

スマートフォンを使用した相互評価に見られる アノテーション傾向と問題点の分析

Analysis of Annotation Trends and Problems in Mutual Evaluation Using Smartphones

北村 雅 則

Masanori KITAMURA

要 旨

本稿では、グループプレゼンテーションの際にスマートフォンを用いて行った相互評価（リアルタイムアノテーション）とその結果をもとにした振り返り活動について、評価傾向と問題点の分析をした。アノテーションは、話し方などの実技に対する観点と種類（良い／惜しい）について聴衆が行うものであるが、分析の結果、それらは概ね妥当なものであった。実技と関連しない観点や種類のものも存在したが、精査すると評価者の評価基準が強く反映されたアノテーションであり、アノテーションの意図を伝える工夫や支援が必要であることが明らかとなった。また、振り返り活動では操作ミスが目立ったが、改善すべき箇所のみを指摘するという手順を逸脱する行為から、ポジティブな評価志向が明らかとなるなど、操作ミスにとどまらない示唆的な自己評価活動であったことが判明した。

1. はじめに

大学においても協同学習、協調学習の必要性が叫ばれるようになって久しい。現在では授業を受動的に受けるのではなく、学習者が協同して学び合うことを促進するためのさまざまな方策がとられ、その成果が報告されている。筆者らは、学習者の相互学習活動を支援するためのシステムとして、FishWatchr（以下FWと称す）とFishWatchr Mini（以下FWMと称す）を開発し、プレゼンテーションを相互評価し、プレゼンター自身が振り返りをする際に相互評価の結果を用いることで協同学習を促進してきた（北村・山口（2017）、（2018）、（2020）など）。

相互評価という手法は以前から存在し、例えば、評価者に評価結果や全体の印象を発言させグループやクラス全体で共有する方法や、評価シートを用いて評価結果を共有する方法などがあるが、これはシステムの支援がなくても成立するものである。倉田他（2018）や田中（2021）などオンラインでの相互評価の実践例が見られるようになってきたもののまだ少数と言える。本稿の実践では

FWMを用いて相互評価を行うことで、プレゼンテーション中の個別・具体的なパフォーマンスについて、ビデオという証拠をもって評価結果をプレゼンターに示すことができるといった利点がある一方で、評価データ（アノテーションの結果）が大量となり、プレゼンターが振り返りを行う際にデータの把握が困難となったり、取舍選択する必要があるなどの問題が生じている。そういった問題を回避するために、FWMではアノテーションの結果をグラフで示す機能などを実装し、結果を直感的に把握しやすくするための工夫を凝らしているが（北村（2021））、評価結果そのものがどのようなものであり、また、振り返りの際にどのように用いられ、問題が生じているのかは十分に分析していなかった。そこで、本稿ではFWMによって大量に得られた評価データがどのようなものであるのか、その内実や傾向を把握し、学習者たちがそれらを振り返りで用いる際に、評価データの取り扱いについてどのような問題が生じているのか分析を試みる。

本稿は教育実践研究に位置づけられるものである。そのため、データを表面的に分析するにとどまらず、学習者たちが行ったアノテーションや振り返り活動が反映されたデータとして捉え、相互評価と振り返り活動の結果という位置づけからデータを分析する。

2. 本研究の背景と分析の観点

2.1. 実践の概要

本実践を行った授業は、大学1年次にアカデミックスキルの習得を目的とした「基礎ゼミ」である。北村（2021）で分析を行った授業実践と同様の実践手法であるが、本稿で示すのは別のクラスとなる。実践の概要は以下の通りである。

【概要】

- ・受講生17名を、3～4名の5グループに分けたグループプレゼンテーション（プレゼンテーションの回数は2回）。
- ・プレゼンテーションの際には、FWMを使用してリアルタイムでアノテーションする。
- ・アノテーションは、6つの観点（①話し方、②アイコンタクト、③ジェスチャー・工夫、④スライド、⑤（内容の）新規性、⑥（内容の）有用性）について、2つの種類（良い／惜しい）を付与する。
- ・プレゼンテーションの次の授業（30分程度）で、FWを使用し、プレゼンテーションの映像とアノテーションを同期したデータを見ながら振り返る。

アノテーションは、聴衆となる学生と教師が行った。プレゼンテーションを見ている間、目に留まった場面において、各自のスマートフォンやPCのブラウザでFWMを操作し、①～⑥の観点と種類（「良い」か「惜しい」か）を選択する。これらすべてのアノテーション結果は時刻情報とともにサーバーに記録されている。プレゼンテーションを録画した映像と同期することで、どの場面でどのような観点と種類のアノテーションが付与されたのかがクラス全体で共有できるようになる。

教師は、授業後、映像とアノテーション結果を同期したデータを作成し、プレゼンテーションの次の授業において、USBメモリによって学生たちに配布し、学生たちにはFWを使って振り

返りを行わせた。その際、学生たちは自分の担当箇所の中で、納得できたアノテーション5箇所程度について、どうしてこのような評価（どのように良かった／悪かった）になったのか、また、どのような改善が必要だと考えるのかをコメントに残してもらった。また、アノテーションが付された以外の箇所（評価を受けていない箇所）において、自分自身で改善が必要だと感じたところには「要改善」というアノテーションを、2回目のプレゼンテーション後に振り返る際、1回目よりも改善できたと感じたところを「改善済」とアノテーションし、その理由や状況等をコメントしてもらった。以上のような授業実践から得たデータは以下ようになる。

【得られたデータ】

- (A) プレゼンテーション中、FWMを用いて行ったリアルタイムアノテーション
- (B) プレゼンテーション後の振り返りの際、評価者から付与されたアノテーションにプレゼンター自身が付与したコメントの箇所
- (C) (B) のコメント
- (D) プレゼンテーション後の振り返りの際、プレゼンター自身が気づいた「要改善」箇所
- (E) (D) のコメント
- (F) プレゼンテーション後の振り返りの際、プレゼンター自身が気づいた「改善済」(2回目のみ)
- (G) (F) のコメント

2.2. 相互評価

2.1 に示した授業実践において行う相互評価では、学生たちは自分がプレゼンターとなるときには評価を受け、聴衆であるときにはFWMを用いて評価を行うという関係にある。

評価と一言に言ってもさまざまな評価観が存在する。想像しやすいのは成績評価であり、これは客観的な基準、根拠に基づいて評価値を決定し、提示しなければならず、これには責任が伴うものである。しかし、本稿の実践においては、聴衆となる学生たち評価とはリアルタイムで行うものであり、熟慮した結果ではなく即時的で、直感的なものとなる。また、学生たちがプレゼンテーションに対する確固たる評価軸を持っているか否かは判然とせず、FWMのアノテーションから得られる評価データとは評価者自身の主観的評価の表れにすぎない可能性がある。

ここで、このような主観的評価が意味をなすのか検討しておかなければならない。本稿における評価とは、宇佐美（2014）に示される評価と同様なものとして位置づけることとする。

主体が持つ内的・暗黙的な価値観に基づいて、対象についての情報を収集し、主体なりの解釈を行ったうえで、価値判断を行うまでの一連の認知プロセス。またその結果として得られる判断(宇佐美(2014) p. 2)

そもそも、プレゼンテーションのパフォーマンスに絶対的な正解はない。最低限押さえておかなければならないこと、つまり、そうすることを避けた方が好ましいことは存在するとしても、個別・具体的なパフォーマンスについてこうすれば必ず良くなるということはない。ある人は良いと感じても他の人はそうではないと感じることもある。また、本実践に当てはめて言うならば、聴衆となる評価者全員が、仮に一定の評価軸を有する者（例えば教師）であったとしたら、主観的評価から

逃れられるかといえそうではないだろう。各々の評価軸から見て妥当な（客観的と思える）評価をする可能性が高いというだけであり、各々の評価軸が完全に一致するものとはならない。

このようにプレゼンテーションの評価は誰が行っても主観性を排することができないものである。したがって、プレゼンテーションにまだ慣れておらず、自身の評価軸が定まっていない学生の直感的な評価だからといって、評価結果に妥当性が見られないと却下されるといったものではなく、アノテーションの結果は一つの意見として尊重されてもよいものであろう。

相互評価は評価をしてそれで終わりではない。評価結果を受けて、プレゼンターが振り返りの材料とし、今後活かせるヒントとなることが求められる。そういう意味で評価結果は被評価者であるプレゼンターを納得させるものであることが望ましい。仮に評価者が教師であれば、立場差という本質とは違った理由になるかもしれないが、プレゼンターは評価結果にある程度の納得をするだろう。しかし、本実践の評価は同じ立場の学生同士である。この相互評価が意義を有する評価活動であったのかを知るためにも、アノテーションの傾向と内容を精査する必要がある。

2.3. データの選定と分析の観点

以上のような評価観のもと、本稿で分析する際の観点とデータについて述べる。本稿で分析する授業実践は、【[プレゼンテーション実技→評価]→振り返り】という流れである。評価活動としては、[]に示す部分と[]を包含する【 】に示す部分に分けられる。[]の部分で行われる評価について着目する点は聴衆である評価者がプレゼンターに対して行ったアノテーションであり、【 】の部分で行われる評価について着目するのは、プレゼンター自身が自分に対して行う評価である。

(C) (E) (G) に示したコメントデータの分析については別稿に委ねることとし、本稿ではアノテーションにのみ着目し、以下の3点について分析を行う。

- ① 2回のプレゼンテーション時に行われたリアルタイムアノテーション→ (2.1に示す (A))
- ② 振り返り時にコメントを付したリアルタイムアノテーション→ (B)
- ③ プレゼンター自身が付した「要改善→ (D)」・「改善済→ (F)」のアノテーション

①に関して言えば、②・③の振り返り時の材料となるものである。どのような観点・種類のアノテーションが付与されたかを知ることにより、そのアノテーションが適当な評価であるのかを検討しておく必要があるだろう。先に本実践における評価とは主観的评价であることを述べた。ここで言う適当な評価とは、アノテーションと実際の映像を照合してなぜここでそのアノテーションが付されたのか、その意図が推測できるか否かと考えることとする。例えば、発話していない場面であるのに、「話し方」のアノテーションが付与されているとしたら、そのアノテーションは適当ではないということである。①には大量のアノテーションが存在する。そのため、これらを振り返りの材料と考えたとき、一つ一つのアノテーションの重みは軽いと言えるが、そうであっても適当ではないアノテーションが多く存在すれば振り返りの際の障害となりかねない。

②に関して言えば、大量に付された自分自身に対するアノテーションの中からどの種類・観点に着目したかを分析する。それにより、①のアノテーションから何を（取捨選択し）、それをどのように判断し、省察したかを知ることができる。

③は、②と同様の評価活動ではあるが裏返しの関係にある。なぜならば、②がアノテーションに

対してコメントを残すという自己評価であるのに対し、③ではアノテーションがないところについて、自己評価を行うというものだからである。③を分析することによって、他者から受けた評価（①のアノテーション）とは異なる、自己評価の視点を見いだすことができる。

分析の対象とするアノテーションについては以上の通りであるが、データの分析を通して、相互評価や振り返りの行為、操作上の問題も明らかとなる。そういった、実践上の問題点も抽出し、分析することが本稿での分析の狙いとなる。

3. 分析

3.1. リアルタイムアノテーションの傾向と適否

はじめに、2.3に示す①について分析を行う。プレゼンテーションは2回とも5グループ分を通して行った。1グループの発表時間は10分、グループメンバー全員が交代しながらプレゼンテーションを行うこととした。また、時間に関しては、意図せず早く終了するのも超過するのも許容した。その結果、プレゼンテーションの時間は、1回目が1時間12分45秒、2回目が1時間7分25秒となった。

表1は、2回のプレゼンテーションにおけるリアルタイムアノテーションの観点と種類の分布である。表1での①とは1回目のプレゼンテーション、②とは2回目のプレゼンテーションを指す。

表1：リアルタイムアノテーションの傾向

観点／種類	①良い	①惜しい	①計	②良い	②惜しい	②計
話し方	106	17	123	217	16	233
アイコンタクト	107	12	119	202	31	233
ジェスチャー	74	15	89	152	20	172
スライド	79	30	109	102	28	130
新規性	141	6	147	175	6	181
有用性	96	86	10	122	16	138
計	593	90	683	970	117	1087

表1の数値は各回のプレゼンテーション全体の数であり、グループごとに付与されたアノテーション数は異なるが、概して1グループあたり少なくとも100以上のアノテーションが付与されたことがわかる。

観点（話し方、アイコンタクト、ジェスチャー・工夫、スライド、新規性、有用性）に関しては6つとも極端な偏りはなく、まんべんなく分布している。また、種類に関しては「良い」の方が圧倒的に多く、「惜しい」は少ないという結果になった。

次に、リアルタイムアノテーションが適当であったか不適当であったかを、分析時に判定した結果を表2に示す。FWではアノテーションとプレゼンテーションの録画映像が同期されており、個々

のアノテーションが付された箇所を映像で見ることができる。適否の判断は筆者一人が行った。適当を○、不適當を×、判定不能を△として、その結果をまとめると以下ようになる。

表2：リアルタイムアノテーションの適否

	1回目	2回目	計
○	660	1072	1732
×	19	15	34
△	4	0	4
計	683	1087	1770

アノテーションと該当箇所の映像を照合した結果、問題があるとは思われないところが多く、適当と判断したが、中には違和感を覚える箇所が存在した。それらについては、その違和感をもって、不適當または判定不能に分類するのではなく、筆者が評価者であれば違うアノテーションをしたであろうが、こういうアノテーションに至る理由が推し量れるものについては適当と判断した。違和感を覚え判断に迷った具体例として、スライドに示されている情報量に関するものが挙げられる。例えば、当該箇所で主張したい内容を箇条書きに示したというような文字列だけのスライドや、逆に、キーワードのみが示されたスライドである。それらに対して、ある学生は良い、別の学生は惜しいとアノテーションした場合があり、判定する際に困難を伴った。スライドの情報量に関しては、文章を隙間なく列挙するようなものは好ましくなくとしても、要点を箇条書きで示してあるものに関しては、文字が多かったとしても聴衆にとっては理解を促進させ、好意的に捉えられる場合がある。しかし、文字が多かったことを理由として、情報の詰め込みすぎと捉える向きもある。キーワードのみのスライドという逆の場合も同様であり、プレゼンターが自身の主張を強調するためにそうしたのであろうが、それを好ましく捉える場合も、不適當と捉える場合もある。判定が困難な例はスライドだけではなく、話し方、アイコンタクトの良し悪しなどもある。したがって、「良い／惜しい」という種類については、確実にそうではないと思える場合を除いて、適当と判断した。

FWMを使ったリアルタイムアノテーションの操作はそれほど困難なことではないが、直感的な操作であるがゆえに、評価という観点から見るとその信頼性に疑念が生じる場所である。アノテーションの判定を行った者はアノテーションをした本人ではないため、アノテーションをした際の判断基準、背景等については推し量ることしかできないが、それがどのような経緯で付与されたかはさておき、映像という証拠から判断すると「適当」のアノテーションがほとんどであったと結論づけられる。

3.1.1. 不適當、判定不能のアノテーションの理由

一方で、判定作業中、違和感を覚えた箇所について、熟慮の結果、不適當(×)、または、判定不能(△)と判断したものの一覧とその理由を表3に示す。

表3：アノテーション×・△の理由

	理由	①×	②×	×計	①△	②△	△計
観点	i. アイコンタクトの間違い	3	2	5	0	0	0
	ii. ジェスチャーなし	3	9	12	0	0	0
	iii. ジェスチャーなし、または、 アイコンタクト	0	0	0	4	0	4
種類	iv. 話し方問題なし	1	3	4	0	0	0
	v. アイコンタクトの問題なし	0	1	1	0	0	0
	vi. スライド変化なし	4	0	4	0	0	0
	vii. 不適切なジェスチャー	1	0	1	0	0	0
不誠実な態度	viii. 時間外	7	0	7	0	0	0
計		19	15	34	4	0	4

不適當は計34箇所、判定不能は計4箇所認められたが、それぞれの判定理由は以下の8つに分類できた。さらに、この8つの理由を種類分けすると、iからiiiは観点、ivからviiは種類、viiiはアノテーション付与の問題ではなく、不誠実な態度（評価に臨む態度の問題）にまとめることができる。以下8つの理由について、それがどのような状況であったのか概要を示すと以下ようになる。

【不適當、判定不能の理由と状況】

- i. アイコンタクトの間違い…「ジェスチャー・工夫」など、「アイコンタクト」以外の観点がアノテーションされていたが、映像からはアイコンタクトの問題と推測される。
- ii. ジェスチャーなし…ジェスチャーやプレゼンテーション中の工夫があったとは見受けられないが、「ジェスチャー・工夫」とアノテーションされている。
- iii. ジェスチャーなし、または、アイコンタクト…iとiiが合わさった形。「ジェスチャー・工夫、惜しい」とアノテーションされているが、「アイコンタクト、惜しい」であると判断できる。
- iv. 話し方問題なし…「話し方、惜しい」とアノテーションされているが、声量、滑舌等、話し方の問題があるとは思われない。
- v. アイコンタクト問題なし…「アイコンタクト、惜しい」とアノテーションされているが、その前後の時間を含め、前を向き、視線を左右に向けておりアイコンタクトの問題はない。
- vi. スライド変化なし…「スライド」とアノテーションされているが、当該箇所に至るまで、長時間、スライドが送られていない。
- vii. 不適切なジェスチャー…「ジェスチャー・工夫、良い」とアノテーションされている箇所でジェスチャーが行われているが、プレゼンテーションの際には不適切であると思われるジェスチャーをしていた。
- viii. 時間外…プレゼンテーションの終了後にアノテーションされている。

不適當なアノテーションは1回目の方が多く、2回目は減少した。これは1回目がFWMを使っ

た初めての相互評価であり、こうした評価活動に慣れていなかったことが理由として推察される。また、評価者として評価に慣れたという側面と、評価を受けて被評価者として振り返りを行うことで、学生たち自身の中で評価活動自体が適正化された可能性がある。

こうした推察のもと、2回目の方が増加したのを見てみると「iv. 話し方問題なし」「ii. ジェスチャーなし」が該当する。3.2に示す振り返り活動と関連するが、「iv. 話し方問題なし」の2回目は3箇所と増えたが、そのうちの2箇所については振り返りのコメントがつけられていた。一つは「話すときに何を話すかわからなくなってもその動揺を体で表さないようにしたい」とコメントし、自分の立ち方、姿勢の問題と捉え、もう一つは「スライドを見やすく作る」とコメントし、スライドの問題と認識されていた。前者は、「話し方」を狭く捉え、声量や滑舌等の問題の有無とすれば問題が存在せず適当な箇所であったが、本実践においては評価の観点として考慮していなかった挙動について、あえて「話し方」という観点で表したと考えると不適當とは言えないものであった。後者については、「話し方」という点については問題がなく、コメントがなければ種類が不適当なアノテーションとしか捉えられないものであったが、被評価者の実際の解釈から想像すると、これは種類の問題ではなく、「スライド」にすべき箇所であって、観点の問題であったと言える。

「ii. ジェスチャーなし」については、1回目は不適當が3箇所、2回目は9箇所となり2回目の方が増加している。しかし、その内訳はまったく異なる様相を見せる。1回目の3箇所についてはアノテーションをした人がすべて異なるのに対し、2回目の9箇所は二人によるアノテーションで、一人が6箇所、もう一人が3箇所をアノテーションしていた。さらに、2回目の同一人物が6箇所をアノテーションした場合が特徴的であり、この6箇所は同一プレゼンターが担当する箇所についてアノテーションされていた。内1箇所について振り返りのコメントがつけられていて「初めの一度以外、ジェスチャーをしなかったから」とあった。ここからアノテーションをした意図を推測すると、「ジェスチャー、惜しい」というアノテーションは、行ったジェスチャーが不適當であったことによるのではなく、当該箇所においてジェスチャーや工夫をする必要があったのに、それをしていないことに対するアノテーションであったと考えられる。

このアノテーションについては、映像からジェスチャーや工夫を行った形跡が見られない以上、不適當と判断されるものではある。しかし、この箇所で、仮にジェスチャー・工夫があったとしたら、より良いプレゼンテーションになった可能性があることは事実である。このように、アノテーションに隠された評価者の意図は表面的にアノテーションの数値だけを追ってはいは見えることはなく、アノテーションを用いて振り返りを行ったことによって見いだされた意義の好例と言ってよいだろう。

3.2. 振り返り活動

次に、リアルタイムアノテーションの結果を用いて、プレゼンターが振り返りを行った活動全体について分析を行う。2.1に示した通り、学生たちは振り返りの際、自分が担当したプレゼンテーション箇所に付与されたアノテーションのうち、5箇所程度についてコメントを残すことと、アノテーション以外の箇所で改善が必要な箇所（要改善）と1回目から改善が見られた箇所（改善済）をアノテーションする。本節では、2.3に示す②と③について分析を行う。

3.2.1. 振り返り時に着目したアノテーション

まず②に関する分析を行う。振り返りを行う授業の回において、学生たちは自分のプレゼンター

ションについて付与されたアノテーションの詳細を見る。その際、自分が納得したり、印象に残ったりするアノテーション箇所について着目するわけであるが、どのような観点・種類のアノテーションに対してコメントが残されたかを見たものが表4である。

表4：振り返り時に着目したアノテーション

観点／種類	①良い	①惜しい	①計	②良い	②惜しい	②計
話し方	3	5	8	5	4	9
アイコンタクト	8	4	12	7	5	12
ジェスチャー	4	3	7	6	3	9
スライド	4	10	14	6	11	17
新規性	11	2	13	9	1	10
有用性	4	0	4	1	2	3
計	34	24	58	34	26	60

観点によって異なるが、アイコンタクトとジェスチャーと新規性については1回目も2回目も種類「良い」の方に着目する傾向にあり、スライドについては2回とも「惜しい」の方に目が向いたようである。話し方と有用性については1回目と2回目に「良い／惜しい」の多少が逆転するという結果となった。

振り返り時に着目する観点と種類に偏りが見られなかったということは、プレゼンテーションの経験が豊富ではない学生たちであったとしても、この観点に関しては自分なりの評価軸を持っており、他者から得た評価をふまえて自分のパフォーマンスを評価することができているということになる。この点は学生たちが評価軸を有していないと考えた、実践前の予想を覆すものであった。

もう少し掘り下げて分析すると、3.1の表2に示したリアルタイムアノテーションの適否のうち、不適當・判定不能と判定したものについては、振り返り時に参照するとミスリードする可能性があるアノテーションであるが、それらに着目した振り返りが5箇所あった。不適當のリアルタイムアノテーションは3箇所、それはすべて「話し方／惜しい」、判定不能は2箇所「ジェスチャー・工夫／惜しい」であった。本稿では振り返り時のコメント内容について考察の対象外とするため、詳細は稿を改めるが、不適當の「話し方／惜しい」に着目した振り返りコメント2例がコメント内容としても不適當であった。

判定不能の「ジェスチャー・工夫／惜しい」に着目した振り返りは、3.1.1に示した「ii. ジェスチャーなし」に該当する。この件については、リアルタイムアノテーションとしては不適當である可能性が高いが、アノテーションの意図を推測することで新たな気づきを得られる箇所である。

3.2.2. 振り返り時に追加されたアノテーション

次に③に関する分析である。振り返りの際、FWには要改善と改善済という2つのボタンを用意し、自分が担当した箇所の映像を見て、該当する箇所があればいずれかのボタンを押してもらった。その際、要改善と改善済は種類のアノテーションとしてリアルタイムアノテーションの中に追加され

る。

この振り返り作業で追加されたアノテーションは、要改善が107箇所、改善済16箇所であった。しかし、振り返り手順には存在しない「良い」「惜しい」「グループ」というタグも計33箇所追加されていた。まずは、新たに追加されたアノテーションが適当か否かを分析した結果を表5に示す。ここでの適否とは、要改善・改善済のアノテーションとそれに対応するコメントが存在するかという意味で用い、コメントの内容については考慮していない。

表5：追加されたアノテーションの適否

	1回目	2回目	計
○	14	21	35
×	77	33	110
△	6	5	11
計	97	59	156

追加されたアノテーションは、不適當（×）が最多であり、適當（○）は35箇所で2割強と多くはない。ここで述べる適當（○）とは、要改善・改善済のアノテーションとそれに対応するコメントが入力されているものことである。不適當（×）と判定不能（△）については、表6にその理由と分類結果を示す。

表6：要改善・改善済アノテーションが不適當な理由

理由	①×	②×	×計	①△	②△	△計
操作ミス	63	25	88	0	0	0
種類変更	14	8	22	0	0	0
種類変更・コメントあり	0	0	0	6	5	11
計	77	33	110	6	5	11

不適當（×）に該当するものは2つに分けられる。一つは、事前に指示をした振り返り手順を守らず、要改善・改善済のアノテーションをしたが、コメントが入力されていないもので「操作ミス」と考えられるもの、もう一つが「種類変更」で、コメントが未入力である点は変わらないが、自動で入力される種類「要改善」をあえて「良い／惜しい」に評価者自身が書き換えたというものである。判定不能（△）としたものは、不適當の「種類変更」に近いが、コメントが入力されているものである。

コメント未入力という「操作ミス」については、手順の周知や理解が至らなかったという点もあるが、FWMを用いて評価者として評価する際、コメント入力を求めず、リアルタイムアノテーションだけであったため、そういったやり方に影響されたという可能性が考えられる。しかし、「操作ミス」のアノテーションの中には同一の人物によって短時間に要改善または改善済が大量に付与さ

れた箇所があったため、単純な操作ミスというだけではなく PC の動作不良やシステムのエラーという問題も否定できない。

「種類変更」と「種類変更・コメントあり」については、こちらが意図した振り返り手順ではなかったが、振り返りの方法としては新たな示唆を得た部分である。振り返りをする際、自分が担当したプレゼンテーションの映像を見て、アノテーションを確認し、アノテーションが付与された以外の箇所について要改善／改善済を探すとこの作業を求めたわけであるが、映像の中に要改善箇所だけでなく、プレゼンター自身が手応えを掴んだ部分を見つけたということを表している。「種類変更」において、「良い」に変更したものは1回目が9箇所、2回目が3箇所の計12箇所、「惜しい」は1回目が5箇所、2回目も5箇所の計10箇所であった。「惜しい」の方は要改善と同様の意図であると思われるが、それを上回る数の「良い」が1回目の振り返りの際につけられている。2回目の「良い」は「改善済」に近い意図でアノテーションされていると思われるため、減少することは予想できるものの、1回目の振り返りの際からポジティブな評価をしたかっ様子うかがえる。

Kolb (1984) の経験学習や PDCA サイクルの流れを想定して振り返りを行う場合は、本実践でも想定したように、実行したことに対し改善すべき課題を自覚し、今後活かすべく克服するための方策を見だし、結果につなげるという発想に陥りやすい。もちろん課題を見つけ、それを克服しようとすることは振り返ることの一つの成果であるが、操作手順から逸脱した学生たちの振り返りからは、達成できた点についても意識するということが評価の視点として必要なことであるという示唆を得た。

3.2.3. 分析結果から見える実践上の問題点と改善点

以上、「①2回のプレゼンテーション時に行われたリアルタイムアノテーション」、「②振り返り時にコメントを付したリアルタイムアノテーション」、「③プレゼンター自身が付した「要改善→(D)」・「改善済→(F)」のアノテーション」それぞれの適否を中心に分析を行った。

①からはリアルタイムアノテーションが概ね適当であり、評価結果として信頼が置けるものであったことが判明した一方で、改善が必要な箇所も見いだせた。3.1 では、映像とアノテーションを照合し、適否を判断する際に困難が生じたことを述べたが、同一観点に対して異なる種類のアノテーションが共存するという場合が存在し、その場合被評価者であるプレゼンターはその結果を取捨選択することを迫られる。多くの場合、相反するアノテーションが共存する場合はアノテーションが集中する箇所で、注目すべき箇所であるのに、振り返りの材料とするには迷いが生じる箇所ともなる。この問題について本稿では指摘にとどめ、別稿において当該箇所の分析を進め、改善点を模索したい。

3.1.1 に見た不適當、判定不能のアノテーションの分析からは、アノテーションの意図がわかるような工夫があると、さらによい振り返りとなる可能性を見いだした。この点は分析してはじめてわかったことであり、本実践の振り返り時には問題として存在したままであった。観点の定義を詳細化し、全体で共有するということが一つの問題解決策として考えられるが、そもそもリアルタイムアノテーションが主観的評価であり、唯一無二の正解を導くものではないため、定義の詳細化よりはアノテーションの意図を伝える工夫を凝らす方策を考えたい。

②からは、不適當なリアルタイムアノテーションに着目し、不適當な振り返りをした場合があるということが確認できた。この件についてもコメントの内容も合わせた分析が必要となるが、不適當なリアルタイムアノテーションを減らす方策を考えなければならない。

③については、操作手順を遵守しないという問題が多く見られ、それについては操作の慣れと周知の徹底で問題解決が図られるとしても、種類の書き換えという操作手順逸脱の問題からは、振り返りの方法自体を検討し直す必要性が示唆された。リアルタイムアノテーションが大量に行われているとしても、振り返り時には被評価者である学生にしか気づかない自分なりの視点、評価軸というものが存在する。これについても、先に述べたアノテーションの意図がわかるようにする工夫と関連するが、評価者の考えと被評価者の考えを対照させるような振り返り結果を提示できるとよい。

4. おわりに

本稿は、FWMを使用したリアルタイムアノテーションと、FWを使用し振り返りをした際の着目点についての内容や適否について分析をした。分析結果からこの相互評価活動が有効に働いたということが確認できた一方、実践時には発覚しなかった問題点も出てきた。本稿ではアノテーションの分析に止まっており、今後を見据えた改善点を提案するところまでは行き着くことができなかったが、現状では問題点も表面的に浮かび上がってきた段階である。今後、コメントの内容と合わせた分析を行い、問題点と改善点のさらなる追究をしなければならない。これは次稿以降への課題である。

謝 辞

本研究は、2020年2月29日、日本教育工学会2020年春季全国大会(大会中止)でみなし発表となった「モバイルデバイスを用いたプレゼンテーション相互評価と振り返りの信頼性」を発展させ、大幅に改稿したものである。改稿にあたり、JSPS 科研費 20K03116 の助成の成果を受けた。

参考文献

- 宇佐美洋 (2014) 『「非母語話者の日本語」はどのように評価されているか—評価プロセスの多様性をとらえることの意義』, ココ出版
- 北村雅則・山口昌也 (2017) 「リアルタイムアノテーションによるプレゼンテーション相互評価の実践」, 『日本教育工学会第33回全国大会予稿集』, pp. 395-395
- 倉田伸・藤木卓・室田真男 (2018) 「携帯型モバイル端末によるビデオプレゼンテーション相互評価支援システムの開発」, 『日本教育工学会論文誌』 41 (Suppl), pp. 201-204
- 北村雅則・山口昌也 (2018) 「知識量の変化に着目した協同学習の学習効果の分析」, 『南山大学短期大学部紀要』 第39号, pp. 161-174
- 北村雅則・山口昌也 (2020) 「モバイルデバイスを用いたプレゼンテーション相互評価と振り返りの信頼性」, 『2020年春季全国大会(第36回大会)プログラム集』, pp. 339-340
- 北村雅則 (2021) 「プレゼンテーション学習における振り返りサイクルを確立するための実践的研究」, 『アカデミア人文・自然科学編』 第21号, pp. 213-226
- 田中克明 (2021) 「オンライン授業におけるレポート文書相互評価方式の構築」, 『埼玉工業大学人間社会学部紀要』

19. pp. 47-52

Kolb, D. A. (1984) *Experiential Learning: Experience as the Source of Learning and Development*, Prentice Hall