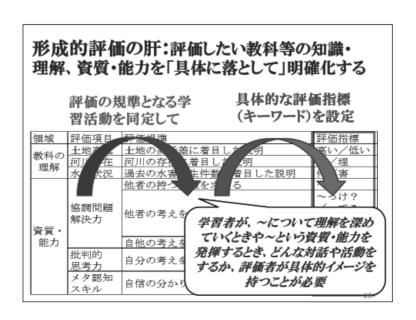
評価リテラシーを上げる教員:

一教員研修の参加者の声より

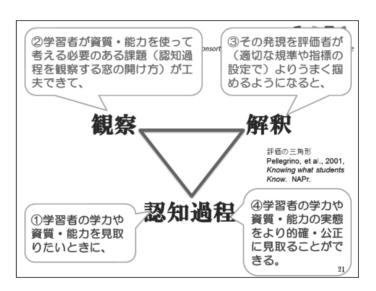
●前後理解比較

- 生徒の学びを見取る力が向上
- →「前後理解比較をすると、途中の変容や姿を印象的に見とれる自分がいてびっくりしました」(ベテラン教員)
- 生徒の具体的変容に基づいて授業デザインを見直せる
- →「課題設定や規準が明確でないと、変化は見とれない」
- 多面的対話分析
 - ●よりよい評価規準&指標を検討できる
 - →「確認 (「なるほど」「そうか」)、疑問 (「どうして」 「なんで」) など 会話の変換点を捉えるキーワードを有効活用すれば生徒の相 互作用のパタンがわかりそう」
 - 郵価をとおして暗黙の前提としていた学習観を見直せる
 - →「班によってわかっていくときに使う言葉が違う」

こんなことをやりながら、とにかく授業の中で取れるデータを増やしていく。Deane 先生はシステムを使っておられましたが、まさに、それを授業の中で自然な形でやれるように、書いたり話したり生徒には自由にやってもらって、システムの力でそれを書き起こして、彼らの学びのプロセスが見えるようにする。そうなりますと、先生方の中にも、「いやあ、前後を比較してみると、結構、子供が最後にどこまで行ったかを取るよりも、変わっていく姿というのが見られるのですよね」とか、非常に大事だと思うのですが、「課題設定や基準が明確でないと、変化というのは見取れませんよね」というような気付きが生まれてきます。あるいは、「確認や疑問の言葉の会話の変換点を捉えるキーワードを活用すると、生徒の相互作用のパターンが分かりそう」とか、「班によって分かっていくときに使う言葉が違う」とおっしゃっていただけます。だから、ラーニング・プログレッションズというのは、本当に多様に、具体的な文脈の中でどのように表れるか、やはりたくさん用意しておかなくてはいけないのではないかと思います。



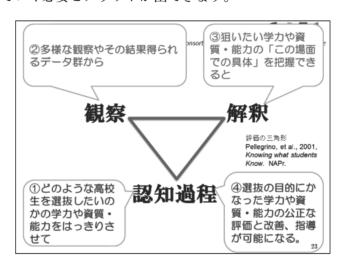
それから、非常に基本的なことの発見なのですが、日本でも海外でも、これまでは学習評価と言えば、書いた文字に関する評価がメインだったので、話し言葉についてはなされてなかった。あったとしても、しっかり何遍も練習したプレゼンのきれいな発話しか評価していなかった。話しながら考えを進める過程で、一体、人はどういう言葉を使うか、そこにどういう力が表れているのかという観点での評価研究が、全く未開の大地なのではないかと思います。



学習者の学力や資質・能力を見取りたいときには、学習者が資質・能力を使って考える必要のある課題を工夫して、その発言を評価者がうまくつかめるようになってくると、学習者の学力や資質・能力がよりしっかり、まさに公正に見取ることができるようになっていくのではないかと思います。このサイクルを何遍も回していく。

月	多列化(評定)への	活用可能性		
	生徒X 回答「神宮→神宮」	生徒Y 「渋谷→渋谷」		
理由 (前)	明治神宮駅方面が高いから。渋谷 方面は水たまりやすそう。	渋谷駅の方が低いと思うけど水の 流れにさからって歩くのはきつい。		
解答 構築 過程	「道筋を考える」という次の課題 を発見して取組み、道筋について の自分の考えを土地の高低という 観点から点検	河川の存在、土地の高低、水害の 有無、距離などを多角的に検討し、 最終的に「距離」を優先して「渋 谷」を選択		
理由(後)	地形が高い(明)。渋谷駅を中心 にして谷になっている!川の近く で水害が多く起こっているから、 川が近くにない神宮駅の方が安全。	渋谷の方が内水氾濫の被害が少ない。 渋谷は明治神宮駅と同じく低いけど最低のところではないから。		
内容	教員や大学が「高等学校地理の履修 学習内容と日常生活への結び付 内容の習得にプライオリティを置くなら、 けや統計データの読み取りに主 生徒X 眼を置くなら、生徒Y 22			

しかし、これは教育であって、入学者選抜に直接関係ないのではと思われるかもしれません。その点については、こんなふうに考えます。生徒 X、Z がそれぞれどういう解答ですかと聞くだけよりも、恐らくそう解答した理由を書いてもらうとよく本人の考えが分かります。さらに、話し合っている真っただ中を見てみると、X というのは結構「真面目に地理の内容を学びます」というタイプなのだということが見えてきます。そういう子供に入学してほしいというのだったら、X を採ればいい。一方、Z は、日常生活に結び付けながら、自分なりの考えを作っています。そんな生徒が欲しいなら、生徒 Z を採ればいい。そう考えると、このようなプロセスデータは、入試に使える豊富なデータだということになるのではないかと考えます。だからこそ、その豊富なデータを活用できるリテラシーを大学側も身に付けていく必要とメリットが出てきます。



●まとめと高大接続改革への含意

評価の原則と今後の課題

Consortium for Renovo



● 評価の3原則

- 人は元来、自分で考えて学ぶことが得意で、状況さえ整えば、学ぶ力を 発揮できる。だから、学習者の発揮する能力は学習環境(文脈)が変わ れば変わる。
- 2. 評価とは、目に見えない認知過程を、目に見えるデータ(記述や発話)に 基づいて、解釈することである。だから、評価の仕方や規準が変わると、 見えてくる学力や資質・能力の様相も変わる。

● 今後の課題

- 1. 質の高い評価を日常的に行うための環境 基盤の整備(ICTによる対話分析を含む)
- 2. 高大接続改革、教育課程改革との緊密な連携
- 3. 大学や産業界、一般社会の学習観・評価観 の捉え直しのためのアプローチ

評価の3原則というものを示します。人は、子供は、状況さえ整えば、結構学ぶ力を発揮できるので、学習環境、まさに contextualization をしっかりやってあげると、子供の力が発揮できます。その評価の仕方や基準が変わると、見えてくる学力や資質・能力の様相も変わります。ですから、どこかで私たちは「個人の評価」から頭を切り替えて、状況一学習環境をうまくデザインして、そこを評価することを考えていく必要があるのではないか。その中で、どんな環境の中で子供が育っているかを見取っていく必要があるのではないかと思います。

最後に、今後の課題として、質の高い評価を日常的に行うための、学習データの収集・ 分析のための ICT と環境基盤を整備していきたいと思っております。

それから、2番目として、高大接続改革、教育課程改革と緊密な連携をしながら、このような授業評価を位置付けたいと思っております。

高大接続改革へのささやかな示唆

- ●児童生徒の全授業における学習過程を記録し評価できる到達点から逆算して、現時点の接続を議論できないか
 - ELSIの観点も加味して(「窓を開ける権利」を学習者に付与する)
- 一回性テストでも観察の窓を全開にするフィールドテスト やCBTを多用して、評価の三角形の質を高められないか
 - センター入試の思考発話実験(白水,準備中)
 - 全国学力・学習状況調査問題の協調問題解決実験(Shirouzu et al., 2015)
- ◆小中高での「実績あるAL」のテストへの活用、及び大学・ 社会でのauthenticなプラクティスの予行演習としての高大 接続をデザインし、全人生段階を緊密に連携できないか
- 評価主体が自らリスクを取ってAPを定め、データを総合的に 活用できないか
 - センター入試のデータを大学側が自らの基準で採点・活用する等

2

この2番目だけ詳しく申し上げますと、この私たちのアプローチ、あるいは西岡先生のポートフォリオ評価のようなことをどんどん進めていくと、最後のゴールとしては、恐らく私が死ぬときぐらいまでには、児童生徒の全授業における学習過程を記録して評価できる、そういう到達点が技術的にはできるようになってきます。そうすると実は、今、ポートフォリオで出しているのが、どれくらい「なりすましの自分」というか、「理想の自分」を出していて、隠しているところに「普段の自分」があるということが見えてきたりする。そうすると、実は、こういう一回性のテストというのは、受験者と私たちの間である種契約を結ぶような形で、今回はここの窓だけ開けて、私の力を発揮しますよというような場面として捉え直せる可能性がある。つまり、いままでテストに苦しんでいたのが、将来はテストがむしろ救いになるのかもしれないという逆説的状況です。

そういう意味では、この到達点から逆算して、一体、今、私たちはこの改革で何をした いのかということを考えていけると将来につながっていくのではないかと考えております。 次は、一回性テストでも観察の窓を全開にするフィールドテストや CBT を多用して、評価の三角形の質を高められないかと思っております。

例えば、センター試験をやっている最中に高3生は何を考えているかというのを取らせていただくために、高3生に、センター試験問題を解いている最中に「今、何考えてる?」というようなことを聞きながらフィールドテストをやってみると、果たして現在の小問や、選択肢の在り方で弊害が本当にどれぐらいあるのかが見えてくるのではないかということです。

それから、全国学力・学習状況調査で、算数の「割合」についての問題は小学生の3割しかできず、あとの7割の子はできないというときに、例えば一人で解き終わったテスト問題を、同じくできていないお隣の子と話し合ってもらうと、子供たちは本当に割合が分かっていないのか、それとも問題文が読めていないのかが分かります。結論から言うと、かなり問題文が読めていないというか、問題文を読み飛ばしているということが見えてくるのです。本質的な概念を二人で話し合ううちに、「ああ、ここら辺につまずいていた。こういうことだったんだ。こういう状況だって分かれば、結構問題は簡単だね」というような流れで子供たちが解いていきます。その子供たちのポテンシャルをうまく測るためのテストになっているかということを問い直していけるのではないかと思います。

最後は、小中高での実績あるアクティブ・ラーニング、仮説実験授業、知識構成型ジグソー法など、もう既にいろいろなプラクティスがありますので、それを模したような形でテストに活用していくような方向。さらに、それが大学・社会で求められる実践の予行演習となるように高大接続をデザインして、全人生段階を緊密に連携していくということも考えられるかと思います。

まとめると、学校教育で育てた力が、適切な接続を通して、社会・大学で求められる力につながっていっているか。そこをデザインしていくことを考えていくと、やはり高校の 先生だけではなくて、いろいろな方が一体としてやっていけるとよいかと思います。

なお、付言しますと、東京大学法学部の推薦入試では、昨年度、グループディスカッション問題で、受験生の論理的思考力・発想力・コミュニケーション能力・チームで作業する能力などを審査するために、難民危機の望ましい対処方法について、「先進国が難民を積極的に受け入れるべきだ」「難民キャンプを作ったらいいのではないか」「そもそも難民が出ないように、国連安保理の決議に基づいた武力介入を行い戦闘を調停させる必要がある」といった三つの案のどれがいいかというのを、55分間、グループで議論させました。

このような形態の入試の中で、実際に子供たちがどんなことを話したり書いたりするか。 知識量と測りたい力との兼ね合いについて、法学部の先生がどういう解釈基準を作って評価していくか。その解釈結果が高校までの教育経験をどうやって反映しているか。成績が果たして大学進学後の活躍につながるか、その predictability(予測可能性)はどのぐらい有力か。こうした点をしっかり検討できると、入試が適切な接続になっているのか、そこの窓はどう開けるべきかを考える一助になっていくのではないかと考えております。

以上で終わります。ありがとうございました。

●質疑応答

- 【山地】どうもありがとうございました。受験生がセンター試験を解いているときにどん なことを考えているか、などというのは実に興味深いですね。フロアの方でご質問 等があれば受けたいと思いますが、いかがでしょうか。はい、お願いします。
- 【フロア】素晴らしいご講演をありがとうございました。私は昭和 34 年生まれで、まさに典型的な昭和の地方都市での教育を小学生から受けて育ってきたのですけれども、白水先生に担任してもらいたかったくらいの気分です。今後ともますますこの感じで、科学的にデータを積み重ねながら、いろいろな角度から子供たちの評価をしていただいて、それが後々この国の未来を作っていけるよう頑張っていただければと思います。ありがとうございました。

【白水】頑張ります(笑)。

【山地】それでは、他に特にないようでしたら、次のご報告に移らせていただきます。白水先生、どうもありがとうございました。



〔報告 3〕

コンピュータで実施するテストの採点をめぐる課題

石岡 恒憲(大学入試センター研究開発部教授)

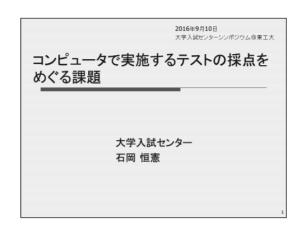
独立行政法人大学入試センター教授。東京理科大学大学院 工学研究科経営工学専攻 修士課程修 了。 株式会社 リコーソフトウエア研究所研究員, 同主任研究員を経て, 1998 年 4 月より大学入 試センター 研究開発部助教授。2011 年 4 月より現職。2000 年, カーネギーメロン大学コンピュータサイエンススクール(大学院)言語技術研究所客員研究員(文部省 長期在外研究員)。工学博士。専門は情報数理。日本語小論文自動採点システム Jess を開発。



【山地】石岡先生は、大学入試センターのスタッフですが、先ほども出ていた自動採点について長く研究してこられ、また自動採点のためのソフトウェアを開発して特許も取られています。本日は、記述式と言われる問題をコンピュータ採点する場合、どの程度のところまで可能なのかという点に焦点を置いてお話しいただくことになっております。

先生は大学入試センターに長くお勤めですが、もともとは株式会社リコーで研究をなさっていたということで、またカーネギーメロン大学などでも研究をされていた経験をお持ちです。そうした広い背景から現在の研究に至っているということですので、大変面白いお話を伺えるかと思います。

では,石岡先生,よろしくお願いいたします。



大学入試センターの石岡でございます。「コンピュータで実施するテストの採点をめぐる 課題 | ということで、発表させていただきます。

AIの三題ばなし

- 1. 囲碁 · 将棋
- » AIがトッププロに勝つ
- ▶ ディープラーニング(深層学習)
- 2. アメリカのクイズ番組(ジョパディ)
- ▶ ワトソンが歴代チャンピオン二人に圧勝
- 3. 東ロボ
- ▶ センター模試で偏差値57.8
- ▶ 私立大の8割でA判定
- ※ AI小説が星新一賞1次審査パス
- → AIが自然言語を(もう/間もなく)理解?

記述テストの採点に、AI(人工知能)を利用するという話があります。昨今、AI 絡みで話題になっているのは、この三つです。

一つ目は囲碁・将棋で、人工知能がトッププロに勝つという事実です。また、そこにはディープラーニング(深層学習)が使われているというのは広く喧伝されているところです。二つ目は、アメリカの「ジョパディ(Jeopardy)」というクイズ番組で、アメリカの IBM が作った人工知能ワトソンが、歴代のチャンピオン 2 人に圧勝したという事実です。大変話題になりました。三つ目は、東ロボです。「ロボットは東大に入れるか」プロジェクトは国立情報学研究所で行っているものですが、直近のセンター模試で偏差値が 57.8、私立大の 8 割で A 判定を取りました。

最近の話題で言いますと、4編の AI 小説が星新一賞の1次審査をパスしたということです。

このようなことから、人工知能が自然言語を既に理解している、または間もなく理解するのではないかと思われている方が多いと思いますが、このあたりの正しい認識、すなわ

ち何ができて、まだ何ができていないのかということをお伝えできればと思っております。

構成
□ 記述式問題のコンピュータ処理・採点の難しさ
■ AI利用の視点から
□ コンピュータでできる採点支援
※ 文科省方針の是非を論じない
※ 一研究者としての見解
※ エッセイタイプの採点は言及しない

今日のスピーチの内容は、大きく分けて二つあります。一つは、記述式問題のコンピュータ処理・採点の難しさを人工知能利用の視点から述べたいと思います。もう一つは、コンピュータでできる採点支援にはどのようなものがあるのかをお話ししたいと思います。なお、記述採点には、大きく分けて短答式とエッセータイプのものがありますが、今日はエッセータイプの採点についてはお話ししません。

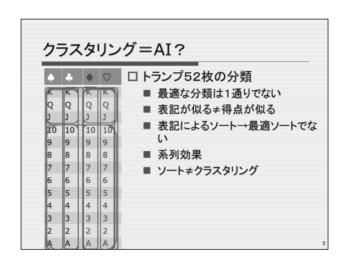
●記述式問題の自動採点の難しさ

囲碁・将棋の難しさ □ 探索量 ■ 将棋:平均合法手80;平均終了手数115→10^220 ■ 囲碁:10^360 ■ チェス:10^120 オセロ:10^60 □ 記述回答 ■ 15字:形態素5~7個の並び→5億~7億通りの表記 ■ 30字:(5億~7億)^2=10^26 ■ 60字:10^52 ■ 300字:10^260 字数が多くなれば解答パターンは↑類型化が困難

さて、人工知能(AI)が囲碁や将棋を行う上で問題になるのは、局面の判断など、いろいろあるとは思いますが、一義的には探索量の多さです。つまり、探索する枝の数が非常に多くなるということです。将棋の場合は、一手ごとに指せる可能手が平均80通りあると言われていて、一局の終了までの平均的な手数は約115ですから、80の115乗で、10の220乗ぐらいの探索量があるわけです。囲碁はもっと多くて10の360乗。チェスだと10の120乗くらいの数字になります。

では、記述採点ではどうなるかというと、たとえば、「グーグルで検索すると」というような句ですと、大体15字で、形態素で言えば、5~7個程度の並びです。これが日本語の

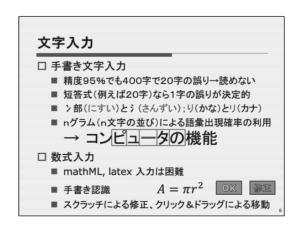
記述だと、5 億から 7 億あると言われています。これは Google のデータベースによるものです。すると、30 字だとこれの 2 乗ですから、10 の 26 乗、60 字で 10 の 52 乗、300 字で 10 の 260 乗になりまして、将棋を超えるわけです。字数が多くなると、解答パターンはこのように指数関数的に増えていって、類型化が極めて難しくなります。



そして、採点の類型化という観点から申しますと、採点にクラスタリングを使うという 話があります。これについて論じますと、今、トランプ 52 枚の分類を考えるとします。

普通はキングからエースまで 13 分類すればいいのですが、もちろんスペード、クラブ、ダイヤ、ハートのマークごとに四つに分けてもいいわけです。赤札と黒札というような二つの分け方でもいいです。絵札と数字という分け方でもいいです。絵札とマークとの混合、ミックスでもいいわけです。

何を申し上げたいかというと、最適な分類方法は一つではないということです。表記が似るということと得点が似るということとは必ずしも同じではありません。したがって、表記によるソート(並べ変え)は、一般には最適ソートではありません。また、並べることによって、(採点した順番によって採点の結果が変わるという)採点の系列効果をもたらすという問題もあります。また、ソートはそもそもクラスタリングでもないわけです。



また、機械による手書き文字の認識という問題が実際の採点では結構ネックになると思います。現在の手書き文字入力ですと、精度が良い条件で大体 95%だそうです。これは、非常に良い数字のように思えますが、400字で 20字の誤字があるということです。このような精度だと通常、誤字が多すぎて読むのが難しくなります。

短答式ですと、例えば20字だとして、平均的に1字の誤りが含まれることになります。この誤りは採点では致命的になります。

また、漢字の部首の〉部(ひょうぶ)は、一般によく「にすい」といいますが、この「にすい」と「さんずい」とは、手書きでは区別がつかないそうです。また、平仮名の「り」と片仮名の「リ」も、字形では区別がつかないそうです。

そこでいわゆる n グラムと言われる,要するに n 個の並びによって,語彙の出現確率を利用して認識の精度を上げることが考えられています。例えばここに「コンピュータの機能」という文字がありますが,この「一」が長音なのか,マイナス記号なのか,罫線なのかを見るとき,三つ組で考えるときに,「ピュー」で見るのと,「ュータ」で見るのと,「ータの」で見るのとで,「一」の出現確率が違うわけです。また,この「」の間に何文字入るかという幅,すなわち n を何個に取るか,この窓のことをウィンドーといいますが,ウィンドーの幅を幾らにすると最も精度がよくなるかということについてもよく分かっていません。このあたりのシミュレーション結果は,まだ明確には得られていないようです。

また、数式の入力に関して言いますと、受験生が MathML や、LaTeX といわれる入力ソフトを使って解答するのは、ちょっと大変です。そこで、手書き文字を機械が認識して、それが「オーケーですか。修正しますか」という感じで進めるのが実際的です。ここでそのままオーケーならばいいのですが、修正する場合に、やはりまた一から入力し直すのではなくて、一部を消してそこだけ修正したくなるわけです。一部を消すことはスクラッチというのですが、ガシャガシャと消して、そこを直すとか、あるいは r^2 の肩書きの 2 を下に持ってくるとか、クリックやドラッグなどによる操作をして修正したくなるわけです。このあたりの標準のインターフェースは、現在のところはまだ決まったものがないようです。

短答式記述

- □ 含意関係認識
 - t1 鎌倉幕府は1192年に始まったとされていた が、現在では実質的な成立は1185年とする説が 支配的だ。
 - t2 12世紀に日本では鎌倉幕府が開かれた
 - t1が成り立つときt2が成り立つか
 - ■「1185年(に)成立した」といった<u>意味理解、時間</u> 情報処理、「開く≒成立した」<u>含意関係知識</u>
- □ NII 「ロボットは東大に入れるか」プロジェクト
 - ■「うそキーワード」発見方式の方が有利(H25)
 - 3つの手法の組み合わせ(H26)

一方,短答式記述について,採点をするための最も正統的な技術は「含意関係認識」と呼ばれるものです。現在,自然言語処理の分野で研究が非常に進んでいて,盛んに行われております。これはどういうものかというと,t1 と t2 という二つの文章を用意して,t1 が成り立つときに t2 が成り立つか,もしくは t1 と t2 が同じ意味かというのを測る技術です。

今ここに、t1「鎌倉幕府は 1192 年に始まったとされていたが、現在では実質的な成立は 1185 年とする説が支配的だ」と t2「12世紀に日本では鎌倉幕府が開かれた」という文章があるとします。普通の日本人なら、t1 が成り立つとき t2 が成り立つ、ことが分かるわけですが、機械にこれをやらせようとすると結構難しい問題になります。

何が難しいかというと、まず、t1の文章から、鎌倉幕府が1185年に成立したということが分からなければいけないわけです。この文は複文になっていますが、最初の文の主語「鎌倉幕府」が、後ろの文の主語にもなっているということが分からなければなりません。普通は必ずしもそうはなりません。また、2番目として、1185年は12世紀だということが分からなければいけません。3番目は、この「成立する」と「開かれた」というのが同じ意味だということが分からなければいけません。「成立する」と「開かれた」は普通、同じ意味ではありません。「本を開いてください」を、「本を成立してください」とは言わないからです。他にも、「実質的な」「支配的だ」といった処理をどうするかという難しい話はありますが、このようなことが分からないと含意関係認識はできないわけです。

国立情報学研究所の「ロボットは東大に入れるか」プロジェクトでも、含意関係認識を使って解くということが試みられています。しかし、平成25年度の段階では、「うそキーワード方式」といいますか、選択肢の中に間違いのキーワードを見つけるというようなやり方が優勢的でした。選択肢に含まれるキーワードの、一つだけが教科書の別の箇所にあるということが分かれば、その選択肢は誤りだということが分かるわけです。この「うそキーワード」を見つけるやり方や、去年の場合だと、三つの方法を組み合わせて解くといったやり方が正統的な含意関係認識に基づくアプローチより良い成績を収めました。

東ロボ・プロジェクト世界史B

- □ 日本ユニシスのアルゴリズム(76/100点)
- □ 前提:固有表現辞書、単語の上位・下位関係認識
- □ 3つの手法の組み合わせ
 - 質問応答で解く
 - 単語の共起確率で解く
 - 構文木のマッチングで解く(格文法)
- ⇒コンピュータによる自然言語理解はまだ難しい

昨年度「東ロボ」で成績が最も良かったのは日本ユニシスの三つの手法、「質問応答で解く」「単語の共起確率で解く」「構文木のマッチングで解く」の組合せです。難しい話は今日は避けたいと思いますが、問題によってこの三つを使い分けします。コンピュータによ

る文章の正しい言語理解は難しいのですが、そうではありながらも、この組合せで 100 点満点中 76 点は取ることができ、現在それくらいの技術水準に至っています。

●採点におけるコンピュータ支援

採点におけるコンピュータ支援 設問ごとの採点基準ファイルに基づく 模範解答との適合性 部分点解答との適合性 受検者解答の表層的表現のチェック 解答形式、字数 + α

記述式問題の解答を全部機械で採点するのは、現時点ではちょっと難しいだろうと思われていて、採点におけるコンピュータ支援といったところが落としどころなのだろうと思われます。現在考えられるのは採点基準ファイルに基づいて、模範解答との適合性、部分解答との適合性、受験者の解答の表層的な表現のチェック、解答形式が合っているか、字数などのチェックという支援です。

30字以内程度の解答 ■ 質問:火星探査機の使命は何ですか? (CBCラジオニュース、NASA) ▶ 火星の大気を調査し水を探すため(14字) ▶ 火星上に生命がかつて 存在していたか否かを科学者に判断させるための手助けをするため(40字) ■ 質問: ――線部「そういう地方」とはどのような地方ですか? (第2回全国学力調査・中学・国語A (基礎),寺田寅彦「茶碗の湯」より、2008) ▶ 日光に特別に温められたために、地面から蒸発する水蒸気の量が特に多い地方(35字)

短答式の記述問題では、理由を問うようなもの、もしくは名詞句を問うようなもので、 30 字程度の解答を求めるものが多いように思われます。

実際に、幾つかの公的な試験に出ている例を見てみましょう。アメリカの大学院入学のための GRA のテストでは、NASA の探査についてのニュースを読ませて、その目的は何ですかというのを書かせます。解答は、日本語換算で 14 字程度とか、40 字程度となります。また学力調査で寺田寅彦の「茶わんの湯」に線が引いてあって、「そういう地方」とはどういう地方ですかと問う問題に、35 字程度で書かす、というような具合です。青字で書

いたのがキーワードで、このキーワードの意味的な一致、表層的な一致、そのあたりを見 れば、記述採点はそれなりにできるのではないかと思います。

短答記述の解答形式(の判別)

- 口 述語文
 - (彼が大阪から来た) ■ 動詞文
 - 形容詞文 (彼女は優しい)
 - コピュラ(=繋辞)文 (彼は学生である)
- □ 名詞文(名詞句)
 - 述語文連体形+名詞
 - 述語文連体形+「ため」(理由の形式名詞)
- □ 接続助詞連用中止文
 - 述語文連体形+「から」(原因・理由の接続助詞)
 - 述語文終止形+「ので」(原因・理由の接続助詞)

また、 短答記述における解答形式の判別は、30字程度の量であればできると思われます。 例えば、その解答が正しい形式の述語文になっているか、もしくは、正しい名詞文になっ ているかの判別です。理由を問うような問題だったら、「から | 「ため | 「ので |、などの語 が正しい文法の様式になっているのかということは機械で判定することができると思われ ます。

「殺す」の同義語検索(日本語ワードネット)

assassinate 弑する/殺す

kill 打ち留める/眠らす/討取る/害う/仆す/殺る/片付ける/弑する /叩殺す/為とめる/ぶっ殺す/殺す/討ち果す/打取る/かた付ける/ 果す/討ち止める/討ち取る/打留める/仕とめる/眠らせる/討止め る/討ち果たす/打止める/殺害/あやめる/打ち取る/死なす/ぷち殺 す/打ち殺す/叩き殺す/殺生/仕留める/討果す/為留める hit 害す/害う/害する/弑する/叩殺す/殺す/討ち果す/危める/ぱ らす/討ち果たす/殺害/あやめる/ぶち殺す/叩き殺す/討果す/害な

mow_down /殺す/鏖殺/虐殺/殺戮

kill 討取る/滅絶/覆滅/殺す/打取る/殪す/掃討/掃蕩/討ち取る/ 無くする/掃滅/打ち取る/斃す/根絶/倒す/退治/剿滅/失くする butcher 屠殺/畜殺/潰す/殺す/屠る

上位語:下位語も検索可能

また、近年使われるようになった日本語ワードネットの利用も有効だと思います。日本 語ワードネットだと、同義語が検索できるのですが、実は、同義語というのは普通の方が 思っている以上にたくさんあります。ちょっと物騒ですが、これは「殺す」という言葉の 同義語です。こんなにあります。センター試験並みの多くの受験者がいれば、これらの単 語はほぼ網羅的に出ることは可能性としてあるでしょう。そうした場合に採点基準を作る ときに、許される同義語と許されない同義語をあらかじめ採点基準ファイルに書くことを すれば、それによって採点を良くすることができると思います。

また、上位語や下位語に置き換え可能な単語があれば、それも利用できると思います。

係り受け解析を用いた格要素の照合

□ 格文法による骨子抽出

原文:日本語のように単語間に 切れ目を 置かない

┗━━ 連体修飾句

膠着言語の文の<u>処理において、形態素解析は</u>第一 述部への連用修飾句 主格

の関門である。

沭部

- 形態素解析器はMeCab
- 構文解析器にCaboCha(格助詞、係り助詞の活用)
- 模範解答との格要素の照合
- 奇妙な「てにをは」、文の正しさの検出

13

今我々の研究グループが考えているのは、係り受け解析を用いた模範解答との格要素との照合というものです。文章を理解するのに格文法という考え方があって、それは文の意味を用言を中心に捉えるものです。下線を付したのが、格文法に基づいて選んだ格要素です。詳細は省略しますが、こういうものを取り出して、模範文と解答文との各要素で照合を行いますと、「てにをは」が不適切な文や、キーワードは押さえられているのに文としておかしいというものを見つけることができるわけです。

採点分類に必要な技術

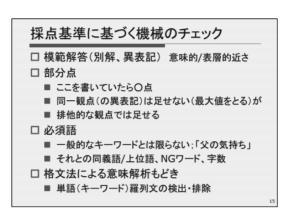
- □ 機械学習
 - 人の採点の仕方を学習して採点を行う
 - SVM(サポートベクターマシン)ではなく
 - RF(ランダムフォレスト)、多層パーセプトロン
- □ 2値分類か多値(kクラス)分類か
 - 記述試験には多くの配点(10点)
 - 2値(0点 or 10点)、3値(0/5/10点)で許されるか
 - 多値(0/3/5/8/10点)→一致しにくくなる
- 利用については習熟:細かなチューニング(パラメータ設定) は必要かも

また、採点をする際には当然、採点分類というものが必要となります。そこで考えられているのが、機械学習です。要するに人の採点の仕方を学習して、採点を行うということですが、自然言語で一般に使われている SVM (サポートベクターマシン) というものではなくて、RF (ランダムフォレスト)、多層パーセプトロンといったものを使おうかと思っています。これは、2 値分類にするのか、それとも多値分類にするのか、機械学習の方は k クラス分類という言い方をしますが、という問題に絡んでいます。

記述試験は、一般的に配点割合が大きいことが予想されます。例えば 100 点満点中 10 点とかです。そうした場合に、要するに 0 点か 10 点かという 2 値や、0 点か 10 点,間をとって 5 点という 3 値で許されるかといったら、多分そうはならないだろうと思います。

多値というか多クラス分類が、恐らく必要になってきて、そういう技術が欲しくなる。そ ういう手法を適用しないと自動採点はうまくいかなくなるだろうと思われます。

ただし、多分類になればなるほど、機械の採点と人間の採点の一致しない割合が高まりますので、より精度を上げるために、細かいチューニング (パラメータ設定など) が必要になるでしょう。



実際の採点は、採点基準に基づいて、それへの適合度の判定をある程度機械に負担してもらおうと考えています。具体的にはどういうことかというと、まず「模範解答」と実際の解答との意味的な、そして表層的な近さを測ります。また、ここを書いていたら何点というような「部分点」が与えられた場合に、それとの近さの度合いを測ります。その場合に、足せる部分点と足せない部分点があるわけです。そのあたりの切り分けが要ります。

またこういうキーワードが解答には欲しいといった「必須語」も採点基準に必要かもしれません。このキーワードは一般的なキーワードとは限りません。例えば国語で「父の気持ち」という言葉が非常に重要な意味を持つこともあるかもしれません。逆に NG ワードのようなものがあるかもしれません。また字数の制限も採点基準に必要です。

さらに、採点に際しては格文法による意味解析もどきによる評価も可能だろうと思います。



これが、我々が開発している途中の採点システムです。画面の上に解答があって、画面中央に自動的に出てくるのが採点基準ファイルです。模範解答や部分解答があります。他には、許される同義語や異表記があれば登録しておきます。あとは必須語があるかとか、意味を成さない文だったら0点にするとか、誤字があったら1点引くとか、そういう基準ファイルの項目があって、それに合致しているかどうかを機械が自動的に判定し初期設定します。そして、これに基づいた機械の採点が出ます。そういうものを基に、人間が最終的に得点を決めます。

もし、機械が出した提案が良ければ、画面下部に確定ボタンがあり、そのまま「確定」 を押すと、次の新しい答案が表示され、それを採点できるという形式になっています。

●まとめ

おわりに

- □ 数学の記述採点支援は難しい
- □ 自然言語の意味理解はまだ難しい
 - ギミックの積み重ね
 - 人工知能≒機械学習(「ルール発見」/「ルール構築」)→例外には対応しきれない
 - 人間採点は不可避
- □ 字数の問題 困難度が指数関数的に増大
 - 30字なら模範解答のキーワードの抜出し、パターンマッチ、若干の言い換え、格文法 →不十分?
 - それ以上なら言い換え辞書,シソーラスの活用 →道具建てが不十分

最後にまとめです。数学の記述採点は現在ではまだ難しいのだろうと思います。自然言語の意味理解も、まだ少し現段階では難しくて、ギミックの積み重ねでそれなりに近づける努力をしているということです。

採点に際して人工知能の活用があるわけですが、これは「機械学習」という言葉にほぼ置き換えていいと思います。機械学習は「ルール発見」「ルール構築」という言葉とほぼ同義語だと思われますが、そのルールを見つけるためには、ルールになるためのある程度の数が必要です。そうでないとルールになりません。ルールベースということは、すなわち例外には対応しきれないわけですから、完全な自動採点はまだ難しく、人間の採点との併用というのは避けられないだろうと思います。

また、字数の問題は結構大きいと思います。字数が多くなると、困難度が指数関数的に増加します。30字程度ならば、模範解答のキーワードの抜き出し、パターンマッチ、若干の言い換え、格文法あたりで済むかもしれません。でも、それ以上になると、シソーラスの活用などが必要で、今の段階ではちょっと道具建てが不十分だと言えるかと思います。

このように自動採点は難しい問題ではありますが、記述式採点の支援ならば現在でも十分可能で、これに向けて我々も研究を続けていきたいと思っております。

ご清聴ありがとうございました
コンピュータ上で実施する記述式試験
ーエッセイタイプ、短答式、マルチメディア利用についてー
〈解説〉電子情報通信学会 会誌
平成28年10月号(10月1日発行)掲載予定

【山地】どうもありがとうございました。ご質問がありましたら手を挙げていただければ と思いますが、いかがでしょうか。だいぶ早口の話が続きまして、結構お疲れの様 子もありますので、もし何もなければここで一旦休憩にしたいと思います。

〔指定討論〕

清水 和弘(福岡大学附属大濠中学校·高等学校副校長)

早稲田大学教育学部 国語国文学科卒。国語教師歴 39 年。進路指導主任、高校学年主任などを経て、2015 年 4 月より現職。その問、全国模試、問題集などに国語問題を多数作問。2013 年、日本テスト学会第 11 回大会企画セッションの話題提供者。



【山地】それでは、ディスカッションに入りたいと思います。最初に、今までのご報告に対して、高校現場からはどのように見えているかという点について、福岡大学附属大濠中学校・高等学校の清水和弘先生に指定討論をお願いしています。清水先生には、事前に資料を作っていただくことはお願いしていませんので、これまでの報告の中で特にコメントしたいというところを選んでお話しいただくことになっております。

清水先生は日本テスト学会でもパネリストをされたことがありますし、国語の先生として模擬試験などの作題の経験も非常に豊富で、学力評価に関してもベテランでおられます。そうした視点から幾つか気付いたことをお知らせいただいて、それらを踏まえて全体のディスカッションを進めたいと思っております。

では、清水先生、よろしくお願いいたします。

こんにちは。福岡大学附属大濠中学校・高等学校の清水と申します。よろしくお願いいたします。まず、こういう発言の機会を与えていただきましたことを心から感謝申し上げます。

今, 高等学校の全国の先生方が高大接続システム改革の成り行きをそれこそ固唾(かたず)をのんで見守っている状況だろうと思います。僕個人は退職間近ですから, このまま改革が遅れることもいささか期待しています(笑)。また, 学力というものがどういうものかという定義の問題はともかくとして, これからは, 思考力・判断力・表現力などを育てることが大切だと言われると, 正直なところ, 今までの教師はそうした思考力も判断力も表現力も育ててこなかったのかと, マイクを使って言うのも何ですが, 一言返したくなるような気持ちも実はあるのです。

まず、今日のお話ですが、西岡先生のお話の中で、京都大学の特色入試の「主に学力型AO入試」の項目の中で、「論述」「小論文」、教育学部は「課題」でした。こういう言葉の使い方が違うのです。この使い分けに何か意味があるのかということが一つです。もし意味があるとしたら、それはどういう意味があるのか。そして、これは西岡先生もおっしゃったとおりで、課題の問題自体が高等学校に対する発信でもあるのです。いわゆる非連続の読解型の論述問題と理解しているのですが、そうした問題を高等学校に発信した意図は何かということが非常に気になります。

これは白水先生が東大の入試にも少し触れられたのですが、東大の入試にも、ディスカッションではあるのですが、複数の資料を提供して、それに基づいてディスカッションさせるというスタイルを取られています。つまり、形としては東大の推薦も、京都大の AO 入試の教育学部の課題と似通っていると考えます。まさか相談なさってそうなったわけではないのでしょうけれども、奇しくもそういう形のスタイルの小論文――「小論文」という言葉は個人的にはあまり好んでいないのですが、小論文の「小」という字が非常に気に掛かっています、あえて言わせていただくのなら「論述」と言っていただきたいのですが――そうした論述の問題を選ばれた意図を、東京大学も京都大学もそうなのですが、お尋ねしたいということです。

また、これも気になるのですが、「小論文」が「英語、数学、国語」と並列的に、入試科目として扱われていることです。「英語、数学、小論文」などと言いますが、小論文というのは科目でも何でもない。方法論でしかないものを科目の項目に入れて、我々もそれをごく自然に受け止めてしまっています。

これも京都大学の課題の内容といいますか、いわゆる合教科・科目型、あるいは教科横断型というスタイルになっています。そうすると、科目を横断してしまったものを高校現場では実際に誰が指導するのかという問題が出てきます。普通に考えると、小論文というのは国語の先生が指導するものだと思っているのです。ところが、本当にそうなのかということです。なぜこのようなことになったのかということを抜本的に考える必要があるのではないかと思います。

その意味で最初に戻りますけれども、京都大学の課題、東京大学のあのディスカッションの資料、高等学校の受験者に対してどういうメッセージを送ろうとなさったのかという

ことが気になります。

それから 2 点目に、採点基準です。「論述」と言う限り、論理的な部分を重視されて採点されているのであろうと思うのですが、例えば我々国語の教師の立場からいくと、そこには表記の問題、漢字の問題、句読点の問題、陳述の問題など、様々な技術的なものが残ります。これをどのように評価されているのかというのは実際には全く分からないのです。採点基準を見ることができない状況で我々は指導しているのですから、大変困難を来します。現実的には、前期の試験が終わって、後期の試験までの2週間ぐらいの間で、一気に小論文のトレーニングをするのです。いろいろな小論文の形態がありますが、わずか2週間ぐらいでは実際には十分に指導・トレーニングができていません。その意味では、嫌味な言い方ではないですが、トレーニングをしていない生来の思考力を試していると言えばそうなのでしょう。

しかし、きちんとしたコンセンサスといいますか、大学側と高等学校の指導する者との 連携がうまくいっていないのではないかという気がしてなりません。この点が非常に大き いです。

それから、白水先生の話には驚き、感心しました。というか、私も、ああいう授業を本来しなければいけなかったのだと。特に気付いたのは、授業の前と後の比較評価ということです。当然我々はそれをしているつもりなのですが、計画的にした覚えがないのです。 僕はあと1年ぐらいで退職ですし、実は授業をしていませんので、一方では、今日は聞きながら「俺は助かったな」と思ってはいるのです(笑)。ああいう授業の管理というか、進行状況は非常に重要だろうと。それは Deane 先生がおっしゃった CBAL の方法にも一致しているのではないかと思います。そういう授業の形態を、今日の課題であるところの高大接続の評価にどのようにつなげるのか、これが非常に大きい問題だろうと思います。こもやはり相当なディスカッションが要るのではないかという気がします。

それから、最後に石岡先生のお話です。どうも僕は機械にも弱いので何とも言いようがないのですが、僕は教師としてずっと作問をすることをやってきましたが、作問をするということは同時に、採点基準を作ることでもあるのです。そして、採点をする者と作問をする者とが一緒であるところに質の高いクオリティが保証されるという感覚が僕にはあります。

ですから、石岡先生はそれは無理だとおっしゃっているのですが、機械採点がもし完全にできたとしたら、採点する能力を持つ教師がいなくなるのではないか。そしてそれは、作問能力を落とすことになるのではないか。ここは Deane 先生もご指摘なさいました。やはりクオリティの高い採点は、人間の採点に頼るべきだということです。そうすると、やはり五十数万人が受けるような大規模な入試では、どうしても記述試験の実施は難しいのではないかという感じがします。

現場の人間としてまず以上のように三つほど疑問点を述べておきます。よろしくお願い します。ありがとうございました。 【山地】ありがとうございました。大きくは、まず高校への発信の意図はどういうものか、 それから白水先生が紹介されたような授業形態で高大接続の評価にはどうつなげ るのか、そして、自動採点に全てなってしまったら作問の質が落ちるのではないか。 そのあたりのコメントをいただいたと思います。

これから先は、報告者の方にまず答えられるところから答えていただいて、その 上で全体のディスカッションに移りたいと思います。

〔全体討論〕

Paul Deane/西岡 加名惠/白水 始/石岡 恒憲/ 清水 和弘/山地 弘起(司会)



【山地】そうしましたら、西岡先生、白水先生、石岡先生、どなたからでも、清水先生の コメントに対してお答えの準備ができた方からお願いします。

【西岡】はい、では発表順ということで西岡からいたします。まず、各学部で出しているのが「論述」「小論文」「論文」「課題」、意味に違いがあるのかということなのですが、それぞれの学部が今回どういう問題を出したかというのは、ウェブ上でご確認いただけます。(追記:http://www.tokushoku.gakusei.kyoto-u.ac.jp/past_issues)特色入試を実施するということを決めたのは全学なのですが、デザインをしているのは各学部です。各学部で使いたい言葉を選んでいるものですから、そちらをご確認いただきたいと思います。

教育学部の場合、なぜ「課題」という言葉を使ったかというと、やはり従来型の 小論文ないし論文と言われる入試問題とは少し違うタイプのものを追求したかった ということがあります。もう一度、「求める人物像」のところをご覧いただきたいの ですが、私たちが今回追求しているものは、ある程度探究的な学習をしっかりやれ るような学生が欲しいということで、それはやはり大学で研究をしていくというこ とにつながるからです。「課題」で問おうとしているものは、複数の資料を勘案し合いながら、それを批判的に検討し、さらには論理的に思考を組み立てて、自分なりの主張を組み立てていけることができるような力ということです。その部分のニュアンスがこの「求める人物像」の中に折り込まれておりますし、それにできるだけ合う課題を何とかデザインできないかということでやってみたということです。

実は、グループディスカッションをやってみようかというのは話として出たのですけれども、そうしますと、たまたま居合わせたグループの相性のようなもので不公平感が出るのではないかというところを私たち自身が心配しまして、その部分は口頭試問で見るということにしました。

それから、採点基準は公開されていないので、申し訳ないのですが、この入試問題の採点基準がどうかというのは申し上げられないのです。一般論として、パフォーマンス評価というものを高等学校の授業改善に役立てていただきたいと、私自身は研究者として提案しています。その絡みということでご説明しますと、高校の先生方はどういうふうに採点すればいいのか分からないということであれば、まず一遍、生徒たちに課題をやってもらって、これはルーブリックの作り方なのですが、ぱっと見の印象で、できれば複数名で、お互いの採点が分からないように採点して、作品群を分類していきますと、レベルの差がそこにあるわけです。そのレベルの差を見比べたときに、なぜこれがレベルの違いとして見えたのだろうかということを先生方同士でディスカッションしていただければ、かなり妥当なルーブリックができるのではないかと思います。非常にいい教員研修にもなるでしょうから、そういうところで見る目をお互いに鍛え合ってみるというのは、一つ、研修としては考えられるかなと思います。

先ほどご覧いただきました「求める人物像」の上のところで、これはポスター発表会をやって、お互いに高校生がやっている探究成果をどういうふうに評価できるのだろうかというようなことを、高校の先生方、大学の教員共に交流しました。これは別に特色入試に向けてということではないのですが、どういう学習が有意義なのだろうかというところで交流をした機会です。そういうものでお互いに見合ってみるというのは非常に大きな共通理解にもつながっていくのではないかと思います。

ちなみに、高校への発信ということで言えば、私たちは2週間でトレーニングできるものを目指しているということではないのです。つまり、こういう探究的な学習を少なくとも高校前半 1~2 年生の間やっている学校が今かなり増えています。そういう取組は価値があるだろうから、私たちとしても応援したいという思いがあります。

高校に関して言いますと、本当に多種多様で、大学入試さえ突破すればいいのだというメッセージを非常に強く高校生に与えている高校と、大学入試がゴールではないよということをメッセージとして高校生に伝えている高校がありますよね。(笑)京大の中で特色入試の導入に熱心だった学部の中には、入試志向で育ててきた高校3校ぐらいからしか入学者が来ないという学部があります。そういうところは「こ

のままの入試ではやっていけない」という意識が強かったのです。教育学部の場合, 私たちは一般入試で入ってきた学生たちに何の不満もありませんでした。本当は, 大学入試改革というのはセンター試験と一般入試がうまく機能すれば,それほど大 きく変えなければいけないということはないのではないかというのが個人的な思い なのです。今は,そういう一般入試を受けている大学入学者が半数ぐらいしかいない,多分それが一番大きな問題なのです。そのあたりは,試験問題,大学入試セン ター試験の問題をいかに工夫していったとしても解決しないという問題が,少し気 になるなと思うところです。

【白水】まず1点目の、複数の資料を議論するような課題がなぜ大事にされているかというポイントですが、東大の法学部の入試自体は、資料を読み合わせたわけではないので、そこは多少ずれています。それから、東大の入試に私が関わっているわけではないので、その点はご了解いただいた上で、一般論としてお話しします。

なぜ大事になってきたのかというと、本当に社会が変わってきて、「ここにソリューション(解)があるので、それを覚えてそのまま適用しよう」というのでは、なかなか問題が解けなくなってきているからだと思います。断片的なヒントはあるのだけれど、それをまとめて編集して答えを作るのが、これからの大人なのだろう。だから、そういう力が必要になってくるということです。例えば代数の問題だったら、今までシリアルに小問が並んでいて、順に解いていったら最後の問題が解けるというようなタイプで学んでは、その学びを適用してくればよかった。ですが、そのようにいかなくなったときに、知識を統合する、知識を自分で構成するということが大事になってきたところが大きな変化なのだと思います。それで「知識構成型ジグソー」という名前を付けて、去年亡くなった三宅なほみなどが中心になりまして、いろいろな授業でやってきております。

2 点目の「前後評価が良かったのだけれど、こういうことの授業の繰り返しの結果をどのように評価につなげているか」という質問にお答えしたいと思います。知識を構成して、最初は分からなかった渋谷の地形について分かってきて、渋谷だけではなくて土地を考えるときにこんな視点があるのだよねということが分かった、そういった学びを日々行う子供たちを評価で救うというときに、大きく分けて二つの評価の仕方があると思います。

一つは一回性のテストです。センター試験, 2 次試験, その他いろいろなテストというのが, 子供たちが日々の授業で作り上げた知識の質, 深い理解を問うような問題に変わっていく, この全体的なトレンドはこれから変わらないのではないかと思います。断片的な知識のピースの方ではなく, そこでまとめたような, 地理の肝をつかめているか, 理科の肝をつかめているか, そういうものを問うような問題が増えていきます。それによって子供たちがアクティブラーニングで学んだ成果を間接的に評価するということが一つあると思います。

もう一つは、今日の授業ですと、事前から事後、45分の中で子供たちは話し合い

ながら結構変わっていきます。この変わっていく力、要するに変化を引き起こす力、 友達と相談しながら目の前に与えられた資料を読み合わせて、組み合わせて、そこ で答えを作っていく力、それをまさに受験、一回性のテストで見ていくものが増え ていくのではないかと思います。それがグループディスカッションのようなところ でやれれば、確かにメンバーによる得意・不得意はいろいろあるかもしれないけれ ども、いずれにせよそのポテンシャルが見えると、大学に入ってきてオーケーでは なく、大学に入った後、伸びていく子供の力が見られるのではないかと思います。 いずれにせよ、大学では新しい知らないことを学ぶわけですから。

そのために変化をその場で引き起こして、CBTで評価する試みも増えると予想します。例えば、事前に書いた答えに対して、いろいろな情報が後から入ってきたり、他の人の議論の様子を読みながら、「あなたは最初の考えをどう変えますか」という形で直接的に評価するものなどです。以上です。

【山地】石岡先生に行く前に少し確認なのですが、清水先生からお尋ねのあった高校への発信の意図ということの中には、ひょっとすると採点基準、もちろん批判性、論理性、その他に内容・構成・表記、いろいろな評価観点があるはずだと。少なくとも内容だけで評価しているのではないだろうけれども、その重み付けなどが分からないし、論述を評価するに当たって、せめてこういう基準、こういう観点を使っていますよ程度のものがないと、なかなかその準備ができないのではないかというようなこともあったかと思います。

加えて、合科的な、科目が連携して授業を作っていくということがまだまだあまり多く行われていない状況で、それが一種の真正な課題として入試で問われるとするならば、それを実際に指導していくのは大変だというようなお話です。その場合にどういうメッセージを受け取るべきかということかと思うのですが、そのあたりはどうですか。あるいは、清水先生からまた追加があるかもしれませんが。

【白水】事前・事後で何が変わるかというときに一番大切にしている基準は、知識・理解のところです。それは、やはり本当に教えたいことがあるので、高校生にこれが分かってほしいということをしっかり事前・事後で比較するというのが一番のベースにあるためです。ただ、その事前・事後の知識・理解の質を押し上げていくときに、それぞれの学習者がどんなふうに文章を読んだか、その文章を理解して人にどうやって伝えて、お互いに情報を持ち寄ってどんなふうに組み合わせてアーギュメントを作っていったかというところのスキルというのは、プロセスで評価できます。

その次です。では、スキルの発現というのを「こんな言い方をしていたらレベル5」というような評価基準を、今、持ち合わせているのかです。それに関して、Deane 先生は「何とかできそう」というお立場でラーニング・プログレッションズを出されています。それに対して私たちは「そこはちょっと待った方がいいのではないか」というので、実際に子供たちの話すデータを見ながら、そこはゆっくり作っていっ

た方がいいと考える立場です。

なぜかというと、子供の考え方、学び方、インタラクションの仕方は非常に多様です。話し合いながら考えているときの子供たちの会話は、誰かが主語を言ったら次の子が述語を言って、最後の子は目的語を言ったりしながら一文を作っていくという過程であることも多い。その会話の仕方は、プレゼンとして見ればプアかもしれないのだけれど、実はそれが人の考えを作っていくときにすごく大事なのだとすると、そこをうまく評価するような指標を、あまり焦って最初につくらない方がいいのではないかと思います。

では、どうするかというと、授業の中で今日教えたいことについて、本当に子供たちが事前から事後で深く学んでいっているかを基軸で考えて、その間のインタラクションの在り方は多様でいいという形で授業をやっていけばいいのではないかと思います。

今日の Deane 先生の話で言えば、広告を禁止にすべきか、賛成にすべきかどうかに関して言うと、私から見ると、最後のエッセー、こういうエッセーを書けるといいよねというのをコンテンツベースで評価できるかどうかが私たちの場合だと一番大事にしたい指標です。その質が上がってくると、間接的に他のスキルもたくさん使っていたのではないかと判断する、推定する、そのような合わせ技でやっております。以上です。

【西岡】二つのことですけれども、高等学校でも総合的な学習の時間というのはとっくの 昔に入っています。今ごろこういう質問が出るというのが、やはり「入試改革まで しないと高校のカリキュラムが変わらない」と文科省が判断した理由なのだなとあ らためて思いました(笑)。

探究的な学習を指導している高校がないわけではないので、そういう先進事例を ご覧いただくと、それぞれの教科の先生方が専門性を生かしながら研究指導的なこ とをやっておられるというのが一つはあるかなと思います。総合学習でやるような 課題は、答えがないから指導をどうしていいか分からないという悩みは、確かに探 究学習を初めて導入される高校からご相談を受けます。

テーマは恐らく、実際に先生方も現在生きていて悩まれていることを選ばれるとか、高校生が実際に考えたい課題を選ばせるということで出てくると思うのです。 答えを見つけるプロセスで、高校生たちと寄り添いながら、対話しながらやっていくということをやるのであれば、高校生よりは先生方が絶対にうまく探究的に学ぶということに慣れておられるはずなので、そういう指導はきっと可能になってくるだろうと思います。それが一つの方向性です。

もう一つの方向性は、先ほど評価方法についてご説明をしたときに、パフォーマンス課題というものをご紹介しました。これは各教科の中で少し探究的な学習というか、知識・技能を活用する学習をやっているわけです。例えば理科で実験をどのようにデザインすればいいのだろうか、人に説得力を持って自分のアイデアを伝え

るためにはどういうふうに表現したらいいのだろうか、社会的なあるアイデアに関して批判的に資料を検討しながら論述するにはどうしたらいいだろうか、そういう各教科の教科内容に即して少し探究につながるような学習を深めていくという課題になるわけです。こういうトレーニングを各教科の中で積み重ねておく、そういうことは恐らく総合的な課題になったとしても生きてくることなのだろうと思います。

本当は教育学部の特色入試でも、各教科のパフォーマンス課題、各教科の内容に即した課題を出したいという思いも個人的にはあるのですが、実は出題者の力量の問題もあります。教育学部の教員というのは大体、文系の教員です。そうなると、数学の問題を作れるかというと、作れないわけです。だから、入試を実施する単位が小さくなればなるほど、その入試を実施する単位の教員が出せる範囲の問題しか出せないという問題も、並行して起こるということが事実としてあるだろうと思います。

【Deane】同じようなやりとりを米国でやっていますから、とても聞きなじみがあります。問題の一つは、受験者へのスキルの期待がカリキュラムの設計や評価に必ずしも十分反映されていないということです。したがって、高校の先生にはフラストレーションがたまるわけです。結局、何で判断されるのか、評価されるのか、十分に分からないところが高校の先生にはあるということかと思います。

そういう問題があったので、アメリカにおいては Common Core State Standards というものを作ることになったわけです。生徒として何を知り、何を理解して college-ready にするべきなのか、それから逆算をしてシステマチックな形で、生徒 はどの段階でどこに到達しなければいけないのか、12 学年、11 学年、10 学年の段階で、college-ready であるためには到達地点はどこでなければいけないかということを考えました。そのスタンダードの実施の仕方でよくなかったのは、テストをあまりに早急に導入したがゆえに、教員側の対応が十分できなかったということです。新たなスタンダードを十分に取り込むだけの準備ができなかったというところがあると思います。

CBAL の考え方とも非常に関連がありますが、大学の教員と高校の教員の本当の 共通理解を醸成して、何をスタンダードとし、そこに到達させるにはどうしたらい いか、をいかに一緒に考えてもらえるかがポイントです。

それから、米国でも聞かれたもう一つの問題は、授業実践とこうしたプロジェクトベースあるいはアクティブラーニング型の学習というのは、密接に関連しているということです。生徒を奨励して、ディスカッションに参加しなさい、探究を通じて自分の考えを作っていきなさいと、生徒自身が責任を持って議論を構築し、深く突っ込んで考えるように誘う教員であるには、教室での然るべき活動が必要です。教員側がそのような活動のやり方を分かっていないと、みんなにとってフラストレーションがたまる結果になって、うまくいきません。

したがって、教員側が自分たちで実践して学習をする、ディスカッショングルー

プに自らも参加する、そして、クラスがどうやったら探究ベースで学べるのかを理解し、クラスを成長に向けてどのようにマネジメントしていくかを学んでいなければ、高大接続は難しいのではないでしょうか。大きな問題になるのではないかと思います。

いま伺っていたフラストレーションに私は共鳴するところがありました。小論文や生徒のプロジェクトのスコアリングの基準が分からないと、基本的なフラストレーションが生じます。生徒にどうやって準備させたらいいのか、自分でもどうやっていいか分からない、何が期待されているかも分からないということですから。

やはり、このあたりはきちんとコミュニケーションをしなければならないのではないのでしょうか。皆がきちんと期待レベルを把握できるように、またどのような指導の活動を通じてそこに到達できるかの共通理解を得られるように。さらに、パフォーマンスで高いレベルとはこういうことだと理解できるように。例えば、ルーブリックの5と4はこれが違うとか、生徒がいちいち先生が付き添わなくても独力でできるようになるためには、こういう指導が必要だ、とか。教育が今まで規範としてこなかったことをやらなければいけない、というのは、大変な苦労を伴う取り組みです。大変なことに取り組んでいるということを自覚するとともに、コミュニケーションや相互理解の醸成が極めて重要と認識する必要があります。

- 【山地】ありがとうございます。アメリカの経験、Common Core State Standards ができて、それに基づいてスタンダードを決めていくのだけれども、それがなかなか高校の先生方には内面化できない、また大学の教員もあまり知らない、ということがあるかもしれません。これは CBAL やアクティブラーニングのトレーニングなどともオーバーラップしていますので、そうした新たな方法でのアセスメントあるいはルーブリックになじんでいくというのが一つの打開策なのかなとお聞きしました。
- 【西岡】関連して Deane 先生に、どのように CBAL で先生方あるいは採点者のトレーニングをしているのかということをもう少し具体的にご説明いただくと、日本の先生方にも参考になるかと思ったのですが、いかがでしょうか。
- 【Deane】まず、非常に豊かな専門職としての教員養成に代わるものはないということです。CBALの一環として、形成的評価に関する研究を多々行っています。形成的評価というのは、教員としては、学習者にどれだけ学習が進んでいるかという情報を得られなければいけないということです。例えば、生徒が明示的に何が学習目標なのかを理解できるようにする、そして評価の基準を明示的にきちんと提示し、それを内面化して、自分の学習の責任をとれるようにする、さらには友達と一緒に課題に取り組むことを目標達成へのリソースと考えられるようにする、ということです。形成的評価に沿った考え方には、こうしたアクティブラーニングを下支えするものがたくさんあるわけです。

CBAL ではいろいろな研修用の資料を作ってきました。アセスメントのために、教員がそのような考え方を取り込めるように助ける教材です。しかし、教材を作ったからといって、それを自分のものにできるかどうかは別の話です。最終的に必要なのは、いろいろな研究をやってきた経験から申しますと、教員はお互いに連携して、教材をいかによく教えるかということを探究し続けている専門職である、という考え方だと思います。形成的評価論の中核には専門職の学習コミュニティの概念があるので、教員研修と並んで、専門職としての授業改善への協働が極めて重要です。

それから、CBALではアセスメント自体がメッセージになると考えてきました。 非常によく体系化されたアセスメントでは、具体的にこのステップを踏んでください、そうしたら到達できますよという説明があり、具体的なルーブリックやヒント、ガイドライン、模範的成果の例などが示されています。このように、基準が何で、何が満たされていれば成功なのか、学習目標の達成とはどういう状況なのか、といったことをはっきり提示していると、教員によっては自分たちが不十分だったところはこの部分だと分かって、こういうことを指導に組み込もうと気付きます。

その例として、私たちは何年もかけて、サマリーを書く基準を作りました。多くの教員は、私たちの経験では、生徒に「サマリーを書きなさい」と言ったら、すぐできると思っているので教えもしないわけです。ところが、「サマリーを書きなさい」と言われても書けない生徒がいるのです。そこで、いろいろと試みまして、一連のタスクを作り、よいサマリーとはどういうものか、またよくないサマリーの例とはどういうものかを具体的に示した上で、あるサマリーの例が要件を満たしているかどうかのテストをし、そして自分で実際にサマリーを書かせるようにしたのです。連携した教員たちからのフィードバックには、「こんなことは教えたことがなかった。これとこれを教えたらサマリーが書けるようになるとは」「この教材を使った後、生徒がいつもよりいいサマリーを書けるようになった。こんなことを教えるのだったのか」というものがありました。

教員のリスクの一つは、やり方をいろいろ知っていて、自動化されている知識が多いということです。当たり前の知識については、説明する必要があるとは思えないのです。したがって、複雑なパフォーマンス課題では、その教え方について注意深く考えを深めておかないと、子供にはこれもあれもできるはずだと思い込み、「はい、やりなさい」と言うだけになってしまいます。ところが生徒側は、頑張っても全然分からないことがあるのです。教員は困惑して、「何でできないのだろう、こんな当たり前のことが」と言うわけです。上位3分の1の成績の生徒であれば、明示的に教えられなくても自然にできるかもしれません。しかし、下位3分の1の成績の生徒だと、明示的に教えられないとできないのです。

非常に有用なのは、明示的なモデルを提示することです。どういうステップを踏んでやるのか、達成しなければいけない大事なことは何なのか。うまくできたといえるものや、よくあるミス・誤解は何なのか、一つのステップからその次に移るの

に役立つ支援はどういうものか、といったことを提示するのです。こうしたことは、カリキュラム開発やアセスメント開発の一環です。よいパフォーマンス評価では、アセスメントの開発とカリキュラムの開発にそうしたつながりができています。というのは、生徒の学習を成功裏に進めるには、その二つは密接に関わっているからです。

【山地】ありがとうございます。教員の評価リテラシーを高めるということに関わってくるようなお話でした。アセスメントがメッセージだというのは、CBALの言葉で言えば、キー・プラクティスの下にある要素タスクをどのように同定して、それを言わば足場掛けに使ってどのように最終的なアウトプットに至らせるか、その辺の組み立て方を教員が自分でやりながら見つけていく、設定してみる、ということも一つのアプローチになりそうですね。

では、お待たせしていました、石岡先生、お願いします。

【石岡】先ほど清水先生から、小論文の採点の評価基準が分からないまま、学校では評価しているというようなお話があったと思うのですが、Deane 先生のいらっしゃる ETS では、e-rater という自動採点システムを作っていて、評価項目が 12 個と決まっています。それぞれに対する重みも全部決まっています。それはトピックによらないで決まっていて、それをあらかじめ公表しているわけです。もし公的な試験にそういう論文が採用されることになったら、説明責任という点から評価基準はやはり明示していく必要があるのではないかと思っています。

もう一つ、論文の機械採点が完璧になったら人間の採点者をつぶしてしまうのではないか、作題の質を落とすのではないかということについてのご回答ですが、そういうことはないだろうと思います。というのは、機械採点は人間の採点に限りなく近づけていくというアプローチを取っているから、というのが答えの一つです。実際に Deane 先生からお話があったように、GMAT と呼ばれるアメリカの経営大学院のための試験では、エッセーが評価されるわけです。機械が全部やるわけではなく、機械と人が採点します。違いがあったらまた別の人間が採点するという形で、最終的な判断は人間に任されています。文法的なことや、機械でチェックできるようなものは機械に任せて、それ以上のことは人間がやってという形で、お互いにコラボすることによって、より良い判断基準をどんどん作って、良い基準システムを作っていくということが目標です。その意味において、人間の採点者がいなくなるとか、作題者の質を落とすのではないかということは杞憂(きゆう)であろうと思われます。以上です。

【Deane】自動採点の部署での私の経験から申し上げたいのですが、どういうことが起きるかというと、集まって作問をしているとき、採点者には分かるだろうと簡単に思い込んでしまうのです。しかし、人が評価する場合、それが問題になるのです。曖

昧で不明瞭であったりして、判断が難しいところが入ってくるからです。したがって、ETSでは、自動採点のモデルの開発をする側と、元々のルーブリックを作る側との連携を緊密にする、ということを旨としています。よく考え抜いて、評価基準をはっきりと明示することができれば、機械にも有用だし、人にも有用なのです。人が分からない状況だったら、機械が分かるはずがないのです。

【山地】清水先生、何かありますか。

【清水】一点だけ思いつきで、申し訳ないです。こうした論述型の問題で幾つかの項目を立てて採点されているだろうと思います。それは大学によったり、あるいは学部・学科によったりということで、その比重は変わる、これはもちろん分かるのです。要は、例えば一つの項目の中で採点基準が設けられていて、これが要するに Goodだとする。ところが、期待した以上の解答が出てきたときに加点、つまり満点以上に点数を加えることはあり得るのか。実際にされているかどうかは公開なさらないで結構なのですが、見解としては、論述の採点という点ではどうなのかということです。

それと、Deane 先生のところでは、アメリカではそういう加点というものをどのようにお考えになられているか、お伺いしたいです。

- 【山地】白水先生が先ほどおっしゃった、あまりラーニング・プログレッションズをはっきり決めないで、いろいろ多様性があるのだから、その辺の柔軟な見取りが必要だということとも関係があるのかもしれませんね。
- 【西岡】すみません,入試に関しては採点基準を公開しないということになっているので, 私の個人の判断では言えません。ただ,入試の場合にやりやすいのは,相対評価で すので,良いものから並べていって上から順に取れば決まるという側面があります。 ご覧いただいたら分かるように,たかだか 12 名の比較なので,そんなに比較は難 しくないというところがございます。

学校現場でパフォーマンス課題に取り組んでいただくときには、先生方に共同で採点基準を作るという作業をワークショップなどでやっていただきます。一般的に言いまして、最初にパフォーマンス課題をやったときに期待以上のパフォーマンスが出てくるということはめったにないです。「何だ、やらせてみたらこれだけしかできなかったのか。テストでこんなにできているのに、英文で作文を書かせたらこんなにできないのか」と驚かれるような例もあります。

ですから、穴埋め問題で試している英語力と、自分のアイデアを例えば2パラグラフで論述する力は違うのだなということをそれで実感していただけるのです。例えばルーブリックのレベル1からレベル4で作ったとすると、その次にレベル5を加えていく、さらに指導が改善され、パフォーマンスが改善されれば、レベル6が

できてくるというのが実際には起こります。ルーブリックを作って頭打ちになることはないというか,ルーブリックで行っているのは,言わば形成的評価なのです。レベル1からレベル3でルーブリックが作られたら,その上に目指すべきレベル4は何だろうということが次にディスカッションできます。それなしで,ただオープンエンドでやらせているだけだと,何がそこで大事にされるべきなのか,あるいは,どういう指導の工夫をすればさらに伸ばすことができるのかというところにつながらないので,その作業は学校で共同研究をやるときなどには必ずやっていただいています。

【白水】ルーブリックやこのフレームワークはエボルビングなもの(進化するもの)であってほしいと思っていたので、今の西岡先生のお話は非常に印象的でした。

加点について、一般的に困ることは、 $A \cdot B \cdot C$ を決めて、想定外の解答を、S にするか D にするかということです。要するに $A \cdot B \cdot C$ の、A を超えた想定外の良いものなのか、それともこれは想定外過ぎて D なのかというようなところで、先生がもめているとやはり元々やろうとしていた評価の質が低いのだろうと考えます。

私たちがジグソーの中で見つけていったのも、創造的な課題をオープンエンドで任せるよりも、例えば三つの資料を読んで、「高校生がいる食卓になるべく良いパーティー料理を作るにはどうしたらいいのか」という具体的な状況設定で創造を求める。そうすると、最後に出てきた解が「高校生になると親と話しにくくなってくるので、大皿料理を作って、どうしても取り分けなければいけなくて会話が起きるようなレイアウトにした方がよい」というものが出てきます。すると、先生は「それは想定しなかった加点したい答えだ」と自信を持って評定できるようになってきます。それは、まずはクローズドエンド的に $A \cdot B \cdot C$ という資料を分かっているかで成績 $A \cdot B \cdot C$ のグレーディングができることによって初めて、それを超えた創造的な要素が含まれているか、組み合わされているかという判断ができたからでしょう。

その意味で、西岡先生がおっしゃった、1、2、3 と決めるのだけれども、それを超えるものがあるかもしれないというのは、しっかり決めるからこそ自由なものが見えてきて、ラーニング・プログレッションズというのがどんどん発達していくということではないかと考えました。

【西岡】ひょっとしたら白水先生と意見が分かれるところなのかもしれませんが、限りなく伸びていくということは、それほどはないです。学校現場では、まず、良い課題を作るのに苦労されます。良いパフォーマンス課題を作るというところで、複数の先生がチームで取り組んでいただいて、良い課題例を共有して、「これは生徒たちが食い付きがいい」とか、「こういうふうに指導するとこういうふうに伸びる」ということが共有されていくと、大体3年ぐらいで、例えば「うちの学校で英語力を付けようと思うと、こういう課題をやるといいね」「これぐらいの評価基準で共通理解が

できそうだね」ということができてくるので、限りなく作り続けなければいけないという印象ではないです。

【Deane】追加で申し上げたいのですが、一般にパフォーマンス課題をやらせるとき、全くなじみも経験もなければ優れた結果は出ません。それが人間というものです。生徒が全くなじみのないタスクをやらされると、戦略もないし、やり方も分からないし、複雑なタスクを一つ一つ細かく分けてやりやすくするということもできないのです。そのタスクになじみがある生徒だけにそういうことができるのです。

ですから、例えば入試で公平なパフォーマンス課題を行うのであれば、生徒が高校で何度もやってきたのと似たようなタスクを与えるべきです。そうでないと、なじみがないためにどうやってアプローチしたらいいか分かりません。重要なのは、たとえ明確な評価基準を開示することが難しくても、似たような一群のタスクがあるわけで、そこに共通した評価基準があるということです。例えば、ディスカッションでは、良いディスカッション行動とそうでないものがあります。一緒に問題に取り組むとき、ディスカッションに慣れている人が自然に使う戦略があるものです。入試で使うものと同じタイプの一群のパフォーマンス課題を高校でやり、よいパフォーマンスと判断できる共通の基準があるのであれば、入試でのタスクと全く同じ評価基準でなくても大丈夫でしょう。重要なことは、生徒が、この種のタスクはこうやって取り組めばよいと学ぶことなのですから。

【白水】ラーニング・プログレッションズというのは、多分、日本で言うと学習指導要領上で、例えば数学や国語、社会、理科のコンテンツについては既に日本にはあると思うのです。Common Core State Standards にせよ、Next Generation Science Standards にせよ、むしろ日本の学習指導要領を参考にしたという面もありますから、コンテンツはオーケーです。

では、スキルの方はどうか。今回の学習指導要領の改訂でも、例えば「考える」 という中に必要な要素、必要なレベルがあるのだとすると、それを学年ごとに分け るといいのではないかという議論は、私の理解では一時期ありました。

でも、これをやってみると、数学でも理科でも社会でも結局やはり「比較する」「関係付ける」「規則性を見つける」「まとめる」というのがすごく大事だね、どこも似ているよね、となったのではないか。そしてこれを具体的にどうするかというときに、スキルの方のプラクティスに関してラーニング・プログレッションズをやると、多分、日本とアメリカのお国事情が違うということを考えに入れた方がよいのです。日本人はすごく真面目なので「こういうものが良いサマリーです、こういうものが良いディスカッションの仕方です、こういうものが良い考え方です」と言われると、そのとおりにやろうとしてしまうというのが、結構、日本の危ない国民性なのかなと思います。

小学校理科で3年生は比べるのが大事,4年生は関係付けるのが大事,5年生は

規則性を見つけるのが大事と言ったら、その学年はそれしかやってはいけないのかと思ってしまう先生が出てくるのです。それは誤解なのですが。だからラーニング・プログレッションズは良いのだけれども、それを個別化せずに、日本の場合は今回、学習指導要領の中で、コンテンツもスキルも学習態度も一体にした「ものの見方・考え方」が、こういうところに到達できるとよい、と出したのではないか。包含的なプラクティスというか、理科というのは「内容もスキルも態度も合わせてこんなことが身に付くといいのではないか」というのを学校段階、学年で配していくというようになったのかと思っています。

【山地】たしかに、学力の3要素と呼ばれていたものが資質・能力の三つの柱と言われるようになっていたり、ちょっと表現が変わっていますね。「生きて働く知識」などからなっているというのも、ある意味でこの三つを別々のものに見過ぎていたのに対して、一体となって動くものとして、言わば知識構成に向けての動きを前提にしているということなのかもしれません。

高大接続ということを、単なる選抜接続で考えるのか、それとも教育接続で考えるかということで、かなりそこでも見方の違いがあると思うのです。今回のテーマは、どちらかというと教育接続の中で入試をどう捉えるかということが重要な観点かと思います。

例えば先ほど白水先生が、もし一回性の入試、1 回でやるのだとしたら二つのやり方があるということをおっしゃいました。一つは、深い知識・理解のアウトプットを問うもの。もう一つは、言わばその場を真正な課題を行う状況にする。それこそ東大や京大で論述させる、ディスカッションさせる、その場でのパフォーマンスを要求するようなことをやる、その中で知識構成の仕方、こういう言い方は非常に悪いかもしれませんが、ある意味でスキル化している、習得されているものを発揮してもらって、そこを窓にして評価する。そういう二つのやり方です。

ただ、これらはいずれにしても一回性ということでもって、どうしてもテスト・ワイズネスの問題、あるいは非常に浅いところでやってしまうなど、当然、高校の側はそれに対する準備をさせるということになってきますから、あまりにもそれが強まってくると、いつもの力を発揮してくれればいいということでは済まないのではないか。ひょっとすると、すでに幾つかの大学のAO入試で行われているように、何回か学習機会を作ってその後で評価の機会を持ってくる、あるいは蓄積されたものの成果を評価する、というような育成型のアプローチが必要かもしれません。教育接続という中で入試を考えた場合に、どういう方法が適切かということですね。

今までのお話はどちらかというと、高等学校の方でどういうふうにやっていけばいいかということです。これを大学教育につないでいく。あるいは、教育接続とはいっても、どこかに選抜の要素が入ってくるので、このあたりの捉え方があるかと思います。

パネリストあるいはフロアの方で、何かコメントなどありましたら、ぜひ伺えた

らと思います。どうでしょうか。はい、どうぞ。

【西岡】ディスカッションの中では課題にハイライトが当たってしまったのですが、私たちは書類審査という点では、高校での学習履歴を非常に重視して見ております。その点では、学校現場が取り組んでおられることをちゃんと大学としても受け止めたいという願いも込めた入試になっています。

それで、先生の手が入るのではないかというのがものすごく心配だったのです。 とにかく中学校からの学習履歴で本人たちがアピールしたいものを出してもらいます。多分、先生方が一人一人の生徒に手とり足とり、例えばファイルに入っている作品を全部先生が書き直すというのは非現実的だろうから、機能しなくはないなという感触は持ちました。実際、例えば彼らが高校時代に書いたレポートや、生徒によってはプレゼンをやっている DVD を入れてきたりするので、それを見るわけです。もちろん、それは見せるようにやっているものではあるのですが、それと、当日やってもらう課題、あるいは口頭試問と重ね合わせれば、大体安心感があるなという感触はあります。

【山地】ありがとうございます。西岡先生が先ほどプレゼンの中でおっしゃっていた、そうやって入ってきた学生たちを迎える大学の授業の側も変わらないといけない、ということで初年度の授業を変えておられる、このあたりも一種の教育接続の在り方として、高校サイドにもメッセージになっていくのだろうと思います。

私が厄介な問題提起をしてしまったかもしれません。残り時間があと 15 分程度なので、この先生にこれは確かめておかなければいけない、これを聞かないと帰れないということがありましたら、ぜひおっしゃってください。いかがでしょうか。はい、どうぞ。

- 【フロア1】随分前から中学高校一貫教育という話が出て、特に私立の名門と言われるような学校でそんな話があります。それが今、高校大学接続ということで、文部科学省的にというか、先生方皆さんはエスカレーター式でずっとつながっていく、そういうのを考えているわけではないのですよね。
- 【山地】高大の関係をちょうど中高一貫のように, 高大一貫でいくようなイメージですか。
- 【フロア1】私立の大学など、怒られるかもしれませんが、生ぬるい感じでだらだらといくような感じになると、まさに京都大学らしさ、東京大学らしさ、そういうものがなくなってしまう怖さのようなものを感じますが、どうでしょうか。
- 【山地】入試というクサビが入ることで、選抜という緊張のあるプロセスを通ることによって生ぬるさが払拭されるのではないかと、そういうことですか。

幾つかいただきたいと思います。今,一つ承りました。他にいかがでしょうか。 あと 10 分ちょっとなので,できるだけ多くの方からご発言があればと思います。 はい、お願いします。

- 【フロア2】高大接続の議論の中で、高等学校からの内申書や報告書が重視されるという ことになりますと、例えば高等学校で先生に非常に嫌われていた生徒、そうした生 徒が不利な採点を受ける可能性というのはどうなのでしょうか。そのあたりは清水 先生にお願いしたいのですが。
- 【山地】ありがとうございます。もう一つぐらい受けたいと思いますが、いかがでしょうか。
- 【フロア3】西岡先生の京都大学の特色入試の関係で、ポートフォリオ、課題も含めて評価基準を公表していないとおっしゃったのですが、特別に公表しない積極的な理由というのがあるのでしょうか。アカウンタビリティといいますか、これは清水先生からご質問があったのですが、もし何か積極的な理由でもあれば教えていただきたいです。非常に簡単な質問です。
- 【山地】ありがとうございます。では、答えやすい方からと言ったら変ですけれども、評価基準を公開できない積極的な理由があるのか、そのあたりはいかがでしょうか。
- 【西岡】私たちとしては、求めている人物像が評価基準なのです。だから、これで公開しているというつもりなのです。
- 【フロア3】そうすると、特色入試のアドミッションポリシーに合っているかどうかという視点ですか。ポートフォリオ、課題の論述、口頭試問も含めて、それで見ているということですか。
- 【西岡】はい、もちろんどの部分でどれを重点的に見るという適合性を見ています。
- 【フロア3】アドミッションポリシーに書いていることに合致しているかどうかで見ていますよ、というだけでは世間は納得しないような気がするのですが、どうでしょうか。他の皆さん方はどうお感じになっていますか。こういう点、こういう点と、先ほど Deane 先生のお話で十幾つぐらいありましたね、そういうルーブリックのようなものが、もちろん中ではきちんと作られているのでしょうが、ここを公表するとまずいという話があるのですか。
- 【西岡】 その辺に関して話をしてはいけないと言われているので、言えないのです。結局、

対策され過ぎてしまっても困るというところがあるのです。実力を見たいというと ころがあるので、今のところは言えないという方針で来ています。

- 【フロア3】すみません、しつこいようで。言えないという理由が全学的に何か整理されているのだったらいいのですが、言えないということで大学で決まっているのを、 西岡先生に教えてくれと言っているわけではないのです。大学のアドミッション全体の中で……。
- 【西岡】本当に一つ一つの情報を公開していいかどうか、例えば過去問にしても出していいかどうかをいちいち全学に確認しながらやっているのです。というのは、何かある学部だけで情報を出してしまって、出し過ぎではないかというようなことになったら困るということでいちいち確認しています。今のところ入試基準に関しては、求めている人物像以上のことは出さないという方針で来ていますが、そういう声をいただいたということは持ち帰ってみたいと思います(笑)。

それから、接続が生ぬるくなるのではないかということですが、全然生ぬるい入 試だとは、私は今やっている入試に関しては思っていません。ただ、一般入試と比 べたときにスタイルが違うなとは思います。特色入試の場合にどういうふうに違う かなと考えると、やはり求めている人物像と高校生たちが学びたい学部像のマッチ ングを一般入試よりも見るという色彩が濃くなるなという印象はあります。

京大の特色入試の場合,提出書類として学業活動報告書というものを求めている 学部が大半です。実は、これは学校が準備するものです。教育学部は恐らく学校に 違和感を覚える子の方が教育学に適しているということもあるのではないかという 思いもありまして(笑)、学びの報告書は本人に書いてもらうという選択をしている というのもあります。

【Deane】一つの視点を提供させていただきたいのですが、ルーブリックで大事なことの一つは、それ自体の個別性です。ETSでは、二つの種類のルーブリックを使っています。ジェネラル・ルーブリックと、一連のスコアリング・ノーツです。ジェネラル・ルーブリックは開示されますが、スコアリング・ノーツは開示されません。特定の問題やトピックにのみ用いられるのがスコアリング・ノーツです。それを開示してしまうと、その問題に対してはこういう答えを書けばよいと教えてしまうようなものです。それに対して、ジェネラル・ルーブリックは一般基準です。いろいろな問題に利用され、何に価値を与えるかを伝えるものです。

皆さんのお話では、求める人物像はジェネラル・ルーブリックのようなもので、評価基準は、スコアリング・ノーツと私たちが言っているものに近いと思いました。 スコアリング・ノーツは、私たちも公表していないです。



- 【山地】私は3月まで長崎大学にいましたが、アクティブラーニングの評価のときにルーブリックを学生たちに先に示すと、今度は学生が手段的に良い点を取ろうとするから、最初は何もないところからやった方がいい、という意見もありました。ただ、それだとフェアな感じがしないというところもあって、なかなか難しいですね。 白水先生は今の質問について何かありますか。
- 【白水】今日の話で、観察と解釈は分けて考えたいという気がしました。その意味で、解釈基準はジェネラル・ルーブリックに当たって、実際の観察の、今日のアイテムだと、どういう指標を作るかというものがスコアリング・ノーツになってくるのだろうと思います。そこを分けて、ジェネラル・ルーブリックまで出すのだけれども、スコアリング・ノーツというのを出してしまうと、次の年にばれてしまうというのもあるし、やはり、すごく多様で個別で、「今回の受験生だとこういう評価だ」というのを出してしまうことになるので、公平性という話になったときにいろいろややこしいことになってしまって、本当の学び、受験の場で学んでくれたことの良さというのが表れにくくなるのだろうなと感じました。

ただ、戻って考えますと、大学の中で先生たちが日々どんなふうに頭を働かせているか、例えば農学部を出たらどういうプラクティスが社会人になったときに要るかというのは、もっとどんどん高校生に伝えていくべきだと思います。

私たちがやっているのは、東大の中の CoREF という機構なのですが、理学部や工学部とタイアップして、一日ものの実践学講座というワークショップをやっています。午前中にその分野の最先端の内容を、ジグソーを使って三つの資料を読み合わせて「ここまで分かった」というのができると、午後にその質問をリサーチクエスチョンにして研究をやっていくためのproposalを一度書いてみようというようなことを、高1から高3ぐらいを呼んでやっています。その日の最後に、若手のファカルティ、大学院生と高校生が一緒にしゃべると非常に面白いのは、高校生の質問に若手の助教が「それ、すごく良い質問なんだけど、答えはまだ分かっていないんだよ」というようなことを答えてくれるので、高校生が「答えの分かっていないんだよ」というようなことを答えてくれるので、高校生が「答えの分かっていないことが世の中にいっぱいあるのね」と感動して学校に帰るということです。高校までは、世の中の全てに答えがあって、そういう問題ばかり解いているのだという教育を受けているのかもしれません。

その意味で、各学部がそれぞれのアドミッションポリシーをしっかり立てて、センター試験のデータを手に入れて、「ここの小問がすごく面白かったので、ここのウエートを高めよう」とか、「正解と誤答で 8 点と 0 点に分けるのではなくて、惜しいやつを取ったやつに、うちの学部は部分点を出してもいいよ」など、アドミッションポリシー、それぞれの大学・学部が自分たちのプラクティスに必要な力に応じて、データを解釈できるようになっていくと、評価のリテラシーも一層上がるのではないかと考えています。以上です。

- 【山地】ありがとうございます。このあたりは白水先生がスライドの一番最後のポイントで示されたところです。「評価主体が自らリスクを取って」、例えばセンター試験の結果を自分たちにいいように使えばいいのだということなのですね。
- 【清水】西岡先生のお話で、うちの生徒もひょっとしたら受験する可能性がありますので、あまり西岡先生のところは敵に回したくないのですが(笑)、例えば今、AO の第 1 期生で入った 3 人なり 5 人なりの合格した子は、当然、学校に報告をしています。どういうレポートを出してどういう内容を書いているのかというのは、報告が行っているのです。そこの学校の後輩はそれを参考にしてトレーニングをする、こういうことなのです。

うちも附属ですから附属推薦を持っていますし、指定校推薦もたくさん持っていますし、推薦で相当行くのですが、面接など全て報告させています。非公開になっているものも、基本的にうちの子が受けたものはデータとしてあるのです。うちのような大規模校は、高等学校の生徒だけでも 1934 人いて、1 学年 600~650 人で全国の大学を受け回ってきます。そうすると、私大も含めてそうした報告書が相当あり、何冊にもなるのです。こういううちのような学校が有利になるというか、ある程度のデータを持ち得ているということです。それだとすると、むしろ公開された方がいいのではないか。それからもう一つは、トレーニングすることはいけないことなのかということも考えていただきたいと思います。

あまりお答えいただかなくて結構です。個人的な意見として,そっと言っている だけですから(笑)。

【西岡】今日はご紹介していないのですが、具体的にどういう資料を入れることができる のかという例示は、かなり詳しく書いています。

(追記:http://www.tokushoku.gakusei.kyoto-u.ac.jp/assets/uploads/pdf/提出書類 様式/H30/h30 activity report educ guide.pdf)

それから,今日ご紹介したような特色入試の合格者に対するインタビューもウェブ 上で公開しています。

(追記:http://www.kyoto-u.ac.jp/ja/mm/jitsuha/160728.html)

そういう点では、できるだけのことはやっているということが一つです。

ここから先は本当に個人的な意見なのですが、探究学習に関してどういう点を評価したらいいかということに関しては、先ほども申しましたように、高校の先生方に集まっていただいて共に議論をするという機会も持っております。その部分も、OCW(オープン・コースウエア)動画で公開もしています。

(追記:http://ocw.kyoto-u.ac.jp/ja/opencourse/92, http://ocw.kyoto-u.ac.jp/ja/opencourse/113) もちろん、そういうのにどんどんアクセスするということを知っている学校なり受験生と、アクセスしたらそういう情報があるということを知らない高校生・学校の皆さんとでは、恐らく有利・不利が出てくるということはあり得

るわけです。できるだけのメッセージをお伝えしたいと思って、今日はこの場に参 りました。

- 【山地】ありがとうございます。さて、時間が少し超過してしまいました。今日のお話で、私としては、高校の先生と大学のスタッフがもう少し近いところにいないといけないという印象を受けました。大学の中のアドミッションズ・オフィスの先生方は当然、高校とかなり密接な関係を持ちますが、それ以外の先生方は、実際に学生たちを受け入れる立場ですが、学生がどういう準備をしてきたか、あるいはどういう入試を経てきているのか、この辺は十分に了解されていないように思われます。高校と大学が、もっと近い距離でもう少し意見交換をしながら、接続の実質を作っていく作業が必要ではないかと思った次第です。
- 【西岡】まずは実際の問題を見ていただきたいです。受験した学生たちは、最初に公開されたサンプル問題では700字で要約しろと書いてあったのに、実際の入試では、自由記述で書け、何字でもいいよとなったので、すごくパニックになったと言うのです。そんなレベルでパニックにはなってほしくないのですよ。どんと「うちの生徒はこれだ、見てみろ」ぐらいの感じで、高校の先生方は自信を持って大学にアピールするぐらいの構えでいていただいた方が、恐らくこの入試は意義が大きいものになっていくと思います。私は高校現場とかなり交流していますので、距離があると言われるとすごく傷付きます(笑)。
- 【山地】大変失礼しました。でも、評価する・されるという立場の違いがやはりあるので はないでしょうか。

というわけで、最初に趣旨説明のところにありましたように、予測どおりというか、期待どおりというか、いろいろな論点が出されてシンポジウムは終了することになりました。今後、新テストに向けていろいろな動きがあるでしょうから、大学入試センターとしても、こういう形の情報提供や意見交換の機会をできるだけ継続して作っていきたいと思っています。またぜひご参加いただけましたら幸いです。最後に、ご登壇の先生方に拍手をお願いいたします。

すみません, 4 分ほど超過しました。最後に、閉会の挨拶を大学入試センターの 大津起夫研究開発部長から申し上げます。

〔閉会挨拶〕

大津 起夫(大学入試センター研究開発部長)



本日は土曜日にもかかわらず、皆さん、このシンポジウムに参加していただき、誠にありがとうございます。ご登壇の先生方からは、非常に野心的でインパクトのあるお話をいただきましてありがとうございました。

センター試験は基本的には教科書に書いてある知識、高校の学習内容を確認するという方針で作っているのですが、それに加えて、人間の学習、人間の知識はどのように構成されるのか、そこのメカニズムに踏み込んでそれを評価するようなシステムを作らなければいけない、そういうものを作りたいという熱いご報告を伺うことができました。ただ、それを共通試験という巨大戦艦のようなメカニズムの上に載せていくにはどうすればいいのか、まだよく見えていないところはあると思います。

一つは、高校生の勉強の総体を評価するセンター試験のようなメカニズム、もう一つは、 白水先生、西岡先生、Deane 先生が言われていたように、日常の学習の中で、テクノロジ ーを使って情報をたくさん集めていって、それを評価の対象にするという二つの流れがあ ると思います。その両輪をどのように組み合わせていくのかが将来的な課題になるであろ うと思いました。

このシンポジウムは、大塚先生、山地先生が大変ご苦労されて設定してくださいました。 また、前川先生をはじめ東京工業大学の皆さん、この会場をお貸しいただき準備にも奔走 してもらいました。さらに、ブースにいらっしゃる同時通訳の方々、大変なスピードの講演 を適切に翻訳していただきました。皆さん、誠にありがとうございました。あまり長くなっ ても仕方ありませんので、これで最後の閉会の言葉とさせていただきます。

シンポジウム企画書

〇日 時:2016年9月10日(土)13:00~17:45 (開場12:30)

○場 所:東京工業大学大学院社会理工学研究科 ディジタル多目的ホール(定員 287 名)

○主 催:大学入試センター ○後援:文部科学省・日本テスト学会

〇テーマ: 「高大接続における学力評価の最新動向」

○趣 旨:去る 3 月末に出された高大接続システム改革会議の最終報告を受けて、いわゆる学力の 3 要素を反映した学力評価や大学入試設計が喫緊の課題となっています。とくに思考力・判断力・表現力や学びに向かう意欲等に関しては、これらを評価するための新たな方法の開発が求められています。そこで本シンポジウムは、思考力・判断力・表現力に焦点を置きつつ学力評価の最新動向を把握し、高校生の学びや成長に資するとともに教育接続としての高大接続にふさわしい学力評価はどのように可能か、具体事例を踏まえて今後の指針を整理することをめざします。内容は、まず基調講演で思考力のモデルに基づいた米国での評価方法の事例を紹介していただき、次いで個別入試でのパフォーマンス評価の活用、学習科学からみた今後求められる資質・能力の育成と評価、そしてコンピュータによる自動採点評価の可能性を報告していただいたうえで、高校現場からのフィードバックをもとに全体討論を行う構成です。

○プログラム

- 13:00~13:10 開会挨拶 山本庸基(大学入試センター 理事長)
- 13:10~13:20 趣旨説明 大塚雄作(大学入試センター 試験・研究統括官)
- 13:20~14:35 基調講演 Paul Deane (米国 ETS 首席研究員) 「認知科学にもとづいたアセスメント―思考力・判断力・表現力の評価・ 改善をめざすアセスメントシステム|
- 14:35~14:50 質疑応答
- 14:50~15:00 休 憩
- 15:00~15:20 報告1 西岡加名恵(京都大学大学院教育学研究科 准教授) 「パフォーマンス評価をどう活かすか―京都大学教育学部における特色 入試の取り組み」
- 15:25~15:45 報告2 白水 始(東京大学大学総合教育研究センター 教授) 「学習科学からみた資質・能力の育成と評価」
- 15:50~16:10 報告3 石岡恒憲(大学入試センター研究開発部 教授) 「コンピュータで実施するテストの採点をめぐる課題」
- 16:15~16:30 休 憩
- 16:30~16:40 指定討論 清水和弘(福岡大学附属大濠中学校・高等学校 副校長)
- 16:40~17:40 全体討論 Paul Deane・西岡加名恵・白水 始・石岡恒憲・清水和弘
- 17:40~17:45 閉会挨拶 大津起夫 (大学入試センター 研究開発部長)
 - *総合司会 山地弘起(大学入試センター試験・研究副統括官)

出席者集計結果

「大学入試センター・シンポジウム 2016」の参加者の集計結果は以下のとおり。

- ○事前申込者 271人
- ○参加者 215人(当日参加者6人を含む。)
- ○参加者内訳

	教員・研究員	事務職員	その他 ^(注3)	計
国共士尚	40 人	12 人	0人	52 人
国立大学	(18.6%)	(5.6%)	(0.0%)	(24.2%)
公立士岗	10 人	8人	0人	18 人
公立大学	(4.7%)	(3.7%)	(0.0%)	(8.4%)
打六十岗	17 人	20 人	0人	37 人
私立大学	(7.9%)	(9.3%)	(0.0%)	(17.2%)
古笙岗桥	7人	0人	0人	7人
高等学校	(3.3%)	(0.0%)	(0.0%)	(3.3%)
÷ 7/# -	1人	14 人	1人	16 人
塾・予備校	(0.5%)	(6.5%)	(0.5%)	(7.4%)
その他 ^(注1)	26 人	32 人	12 人	70 人
その他 ~/	(12.1%)	(14.9%)	(5.6%)	(32.6%)
上 兴了三 朴 L 、	15 人	0人	0人	15 人
大学入試センター (注2)	(7.0%)	(0.0%)	(0.0%)	(7.0%)
≡↓	116 人	86 人	13 人	215 人
計	(54.0%)	(40.0%)	(6.0%)	(100.0%)

- (注1) 教育情報産業,マスコミ,無所属の個人等
- (注2) 客員教授については、本務校で計上
- (注3) 会社役員,無所属の個人等

アンケート集計結果

「大学入試センターシンポジウム 2016」の会場で実施したアンケート調査の集計結果は以下のとおり。

- シンポジウム出席者 215人 アンケート回答者 112人 (回収率 52.1%)
- 1 あなたのお立場についてお答えください。

① 性別

男	女	未記入	計
92 人	19 人	1人	112人
(82.1%)	(17.0%)	(0.9%)	(100.0%)

② 年齢

~29 歳	30~39 歳	40~49 歳	50~59 歳	60 歳~	計
5人	19 人	27 人	46 人	15 人	112 人
(4.5%)	(17.0%)	(24.1%)	(41.1%)	(13.4%)	(100.0%)

③ 所属 ④ 職種

		教員・研究員	事務職員	その他	未記入	計
1	国公立大学	30 人	15 人	0人	0人	45 人
2	私立大学	11 人	12人	0人	0人	23 人
3	塾・予備校関係	1人	7人	2人	2人	12 人
4	国公立高校	3人	0人	0人	0人	3人
5	私立高校	1人	0人	0人	0人	1人
6	民間企業	7人	4 人	7人	3人	21 人
7	マスコミ関係	0人	0人	1人	0人	1人
8	省庁・役所等	0人	0人	1人	0人	1人
9	その他	1人	1人	2人	1人	5人
	計	54 人	39 人	13人	6人	112人

2 本シンポジウムが開催されることについて、何を見て知りましたか。 (複数回答可)

チラシ	チラシ 大学入試センターHP		計	
41 人	42 人	35 人	118人	
(36.6%)	(37.5%)	(31.3%)	(105.4%)	

3 本シンポジウムについての感想を4段階で評定してください。

	4	3	2	1		
	当てはまる	やや当ては まる	あまり当て はまらない	当てはまら ない	未記入	計
 わかりやすか 	27 人	73 人	11 人	0人	1人	112人
った	(24.1%)	(65.2%)	(9.8%)	(0.0%)	(0.9%)	(100.0%)
②新たな発見が	67人	41 人	3人	0人	1人	112人
あった	(59.9%)	(36.6%)	(2.7%)	(0.0%)	(0.9%)	(100.0%)
③ 興味深かった	66 人	39 人	6人	0人	1人	112 人
③ 興味/未がりた	(58.9%)	(34.8%)	(5.4%)	(0.0%)	(0.9%)	(100.0%)
④ 構成は適切で	41 人	51 人	18 人	0人	2人	112 人
あった	(36.6%)	(45.5%)	(16.1%)	(0.0%)	(1.8%)	(100.0%)
⑤ 有益であると	50 人	49 人	11 人	0人	2人	112人
思った	(44.6%)	(43.8%)	(9.8%)	(0.0%)	(1.8%)	(100.0%)
⑥ 集中して参加	35 人	60 人	15 人	0人	2人	112人
できた	(31.3%)	(53.6%)	(13.4%)	(0.0%)	(1.8%)	(100.0%)
⑦ 時間の長さは	33 人	53 人	23 人	2人	1人	112人
適当であった	(29.5%)	(47.3%)	(20.5%)	(1.8%)	(0.9%)	(100.0%)
⑧ 開催日時は適	48 人	58 人	5人	0人	1人	112人
切であった	(42.9%)	(51.8%)	(4.5%)	(0.0%)	(0.9%)	(100.0%)
⑨ テーマはよか	68 人	36 人	6人	1人	1人	112人
った	(60.7%)	(32.1%)	(5.4%)	(0.9%)	(0.9%)	(100.0%)
⑩ シンポジウム	55 人	44 人	12 人	0人	1人	112人
会場はよかった	(49.1%)	(39.3%)	(10.7%)	(0.0%)	(0.9%)	(100.0%)
⑪ 総合的に満足	52 人	53 人	5人	0人	2人	112人
できた	(46.4%)	(47.3%)	(4.5%)	(0.0%)	(1.8%)	(100.0%)

4 本シンポジウムについての御意見・御要望,特に印象に残った点・疑問に感じた点,大学入試改革についてのお考えなど、御意見・御感想等がございましたら、御自由に御記入ください。

(*は判読不能箇所)

- ○思考力や表現力を入試はもとより、小中高大の教育で評価していくことには賛同する。しかし、教科 あるいは領域固有の知識(や考え方)についてはどう扱うのか?/志望学部、専門によって求める科 目=知識(や考え方)にちがいがある。コンテンツあっての思考・表現であり、知識の教育(学習)・ 評価を過度には軽視できないと思う。(国公立大学・教員)
- ○「高大接続における学力評価」についての文科省の方針が定まらないので、各論を聞いても、それら をどのように生かせば良いかわからない。もやもやしたシンポジウムでした。(国公立大学・教員)
- ○変えるなら5教科7科目のセンター試験ではなく、ACT や SAT のようなテストに変更できないのでしょうか。センターの実績からすれば可能だと思われますが、SAT 的なセンター試験プラス調査書+ 個別試験(加えたい大学だけで)+面接(これも加えたい大学だけ)で入試ができればと思います。(国公立大学・教員)
- ○どのように「評価」するかについて平易な言葉で伺いたかった。入試として用いるのに適する問題は だいたいイメージしていたものと一致していたので、どのように評価するかといった観点から話を聞 きたかった。(国公立大学・教員)
- ○今回は国語(英語)についてであったが、理科や数学では、少し内容が違うのではないかと思う。そのような点も聞きたかった。(国公立大学・教員)
- ○入試改革としては直近というより少し先の話が中心だったと思います。大変勉強になりました。どちらかと言うと、授業構成や学生の学力の可視化に参考になったと思います。よく理解を深め、可能な所から、FD 等に活用してゆきたいと考えます。/西岡先生の言葉にもかかわらず、やはり高大のギャップを感じました。(国公立大学・教員)
- ○グループ学習時の会話の記録から思考回路(経路)を抜き出す方法は、大変面白いと思いました。一方で、発言力、コミュニケーション能力は低くても論理的思考能力の高い学生をどのように見出すのか、考えさせられました。理工系大学教員なのでCBALの理科についての紹介があるとよかったです。 / 高大接続(AO)入試の観点から、高校での評価基準の統一やポートフォリオ作りは、大学側としては、有用なのですか?実現の可能性はあるのでしょうか。(国公立大学・教員)
- ○理論・理想や方向性の発表・議論が多く、今後の具体的内容を期待していた参加者が多かったと思える。休憩時間に 50 人程度の参加者が退席された原因の一つだと考える。高校の教員や教育センターの職員等が参加すると意見がある内容であるので、教育現場への広報がよりあれば良かったと思う。(国公立大学・教員)

- ○評価の問題の難しさがわかりました。これからもこのような機会を開いてください。コンピュータによる評価はまだ時間がかかるようです。ただ、この開発にかける時間と成果、つまり費用対効果に疑問があります。つまり、マークシート方式である程度思考過程は評価はできるようになる。その他の能力は、大学入学後に育成することで良くはないか。(国公立大学・教員)
- ○現在進められている高大接続は、入試に対する考え方の転換を伴うものである。こうしたシンポジウムを、複数地域で実施すべきだと思う。(国公立大学・教員)
- ○この記入がコンピュータで採点できるか知りたい。(国公立大学・教員)
- ○事前にもう少し詳しい資料がほしいです。(国公立大学・教員)
- ○理想が先行してシステムがついていかない現状かと思います。最終的に生徒, 先生方の負担であったりしてはならないと思います。本末転倒にならない改革を望みます。(国公立大学・教員)
- ○大学の個々の教員と高校の個々の教員とのおつきあいが大変と感じたシンポジウムであった。(国公立大学・教員)
- ○背景事情と合わせて解釈すると、面白かったです。もう少し広い会場を御用意いただけるとありがたいです。日本の高校の先生は(おおむね)ずっと丁寧に指導しているし、研修の機会も豊富にあります。米国の教員の指導力が低いだけなんじゃないかなぁ……と思いました。失礼ながら。(国公立大学・教員)
- ○基調講演: Native で難しかったが、相当なスピードでの*専門用語のため、日本語で多く聴きとりとしたが、こちらも和訳が早口と*ため、十分な聴き取りはできなかったが、CBALのアーキテクチャ、課題、***に対してのとらえ方などが、十分理解できた。/2 特色入試は大学入試の一つの理想形とも思われるが、***など、地方の国立大学が導入していくには***が多く、また、それだけの入試であって志望者にとって入りたい大学・学部となるよう、また、その***にして、人材を受け入れ育てていくということを、大学人は努力が必要と感じた。/3 石岡先生の話が大変興味を持って聴くことができました。共通一次の開始した昭和54年からの***に大きく仕事で係わってきたが、現在の状況は全く想定できなかった。マンマシンの使い分けとしてアーキテクチャは当時も現在も同じと***るが、システムの***援用で、マン(人力、労力、時間)の負担を下げることが可能と思われ、公平性の確保***でにマン(人力)を有効的にあてることが可能と思われます。/4 コンピュータを利用した学力評価システムを大学入試センターと大学(国立大学)とで、共同開発することを検討してはどうか。参考までに、医療の分野では、平成14年から2年かけてDPCシステムを開発し、大きな変革を遂げてきた。/5 ***を完成させていくことが大事だなと思っています。(国公立大学・事務職員)

- ○もっと広い会場にしてほしい。(国公立大学・事務職員)
- ○全体討論での高校側の現状からの意見もあり、高校と大学側の意見が出て、興味深かったです。(国公立大学・事務職員)
- ○高校現場からの率直なご意見がおうかがいでき、また、その中での討論は大変有意義でした。高校現場の声をしっかり聞きながら進めていけたらよいと思います。(国公立大学・事務職員)
- ○高校側とのパネルディスカッションでのやり取りが、予定調和でなく面白かった。今後もこのような本音での議論をぜひ!! (私立大学・事務職員)
- ○高校現場の方のご質問が一番有益であったと思う。高大接続といわれているが、大学側と高校側との 温度差を感じた。(私立大学・事務職員)
- ○大学と高校の意識のギャップが明らかになって大変興味深かった。(民間企業・研究員)
- ○テーマと内容が一致していないように感じました。「高大接続の」という観点が薄かったように思います。しかし、多分野からの発表を一度に聞けたのはよかったです。/高大接続の議論も、京大・東大の AO 入試も、CBAL に比べて乱暴だと感じました。生徒に身につけてほしい能力についての議論検証が不十分で、そのコミュニケーションがうまくできていないのではないでしょうか。抽象論ではなく、科学的に、具体的に能力について考え、検証し、その後で評価について考えるべきだと思います。/白水先生のご研究は、学び方が多様とされつつ、対話以外の方法で学ぶことが得意な生徒をとりこぼすことになっていることについてどうお考えでしょうか。/いずれにせよ、評価のあり方について考えさせられる良い機会となりました。指定討論者の先生による発問のおかげが大きいです。(民間企業・研究員)
- ○学力評価の最新動向とは言え、最新動向が適切な方向に向かっているかどうかは疑問が残る。討論の中で内容が深まったのは指定討論者の高校の現場の意見をもとにしたからだと考える。/ディーン先生のルーブリックの表は大変、参考になりました。/校長名で出す調査書に、先生に嫌われて不利な内容など書くことはありえません。(国公立高校・教員)
- ○指定討論者の着眼が一般人にも共感できて、すばらしかった。(国公立大学・事務職員)
- ○同時通訳だけでなく、Paul 氏の資料の和訳も用意してもらえて、とても助かりました。/指定討論者 の指定は困難なのはわかりますが、もう少しまともな方というか、本筋の討論ができる人を準備して もらいたい。指摘点が些末すぎる。/Paul 氏への質問は通訳を通さずに、演者が直接英語でした方が、 齟齬が無くてよかったのでは。/山本理事長は、学長をされた大学で、採点基準を、公表していたのか。/Paul 氏の話は、基本的には教育評価に基づきすぎていて、入試に及ぶ話だとちょっとズレる。/

フロアに変てこな質問をさせないように、DNC のスタッフは質問をがんばって欲しい。(国公立大学・教員)

- ○パネリストの中に「高校生は世の中の問題には必ず正解があると思っている」旨発言している者がいたが、それは嘘だ。ほとんどの高校生は、正解がない問題もあることは知っている。嘘の前提に基づく議論は不毛。一方で、社会に出ると学校のように問題には必ず答えが用意されているわけではないと言うのも嘘。社会に出るとほとんどの問題で、答えは決められていて、多くの人々はそれを受容しながら生きている。それを知らないのは大学教員だけだ。(国公立大学・事務職員)
- ○思考力,判断力,表現力を評価する入試を実施した場合,ルーブリックを明確にしていないと,成績 開示の請求があった場合に回答が難しい。今後はルーブリックの例示等を講演願います。(国公立大 学・事務職員)
- ○同時通訳は非常に良かった。(特に女性) それに対して、スライドの和訳がひどかった。かえって理解 に苦しむ結果となった。/CBAL のような手法が実用化されると、現行の集団教育が終わる、もしく は無意味になると感じた。/日本は、もっと国を率いる人材の育成、エリートの育成に教育資源を傾けるべきだと思うが、そうした形が高大接続以下の諸改革の中に全く見えてこない。(私立大学・教員)
- ○Learning Progression という考え方が非常に納得できました。普段の教育活動の中でもやもやしていた点のうちのいくつかが言語化できたように思います。(私立大学・教員)
- ○生徒(学生)の多種多様な特性を評価する姿勢、方法の開発に関する具体的な提言という点で、大変興味深く拝聴した。最近同類のシンポジウムでも評価方法の多様化に関する提言が多いが、問題は次の二点にあろう。①生徒(学生)の基礎知識、教養の差をどのように関連付けるか。想定している生徒のレベルが微妙に異なり、その事で聴衆により違和感を抱く場合が少なくない。②構成の多くが「どんな点を評価したいか」「どんな学生を採りたいか」、つまり、試験を執行する側のスタンスに基づいている。生徒の側からして、「どんな点を評価して欲しいか」「どんな環境(大学)に身を置きたいか」という視点が反映された提言は比較的少ない。今後はこの点が重要になるものと思われる。即ち、受け入れた多様な個性をどのようにして、満足感を与えながら進展させるかという事である。真の「高大接続」を目指すのであれば、選抜の課題のみならず、進学後の質保証、アカウンタビリティーこそが問われるべきであろう。(私立大学・教員)
- ○2011 年 11 月に行われたシンポジウムに参加しました。当時中教審の答申が出るころで、その中の学力評価テストに対する論調で、文科省の主張は矛盾している、意味をなさないという論調が多く聞かれました。つまり「一点刻みでない評価と言うが、そのような評価では、レベルの高い大学では全員合格となってしまい、選抜テストとしては、機能しない」という論調でした。講演者の多くの方が同じ論調でした。しかし、その後何かが変わったかというと、何も変わりません。つまり一昨年のシンポジウムは単に不満やボヤキを並べていただけでした。何の意味もない「シンポジウム」でした。今

回は、少し学力評価に関する事例紹介等があり、参考になるという気がします。特に京大の特色入試が参考になりました。今後も、不満のガス抜きのためのシンポジウムではなく、参加者の考え方が拡大発展するようなシンポジウムをお願いいたします。(私立大学・教員)

- ○多面的な観点からテーマを深めていた。特に一定の見解にこだわらない点がよかった。(私立大学・教 員)
- ○基調講演は特に、内容としては事務系にはハイレベルであったと思います。次回は是非、教員にも紹介して、参加していただければと思います。(私立大学・事務職員)
- ○CBAL のアプローチが非常に興味深いが、全ての小中高に適用できるかが、心配。(進学校ならともかく、あまり学力レベルの高くない学校では、識字レベルが高くないので、まず「書く」力を身につけることが重要ではないか?※映像の学生(中?高?)は比較的学力レベルが高い学生と見受けました。個人差をどう解消するのか?それとも個人差を測定したテスティングはその意味で難しいのでは→大学入試センターの問題の理想的。/CBAL のアプローチが入試に適応できるかは別問題とありますが、思考力(表現力、論理的、批評的思考etc.)とエッセイは割と今後の入試に応用できるパースペクティブだと思います。/CBAL はプレゼンテーションスキルも測定可能とあるが、特にルーブリックやガイドラインがしっかりしていても、個人のプレゼンテーションスキルは、パフォーマンス測定は難しいのではないか。/CBAL と SAT、ACT の相関性は?どこまで通じているのか?(CBAL の問題は、やはりSAT や ACT(特に ACT)を意識したものだと感じました。ただし、採点(システムによらない human task)が難しい。日本でも IB コースなどの高校でないと難しいと思います。(私立大学・事務職員)
- ○初等教育・中等教育を経て高等教育へ、まさに learning progressions がいかに重要になるか再認識させて頂きました。大学の 4 年あるいは 6 年間でいかに人間力を涵養し、人間性を形成していくか、これからの大学が求められる大きなテーマだと思います。その大学までに入学する 6・3・3の課程でどのような学習形態や自我、主体性等を確立させてゆくかその観点からも高大接続がより重要になってくることは言うまでもありません。少子化が続き、これからの日本社会を担うネイティブジャパニーズは東アジア、東南アジアの若者はもとより世界の若者と切磋琢磨していかなければなりません。世界的にも「おとなしい」「控えめ」と言われる日本人、特に若者は自己主義、利己主義的な傾向にありがちです。さまざまな価値観、人生観を相互理解、共有し、世界の人々と(流行の若者言葉で言うならば)「ガチ」でコミュニケーション出来る人材を多く輩出していかなければならないと感じています。私自身が地方の大学ですので有能な人材が県外へしかも都市部に流出する傾向がより進むのではないかと懸念しております。「知」の格差も経済の格差と同様拡がっていくのではないでしょうか。(私立大学・事務職員)
- ○大学で教務課に所属しているが、学力評価の方法に疑問を持っており、参考になる話を聞け大変有意義であった。/特に Deane 氏の語る学習目標の明確化と明示化を進める必要性を強く感じた。(私立大学・事務職員)

- ○ディーン先生の話は興味深かったが、難しかった。同時通訳を聞くというところが限界がある(通訳者が専門的ではない)(塾・予備校関係・事務職員)
- ○採点基準の話に終始したのは残念だった。詳細まで公開しなくてもよいのでは、もっと骨太の力を学校では育成して***。もう少し幅広い視点のギロンを聞きたかった。(塾・予備校関係・事務職員)
- ○テーマに期待したが、広範囲に話が及び必ずしも「評価」について深められていないと感じた。/京 大の特色入試が京大全体の入試のイメージではない!(塾・予備校関係・事務職員)
- ○大変意義深い内容でした。今後の教育について非常にワクワクさせられましたが、一つ懸念があるとすれば、よりきめ細かい指導・評価となることになって、逆に手取り足取りになりすぎて、ある種「弱い」人間を作り出していかないかということです。大学に入ってから、また、社会において一人で黙々と文献を読んで問題を(紙の上で)解き続けるということを求められたとき、それを実現できる人が少なくなってしまうのでは、といことを少しだけ思いました。(塾・予備校関係・事務職員)
- ○今春突然復活した記述式について、採点は本当にできるのか、懐疑的だったが、石岡先生の話を伺って少し現実味を帯びてきた。(塾・予備校関係・事務職員)
- ○基調講演については、「高大接続における」が「学力評価の最新動向」と強くむすびついてはいないような感をもちましたが、Discussionで氷解しました。個々のテーマはそれぞれ興味深いものでした。 (塾・予備校関係・その他)
- ○Deane 先生の話は、たいへん参考になりました。報告 $1 \sim 3$ も、コンパクトにまとまっていて、良かったです。全体討論は、京大の話が中心になったのは、感想としては、良かったとも悪かったとも言えない微妙な感じです。全体としては、来てよかったし、また、開いていただければと思います。確かにインパクトのある話でした。(塾・予備校関係・その他)
- ○時間の都合もあるとおもうが、思考力、判断力、表現力の評価の観点が部分的かつ拡散的である気が した。どのようなものがより良いのかより今後の議論が必要だと思いました。ありがとうございます。 (塾・予備校関係・事務職員)
- ○白水先生のお話が、とくにストンと頭に入ってきました。アクティブラーニングを手さぐりで始めてみても、やはり現状はテストでの評価になりがちであり、まだまだアクティブラーニングの評価はあとまわしになっている感が強いように感じています。西岡先生の取り組みも含め、入試で見たいところ=学校で見るべき(育てるべき)ところなのだと改めて感じることができました。結局はすべての教育に関わる人々がまだまだ手探りなのだ、と感じつつ、私自身ももがいていきたいと思います。本日はそのためのヒントもたくさん得られたと思います。ありがとうございました。(私立高校・教員)

- ○個別の talk は大変勉強になりましたが、「中・高」の教育・評価の実践がどのように「大」につながるのかという大きなところ(高大接続)が、まとまり切らなかったように思います。/山地先生の討論のマネジメントが素晴らしかったです。(お疲れさまでした。)(民間企業・研究員)
- ○西岡先生、白水先生のお話は教育改革のなかで一つの手法として有効だと思った。勉強になった。 / 一方で(私は教師ではないが)、どのような能力・スキルを大学が求めようとしているか今一つ伝わらない。入試である以上、何を問うかだけでなく、評価の観点くらいは公開する必要があろう。 / ジグソーと CBAL の大きな違いは、CBAL は子供に応じて、認知レベルやステップを考慮してトレーニングする方法も尺度上の位置もわかることかと思う。 やはり、アセスメントが組み込まれていないと現場では使いにくい。 / センター試験は、ぜひ共通テストとしての役割を大切にするとともに、測定すべき能力・スキルを正しく評価して欲しい。 「ある程度機械クラスタリングができるのが30字までだから、30字の記述で」というのはやめて欲しい。 / 指定討論が高校の先生だったのは良かった。現場の多くの人の感覚と、大学の研究者の意識の違いがよくわかった気がする。 CBAL は両者の溝を埋める工夫もあるのではないか。 (民間企業・研究員)
- ○石岡先生のお話がとても興味深かったです。(民間企業・研究員)
- ○基調講演は大変興味ある内容であったが、通訳者による差なのか理解しづらいところがあったのが残 念です。(民間企業・事務職員)
- ○大変勉強になりました。表現や思考力の教育は、本日聞いたことがもっともっと現場に下りていくべきだと痛感しました。大学が求める能力や人物像を、小中高に向けて強く発信し、教育現場に影響を与えて行って欲しいと思います。私も、入試向け教材を作る立場ですが、学習を通して広い視野やスキルが身につくものを考えて行かねばと思います。今が変革の時期であると考えます。逆に、現場の変革は、入試が変わらなければ困難な状況だと言えます。痛みは伴うと思いますが、これからの子供達の為にも、本日伺った取り組みが入試改革へつながることを祈ります。(民間企業・その他)
- ○大人の立場からの議論しかない?子どもの側からみた入試は、こうあって欲しいとか言う機会はないのでしょうか。何のための入試なのか、大学の都合で採点基準を操作する?ともかく疑問点の多い話でした。/新しい課題にあたった時に、その解決ができる子どもたちを育てたいといいながら、入試側は、そのことを問うていない気がします。主体性を育てる入試のあり方という観点でみてみたいと思いました。/白水先生に質問です。将来、授業の流れの中で、子どもたちは伸びていくと思うのですが、その能力は、どのようにみとればいいのでしょうか。知識構成力は、どう身についているといえばいいのでしょうか?一つ一つの事例だけで、入試の判断にまで使う……という観点は狭い気がしました。また、いずれかの機会でお話しに含んで御教授いただければ幸いです。(民間企業・未記入)
- ○ポール・ディーン先生の講演はアメリカの事情がわかって、よかったと思います。 / 西岡先生の京大 AO 入試も、AO 入試について、よくわかりましたが、受験生が*るかについては疑問に思いました。

/白水先生の講演は、とても面白い内容でした、ほんとに授業の中での生徒の発言が記録できるのかと思いましたが、それができるとすれば、すばらしいと思います。/石岡先生のコンピュータによるテスト採点です。期待した記述(エッセイ)の記述の可能性がなかったのは残念。また、用語も聞きなれないものが多く、かんたんな解説を入れていただけると、わかりやすくなったと思います。(マスコミ関係・その他)

- ○大学入試改革が「豊かな」人間を育てるところにその目的の一つがあるのだとすれば、なんとなくそれらしい、普段の授業とは異なる、生徒も先生も何かをやった気になるような「探求的な」授業を高校以下で行うのでは意味がない。また、大学入試、大学側が高校の現場を振り回すような状況は好ましくない。一連の講演の中ではこの点が軽視されていたように感じた。全体討論中での司会進行、ディーン先生の話の辺りで若干触れられたのでよかったが、「豊かな」人材を育てる方法について、大学入試、大学側から、その選抜方法や内容から、明示的に進めていければ望ましいと感じた。/質問はすべてなんとかひろってもらいたい。(その他・事務職員)
- ○全てのお話が大変参考になりました。報告の時間が短かったので、もっと時間の余裕を持って聴きたかった。ありがとうございました。(その他・その他)
- ○Paul Deane 氏の基調講演は、きわめて刺激になった。米国のスタンスが良くわかった。また、西岡先生、白水先生の報告は短時間であったが、現在の「評価」に関わる具体的な動勢を認識する上で大変参考になった。京大(教育)・東大(法)の推薦入試についての観点もよく理解できた。石岡先生の報告もコンピュータで採点することの可能性がどこまで広がっているのか、充分に興味深いものであった。「報告」は20分ではかけ足であり、やはりもったいないように思う。清水先生の高校側の立場もよく理解できたが、やや今回のシンポジウムの内容には少し的外れであったように思う。全体討論は、大きな観点から、それぞれの問題意識の在り処が確認できて深く考えさせられた。このシンポジウムは高校側の声がきわめて届いていないように思う。もう少し「高大接続」のテーマを大学側でない人間を招いて議論したらよい。3月末まで地方の高校の校長職にあったものとして大変残念に思う。(その他・その他)
- ○「高大接続における学力評価」というテーマは非常に良い。ただし、まだ「最新動向」という段階ではなく、基本事項を共有する段階であるように思う。例えば以下の2点。 ①「学力評価テスト」における思考力の評価および条件付記述式 ②高校授業および大学入試における主体性の評価 特に②はそもそもルーブリックの手法も考え方も浸透しておらず重要と考える。本日の各テーマは有意義であったが、「最新動向」の各論の前に、こうした基本的なテーマがあるとよりよかった。(「入試センター」の「高大接続」の「評価」のシンポで入試改革、学力評価テストに1ミリも触れないのも「神業」だ)……それにしても各講演者の方々は、なかなか質問に的確に答えませんね。(その他・未記入)

- **5** シンポジウム等において、今後取り上げてほしいテーマや形態などがありましたら、御自由に御記入ください。
- ○文型・理系を問わず、国語(現代文)と英語は最小限残ると思う。それらのテスト形式がどう変わるのか示していただくと、新テストがどんなものなのか分かりやすいと思う。(国公立大学・教員)
- ○ア・ポリシーについて高校側あるいは出口(産業界等)の視点から改革課題を考えることもお願いします。/京大のポートフォリオ入試は興味があります。ポートフォリオを介した高大接続も取り上げていただきたい。(国公立大学・教員)
- ○大学の側としては、高校がどのように対応しようとしているのかという情報がそろそろほしいです。 (国公立大学・教員)
- ○高大接続教育における効果の高い具体的プログラムについての事例発表及び理想的内容の例示。(国 公立大学・教員)
- ○今回のテーマの続きを希望 (国公立大学・教員)
- ○センター試験の意義は? (国公立大学・教員)
- ○大学の現場では、改革が求められているなかで、少人数に対して、時間をかける選抜より、一般入試 に生かせるような取組みが聞きたい。(国公立大学・教員)
- ○継続して情報提供いただきたい。(国公立大学・事務職員)
- ○入試改革と初年次教育の連動性(私立大学・教員)
- ○2020 年度における大学入学希望者学力評価テストの全容と高大接続との相関性について/入試センター職員(Admission officer)の専門化について/教職員に共通する Application Review の重要性について(私立大学・事務職員)
- ○加速する少子化にともなう日本の高等教育をはじめとした教育のあり方について/都市部の大学と 地方大学の様々な格差/淘汰される大学~大学の生き残り戦略~(私立大学・事務職員)
- ○マーク式出題について、改革の課題である「思考力・判断力」を問う出題の可能性、研究状況(塾・ 予備校関係・事務職員)
- ○評価テストの準備状況。支援するシステムの構築など。ICT についてのセンターの研究。(塾・予備校

関係・事務職員)

- ○問題の作り方(作問)(民間企業・研究員)
- ○一昨年の ETS 山本先生の方が圧倒的におもしろかった。テスト分野での専門家(池田先生,前川先生,南風原先生,柴山先生,村木先生)を一同に集結して,最先端のテスト研究について語ってほしい。→テストの専門家の養成につながる。(民間企業・研究員)
- ○石岡先生の採点システムについて深く取り上げてほしい。(民間企業・研究員)
- ○新テストの作問例(民間企業・その他)
- ○試験の作問の仕方/具体的な評価の方法(民間企業・その他)
- ○高大接続の方向性と入試センターがその中で果たす役割/コンピューターによる文章採点について (マスコミ関係・その他)
- さらなる新しい情報をお願いします。(その他・その他)
- ○新テストの具体的な問題内容についての大学・高校現場からの作問試案について討議するような機会をこのシンポジウムに加えてほしい。(その他・その他)

以 上



