

オランダにおける自己評価ツールZEBOの意義と限界

— 学校の自己評価を充実させる試み —

奥村 好美

はじめに

本稿は、オランダにおける初等学校の自己評価ツール ZEBO（初等教育の自己評価）（ZEBO: Zelfevaluatie Basisonderwijs；以下、ZEBO）を検討することを目的とする。

オランダでは、「教育の自由」が憲法で認められている。それは、誰でも学校を自由に設立することができる学校設立の自由、学校が宗教的、哲学的そして教育学的理念を保持できる教育理念の自由、教育内容（leerstof）や教育方法の自由を意味する教育を組織する自由である¹。こうした「教育の自由」のもと、学校の多様性の保証に対するリスク管理のために、教育監査と呼ばれる学校評価が教育監査局（Inspectie van het Onderwijs）によって行われている²。

オランダで、定期的に全ての学校における教育の質が監査されることとなったのは、1990年代後半以降のことである。その後、2002年に整備された教育監督法（Wet op het Onderwijstoezicht；以下、WOT）では、学校の自己評価に多くの関心が払われていた。しかしながら、特に2000年代後半以降、オランダの学校評価は学力テスト重視の方向へ進んでいったことが指摘されている³。2012年に改正されたWOTでは、プロセスに関わる学校の自己評価は以前ほど重視されなくなった。その原因の1つとしては、学校の自己評価が2002年WOTで期待されていた役割を果たせなかったことがあげられよう⁴。

こうした流れの中で、1990年代後半、すなわち2002年WOTに向けてオランダの教育監査が整備されゆく頃、学校の自己評価に対して何の対策もなされなかった訳ではない。既存の学校の自己評価ツールの多くは妥当性や信頼性が確かめられていないという調査結果を受けて、より良い自己評価ツールを開発するプロジェクトがオランダで立ち上がっていた。それがZEBOプロジェクトである。ZEBOプロジェクトは、学校効果研究を基盤に、妥当性・信頼性を有する包括的な自己評価ツールを開発することを目的として、1995年から約5年にわたって行われたプロジェクトであった。それには、SVO（教育調査機関）（Stichting voor Onderzoek van het Onderwijs）⁵が資金を提供し、Cito（テスト開発中央機関）（Centraal Instituut voor Toetsontwikkeling；以下Cito）、トゥベンテ（Twente）大学教育学部の研究機関であるOCTO（応用教育学研究センター）（Onderzoek Centrum Toegepaste Onderwijskunde；以下、OCTO）、SLO（カリキュラム開発機関）（Stichting voor Leerplanontwikkeling；以下SLO）という3つの機関が参加していた。プロジェクトの結果、当初想定されていたような包括的な自己評価ツールの開発は実現しなかったが、ツールの一部を担うはずであった学校の教育プロセスを評価

できるような ZEBO という自己評価ツールが OCTO を中心として開発された。ZEBO は、想定通りのツールとしては完成しなかったこと、商業ベースではなく大学の研究機関である OCTO によって開発されたことなどから、一般的な認知度は低く、学校の自己評価に実質的な貢献をしてきたというよりはむしろ主に研究的に利用されてきた。学校の自己評価が 2002 年 WOT 下で期待された役割を果たせなかったことをふまえれば、その制度が整備されゆく頃に、学校の自己評価をサポートするために開発されたツールについて、その開発の経緯を含めて改めて検討することの意味は大きいといえるだろう。

ZEBO についての研究は、その開発にあたったトゥベンテ大学を中心に行われてきた。特に、開発に関わったヘンドリックス(Hendriks, M.)らによって、開発の経緯⁶や使い方⁷、他ツールとの比較分析⁸などが行われた。ZEBO プロジェクト後は、同大学のスヒルトカンブ(Schildkamp, K.)によって、ZEBO を用いた自己評価ツールの使用や効果、自己評価に影響を与える要因の研究⁹、国際比較研究¹⁰などが進められた。一方、日本でオランダの学校評価が研究される際には教育監査に焦点があてられることが多く、ZEBO についての先行研究は存在しない。

以上より、本稿では ZEBO という自己評価ツールについて開発の経緯を含めて検討し、その意義と限界を明らかにすることを目的とする。それにより、学校の自己評価が義務づけられた日本の現状への示唆も得られると考えられよう。

1. ZEBO プロジェクト

1-1. OCTO による ZEBO の開発

最初に ZEBO 開発の背景をみていこう。オランダでは、もともと歴史的に憲法で「教育の自由」が保障されており、学校が高い自律性をもっていた。そうしたオランダでも、1980 年代には、学校の自律性の向上や地方分権・規制緩和が強調されてきた。1993 年には、各教科で子どもが達成することが望ましい目標が中核目標(kerndoelen)として設定された。こうして中核目標が国によって規定された一方で、学校は教育の質に責任を負うことが求められていった。

学校が行う教育の質の管理に関しては、1993 年に教育科学省の大臣と教育界代表との間で開催された「行政刷新についてのスヘフェニングス協議(Schevenings Beraad Bestuurlijke Vernieuwingen)」において、初等・中等学校は自身の教育の質に責任を持ち質の管理について積極的な方針をたてることや、学校は自己評価のための方法やツールを選ぶことなどが示された。しかし、当時、自己評価のためにどのような方法やツールが存在するのか、また、これらの方法やツールの質はどうかについては明らかではなかった¹¹。そのため、これをきっかけに、教育省が資金を出し、学校の自己評価ツールの調査研究などが行われた。

こうして行われた研究の 1 つに OCTO のクレイメルス・ファン・ベース(Cremers-van Wees)らによる学校の自己評価ツールの調査研究があった¹²。31 のツールが選ばれ、その相違点の検討などが行われた。そこでは、学校にとってどのツールが良いかは、その学校の状況や自己評価の目標などによって左右されることなどが指摘された。加えて、調査の結果、妥当性や信頼性がしっかりテストされたツールはあまり多くないことが明らかになった¹³。こうした研究結果を受け、より良いツールを開発するために 1995 年 ZEBO プロジェクトが始められた。

ZEBO プロジェクトは、妥当性・信頼性をもつ包括的なツール、すなわち精神測定理論や個

に応じた教授を重視した児童のモニタリング (Cito が担当)、カバーされた教育内容の評価 (主に SLO が担当)、さらに学校効果・学校改善研究を基盤にした学校プロセスの指標の測定 (OCTO が担当) を行なえるような包括的なツールの開発を目指すものであった。しかしながら、結果的に当初目指されたような包括的なツールが開発されることはなかった。最終的に開発されたツール ZEBO は OCTO が担当した教育プロセスの部分だけを反映したものであった¹⁴。

OCTO による ZEBO 開発の経緯は次の通りである。まず、ZEBO は、1995～6 年 ZEBO-SK-0 と呼ばれる段階で基本となる部分が開発された。その後、1997～9 年 ZEBO-SK-1～3 という段階それぞれにおいて、学校での実地調査と文献調査を繰り返すことで具体化された¹⁵。また実地調査の度に信頼性・妥当性の調査が行われた。

ZEBO 開発の基礎をなす ZEBO-SK-0 は、既存の自己評価ツールの調査から始められた。その調査の枠組は、オランダの代表的な学校効果研究者であり、ZEBO プロジェクト中間報告書等の編集者でもあるスヘーレンス (Scheerens, J.) の研究¹⁶にもとづき、若干の修正を加えて設定された。この枠組については詳しくは後述する。プロジェクトでは、この枠組を用いてオランダの学校の自己評価ツール 5 つ、効果のある学校研究で使われたツールや方法 9 つが調査された。加えて、5 つの学校の学校長や教師へのインタビューや、初等教育評価委員会の 1994 年のレポート¹⁷なども参考にされた上で、質問項目となる主な教育プロセスの指標が決定された。

こうして ZEBO-SK-0 において設計されたおおまかなツールをもとに、ZEBO-SK-1～3 において、学校での実地調査や妥当性¹⁸・信頼性¹⁹の調査が行われた。質問項目は、実地調査を繰り返す中で、学校の声やさらなる文献調査にもとづいて具体化されていった。このように、ZEBO は主には学校効果研究にもとづきながら、実際の学校の意見が反映された、妥当性・信頼性についても確かめられたツールとして開発された。

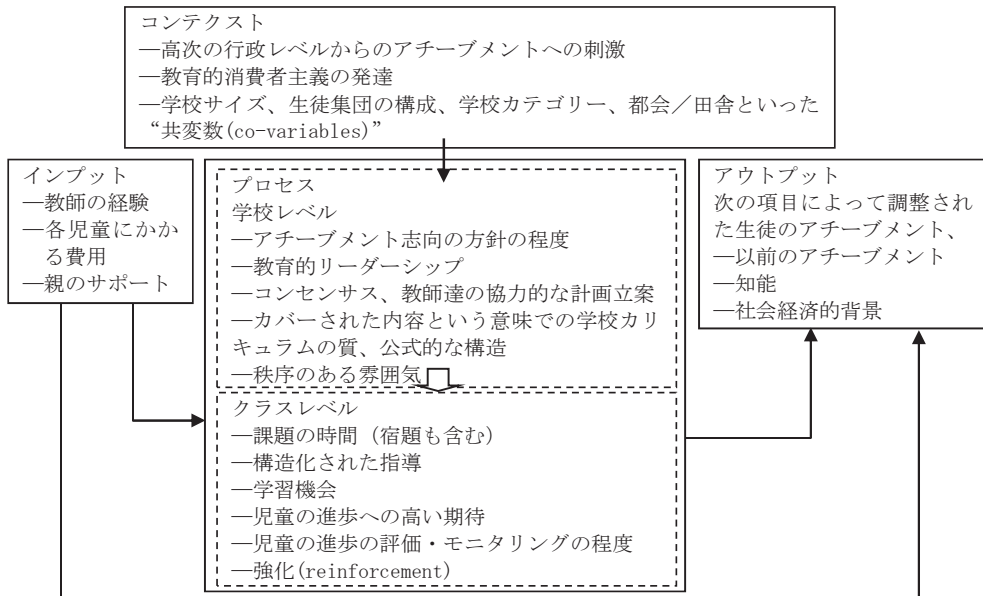
1-2. ZEBO の理論的背景—学校効果研究—

前節でみたように、ZEBO 開発の際、スヘーレンスのモデル (図 1) が調査の枠組の基盤となった。ここでは、このモデルをスヘーレンスがどのように開発したのかをみていきたい²⁰。

まず、学校の機能に関するプロセスの指標を決定する際に、スヘーレンスが学校効果研究の知見に着目した理由について確認しておこう。それは、スヘーレンスが、プロセスの指標は常にアウトプットの指標と結び付けられていなくてはならないと考えるためである。これらと切り離されると、「手術は成功したが、患者は死んだ」という言葉のように、プロセスの指標そのものがゴールへと置き換わってしまうことにもなりかねない。したがって、一般的に「生徒のアチーブメントとして測定される学校のアウトプットと積極的に結び付きがある学校の特徴を発見することを目的としている」とされる学校効果研究にスヘーレンスは着目したのである。

具体的にプロセスの指標を選択するに際しては、スヘーレンスは学校効果研究を次の 4 つのタイプに整理し、そこからの知見を取り上げている。それは、コールマン・レポートに代表される教育の (不) 平等と学校効果についての研究、インプット・アウトプット分析とも呼ばれる教育の生産機能についての研究、効果のある学校研究、教授の効果についての研究である。スヘーレンスはこれら 4 つのタイプの研究の特徴を整理し、それを図 1 のモデルへ反映させている。その中でも、プロセスの変数と関わりが深いのは、効果のある学校研究、教授の効果についての研究であるといえよう。

図1 スヘーレンスが開発した学校効果の統合モデル



出典：Scheerens, J., “School Effectiveness Research and the Development of Process Indicators of School Functioning”. *School Effectiveness and School Improvement*, Vol.1, No.1, 1990, p.73.を基に筆者が訳出。

効果のある学校研究は、コールマン・レポートに対する反発から行われ始めたものである。コールマン・レポートの頃の学校効果研究との違いは、当時は学校内において何が起きているかがブラック・ボックスになっていたのに対してそれらがオープンにされている点、そして学校組織、学校文化、教育技術といった学校の変数が明らかにされている点である。効果のある学校研究において、一貫してアチーブメントに効果があるとされたものとしては、安全で秩序ある学校の雰囲気、児童のアチーブメントへの高い期待と学校のアチーブメント志向の方針、教育的リーダーシップ、児童の進歩の頻繁な評価、基礎的な技能に関する明確な目標、教師間の継続性とコンセンサスがあげられている。図1をみるとプロセスに、これらほぼ全てが含まれていることがわかる²¹。基礎的な技能に関する明確な目標のみが含まれていないように見えるが、そうした目標に対応する内容がカリキュラムでカバーされているかどうかという点で、図1における学校レベルの「カリキュラムの質」と関わっていると考えられよう。

次に、教授の効果についての研究をみてみよう。特徴としては、教師レベル、クラスレベルで教育プロセスの特徴を研究することがあげられている。スヘーレンスは、他の研究者らによる研究レビューを参照し、アチーブメントと積極的に結びつきがあることが繰り返し示された変数として、次のものをあげている。それは、効果的な学習時間ないし「課題のための時間」、構造的ないし「直接的」指導、学習機会ないし「カバーされた内容」、教師の態度と期待、生徒の動機を高めること、家庭での変更可能なカリキュラム²²である。スヘーレンスは、生徒の動機を高める強化(reinforcement)を実現するためには進歩の頻繁なモニタリングやフィードバックが重要であるという研究も参照している。これらについても、家庭での変更可能なカリキュラム以外全てが図1におけるプロセスのクラスレベルの指標に含まれていることがわかる。

スヘーレンスは、これらの学校効果研究を統合して、図1のような枠組とするために、まずコンテキスト・インプット・プロセス・アウトプットという分析システムモデルを採用している。さらに、プロセスの指標は学校環境レベル、組織レベル（学校レベル）、クラスレベルで定義されうることから、マルチレベルの枠組が必要であると考えた。そこでは、児童レベルも考慮に入れられなくてはならないという²³。最後に、異なるレベルでの変数間の相互関係を見るための理論的視点として、コンティンジェンシー理論(contingency theory)を採用している。

一般的に、コンティンジェンシー理論とは、特定の組織構造や経営プロセスが効果的であるかどうかは、環境的制約や技術的制約といった文脈的な特徴に依存しているという考え方である。学校効果研究にこの考え方を適用するに際しては次の2つが示されていた。1 つめは、学校効果研究において効果といえば生産性だけが考えられてきたが、学校がおかれる状況などに応じて目指す効果は変わるという考え方である。もう1つは、学校がアチーブメント志向の方針を取るにあたって管理者、消費者やその他のステイク・ホルダーや政策などの外部刺激が重要であるという考え方である。さらに、コンティンジェンシー理論からスヘーレンスはメゾ・ミクロ関係という考え方も参考にしている。ここで、メゾは学校レベル、ミクロはクラスレベルを指す。学校効果研究では、学校レベルでのプロセスの特徴を明らかにする研究と、教師／クラスレベルでの教授的效果についての研究とが、別々に効果があるとされた特徴を特定してきたとスヘーレンスは考える。そこで、これらを統合するために、メゾ・レベルの状況がミクロ・レベルの状況を促進するという考え方を援用した。すなわち、学校レベルの状況がクラスレベルの状況を促進するという関係で2つの学校効果研究の知見の統合を考えたのである。こうした考え方にもとづくと、学校レベルと教師／クラスレベルの両方で意味を持つ変数があるとスヘーレンスは述べる。例えば、クラスレベルでの構造化された指導は、学校の発達計画にしたがって行われるといった点で、学校レベルでの明確なカリキュラムの方針によって刺激を受けうるといふ。図1はこうした考え方をふまえて図式化された。

最後に、こうしたプロセスの指標を活用するに際してのスヘーレンスの考え方を確認しておこう。プロセスの指標を活用する際には、アカウンタビリティ志向と改善志向の2つが考えられる。スヘーレンスは、両者ははっきりと線引きされるべきではないものの、プロセスの指標は、国や学区レベルで用いられる場合にはアカウンタビリティ志向となる一方で、学校の自己評価に用いられる場合には改善と強く結びつくと考えた。

以上より、スヘーレンスは、学校効果研究の知見にもとづきながら、コンティンジェンシー理論などを参考にしつつ、学校レベルとクラスレベルのプロセスとの結びつきを考慮してモデルを作成していた。さらに、プロセスの指標が学校の自己評価に活用される場合、改善と強く結びつくと考えていた。こうしたスヘーレンスの考え方は、学校評価・カリキュラム評価・授業評価を個別に捉えるのではなく、それらを全体として含む学校評価を実現し、改善へつなげる道を拓くものであると考えられる。ただし、実際に ZEBO プロジェクトで用いられた枠組は、スヘーレンスのモデルそのままではなく修正が加えられている²⁴。また、ツールが完成するまでには選ばれた指標に変更点もみられる。続いては、こうしたスヘーレンスの考え方が、ZEBO にどのように反映されているのかについてみていこう。

2. 「ZEBO」の内実

2-1. ZEBO 全体の構成

まず、ZEBO の全体的な構成を見ていこう。ZEBO は、学校長用・教師用・児童用の質問から成る。ZEBO はプロジェクト後にプログラム化され、学校は ZEBO を必要な時に必要に応じて使用できるようになった。最終的に作成された質問の大項目は図2の通りである。項目の内容の多くは、スヘーレンスのモデルの指標と一致している²⁵。また図2を見ると、学校レベルの質問は学校長と教師、クラスレベルの質問は教師と児童といったように、異なる立場にある人々から情報が集められている。これは、回答者の立場の違いによって回答に違いがでるといった問題に対応するためである²⁶。質に関連するプロセスの特徴は、事実だけでなく意見にもとづきうるという点でこうした問題を完全に回避することは難しい。そこで、できる限りバイアスを除く方法として二重チェックが採用された。すなわち、学校レベルでは、学校長は自己報告という形で情報を提供する一方で教師によって判断されるという形が取られたのである。

図2 ZEBO の質問の大項目タイトル

学校レベル		クラスレベル	
学校長・教師用の質問		教師用の質問	児童用の質問
1. 公式な会議	6. チームの団結	10. 教授行為	3 年生
2. 共働	7. 学校経営(者)	11. 評価形態	4~8 年生
3. 教育的リーダーシップ	8. 仕事量	12. 学力が低い児童と高い児童のためのケア	1. 学業成績へのプレッシャー (prestatiedruk)
4. 専門職性	9. 計画的な行動	13. クラスの作業雰囲気	2. 先生
5. 目標設定と期待			3. 先生
			4. 先生
			5. 作業
			6. 同級生
			7. 同級生
* 学校長・教師向けの質問は 7. 学校経営(者)で多少表現が異なる以外、同じである。		* 質問 10 以降は教師のみへの質問である。	1. 学業成績へのプレッシャー (prestatiedruk)
			2. 先生
			3. 教授行為
			4. 同級生
			5. 作業：個に応じている
			6. 作業：学習時間
			7. クラスの作業雰囲気

出典：Hendriks, M. に提供を受けた資料 (2009. 04. 28) を基に筆者が作成。

ZEBO では、学校レベルに関わる学校長用・教師用の質問 1～9 から学校報告書が、クラスに関わる教師用の質問 10～13 と児童用の質問からクラス報告書が作成される。ZEBO の児童用の質問は 3～8 年生を対象としている²⁷。3 年生向けの質問は、4～8 年生向けの質問より簡単に答えられるよう工夫されており、質問数も少ないといった違いがみられる²⁸。

2-2. 具体的な質問

それでは、具体的な質問例をみてみたい。スヘーレンスは例の 1 つに、クラスレベルでの構造化された指導は学校レベルでの明確なカリキュラムの方針によって刺激を受けうることをあげていた。したがって、まずはクラスレベルでの児童向けの質問 3「教授行為」をみてみよう。

図3 質問3.教授行為

a. 授業の始めに、先生は私たちが前回行ったことを繰り返す。
b. 先生は、たくさんの異なった方法で（算数の）計算問題を説明することができる。
c. 先生は、授業の始めに、私たちが今から行うことを話してくれる。
d. 先生は、新しい計算問題の説明の時、たくさんの例をくれる。
e. 先生は私たちが理解していない時、もう一度説明してくれる。
f. 先生は、私がどのように計算問題を解いたか、よく尋ねてくる。
g. あなたが分からない時、先生はもう一回何かを説明することを嫌だとは思っていない。
h. 私たちがクラスで自分たちのための作業をしている時、先生は助けるために歩き回っている。
i. 先生は、私たちがすでに以前に練習を積み熟練した計算問題をしばしば繰り返す。
j. 先生は私がどのように計算問題を解いたか決して尋ねてこない。

出典：Hendriks, M. に提供を受けた資料 (2009. 04. 28) を基に筆者が訳出。

図3をみると、前回の授業の復習・授業の導入・例の提示・説明・机間指導など具体的に教師の教授方法が取り扱われていることがわかる。先述したように、各質問内容は立場の異なる者から二重チェックが行われるようになっている。したがって、例えば、児童向けの質問「a. 授業の始めに、先生は私たちが前回行ったことを繰り返す」にあたる内容は、教師向けの質問10でも「a. 授業の始めに、私は前の授業で扱われたことを繰り返す」「b. 授業の始めに、私はそれ以前の授業の短い要約を児童に伝える」という形で問われている。

図2からわかるように、学校レベルでの質問に明確なカリキュラムの方針という質問項目はない。しかしながら、実際には教授行為やカリキュラムの方針に関わる質問は数多くみられる。例えば、質問1「公式な会議」では、「a. 教育内容を提供する方法」「j. 異なるクラスの教育内容の活動をお互いに調整する」などのテーマが公式な会議でどのくらいの頻度で議題になるかが問われている。また、質問2「共働」でも、「a. 扱われる授業内容について取り決めるすること」「b. (一連の) 授業の準備」「c. 指導 (lesgeven) でのお互いのサポート」「d. 指導でのお互いの観察」「e. 授業後共に話し合うこと」などについてどの程度非公式に教師たちが協力できているかが問われている。教師たちが教授行為について話し合い、助け合っているかが問われていることがわかる。

さらに、質問9「計画的な行動」においては、「a. 私たちの学校では各学年の最終目標が決められている」「i. 私たちの学校では、学年計画²⁹に沿って働く」「j. 私達の学校では、学年計画は規則的に評価され、もし必要なら修正される」などが含まれており、学校が具体的な目標を設定し、それにもとづく計画を立て、それが計画通りに実施されているか、必要であれば計画が修正されるかなどが問われている。授業に影響を与えるカリキュラムの自主編成が実現されているかどうかが問われていることがわかる。

ただし質問5「目標設定と期待」をみると、「a. 私たちの学校では、できるだけ多くの時間をオランダ語と算数／数学に使うことができるように時間割を編制している」「b. 学校の最も重要な目標はオランダ語と算数／数学の学習である」「c. 私たちの教育では、認知的発達と学業成績 (leerprestaties) に強調点を置く」などの質問が含まれている。このように認知的発達や学業成績が重視されているのは学校効果研究の影響であろう³⁰。このことから、ZEBO を利用すると、学校方針としてそれらに重点をおくことが推奨されると考えられる。これについては、ZEBO の評価結果の在り方とも関わってくるため、詳しくは次節で考察したい。

以上より、ZEBO では、学校レベルでカリキュラムを編成し、授業での取り組みを支えることができているかが問われるとともに、クラスレベルでどのような教授行為が実現されているかが問われていた。スヘーレンスが例にあげていたような両方のレベルで意味を持つ内容が質問項目に含まれていたという点で、両レベルの関連が ZEBO でも考慮されているといえる。ただし、ZEBO を利用すると、学校方針として認知的発達や学業成績に重点をおくことが推奨されると捉えられる側面もあった。

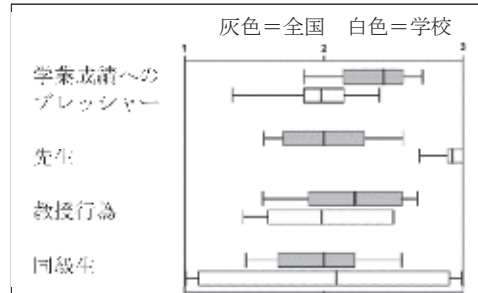
2-3. クラス報告書

次に、報告書を具体的にみていこう。ZEBO を全て使った場合、学校報告書とクラス報告書の2種類の報告書がコンピュータで自動的に作成される。学校報告書・クラス報告書の形式は同様であり、報告書はグラフと文章、各スコアの形で作成される。

ここでは、クラスの報告書をみていこう。図4のように、グラフは箱髭図を用いて作成される。ここでは、各質問内容における児童間の意見のバラつき（白い箱と線の横幅）、児童の意見の平均（白い箱の中の黒い縦線）、オランダの他の学校サンプルにおける児童の意見のバラつき（グレーの箱と線の横幅）、その平均³¹（グレーの箱の中の黒い縦線）があらわされる。

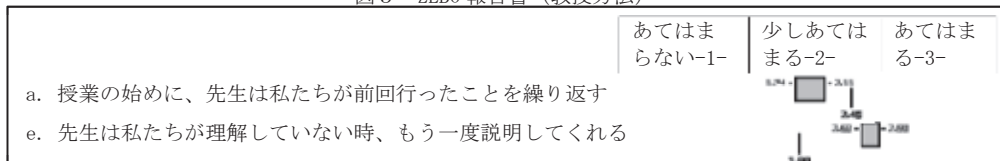
例えば図4をみると、「学業成績へのプレッシャー」について、この学校は普通程度であるようにみえるが、オランダの他の学校と比べると児童のプレッシャーが低いことが分かる。また、「同級生」については、平均は普通程度で他の学校と変わらないものの、白い箱が横長くなっていることから、多様な意見が児童の中にあることがわかる。このように、この学校では、「学業成績へのプレッシャー」「同級生」について、より念入りに検討する必要があることがわかる。

図4 ZEBO 報告書例（クラスの概観）



出典：Hendriks, M. に提供を受けた資料（2009. 04. 28）を基に筆者が作成。

図5 ZEBO 報告書（教授方法）



出典：Hendriks, M. に提供を受けた資料（2009. 04. 28）を基に筆者が作成。

また、各個別の質問への児童たちの回答は図5のようなグラフで示される。グラフでは、オランダの他の学校のスコア（グレーの四角）と、その学校における児童の意見の平均値（黒い縦線）が示される。児童の意見は平均値のみが示され、各児童がどの質問にどう答えたかは分からないよう匿名性に配慮されている。図5を見ると、質問aより、この学級の先生は、授業の始めに前回の復習を行っていることがわかる。しかし、質問eより、子どもが分からないと感じていても、もう一度説明することがあまりないようである。このように、報告書では、各質問結果から、教師は教授行為について具体的に改善点等を参考にすることができる。さらに、ZEBO の報告書には文章も添えられる。クラスの報告書では、他の学校における同学年の児童の結果との比較、そのクラスの教師と児童の結果との比較が簡単な文章であらわされる。

このように、ZEBO の報告書からは、学校で検討されるべき3種類の項目を特定できる。①平均の結果が良くない項目、②意見の相違やバラつきの多さのため検討されるべき項目、③他の学校と比べて良くない項目である。ZEBO の報告書では各結果の平均値だけでなく、意見の相違やバラつきが示される。そのため、そうした項目についての学校内でのコミュニケーションが促される。ZEBO を使用した結果、必ずしも ZEBO の平均の結果を上げることが目指される訳ではなく、自己評価後にどのような改善を行うかについては学校に委ねられている。ZEBO ではできる限り学校が自身で改善計画を実行できることが意図されている³²。これは、学校がおかれる状況などに応じて目指す効果は変わるという考え方をスパーレンスがコンティンジェンシー理論から採用していたことと無関係ではないように思われる。

このことから、ZEBO が認知的発達に重点をおくことを推奨していても、学校が必ずしもそれに従う必要がないことが示唆される。認知的発達に重点をおくことについて、教職員のコミュニケーションが促された結果、ZEBO の質問項目とは異なる方向での改善策が導き出されることも十分に考えられる。スヘーレンスによって自己評価は改善と強く結びつくと考えられていたが、この過程での学校の自律性が尊重されていることがわかる。

2-4. ZEBO の使用状況

最後に、ZEBO 開発後の使用状況について確認しておこう。まず、ZEBO はプロジェクト後もトゥベンテ大学のスヒルトカンパによってその実際の使用や効果などについて研究が進められた。特に、スヒルトカンパは、2001-2 年度から情報収集を始め、2002-3 年(2003-4 年)、2004-5 年、2005-6 年の 3 回、学校での ZEBO の使用を調査することで、①どのように、どの程度学校は ZEBO を使うのか、②(意図されていない効果も含め) ZEBO の効果は何か、③どんな要因が ZEBO の使用に影響するのかに関して、フィッシャー(Visscher)³³が開発した理論枠組を援用して分析を行い、博士論文にまとめている³⁴。

分析の結果、スヒルトカンパは、ツールの選択といった決定プロセスに教職員全員が関与する必要、自己評価プロセスの結果が教職員に明確である必要、学校長が自己評価ツールの使用を奨励する必要といった要因が自己評価ツールを成功裏に使うために重要であることなどを指摘している。ただし、ZEBO の効果に関しては、ZEBO の使用と児童のスペリングや算数のアチーブメントには重要な関連はみられなかったという。これについて、スヒルトカンパは、ZEBO のアウトプットが限定的にしか学校で使用されなかったためであろうと考察している。一方で、ZEBO を使用する学校では、特に、学校内での協議やコミュニケーションといった学校改善のための重要な必要条件に関して効果が認められている。これは先にみたように、ZEBO の報告書において、各結果の平均値だけでなく意見の相違やバラつきが示されるためであろう。

また、スヒルトカンパは、学校が ZEBO のアウトプットを改善のために使用することの難しさを明らかにしている。ZEBO では、その結果をどう使うかについては学校側に委ねられている。先に見たように、ZEBO の評価項目は具体的であり、改善方策は比較的考えやすいと思われる。それでも、多くの学校は評価結果を改善のための方策につなげることができなかったという。これは ZEBO に固有の問題というよりは、自己評価の実施一般にいえることである³⁵。そこで、スヒルトカンパは、自己評価ツール設計者への今後の示唆として、自己評価結果の使い方のガイドラインや、改善のために取られうる活動例の提供などを提案している。

次に、オランダでの ZEBO の位置づけを確認してみたい。2002～2007 年、Q*Primair という国のプロジェクトグループが設置され、自己評価ツールの調査などが行われている。2005 年に Q*Primair から出された調査報告書によると、教育監査の評価指標にもとづく WMK というツールはシェアが 30.1%であるのに対して³⁶、ZEBO は 2.1%である³⁷。より良い自己評価ツールを目指して開発された ZEBO であるが、実際的には、教育監査の評価指標にもとづくツールなどと比べてシェアはあまり高くなかったことがわかる。さらに、報告書によれば、ZEBO の信頼性は 5 段階中 4 (良い)、妥当性が 5 段階中の 2 (並み) と評価されている。一見、妥当性が低いように思われるが、他のツールと比べれば妥当性も信頼性も高い方である。なお、補足しておけば、ここでの妥当性の低さは「妥当性」解釈の違いによるものと考えられよう³⁸。

このように ZEBO に対しては、学校内でのコミュニケーションを促すといった意義が認められた一方で、その結果を改善へつなげることの難しさ等が明らかになっていた。また、ZEBO は他のツールと比べ、妥当性・信頼性は比較的高い一方で、シェアは低いことがわかった。

おわりに

本稿では、オランダにおける学校の自己評価ツールの 1 つ ZEBO を検討してきた。以上の検討から、ZEBO の意義と限界を整理してみよう。まず、意義としては、ZEBO は示唆的な学校評価の在り方を提示している。ZEBO はスヘーレンスの考え方にもとづくことで、学校評価・カリキュラム評価・授業評価を個別に捉えるのではなく、それらを全体として含む学校評価を実現し、改善へつなげていく道を拓こうとしていた。さらに、どうしても主観性を含むことを免れえない学校評価において、できる限りバイアスを除くために、立場の異なる者に同じ内容を評価させる二重チェックという方法を取り入れていた。加えて、学校評価における評価項目を絶対視することなく、意見の相違が高いものを示し、学校内におけるコミュニケーションの糸口を提示するという評価の在り方を示していた。

次に、限界について考えてみたい。ZEBO に対しては、ZEBO の使用とアチーブメントには重要な関連がみられないことや、ZEBO のアウトプットを改善に使用することの難しさ等が指摘されており、シェアも低いことがわかった。その要因としては、次の 2 つが考えられよう。

まず、学校効果研究に立脚して選ばれた評価指標が必ずしも学校のニーズと合うとは限らないことである。「教育の自由」が保障されているオランダでは多様な学校が存在するため、例えば情意面にも重点をおく学校では ZEBO の評価指標だけを用いた自己評価では不十分である³⁹。そうした要因がシェアの低さにつながっていると想定される。ただし、評価指標が学校のニーズに沿っていることがただちに良い自己評価へつながる訳ではない点には留意が必要であろう。

次に、ZEBO にはできる限り学校が自身で改善計画を実行できるようにとの意図があり、評価の後に ZEBO の評価指標に沿って学校が変化を起こすことが必ずしも求められてはいないため、ツール開発時に想定された学校効果の実現が期待できるとは限らないことである。もちろん、ZEBO は妥当性・信頼性を持って評価を行うためのツールであり、改善活動とは分けて考える必要はあろう。しかし、学校効果研究は、効果があるとされる学校の特徴を発見することを目的とした研究である。その場合、そこで評価された項目に沿って改善が行われなければ、想定された学校効果の実現も期待できるとは限らないことになる。ここに、「教育の自由」のもとで学校が自主的に改善計画を立てることを重んじながら、研究上効果が実証された特定の教育活動を奨励することの難しさが生じてくる。そうした中で、自己評価の結果を改善へつなげるためにどのようなサポートができるのかについて、ZEBO には未だ課題が残るといえよう。

以上をふまえれば、ZEBO は学校効果研究に立脚しながら妥当性・信頼性を確保することを目指して開発されたツールであったがゆえにこそ、その限界を生じさせてしまったとも捉えられる。それでも、先にみたような ZEBO が提起する学校評価の在り方は、日本の学校評価を考える上でも示唆に富んでいる。日本では一般的にアンケートによって学校の自己評価が実施されている。日本の現状に対しては、自己評価の結果、評価の分かれた事項については「中程度」に平均され、「問題」として見えなくなり議論の対象から除外される傾向があるとの指摘や⁴⁰、

教育課程や授業の評価は学校改善にとって重要であるにもかかわらず、学校評価において軽視されているとの指摘もある⁴¹。こうした点に関して、ZEBO が示す学校評価の在り方は、二重チェックという方策を用いながら、平均値だけでなく意見の相違や有無を示すことで、同じ立場内（例えば児童間）・違う立場間（例えば教師と児童）で評価の分かれた事項を議論の俎上に置くことを可能にしたり、学校レベルでもクラスレベルでもカリキュラムや授業といった同様のテーマを取り上げたりするなど、多くの示唆を含んでいる。

今後は、それぞれの学校らしさを生かしたより良い学校評価の在り方を探るべく、オランダのオルタナティブ・スクールが考える新しい学校評価の形について研究を進めていきたい。

¹ オランダの教育・文化・科学省ウェブサイト

[<http://www.rijksoverheid.nl/onderwerpen/vrijheid-van-onderwijs/zelf-een-school-oprichten>] (2012 年 9 月 20 日確認)。

² リヒテルズ直子「オランダにおける第三者評価制度」株式会社三菱総合研究所『学校の第三者評価の評価手法に関する調査研究—最終報告書—』2007 年、p.54。

³ 奥村好美「オランダの初等教育における学校評価政策の動向—学カテストの位置づけに注目して—」京都大学大学院教育学研究科『京都大学大学院教育学研究科紀要』第 59 号、2013 年、pp.583-595。

⁴ 2002 年 WOT によれば、監督枠組で示されているような監査で判断が行われる際に関係する全ての質の側面が自己評価において取り上げられており、評価の実行方法や質(hoedanigheid)が十分に信頼でき、その機関自身が設定した質の目標が十分なレベルである場合に、監査は自己評価の結果から始められ、自己評価の結果は監査の判断に指針を与えるという。それにも関わらず、ほとんどの学校の自己評価はその条件を満たせなかったことが指摘されている (Smeets, G., & Verkroost, J., *Selectief en slagvaardig: werken met de WOT (2000–2010)*, Utrecht: Inspectie van het Onderwijs, 2011, p.77)。

⁵ SVO は 1997 年に NOW(de Nederlandse organisatie voor Wetenschappelijk Onderzoek)に合併。

⁶ Hendriks, M.A., Doolaard, S., & Bosker, R.J., “Using school effectiveness as a knowledge base for self-evaluation in Dutch schools: The ZEBO-project”, (In A.J. Visscher & R.Coe(Eds.), *School improvement through performance feedback*, Lisse/Abingdon/Exton/Tokyo: Swets & Zeitlinger, 2002, p.115-142)など。

⁷ Hendriks, M. & Bosker, R., *ZEBO instrument voor zelfevaluatie in het basisonderwijs. Handleiding bij een geautomatiseerd hulpmiddel voor kwaliteitszorg in basisscholen*, Enschede, The Netherlands: Twente University Press, 2003。

⁸ Hendriks, M.A., “Schooleffectiviteitsonderzoek en kwaliteitszorg in scholen”, *Nederlands Tijdschrift voor Onderwijsrecht en Onderwijsbeleid* (jrg.15, nr.4, 2003, pp.195-216)など。

⁹ Schildkamp, K., *The utilization of a self-evaluation instrument for primary education*, Enschede: Ph.D thesis for Universiteit Twente (2007)など。

¹⁰ Schildkamp, K. & Teddlie, C., “School performance feedback systems in the USA and in the Netherlands: a comparison”, *Educational Research and Evaluation* (Vol.14, No.2, 2008, pp.255-282)など。

¹¹ Cremers-van Wees, L.M.C.M., Rekveld, I.J., Brandsma, H.P., & Bosker, R.J., *Instrumenten voor Kwaliteitszorg: Inventarisatie en beschrijving*, Enschede: Universiteit Twente, Onderzoek Centrum Toegepaste Onderwijskunde, 1996a, voorwoord。

¹² Cremers-van Wees, L.M.C.M., Rekveld, I.J., Brandsma, H.P., & Bosker, R.J., *op. cit.* (1996a)及び Cremers-van Wees, L.M.C.M., Rekveld, I.J., Brandsma, H.P., & Bosker, R.J., *Instrumenten voor kwaliteitszorg: Beschrijving van 31 instrumenten* (Enschede: Universiteit Twente, Onderzoek Centrum Toegepaste Onderwijskunde, 1996b)など。

¹³ ここでの信頼性は、そのツールが例えば一カ月の間を置いて再度使われた場合同じ結果が得られるかどうか、妥当性はツールの設計者が測りたいものが実際に測れているかどうかである(Cremers-van Wees, L.M.C.M., Rekveld, I.J., Brandsma, H.P., & Bosker, R.J., *op. cit.*, 1996b, voorwoord)。

¹⁴ オランダでは Cito の初等教育最終学年のテストやモニタリング・システムがかなり普及しているため、教育プロセスを評価できる自己評価ツールが開発されれば、学校が自己評価の結果と Cito によるテストの結果を合わせて用いることは可能である。

¹⁵ 学校の負担を軽減するため、ZEBO-SK-1 では 4・5・6 年生、SK-2 では 3・5・7 年生、SK-3 では半分の学校が 3・6 年生、もう半分の学校が 4・5・7 年生を対象に実地調査は行われた。

¹⁶ Scheerens, J., “School Effectiveness Research and the Development of Process Indicators of School Functioning”, *School Effectiveness and School Improvement*, Vol.1, No.1, 1990, pp.61-80。

¹⁷ Commissie Evaluatie Basisonderwijs, *Zicht op kwaliteit. Evaluatie van het basisonderwijs. Eindrapport* (Den Haag: Sdu, 1994)のこと。

- ¹⁸ 妥当性の調査は、収束的妥当性（立場の違う者（例：教師と学校長）の間で同じ概念が測定されているかを調べるために二重チェックでの相関関係を算出）・弁別的妥当性（同じ立場の者（例：児童同士）のなかで、同じ概念が測定されているかを調べるために変数間の相関関係を算出）・予測的妥当性（半年間で児童が修める成績(*leerwinst*)と測定されるプロセスの変数の間の関係を推定。ここでは Cito のモニタリング・システムの成績が使われた）に関して分析が行われた。
- ¹⁹ 信頼性の調査は、クロンバックの α 係数を用いて、個人レベル（学校長/教師/児童）、集約レベル（学校レベル/クラスレベル）共に調べられた。
- ²⁰ 本節での記述は Scheerens, J., *op. cit.* (1990) にもとづく。
- ²¹ 「児童のアチーブメントへの高い期待」は「児童の進歩への高い期待」とほぼ同義であると思われる。
- ²² 学校の成績に影響を与えようと考えられるような家庭状況のことである。例えば、学校で子どもが何をしているかに親が関心を持つ、家庭で読み聞かせを行う、テレビの視聴を抑えるなど。
- ²³ スーパーレンスはクラスレベルでの児童レベルも考慮しているが、一般的にはクラスレベルは教師レベルを指すことが多い。そのため、児童レベルが含まれない場合は、教師/クラスレベルとしている。
- ²⁴ 教育プロセスを評価するツールの開発が目指されていたことから、スーパーレンスのモデルからもプロセスの変数だけが取り上げられたこと、スーパーレンスの研究以降の効果的な学校研究から「親の関与」や「作業雰囲気」といった変数に加えられたことなどである。
- ²⁵ Hendriks, M.A., Doolaard, S., & Bosker, R.J., *op. cit.*, p.124.
- ²⁶ Hendriks, M.A., Doolaard, S., & Bosker, R.J., *op. cit.*, p.123.
- ²⁷ オランダの初等学校 3 年生は日本の小学校 1 年生に該当するため、実際には日本の小学校 1～6 年生にあたる児童を対象としている。
- ²⁸ 大項目 1、5「作業」では、1 で作業が簡単すぎることはないか、5 で作業が難しすぎることはないかが問われている。大項目 2、3、4「先生」では、2 で先生の印象、3 で先生が自分をどう思っているか、4 で教授行為を自分がどう感じているかが問われている。また、大項目 6、7「同級生」では、6 で同級生が自分をどう思っているか、7 で自分が同級生をどう感じているかが問われている。
- ²⁹ 学年計画：全学年で提供される教育内容、活動、練習問題と特別な助けを書き起こしたもの。
- ³⁰ 認知的発達重視に関しては、ペーパーテストのようなテストのみが想定されている訳ではない。質問 11.「評価形態」をみると「d.製作品」「e.口頭試問や指名」「f.クラス全体での授業での会話」「g.授業時間での観察」などが含まれており、多様な評価方法が考えられている。
- ³¹ ただし、ここでの国の平均値は、1999 年 ZEBO プロジェクト最終段階において 121 校から集められたデータであり、毎年更新されているわけではない。
- ³² Hendriks, M.A., Doolaard, S. & Bosker, R.J., *op. cit.*, 2002, p.135.
- ³³ トッペンテ大学の准教授。学校効果研究、特に学校での質の管理の情報技術的側面や効果の改善を支援するパフォーマンスの質の高いフィードバックの提供方法などを研究している。
- ³⁴ Schildkamp, K., *op. cit.*, 2007.
- ³⁵ スヒルトカンプは教育監査局の調査でも、自己評価を定期的に行い、その結果を質の改善のために使える初等学校は 3 分の 1 であったと紹介している。
- ³⁶ 2010 年に筆者が WMK 開発者に行ったインタビューによれば、WMK の利用者はその後増え続け、2010 年の時点では、半数以上の初等学校が利用している。なお、ZEBO についても、ZEBO プロジェクトの最終段階において、ZEBO の結果が監査の結果とかなり対応することが指摘されている。
- ³⁷ Dijkstra, N., Linde A.van der & Majoor, D., *Kiezen voor Kwaliteit: tweede versie 2005*, Instrumenten de maat genomen. Q*Primair, 2005, p.72.
- ³⁸ Q*Primair の調査では、「妥当性と規準(*normering*)」については、妥当性の程度、内容の妥当性（理論的基礎）、基準のタイプ、周囲との明確な比較可能性(*Omgevingsspecifieke vergelijkingsmogelijkheid*)、外部の集団の関与、ツールの認定(*accreditatie*)があげられていた（Dijkstra, N., Linde A.van der & Majoor, D., *op. cit.*, 2005, p.11）。
- ³⁹ 比較的シェアが高い WMK という自己評価ツールは、教育監査の評価指標にもとづきつつも、一般的な学校用だけでなく、多様な学校種に対応できるようモンテッソーリ用やイエナプラン用のツールも開発してきている。さらに、WMK では学校独自の評価指標をツールに加えることも可能である。
- ⁴⁰ 木岡一明「豊かな学校評価を共に作り出す—学校組織開発に生きる学校評価に向けて—」長尾彰夫他編『学校評価を共に創る—学校・教委・大学のコラボレーション』学事出版、2003 年、p.22.
- ⁴¹ 田中統治「日本の学校評価に必要なカリキュラム評価とは何か」田中統治編『カリキュラム評価の考え方・進め方』（教職研修 12 月号増刊号）教育開発研究所、2005 年、pp.48-51.

（日本学術振興会特別研究員 教育方法学講座 博士後期課程 2 回生）
 （受稿 2013 年 9 月 2 日、改稿 2013 年 11 月 28 日、受理 2014 年 1 月 21 日）

A Study of the Significance and Limits of ZEBO in the Netherlands: The Effort to Enrich School Self-Evaluation

OKUMURA Yoshimi

A key feature of the Dutch education system—guaranteed under article 23 of the constitution—is freedom of education (i.e., the freedom to found schools, organize teaching, and determine operating principles). This paper discusses the significance and limits of the school self-evaluation instrument ZEBO (Zelfevaluatie Basisonderwijs in Dutch; self-evaluation in primary education) in the Netherlands. ZEBO is designed to measure school process indicators, with school effectiveness and school improvement research as the conceptual background. It was developed as a direct result of evaluation studies that noted a general lack of attention to the reliability and validity of school self-evaluation instruments in the Netherlands. This study provides general information on the ZEBO project and its theoretical basis. In addition, it describes the content of the instrument in detail. Finally, it shows that ZEBO is suggestive as a useful tool for school self-evaluation, but, at the same time, it is limited from the perspective of autonomous school improvement under freedom of education because it is based on school effectiveness and school improvement research.