

教員研修受講者の反応は記名調査によって測定可能か

岡山県総合教育センター 上岡 伸

研究の概要

本研究の目的は、教員研修受講者の反応を記名調査によって測定した場合に、妥当な測定結果が得られるかを検討することである。記名調査においては特に作為回答が生じやすいことが知られており、研修の特徴を捉えた測定結果とはならない可能性がある。調査 1 では、複数の研修内容に対する同一受講者の反応をプールして級内相関を算出したところ、研修内容間の分散は 11% 以下である一方、受講者間の分散は 42% 以上にのぼることが明らかとなった。調査 2 では、2 つの研修内容について匿名調査を試行したところ、前年度同一研修の記名調査でみられた分布の偏りや強い項目間相関が一定程度緩和されていた。これらの結果は、少なくとも記名調査による反応測定では、受講した教員が付度することにより、研修の特徴を捉えた妥当な測定結果が得られないことを示唆している。

キーワード

教員研修、効果測定レベル、社会的望ましさ、記名調査、マルチレベルモデル

1. 問題と目的

(1) 改めて問われる教員研修評価の在り方

新たに構築された教員研修プラットフォームに関連し、令和 4 年度第 2 次補正予算では「教員研修の高度化に資するモデル開発」に 10 億円超が措置されるなど、教員研修評価の在り方に関心が高まっている。現在、国内の教員研修機関では Kirkpatrick の効果測定レベル (Kirkpatrick and Kirkpatrick 2006) が広く認知されており (e.g., 独立行政法人教職員支援機構 2018)、この 4 段階モデルに基づいて研修改善に取り組んだ実践研究も報告されている (小清水ほか 2016)。

4 段階モデルでは、レベル 1 (反応)・2 (学習)・3 (行動)・4 (業績) は段階的に達成される想定であるが、先行研究によるとレベル 1 の反応は他のレベルとの間に正の相関がないか、あるとしても小さい効果量である (e.g., Blume *et al.* 2010; Ebner and Gegenfurtner 2019; Sitzmann *et al.* 2008)。また、研修での学び (レベル 2) が仕事に転移する際には個人特性や職場環境など様々な要因が複雑に影響するため (Blume *et al.* 2019)、レベル 2 とレベル 3 の関係も小～中程度の効果量である (Blume *et al.* 2010)。こうした先行研究の指摘を踏まえ、レベル 1・2 は研修効果の指標というより、効果的な研修であったかどうかの指標だとみなす主張がある (レビューとして、中原ほか 2022)。

他方で、受講者の行動変容度合（レベル3）と業績（レベル4）との正相関については大きい効果量が報告されている（ $r=.63$: Saks and Burke 2012）。そして、実験の手続きによるレベル4の検証が「労多くして益少なし」であるため、レベル4との結びつきが強く、なおかつより低いコストで測定可能なレベル3は、研修の評価・改善の実務において重要な指標であると考えられている（中原ほか 2022）。

（2）反応測定の目的

しかし、平成18年度時点で一部の研修なりとも追跡調査を行っていた都道府県・政令指定都市の教育センターは62.1%に過ぎず（全国教育研究所連盟 2007）、その割合は令和2年度時点で49.2%に低下している（大阪府教育センター 2020）。代わりに普及しているのはレベル1であり、平成18年度時点では93.9%の教育センターが全ての研修でアンケートを実施していた。

レベル1が最も一般的な評価指標であることは、教員研修に限った話ではない（徳村ほか 2006）。研修評価研究においては、業績向上に益する研修であるかを評価する目的でのレベル1の価値について、長年議論が続いている。この議論は Alliger and Janak (1989) のレビューが端緒となっており、レベル1と他のレベル間の正相関の有無・効果量の大小がその焦点である。Alliger *et al.* (1997) のメタ分析では、レベル1を感情的反応と実用的反応とに分類することで、実用的反応とレベル2との正相関が見出された（mean weighted $r=.26$ ）。ところが、Sitzmann *et al.* (2008) のメタ分析ではより小さい効果量（宣言的知識 $\rho=.16$ 、手続的知識 $\rho=.15$ ）が報告されており、「感情的反応と実用的反応を区別することにはほとんど意味がない」と結論付けられている。また近年は、研修実施形態（ウェビナー・対面・オンライン非同期）間の標準化平均差を基に、レベル1とレベル2の負の相関を報告した研究もある（Ebner and Gegenfurtner 2019）。

なお、業績向上に益する研修かどうかとは別の観点として、4段階モデルを提案した Donald L. Kirkpatrick は、研修事業者が新規顧客とリピーターを獲得する上での反応測定の重要性を指摘している（Kirkpatrick and Kirkpatrick 2006）。教育センターは希望者向けの研修を多数行っているため、こうした顧客獲得を目的としてレベル1を測定する観点も無関係ではない。

（3）教員は反応測定時に付度するのか

いずれの目的にしろ、レベル1を測定する際には、「反応は研修の特徴を捉えている」という仮定の成立が前提となっている（Sitzmann *et al.* 2008）。Sitzmann *et al.* (2008) はメタ分析を通じ、レベル1は講師のコミュニケーションの上手さ（immediacy: $\rho=.66$ ）と研修内での人的相互作用の量（human interaction: $\rho=.56$ ）という研修由来の要因との間に大きい効果量の関連を持つことを示し、当該仮定の成立を裏付けた。

一方で、「よほどひどい研修でなければ（あまり学んでなくても、気持ちよく過ごただけでも）『この研修は良かったですか』と聞かれて5段階の3より下をつけないのが大人の反応というものだ」（鈴木 2015）との指摘があるように、レベル1の測定結果の背後には社会的望ましき特性に由来する作為回答（堀尾・高橋 2004）の影響が懸念される。また、匿名性の低い調査において社会的望ましきのバイアスは増幅することが知られており（Krumpal 2013）、正直な反応を引き出すためには匿名調査とする必要があることを Kirkpatrick and Kirkpatrick (2006) は指摘している。

ところが、日本語で著されている実務者向けの解説書には、レベル1の測定方法として記名欄のあるアンケート様式例を示すものが散見されるほか（平松 2010; 鈴木 2015; 堤ほか 2007）、中には「原則とし

て、記名を推奨」するものも見られ（平松 2010）、現に多数の教育センターが記名調査による評価を実践している（e.g., 兵庫県立教育研修所 2009; 磯部ほか 2023; 奈良県立教育研究所 2008; 上岡 2022; 上岡・中藤 2024）。Sitzmann *et al.* (2008) のメタ分析では、レベル 1 測定時の匿名性がもたらす調整効果は検討されておらず、記名調査の場合でもレベル 1 が研修の特徴を捉えて変動するのか、それとも「大人」度合のような個人特性的要因に強く影響を受けるのかはこれまで検証されていない。

2. 調査 1：研修内容レベルの級内相関

岡山県総合教育センターでは、それまでに実施してきた多くの研修の反応に偏りが見られたことから、測定結果の妥当性が懸念視されていた。もしも反応の測定結果が妥当でないならば、反応に基づく研修改善の取組もその根拠を失うことになる。岡山県総合教育センターでは、研修事業の改善を図る上での必要なプロセスの一つとして、反応項目の妥当性を検証することになった。

「反応は研修の特徴を捉えている」（Sitzmann *et al.* 2008）という仮定が成立するならば、ある研修内容についての受講者同士の反応評定値は相当程度似た傾向を示すはずである（研修内容レベルの級内相関）。しかし、鈴木（2015）が指摘するように個人の特性的要因が色濃く影響するならば、各受講者の回答はどんな研修内容であっても毎回似た値になるだろう（受講者レベルの級内相関）。そこでまず、同一受講者を対象に実施した複数の研修内容についての反応データをプールし、研修内容レベルと受講者レベルの級内相関係数（ICC: Intraclass Correlation Coefficient）を算出することで、それぞれのクラスレベルでどの程度の分散が生じているのかを検討することにした。

(1) 方法

1) 対象者と対象研修

分析対象者は、初任者研修または中堅教諭等資質向上研修を受講した公立学校（国立を含む）の教諭である。初任者研修と中堅教諭等資質向上研修はどちらも教育公務員特例法の定める法定研修であり、年度内に複数の研修日が設けられる。岡山県総合教育センターでは、初任者研修は採用 1 年目、中堅教諭等資質向上研修は 8 年目から 10 年目に受講するよう指定しており、小学校・中学校・高等学校・特別支援学校の校種別に実施している。なお、対象に当該二種類の研修を選択した理由は、法定研修であるがゆえに得られた知見を他の教育センターに一般化しやすいと考えられたことと、数十名規模の同一受講者を対象として多数の研修内容を設ける研修が他に存在しなかったことにある。

初任者研修は共通研修 14 日・選択研修 1 日で構成され、中堅教諭等資質向上研修は共通研修 5 日・選択研修 2 日で構成される。選択研修では受講者が自ら選択して複数の研修会場（e.g., 企業、学校）に分散するため、全員が同じ研修プログラムを受講する共通研修のみを分析対象とした。共通研修の 1 日の中では、全日を通じて一つの研修内容が扱われる場合と、複数の研修内容が扱われる場合があるが、反応測定は研修内容ごとに行われていた。

研修内容の概要は表 1・表 2 に示すとおりである。いずれの研修も 2022 年 4 月～2023 年 2 月までに行われた。各研修内容は、講師が主導する「講義中心」の内容と、ペアワーク・グループワークが中心となる「協議演習」、そして、受講者が個別にオンデマンド（OD: On Demand）形式で受講する「e ラーニング」とに分類され、人的相互作用の量にはかなりの幅があったほか、e ラーニングでは講師と双方向のコミュニケーションを行う術はなかった。従って Sitzmann *et al.* (2008) の仮定が成立するならば、レベル 1 の測定結果は研修内容間で相応に分散するはずである。

表 1 初任者研修の概要

研修内容	時間配分				研修内容	時間配分			
	講義 中心	協議 演習	e ラーニング	形式		講義 中心	協議 演習	e ラーニング	形式
小学校					高等学校				
自己成長のマネジメント			1:00	OD	自己成長のマネジメント			1:00	OD
学級経営・生徒指導1	1:05			集合	学級経営1	1:00			集合
教科指導の基礎技術1算数	0:20	1:25		集合	学習指導案の書き方	1:25			集合
学級経営・生徒指導2	1:00			遠隔	授業づくりの基礎・基本	1:00	1:05		遠隔
発達障害の特性理解	1:10			遠隔	教育の情報化			0:50	OD
GIGAスクール対応基礎	1:00	3:45		集合	GIGAスクール対応基礎研修	2:20	2:35		集合
教科指導の基礎技術2国語		5:10		集合	中堅教諭に学ぶ授業づくり		4:45		集合
教科指導の基礎技術3		4:20		集合	キャリア教育	1:45	1:35		集合
生徒指導・教育相談			2:00	OD	発達障害の特性理解	1:20			集合
教職員としての成長	1:25			遠隔	生徒指導1			2:00	OD
学級経営・生徒指導3	0:50			遠隔	人権教育			1:00	OD
総合的な学習の時間			1:40	OD	教職員としての成長	1:00			遠隔
特別活動			0:50	OD	学級経営2・生徒指導2	1:30			遠隔
教科指導の基礎技術5外国語	2:45			集合	学校講師に学ぶ授業づくり		5:10		集合
教科指導の基礎技術5道徳		2:35		集合	郷土理解・伝統と文化	2:00	2:00		集合
教科指導の基礎技術6	0:50	3:50		集合	授業づくりのまとめ		5:05		集合
キャリアステージと教職生活	2:40			集合	カリキュラム・マネジメント			1:00	OD
自己成長のマネジメント		1:55		集合	総合的な探究の時間			1:00	OD
					道徳教育			1:00	OD
					異業種からの学び	0:50			遠隔
					新たな教育課題(SDGs)	1:10	0:30		遠隔
					特別活動			0:40	OD
					キャリアステージと教職生活	2:15			集合
					自己成長のマネジメント		2:25		集合
中学校					特別支援学校				
自己成長のマネジメント			1:00	OD	自己成長のマネジメント			1:00	OD
学級経営・生徒指導1	1:05			集合	授業設計の基礎技術1	0:15	1:50		集合
教科指導の基礎技術1	2:35			集合	授業設計の基礎技術2			3:15	OD
学級経営・生徒指導2	1:00			遠隔	教育の情報化			1:00	OD
発達障害の特性理解	1:10			遠隔	授業設計の基礎技術3	2:25	2:45		集合
教科指導の基礎技術2	5:10			集合	GIGAスクール対応基礎	2:20	2:20		集合
GIGAスクール対応基礎	2:35	2:35		集合	授業設計の基礎技術4		5:10		集合
教科指導の基礎技術3		4:50		集合	授業設計の基礎技術5		5:15		集合
生徒指導・教育相談1			2:00	OD	授業設計の基礎技術6	2:40	2:30		集合
教職員としての成長	1:00			遠隔	授業設計の基礎技術7	2:35	2:25		集合
教職員に求められるマナー			1:20	OD	生徒指導・教育相談			2:00	OD
人権教育			1:20	OD	人権教育			1:00	OD
生徒指導・教育相談2	2:35			集合	郷土理解・伝統と文化	2:00	2:00		集合
道徳教育	1:15	1:20		集合	教職員としての成長	1:00			遠隔
総合的な学習の時間			1:00	OD	学級経営・生徒指導		1:10		遠隔
協働的な人間関係づくり			1:30	OD	授業設計の基礎技術9		2:20		遠隔
教科指導の基礎技術5	2:10	2:15		集合	キャリアステージと教職生活	2:35			集合
キャリアステージと教職生活	2:35			集合					
自己成長のマネジメント		1:50		集合					

注) 上記外で「開会行事」と「研修の振り返り」に時間配分があり、一日の研修日は5時間30分を上限に企画構成される。

表 2 中堅教諭等資質向上研修の概要

共通研修	研修内容 (校種共通)	時間配分		
		講義中心	協議演習	eラーニング
1日目	協働的な人間関係づくり	1:00		集合
1日目	GIGAスクール対応基礎研修	3:00		集合
2日目	学校組織マネジメントと人材育成			2:45
2日目	OJTを通した若手育成	2:30		OD
3日目	B&S研修		5:10	遠隔
4日目	学校の課題解決に向けた取組の提案【中間】		2:35	集合
4日目	異業種から学ぶ人材育成	2:35		集合
5日目	学校の課題解決に向けた取組の提案【最終】		3:35	集合
5日目	自己成長のマネジメント		1:00	集合

注1) 研修内容は小中高特の校種別で同じであるが、研修資料はそれぞれ異なったほか、講師を担当した指導主事も異なっていた。また、3日目の協議演習時間として表示している5:10は小学校のものであり、中学校は4:50、高等学校は4:45、特別支援学校は4:25であった。その他の研修内容の時間配分は校種共通である。

注2) 上記外で「開会行事」と「研修の振り返り」に時間配分があり、一日の研修日は5時間30分を上限に企画構成される。

2) 調査の実施方法

研修効果レベル 1 及びレベル 2 を同時に測定するため、対象となった全ての研修において Google Forms を用いた記名式のオンライン調査が行われていた。各研修の運営担当者は、この調査の目的が、研修の振り返りを通じて受講者に学びを深めてもらうことと、今後の研修事業の改善に資することの二点であり、回答が目的外に使用されることはないことを説明した上で、研修当日から一週間以内を目安とした任意の回答協力を依頼していた。調査の結果、対象研修の対象受講者547名（初任者267名、中堅280名）のうち、544名（初任者265名、中堅279名）から少なくとも 2 つ以上の研修内容について回答協力が得られていた。

3) 調査内容

レベル 1 測定に用いられた項目は次のとおりである。

① **研修の有効性**：「研修の内容は、自らの力量向上に有効であったと感じますか。」で測定された。研修がその後の仕事の遂行に役立つかどうかを問う実用的反応項目の典型例として、Alliger *et al.* (1997) は「To what degree will this training influence your ability later to perform your job? 」を挙げており、この例に準じて設けられた項目であった。1。有効でなかった～5。有効だった、までの 5 段階で評価が求められた。

② **資料の有用性**：「本研修で用いた資料は、内容を理解するのに役立ちましたか」で測定された。1。役立たなかった～5。役立った、までの 5 段階で評価が求められた。

③ **時間配分の適切性**：「本研修プログラムの時間配分は、内容の理解を深めるのに適切でしたか」で測定された。1。適切でなかった～5。適切だった、までの 5 段階で評価が求められた。

④ **ニーズとの適合性**：「研修の内容は、あなたが期待していた内容でしたか」で測定された。1。期待した内容でなかった～5。期待した内容だった、までの 5 段階で評価が求められた。

これらの項目は、知識技能等を獲得する上で効果的な研修内容であったかを把握するため、岡山県総合教育センターの研修企画担当部署の指導主事らが先行研究（兵庫県立教育研修所 2009; 奈良県立教育研究所 2008）を参考にして作成した項目であり、上岡（2022）においても使用されている。調査方法について教育センター内で所属長までの稟議が行われ、記名調査であることを含めて内容的には妥当であると判断されていた。

(2) 結果と考察

1) 各項目の度数集計と項目間相関

まず、544名から得られた延べ7,036件（初任者4,808件、中堅2,228件）の回答から、全ての研修内容について 4 項目とも同じ数字を選択している受講者66名（初任者40名、中堅26名）の回答延べ951件（初任者744件、中堅207件）を除外した分析を行うことにした。除外者のほとんどは全て 5 を選択した者である（表 3）。このような偏った回答が含まれていると、計算上では研修内容間の分散が小さくなることで研修内容レベルの ICC は低下する。研修運営者にとってオール 5 評価はむしろ喜ばしい結果であるため、研修評価の実務において上記のような除外処理が施されることはまずないと考えられるが、ここではあえて分析から除外した。残った478名（初任者225名、中堅253名）の回答延べ6,085件（初任者4,064件、中堅2,021件）を研修別で度数集計すると表 4 のようになった。

表 3 除外した回答				
初任者研修	除外した受講者数			除外回答数
	全て4	全て5	計	
小学校	0	12	12	201
中学校	1	10	11	185
高等学校	0	11	11	260
特別支援学校	0	6	6	98
計	1	39	40	744
中堅教諭等資質向上研修				
小学校	1	8	9	72
中学校	1	5	6	45
高等学校	0	8	8	66
特別支援学校	1	2	3	24
計	3	23	26	207

表 4 回答の度数集計																						
校種	n	回答数	研修の有効性					資料の有用性					時間配分の適切性					ニーズとの適合性				
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
初任者研修																						
小	97	1,646	3	15	92	460	1,076	4	14	93	475	1,060	3	18	151	515	959	4	20	121	529	972
中	47	803	5	2	39	173	584	6	－	32	196	569	6	5	55	193	544	5	10	52	203	533
高	52	1,157	9	14	68	341	725	9	7	58	337	746	12	11	108	353	673	10	21	89	355	682
特	29	458	1	6	28	129	294	1	8	30	133	286	－	9	46	118	285	2	9	29	136	282
中堅教諭等資質向上研修																						
小	83	688	4	10	99	307	268	1	12	113	301	261	5	26	139	282	236	6	13	126	301	242
中	48	364	1	7	45	137	174	－	3	55	144	162	－	9	62	135	158	－	11	63	138	152
高	80	643	1	25	129	237	251	－	13	148	229	253	2	27	139	229	246	2	28	151	239	223
特	42	326	－	1	76	139	110	－	8	73	138	107	1	13	69	137	106	－	6	82	137	101

表 5 項目間相関(Spearmanの順位相関係数)			
	研修の有効性	資料の有用性	時間配分の適切性
資料の有用性	.789		
時間配分の適切性	.751	.758	
ニーズとの適合性	.844	.783	.757

次に、項目ごとに全ての研修内容の回答をプールして項目間相関を算出すると、表 5 のとおりそれぞれ強い正の相関を示した。

2) 級内相関係数の算出

以降の分析には R (4.2.2) を用いた。初任者研修・中堅教諭等資質向上研修の各研修内容の反応を校種別にプールすると、そのデータは研修内容ごとにネストされていると同時に、受講者ごとにもネストされた交差的な階層構造となる。このような複数のクラスタレベル間で完全な階層関係が保たれない特殊なマルチレベルデータに適用できる分析モデルとして、交差分類モデルがある (cross-classified model: Goldstein 1994)。Sharma *et al.* (2022) を参考に、交差分類モデルの帰無モデルを用いる方法によって研修内容レベルと受講者レベルで生じている ICC を求めたところ、表 6 に示す値が得られた。なお、4 項目はいずれも分布が右に偏っているため (表 4)、分析には線形混合モデルではなく ordinal パッケージの clmm 関数によるマルチレベル順序ロジスティック回帰モデルを採用した。

Hox (2010) は ICC の効果量について、教育の文脈では ICC=.05 を小、ICC=.10 を中、ICC=.15 を大とする基準を提案している。この基準に照らすと、研修内容レベルの ICC は初任者研修で中程度以下、中堅教諭等資質向上研修では小程度以下となる。初任者研修の方がやや高い傾向にあるのは、初任者の方が付度なくより素直に回答していることを意味するのか、それとも研修内容の特徴がよりばらついていていたことを意味するのかは、この結果だけでは判断できない。

表 6 研修内容レベルと受講者レベルの級内相関係数：偏った回答除外後

	研修の有効性		資料の有用性		時間配分の適切性		ニーズとの適合性	
	研修内容 レベル	受講者 レベル	研修内容 レベル	受講者 レベル	研修内容 レベル	受講者 レベル	研修内容 レベル	受講者 レベル
初任者研修								
小学校	.047	.544	.030	.523	.037	.434	.069	.497
中学校	.105	.450	.077	.485	.076	.443	.116	.498
高等学校	.119	.422	.091	.450	.065	.449	.136	.395
特別支援学校	.104	.522	.084	.491	.051	.448	.138	.480
中堅教諭等資質向上研修								
小学校	.016	.480	.023	.512	.078	.336	.031	.504
中学校	.000	.476	.013	.441	.025	.325	.002	.441
高等学校	.061	.523	.040	.495	.050	.456	.072	.506
特別支援学校	.019	.544	.046	.588	.047	.466	.028	.572

3) 実務条件での級内相関係数

それでは、研修評価の実務に近い条件ではどうだろうか。除外処理を施す前のデータセットを用いて再度計算すると、初任者研修は $.027 \leq \text{ICC} \leq .109$ 、中堅教諭等資質向上研修は $.000 \leq \text{ICC} \leq .062$ となり、研修内容レベルの ICC は概ね小程度以下の水準となった（表 7）。ICC はデータ全体の分散を占めるクラスタレベルの分散の大きさであるため（清水 2014）、研修内容の違いによって説明される分散は11%以下ということになる。

その一方、受講者レベルの ICC は $.429 \leq \text{ICC} \leq .632$ と研修内容レベルをはるかに上回る大きい効果量を示した。反応の分散の42%以上が受講者レベルの分散であることになり、その影響がいかに大きいか分かる。以上の分析結果から、初任教員・中堅教員は研修内容の特徴（e.g., 講師のコミュニケーションの上手さ、研修内での人的相互作用の量：Sitzmann *et al.* 2008）に関わらず、毎回似たような反応評定を行う傾向にあることが示唆され、「反応は研修の特徴を捉えている」（Sitzmann *et al.* 2008）という仮定の成立には疑念が生じた。

また、項目間相関（表 5）は偏った回答を除いた上でもいずれも $r > .70$ と高い値であり、4 項目の背後に何らかの共通因子が存在することが示唆された。

表 7 研修内容レベルと受講者レベルの級内相関係数：偏った回答込み

	研修の有効性		資料の有用性		時間配分の適切性		ニーズとの適合性	
	研修内容 レベル	受講者 レベル	研修内容 レベル	受講者 レベル	研修内容 レベル	受講者 レベル	研修内容 レベル	受講者 レベル
初任者研修								
小学校	.040	.605	.027	.583	.032	.519	.059	.573
中学校	.081	.567	.062	.587	.059	.558	.088	.607
高等学校	.091	.559	.070	.579	.049	.586	.104	.541
特別支援学校	.083	.623	.068	.599	.041	.560	.109	.596
中堅教諭等資質向上研修								
小学校	.013	.564	.019	.593	.062	.456	.025	.589
中学校	.000	.554	.010	.532	.020	.429	.001	.534
高等学校	.050	.604	.033	.576	.041	.548	.058	.595
特別支援学校	.016	.593	.040	.632	.039	.523	.024	.621

3. 調査 2：匿名調査との比較

調査 1 において研修内容レベルの ICC が11%以下にとどまった理由として、研修内容がいずれも真に優れていたために、天井効果が生じた可能性も排除しきれない。もしも反応の背後に教育センターへの付度があるとすれば、匿名調査として測定することで社会的望ましさのバイアスは弱まり（Krumpal

2013)、分布の偏り（表3・表4）や強い項目間相関（表5）は緩和されると考えられる。そこで調査2では研修評価の改善に向けて匿名調査による反応測定を試行し、前年度の記名調査による測定結果と比較することにした。

(1) 方法

1) 対象者と対象研修

調査の対象者は、2023年度に中堅教諭等資質向上研修を受講した公立中学校（国立を含む）・高等学校の教諭である。対象研修は、共通研修5日目の「学校の課題解決に向けた取組の提案」と「自己成長のマネジメント」とした。当該2つの研修内容は、実施時期・時間配分・配付資料等が2022年度から大きく変わっていないことから、比較対象に適していると考えられた。ただし、「学校の課題解決に向けた取組の提案」のテーマは、2022年度が「GIGA スクール構想の実現」のみであったのに対し、2023年度は「GIGA スクール構想の実現」または「PBL（Project Based Learning）」のいずれか一つを受講者自身が選択する方式に変更されていた。いずれの研修内容も協議演習が主体であるが、配分された研修時間の違いから（表2）、研修内での人的相互作用の量（Sitzmann *et al.* 2008）は大きく異なった。

岡山県総合教育センターは中堅教諭等資質向上研修の受講者に対し、一連の研修課程から得る学びを活かして、所属校の課題解決に向けた取組提案書を作成することを課している。共通研修5日目の「学校の課題解決に向けた取組の提案」は、作成した提案書の最終発表会という位置づけであり、この日に至るまでに中間発表会（共通研修4日目）が設けられるなど受講者は事前に相応のリソースを投入して臨む。この研修内容には3時間35分の研修時間が配分され、受講者は6・7名のグループ内で相互に発表とフィードバックを繰り返した（1人30分または25分）。

「自己成長のマネジメント」は、目指す教員像を確立し、実現に向けた手立てを見出すことを目的として行うグループ協議主体の研修内容である。受講者は事前にチェックシートを用いて現在の資質能力を再認識しておき、研修当日はその内容をグループ内で共有することでより深い省察を目指す。研修時間の配分は説明と個人思考の計約15分を含めて1時間であり、受講者同士の相互作用にあてられた時間（約45分）は、「学校の課題解決に向けた取組の提案」と比べると4分の1程度である。

2) 調査の実施方法

調査は前年度（2022年12月実施）と同じ手続きによって2023年12月に行われたが、大きく異なったのは調査フォームの構造である。受講者は記名式のWeb調査フォームで研修の振り返りを入力した後、フォーム上のハイパーリンクから匿名式の別フォームに遷移して本調査に協力する形となった。別フォームの冒頭では、調査の目的とともに「この調査フォームは匿名（無記名）です。率直な感想をお聞かせください。」という教示を示した。調査の結果、対象受講者125名（中学校54名、高等学校71名）のうち、113名（中学校49名、高等学校64名）から回答協力が得られた。

3) 調査内容

用いた項目は調査1と全く同じである。

(2) 結果と考察

1) 平均値の比較

以下の集計・分析には HAD18 (清水 2016) を用いた。研修評価の実務に近い形で反応の測定結果を比較するため、集計は偏った回答の受講者を除外せずに実施した。各項目の記述統計と t 検定の結果を表 8 に示す。中学校・高等学校とも、全ての項目で2023年度の反応は2022年度を下回った。一部には統計的に有意な差も見られ、効果量は最大 $d=-0.52$ と中程度であった。

反応が悪化した理由は三つ考えられる。一つ目は付度のバイアスが緩和されたこと、二つ目は受講者の性質が2022年度と異なっていたこと、そして、三つ目は単純に研修内容が悪くなったことである。まず、付度のバイアスが緩和されたのだとすれば、本研究が懸念するとおり記名調査では妥当な測定結果が得られないことを意味する。次に、受講者の異質性が原因だとすれば、そもそも今回用いた反応項目は研修改善に用いる指標としては評定者間信頼性に欠けることを意味する。最後に研修内容が悪化したかどうかであるが、一般論として、前年度を踏襲した研修を実施する場合には、前年度の反省に基づく研修改善に取り組むものであり、効果測定はそのために行われている。時間配分や資料はほぼ同じでも、担当の指導主事らは当日に至るまでの段取りやグループ割、当日の口頭説明など細かな点を調整してよりよい研修となるように努める。このことを踏まえると、2023年度の変化は全体的に逆方向のトレンドであり違和感がある。

表 8 記名調査と匿名調査の比較

	学校の課題解決に向けた取組の提案				自己成長のマネジメント			
	2022年度:記名 <i>Mean</i> <i>SD</i>	2023年度:匿名 <i>Mean</i> <i>SD</i>	効果量 d (t 検定)		2022年度:記名 <i>Mean</i> <i>SD</i>	2023年度:匿名 <i>Mean</i> <i>SD</i>	効果量 d (t 検定)	
中学校								
研修の有効性	4.47 0.68	4.43 0.68	-0.06		4.22 0.71	4.12 0.83	-0.13	
資料の有用性	4.37 0.67	4.14 0.74	-0.32		4.14 0.74	3.94 0.85	-0.25	
時間配分の適切性	4.43 0.71	4.18 0.91	-0.30		4.18 0.81	3.78 0.98	-0.45 *	
ニーズとの適合性	4.37 0.67 <i>n</i> =49	3.96 0.87 <i>n</i> =49	-0.52 *		4.16 0.77 <i>n</i> =49	3.84 0.69 <i>n</i> =49	-0.44 *	
高等学校								
研修の有効性	4.21 0.86	4.20 0.80	-0.01		4.12 0.87	3.92 0.90	-0.22	
資料の有用性	4.22 0.86	4.17 0.75	-0.06		4.09 0.89	3.78 0.88	-0.35 *	
時間配分の適切性	4.27 0.87	4.20 0.78	-0.08		4.14 0.87	3.81 0.87	-0.37 *	
ニーズとの適合性	4.17 0.87 <i>n</i> =86	4.11 0.78 <i>n</i> =64	-0.08		4.06 0.95 <i>n</i> =86	3.86 0.96 <i>n</i> =64	-0.21	

* $p < .05$

2) 項目間相関

次に、記名調査と匿名調査の項目間相関の違いを確認するため、独立した2つのグループの相関係数の差の検定を行った。その結果、予想どおり匿名調査とした2023年度の方がいずれの項目の組み合わせでも相関係数は低い値となり、その差は多くの組み合わせについて統計的に有意であった (表 9)。表 8 において反応の悪化幅が比較的小さかった高等学校の「学校の課題解決に向けた取組の提案」でも、全ての組み合わせで有意な相関係数の低下が生じている。この変化は、受講者が各反応項目に対してより正直に回答した結果と考えれば解釈が容易である。

なお記名調査では $r > .90$ となるなど強い相関が生じており、これでは項目を複数用いる意味がない。当該4項目を、研修の反応という一概念を捉える尺度とみなすにしても、心理尺度において望ましい項目間相関は $r = .15$ から $r = .50$ の範囲とされており (レビューとして、高本・服部 2015)、匿名調査の項目間相関でもまだ高すぎる値である。

表9 相関の差の検定

	学校の課題解決に向けた取組の提案			自己成長のマネジメント		
	2022年度	2023年度		2022年度	2023年度	
中学校	<i>r</i>	<i>r</i>	Δr	<i>r</i>	<i>r</i>	Δr
研修の有効性・資料の有用性	.759	.460	-.300 *	.888	.687	-.201 **
研修の有効性・時間配分の適切性	.829	.549	-.281 **	.865	.644	-.220 **
研修の有効性・ニーズとの適合性	.851	.706	-.145 +	.912	.764	-.149 *
資料の有用性・時間配分の適切性	.719	.616	-.102	.866	.604	-.261 **
資料の有用性・ニーズとの適合性	.906	.500	-.406 ***	.947	.659	-.288 ***
時間配分の適切性・ニーズとの適合性	.763	.648	-.115	.918	.683	-.235 ***
	<i>n</i> =49	<i>n</i> =49		<i>n</i> =49	<i>n</i> =49	
高等学校	<i>r</i>	<i>r</i>	Δr	<i>r</i>	<i>r</i>	Δr
研修の有効性・資料の有用性	.880	.684	-.196 **	.940	.822	-.118 ***
研修の有効性・時間配分の適切性	.932	.594	-.338 ***	.892	.856	-.036
研修の有効性・ニーズとの適合性	.946	.829	-.116 ***	.956	.875	-.081 **
資料の有用性・時間配分の適切性	.892	.757	-.135 **	.895	.856	-.039
資料の有用性・ニーズとの適合性	.891	.731	-.160 **	.898	.810	-.089 *
時間配分の適切性・ニーズとの適合性	.913	.720	-.193 ***	.873	.806	-.067
	<i>n</i> =86	<i>n</i> =64		<i>n</i> =86	<i>n</i> =64	

*** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$

4. 総合考察

(1) 「反応は研修の特徴を捉えている」のか

本研究の第一の意義は、記名調査では正直な回答が得られないという Kirkpatrick and Kirkpatrick (2006) の主張を検証することにある。今も昔も、研修評価研究におけるレベル1を巡る議論は、4レベル間の因果関係を中心に行われている。この議論において、測定結果が研修の特徴を捉えていることは言わば暗黙の前提となっており、調査が匿名で行われたかどうかにはほとんど注意が払われてこなかった。Kirkpatrick and Kirkpatrick (2006) の共著者である James D. Kirkpatrick の後の著書にすら、記名を必須とするアンケート様式の使用事例が掲載されている (Aggett 2016; Steer 2016)。

調査1では、プールした反応データ全体に占める研修内容レベルの分散が11%以下である一方、受講者レベルの分散は42%を超える高い水準にあることが分かった (表7)。そして調査2では、記名調査において大きく偏っていた分布や極めて高い項目間相関は、匿名調査とすることである程度緩和されることが示された (表8・9)。これらの結果は Kirkpatrick and Kirkpatrick (2006) の主張を支持するものであり、「反応は研修の特徴を捉えている」(Sitzmann *et al.* 2008) という仮定は、教員研修の反応を記名調査によって測定する場合には成立しないことを示唆している。4段階モデルに関する将来のメタ分析においては、反応測定時の匿名性を調整変数の一つとして考慮するべきであろう。また、Sitzmann *et al.* (2008) が示した反応と研修の特徴 (講師のコミュニケーションの上手さ・研修内での人的相互作用の量) との高い正の相関にも、忖度特性がもたらす疑似相関が含まれている可能性がある。Sitzmann 自身、受講者間の異質性のバイアスを問題視し、近年は受講者個人内の変化に着目した研修評価を行うべきだと提案している (Sitzmann and Weinhardt 2019)。

第二の意義は、教員研修評価の実務で広く採用されている記名調査に疑問を投げかけることにある。今年の反応が去年より良いのは研修改善の成果ではなく、今年の受講者に「大人」が多いだけかもしれない。新たに構築された教員研修プラットフォームにおいてレベル1を測定する場合も、受講者のアカウント情報と紐づくような匿名性の低い仕組みでは、妥当な測定結果は得られない可能性が高い。

研修評価の実務においては、レベル1とレベル2を研修直後に同一調査フォームで測定すると手間が省ける (上岡 2022)。しかし、忖度のバイアスを抑えるためには、少なくとも調査2で試行したようにレベル1とレベル2の測定に別々の調査フォームを用い、レベル1の調査フォームを匿名とする対応を

検討する必要がある。あるいはレベル2もまとめて匿名調査とすることも考えられるが、知識技能を測定する際の真剣さに影響する懸念があるほか、学びの不十分な受講者に対する個別のフォローアップが不可能になるという課題がある。

(2) 今後の課題

本研究の一連の調査結果は、記名調査から得られた反応の測定結果が妥当でないことを示すに止まっており、匿名調査の結果が妥当であることを証明するものではない。現に、匿名で実施された調査2においても項目間相関は依然として高い水準にあり、何らかのバイアスの影響が疑われる。また、本研究で扱った反応データは、複数の事業を通じて学校・教職員と接点を持つ教育センターが収集したものであって、受講者にとっては作為回答を行うための十分な動機づけのある測定環境であったと言える。同様の社会的文脈がない、例えば大学主催の研修を自ら希望して受講するような場合でも今回の結果が一般化されるのかについては、更なる検討が望まれる。

謝辞

調査に御協力いただいた研修受講者の皆様と、調査結果をまとめる上で重要な示唆を賜った環太平洋大学の安永和央准教授に厚く御礼申し上げます。

参考文献

- Aggett, P. (2016) Common practice in leadership program: Greencore Northampton. In Kirkpatrick, J. D. and Kirkpatrick, W. K. *Kirkpatrick's Four Levels of Training Evaluation*. ATD Press, pp.157-166
- Alliger, G. M. and Janak, E. A. (1989) Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2): 331-341
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., and Shotland, A. (1997) A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50(2): 341-358
- Blume, B. D., Ford, J. K., Baldwin, T. T., and Huang, J. L. (2010) Transfer of training: A meta-analytic review. *Journal of Management*, 36(4): 1065-1105
- Blume, B. D., Ford, J. K., Surface, E. A., and Olenick, J. (2019) A dynamic model of training transfer. *Human Resource Management Review*, 29(2): 270-283
- 独立行政法人教職員支援機構 (2018) 教職員研修の手引き2018—効果的な運営のための知識・技術—. https://www.nits.go.jp/materials/text/files/index_tebiki2018_001.pdf (参照日 2023.08.31)
- Ebner, C., and Gegenfurtner, A. (2019) Learning and satisfaction in webinar, online, and face-to-face instruction: A meta-analysis. *Frontiers in Education*, 4: 00092. <https://doi.org/10.3389/feduc.2019.00092> (accessed 2023.08.31)
- Goldstein, H. (1994) Multilevel cross-classified models. *Sociological Methods & Research*, 22(3): 364-375
- 平松陽一 (2010) 教育研修の効果測定と評価のしかた (改訂新版). 日興企画
- 堀尾志保, 高橋潔 (2004) 作為回答場面での5大因子性格検査に関する社会的望ましき尺度の役割. 産業・組織心理学研究, 17(2): 65-77
- Hox, J. J. (2010) *Multilevel analysis: Techniques and applications (2nd ed.)*. Routledge

- 兵庫県立教育研修所(2009)研修の事業評価システムの構築に関する研究—研修の有効性を評価する—。
兵庫県立教育研修所紀要, 119: 1-14
- 磯部征尊, 伊藤大輔, 倉本哲男 (2023) 学び続ける教員像の確立に向けた研修体系表の提案と研修プログラムの検証—愛知教育大学教員研修連絡協議会の概算プロジェクトを通して—。学校改善研究紀要, 5: 41-55
- Kirkpatrick, D. L. and Kirkpatrick, J. D. (2006) *Evaluating Training Programs: the four levels (3rd ed.)*. Berrett-Koehler
- 小清水貴子, 藤木卓, 室田真男 (2016) ICT 活用推進リーダーを対象にした集合研修の改善と評価。日本教育工学会論文誌, 40 (2): 113-126
- Krumpal, I. (2013) Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity: International Journal of Methodology*, 47(4): 2025–2047
- 中原淳, 関根雅泰, 島村公俊, 林博之 (2022) 研修開発入門—「研修評価」の教科書—。ダイヤモンド社
- 奈良県立教育研究所 (2008) 教員研修の在り方—研修の工夫・改善—。平成19年度研究紀要。
https://www.e-net.nara.jp/kenkyo/index.cfm/21,2469,c,html/2469/02_kiyou_h19.pdf (参照日 2023.08.31)
- 大阪府教育センター (2020) 研修評価についてのアンケート結果概要。 https://www.osaka-c.ed.jp/category/forteacher/investigate/publication/r02/R02_kikaku_summary.pdf (参照日 2023.08.31)
- Saks, A. M., and Burke, L. A. (2012) An investigation into the relationship between training evaluation and the transfer of training. *International Journal of Training and Development*, 16(2): 118–127
- Sharma, R., Dwivedi, L. K., Jana, S., Banerjee, K., Mishra, R. *et al.* (2022) Survey implementation process and interviewer effects on skipping sequence of maternal and child health indicators from national family health survey: An application of cross-classified multilevel model. *SSM - Population Health*, 19: 101252. <https://doi.org/10.1016/j.ssmph.2022.101252> (accessed 2023.08.31)
- 清水裕士 (2014) 個人と集団のマルチレベル分析。ナカニシヤ出版
- 清水裕士 (2016) フリーの統計分析ソフト HAD—機能の紹介と統計学習・教育, 研究実践における利用方法の提案—。メディア・情報・コミュニケーション研究, 1: 59-73
- Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., and Zimmerman, R. (2008) A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, 93(2): 280-295
- Sitzmann, T., and Weinhardt, J. M. (2019) Approaching evaluation from a multilevel perspective: A comprehensive analysis of the indicators of training effectiveness. *Human Resource Management Review*, 29(2): 253–269
- Steer, L. (2016) Sales graduate program pilot: ArjoHuntleigh Getinge group. In Kirkpatrick, J. D. and Kirkpatrick, W. K. *Kirkpatrick's Four Levels of Training Evaluation*. ATD Press, pp.174-190
- 鈴木克明 (2015) 研修設計マニュアル—人材育成のためのインストラクショナルデザイナー—。北大路書房
- 高本真寛, 服部環 (2015) 国内の心理尺度作成論文における信頼性係数の利用動向。心理学評論, 58(2): 220-235
- 徳村朝昭, 前川朝康, 鈴木克明 (2006) ICT スキル研修改善のためのアクションプラン導入手法。日本

教育工学会論文誌, 30 (3): 203-212

堤宇一, 久保田享, 青山 征彦 (2007) はじめての教育効果測定—教育研修の質を高めるために—. 日科
技連

上岡伸 (2022) 改訂版キーワード法による研修効果測定の妥当性に関する検討. 日本教育工学会論文誌,
46(2): 289-302

上岡伸, 中藤路子 (2024) ハイフレックス型研修の学習効果は受講形態によって差があるか. 日本教育
工学会論文誌, 早期公開版, DOI: <https://doi.org/10.15077/jjet.47090>

全国教育研究所連盟 (2007) 平成18年度教育課題調査結果.

<https://www.nier.go.jp/zenkyou/zenkyou/chousa/06chousahokoku.pdf> (参照日 2023.08.31)

令和 6 年 3 月発行

岡山県総合教育センター 研究紀要 第 14 号 研究番号 23-03

「教員研修受講者の反応は記名調査によって測定可能か」