

読みのレベルを診断する
ベンチマーク・アセスメント・システム

An Analysis of a Benchmark
Assessment System for Reading Diagnosis

足 立 幸 子
Sachiko ADACHI

1. 問題の所在

読むことは、国語科教育の根幹でありながら、読むことを教える前提として、子どもたちの読む能力をどのように評価すればよいのか、十分に研究されてきたとは言い難い状況にある。教員は、経験的に自分の学級に様々なレベルの子どもたちを混在しているのを知っているが、その違いや個々の子どもたちの読みの実態を明らかにせず、何となく学級全体に対して読むことを教えてきているのではないだろうか。特に、我が国では、読むことについては読解と読書という分け方が行われており、ある教材に対してその教材がどれくらい読めているかという読解テストは開発されてきたが、個々の子どもたちが日々の読書生活の中で、どのような読書能力を養ってきているかを診断する¹評価ツールはほとんどと言ってない。本来読む力は多面的であり、その評価の方法も様々な開発されるべきである。

このような問題意識のもと、これまで筆者は比較研究手法を用いて、英語圏やスペイン語圏の様々な読むことの評価法を研究してきた。表1は、評価の実施規模による分類である。実施規模の大きいもの、すなわち、国などの大きな規模で実施される学力テストをマクロ・レベル、学校や教室といった小さな規模で実施できるテストをマイクロ・レベルと呼ぶ。

マクロ・レベルの評価は、「達成」を測定するために用いられ、マイクロ・レベルの評価は、「診断」のために用いられる傾向がある(表2)。そして、上記のような問題意識からは、重要なのは、読みのレベルを「診断」するマイクロ・レベルの評価である。筆者はこれまでも、Accelerated Reader, PPVT (Peabody Picture Vocabulary Test), QRI (Qualitative Reading Inventory) などについて研究を行ってきたが(足立, 2005, 2007), 本稿では、読みのレベルを診断するベンチマーク・アセスメント・システムに焦点をあてる。

ベンチマーク・アセスメントというのは、予め設定されている指標(benchmark)から見て、子どもたちの能力がどの程度であるかを測定していく評価である。読むことの能力のベンチマーク・アセスメントでは、一般的に、指標を実現した複数の本を用いて、その本をある子どもがどのように読むかを、決められた複数の評価方法によって判断していくというものである。本稿では、アメリカの小学校で用いられているFountas & Pinnell Benchmark Assessment Systemを対象として、その評価ツールの実際を検討する²。

表1 評価の実施規模による分類

規 模		評 価 例 (テスト・評価ツール名)	サンプルと評価を実施する目的
マクロ レベル	国際	<ul style="list-style-type: none"> • PISA (Programme for International Student Assessment) • PIRLS (Progress International Reading Literacy Study) 	<p>ある特定の学年・年齢の子どもたちをサンプリングする。PISA は15歳, PIRLS は4年生である。参加国及び地域は増加する傾向にある。</p> <p>目的は, 各国の読書力を国際比較することである。</p>
	国	<ul style="list-style-type: none"> • NAEP (National Assessment of Educational Progress) 	<p>米国全州からサンプリング, 4年生・8年生・12年生。</p> <p>目的は, 連邦政府が, 各州における教育の達成度を見ることである。</p>
	州	<ul style="list-style-type: none"> • ISAT (Illinois Standards Achievement Test) • ITBS (Iowa Tests of Basic Skills) • ISEL (The Illinois Snapshot of Early Literacy-English) 	<p>その州で教育を受けている子ども全員が受験する。学年は州によって異なる。例えば, イリノイ州の ISAT は3年, 5年, 8年。多くは, High-stakes Test として用いられる。</p> <p>High-stakes Test とは, その子どもが当該学年の内容を習得し, 次の学年に進めるかどうかを確認するテストである。これに合格できない場合は, 夏休みに補習を受けたり, 落第したりする。</p>
ミクロ レベル	校区 学校 学級	<ul style="list-style-type: none"> • Accelerated Reader • PPVT (Peabody Picture Vocabulary Test) • QRI (Qualitative Reading Inventory) • Fountas & Pinnell Benchmark Assessment System 	<p>校区・学校・学級の全員が受験する。</p> <p>子どもがどのような読書力を持っているかを評価するために行う。結果は, 子ども自身や親にも知らされる。指導に生かすことを前提とし, 診断(表2参照)テストが中心となる。</p>

表2 評価する性質による分類

	性 質 の 説 明
達成 (achievement)	<p>目標に子どもが達しているかどうかを見る。達した/達していないの2段階だけではなく, 達し方の程度を数段階に分ける場合もある。多くのサンプリングに基づいて, 妥当性が図られている。</p>
診断 (diagnostic)	<p>その子の状態を見る。できているかどうかよりも, どのようにできていないか(どこに問題があるか)を発見するために行われることが多い。ミクロレベルテストには, このタイプのものが多い。</p>

2. ベンチマーク・アセスメント・システムの 内容 -Fountas & Pinnell Benchmark Assessment System を事例として-

2. 1 出典と位置付け

以前は小学校教師だった Irene C. Fountas と Gary Su Pinnell が開発したベンチマーク・アセスメント・システムは、数々の教員研修 (professional development session) を開発しているアメリカの教育書の出版社 Heinemann から出されている。教師が自分の子どもたちを診断していくための教育ツールとしての位置づけで、評価ガイドだけでなく、あるレベルと判断された子どもたちに、どのように読むことを教えていけばよいのかという指導ガイドがついている。

2. 2 評価に必要な物

このシステムには以下の物が含まれている。

- ① 指標となる本 (Benchmark Books)
- ② 音読の状況を記録している書式 (Recording Forms)
- ③ 子どもが本を読んだ後に書く作文の教材 (Student Writing Materials)
- ④ 計算機とストップウォッチ (F&P Calculator/Stopwatch)
- ⑤ 評価要約書式 (Assessment Summary Form)

ここでは、①指標となる本について、詳しく説明する。本は、この評価のために開発され・書き下ろし

表3 指標となる本とレベルと想定される学年の関係

本のレベル	想定される学年
A~C	K(※)
B~I	1年生
H~M	2年生
L~P	3年生
O~T	4年生
S~W	5年生
V~Y	6年生
X~Z	7・8年生以上

※Kとは Kindergarten の略だが、日本の幼稚園とは違い、通常小学校の中の最初の学年として設定されている。

となっているものである。表3のように、レベルが設定されている。

2. 3 評価の手順

次の13の手順で、教師が評価を進めていく。したがって、それぞれの手順を行っていく動作主は評価者である教師である。

- ① 記録書式に子どもの情報を記録する。
- ② 本の題名と導入を子どもに対して読みあげる。
- ③ 計算機上のタイマーを作動させる (始めの時刻を記す)。
- ④ 子どもが音読 (reading orally) を始めるようにさせる。
- ⑤ 記録書式に読む行動を記号を用いて記録する (code)。
- ⑥ 計算機上のタイマーを止め (終わりの時刻を記し)、その時刻を記録書式に記録する。
- ⑦ 正しく読めた単語 (running words(RW)) の数、間違いの数 (errors(#Errors)), 自分で修正した数 (self-corrections(#SC)) を計算機上に入れる。
- ⑧ 読みの流暢さ (fluency) についての簡単なノートをつける。加えて/または、流暢さの割合に○をつける。
- ⑨ 読んだ本について子どもと会話をする。子どもが言ったことについて、チェックをしていく。理解の度合いを測るために必要な質問を行う。会話の後ですぐに、それぞれの領域の得点をつけ、追加得点を加えるかどうかを決定する。
- ⑩ 書くための質問を子どもに対して読みあげる (もしも必要ならば)。子どもに書かせ始める (最大約20分間)。
- ⑪ 得点を獲得・記録するために計算機のボタンを押す (かかった時間 (WPM), 正しかった単語数のパーセンテージ (Accur.%), 自分で修正したもの (SC) を記録する)。
- ⑫ その子どもが独りで読めるレベル (independent level) か、指導されれば読めるレベル (instructional level) か、読むのには難しいテキストのレベル (hard text level) かを決定するまで、この過程を繰り返す。
- ⑬ 評価要約書式に結果を記録する。

これらの手順は、複数種類のものを含んでいる。その種類に応じて、解説する。



図1 指標となる本（3～8年向け）と計算機

2.3.1 ランニング・レコード (running record)

手順の①～⑦は running record という、英語圏で非常によく使用される評価法である。子どもにテキストの音読をさせ、いくつかの語を正しく音読することができたか、いくつかの語を間違ってしまったのかを計算するというものである。この評価法は、綴りと発音が一致しにくい英語において、これを一致させていく初歩の読みの教育 (phonics) を反映している。しかし、間違ってしまった語はどのように間違っただのかを丁寧に見ていくことで、子どもの読みの能力や習性を見ていくというものである。F&P Benchmark Assessment では、意味 (M=meaning) 上の間違い、構造 (S=structure) 上の間違い、見た目 (V=visual) の間違いがある。これらのキューを用いて、自分で修正するということもある。例えば、子猫についての物語を子どもが読んでいるところで、cat と読むべきところを kitten と読み間違えたとする。そうすると、意味的な間違いではなく、その部分に差し掛かるまで子どもが内容を理解していたことが分かる。また、kitten の冒頭の発音は cat の冒頭の発音と同じ [k] であり、その発音を使いながら音読したということが出来る。あるいは、hat と読んだとする。そうすると、その子どもは、-at の [t] という発音を理解して音読することができているが、内容的な理解³はできていないということになる。さらには、hat と発音してしまってから、すぐに cat と自分で（言われる前に）読みを訂正したとする。その場合は、見目で hat と言ってしまったが、意味上のキュー

を用いて修正ができたということになる。このようにして、機械的に phonics を学んで音読（文字体系に現れたもののコードを解体して発音するという意味で、decode という）できるだけではなく、内容理解が伴っているかを知る重要な手掛かりとなるのである。

2.3.2 読みの流暢さ (fluency)

⑧の読みの流暢さ (fluency) とは、どのくらいの速度で音読ができたかということを示す。我が国の国語科教育で使用される言葉としては、「すらすら音読」に近いであろう。流暢さも英語圏で話題になる読みの評価法である。ある程度の速さで decode して音読ができないと、内容理解が伴わないという研究成果に基づいて、このような評価法が重視されている。

2.3.3 音読後の会話 (conversation)

⑨の会話は、音読しながらどのくらい内容が理解できたかを、教師の口頭の質問に対して口頭で解答させることによって、測定するものである。教師が 0, 1, 2, 3 の4段階によって得点をつける。理解には、テキストの中 (Within the Text)、テキストを越えて (Beyond the Text)、テキストについて (About the Text) の3種類がある。例えば、ホッキョクグマについてのノンフィクションの本 (Level Q: Not Too Cold for a Polar Bear) であれば、次のような質問が理解を測定するために行われる。

テキストの中：

「ホッキョクグマについて何を学んだか言いなさい。」

「他に何を学びましたか。」

テキストを越えて：

「なぜホッキョクグマにとってこの場所に生きることが重要なのですか。」

「ホッキョクグマについてまだあなたが持っている疑問は何ですか。」

テキストについて：

「この本の冒頭を見なさい。筆者は、この話題であなたの興味をひくために、何をしましたか。」

「この本の最後のところを見なさい。この終わり方はよいと思いますか。それはなぜですか。」

3点満点の3種類で9点満点だが、その他に追加得点を1点加えることができ、合計得点の最高点は10点ということになる。

2.3.4 書くこと

読んだ本をもとにした質問に基づいて、書くことを行う。書く量は最大レターサイズ（A4判にほぼ同じ）1枚で、20分以内である。先ほどのホッキョクグマであれば、次のような質問である。

「ホッキョクグマについて、最も興味深かったことは何ですか。また、それはなぜですか。スケッチなどを描いてもかまいません。」

いわゆる読書感想文のような長文で練り上げた文章ではなく、短時間で書かせた文章によって、読みのレベルを診断するのである。

3. 考 察

3.1 我が国への示唆

以上、Fountas & Pinnell Benchmark Assessment System を概観した。そこで、このシステムからどのようなことが考察できるか、我が国への示唆という視点から考察してみたい。やや羅列的であるが、次の6点が考えられる。

1点目として、このような評価法の発想である。まず、このような評価システムが存在する背景として、アメリカやオーストラリアのような英語圏の国では、個々の読みのレベルに応じて、読みの補習教育が行われているという事情がある⁴。このようなシステムティックな評価は、我が国ではほとんど行われていない。その理由として、各学年で1種類の授業が行われ、子どもたちが持っている読みのレベルというものに関心が払われていないという問題がある。言わば、我が国の教育は「身の丈に合った服を着る」のではなく、「身の丈に合わない服に無理に体を合わせようとしている」ようなものであると言える。身の丈を知るために、このような評価法があるという発想を知るだけでも、まずは有意義であると言える。

2点目に、指標となる本も、我が国にとって意味深い。単語数、一文中の単語数、語彙の難しさなどで、指標となる本が設定できるという発想が重要である。我が国の国語科教育の教材には、そのような発想がない。教材をどのように教えるかということの前提に、この子どもはこのテキストを読むのに相応しいレベルであるか診断することが、本質的であると言える。

3点目に、評価の内容について、音読の見直しが挙げられる。このシステムでは、running record、fluency という音読を用いたものを、一般の教師で

も使える形にしているところがよい。我が国では、小学校1～2年生ではすらすら音読が重視されるが、3年生以降は音読自体をあまり重視しなくなってきた。何度も読解した後、それをどのように感情をこめて表現するかといった「朗読」の方に傾倒していってしまう。しかし、初見の音読の意味をとらえ直す必要があるのではないだろうか。確かに、日本語のように、ひらがなの綴りと発音が一致しやすい言語のもとでは、音読に対する関心は払われにくい。あるいは、音読をすると、音読をする方に気が行って、内容が頭に入ってこなくなるという意見も聞く。しかし、経験的には、初見の音読と内容の理解はある程度の整合性があると考えられる。この点について、今後さらに研究が必要である。

4点目として、やはり評価の内容についての会話という手法も魅力的である。1対1の会話の時間をとることは、通常の国語科教育ではできないが、いわゆる「読書の時間」に全員が個人で読書をしている時に、1人ずつ呼んで1対1で本を読んで会話をするということが可能である。会話という手法は、個々の子どもの読みのレベルの実態を見ようという時に、手軽で有効である。他に、短い時間であれば、朝の読書の時間なども使用可能である。

5点目として、会話で用いる質問が、「テキストの中」「テキストを越えて」「テキストについて」の3種類であることも興味深い。我が国の従来の読解テストでは、設問は「テキストの中」に集中していた。「テキストを越えて」や「テキストについて」も読みのレベルを診断する時に必要な項目となっていることは、今後読む能力を考えていく際の参考になるであろう。

6点目として、読んだことをもとに短時間で書くという評価法が挙げられる。しかし、書かれたものをどのように評価すればよいかは、さらに検討の余地がある。

3.2 今後の課題

以上、Fountas & Pinnell Benchmark Assessment System をとらえて、我が国への示唆を考察してきた。しかし、我が国の国語科教育における「読むこと」の指導は、影響を受けているとは言え、必ずしも英語圏の English や Language Arts や Reading という科目と同じではない。したがって、このような比較研究を行う際に、今後の課題として4点挙げられる。

1点目は、開発研究上の課題である。もしも、こ

のようなシステムを我が国で開発するとしたら、指標となる本の開発上の困難がある。我が国には、テキストの難易度を測定する readability という考え方がない。漢字の使用によっては、同じテキストであっても、難易度が変わってしまう。このことをふまえて、指標というものを設定することは難しい。しかし、指標を設定するという自体は、読むことを指導していく前提として重要なことであると考える。

2点目は、生活年齢(学年)の一斉指導の問題である。学年で教えられるべき教育目標が決まっている、それに固定された教材があると、実際問題として個々の子どもの実態を見るという発想にはならないであろう。子どものレベルを見、それから教材を考えるという順番にすることが重要である。このことから、「独りで読めるレベル」「指導されれば読めるレベル」「読むのには難しいテキストのレベル」という視点から、教材を見直してみる必要があるであろう。

3点目は、評価をする教師の育成である。筆者は英語圏の国々の小学校教師が、短時間の説明で running record を行う様子を見てきたが、我が国に適用するには、教師への説明と訓練が不可欠である。

4点目は、指導ガイドの分析である。F&P Benchmark Assessment System は、指導に結びつけるための診断的評価の性質を持っているのであるが、本稿では、指導ガイドについての分析を扱うことができなかった。診断されたレベルと、推奨される指導法との間にどのような関係があるのか、さらに追究していきたい。

謝 辞

本研究は、平成19年度～22年度文部科学省科学研究費補助金若手研究(B)「知識基盤社会における読書力を評価するマイクロ・レベル・テスト及び質的分析手法の開発」の研究助成を受けている。

文 献

- 足立幸子(2005) アメリカの読書力評価 全国大学国語教育学会発表要旨集 108, 35-38.
- 足立幸子(2007) 読書力を評価するマイクロ・レベル・テスト 日本読書学会第51回研究大会発表資料集15-24.
- Clay, M. M. 1993) Reading Recovery: A Guidebook for Teachers in Training. Portsmouth, NH: Heinemann
- Fountas, I.C. & Pinnell, G.S. (2008) Fountas & Pinnell Benchmark Assessment System 2 Assessment Guide. Portsmouth, NH: Heinemann.
- Smith, A., Giles, J., & Randell, B. (2000) PM Benchmark Level 1. South Melbourne, Victoria: Nelson Thomson Learning

注

- ¹ 我が国においては、診断的評価という、ブルームの提唱した評価の時期による分類(診断的評価、形成的評価、総括的評価)がよく知られており、事前評価と考えることが多い。しかし、新しい教育課程に入る前であっても、何かの教育行為の途中であっても、子どもの状態を診断するために診断的評価(diagnostic assessment)が行われている。本稿でも、そのような意味で「診断する」という言葉を用いている。
- ² 例えば、Fountas & Pinnell Benchmark Assessment System の他に、オーストラリアの小学校で用いられている Nelson Thomson Learning の PM Benchmark などがある。
- ³ 内容理解(understand)の他に、意味が分かるあるいは意味がとれるという Make sense あるいは meaning という言葉を用いて表される場合もある。
- ⁴ この方向の Clay, M はニュージーランドの研究者であるが、彼女の提唱した小学校1年生に対する徹底した個人指導 Reading Recovery は、アメリカ・オーストラリアでも行われている。