

市販の学力検査における現状と展望

村山 航*

2003年2月17日

於：東京大学大学院教育学研究科361教室

1 本報告の意義

市川：今日は教育心理学コースの博士課程の村山君の方から「市販の学力検査における現状と展望」について報告をしてもらうとことになっています。「学力検査」といっても学習行動に関するようなものも含まれています。

私たちは心理学を専門としていますが、市販の学力検査としてどんなものがあるのかというのはあまり知りませんし、どれくらい使われているのかということも十分に把握をしていません。そういうものを作っているのはもちろん心理学者です。いろいろな大学の教官が開発して、出版社・書店などから市販されています。今回はカタログなどを中心にして、できるだけそういうものを集めてみて、どんなものがあるのかということ村山君にサーベイしてもらいました。今後我々が新しいテスト・学力検査を開発したりするときの参考にもなるだろうと思います。どういうものが欠けているかというだけでなく、これから私たちが学習相談をしたりするとき、あるいは学力調査をするようなときにも参考にできることもあるかもしれませんので、その辺りをまずレビューしてもらおうということになりました。

それでは、村山君からお願いします。

2 報告の構成

村山：教育心理学コースD1の村山です。よろしくお願いします。

今、市川先生からお話があったように、市販の学力検査というものはどういうもので、どういう問題点があるかということをお話させていただきたいと思います。レジュメがB4のもの2枚と、あと資料が30数枚ある分厚いものが1部ずつあるか確認してください。資料が厚いことからわかりますように、できるだけ具体的に覚えてもらって、どのようなものかというイメージを持ってもらえればと思います。そして、皆さんのご意見などお聞かせいただければ幸いです。

では、レジュメの方を見てください。今回の発表では、田研出版と図書文化（教研式）などが発行している市販の学力検査に関して、その現状をまとめて、今後の展望を記すことを目的とします。ここに田研出版と図書文化というのがありますが、これが今、日本における2大心理検査のメッカのようなものです。田研というのは田中式知能検査の田研です。ですから、非常に知能検査と関係が深いです。

まず、どのような構成になっているかといいますと、学力検査は大きく3つの種類に分けられると考えられるので、それぞれ便宜的に、「領域別学力検査」、「観点別学力検査」、「学習状態診断検査」として、それぞれがどのような特徴を持っているのかということを紹介して、問題点についても考察したいと思います。4つ目のセクションで、学力検査における今後の展望についてお話ししていきたいと思っています。

* 東京大学大学院教育学研究科博士課程

3 領域別学力検査

3.1 「領域別学力検査」とは

それではまず3つのグループのうち、最初の領域別学力検査というものについて、これがどのようなものかということをお話ししたいと思います。領域別学力検査というものは内容領域ごとに問題を作成したものです。これを聞いてもよくわからないと思いますので、資料1の一番最初の所を見てください。

学習指導要領には、この学年のときにこういうことをやる、ということが領域ごとに記されているのですが、その領域ごとに問題を作成してテストが作られています。たとえば、社会のところを見ると、非常に大雑把ですが地理分野・歴史分野・公民分野と分類されています。このように領域ごとにまず分けて、そこから問題を作成するというものが、この領域別学力検査です。

もう少し具体的にどういうものがあるかということ、1枚目の資料の右側の方を見てください。これは2年生の英語の学力検査です。たとえば、「聞くこと」に関する問題があるのですが、それらが指導要領とどういう関連があるのかということが書かれています。このように指導要領から満遍なく出されているというのが領域別学力検査です。

3.2 領域別学力検査の特徴

特徴としては評価形式が相対評価ということがあげられます。問題は標準化されていまして、普通の標準学力テストと考えていただければいいのですが、平均は50点程度で、ある程度分散も収まりがあるように標準化されています。

先ほど、中学1年生や中学2年生などとあったのを見ればわかるように、テストは学年別に作られていまして、だいたい年度末か、もしくは新学期に、前の学年の内容がどれくらい習得できているかチェックするために使用されています。

そして、領域別学力検査には他にも1つ特徴がありまして、普通の学校のテストなら50分かけてテストをやれば終わりですが、ここでは先ほどの領域ごとに時間が区切られています。例えば、1枚目の資料のところをもう1回見ていただくと、言語だったら15分、表

現だったら15分というように領域ごとにかかる時間に偏りが出ないような工夫が凝らされているのがわかります。

3.3 検査の標準化—妥当性について—

具体的にどのような標準化がなされているか、ということですが、これについては資料2を見てください。最初に教研式と田研の2大メッカがあると言ったのですが、ここには田研の標準化の内容が書かれています。だいたい、信頼性と妥当性を確認して標準化をするという形になっていますが、ここでは特に妥当性について載っています。

内容的妥当性というのは、学習指導要領にきちんと沿ったものが出ているかということを見ることで、それに関する詳細は、この学力検査の資料には載っていませんでした。そのほかに正答率などの観点から項目を精選ということで、田研のものにはこのようなことが結構詳しく書かれていました。「妥当性を高める方法」というのが2枚目の資料の下側に書いてあります。先ほど述べたように正答率が50%になるようにしながら、できるだけ弁別力の高い項目を選ぼうとしています。これについては、次のページの、第2図というところを見ていただくと一番わかりやすいと思います。(ア)、(イ)、(ウ)という3つの曲線がありまして、これは学力が高いと正答率がどうなるかということを表わしている曲線ですが、その中である程度角度が急な項目というものを選ぶということです。ただし、それなりに出題内容の意味というものもあるので、そういうものも含めながら選んでいます。どのような問題を選んだのかということに関しては、やはり具体的には書いてないのですが、とりあえずこのように項目特性曲線のような形で、問題を選んでいるということが書かれました。そして、このような問題については、資料がまた次のページに続きます。このようなことが例としていくつか載っていて、全体の得点がこれくらいの人はこの問題ではどのくらいの正答率であったか、ということがわかります。

3.4 検査結果のフィードバック

結果がどのようにフィードバックされるかということについては、学校の方で答案を集めて先生が採点す

ることもできるのですが、基本的には返送してコンピュータ採点となります。それで、返されるのは学力偏差値です。これは領域ごとの学力偏差値であったりとか全体の学力偏差値であったりとかします。これは普通の模試と同じようなものです。

他に返ってくるものとして特徴的なのが、「総合成就値」というものがあります。これはいわゆる「アンダーアチーバー」とか「オーバーアチーバー」を見つけるための指標です。田研が作っていることからわかりますように、市販の学力検査というのは非常に知能というものの意識が強くて、知能検査との差がどれくらいかということに関する指標がかなり事細かに説明されています。これらが先生の方に返されます。それから、学級・学年ごとの集計表や、小問ごとにクラスでの正答率・誤答率などを表わしたりして、教師としては、「このクラスにはこういう弱点があるんだ」というのが一目でわかりやすいような資料となっています。

3.5 領域別学力検査の問題点

考えられる問題点の方に行きたいと思います。

3.5.1 未公開情報の多さ

このように標準化がなされているということですが、問題制作過程や調査協力校など、標準化テストとして必要だと思われる資料が未公開であるということが挙げられると思います。そのため、たとえば、「平均が50点に標準化されています」といわれたときに、どこにとって50点であるかということがわからないと、使う方としても困ることになると思うのですが、そのようなことが書かれていない。少なくともランダムサンプリングをしたということも書かれていないので、おそらくそういうことはやっていないのではないかと思います。このように非常に未公開の部分が多い、ということが挙げられます。

それから、同じようなことなのですが、どの学校ということだけではなくて、どの時期のデータを使って偏差値が算出されているかということもこの資料には書かれていません。特に近年は学力というものは経年的に変化しているという話もありますので、どの年のデータをもとに標準化したかということ是非常に重要

な情報で、頻繁に更新していく必要性もあると思います。しかし、そのような更新の過程なども全く何も記されていません。

さらに、テスト制作者が明記されていないということもあり、過程が不明確であるということに、全てつながっていきます。本当にどのように作られているのかということに関することについての検討、それから情報公開というものが弱いのではないかと思います。

3.5.2 解答形式の偏り

それから次の所ですが、多くの問題が多肢選択形式、もしくは単答形式より構成されている、ということがあります。ほとんど多肢選択問題です。都立入試などでもほとんど多肢選択と穴埋めとなっているのですが、それと同じような形になっています。例えば計算問題で計算の過程を聞くということはまずありませんので、結果としてどのような答えが出てきたかということだけをみて、どのような過程を経てそのような答えが導き出されたかということは、全然わかりません。

それで、次のefというところを見てください。英作文の問題もあるのですが、整序問題なので、十分に理解していなくても感覚で解答できる可能性が十分にあります。算数・数学は計算過程が問われず、また証明問題というのもあります。都立高校入試などではほとんど穴埋めですが、それ以外にも証明くらいはやらせるのですが、そのようなことも全然させていません。

3.5.3 出題内容の制約

それから、次は一長一短だと思うのですが、学習指導要領の教科内容による拘束が大きいということがあります。例えば、今、学習指導要領というのはそれを最低基準として、それ以上のことをやってもいいということになっています。しかし、学習指導要領による拘束が大きいので、それ以上のことをやった人の学力というものが何も測られていないということになります。特に私立学校などではそれ以上のことを教えるということはかなり積極的にやられているように思われます。しかし、そのような部分は見ずに学習指導要領の範囲で、どのくらいできているかということが問題となっています。したがって、そのような中で偏差値を出すことにどういう意味があるのかと思われる。

これまで紹介したのは、領域別といったのですが、どちらかというと古いタイプの学力検査で、教育改革が行われる前の時代の学力検査が今でも続いているという感じのものです。近年の教育改革に伴って新しい学力観というものが出てきたのですが、それに対応した形式というものが、これから述べる観点別学力検査になります。

4 観点別学力検査

4.1 「観点別学力検査」とは

次は観点別学力検査についてお話ししたいと思います。こちらの方も皆さんにじっくり見てもらいたいと思います。観点別学力検査というのは、新学習指導要領で取り入れられた「観点別評価」の各観点を評価するために問題を構成したものです。「観点別評価」というのは例えば「関心・意欲・態度」とか「思考・判断力」などです。

4.2 問題例

具体的な問題を見ていただければと思いますが、資料4が比較的長く続いているのでざっと見てください。おそらく皆さんが中高時代に受けてきたテストとは随分違ったものがあると思います。

4.2.1 数 学

たとえば、資料4の一番最初は計算問題で、③の問題では計算過程を一応問うている問題になっています。ただし、答えがわかればあまり計算過程は関係ない問題です。このように計算過程を問う問題であったり、資料4の2枚目には山本君、田中さん…というマンガが描かれている問題がありますが、答えだけを聞くのではなくて、なぜそうなるのかということを知っている問題になっています。また今のページの左側の④のところと同様に過程を知っている問題があります。それから、また次のページを見ていただくと、たとえば右下のなど、右側のページは全部そうになっています。③、④も計算の過程を知っている問題です。

4.2.2 英 語

それから次のページは英語です。これは特に普通の部分もあります。これに関して個人的に思ったのは、右側にteaという単語に脚注がついているように、英単語の量が最近減ってきているというのが非常に感じられるということです。

次のページを見てください。次のページで特徴的だったのが、一番右下の問題に、George Washington…というのが載っているのですが、ワシントン大統領はアメリカの何代目の大統領か、という問題になっています。これはfirstとfastestというのが入っていてわかるのですが、何のための問題なのかよくわかりません。後でも述べますが、このようにいろいろと新しい問題を作ろうというのは非常によくわかるのですが、何を聞いている問題なのかわからなくなるような問題も増えているような感じがしました。例えば、今述べたジョージ・ワシントンの問題の上にある3番では、アメリカの学生は土曜日に学校に行かない、という内容の問題ですが、そのようなことを尋ねる問題は、英語の問題なのか社会の問題なのか、わかりません。このように、普通のいい問題と見受けられるものもあれば、いまいち何を聞いているのかわからない問題があることが特徴だと思います。

4.2.3 社 会

それから次のページは社会です。これはグラフが増えているとか写真を使った問題が増えているという程度です。

4.2.4 関心・意欲・態度を測定する質問紙（教研式）

それから次のページを見てください。後でも述べるのですが、関心・意欲・態度というものをどうやって測定するかという問題が出てきます。観点別評価というのは普通、4つの観点がありますが、そのうち特に重要だといわれているのが、関心・意欲・態度です。これに関して、会社によってスタンスが違うのですが、この教研式では質問紙で測っています。そして下のプロフィールというところを見ていただくとわかりますように、「関心・意欲・態度」得点と「社会的な思考・判断」得点と、「資料活用の技能・表現」得点、それか

ら「知識・理解」得点というものが算出されるといった形式になっています。この表の見方についてはまた後で、やっていきたいと思います。

4.3 評価形式

今資料4を見ていただきましたが、またレジュメの方に移ります。評価形式についてですが、先ほどの領域別検査では偏差値を使った相対評価だったのですが、ここでは到達度絶対評価ということになっています。先ほどは平均点が50点でしたが、ここでは平均点は70～80点ということです。平均点は学習者の意欲を保たせるようにある程度高目が望ましいということが記されてありました。

4.3.1 関心・意欲・態度の評価（田研）

関心・意欲・態度というものは今述べたような質問紙で測定するものもありますが、次のページの資料5というのを見てください。これは田研の方ですが、問題で関心・意欲・態度を測定しようとしています。それで資料5はこの問題が何を測定しているのかということに関するものです。問題番号①で、「関」というのは「関心・意欲・態度」を測定している問題だということで、「資」というのは上に書いてあるように「資料活用」を測定するための問題だということです。

次の右側に「DR T 数学1年」とあるところを見ていただくと、具体的な数学の方の問題があります。これは今回の資料にもありまして、資料4の一番最初のページです。ページ番号が書いていなくて申し訳ないんですが、左側に「資料③」と書いてあって右側に「資料④」と書いてあるところです。この①が、この資料5の右側の問題に対応しています。すると1番は単なる計算の答えを聞いているというわけではないのですが、 $(-5) - (-8) = (-5) + ()$ となっていて、これは知識を問うている問題であるということです。2番は知識を問う問題です。3番は計算の過程が日本語で書かれていて、ア、イ、ウ、エから選ぶ形になっていますが、この問題は、表現・処理という観点を測定していると同時に、関心・意欲・態度も測定している問題であるということになっています。

そのように関心・意欲・態度を単独で測定する問題はないのですが、それを測定している問題が時々紛れ

込んでいて、問題を通して関心・意欲・態度を測定しようというようなスタンスです。

4.4 特徴

またレジュメの方に戻ります。先ほどと同じようにテストは学年ごとに作られていて、年度末もしくは新学年に昨年の分を実施するようになっています。そして先ほどの領域別では領域ごとに問題が分けられていて、時間制限がありました。こちらの方は、やはり田研と教研式で少しスタンスが分かれています。教研式の方は知識とか思考判断とかを別々の問題として領域ごとに制限時間を設定しています。資料5にあるのは田研ですが、この問題を見ていただければわかりますように、田研はどの問題が「知識」で、どの問題が「関心・意欲・態度」で、どの問題が「思考・判断」かというのは特にグルーピング化されているわけではありません。このように全部一緒くたにして一度にやってもらうというのが田研のスタンスです。

それからもう1つ特徴的だったのが、これは両方ともそうなのですが、時間の制約がルーズであって、余裕をもって終えられるようになっています。しかも場合によっては時間延長してもよい、ということは資料に明記されていまして、おそらく教師もそれに合わせて時間延長を適宜行っているのではないかと思います。

4.5 問題作成の方法（教研式）

では、このような問題がどのように作られたか、ということですが、それについては資料6を見てください。これは教研式です。この観点別学力検査というのは、田研に比べて教研式の方がかなりしっかりと作っているという印象を受けました。そのままコピーしてきたものですが、どのようなことが書かれてあるかというと、資料6の2枚目を見てください。

4.5.1 項目の精選

「どのようにして標準化されているか」というところをみていただくと、(1)からこの過程が記されています。それを要約したのがレジュメの方にもあります。観点ごとにまず「目標の具体化表」というものを作

成するようです。そして、「この具体化表に含まれる全体の目標を一つひとつテストすれば、“真の到達状況”がつかめるが、現実問題としてそれはできないので、以下のように考える。」と書いてあります。この教科で「知識」とは、あるいは、どういうことができれば「知識」を習得したことになるのか、という抽象的な目標を立て、その目標を体現するような、さらに具体的な目標を選び出して、それに対して作問するという形になっています。非常に抽象的ですし、どのような具体化表があるのかということが書かれていませんでした。後でもう1回、一番最後に、似たようなことをやっている具体例を出しますので、また戻ってきますけども、そのようなことが書かれてあります。

そして、そのアイテムを複数のジャッジによって、いいか悪いかというものを審査してもらい、不適切なアイテムは捨てたり作り直したりして、アイテムの妥当化を行うということです。

4.5.2 到達度基準

そしてその次ですが、資料の(4)のところを見ていただくとわかりますように、到達度評価というのは絶対評価なので、本来なら、みんなに実施して、みんなこれくらいできるから、これくらいの点数がとれたらA、という判断をするのが常識だと思うのですが、そのようなことは一切していません。到達基準分割点というのは、やはり専門家が決めるというような手続きになっています。ですから専門家が、「これくらいできればこの目標は何パーセント達成ということになる」ということを決めて、基準を定めているわけです。

具体的にどのようにやるかということですが、下はかなり細かいことが書いてありますので、興味のある方は見ていただければと思います。到達度基準に関しては一定の手続き、「修正エーベル法」というものが、この資料の続きに書いてあります。これを一言でまとめますと、「原則として見識ある人の良識的判断に基づく事前設定」となっています。ですから、今述べたように、専門家が、これくらいできれば何パーセント到達したことになる、ということを複数のジャッジをもとに行き、これくらいの点数がとれればA、これくらいとれればB、ということを決めているということになります。

資料の前の方に書いてありますが、資料4の最後に

領域ごとの分割点というものが記されています。資料4の一番最後、右側に「TK式観点別到達度学力検査」という大きな字の表紙が書いてあるページです。この左下の所をみていただくと、先ほど観点ごとに得点が出ると述べましたが、何点以下だったら「努力を要する」、何点以上だったら「おおむね満足」、何点以上だったら「十分満足」というのが右側に書いてあります。この分割点が今のような過程を経て作られたということになります。たとえば「社会的な思考判断」は7.5点以上だったら「概ね満足」、10.5点以上だったら「十分満足」となっています。そして、これはどのようにして定められたかということ、複数の専門家がジャッジをして定めたということです。以上が項目の精選、いわゆる妥当性の検証ということになります。

4.5.3 信頼性の検証

次に、信頼性が報告されています。信頼性はどのようになっているかといいますと、資料7です。折半法と再テスト法による信頼性が書かれています。それなりに高い値ですので信頼性の方はなかなかしっかりしているのではないかと思います。このように資料として、信頼性が明記されています。

4.5.4 結果のフィードバック

結果のフィードバックですが、レジュメの方に戻りますけれども、先ほどと同じように基本的にコンピュータ採点ですが、学校での採点も可能です。フィードバックされるものは先ほどの表にもありましたように、観点ごとに「十分満足」、「おおむね満足」、「努力を要する」の3段階の評価です。それから、教研式は先ほど述べたようにジャッジが決めるのですが、田研だけは標準得点にこだわっていて、標準得点もあわせて提出されます。このあたりが出版社ごとのスタンスの違いだと思ってください。

それから教師のみに返ってくるものとして、学級・学年集計表があります。また教研式は相対評価をしていないのですが、全国の正答率がどれくらいかという資料が配られ、実質的に相対評価と同じではないかという気がします。あとは小問分析表というものも配られます。

4.6 考えられる問題点

考えられる問題点は以下の通りです。

4.6.1 解答形式の限界

4つの観点を選定するために、いろいろと工夫して問題も検討されているように感じます。それはここまでのをずっと見ていただければわかります。しかし、その反面、基本的にマークシート方式で、特に教研式はそうになっていますが、理解力や表現力を測定する限界があると思います。

4.6.2 妥当性の統計的根拠がない

レジュメの次のページに移ります。特にやはり問題となっているのは妥当性の問題だと思います。こんなに頑張って作りましたということは書かれているんですけども、それは内容的な部分に関してだけで、何かの基準とこういう関係があるという、基準連関的な妥当性が示されていないために、本当にこのような問題で、思考・判断力というものや理解というものを測定できているのかということが不明確です。あくまで専門家の良識というものに任せて作られていますので、統計的な根拠というものがありません。つまり、内容的妥当性を非常に強調した書き方になっていまして、何か他の尺度との関係があるということは一切書かれていないわけです。

4.6.3 観点別評価の方法

また、教研式と田研で考え方が大きく違っているということがありました。たとえば、関心・意欲・態度というものを問題で測るか、質問紙で測るかということ、それから標準得点を積極的に取り入れるか否かなどです。これは逆にいうと、新学習指導要領で観点別評価というのが導入されたのですが、これをどのようにして評価・測定するかということに関して、出版社ごとの合意が得られていない、未だ混乱が続いている状態であるといえます。

4.6.4 観点と領域の交絡

それから、1つの観点に対して全ての領域が満遍なく問題構成がされているわけではなく、観点と領域との交絡が生じていることが考えられます。よくあることだと思いますが、この問題だったらじっくりと思考・判断ができる、この問題、この領域だったら理解はできるけれども、この問題だったらできない、ということがあります。しかし、この学力検査では、たとえば思考・判断を問う問題では同じ社会科の中では歴史に偏っていたりとか、そのようなことがあったりするわけです。このように、あまりに観点別に測るということに特化しすぎたために、そのほかの観点がかなり見捨てられてしまっているという感があると思います。どの領域の問題を用いるかということをもう少し系統的に考えて問題作成をした方がいいのではないかと思います。

4.7 その他の学力検査

4.7.1 学習進度指導検査

ここまでのものが2つの大きな学力検査である、領域別と観点別の学力検査です。その他の学力検査として何があるかということだけ付け加えておきました。学習進度指導検査というのがあります。今までは1学年分の問題だけだったのですが、これは3～5学年前にまで遡って問題を作成ということで、いわゆる低学力者に向けての学力検査です。ただし、語彙と計算に限るということです。

4.7.2 読み能力検査

それから読み能力検査というのがあります。これは語の認知とか文の理解とか音節の抽出など、いくつかのコンポーネントに分けてそれぞれを測定するという、非常に認知心理学の知見を活かした学力検査であると思いました。しかし、下に書いてあるように、10年から20年以上前に作成されて改訂されていないので、その妥当性には疑問が生じます。

資料8を見ていただくとわかると思います。これは金子書房、読み能力検査ですが、これは語彙のところを問う問題です。27番「かち」というのがいきなり出

てきています。これは「徒歩」です。それから、「評定（ひょうじょう）」が29番にあったり、ちょっと僕でもよくわからない43番「局方薬（きょくほうやく）」というものがあったり、「相称（そうしょう）」など、東大生でもわからないような問題が出ていたりします。これは中学生向けの問題です。中学生でも、今、金子書房に「読み能力検査」を取り寄せるとこういう問題が送られて来るというわけで、全然更新されていない実態というものがわかります。

これはおまけの部分なんですけども、読み能力検査というのが特に他の分野より遅れているような感がありましたので書いておきました。

5 学習状態診断

5.1 「学習状態診断」とは

最後の3番目の検査、学習状態診断としてどのようなものがあるかということですが、これは質問紙です。主に学習意欲・身体の健康・学習環境・学習方法・自己学習力などを測定する質問紙です。測定している構成概念は質問紙ごとに違っています。たとえば、資料9の1枚目に、教研式SETと教研式AAIを左側と右側に載せておきました。たとえば教研式SETというのは下に書いてあるような集中力、計画性、自主性、課題意識、主体的思考などから成り立っています。右側のほうは教研式のAAIというものですが、これは学年ごとに測定する構成概念が違います。勉強の意欲からテストの受け方、勉強の技術、ノートの取り方、友人関係、学校環境などが記されています。具体的な質問項目としては次のページに載っているのを見てください。

これは例えば、21番、22番というのは学習の仕方を測定する質問項目です。それから次の方には田研の質問項目が載っています。右側には例えば学習環境を測定する項目として親との関係を問う質問があります。

「私は親にしかられたときにどうしますか」など、父母に関する質問がいろいろとあります。このような質問から成り立っていて、学習環境の測定がなされるわけです。

5.2 質問紙の作成—信頼性と妥当性の検討

では、どのようにこの質問紙が作成されているかということですが、これは僕が知っているような心理の専門家の先生方が入ってやっけていることが多いようです。項目の収集は専門家が作成したものについて、因子分析、信頼性・妥当性などの検討を行っています。それに関しては資料10というところを見てください。資料10の右側のページです。妥当性の検討として、アンダーアチーバーとオーバーアチーバーで得点が違うのでこの尺度は妥当であるという形で妥当性の検討が行われています。また、資料10の次のページには一応信頼性について書いてあります。それから次のページにも同じく教研式の信頼性（信頼度）というものが書かれています。右側には具体的な数字が載っています。このように、因子分析によって信頼性・妥当性などの検討が行われているようです。資料としては因子ごとの平均値や標準偏差、学年ごとのものもあるということです。

5.3 考えられる問題点

5.3.1 教育心理学の知見が生かされていない

考えられる問題点としては、まず、教育心理学の専門家が入っていると思うのですが、それにしてもあまり教育心理学の知見が生かされていないというようなことがいえます。動機とか学習方略というものは、教育心理学の分野ではかなり細かな概念に分かれています。しかし、たとえば学習方略では「よい勉強方法はこれです」というのが1個だけあって、それをしていくかどうかということ、複数の項目で測定しているだけです。動機づけというのも非常に漠然としていて、授業が楽しいとかそのような感じのもので1つの「動機づけ」という形で取り上げられていたりします。

教育心理学の方で「原因帰属」という、動機づけの中で特定の構成概念があります。これに関してだけは独立した1つの尺度になっていたのですが、その他に関しては、やはり漠然と意欲があるのかどうかという形で測定されているということがありました。

5.3.2 学力との相関が低い

さらに問題点は続きます。資料11を見てください。資料11の方には妥当性の検証として、先ほどのSETというものと他の認知的特性との関連を示した関連の表があります。向きを変えないとわかりにくいのと、次のページにもあって見にくいですが、縦に並んでいるのがSETの項目です。右に並んでいるのが英数国の学力や、次のページにいくと知能とか創造性など、その他の認知的な構成概念が並んでいます。これを見ると全般として相関が非常に低いということがわかりかと思えます。ほとんどが0.1台です。書いていないところは無相関だということで、非常に低いです。これが「妥当性」という所を書いてあるのが、なかなかすごいと思うのですが、こういう資料が書いてあります。

今の資料11の左下の文章、20ページの下の方、一番最後の段落を見てください。「しかし、全般的には相関値が低いことがわかるが…」という一連の文章がありますが、「このことは今日の学校では、生徒の自己教育力がそのまま発揮される場が少なく、むしろ周りからの圧力によって学習させられ、徹底した管理の下に、他律的に学んでいるという状況を反映しているとみなされる。」と書いてありまして、妥当性の検証なのか学校の現状批判なのかよくわからない、非常に本末転倒な感じの書き方になっています。原因としてはやはり構成概念として、漠然としたものを測っているために、全体としての相関が薄まったのではないかと考えられます。

したがって、市販の学習状態診断検査というものは、結構いろいろと広まってはいるようですが、それほどきちんとしたものが測れているかということ、かなり疑問の余地があると思われまます。

5.3.3 解答のしにくい項目

それから問題点としまして、先程挙げたように親子関係など、解答のしにくい項目があるということがあります。

5.3.4 妥当性の検証の仕方が弱い

そして、妥当性の検証が行われているが、全般的に

弱い。弱いというのは相関が低いということだけではなく、先ほどあったアンダーアチーバーとオーバーアチーバーを比べる程度のことしかなされていないので、もうちょっといろいろなことがやれるのではないかと思います。そういうところも見ていけばいいのということを感じました。

6 今後の展望

最後に、今後の展望について述べます。

6.1 「学力」の意味の多元化に対応するテスト作成

今後の展望で、学力の意味の多元化に対応するテスト作成ということを書いたのですが、今回学力検査をいろいろ見ていただいて、とりあえず評価するために3つの次元というものが見えてきたかと思えます。1つ目は内容領域別の学力というものです。たとえば、あるところではできるけど、あるところではできない、「2次関数は大丈夫だけど図形はダメ」というような内容領域別の学力というものがあります。それから観点別の学力というものがあります。「知識はあるけれど思考力がない」、「知識はないけど意欲はある」というような観点別の学力というものへ変わってくるということも確かにあります。3つ目は基準によって変化する学力です。これはちょっとおまけとして書いたのですが、要求した基準に到達していないけれども偏差値は高いということも考えられますので、到達度別という考え方と、相対評価という考え方の、2種類出てきたと思います。

それと、みなさんも感じていらっしゃると思うのですが、どの次元も学力を捉える上で重要であり、不必要な次元はないといえます。最近特に偏差値評価などが叩かれる側に回っていると思うのですが、やはり偏差値評価は偏差値評価で重要な部分も多々あると思います。観点別とか到達度という考え方が入ってきたからといって、内容領域別の学力や相対評価というものが決して不要になったわけでは決してありません。したがって、次元がシフトしたのではなくて、多元的になっているととらえるのが正しいと思います。ところが、今の学力検査、新しい方の観点別学力検査では、

どちらかという次元をシフトさせていて、残りの部分はほとんど見ないという方向に向かっている部分があると思います。しかし、それらの次元は全て大切なものですから、それらの次元を交絡させることなくテストは作成されることが望ましいのではないかと思います。

先ほど述べたように、観点別学力検査も内容領域別の学力が交絡しているのです、テストで測定しているのが何の次元における違いなのか明確ではないということがあります。そのため、思考力があるという結果が出たといっても、それがたまたま図形の問題が得意だっただけ、という可能性も十分あるわけです。そのようなことが見極められないのが問題ではないかと思えます。したがって、このような上記の次元を意識した上で、テストのバリエーションを増やす必要があると思います。ただ、こうすることによってテストの数が増えるという問題が出てきます。しかし、今の学力検査というものは、次元を多次的に意識しているというよりは1次元的に見て、とりあえずそれだけ見ようというような感が強いと思われまますので、むしろ複数の次元があるんだということをもっと意識した上でバリエーションを増やすことが大切なのではと思います。特に観点別学力検査では、1つのテストで多くのことを測定しようとしすぎている感があります。ただし、今述べたように、実際に学校で実施する場合には時間の問題など、現実的な制約も多く、それらとの兼ね合いを考える必要があると思います。今言ったことは1つの提言でありますので、本当は現場の先生方がいらっしゃったら、どういうことを考えているのかということも聞きたいところです。

それからもう1つ、今回、これの整理に追われて時間がかかってしまった理由として、会社ごとに好き勝手な名前を付けて、1つの会社から複数のテストが出ていることが挙げられます。これらのテストにどういう違いがあって、どういようなものなのかということは、カタログを見ただけでは全然わからない。今回こういうふうに簡単にラベル付けしたのですが、もう少し統一感があって、しかも一見しただけでわかりやすいようなラベル付けや構造化ということをしないと、現場の先生にとっても使いにくいのではないかと考えます。

6.2 「新しい学力観」を適切に反映した妥当性のあるテストの作成

6.2.1 関心・意欲・態度の測定

それから次の(2)ですが、新しい学力観を適切に反映した妥当性のあるテストの作成ということが挙げられます。まず、新しい学力観で1つだけ別物として扱わなければいけないものが関心・意欲・態度です。これは、どの会社も無理矢理入れているのですが、これはむしろ現場で見ればわかることなので、あえて測定する必要もないのではないかと、質問紙で測定するというよりは、実際の場合でどうなのかということ先生が詳細に見ていくことが必要なのではないかということが考えられます。

6.2.2 思考・考え方・表現力・判断力の測定

次に、思考・考え方・表現力・判断力はどうするかということですが、2つの方向性があると思います。1つはマークシートをやめて記述式などで測定ということで、評価の形式などを変えていくということです。ただ、コンピュータ採点をしていた市販の学力検査では、こうすることによって採点に伴うコストの問題も大きく絡んでくると思われます。また評価の客観性という問題も出てきますので、これらの点についてある程度クリアした上でないと、これはなかなか現実的なようで現実的ではないものになってくるという気がします。

では、あくまでマークシート方式で測定するというのですが、現在、教研式も田研もこの路線です。資料4のところで見ただけだと思いますが、それなりに頑張って作っているという印象はひしひしと伝わりますが、不十分な感が否めません。

それで、例えばどのようなことができるかということいろいろ考えました。これは観点別というよりは、こういうような学力のとらえ方もある、ということで1つ挙げられるものとして、例えば教育心理で「誤概念」や「バグ」に関する研究というものがあります。資料12という所を見ただけだと、教育心理の方はご存じと思うのですが、たとえば一番左の図で歩いているときにボールがどのような落ちるかというものです。これは慣性の法則を考えれば、Aというような考えが

出て、Aが正答になるわけなのですが、解答のパターンとして、Bと間違える人が多いことがわかっています。また、左下の問題でも、正答と最も多い誤答というのが書かれていて、チューブに水を入れたらどこに水が出るかというものでCと解答している人が多いという知見があったりするわけです。それから右側の方では電流の流れ方について子供たちがどう考えているのかということが出てきています。それから次のページを見ていただきますと、左にあるような繰り下がりのある計算の時に、人はどういう間違いをするかというのはやはり教育心理学でかなり系統的な知見が導き出されていて、誤答のパターンというのはわかっています。右側に誤答のパターンがいろいろと記されているのですが、せっかくこのように教育心理学で誤概念という形で、人がどのように間違いやすいのかということがわかっていますので、そのような方面から学力というものを測定することも考えられるのではないかと、思います。そしてそのような段階に達しているか達していないかという観点から測定することも、1つのアイデアとして可能なのではないかと思います。

6.3 テスト評価後のフォローを考える重要性

それから(3)はおまけですが、テスト評価後のフォローを考える重要性ということで、「Plan-Do-See」という評価観から、「Plan-Do-Check-Act」という、評価後のActというものが重要だということがよく言われるようになりました。それで、「〇〇の部分がよくなかった」ということがわかったときに具体的にどうすればいいのかということは何らかの形で検査セットの中に取り込む必要があるのではないかとことです。特に学習状態検査のような場合には、そのような配慮が必要です。一応、学習状態検査では、副教材としてそのようなものが販売されていますが、どちらかというと、その先生が書いた本を売りたいがために内容の薄い本を売りつけている感もありますので、こういう学力検査を出しているところは、もう少しきちんとフォロー、ケアというものを考えていくことが大切なのではないかと感じました。

ちょっと資料がいっぱい急いで進めてしまい大変申し訳ないのですが、これで発表を終わらせていただきます。

7 討 論

市川：はい、ありがとうございました。この後30分くらいを目安としてディスカッションに入りたいと思います。ご意見や質問など、どんなことでもどうぞ。

7.1 学力検査の使用目的について

清河：質問なんですけど、学校で行われている市販の学力検査を使いたいというときには、どういう目的で使われていることが多いのでしょうか。実際の学校でもこういうのではない形としてテストをして、子供たちの習得の度合いを測って、そちらの方は割と授業につなげていたりすると思うんですが。資料として残すためにとりあえず形式的にやっているというのがあったのですが、用途としてはどういうものがあるのか、また、実際としてはどうなのか。

村山：実際としてはやはり学年末か新学期が始まる前に、まず学年末でしたら、授業の内容がどれくらい習得できているかということ、それから新学期でしたら、新しい担任の先生が、この学校の生徒はどのような学力を持っているのかということを判断するために使っていると思います。もちろん、その学校でやった授業内容に即した問題作成というのが重要だと思うんですけど、1つには相対的な基準を見るため、これだと到達度の部分が怪しくなってくるんですけども、そういうためにあるということ。

もう1つはやはり先生方も自分の作ったテストよりは、こういうある程度統計的な観点に基づいて標準化されていると安心できる。しかもこれは指導要録に参考資料として書き込んで、その後の指導にも生かせるし、引き継ぎの資料にも生かせるということで使われていたり。実際に使われている普及率とかは会社に聞いても教えてもらえなかったもので、ちょっとよくわからないですけど。

市川：今の話ですけど、少なくとも学校の先生は自分でテストを作って実施したり、それからいろんな教科書会社、あるいは教科書会社に限らないですが、教科書に準拠したドリルやテストをいっぱい作っているじゃないですか。かなりやられているものだと思うんですね。そういうのとちょっと違いは出したいですね。見てみると、結局、これはある意味で教科書フリーでしょ。どの教科書を使っていようが、これをやる

と全国的な、例えば、位置というか、相対評価で偏差値が出るとか、それから少し細かく診断結果が出るみたいに。それから学力診断。学校の先生は自分で作ったのだと何もそこまで細かい点数が出てくるわけではないし、教科書に準拠しているテストでもそんな点数は出てきませんよね。要するに100点満点中何点だったかということしか出てこない。そういうのに比べるとそういう情報が得られますよっていうのが売りなんじゃないですかね。

ただ、現場の先生が本当にそういうことをこれを使ってやっているかということ、最近はどうも疑わしい。これで学力診断をやっている学校って僕はあまり最近多くないんじゃないかなという気がしますよね。知能検査ですらやらなくなっているし。だから乗りは非常に知能検査に近くて、これをやっていると標準的な学力とか、あと知能検査だと、サブカテゴリごとによっぱい、いろいろな点数が出てきますよね。そういうような形で観点別の方が客観的な点数として出ていますよ、というのを売りにしているんじゃないですかね。

村山：教科書ベースということなんですが、教研式は教科書別です。2つのカテゴリに分けて。

市川：教科書準拠なんですか。

村山：はい、ある程度。田研はそういうのは検査ではないといって、暗に教研式を批判して書いてあるようなところが。

市川：じゃあ教科書が改訂されるたびに教研式も全部改訂しているのかな。それはものすごい作業ですよ。よほど売れていないとできないことですよ。

7.2 妥当性の問題

南風原：まあ実際どれだけ使われているのかわからないけど、あまり使われていないだろうというテストだし、あまり役に立たないだろうという感じですね。今、村山君の方からはだいぶ工夫して作っているというのはわかるということ、内容的な妥当性は調べてあるということが言われたけれど、君が見た範囲でもジョージ・ワシントンの問題だとか、何を聞いているのかわからない、なぜそれが「関心・意欲・態度」なのかわからない問題だとかといったような、工夫しているというよりは、何かこれが教育心理学者の学校教育との接点かと思うとハラハラしてきて落ち着かない。ちょっとアカデミズムよりはコマーシャル

イズムで動いているし、片手間でちょっと恥ずかしい仕事をしているなという感じですね。そういう実態を知るというのも1つ大事だと思うし、こういうテストの実態を調べ、また活用の実態を今度は学校に入って聞いてくるということも大事だと思うんですが、やはりこの領域というのはアメリカの水準との違いがものすごく大きいんですね。

アメリカでは心理検査でもそうですが、教育検査、教育テストに関して、テストレビューというのが非常にクリティカルになされて、専門家の観点からの評価がなされている。だから、テスト会社は生き残るために、本当にドクターレベルの専門家を何人も雇って改訂などしているんですよ。その水準との違いがものすごく大きいので、今度のプロジェクトとの関連も考えなくてはいけなくても、米国での学力検査の開発と利用の実態というのも並べてみると非常にそれがわかってくるんじゃないかな、と思いますね。

内容的な妥当性という観点では、指導要領との対応付けは見ているけれども、内容的というためには専門家が見て少なくとも納得するものでなくてはいけなくても、専門家でない人がみても納得しないという部分が出てきた。そういうのは内容そのものについても項目の内容のクリティカルなレビューも十分でない。専門家というのはある意味仲間ですからなれ合いになってしまう。

あとは、資料7の2枚目の所に妥当性というのがあって、「②知能検査との相関。妥当性を見るためには知能検査との相関を見ることも重要です。知能と学力との相関は一般的にはこれくらいで、…低学年は低めですが、高学年になるにしたがって高めになります。」と書いてあるけど、だからどうなんだ、ということがわからないんですね。知能と相関が高かったら何なのか、高学年になると高くなっていくというのは何なのか。調べることの意味もよくわからないですよ。

本当に水準が問われるような検査が作られて売られている。実際に使われていなければいいですけど。アメリカでは学校で作ったテストも意味があるのかもしれないけど、やはり学校教育のアカウントビリティとしてどれくらいプログ्रेसしているのか、学年進行に応じてどれだけ変化し、他の地域と比べてどうなのかということがすぐに問題になるので標準化したテストでないと話にならないんですよ。学校で作ったテストではダメなので。そういう位置づけも全く違う。

社会的なプレッシャーだとかそういう違いもありますね。

7.3 アメリカの学力検査について

秋田：南風原先生に質問なんですけど、今、アメリカの学力検査の話が出ましたけど、たとえば観点別も領域別も教科ごとですよ。学習診断は教科フリーですか。アメリカの学力検査もやっぱり基本的には教科ディペンデントですか。たとえばOECDは直接教科ではなかったですよ、リテラシーですよ。そのあたりがアメリカの場合の学力検査はどんなものがあるのか私はよく知らないのです。

南風原：僕も詳しくは知らないんですけど。僕のいたところはアメリカでそういう学力検査を作るところの1つのメッカだったんですね。アイオワテストっていうところが。それがやったところでは、教科という区分けをはっきりしていませんでしたけど。例えばソーシャルスタディといったような非常に大きな枠組みを、あるいはアリスメティック・リーズニング、数量の扱いとか。4つか5つの大きなくくりですね。

秋田：将来どうなるのか、そういうのがいろいろと説明ができますよね。小学校中学校で教科が分化したり、いろいろなことが起こってくると、細かければ細かいほど果たして将来にどれくらい学力が意味があるのかっていうこととか。

恒吉：アイオワもそうですけどテキサスとか非常に大きなテストで業者が作った州テストもありますよね。今の観点別でないですけど、州によってかなり変わっていますよね。

南風原：アイオワのテストは標準学力テストで、いわゆる偏差値を見たり、あるいは…

苅谷：この標準というのはどっちの。ここでいうとCRTの方か。

南風原：NRTですね。基本的にノームがあるんで。それからノームの表し方の特徴としてグレード・イクイバレンスというのがあって、その子供がソーシャルスタディに関しては、小学校3年生レベルとか、小学校5年生レベルというような昔の知能指数のような考え方で、精神年齢のような考え方をしているんです。それから今出た州レベルのテストというのは、ひとつは州ごとに高等学校卒業資格としてのコンピタンスーテストというのがあって、ちょっと確認しなくてははいけ

ないけど、それは大抵州レベルでやっているんじゃないですかね。相対的にどうじゃなくて、カットポイントを設けて。あとカットポイントの設け方でも非常に測定論的な議論があるんですけども、「えいやー」というのも含めて、そこでパーセントで切っていくということですね。学生生徒にとって、学校以外で作られたテストというのが卒業資格として非常に重要な意味を持っている。これはコンピタンスー・テストの方ですね。学力テストの方は、むしろ学校や教師に重要な意味を持っているというところですかね。

恒吉：州テストは見たことがあったんですけど、いわゆる「学力」という意味でなくて記述式もかなりあったりとか、協同学習をグループでやって、それで点数を出したりとかいろいろありますね。ただ、それがうまく先程の観点別みたいに評価できないとかあって、もとに戻ったりとかいろいろあるみたいですね。

南風原：今言ったような、協同学習というのは初めて聞きましたけど、今度はそういう最前線での広がりを見たいというのがこのプロジェクトとも関係して面白いところかと。

7.4 到達度の判断

苅谷：たぶん、心理学と教科教育との境界線上で、僕が言っていた標準化を考えるとクリティカルになると思うんですよ。どんなに心理学的な統計的なスキルを使っても、最後のところでこの問題はどれくらいの人ができるかという判定自体はおそらく教科の専門家の判断が重要になると思うんですけど。それを考えるに、最初に村山さんが言ったことは大事だと思うんです。どの時点の集団を念頭に置くかで、あるタイムレンジで変化ということまで考えて、その中で、たとえば日本の子供だったら4年生はこれくらいのことが10年前でも今でもできるという想定をするのか、それとも現在の子供だけ考えて、その中で見れば、まあこのくらいまでできればいいだろうと見るのかで変わってくる。それはもちろん、改訂とも関連してくるけど、その判断自体はどうやったらできるのか。ここでいうと、妥当性の話のところでは修正エーベル法というので、マトリックスの中にいろいろな専門家がこの問題はこれだとかいって行くわけですよ。それで正答率が65%だとか15%だとか。あとは機械的に計算されるけど、これの決め方が、なんで正答率が60になるんだと

いうこと。

村山：いかにも客観的に決めているように書いてあって、読んでいて、「おや？」と思うんですけど。

苺谷：この間の全国の文科省の調査でも正答率の設定にたぶんこれと同じようなことがやられて、叩かれたんじゃないかと思う。

南風原：3種類くらい代表的なものがあるんですけど、全部そういう判断ですよ。たとえば5肢選択なら、ミナマムに達している子供だったら、5肢選択の選択肢のうち、どれとどれは間違いだと気がつくか、という感じに集計して。いずれにしてもミナマムレベルというものを想定するのが専門家だから。

苺谷：そういう情報がやっぱりあるわけですね、アメリカだったら。

南風原：主なものが3種類くらい。Ebelの方法の他に Nedelsky、Angoffという3つくらいかな。その辺は教育測定のテキストなんかに出ているけど。

苺谷：日本ではあんまり使われていないんですか。

南風原：日本では使われていないんじゃないですかね。

苺谷：何人くらい専門家がいて「えいやー」ってやっているんだろうか。複数とかいうけど2人でも複数。

南風原：何人いてもあんなやり方で曖昧なものを集めても。安定はするかもしれないけど、安定したら何なのかと。

村山：資料には書いていないのでそのときの学者の人たちが、どの時期を考えてラインを引いているかとか細かいことまではわからないです。それにお年寄りの人とか高めに考えがちのような気がします。

7.5 学力検査の結果と教育現場における行動との関係

苺谷：何らかの形で今回このプロジェクトでテストを開発するとして、その中にはここでいうと、判断力とか思考力みたいなものも測定できるならば、どうやって開発できるか。関心・意欲・態度も測るかどうかはともかく、何らかの形で、仮にテスト可能だとしますよね。そういうことをやってみたときに、こういう市販のものとは違うこととしてできそうなのは、実際の教育現場で、そのテストを受けた結果を見て子供たちを観察したり、あるいは授業の中で先生からそのフィードバックをもらおうとかして、この評定自体が教育の中でどういう意味を持っているのかということ調べ

ることとか。そういうことはたぶん他のいろいろな方法、多元的な方法を使うことでテストの持っている意味合いをもうちょっと実際の教授場で評価するということは可能じゃないかと思うんですよ。

そういうことはたぶんこのプロジェクトで大事になってくると思うんですけど、そういうことを考えたときに、そのときに関心とか思考力とか判断力みたいなものについて、どうしてもテストという方法だと、その受けている時点での能力を時間を止めて切り取ったように見える。関心などは一番わかりやすい例だけど、どれくらい持続する関心なのかというのは学習の中で本当は大事ですよ、そのときたまたまやっているある項目について面白いと思うんだけど、そのあとその教科については興味がなくなったら意欲とか関心とかには全然つながらないわけで。判断力みたいなものも、どれくらいそれを他に共用できるのか。さっきも、数学の例でいろいろとアルゴリズムを分解して行って、こういう認知過程があるというのはわかる。ただ、それが本当に次に他のところで生かせる判断力かどうかみたいなことは、たぶん、今言った他の場面の観察とか、そういうことによって評価が可能になってくるのでは。そういうテストの持っている一面性とか一時点性みたいなものを考慮して、どうやって他の情報によって付け加えていけるかというのが、この全体のプロジェクトでやっていけると面白い。

村山：それは非常にありますね。そういうことをよく考えて、研究をやっているんですが、ひとつ悲観的なこととして、そういうなんらかのモチベーションを授業観察を通して行動指標としてとったり、そういう認知機能と別な教科の授業行動との関係を見るという研究はおしなべて結果が出ない。たとえば授業行動というものを単位としてモチベーションというものを測定したときに、それがその他の指標と相関を持たないとか、そういう研究が結構ありまして、アイデアとしては僕も面白いと思うんですけど。

南風原：観察された授業そのものが切り取られたもの、上辺だけのものかもしれないというのもひとつある。

今、苺谷先生が言ったのは、テストというものを行動のようなもので、妥当性検証するというような考え方ですが、それだけでなく、こういう基準というのがあれば、これを組み込んだ形の評価、アセスメントがいいわけですよ。授業での行動にバリデーション

の対象となるような変数、そういう側面があるならば、むしろそれを組み込む形でアセスメントできないかということを考えて方がいい。少なくとも関心・意欲で、「予習してきますか」とか「理科は好きですか」というのは、意味ないですよ。あなたは何れくらい正直ですか」みたいなのも同じで、「はい」と答えたからといって何なのか、といったもので。たぶん、こういうのを作ってはいけないんだと思う。

村山：この組み込んだ形でのアセスメントをするときに、どういうふうに組み込むかというのもまたバリデーションしないといけないと思うんですけど、そこに無限の…。

南風原：そうだけれど、バリデーションというのは最後はやらなくていいと思うんだよね。そういうふうに組み込んで、みんなが納得して実際に使えてるという感覚が出てきたら、バリデーションのために何かをとっているというのは、誰かのため、証明のために置いておくだけで。むしろ自分たちが納得するような形にやっていって、そういう共感が得られて、使っていく中で確認されていけばいいと思うので。バリデーションの形式に持つていくために何かを残していくというのは、これはあまり本質的な部分ではないんじゃないかな。

秋田：私は、今苅谷先生が言われた発想って、アセスメントする側から見ると妥当なんだけど、授業展開する側からいうと全く逆だな、と思うんですよ。テストに何か関係ありそうな情報を、要するに授業とか、実際にやっているもののなかで、何か妥当なものを付け加えていけば、正しい学力が見れるか、適切に評定できるかということ、必ずしもそうではない。たぶん授業のプロセスそのものと、いわゆる一時的に学力として測れるものというのが、そんなに簡単に結びつくものかということ、ある授業というのはある固有の概念だったり、ある単元の特定の学習の積み重ねとしていくわけですよ。だからテストのある一部分は、いわゆる教科関連のテストとの関係は説明できるんだけど、授業観察を少し入れれば学力との関係を見ることができるとするのは希望的観測。現実的に授業研究をやっている人たちから見ると、それは実際には直接的には結びつかない。でも何らかの関係があるのも事実だから、何とかその関係を模索したいな、と思うという部類じゃないかな。逆にこのテスト情報がいかに活かされていけば授業が変わっていったりできるのかとい

う方向の議論ができるかなと思うんですよ。授業情報を評価に入れるというより、評価を授業に組み込むという方向の方が使えると思うんだけど。

苅谷：関心とか意欲とか、そっちも。

秋田：「ああ面白い」と思って、ある実験には関心があるけど、ある実験には関心がない子が、じゃあ理科について関心がないかといえばそんなことじゃないわけだから。さっき南風原先生がいったように、ある一場面で切り取られるんですよ。その観察情報がどれくらいある種の学力といわれるものを測定できるかというと、こういう指標とするのは非常に難しい。

苅谷：今現実に行われている絶対評価で動いている評価の仕方は、むしろ今いったような方向に変わりつつありますよね。生徒たちが授業ごとにどういう関わりをしているのかを得点か何かにして、最後とかに集計してみるというのはむしろそういう発想ですよ。やっぱり測定しない方がいいってこと？

秋田：そうではなくて、測定はなんらかの形でして、それが活かされた方がいいと思うし、ある種のもののどこが弱いかという、そういうのを生かすという方向はあり得ると思うんだけど、授業観察が学力テストのある種の保証になるかということと非常に難しい。ある単元で、どういう方略を使っているかとか、どんな概念の形成の段階にあるかということと、ある種の限られた部分での関係を見ることはできると思うんだけど。

村山：動機づけをやっている方から見ると、質問紙は切り取られたものとか、いろいろと批判されているんですけど、それなりに妥当だと思うんです。やはり、かなり長い年月をかけて概念が明確化されて、それなりに教育場面におけるいろいろな行動との関係が示されてきて、それに関する研究というのは本当に数え切れないくらいあるわけです。

僕が今回強調したかったのは、今回の学力検査ではそういうような話は全然生かされていなくて「動機」という概念がすごく矮小化されておかれてしまっているところ。質問紙で測定できないというのは悲観的なんですけど、むしろ逆にその辺は楽観的なものかもしれないんですが、もう少し教育心理学で積み重ねてきたものをきちんと出せばそれなりに妥当性のある関心・意欲・態度の測定というのはできるのではないかとも思っています

南風原：この現状を大きく改善することは容易にでき

ますよね。本当に学問をやっている人がちゃんと取り組めばね。

7.6 教育心理学者と学力検査の歴史的背景

市川：今回、こういうのを村山君にやってもらったきっかけというのは、COEの話が来る前から、認知カウンセリング研究会でずっとやっていたことで、われわれは個別に相談に来た子には教えるんだけど、こういうテストを使ってアセスメントみたいなことはほとんどやるということはないですよ。臨床心理だとよくやるじゃないですか。あとLDの子に対してもそういうアセスメントはいろいろやっているし、開発されているテストもある。普通の子が来た場合に、何らかのアセスメントとして何かいい方法があるんだろうか、とりあえず、市販でどんなものがあるか、というのを見てみよう。こういうのは何もクラスで一斉にやってクラスの平均値はどうだなんていうのを見ると、ということよりも、どちらかというと個別診断のときに、この子はどういうところが問題なのかを見ていこうという意図で始まったという面も多分にあつたんです。

とりあえずどんな方法があるのかを調べてみよう。調べてみたところでちょっとびっくりしたところがある。あ、あ、こんなものなのかと。大体30年、40年くらい前に作られてそれっきり、みたいなものもかなりありますよね。その後、認知心理学でのいろいろな発展もあるし、動機づけとか学習スタイルの研究の発展もあるのに、その辺は全然生かされていない。僕たちから見ると大御所のような先生がずっと昔に作ったものが、そのままちょっと改訂されているくらいで残っていると深く感じたんですね。

同時に、他の近い分野ではどうかというと、たとえば、これはつい最近送られてきたベネッセのものです。これは学力診断、学習スタイル、学習動機づけとかになっていて、私もちょっと学習動機の所は向こうからインタビューに来ていて、いろいろ関わっていたんですけど。こういうところだと、目的がはっきりしていて、要するにこのテストをやることによって点数がフィードバックされていきますよね。結局、学力の特徴はこうです、こんなところが弱いようです、ということがフィードバックされて、もっとこういう風にやるといいですよ、というアドバイスがついてくる。こういうのは僕たちがもともと考えていたことと合ってい

るんですよ。つまり、こういう市販の学力テストというのは、生徒に一体どういう風に働きかけたらいいのかということとはほとんど見えてこない。先生にすらそのアクトの部分がありません、先生はそれをもとにどういう指導をしたらいいかがほとんど見えてこない。ましてや生徒の方にこれをフィードバックしても、生徒の自己学習力の足しになるという感じがほとんどしない。こういう通信教育では確かにその辺では力を入れていて、問題としても、多分普段問題を作っている人たちだから一応そんな古めかしい問題ももちろんない。あと動機づけとか学習スタイルについてもこれよりは一応新しい知見に割と取り組もうとしている、という姿勢は見られるんですね。

ですからこういう心理テスト業者に加えてこういう分野がある。それから、ほかにあと検定試験とかそちらの分野でも、結構、心理学者が出てきているところが。特に最近、法科大学院ができて、あの適性試験は肥田野先生を委員長に駆り出して入試センターでやっています。柳井先生とか、私もちょっと関わっているんですけど。そういう適性検査がらみで学力診断のようなことがかなりやられている。そういう分野と比較しながら、我々独自の小中高生向けの学力とか学習方法のテストというものを考えていきたいということで。そのための第一歩という位置づけで今日のをやりました。

どうでしょうか。今後に向けて、とにかく現状がこんな感じなんで随分改善の余地はあるだろうと。

苅谷：やはり相当ダメですか。すぐに改善できるというのは、それくらい学問的レベルが低いままだということですか。

南風原：そうですね。妥当性の部分とかマニュアル自体も全然違いますよね。教師用のマニュアル、それからアドミニストレータのマニュアルとか。

苅谷：たまたま最近、うちの院生が書いた修論で、戦後日本での知能検査の歴史を調べた学生がいるんですけど、1960年代で、もう研究が止まってしまったと言っている。日本の教育心理学では、やはり教育界でさんざん知能検査を使うことに批判を受けて脅えてしまっていて、もうそれ以来、多少学力論では関係して出てきているけど、それ以外は新しい研究が出てこなくなるとレビューしている。たぶん学力テストもそれに近い歴史が過去にあったんでしょうね。それとも教育心理学者の関心自体がそういうところになくなってし

まったということなんですか。つまり、なぜ昔のまま止まってしまって何十年も前のことを。今、知能を別の例として挙げてしまったけど、学力テストでも同じことがいえるんじゃないかと。

南風原：そうですね。さまざまな心理特性の尺度作りが盛んなのに比べたら、学力というものや学力調査というものに対する関心はものすごく薄いですね。それに関わろうとしている人たちがあまり出ていない。むしろ出版社が中心となって、これを作ってみようかという感じで、一晩か二晩合宿に行って作って見ましたという感じ。

苺谷：そういうところに心理学者が名前だけを載せている。

南風原：同じような名前がしょっちゅう出てくる。昔の偉い人とか。

市川：だから学校でもかなり心理テストが叩かれたという面があって、叩かれたのは心理学者にも問題があったわけ。それを使うとどんないいことがあるんだということを示さないまま、いかにもそういうテストは科学的にやっているというように見せて。だから知能テストにも反論ができないまま反対だけ受けたし、性格テストもそうです。新聞でも、こんなものやって何になるんだって随分と叩かれて、反論はほとんどできなかった。学力テストについていえば、大体市販テストそのものが随分叩かれましたよね。ましてや心理屋が作った学力テストなんてやって何になるんだと。実際、やってもほとんどその先が見えない。ニーズも減るし、ニーズが減れば当然作る方もモチベーションがどんどん低くなりますよね。

南風原：そうですね。「関心・意欲・態度」みたいなものが文部省から降りてきて、そういうものを取り入れれば売れるというような出版社の考えがあって、何人かの人が大学から呼ばれて、それを作る。だからその程度のものしかできないですよ。学問から出てきたものじゃないから。

市川：「関心・意欲・態度」を出したときには、心理学者も当時の大御所は随分と裏で動いたんじゃないかという気はしますけどね。委員会などで、そういうものをこれからの教育のメインとして、しかもちゃんと評価もしてはと。

南風原：そういう人はいましたけど、強い下からの研究が盛んだということではなかった。

市川：そうですね。これから我々はどうするかなんて

すけどね。

7.7 到達度評価としての得点の持つ意味

苺谷：やはり、1つには、診断的なテストの開発は大事だと思うんだけど。もう1つは個人のアセスメントではなくて、学校とか制度のアセスメントとしての側面というのも、このプロジェクト全体の中では、基礎学力形成といった場合にある。どういう調査方法が可能かという問題はさておくんだけど、そういうことで考えると、やはり標準化っていうか、標準のノルムの部分をどうやって設定できるのかっていうこと。さっきの修正エーベル法とか、これに代わるそういうものをこのプロジェクトで、何らかの教科について我々にはできるんですか。何らかの絶対尺度、絶対的なあるカッティングポイントを作る余地ってあるんですか。これはいい加減にやっているのかもしれないけど、本当に専門家を入れて、アメリカのような研究をしてこの程度できれば、数学なら数学について、これくらいの問題はこれくらいできて欲しいっていう基準を入れるような。

南風原：「欲しい」という主観的な判断を入れた形で持っていくのか。

苺谷：「欲しい」じゃなくてもいいけど。

南風原：あるいは「このテストでこのレベルのものは何々ができる」というふうな別の証拠と付き合わせて関連づけることはできる。

苺谷：後者の方が大事ですよ。そういうことが、もしできるとすごくいいね。

南風原：到達度評価というのも、できる・できない、というふうにかットする手法だと思われているけど、ある別の見方は、何ができて何ができないのか明らかにする方法だということなんですよ。

苺谷：それは易しいか易しくないかじゃなくて、難しいことによって、他のこれができる。その辺の関連性がわかればいいわけですね。それは大事だと思うんですよ。

村山：それは現実的にまだまだ基礎の部分が…

南風原：とは思うんだけど。それをあんまり抽象的な言葉で言われるとダメなわけですよ。「文章を読んで理解することができる」というようなことを言っても文章なんていろいろある。だから、例えば、この程度の文章が理解できるとか、サンプルを示すことができる

わけですよ。何ができるかということの伝え方というのはいろいろあるんだけど、それが今は学習指導要領のようなお題目が並んでいるからそれができると言われても何ができるのかわからない。絶対評価で◎とかつけるというの、観点を見ても、何々に興味を持っている、といってもそれが具体的に何を意味しているのかわからない。実際の行動とか、読む課題、書ける問題とか、そういう形で整理していくことはできる。算数のこの問題ならできるけれども、方程式がこれくらい複雑になるとできないというレベルが何点かということ。意味がわかるような。

苅谷：そういうのは大事だよ。要するにこの何十年もそういう話は日本では全然やってこなかったってことか。教育の測定とかって、かなり基本的なことですよ。

市川：アカデミックな人たちはあまりそういうことはやってこなかった。

苅谷：というか、そういうのがなくてもそれで日本の教育の現実がそれで許してきたわけですよ。

市川：入試センターの心理系の人なんてどうなんですかね。それなりに点数の持つ意味、この点数くらいの人はこのことができるんだということをやってきたんじゃないですか。

苅谷：すごい膨大なデータを持っていますからね。

南風原：付き合わせるような研究はしていないんじゃないですかね。むしろ心理学者といっても心理統計の人たちが多くて、統計モデルのこととかはやっているんでしょうけど。あるいは、等化など技術的なことはしてるんでしょうけども。

苅谷：それでもあのデータを使えば、そういうことも開発可能なんですか。センター入試の過去の得点という問題との。

南風原：問題を比較していけば。

苅谷：こういうこともできる、こういうこともできる、というような。

南風原：相関的なデータをとってあげばいいわけですね。

市川：検定試験なんかもそうですけど、少なくとも物差しの意味ははっきりさせるといことは、もっとやらなくてはいけないと思うんですよ。それも、例えば作文であるとか、さっき恒吉先生もおっしゃったけど、イギリスはかなり作文をきちっと書かせて、それを客観的に評価するのをやっていますよね。何点はこ

れくらいのものだというサンプルがあって。それをもとにやっているから、非常に信頼性も高くなるし、何点をとる人はどの程度ものなのかということがわかる。ということをやっていかなければならないんですよ。

南風原：この程度の文章を書ける人が減ってきているというのだったら、学力低下というのがよくわかるわけ。

市川：作文力の低下というのでも、例えば20年前の作文と今の作文とをランダムにして、ブラインドにして、そういう基準で評定をさせてみれば、作文力が本当に落ちていることがいえることができるわけですよ。

今日のところはこれくらいでよろしいでしょうか。今回購入した学力検査については一覧表にしたものがありますので、必要に応じてご覧下さい。

村山くん、今日はどうもありがとうございました。

配布資料

市販の学力検査における現状と展望

教育心理学コース博士課程 1年 村山 航

本稿では、田研出版・図書文化（教研式）などが発行している、市販の「学力検査」に関して、その現状をまとめ、今後の展望を記すことを目的とする。まず最初に、学力検査を「領域別学力検査」「観点別学力検査」「学習状態診断検査」の3つに分け、それぞれがどのようなものであり、どのような問題点を持っているかについて考察を行う。最後に、これらの考察をもとに、学力検査における今後の展望に言及していきたい。

1. 領域別学力検査

- ・学習指導要領に準拠した領域を設定し、領域（単元）ごとに問題を構成（資料1）。
- ・評価形式は相対評価：問題は標準化されている（平均点は50点程度）。
- ・テストは学年ごとに作られている。年度末もしくは新学期に実施する想定。
- ・領域ごと（1教科3領域）に制限時間が設定されている。

《標準化の方法》：資料2（田研）を参照。

- ・内容的妥当性の確保：その過程などは明示されず
- ・正答率・弁別性などの観点から項目を精選（田研）

《結果のフィードバック》

- ・基本は返送してコンピュータ採点だが、学校で採点も可能。
- ・フィードバックされるもの（フィードバックの利用方法については資料3参照）。

- | | | |
|------|---|--|
| 教師のみ | { | 1) 学力偏差値（領域ごと・全体） |
| | | 2) 総合成就値：知能指数から期待される学力との差 ¹ |
| | | 3) 学級・学年集計表 |
| | | 4) 小問分析表：小問ごとにおけるクラスの正答率・誤答率 |

《考えられる問題点》

- ・問題製作過程や調査協力校など、標準化テストとして必要だと思われる資料が未公開。
- ・どの時期のデータを使って偏差値などが算出されているかが不明確：学力低下などが叫ばれている近年、少なくともどの年のデータをもとに標準化したのかは重要な情報であるし、頻繁に更新していく必要があると思われる。
- ・テスト制作者が明記されていないものがある。
- ・多くの問題が多肢選択形式、もしくは単答形式より構成：それによって測定するのがふさわしい種類の問題と、ふさわしくない種類の問題がある。結果主義に陥りやすい。

¹ 総合成就値=学力標準得点-{(知能指数と学力テストの相関係数)×(知能標準得点-50)+50}。±7あたりが、アンダーアチーバー・オーバーアチーバーの目安。市販の学力検査では、知能との関係を議論する傾向が強いようである。

cf. 英作文：整序問題であり，十分に理解していなくても，感覚で解答できる可能性。

算数・数学：計算過程が問われない。また証明問題がない。

- ・学習指導要領の教科内容による拘束が大きい：学習指導要領がミニマム・スタンダードとなっている現在における意味がやや希薄化。

2. 観点別学力検査

- ・新学習指導要領において取り入れられた「観点別評価」の各観点（関心・意欲・態度や思考，判断力など）を評価するために問題を構成（資料4）。
- ・評価形式は到達度（絶対）評価：平均は70～80点程度。
- ・関心・意欲・態度は問題で測定するもの（田研：資料5，資料4も併用）と質問紙によるもの（教研式）がある（資料4）。
- ・テストは学年ごとに作られている。年度末もしくは新学期に実施する想定。
- ・領域ごと（1教科4領域）に制限時間が設定されているもの（教研式）と，設定されていないもの（田研）に分かれている。また，領域別テストに比べて，時間の制約がルールであり，余裕を持って終わられる，もしくは時間延長が可能になっている。

《テスト作成の方法》：資料6（教研式）を参照

- ・各観点ごとに目標の具体化表を作成，そこから具体的な項目を作成する。
- ・作成された項目は複数のジャッジによってさらに精選される。
- ・到達基準に関しても一定の手続（修正エーベル法）に則り，できるだけ客観的に定める：「原則として見識ある人の良識的判断に基く事前設定」²。
- ・領域別テストとは異なり，信頼性の検討も報告されている（資料7：田研は折半法・教研式は再テスト法）。

《結果のフィードバック》

- ・基本は返送してコンピュータ採点だが，学校で採点も可能。
- ・フィードバックされるもの

- | | | |
|------|--------------------------------------|---------------------------------|
| | 1) 観点ごとの到達度：3段階（十分満足・おおむね満足・努力を要する）。 | |
| | 2) 教科ごとの標準得点：田研のみ | |
| 教師のみ | { | 3) 学級・学年集計表：学級内での相対的な評価 |
| | | 4) 到達度基準の全国的な通過率：参考資料として（教研式のみ） |
| | | 5) 小問分析表：小問ごとにおけるクラスの正答率・誤答率 |

《考えられる問題点》

- ・4つの観点を測定するために，かなり本格的に検討された問題が使われているように感じる（資料4）。しかし反面，マークシートで理解力や表現力を測定する限界が露呈。

² 以上は教研式に記してあったものだが，田研のほうは問題作成過程が不明。

- ・妥当性の問題：何ら基準連関的な妥当性が示されていないため、本当にこのような問題で「理解」や「思考」能力を測定できているのかが不明確。基準連関妥当性があまり示されていない反面、内容的妥当性が強調されすぎている。
- ・教研式と田研で、考え方が大きく違っている（関心・意欲・態度の測定法/標準得点を積極的に取り入れるか否か、など）：新学習指導要領における「評価」を、いかにすべきかに関して、未だ混乱状態が続いている部分がある。
- ・1つの観点に対して、すべての領域から万遍なく問題構成がされているわけではなく、観点と領域の交絡が生じている：どの領域の問題を用いるかに関しても、もう少し系統的に行う必要がある。

※その他の学力検査

- ・学習進度指導検査：3から5学年前にまで遡って問題を作成。どの学年の段階で子どもがつかずいたかがわかる。ただし、語彙と計算に限る。
- ・読み能力検査：読みの能力を、「語の認知」「文の理解」「音節の抽出」など、いくつかのコンポーネントにわけ、それぞれを測定する。標準化されており、読書力偏差値や読書年齢などを算出可能。ただし、10から20年以上前に作成され、改訂されていない。よってその妥当性には疑問が生じる（資料8）。

3. 学習状態の診断

- ・学習意欲・身体健康・学習環境・学習方法・自己学習力などを測定する質問紙。
- ・測定している構成概念は、質問紙ごとによって異なる（資料9）。どの質問紙も10因子以上・100項目以上のもの。

《質問紙作成の方法》

- ・項目の収集（専門家による作成が多い）
- ・因子分析・信頼性・妥当性の検証などを行う（資料10）。
- ・因子ごとの平均値や標準偏差、学年ごとの変容なども資料として報告。

《考えられる問題点》

- ・教育心理学研究で明らかになった動機・学習方略の概念が十分に活用されていない
e.g. 学習方略：深い処理の方略のみ
動機づけ：漠然としており、特に取り上げられているのは原因帰属のみ。
- ・上記の理由のためか、教研式SETでは、学力などとの相関も低い（資料11）。
- ・親子関係など、回答のしにくい項目がある。
- ・妥当性の検証が行われてはいるが、全般的に弱い感がある。

4. 今後の展望

(1) 「学力」の意味の多元化に対応するテスト作成

- 今回の学力検査をから、3つの次元を抽出可能.
 - 内容領域別の学力：「2次関数は大丈夫だけど図形はダメ」
 - 観点別の学力：「知識はあるけど思考力がない」「知識はないけど意欲はある」
 - 基準によって変化する学力：「要求した水準に到達していないけど偏差値は高い」
- どの次元も学力を捉える上で重要であり、不必要な次元はない：「観点別」「到達度」という考え方が入ってきたからといって、「内容領域別の学力」「相対評価」が不用になったわけではない ⇒次元がシフトしたのではなく、多元的になっている
- この次元を交絡させることなくテストは作成されることが望ましい：
cf. 観点別学力検査：内容領域別（単元）の学力が交絡。テストが測定しているのが、何の次元における違いなのかが明確ではない。
- 上記の次元を意識した上で、テストのバリエーションを増やす必要性：特に観点別学力検査では、1つのテストで多くのことを欲張って測定しようとし過ぎている。
- ただし、実際に学校で実施する場合には、時間の問題など現実的な制約も多く、それらとの兼ね合いを考える必要性。
- 現場の先生に分かりやすいようなラベルづけ・構造化の必要性。

(2) 「新しい学力観」を適切に反映した妥当性のあるテストの作成

- 関心・意欲・態度はどう測定するか？：あえて測定する必要性も低い（例え質問紙でも）。
- 思考・考え方・表現力はどう測定するか
 - i) マークシートをやめ、記述式などで測定：評定の客観性・採点に伴うコストの問題
 - ii) あくまでマークシート方式で測定するようにする：現在はこの路線。
 - ⇒それなりに考えているようだが、まだ不十分な感は否めない。
 - ⇒教育心理学の知見（例：誤概念・バグ研究（資料12））などを生かした問題作成

(3) テスト評価後のフォローを考えることの重要性

- 近年の評価観「Plan-Do-See」から「Plan-Do-Check-Act」へ：評価後のActの重要性
- 「〇〇の部分がよくなかった」ということが分かったときに、具体的にどうすればよいのかということ、何らかの形で検査セットの中に取り込む。
- 特に「学習状態検査」の場合、そのような配慮が必要：現状でも副教材として、販売されている模様。

參考資料

学力に関する心理検査一覧表（会社別）

作成責任者：村山 航

（購入時：2002年11月：合計28項目）

《田研》

検査名	セット	実施時間	内容
TK 式領域別標準学力検査 小学校標準学力検査（移行措置版）	小1～小2（国・算），小3～小6（国・社・算・理）	40分	全国の小学生を用いて標準化。各教科に関し、領域別に得点が算出される。学習指導要領準拠。診断は田研。手引書あり
TK 式領域別学力検査 中学校標準学力検査（移行措置版）	中1～中3（5教科） 新入生（数・国・理・社） 中1・中2 社会科（並列型）	45分	全国の中学生を用いて標準化。学習指導要領準拠。各教科の領域別の得点が算出される（e.g. 数学：「数と式」「図形」「数量関係」）診断は田研。手引書あり。
TK 式 DRT 観点別到達度学力検査（移行措置版）	小1～小2（国・算），小3～小6（国・社・算・理）	40分	観点別（思考判断・関心意欲態度・表現・知識理解技能）で標準得点を算出。換算表有り。コンピュータ診断有り。手引書有り。
TK 式中 DRT 観点別到達度学力検査（移行措置版）	中1～中2（5教科）	45分	観点別（思考判断・関心意欲態度・表現・知識理解技能：関心意欲態度も普通の問題で測定）で標準得点を算出。換算表有り（と手引書には書いてあるが、見当たらず）。コンピュータ診断有り。手引書有り。
TK 式 20分中入 中学新入生学力検査	新中学生：数・国・理・社	各教科 20分	中学新入生用の学力検査。各教科 20分。標準得点換算表あり。
TK 式 40分中入 中学新入生学力検査	新中学生：数・国・理・社	4教科で 40分。	中学新入生用の学力検査。
TK 式 LSI 学習状態診断検査	小1 and 2・小3 and 4・小5 and 6・中学生の4部	45分 あれば十分	学習意欲・身体健康・友人や家庭、学校環境、親子関係などを問う質問紙。因子得点の平均値とSDあり。手引書あり。
TK 式 学習進度指導検査	小3～中1（それぞれ語彙編・計算編）	45分	3から5学年前まで遡って問題作成。個々の学習進度を測るのが目的。診断は田研。
TK 式 読み能力診断検査	小3～中1で1部。	約10分	読みの能力を識別・理解・記憶・推理の4つにわけ測定。偏差値換算表・利用の手引き有り。

《図書文化社(教研式)》

教研式全国標準診断的学力検査 小学 T 形式 移行措置 2 年次版	小 1 ～ 小 2 (国・算), 小 3 ～ 小 6 (国・社・算・理)	40 分	全国の小学生を用いて標準化. 学習指導要領準拠. 各教科の領域別の得点が算出される. 標準得点換算表・5 段階得点換算表有り. 学年ごとに手引書有り.
教研式全国標準診断的学力検査 中学新 1 年 T 形式	新中 1 (数・国・理・社)	45 分	全国の新中学生を用いて標準化. 学習指導要領準拠. 各教科の領域別の得点が算出される. 標準得点換算表・5 段階得点換算表有り. 手引書有り.
教研式全国標準診断的学力検査 中学 T 形式 移行措置 2 年次版	中 1 ～ 中 3 (英・数・国・理・社)	45 分	全国の中学生を用いて標準化. 学習指導要領準拠. 各教科の領域別の得点が算出. 標準得点換算・5 段階得点換算表有り. 学年ごとに手引書あり.
教研式中学新入生学力検査 中入 LII 形式	新中 1 (国・数・理・社)	20 分	手引書が見つからない
教研式高校新入生学力検査 高入 AII 形式	新高 1 (英・数・国・理・社)	45 分	手引書品切れ
教研式新観点別到達度学力検査 新 CRT-III A 移行措置 1・2 年次版	小 1 ～ 小 2 (国・算), 小 3 ～ 小 6 (国・社・算・理)	40 分	観点別 (思考判断・関心意欲態度・表現・知識理解技能: 関心意欲態度は質問紙) で到達度 (標準化得点ではなく, 達成のパーセンテージ) を測定. 学年ごとに手引書あり. 教科書別に 2 部あり.
教研式新観点別到達度学力検査 新 CRT-II 移行措置 1・2 年次版	中 1 ～ 中 3 (英・数・国・理・社)	45 分	観点別 (思考判断・関心意欲態度・表現・知識理解技能: 関心意欲態度は質問紙) で到達度 (標準化得点ではなく, 達成のパーセンテージ) を測定. 学年ごとに手引書あり. 中学 1, 2 年生用は教科書別.
教研式新版学習適応性検査(New AAI)	小 1～2 用 1 部, 小 3～6 用 1 部, 中 1～3 用 1 部, 高 1～3 用 1 部.	長くて 40 分	学習方略に関する検査. 採点はすべて教研が行う. 下位尺度は勉強の態度・勉強の受けかた・勉強の技術・家庭環境・学校環境・自主性・根気強さ・心身の健康・原因帰属・学習スタイルなど. 標準済み. 辰野千壽著. 手引書あり
教研式サポート 学習支援システム	中 1 ～ 中 3	約 1 時間	ギルフォードの MI に基づく一種の知能検査. 標準化済み. 図形的情報・意味的情報・認知・拡散思考・集中思考・評価. マークシートに記入. 標準化得点換算表あり. 利用手引書品切れ (実施手引きはあり).
教研式サポート 学習適性確認システム	高 1	約 1 時間	ギルフォードの MI に基づく一種の知能検査. 標準化済み. 図形的情報・意味的情報・認知・拡散思考・集中思考・評価. マークシートに記入. 採点は教研が行う. 手引書有り.

教研式全国標準読書力診断検査 A 形式	小学校低・中・高学年各 1 部, 中学校 1 部	小学校用 27 分, 中学校用約 50 分.	読字力・語彙力・文法力・読解や鑑賞力の領域別診断. 読書力偏差値算出可能. 換算表有り. 手引書有り (小学校 1 部, 中学校 1 部). 中学校用はマークシート方式.
教研式 SET 自己教育力指導検査	小学 4~6 年で 1 部・中学校で 1 部	多くて 40 分程度	学習 (課題意識・主体的思考・学習方法・自己評価)・性格 (集中力・計画性・自主性)・生活 (生活体験・自立生活・自己実現) の領域に関する質問紙. 手引書有り (小学校 1 部・中学校 1 部). 換算表なし. 採点は教研が行う.

《金子書房》

幼児・児童読書力テスト (金子書房)	4~8 歳で 1 部	約 30 分	語の理解・図形の弁別・音節の分解・音節の抽出・文字の認知・文や文章の理解に関する領域別診断. 総合的な読書力偏差値・読書力の段階などを算出可能. 換算表有り. 手引書有り.
標準読書力診断テスト B I 型・BII 型 (金子書房)	小 1~3 で 1 部	40 分	語の認知・文の理解・節の理解・漢字の読みの領域別診断. 読書力偏差値・読書年齢・読書指数なども算出可能. 換算表有り. 手引書有り. B I 型と B II 型は問題は違うが, 全く同質・同形のテストで, プレ・ポストデザインなどに使用可能.
標準読書力診断テスト C 型 (金子書房)	小 4~小 6 で 1 部	約 40 分	速読・読解・読字・単語の 4 領域による診断. 読書力偏差値・読書年齢・読書指数なども算出可能. 換算表有り. 手引き書あり.
標準読書力診断テスト D 型 (金子書房)	中 1~中 3 で 1 部	約 40 分	速読・読解・読字・単語の 4 領域による診断. 読書力偏差値・読書年齢・読書指数なども算出可能. 換算表有り. 手引き書あり.
自主性診断検査解説 DTI (金子書房)	小 5~中 3 で 1 部		自発性・主体性・独立性・自己主張・判断力・独創性・自律性・自己統制・責任制・役割認知を質問紙で測定. パーセンタイル換算表有り. 採点盤つき.

《その他》

LS 学習スタイル発見調査 (大阪心理出版)	中学校~高校で 1 部	40 分程度.	思考-動作タイプ, 実行-意識レベルの 2 軸で学習スタイルを表現. 大阪心理出版が採点.
FAT 学力向上要因診断検査 (日本文化科学社)	中学~高校		学習環境や学習習慣に関する質問紙. 手引書なし