

小学校外国語科の「話すこと(発表)」における「思考・判断・表現」の評価に関する研究

～比較可能性に基づいたより信頼性の高いルーブリックの開発を目指して～

さいたま市立下落合小学校 教諭 有江 聖

I はじめに

令和2年度から小学校での新学習指導要領の全面実施がされた。高学年では外国語活動が教科化され、他教科同様「知識・技能」(以下、「知・技」)「思考・判断・表現」(以下、「思・判・表」)「主体的に学習に取り組む態度」(以下、「主体性」)の3観点で評価を行うことが決められた。それに伴い、評価に関する問題は現在も多岐に渡っている。

II 評価を取り巻く現状

中教審(2009)によると、小・中・高の教員対象に行った観点別評価の実施状況調査において「思考・判断」は最も「円滑に把握ができていない」項目として回答が多いことが分かった(図1)。

(1)「思考・判断」に関する評価が難しい

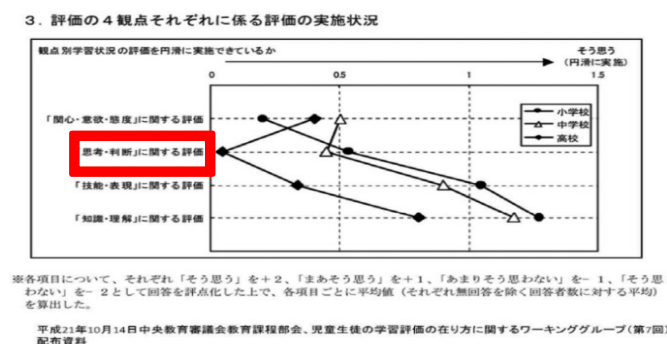


図1 小・中・高教員対象の観点別評価実施状況調査

III 小学校外国語科における評価の現状

3-1 先行研究

株式会社イーオンが小学校教員に行った「小学校の英語教育に関する教員意識調査2021」では、「指導が難しいと感じている項目」に「話すこと」がトップに挙がり、小学校英語の教科化・早期化について、「予定どおり進めることが難しかった点・課題と感じた点」では、「児童の評価の仕方」がトップに挙げられた(図2)。

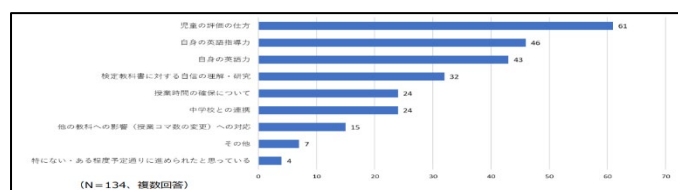


図2 予定どおり進めることが難しかった点・課題と感じた点

3-2 筆者による教員アンケート調査

筆者の勤務校で2021年7月に行った「グローバル・スタディ科(以下GS科)の指導と評価に関するアンケート」(n=34)の結果の一部を掲載する(さいたま市では外国語科の代わりにGS科を設置しており、3年生から観点別評価が導入されている)。

まず、「GS科の評価で困り感を感じたことはありますか」という項目に対して9割以上の教員が「ある」と回答し、先行研究と同様の傾向を示した。次に、「評価の3観点の内容をどのくらい理解していますか」という項目については、「知・技」は「分かる」と回答をした教員が多い一方、「思・判・表」「主体性」の項目は「分からない」傾向の回答をした教員が多かった(表1)。

表1 3観点の評価に関する自己評価(単位:人)

	分かる	まあまあ分かる	あまり分からない	ほとんど分からない
知・技	4	26	3	1
思・判・表	0	5	20	9
主体性	0	7	20	7

これらの結果から、小学校教員は外国語科の「話すこと」の「思考・判断・表現」の評価に特に強い不安を感じていると考えられる。しかし、児童の学習成果を把握し、指導改善を行うために評価は欠かせない。

IV 「思考・判断・表現」の評価について

西岡他(2201)は、「知識やスキルを課題や場面に応じて自分で考えて総合させて発揮する力(思考力、判断力、表現力等)の評価にはパフォーマンス評価(以下PT)が有効である」と述べている(p.113)。また、「ルーブリックは、思考力や判断力といった高次の目標を質的に把握する評価の方法論である」と述べている(p.116)。この理論を基に本実践では、「思・判・表」の測定のためにPT(プレゼンテーション)を実施し、ルーブリックを用いて評価を行う。

4-1 ルーブリックの分類と記述形式

小泉(2016)によるとルーブリックには以下のような種類があり、タスクを中心とする小学校外国語科においては「総合的・タスク特有型ルーブリック」がより適しているとされている(表2)。

表2 ルーブリックの分類(小泉(2016)より抜粋)

評価方法による分類	
(1) 総合 (holistic)	発話を全体的に評価、1つの得点を出す。
(2) 分析 (analytic)	発音、流暢さなど各規準で評価、複数の得点を出す。
測る能力の種類による分類	
(1) 一般型 (generic)	広く使用可能で、結果の一般化がより可能。一般的能力を図るのが意図。
(2) タスク特有型 (task-specific)	タスク形式ごとに作成。特定のタスクに限定される能力を測る。

また、ルーブリックの評価基準の記述形式についてダネル他(2014)は、最高到達点(A評価)のみを記載した「採点指針ルーブリック」と、各段階の到達点を記載した最も一般的な「3段階ルーブリック」があると述べている。それぞれの特徴や利点については著書を参考に以下のように整理できる(表3)。

表3 各ルーブリックの特徴(著書を基に筆者が表に整理)

採点指針ルーブリック	3段階ルーブリック
・学生の到達点を常に念頭に置いて採点に集中できる。 ・時間がかからない。 ・採点のブレが少ない。	・各段階の基準が詳細に記述されているため、作文や作品の評価に適している。 ・該当箇所を探すのに手間がかかる。

以上のことを踏まえると、プレゼンやスピーチのように短い時間の中で素早く正確な評価が求められる場面では、採点指針ルーブリックを使用することが適していると考えられる。

4－2 PTとルーブリックによる評価の問題点

鈴木(2019)は「ルーブリックによるPT評価は、一般に評価の妥当性は高いといわれている。(中略)逆に、評価者間信頼性の問題がある。異なる評価者で評価結果がかなり不一致になってしまうことである」(p.25)と述べている。これには、評価基準の曖昧さや評価者個人の先入観が関係している。さらに、宮本(2017)は、「パフォーマンス評価のような総括的評価は選抜の資料として用いられることもあるので、(中略)特に評価者間信頼性が高いことがまず必要である」と述べている(p.30)。

それらの解決の手立てとして近年、西岡(2021)では、信頼性を発展させた「比較可能性」とよばれる概念が提起されている。これは、複数の評価者間で評価規準・基準を共通理解し、同じ採点規則に従うことによって、評価の一貫性が確保されているかどうかを検

討する視点である。

外国語科ではPTを実施する機会が非常に多い。児童の到達度把握と指導改善のためには、PTの実施とルーブリックによる評価は欠かせない。一方で、小学校教員の評価への不安度の高さやルーブリックによる評価の信頼性の問題が存在している。信頼性の確保ができないまま、不安感の強い状態で評価をしても、本末転倒ではないだろうか。しかし、小学校外国語科におけるルーブリックの開発や評価の信頼性に関する先行研究は少なく、喫緊の課題である。

V 実践研究について

5－1 研究の目的と仮説

本研究は、これまでの経緯を踏まえ、小学校外国語科の「話すこと(発表)」における「思・判・表」のより信頼性の高いルーブックの開発を目的としている。本稿では小泉(2016)とダネル他(2014)の理論から、「総合的・タスク特有型採点指針ルーブリック」に高い信頼性があるという仮説の元、比較可能性の概念を用い、信頼性の高いルーブリックの開発を目指す。仮説は以下のとおりである。今回は「信頼性＝評価者間信頼性」と定義する。

仮説：外国語科の「話すこと(発表)」において「思考・判断・表現」の評価を比較可能性に基づき、総合的・タスク特有型採点指針ルーブリックを用いることで、より信頼性の高い評価が行えるだろう。

なお、本研究は令和3年度の10月～3月に勤務校の4年生(128名)を対象に行った実践をまとめたものである(筆者は6年生担任)。今回はG・S科の中の3つの単位における評価の実践について掲載する。

5－2 研究方法

仮説検証のために、鈴木(2019)等で言及されていた「評価者間信頼性」(複数教員による採点の一致度)を判断基準に用いる。今回は担任、ALT、筆者の3者による評価を行い、それをIBM SPSS Statisticsの「信頼性分析」にかけ、クロンバックの α 係数を算出する。 α 係数は1に近いほど評価者間信頼性の高さを示す。一般的に、0.7以上で中程度、0.8以上で高い信頼性、0.9以上で理想的な信頼性と判断される。なお、採点指針ルーブリックの信頼性の高さを検証するために、初めに3段階ルーブリックとの比較検証を行う。

5－3 研究の流れ

本実践は、4年生の「日本の美しい四季」「ふるさと再発見」「大きくなったら」の3単元で行い、研究の流れは比較可能性の概念に基づき、全ての単元を通じて図3のように統一した。

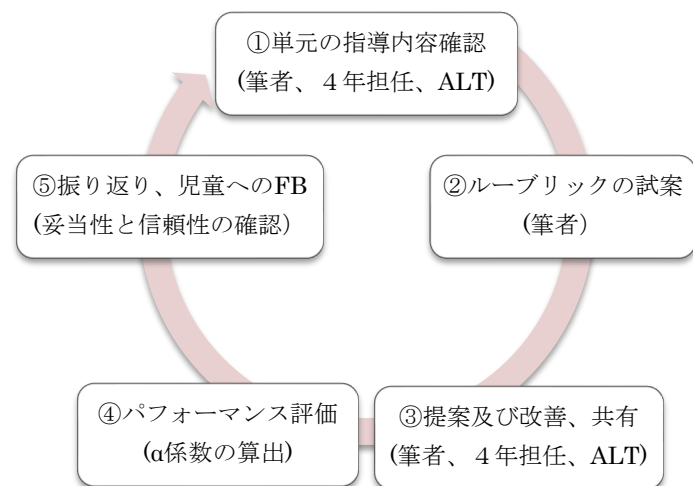


図3 研究の流れ（比較可能性に基づいて）

大前提として、評価のための活動にならないよう、単元導入前に該当単元の目標と指導内容の確認を4年生担任、筆者、ALTで行った。それを踏まえた上で、筆者が単元末プレゼンで使用する採点指針ループブリックと3段階ループブリックを試案する。それを4年生担任、ALTに提案及び改善し、到達点と採点基準の共通理解を図る。その後、指導計画の元で授業を進め、2クラスごとに違うループブリックを用いて単元末のプレゼン評価を行う。なお、担任とALTは授業内で、筆者はプレゼンを撮影したビデオを観て各クラスと同様のループブリックで評価を行う。そして、それぞれの評定を点数化し（A：3点、B：2点、C：1点）、「信頼性分析」にかけて α 係数を算出し、比較検証を行う。使用するループブリックは以下のとおりである（表4）。

表4 単元毎に使用するループブリックの種類

	1組	2組	3組	4組
日本の美しい四季	採点指針ループブリック		3段階ループブリック	
ふるさと再発見	3段階ループブリック		採点指針ループブリック	
大きくなったら	より評価者間信頼性の高いループブリックを採用			

これを2単元どちらのパターンも行った後に、2つのループブリックに関するアンケートを担当とALTに行う。 α 係数とアンケートの結果から、採点指針ループブリックの信頼性について考察する。なお、ループブリックは田村（2021）と瀧沢（2021）を参考に作成した。

5-4 「日本の美しい四季」での実践

まず、「日本の美しい四季」では、自分の好きな季節について理由等を交えながら説明でできるようになることが目標である。しかし、4年生では既習事項が少なく、表現の幅に限界がある。そこで既習事項に加え、内容に関するジェスチャーを「思・判・表」の項目として追加した。作成したループブリックは以下のとおりである（表5、6）。

表5 「日本の美しい四季」の採点指針ループブリック

プレゼンテーションループブリック	
①好きな季節を紹介するために、②becauseや③I like ～、I want to ～の表現を活用するとともに、④既習事項を用いた追加の説明や⑤内容に関するジェスチャーを加えるなどの工夫を通じてその季節の魅力詳しく説明している。	
下線部が4つ以上入っている	A
下線部が3つ入っている	B
下線部が1～2つ入っている。	C

表6 「日本の美しい四季」の3段階ループブリック

	プレゼンテーションループブリック
A	好きな季節を紹介するために、becauseやI like ～、I want to ～の表現を活用するとともに、既習事項を用いた追加の説明や内容に関するジェスチャーを加えるなどの工夫を通じてその季節の魅力詳しく説明している。
B	好きな季節について、becauseを用いて理由の説明をするとともに、I like ～やI want to ～などの既習表現を活用して、その季節の魅力詳しく説明している。
C	好きな季節について、becauseやI like ～やI want to ～などの既習表現を用いて魅力を説明することが難しい。

比較検証のために、A評価の記述を同様とし、BとCはAの記述から該当部分を抜いていく形で作成した。このループブリックと各クラスの名簿を記載した評価シートが以下のとおりである（図4）。

プレゼンテーションループブリック	
①好きな季節を紹介するために、②becauseや③I like ～、I want to ～の表現を活用するとともに、④既習事項を用いた追加の説明や⑤内容に関するジェスチャーを加えるなどの工夫を通じてその季節の魅力詳しく説明している。	
下線部が4つ以上入っている	A
下線部が3つ入っている	B
下線部が1～2つ入っている。	C

プレゼンの様子は
評価の際の注意点

クラス名簿

ループブリック

評価に悩んだ児童がいた場合は出席番号を○で囲み、

名前	評定	名前	評定	名前	評定
1 児童①	A	12 児童④	B	23 児童⑦	A
2 児童②	B	13 児童⑤	A	24 児童⑧	A
3 児童③	B	14 児童⑥	B	25 児童⑨	A

図4 PTで使用した評価シート（採点指針ループブリック）

5-5 ふるさと再発見での実践

次の「ふるさと再発見」は、埼玉県の名所について調べ学習を行い、行きたい理由や魅力についてプレゼンを行う。本単元では自分の考えや気持ちを表す表現を学習する。それを踏まえた上で以下のようなループブリックを作成し、評価を行った。今回は採点指針ループブリックのみを掲載する（表7）。

表7 「ふるさと再発見」の採点指針ループブリック

プレゼンテーションルーブリック	
① <u>埼玉県</u> の <u>施設や場所</u> のことをよく知ってもらうために、② <u>I want to go to～</u> や③ <u>because</u> を使って名所や理由を話すとともに、④ <u>追加の説明</u> や⑤ <u>自分の考えや気持ち</u> などを含めてより詳しく話している。	
下線部が4つ以上入っている	A
下線部が3つ入っている	B
下線部が1～2つ入っている。	C

5－6 評価者間信頼性の分析結果

2 単元での 3 者による評価をクラスごとに分析し、算出した α 係数が以下のとおりである(表 8)。どちらの単元でも採点指針ルーブリックを活用した 2 クラスで高い α 係数が算出された。

表 8 各単元におけるクラスごとの α 係数

	1 組	2 組	3 組	4 組
日本の美しい四季	採点指針		3 段階	
	.928	.754	.719	.741
ふるさと再発見	3 段階		採点指針	
	.484	.614	.819	.911

さらに、それぞれのルーブリックを使用した場合の評価者間信頼性について、詳細分析をかけた結果が表 9、10 のとおりである。今回は「ふるさと再発見」の 2 組（3 段階ルーブリック）と 4 組（採点指針ルーブリック）の結果を一部掲載する。

表 9 3 段階ルーブリック評価の詳細分析（2 組）

項目合計統計量			
項目が削除された場合の尺度の 平均値	項目が削除された場合の尺度の 分散	修正済み項目 合計相関	項目が削除された場合の Cronbach のアルファ
担任 2	4.89	.692	.581
ALT2	4.64	.683	.311
筆者 2	4.75	.713	.419

表 10 採点指針ルーブリック評価の詳細分析（4 組）

項目合計統計量			
項目が削除された場合の尺度の 平均値	項目が削除された場合の尺度の 分散	修正済み項目 合計相関	項目が削除された場合の Cronbach のアルファ
担任 4	4.47	.602	.902
ALT4	4.40	.593	.807
筆者 4	4.53	.740	.775

表の一番右にある「項目が削除された場合の Cronbach のアルファ」（太枠内）に注目してもらいたい。これはそれぞれの評価者の結果が分析に入らなかった場合、 α 係数がどのように変化するかを示す数値である。例えば、表 10 の「担任 2」による評価が削除された場合の α 係数は.321 となり、これは ALT と筆者の評価の一致度を表している。それを踏まえた上でこの項目を見ると、3 段階ルーブリックを使用した

2 組は数値にばらつきがあることが分かる(.321～.705)。一方で採点指針ルーブリックを使用した 4 組は、どの項目を削除しても 0.8 以上の α 係数を維持している(.802～.913)。つまり、採点指針ルーブリックを使用した場合、どの組み合わせで評価をしても高い評価者間信頼性を維持できる可能性を示唆している。

5－7 教員アンケートの結果

また、次に 4 年生担任 4 名と ALT に行った 2 つのルーブリックに関するアンケート調査の結果と自由記述の内容を以下に掲載する(表 11、12)。

表 11 2 つのルーブリックに関するアンケート

n=5 単位：人

	採点指針	3 段階
①評価がしやすい。	4	1
②時間がかからない。	5	0
③正確に評価できる。	4	1
④フィードバックがしやすい	4	1
⑤指導改善に繋げやすい	4	1
⑥ALT との連携が図りやすい	4	1

表 12 ルーブリックに関する自由記述（一部抜粋）

- 基本的に採点指針の方がやりやすかったです
が、最重要項目などの重みがついているとより
やりやすかった。
- 採点指針の方が正確で簡単。説明責任を果た
せ、児童へのフィードバックにも使える、所見
にも使いやすい。

アンケートの結果からも、採点指針ルーブリックは妥当性や信頼性、実施容易性の観点からも使用しやすいと考察される。以上のことから、比較可能性に基づいた採点指針ルーブリック評価には、高い信頼性があるという示唆を得た。ここままで仮説の立証にある程度の効果があったと考察できる。しかし、採点指針ルーブリックはどの項目ができているか分からず、担任によっては話しぶりなどを無意識に評価に考慮しており(図 5 下線部)、評価の信頼性に多少なりとも影響を与えている、また、表 8 から見て取れるように、採点指針ルーブリックを活用しても.754～.928 と α 係数にばらつきがあるなどの課題が明らかとなった。

1		B	12		話しぶり(下線部)が23		B
2			13		B	24	B
3		話しぶり(下線部)が23	14		B	25	R

図 5 話しぶりを評価に考慮している可能性のある部分(一部抜粋)

Ⅵ 安定した信頼性を実現するために

6-1 チェック式採点指針ループリックの開発

前述を踏まえた更なる信頼性獲得のため、ダネル他(2014)で言及されているチェックボックスを取り入れた採点指針ルーブリックを以下のように開発した(図6)。

評価基準	基準	評定
自分の夢を知ってもらうために、①なりた職業について②I want to be ~や③“because”を使って紹介するとともに、④I wantや⑤I like、⑥It is ~などの表現を活用して、より詳しく自分の考えや気持ちなどを含めて話している。	下線部が5つ以上入っている	A
	下線部が3~4つ入っている	B
	下線部が1~2つ入っている。	C

できた部分にチェックを付ける（知識・技能は必要に応じて使ってください。＊ジェ＝ジェスチャー）

		知識・技能	思考・判断・表現（太枠は重要項目）						
Name	声	ジェ	①	②	③	④	⑤	⑥	評定
1									
2									
3									
4									
5									

図6 チェック式採点指針ループリック(黒塗りは児童名)

ダネル他(2014)は、「チェック式ルーブリックは評価のスピードやその容易さという点では最も優れている。これはプレゼンテーションのように短時間で詳細に評価しなくてはならない場合に適している。発表を聴きながら、それぞれの該当項目をチェックすればいい」と述べている。

これまでと同様に、名簿上部に採点指針ループリックを掲載する。さらに、名簿には下線部毎のチェックボックスを設け、できた項目にはチェックを付けることができるようにした。これにより、どの項目ができていたかを可視化でき、児童への具体的フィードバックが可能となる。また、話しぶりが評価に混同されぬよう、独立した項目として設定した。このチェック式採点指針ループリックを使用し、3学期末の単元である「大きくなったら」のプレゼンテーションの評価を行う。

6-2 「大きくなったら」での実践

研究の流れはこれまでと同様である。「大きくなったら」は自分の将来の夢について理由等を交えて言えるようになることが目標の単元である。この単元では、これまで学習した表現を駆使して多様な表現をできる。3月上旬に「大きくなったら」の単元末プレゼン評価に上記のルーブリック(図6)を用いて行い、 α 係数を算出した。なお本単元は、コロナウイルス感染症拡大による学級閉鎖や職員の出勤停止等の事情により、通常実施ができた3クラス分のデータを掲載する(表13)。

表 13 チェック式採点指針ループリックによる α 係数

	2組	3組	4組
大きくなったら	.857	.950	.918

チェック式採点指針ループリックを使用して評価を行った結果、どのクラスでも高い信頼性を示す α 係数

が算出された。また、最低値が.857 と安定した評価者間信頼性を示した。

最後に、実際に使用した評価シートを記載し、評価の実際について論じていく。今回は担任 A、B と筆者による評価シートの一部を掲載する（図 7～10）。なお、個人情報保護の観点から、担任名と学級名は伏せて A や B のように表記する。

27							O	O	O	A
28							O			B
29										
30				@				O	O	A
31								O	O	A
32								E		B
33				@	O	O	O	O	O	#A

図7 担任Aによる学級Aの児童の評価(一部抜粋)

[illegible]

図8 筆者による学級Aの児童の評価（一部抜粋）

担任 A は①～③の項目ではとりわけ突出した部分のみチェックし、④～⑥の重要項目のできた部分にのみチェックをつけるようにした（太枠内）。筆者も同様の方法を採用しており、太枠内を見ると、チェック部分と評定がほとんど一致していることが分かる。

17		0	0	V	V	V	V		B
18		V	V	V	V	V	V		B
19		0	0	V	V	V	V		A
20		0	0	V	V	V	V		A
21		0	V	V	V	V	V		A
22		V	V	V	V	V	V		A

図9 担任Bによる学級Bの児童の評価(一部抜粋)

17			✓	✓					✓	✓	B
18									✓		B
19			✓	✓				✓	✓		A
20									✓		A
21								✓	✓		A
22								✓	✓		A

図 10 筆者による学級 B の児童の評価（一部抜粋）

次に、担任 B はできた項目全てにチェックを付け、よくできた項目には○を付けている。この例でも担任 B と筆者のチェックと評価は一致しており、話しぶりなど他の要因との評価の棲み分けがされていると考えられる。

以上のことから、教員によって活用方法に多少の違いがあっても、高い評価の一致度が伺えた。

VII 成果と課題

7-1 成果

今回の実践では、「話すこと（発表）」における「思・判・表」のより信頼性の高いルーブリック開発のために、比較可能性に基づいた比較検証の結果、採点指針ルーブリックによる高い評価者間信頼性が算出された。しかし、いくつかの課題も散見され、更なる

信頼性獲得のために、評価項目ごとにチェックボックスを設けたチェック式採点指針ループリックを開発した。その結果、どのクラスでも安定して高い評価者間信頼性が算出された。これらから、仮説の実証に一定の効果があつたと考える。アンケートからも、採点指針ループリックは指導改善や説明責任の面からも活用しやすいという示唆を得た。

また、何をもって「思・判・表」を評価するかという議論は尽きないところではあるが、言語的側面「思考・判断・表現」、非言語的側面（ジェスチャーや声の大きさなど）は「主体的に学習に取り組む態度」の評価に組み入れた。担任群の評価を見ると、無意識のうちに話しぶりを考慮してしまうことも見受けられたため、このように棲み分けをすることでよりクリアな評価ができるのではないかと考える。

このループリックを「話すこと（発表）」の評価に導入していくことで、英語を専門としない担任でも高い信頼性の元での評価と指導改善が可能になるのではないかと考える。しかし、そのためには比較可能性の元、単元の目標と指導内容を理解し、評価基準を共有しておくことが欠かせない。主観的な評価に陥らないためにも、基準の共有は重要である。そうした前提でループリックを用いることで、より信頼性の高い評価が実現する。これはどの教科にも言えることだろう。

7-2 課題と今後に向けて

更なる評価の信頼性や妥当性獲得に向け、評価基準の表記について見直す必要がある。既習事項について“it is～” “I want～”という記載をしたが、この場合、「知・技」に意識が移ってしまう恐れがある。

「自分の気持ち」や「したいこと」という概念的記載を採用していく方がより「思・判・表」の評価記述として適しているだろう。

次に、短い時間の中での評価には限界があり、複数の観点を同時に一人で見取することはふさわしくない。そこで、今回作成したループリックを用いて、担任は「思・判・表」、ALTは「知・技」という評価の棲み分けをするとよいのではないだろうか。

また、評価の信頼性を高めるためには明確な評価基準の作成と並んで、「評価者トレーニング」が欠かせない。今回の研究結果を活用し、教員の不安感解消のためにも、より効果的なトレーニングを校内研修等に取り入れていく必要がある。

今回は「話すこと（発表）」の評価について研究を行った。プレゼンはある程度の型があり、評価が「やり取り」と比べて容易である。今後は教員が最も困難を感じている「話すこと（やり取り）」に適したPTやループリックの開発に取り組んでいく決意である。

そして、評価基準や内容は児童と共有し、指導改善

に適切に活用してこそ意味を成す。今後は、単元導入時やパフォーマンステスト実施の際に、ループリックを児童と共有し、ループリックによる波及効果や動機付けについて研究を進めていく必要がある。

Ⅷ 終わりに

繰り返しになるが、評価を研究するにあたり重要なことは、「評価のための活動」にならないことである。身に付けさせたい資質能力を明確にして指導に臨み、評価及び改善を行うPDCAの流れを意識することが何よりも重要である。

今回はあくまでも、短い時間で正確な評価の求められる「話すこと（発表）」の評価には総合的・タスク特有型チェック式採点指針ループリックが適しているという示唆を得た。これは、決して他のループリックを批判するものではない。信頼性と妥当性、実施可能性はトレードオフの関係にある。それぞれのループリックの特徴や利点を理解し、使い分けることが大切である。

そして、本実践が評価に困り感を感じている先生方や子どもたちの更なる育成に少しばかりの一助になることを願い、結びとさせて頂く。

【引用参考文献】

- 株式会社イーオン. (2021). 『小学校の英語教育に関する教員意識調査 2021』.
- 小泉利恵. (2016). 「ループリックを使ったスピーキングの評価」. 『英語教育』12月号, 34-35.
- 瀧沢広人(著). (2020). 『小学校英語サポート BOOKS 単元末テスト・パフォーマンステストの実例つき！ 小学校外国語活動&外国語の新学習評価ハンドブック』東京：明治図書.
- 田村学(著). (2021). 『学習評価』東京：東洋館.
- ダネル・スティーブンス/アントニア・レビ(著). (2014). 『高等教育シリーズ 163 大学教員のためのループリック評価入門』東京：玉川大学.
- 中央教育審議会. (2009). 「児童生徒の学習評価の在り方に関するワーキンググループ(第7回)配布資料」.
- 西岡加名恵・石井英真(編). (2021). 『教育評価重要語辞典』東京：明治図書.
- 日本図書文化協会. (2017). 『指導と評価 3月号』東京：日本教育評価研究会.
- 日本図書文化協会. (2019). 『指導と評価 9月号』東京：日本教育評価研究会.
- 文部科学省. (2017). 「小学校学習指導要領解説 外国語編」.