



Title	算数科の学習評価におけるペーパーテストの可能性と限界に関する一考察
Author(s)	石井, 洋
Citation	北海道教育大学紀要. 教育科学編, 71(1): 213-225
Issue Date	2020-08
URL	http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/11380
Rights	

算数科の学習評価におけるペーパーテストの可能性と限界に関する一考察

石 井 洋

北海道教育大学函館校数学教育研究室

A Study on the Possibility and Limitation of the Paper Test on Learning Assessment in Mathematics

ISHII Hiroshi

Department of Mathematics Education, Hakodate Campus, Hokkaido University of Education

概 要

本稿では、児童の学力を捉えるための評価方法の一つであるペーパーテストに焦点を当て、現在我が国の小学校で用いられている業者テスト、標準式学力テスト、全国学力・学習状況調査を取り上げて分析した。旧来、ペーパーテストは、学習に向かわせるための動機づけや、学習結果を把握する資料として用いられ、信頼性や妥当性が高い評価方法である。それは、「知識・技能」を測定するためには問題なく使用でき、便利なものであるが、「思考・判断・表現」を測定する上で、テストによって評価できる範囲の限界があることが明らかとなった。特に、全国・学力学習状況調査においては、現実的な問題状況を設定するための詳細な文脈の説明から読解力の影響が考えられること、正誤の二分法で評価することによる思考力・判断力・表現力の深浅さが判断できないこと、ICTを含むツールの活用ができないことといった制限があることを課題として提示した。

1. はじめに

昨今、異常気象の増加やCOVID-19等、これまで人類が経験したことのない社会を迎えており、予測困難な問題に対応できる資質・能力を身につける必要性が一層高まってきている。中央教育審議会答申（平成28年12月21日）においては、そのような社会の変化に主体的に関わり、よりよい社会と幸福な人生の創り手となる力として「生きる力」をより具体化し、教育課程全体を通して育成

を目指す資質・能力として「知識及び技能」、「思考力、判断力、表現力等」、「学びに向かう力、人間性等」の三つの柱に再整理している。そして、その資質・能力を捉えるための観点別学習状況評価についても、目標に準拠した評価の実質化や、教科・校種を越えた共通理解に基づく組織的な取組を促す観点から、「知識・技能」、「思考・判断・表現」、「主体的に学習に取り組む態度」の3観点到に整理されることになった。

このような観点において資質・能力のバランス

のとれた学習評価を行っていくためには、論述やレポートの作成、発表、グループでの話し合い、作品の制作等といった多様な活動に取り組みさせるパフォーマンス評価を取り入れ、ペーパーテストの結果にとどまらない、多面的・多角的な評価を行っていく必要性が指摘されている。そうしたことから、求められている学力を捉えることは一層困難になってきたといえる。

しかし、学習評価においては、従前としてペーパーテストの影響力は大きい。毎年、全国学力・学習状況調査の結果が公表される時期には報道等で大きく取り上げられ、3年毎に行われるPISA調査も教育指針の方向性を検証している実態がある。また、高等学校においては、客観的で信頼性や妥当性が高いとされ、未だペーパーテストの評価を使い続けていると指摘されている（工藤，2011）。高校、大学の進学のための試験では、AO入試等で学校での成績や小論文、面接などで人物を評価することも増えているが、大多数は依然としてペーパーテストによる選抜を行っており、センター試験に変わって実施される大学入学共通テストに注目が集まるなど、我が国では依然として学力評価をペーパーテストに依存している現状にある。

一方、入試対策をほとんど必要としない小学校においても、学習指導要領に多様な評価方法を用いて適切に評価を行うことが述べられているにも関わらず、従来より用いられている業者テストで評価しているケースが少なくない。新しい学力観に対応した評価を行うべく、パフォーマンス評価等、多面的・多角的な評価方法が検討され、ここではペーパーテストの限界が述べられているが、どのような点に限界があるのかが十分に議論されていない現状がある。

そこで、本研究では児童の学力を捉えるための評価方法の一つであるペーパーテストに焦点を当て、現在我が国の小学校で用いられている業者テスト、標準式学力テスト、全国学力・学習状況調査を取り上げて分析する。ここでは、それぞれの可能性や限界について考察し、限界を一部解消す

るものとして注目されている「パフォーマンス評価」についてペーパーテストで扱う際の方向性について提示する。

2. 今日の学力評価の動向

(1) 今日の学力の捉え方

1990年代、ゆとり教育による学力低下論争が起こり、それに追い打ちをかけるような形で巻き起こったPISA2003の「PISAショック」により、学力向上が現在の教育政策において強調されている。その後、PISAの影響もあり、我が国では学力に関して「新しい能力」概念が提案されるようになった（表1）。

表1 「新しい能力」概念

名称	機関・プログラム	出典	年
生きる力	文部科学省	中央教育審議会答申『21世紀を展望した我が国の教育の在り方について－子供に「生きる力」と「ゆとり」を－』	1996
リテラシー	OECD-PISA	国立教育政策研究所編『生きるための知識と技能』	2001 (2004・2007)
人間力	内閣府（経済財政諮問会議）	『人間力戦略研究会報告書』	2003
キー・コンピテンシー	OECD-DeSeCo	ライチエン&サルガニク『キー・コンピテンシー』	2006 (原著2003)
21世紀型スキル	21世紀型スキルプロジェクト	グリフィン・マクゴーケア『21世紀型スキル：学びと評価の新たなかたち』	2014 (原著2012)
21世紀型能力	国立教育政策研究所	『教育課程の編成に関する基礎的研究 報告書5』	2013

松下（2010）に加筆

新学力観から端を発した「生きる力」については、平成8年の中央教育審議会答申において、変化の激しいこれからの社会の中で子供たちに必要となる資質・能力として用いられ、現在に至るまで学習指導要領の目標の核となっている。その後、OECD-DeSeCoが提案したキー・コンピテンシーや、21世紀型スキルプロジェクトの21世紀型スキル等、諸外国の資質・能力目標を踏まえ、我が国では「知識及び技能」、「思考力、判断力、表現力等」、「学びに向かう力、人間性等」の三つの柱に再整理されている。この学力の3要素につい

ては、学習指導要領の各教科における目標や評価の観点としても採用されている。

一方、石井（2015）は、ブルーム・タクソノミーの教育目標の分類を拠り所とし、能力の質的レベルを明確化することを提案している。そこでは、個別の知識・技能の習得としての「知っている・できる」知識と概念の意味理解としての「わかる」知識、知識・技能の総合的な活用力としての「使える」知識をレベルとして設定している。

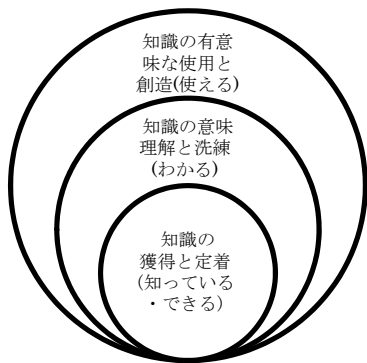


図1 めざす学力・学習の質の明確化（石井，2012）

このように、学校教育において育む資質・能力は、3つの要素「知識」「スキル」「情意」と3段階の質的な分類「知っている」「わかる」「使える」がなされている。それを石井（2015）は2次元の枠組みを用いて整理している。

表2のように学力の要素と質は、学習指導要領において文部科学省が「確かな学力」として示している育成すべき資質・能力と対応している。学校教育において、これらの学力をどのように身につけさせるか、また、どの程度身についたのかを把握することが必要となる。

本稿で焦点を当てるペーパーテストは、学習に向かわせるための動機づけや、学習結果を把握する資料として元来から用いられ、信頼性や妥当性が高い評価方法であるが、本節で取り上げた学力をどこまで捉えられるかについては明確になっていないという課題がある。

(2) 評価の捉え方

鹿毛（2000）よれば、評価活動には「価値判断

表2 学校で育成する資質・能力の要素の全体像を捉える枠組み

能力・学習活動の階層レベル（カリキュラムの構造）		資質・能力の要素（目標の柱）			
		知識	スキル		情意（関心・意欲・態度・人格特性）
教科等の枠づけのなかでの学習	知識の獲得と定着（知っている）		認知的スキル	社会的スキル	
	知識の獲得と定着（知っている）	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感
	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、関連づけ、構造化、比較・演繹的推論		
知識の有意意味な使用と創造(使える)	見方・考え方(原理、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、美的表現(批判的思考や創造的思考が関わる)	プロジェクトベースの対話(コミュニケーション)と協働	内容の価値に即した内発的動機、教科への関心・意欲	

石井（2015）を一部抜粋

としての評価」と「問題解決としての評価」といった2つの側面があるという。前者は、生徒一人ひとりを対象として、その性質（性格や能力等）や達成（学習成果等）を具体的に値踏みする作業である。それは、教師にとってつける作業として手続き化されており、評価資料としてテスト、ノート、提出物を用いて判断し、学習成果として示す。そのため、この評価には、評定主義、結果主義といった考え方が潜んでいる（鹿毛，2000）。

一方、「問題解決としての評価」は、価値判断が最終目的とはならない評価である。まず教師は評価の対象を把握し、それを値踏み（判断）する。ここまでの過程は、「価値判断としての評価」と同様である。その後、価値判断した内容を改善への行為の意思決定に活用し、そのサイクルを循環させていくのが「問題解決としての評価」である。ここでの評価の目的は、値踏みをするのではなく、よりよく生徒への指導を進めることにある。これらは、その評価を行う時期と機能から、前者を総括的評価、後者を形成的評価といい、近年では、「学習の評価」、「学習のための評価」として再定義されている。

本稿では、前者の「学習の評価」について扱うが、我が国では「目標に準拠した評価」を採用し

ている。「目標に準拠した評価」とは、学習指導要領に示された育成すべき資質・能力としての教育目標をどの程度身につけているかを評価するものである。そのため、その目標として何をどのように設定するかが問われている。

次に評価において問題になるのは、その方法であるが、一般的に想定されているのは、ペーパーテストである。しかし、梶田（1994）が示すよう

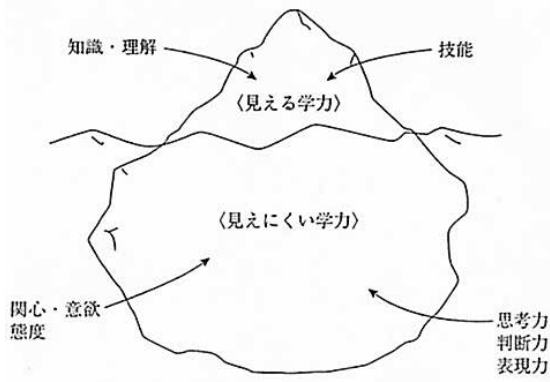


図2 学力の冰山モデル（梶田，1994）

に学力には「知識・理解」「技能」のように捉えやすいものと「関心・意欲・態度」「思考力・判断力・表現力」のように捉えにくいものがあり、従来型の客観的なペーパーテストでは前者の学力測定に偏っていることが多くの先行研究で指摘されている。

国立教育政策研究所（2019c）は、新学習指導要領に対応した学習評価の指針を示すべく「学習評価の在り方ハンドブック」を発行している。ここでは、観点別学習状況の評価として「知識・技能」、「思考・判断・表現」、「主体的に学習に取り組む態度」の3観点に整理し、それぞれの具体的な評価の方法について示している。「知識・理解」については、事実的な知識の習得を問う問題と、知識の概念的な理解を問う問題とのバランスに配慮したペーパーテストの工夫改善が示されている一方で「思考・判断・表現」については、ペーパーテストのみならず、論述やレポート、発表、グループでの話し合い、作品の制作や表現等の多様な活動を取り入れたり、それらを集めたポートフォリオ

を活用したりする必要性が述べられ、「主体的に学習に取り組む態度」については、ペーパーテストが一切触れられておらず、ノートやレポート等における記述、授業中の発言、教師による行動観察や、児童生徒による自己評価・相互評価の状況の考慮など、多様な方法が示されている。ここから、文部科学省は評価方法としてのペーパーテストの可能性について、「知識・技能」と「思考・判断・表現」の一部として捉えていることがわかる。

一方、西岡・田中（2009）は、評価方法について「単純・複雑」「筆記・実演」の2つの軸でまとめ、それぞれの評価方法の位置づけを明確化している。

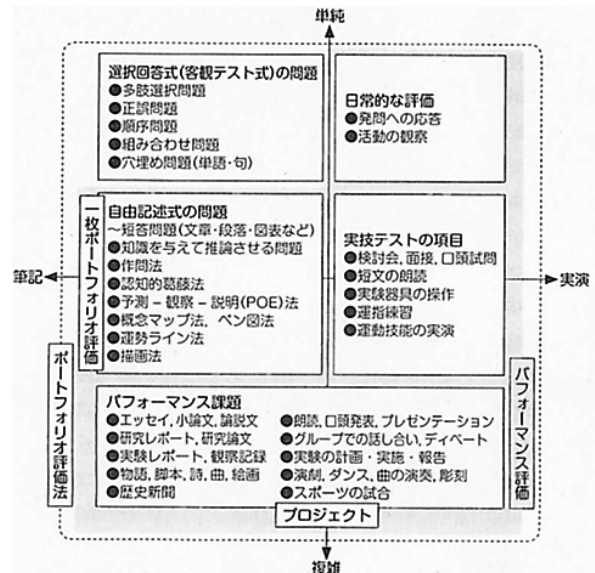


図3 さまざまな評価方法（西岡・田中，2009）

ペーパーテストに限定すると、左側の「筆記」の3項目が該当する。上段の選択回答式（客観テスト式）の問題は、センター試験のようなマーク式が該当する。自由記述式の問題については、一般的な短答問題をはじめ、推論させる問題や作問法等がある。下段のパフォーマンス課題は、説明文やレポート等、文章化が必要な評価方法となっている。

「知識・技能」については、選択回答式の問題である程度捉えられるかもしれないが、「思考・

判断・表現」といった学力を評価するには、自由記述式の問題を用いることが求められる。

また、テストのために特別に設定された状況ではなく、現実の状況を模写したりシミュレーションしたりしながら評価することの重要性を強調する「真正の評価」という立場も出てきている。これは、学校で育成する資質・能力の要素の全体を捉えようとする評価でパフォーマンス評価がそれにあたる。

このように評価方法は、「確かな学力」として整理された幅広い資質・能力を捉えるために、多様で複雑なものになってきている。ペーパーテストにおいては、「知識・技能」の評価が主とされるが、どこまでの学力が測定可能なのか検討する必要がある。

3. 算数科の学力評価におけるペーパーテスト

算数科においては、今回の学習指導要領改訂によって、表3のように3つの観点別評価規準が示されている。

表3 算数科の観点別評価規準

	趣旨
知識・技能	・数量や図形などについての基礎的・基本的な概念や性質などを理解している。 ・日常の事象を数理的に処理する技能を身に付けている。
思考・判断・表現	日常の事象を数理的に捉え、見通しをもち筋道を立てて考察する力、基礎的・基本的な数量や図形の性質などを見だし統合的・発展的に考察する力、数学的な表現を用いて事象を簡潔・明瞭・的確に表したり目的に応じて柔軟に表したりする力を身に付けている。
主体的に学習に取り組む態度	数学的活動の楽しさや数学のよさに気づき粘り強く考えたり、学習を振り返ってよりよく問題解決しようとしたり、算数で学んだことを生活や学習に活用しようとしたりしている。

これらの資質・能力をペーパーテストにおいて測定しようとする際、ペーパーテストの問題の形

式や質が問われる。ここではまず、算数科のペーパーテストで扱われている問題について分類し、その上で、現在我が国の小学校の現場で用いられている問題を分析する。

全国学力・学習状況調査においては、算数科の問題形式として選択式、短答式、記述式の3つが採用されている。それに対して、我が国で古くから数学技能の検定を行ってきた日本数学検定協会(2020)は、記述式について論理読解力を問う問題と論理構成力を問う問題に分けて、それぞれを次のように説明している。

- ・選択式... 複数の選択肢から正しいものを選択する。
- ・短答式... 数値や用語など主として単語で答える。
- ・記述式... 事柄について文や図・表・グラフなどで説明する。
 1. 論理読解力：解法が示された問題に対してその筋書きを読み取って答える力
 2. 論理構成力：解法が示されていない問題に対して自分で筋書きをつくって答える力（論述式）

算数の教科書にある問題の多くは、短答式になっており、特に数と計算領域では、式と答えを解答する形式が大多数を占めている。そしてそれらは、知識・技能を問うものとなっている。一方で選択式や記述式は単元末の練習問題で数問ある程度であるが、授業の中で児童に選択させたり、記述させたりすることは頻繁にされており、そこで思考力・判断力・表現力を高めることが目指されている。

次に問題の質的な側面を整理するためにTIMSSとPISAの分類を参照する。TIMSS2003によれば、評価の枠組みとして内容領域と認知的領域の二次元で整理している。内容領域は「数」「代数」「測定」「幾何」「資料の表現・分析」「確率」となっており、日本の算数の領域と大きな違いは見られない。認知的領域としては、以下の表4のように4つのカテゴリーで分類している。

表4 TIMSS2003の認知的領域

カテゴリー	下位項目 (具体的な行動)
事実や手順を知る こと 〔「知る」〕	思い出す, (等しい・等しくないなどが) わかる, 計算する, 道具を使う
概念を用いる こと 〔「用いる」〕	(概念や関係を) 知る, 分類する, 表す, 式で表す, 区別する
決まりきった問題 を解くこと 〔「解く」〕	(問題を解く方法を) 選ぶ, (適切なモデルを) 作る, (数学的モデルを) 解釈する, (問題を解くのに事実や手順や概念を) 応用する, (解の正しさを) 検証する
推論を行うこと 〔「推論」〕	適切な推測を行う, 分析する, (数学的アイデアを) 評価する, 一般化する, (新しい知識を既知の知識と) 関連付ける, (異なる手順や結果を) 結び付ける, 決まりきった解き方のない問題を解く, (行為や説明が) 正しいことの証拠を示す

(石井, 2008)

このカテゴリーと内容領域をクロスさせた形で問題が分類されている。

表5 内容領域と認知的領域のクロス表

		認知的領域			
		知る	用いる	解く	推論
内容領域	数				
	代数(きまりと関係)				
	測定				
	幾何				
	資料の表現・分析, 確率				

(石井, 2008)

「知る」「用いる」については、日本の教科書レベルの問題として、多く見られるものであるが「解く」「推論」については、解釈や分析が必要となり、難易度が高いものとなっている。

一方、PISA2012においては、15歳児の数学的リテラシーを捉えるため「数学的な内容」「数学的プロセス」「数学が用いられる状況」の3つの側面から作成されている。「数学的な内容」につ

いては、「空間と形」「変化と関係」「量」「不確実性とデータ」となっており、我が国の数学で学習する内容と親和性が高い。生徒はそれらの内容について、実世界の文脈に基づく問題に取り組み、数学的な能力を用いることで解決する。この能力とは、思考と推論、論証、コミュニケーション、モデル化、問題設定と問題解決、表現、記号による式や公式を用い演算を行うこと、テクノロジーを含むツールを用いることといった8つの能力が関わっている。問題に取り組むときには、これらの能力が同時に機能し、複雑に絡み合うが、この中の1つか2つの能力が特に顕著に関わっている。これら8つの能力を含む認知的活動は、表6のように「定式化」「適用」「解釈」の3つの数学的プロセスで構成されている。

表6 PISAの数学的プロセス

カテゴリー	数学的プロセス
①定式化	数学を応用し、使う機会を特定することを含めて、提示された問題や課題を数学によって理解し、解決することができること
②適用	数学的に推論し、数学的概念・手順・事実・ツールを使って数学的に問題を解決すること
③解釈	数学的な解答や結果を検討し、問題の文脈の中でそれらを解釈すること

国立教育政策研究所 (2013) を基に筆者作成

これは、PISAが表現している数学化サイクルで、現実世界の問題を数学を用いて定式化し、その数学的問題を解いた上で解答を現実の状況に照らして解釈するというプロセスとなっている。そして、このプロセスで用いる能力については、直面する問題状況の質に応じて3つのまとまりに分類して捉えようとしている (表7)。

このように様々な能力を捉えるために問題状況の質を分類して出題していることがわかる。そして「数学が用いられる状況」については、表8のように「私的」「職業的」「社会的」「科学的」の4つのカテゴリーで構成され、PISAで出題される問題文には幅広い文脈が設定されていることが

表7 PISAの問題状況の質

クラスター	認知的な質
再現	よく見慣れた場面で練習した知識を再現すれば決まりきった手順で解ける問題を解く能力
関連付け	どちらかといえば見慣れた場面だが手順があまり決まっていない、したがって、いろいろな表現や場面を関連付けることが必要になる問題を解く能力
熟考	あまり見慣れない場面で、問題場面の洞察、解法の熟考、問題解決過程における反省的思考、そして、結果の一般化や正当化が要求される問題を解く能力

石井（2008）を基に筆者作成

わかる。

表8 PISAの数学が用いられる状況

カテゴリー	数学が用いられる状況
①私的	生徒の日々の活動に直接関係する文脈
②職業的	職業の場面に現れるような文脈
③社会的	生徒が生活する地域社会における文脈
④科学的	より抽象的な文脈で、技術的な過程、理論的な場面、明らかに数学的な問題についての理解に関連する

国立教育政策研究所（2013）を基に筆者作成

算数・数学におけるペーパーテストでは、大きく問題形式による分類、内容の分類、質的な分類、プロセスの分類、状況の分類があることが明らかとなった。これらの分類を用いて、現在の小学校で広く実施されている各ペーパーテストについて分析していく。

4. 各ペーパーテストの問題の分析

本稿では、児童の学力を測定するペーパーテストとして、現在我が国の小学校で用いられている業者テスト、標準式学力テスト、全国学力・学習状況調査の問題を取り上げて分析する。その際、前節で整理した各分類を参照して考察する。ただし、業者テスト及び標準学力テストについては、

民間業者が商業的に作成しているものであるため、具体的な問題例を示して分析することは避ける。

(1) 業者テスト

業者テストとは、小学校でよく用いられている出版社などが作成したテストのことである。文溪堂、光文書院、青葉出版、日本標準等、複数の専門業者から発行されており、どの地域でも使えるように標準化されている。その特徴として挙げられるのが、表9の点である。

表9 業者テストの特徴

	特徴
実施時期	単元の学習が終了した後に実施することを想定している
内容	教科書の単元に沿った問題となっており、逸脱した内容がほとんど含まれていない
形式	問題形式は短答式、次いで選択式が多く、客観的な評価が難しい記述式の問題は少ない
難易度	各問題の難易度に大きな差はなく、設定されている目標点や平均点が80点程度になっている
配分	観点別評価の各項目「知識・技能」「思考・判断・表現」「主体的に学習に取り組む態度」ごとに問題が名目上配分されている

業者テストは、客観的な評価ができるようになっており、観点別に「知識・技能」「思考・判断・表現」「主体的に学習に取り組む態度」を測定する短答式の問題が多く設定され、教師によって採点に相違がないように作成されている。

一方で問題の質をPISAの認知的な質に照らして分類すると、殆どの問題が「再現クラスター」（よく見慣れた場面で練習した知識を再現すれば決まりきった手順で解ける問題）に該当しており、その点を踏まえると、「知識・技能」の学力を測定しているに過ぎないといった批判的な見方もできる。児童の学力を真に測定できているかは不透

明であり、それは、教師の評価行為をエスノグラフィを用いて分析した佐藤（1996）も指摘している。佐藤（1996）は、業者テストの評価基準の妥当性を問題視している教師の姿をもとに、学校の制度的・実践的環境の制約等の理由があることを明らかにしている。

選抜するための学力評価をほとんど行う必要がない小学校では、児童の算数に対する情意的な側面を低下させないためにも、現在の業者テストの形も意義はある。しかしながら、「知識・技能」以外の学力を真に測定できていないことや、PISAの認知的な質として再現クラスター以外のあまり見慣れない場面における問題がないこと、単元の学習直後に行うことから、短時間で学習した内容の習得は把握できるものの長期的な学力の定着を測定できるようになっていないことなど、限定的なものとなっている。そのため、現在我が国が育成することを目指している生きる力を核とした新しい学力観からは、大きくかけ離れていると言わざるを得ないのが現状である。

(2) 標準式学力テスト

標準式学力テストとは、業者テストと同じように民間業者が作成する学力テストで、学習指導要領に準拠した形で数社が出版し実施している。以下表10がその特徴である。

業者テストとの大きな違いは、年に1回の実施であるため、長期的な視点で学力を測定できる点が挙げられる。採点は実施後に業者が一括して行うことから、問題形式は、選択式が大部分であり、客観的な評価ができるようになっている。一部は記述式で、「思考・判断・表現」を評価することを想定しており、問題の工夫において、「実生活に関する問題場面を増やしたこと」「言語活動を通じて評価する問題を採用したこと」の2点が挙げられている。しかしながら、採点者によって採点に相違がないような文章量で設定されており、日本数学検定協会が示している論理構成力を捉えるまでには至っていない。

一方、問題の質をPISAの認知的な質に照らし

表10 標準式学力テストの特徴

	特徴
実施時期	学年末や2学期終了時など、年に1回、それまでの学習状況を把握するために実施することを想定している
内容	教科書の学習内容が基本だが、一部、教科書では見られない問題もある
形式	問題形式は選択式が多く、一部客観的な評価が難しい記述式の問題も設定されている
難易度	問題によって難易度に差があり、児童の学力によって正誤のばらつきが大きい
配分	観点別評価の各項目「知識・技能」「思考・判断・表現」「主体的に学習に取り組む態度」ごとに問題が名目上配分されている

て分類すると、多くの問題が「再現クラスター」に該当しているが、一部で「関連づけクラスター」や「熟考クラスター」に至る問題状況があった。「思考・判断・表現」の学力を測定している問題と設定されているが、その実態は限定的なものとなっていると言わざるを得ないのが現状である。

(3) 全国学力・学習状況調査

PISAの結果や学力低下についての世間の注目などを受け、文部科学省は2007年から全国学力・学習状況調査を行っている。この調査の目的は、義務教育の機会均等とその水準の維持向上の観点から、全国的な児童生徒の学力や学習状況を把握・分析し、教育施策の成果と課題を検証し、その改善を図ること、学校における児童生徒への教

表11 調査問題作成の基本理念

主として「知識」に関する問題 (A)	身に付けておかなければ後の学年等の学習内容に影響を及ぼす内容や実生活において不可欠であり、常に活用できるようになっていることが望ましい知識・技能など
主として「活用」に関する問題 (B)	知識・技能等を実生活の様々な場面に活用する力や様々な課題解決のための構想を立て、実践し、評価・改善する力などに関わる内容

(国立教育政策研究所, 2018)

育指導の充実や学習状況の改善等に役立てること、そのような取組を通じて、教育に関する継続的な検証改善サイクルを確立することとされている。算数科においては、2018年まで調査問題作成の基本理念を以下の2つに整理していた。

知識の問題については、学習指導要領に基づき、第5学年までに身に付けておくべきものを焦点化し、「知識・理解」「技能」に関わるものが出題されている。他方で活用の問題については、「数学的な考え方」に関わる問題を出題しており、以下、表12の4つの観点を踏まえて作成されている。

表12 「活用」の問題と4つの観点

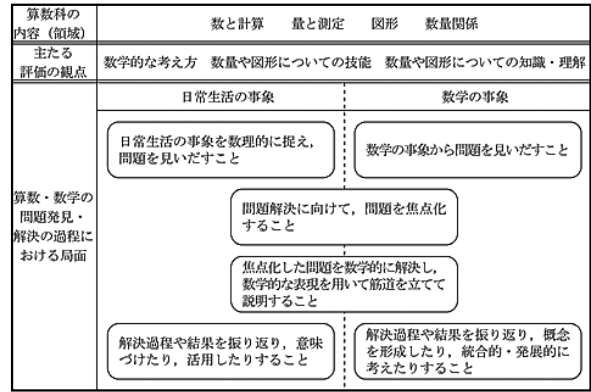
物事を数・量・図形などに着目して観察し的確に捉えること	
与えられた情報を分類整理したり必要なものを適切に選択したりすること	
筋道を立てて考えたり振り返って考えたりすること	筋道を立てて考えること 振り返って考えること
事象を数学的に解釈したり自分の考えを数学的に表現したりすること	事象を数学的に解釈すること 自分の考えを数学的に表現すること

(国立教育政策研究所, 2018)

ここでは、観察、選択、思考、解釈、表現などの問題解決のプロセスが示され、PISAの「定式化」「適用」「解釈」といった数学のプロセスが内包している。そして、問題の状況としては、算数科固有のもの、他教科等の学習に関わるもの、日常生活に関わるものを基に作成されている。

一方、2019年からは、新学習指導要領の考え方を受けて、「知識」「活用」の分類が一本化されているが、その内容や質的な側面に大きな違いはない。学力の3要素は相互に関連しているため、「知識」「活用」をバラバラに捉えないようにする新しい学力観を明確に打ち出した形と捉えることができる。そして、調査問題の枠組みを表13のように提示している。

表13 調査問題の枠組み



(国立教育政策研究所, 2019)

新学習指導要領の考え方を考慮して、「算数科の内容(領域)」「主たる教科の観点」「算数・数学の問題発見・解決の過程における局面」の3つの視点で問題が作成されている。

以下、全国学力・学習状況調査の特徴は表14のとおりである。

表14 全国学力・学習状況調査の特徴

	特徴
実施時期	第6学年の4月に全国で一斉に実施されている
内容	教科書では見られない問題が多い
形式	問題形式は選択式5問、短答式5問、記述式4問となっており、それぞれ同程度の割合で出題されている
難易度	「知識・技能」に関する正答率が高い一方で、記述式の正答率が低いなど、難易度に差がある
配分	観点別評価の各項目では、「知識・技能」が6問、「思考・判断・表現」が8問、「主体的に学習に取り組む態度」が0問となっている

問題形式は記述式も一定数あり、記述する事柄としては「事実」「方法」「理由」の3つが設定されている。このような記述問題は、算数科の各種テストや問題集においてあまり出題されていないため、児童にとって見慣れない問題と考えられ、正答率が選択式や短答式よりも低くなっている。

また、教科書等ではあまり見られない日常生活

場面を取り上げた問題が多く、「関連づけクラスター」や「熟考クラスター」に至る問題状況が確認できる。その状況を説明するために問題文が長文となっており、解答者には問題場面の把握が求められている。

具体的な問題例として、平成31年度実施の問題から1問を取り上げて分析する。取り上げる問題④(3)は図4のものであり、場面の状況から、単位量当たりの大きさを基に、求め方と答えを記述し、その結果から判断する問題である。報告書によってまとめられた正答率は、62.8%であり、正答の条件や正答例は以下の通りとなっている。

この問題は、4割程度の児童が不正解となっており、報告書の中でも課題として挙げられているものである。技能面では、1位数の乗除ができれば解くことができる問題であり、解答も2つの演算を踏まえて判断する「方法」を記述するもので、

表15 正答条件及び正答例


正答の条件	次の①、②の全てを書き、答えを21と書いて、番号を1と選んでいる。 ① 1ポール分当たりにかかる時間を求める式や言葉 ② 7ポール分進むのにかかる時間を求める式や言葉
正答例	$9 \div 3 = 3$ で、1ポール分には3分間かかります。残り7ポール分なので、 $3 \times 7 = 21$ で、21分間かかります。

(国立教育政策研究所, 2019)

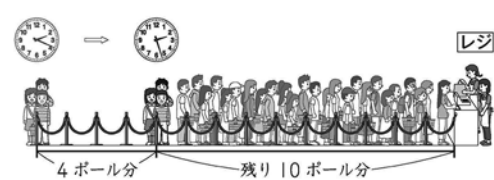
それほど表現力が問われているわけではない。しかしながら、示された場面の状況を捉えることができている児童の誤答が報告されている。

全国学力・学習状況調査は、業者テストや標準式学力テストとは異なり、「思考・判断・表現」について、ある程度対応しているように考えられるが、以下の3点についての限界が指摘できる。

(3) はるとさんたちは、限定商品を買いたいと思っています。次の予定があるので、午後3時までにはレジに着きたいと考えています。
列に並ぶと、レジまでは14ポール分ありました。ポールとポールの間の長さはどこも同じです。



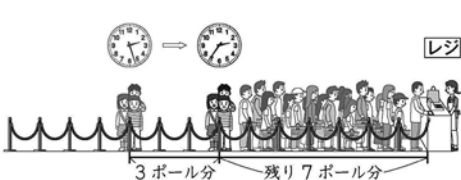
はるとさんたちが並んでから、4ポール分進むのに8分間かかり、残り10ポール分になりました。午後3時までは、残り33分間です。そこで、33分間以内にレジに着くことができるかどうかを考えてみました。



はると

4ポール分進むのに8分間かかったことから、残り10ポール分も同じ進みぐあいでも進むとして考えます。
 $8 \div 4 = 2$ で、1ポール分には2分間かかります。
 残り10ポール分なので、 $2 \times 10 = 20$ で、20分間かかります。
 だから、33分間以内にレジに着くことができます。

ところが、レジにいる店員さんが減ってしまいました。それからは、3ポール分進むのに9分間かかり、残り7ポール分になりました。午後3時までは、残り24分間です。
そこで、はるとさんたちは、24分間以内にレジに着くことができるかどうかを、もう一度考えてみました。



あかり

3ポール分進むのに9分間かかったことから、残り7ポール分も同じ進みぐあいでも進むとして考えます。

3ポール分進むのに9分間かかる進みぐあいでも進むとすると、残り7ポール分進むのにかかる時間は何分間ですか。
 求め方を言葉や式を使って書きましょう。また、答えも書きましょう。
 さらに、24分間以内にレジに着くことができるかどうかを、下の**1**と**2**から選んで、その番号を書きましょう。

1 着くことができる。
2 着くことができない。

図4 平成31年度全国学力・学習状況調査 小学校算数 問題④(3) (国立教育政策研究所, 2019)

①読解力

図4の問題例に代表されるように、全国学力・学習状況調査の問題は、1題当たりの問題文の文章量が長い傾向にある。この問題だけで636字あり、一般的な大人でも2分程度かかる文章量である。「関連づけクラスター」や「熟考クラスター」に至る現実的な問題状況を設定するためには、どうしても詳細な文脈の説明が不可欠となる。これらの問題には、算数科の評価の観点が設定されているが、読解力が影響している可能性も否定できない。それは、新井(2018)が、大学生に数学基本調査を行った結果を踏まえ、読解力が数学の正答に結びついていることを主張していることから十分に考えられる。そこでは、問題文を理解する過程である変換過程(問題文を読み、一文ごとの内容を理解し、文単位の表象を形成する過程)や統合過程(一文一文を関係付けて、問題全体としての表象を形成する過程)でつまづいている児童が多く存在すると推測される。

②思考力・判断力・表現力の深さ

全国学力・学習状況調査は、業者テストや標準式学力テストとは異なり、「思考・判断・表現」の評価についてかなり考慮されていると言える。しかしながら、ペーパーテストという性質上、結果のみに焦点化されている点が否めない。もちろん、数学化サイクルにおいて、定式化、適用、解釈のプロセスは行っているものの、採点では正誤の二分法で評価する以外に、プロセスの一部しか見ることができていない。

これは、清野(2015)が全国学力・学習状況調査の課題として挙げている思考力・判断力・表現力の「深さ」を測定することができていないという指摘とも関連する。評価する時期や対象等が事前に設定されたかなり限定的なものとなっている。また、評価者間の信頼性を担保するために記述量も限定されており、採点コードに従って評価さざるを得ない等、思考力・判断力・表現力の一部しか捉えられていないのが現状である。

③ICTを含むツールの活用

算数の授業では、具体的な教具を用いて問題解決を行うが、ペーパーテストでは基本的に筆記用具しか用いられないケースが多い。OECD-DeSeCoが提案したキー・コンピテンシーでは道具を相互作用的に用いる能力を掲げているように、ICTを含むツールの活用は、現在求められている新しい能力である。そのため、清野(2015)が指摘するように思考の道具が整備された状況下での「思考力・判断力・表現力」の評価問題の設定が求められていると言える。

5. ペーパーテストの可能性についての考察

ここまで、ペーパーテストの限界について述べてきた。現在、この限界を一部解消するものとして注目されているのが、「パフォーマンス評価」である。パフォーマンス評価とは、文字通り、パフォーマンスによる評価で、「ある特定の文脈のもとで、様々な知識や技能などを用いて行われる人のふるまいや作品を、直接的に評価する方法」のことである(松下, 2007)。そのためにパフォーマンス課題を与えて解決・遂行させ、それを複数の評価者がルーブリックと称した評価基準表を用いながら評価するものである。パフォーマンス課題の一例は図5のとおりである。

そして、パフォーマンス課題の要件として次の4つを挙げている。

《(a)思考のプロセスを表現することを要求する

子ども会でハイキングに行ったところ、ある地点でコースが二手に分かれていました。さつきコースが全長3 kmで、けやきコースは全長5 kmです。どちらのコースをとってもレストハウスへ行けます。そこで2つのグループに分かれて、レストハウスで合流することになりました。ゆう子さんのグループは、さつきコースにしました。あきお君のグループはけやきコースにしました。

10時に二手に分かれて、ゆう子さんのグループがレストハウスについてのは11時でした。その時、あきお君たちのグループはまだ到着していませんでした。「距離が長いから当然だね。あきお君たちが着くまでのくらの時間がかかるのかはかってみよう。」ということで、時間をはかっていたら、30分後にあきお君のグループがレストハウスに到着しました。ゆう子さんはあきお君に「どこかで休憩していたの？」と聞きました。あきお君は「休憩なんかしてないよ。ずっと歩いてたんだよ」と答えました。どちらのグループも休憩したりせず、一定の速さで歩いていました。

そこで、みんなはどちらのグループのほうが早く歩いたの知りたくなりました。あなたは、どちらが早く歩いたと思いますか。考えたこととその理由を書いてください。

図5 パフォーマンス課題の例(松下, 2007)

- (b)多様な表現方法（式、言葉、図、絵など）が使える
- (c)真実味のある現実世界の場面を扱っていて、そこから数学化するプロセスを含んでいる
- (d)複数の解法がとれる》

松下（2007）によれば、一般的な標準テストでは測りにくい質の学力を多面的・総合的に把握することが目的とされている。現在、我が国で行われているふつうのテストとの違いについて、表16のようにまとめている。

表16 ふつうのテストとパフォーマンス評価の比較

	ふつうのテスト	パフォーマンス評価
評価手段	ペーパーテスト	具体的活動、ペーパーテスト
解答形式	択一式、簡単な記述式	自由記述式（ペーパーテスト）
問題数	多い	少ない
テストの性格	スピードテスト的性格	パワーテスト的性格
評価尺度	比較的限定された学力を、一元的な尺度で、評価する（正誤の二分法）	複合的な学力を、多次元的な尺度で、評価する（複数のレベル）
評価観点の設定時期	どんな学力をみる問題化をあらかじめ決めておく（採点コードを用いる）	どんな学力が発揮されているかは、事後的に個人ごとに明らかになる（採点と同時並行で作成したルーブリックを用いる）

（松下，2007）

パフォーマンス評価は具体的な活動を評価対象とする一方で、ペーパーテストでも対応できるとされている。その際、全国学力・学習状況調査などとは違い、自由記述式である必要がある。そして、パフォーマンス評価の最も大きな特徴として複合的な学力を、多次元的な尺度で、評価できることが挙げられる。

もちろん、ペーパーテストにおいても多様な問

題を何題も作成し、複合的な学力を測定しようとすることも不可能ではない。しかし、鈴木(2020)は 労力や妥当性の観点からパフォーマンス評価導入の必要性を指摘している。思考・判断・表現の評価は、正誤の二分法的評価ではできないため、多次元的な評価基準を作成して解答を判定していくことになる。

ただし、パフォーマンス評価にはデメリットもあり、松下2007は以下の4点を指摘している。

- 《①多くの時間と労力を要する
- ②課題数が制限される
- ③客観テストに比べると「信頼性」が低いこと
- ④ルーブリック作りが難しいこと》

パフォーマンス課題は、問題例でも挙げたように現実世界の場面を扱っているため、作成するのに時間や労力がかかる。そのため、出題する課題数も少なくせざるを得ない。そして、評価する際には、解答が多様な表現方法でかつプロセスを含んだ記述式のため、例え質的な評価基準を設けたとしても評価者によるぶれが生じ、信頼性が低いと指摘されることもある。また、ルーブリックにある評価基準の記述は一般的な表現になり、具体性に欠くため、判断に迷うことも想定される。そのため、事前に各段階を代表する典型的な記述例を揃えておき、適宜これを参照することで、評価者間の信頼性を高めるなどの対策も必要となる。また、複数の評価者間で採点結果の調整(モデレーション)をするなどの工夫も求められている。

6. おわりに

本稿では、児童の学力を捉えるための評価方法の一つであるペーパーテストに焦点を当て、現在我が国の小学校で用いられている業者テスト、標準学力テスト、全国学力・学習状況調査を取り上げて分析した。旧来、ペーパーテストは、学習に向かわせるための動機づけや、学習結果を把握する資料として用いられ、信頼性や妥当性が高い

評価方法である。それは、「知識・技能」を測定するためには問題なく使用でき、便利なものであるが、「思考・判断・表現」を測定する上で、テストによって評価できる範囲の限界があることが明らかとなった。特に、全国・学力学習状況調査においては、現実的な問題状況を設定するための詳細な文脈の説明から読解力の影響が考えられること、正誤の二分法で評価することによる思考力・判断力・表現力の深淺さを判断できないこと、ICTを含むツールの活用ができないといった制限があることが課題として挙げられる。

一方、ペーパーテストによるパフォーマンス評価は、「思考・判断・表現」を測定する上で、現状では最も有効なものである。しかしながら、時間と労力、評価の信頼性といった問題点もあり、導入へのハードルが高い状況となっている。今後、教師の負担を減らしていこうとしている社会的な背景を鑑みると、パフォーマンス課題やルーブリック、評価事例集については、組織的な作成を進めていくことが求められているといえる。

本稿では、各種テストの問題を分析したにすぎず、児童を対象にテストを実施して実証的に検討することが今後の課題として挙げられる。

引用文献

- 新井紀子(2018).『AI vs. 教科書が読めない子どもたち』. 東洋経済新報社.
- 石井英真(2008).「算数・数学の学力と数学的リテラシー」. 田中耕治編著『新しい学力テストを読み解く』. 日本標準.
- 石井英真(2012).「学力向上」. 篠原清昭編著『学校改善マネジメント』. ミネルヴァ書房.
- 石井英真(2015).『今求められる学力と学びとは—コンピテンシー・ベースのカリキュラムの光と影—』. 日本標準.
- 鹿毛雅治(2000).「学びの場で経験される評価：豊かな学びが生まれるために」. 長尾彰夫, 浜田寿美男編『教育評価を考える：抜本的改革への提言』. ミネルヴァ書房.
- 梶田叡一(1994).『教育における評価の理論Ⅰ 学力観・評価観の転換』. 金子書房.
- 工藤文三(2011).「高等学校における学習評価の実態把

- 握と改善に関する研究」. 科学研究費補助金基盤研究(C)研究成果報告書.
- 国立教育政策研究所(2018).『平成30年度全国学力・学習状況調査解説資料小学校算数』.
- 国立教育政策研究所(2019a).『平成31年度全国学力・学習状況調査解説資料小学校算数』.
- 国立教育政策研究所(2019b).『平成31年度全国学力・学習状況調査報告書小学校算数』.
- 国立教育政策研究所(2019c).『学習評価の在り方ハンドブック 小・中学校編』.
- 佐藤真理子(1996).「教室の評価行為のエスノグラフィ—小学校における「業者テスト」の存続メカニズム」.『東京大学大学院教育学研究科紀要』36.
- 鈴木秀幸(2020).「ここまでは押さえたい学習評価1」. 日本図書文化協会『指導と評価 4月号』.
- 清野辰彦(2014).「算数・数学科における「思考・判断・表現」を評価するための活用型評価問題の開発」. 日本教材文化研究財団調査研究シリーズ『算数・数学科における「思考・判断・表現」の評価に関する研究』.
- 西岡加名恵・田中耕治編著(2009).『「活用する力」を育てる授業と評価 中学校—パフォーマンス課題とルーブリックの提案』. 学事出版
- 日本数学検定協会(2020).「記述式について」. www.sugaku.net/common/pdf/kijutsushiki2.pdf (最終参照日2020年3月31日).
- 松下佳代(2007).『パフォーマンス評価—子どもの思考と表現を評価する—』. 日本標準.
- 松下佳代(2010).『〈新しい能力〉は教育を変えるか—学力・リテラシー・コンピテンシー—』. ミネルヴァ書房.
- 文部科学省(2016).『幼稚園, 小学校, 中学校, 高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について(答申)』. 中央教育審議会.
- 文部科学省(2017a). 小学校学習指導要領(平成29年告示).
- 文部科学省(2017b). 小学校学習指導要領(平成29年告示)解説算数編.
- 文部科学省(2017c). 小学校学習指導要領(平成29年告示)解説総則編.

(函館校准教授)