

子どもアバターとの模擬面接とフィードバックを組み合わせた 面接訓練が警察官の発問スキルに与える影響

萩野谷俊平*

The Effect of Simulated Interviews with Child Avatars Combined with Feedback on Police Officers' Questioning Skills

Shumpei HAGINOYA*

We tested the efficacy of a training protocol consisting of simulated child sexual abuse interviews with feedback in improving interviewers' questioning skills. Twenty-one police officers were randomly assigned to a four interviews or eight interviews group. The four interviews group conducted four simulated interviews while receiving feedback after each interview. The eight interviews group conducted four interviews without feedback, followed by another four interviews together with feedback. The feedback consisted of information of the outcome (i.e., what memories of the event the avatar really had) and of feedback on the quality of the questions asked in the interviews. The results showed that the interviews together with the feedback (vs. the ones without feedback) increased the proportion (four interviews group: from 49% to 82%, eight interviews group: from 41% to 85%) of recommended questions as a function of the number of interviews. Meanwhile, the number of relevant and wrong details elicited from the avatars was not clearly improved as the questioning skills.

key words: child sexual abuse, investigative interviewing, forensic interviewing, simulation training, avatar

問題と目的

子どもに対する性的虐待では、子どもの証言が唯一の証拠であることが少なくない (Lamb, 1995) にもかかわらず、被害者である子どもの返答には、経験した出来事の前後に与えられた情報の影響を受けやすい傾向があり (Ceci & Bruck, 1993)、事実と異なることを尋ねる質問 (e.g., 犬のぬいぐるみを受け取った子供に「あなたは猫のぬいぐるみをもらいましたか」と尋ねる) や正しい選択肢を提示しない選択

式の質問、さらにはオープン、クローズドを問わず同じ質問を繰り返すといった面接方法によって、証言が歪められる可能性が指摘されている (Finnilä et al., 2003; Krähenbühl et al., 2009)。一方で、子どもはオープン質問 (e.g., 「お話しして」、「それから」) に対して正確な情報を提供できることが先行研究から示されている (Allwood et al., 2008)。

子どもから歪みのない情報を引き出す面接を実現するために、オープン質問を主体とする NICHD プロトコル (Hershkowitz et al., 2007) などの面接手法

* 明治学院大学

Faculty of Psychology, Meiji Gakuin University, 1-2-37 Shirokanedai, Minato-ku, Tokyo 108-8636, Japan.
(haginoya@psy.meijigakuin.ac.jp)

が開発され、その有効性も実証されてきた (Orbach et al., 2000)。推奨される手法に基づいて面接を実施するためには、適切な訓練の受講が必要である。これまでに面接技術の習得と維持には、集中的な訓練だけではなく、その後の継続的なスーパーヴァイズも重要であることが示されてきた (Lamb et al., 2002)。しかし、これには人的・時間的に大きなコストを要するという課題がある。

この課題を考慮して、子どものアバターを使用した面接のシミュレーション訓練であるアバタートレーニング (Avatar Training) が開発された (Pompedda et al., 2015)。アバタートレーニングは1時間から2時間程度で実施される訓練プロトコルであり、受講者は、被面接者のアバターを使用した実践的な面接演習を通じて、比較的低位年齢 (4歳と6歳) の子どもの面接における被害の有無を聴取する段階について、オープン質問を主体とした面接技術を学習する。受講者は、アバターとの模擬面接において、実際の子どもと面接するように口頭で自由に質問をして、情報を引き出すことができる。アバタートレーニングに関する最初の研究である Pompedda et al. (2015) は、模擬面接に続いて、事案の結末 (虐待ありの事例では加害者による犯行内容を説明する文章、虐待なしの事例では虐待が疑われた状況で実際はどのようなことが起きていたかを説明する文章) と質問方法 (面接中に使用された望ましい質問2問と望ましくない質問2問の計4問について該当する質問タイプの説明およびその質問で得られる返答の信頼性に関する情報) に関するフィードバック (いずれのフィードバックもテキスト形式で提示された) の組み合わせが、フィードバックを受け取らずに面接を繰り返す場合に比べて、望ましい質問の増加と、はい・いいえ質問や暗示的な質問といった望ましくない質問の減少をもたらすことを示した。その後も一連の研究が行われており、上記のフィードバックをともなう模擬面接を重ねるほど、発問スキルが向上することが示されている (Krause et al., 2017; Zhang et al., 2023)。

子どもからより多くの開示を引き出す上で、オープン質問が重要であることは日本においても同様である (仲, 2012)。日本では NICHHD プロトコルに基づく研修が2007年から実施されており、2014年までの期間で2500人以上が参加している (仲, 2015)。し

たがって、日本においてもオープン質問を主体とした面接技術を身につける訓練および維持のためのツールは重要であり、実践的な模擬面接を低コストで実施できるアバタートレーニングを活用できると考えられる。日本におけるアバタートレーニングでは、大学生 (Haginoya et al., 2020) や公認心理師・臨床心理士 (Haginoya et al., 2021) を対象として訓練の有効性が示されており、近年は警察官を含むサンプルを対象とした訓練でも発問スキルの改善 (望ましい質問の使用の増加) を示している (Haginoya et al., 2023)。

警察官は、子どもの心理的負担等に配慮して検察、警察、児童相談所といった多機関が連携して行う司法面接の90%以上に関わっており、聴取者として面接を行うことが検察官に次いで多いことから (法務省, 2022)、面接訓練の必要性が特に高いと言える。しかし、警察官を訓練した先行研究 (Haginoya et al., 2023) が実施した模擬面接は2回であり、介入としてのフィードバックも1回 (模擬面接間で提供される一連のフィードバックを1回とした場合) しか行われていない。上記の通り、フィードバックを用いたアバタートレーニングについてはフィードバックの回数を重ねることでより大きな発問スキルの向上が示されているが、警察官についてはこうした検討は不十分である。

そこで本研究では、フィードバックを用いた模擬面接訓練について、複数回のフィードバックを行うことによる訓練効果を検討することを目的とした。具体的には、以下の3つの仮説を検証した。まず、複数回のフィードバックを行うアバタートレーニングの訓練効果を調べるため、以下の2つを検討した。

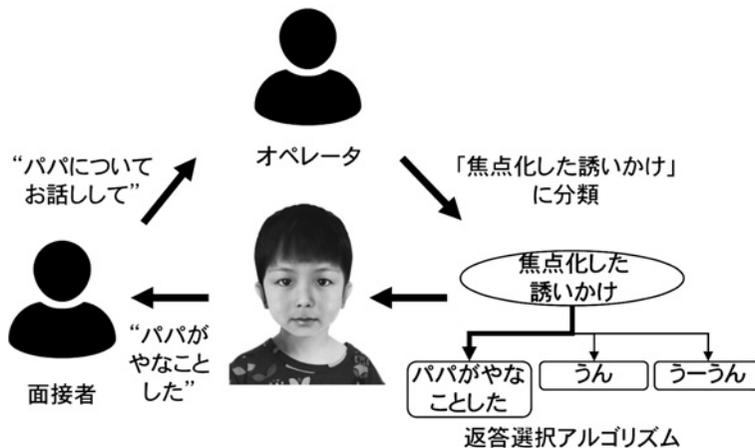
仮説1. フィードバックをともなう模擬面接において、フィードバックを行わない場合に比べて望ましい質問の割合が増加する。

仮説2. フィードバックをともなう模擬面接により発問スキルが改善することで、フィードバックを行わない場合に比べてアバターから引き出される関連情報 (事案の解明に関連する情報) の数が増加し、アバターが面接者に誘導されて誤情報を生み出す数が減少する。

また、フィードバックを重ねることによる訓練成果の変化を調べるため、以下の仮説を検討した。

仮説3. 望ましい質問の割合および引き出された

Figure 1 アバターとの模擬面接の流れ



関連情報と誤情報の数におけるフィードバックの有無の差の効果量は、面接回数を重ねるほど大きくなる。

方 法

参加者

27歳から43歳 ($M_{age} = 35.5, SE = 5.2$) の警察官21名(女性5名)が面接訓練の一環として実験に参加した。15名(71%)が子どもへの面接に関する何らかの研修の受講経験を有していた。

材料

実験では、アバターとの模擬面接を実施するアバタートレーニング専用のソフトウェアを使用した。ソフトウェアには、年齢(4歳, 6歳), 性別(女性, 男性), 虐待被害の有無(あり, なし)が異なる8つの条件についてそれぞれ2体のアバターで構成される, 合計16体のアバターが実装されている。各アバターは、虐待被害を調査する面接を受けるに至った背景のシナリオ(アバターの年齢・性別, 家族構成, 住居・生活環境, 虐待が疑われるに至ったきっかけを含む200字から300字程度の文章), 事案の結末, および面接者の質問に対して準備された返答のビデオクリップを持つ。アバターとの模擬面接の流れをFigure 1に示す。アバターは、質問タイプ(誘いかけ, はい・いいえ質問など)と子どもの返答の関係を検討した過去の心理学的知見に基づく返答選択アルゴリズムによって動作する(Haginoya et al., 2020, 2021)。模擬面接中に、オペレータは参加者が尋

ねた質問を聞き取って特定の質問タイプに分類し、分類結果をソフトウェアに入力する。返答選択アルゴリズムには質問タイプごとに複数の返答とそれぞれの選択確率が設定されており、入力された質問タイプに対応する返答のうち1つが確率に従って選択され、返答のビデオクリップが再生される。

手続き

各実験は2時間から3時間を要した。参加者は最初に同意書に署名し、属性(性別, 年齢など)に関する質問に回答した。次に、子どもに対する望ましい面接方法に関する教示文を確認し、教示の理解を確認する2つの質問に回答した。教示文は、望ましい質問と望ましくない質問について解説する900字程度の文章であり、ソフトウェア上にテキスト形式で参加者に提示された。教示の理解を確認する質問もソフトウェア上にテキスト形式で提示され、参加者はプルダウン式の回答欄で選択肢を選んで回答した。例えば「子どもが虐待の事実について、『男の人が私のこと叩いた』と述べた場合、それに対して以下のどの質問をするのが最も適切だと思いますか。」という質問に対して「痛かった?」「誰が叩いたの?」「その男の人はお父さんかな?」という3つの選択肢が提示された。回答はソフトウェアで処理され、2問のうち一つでも不正解があった場合はその旨がメッセージとして表示されて次の段階へ進むことができず、2問すべてに正解した時点で次の段階へ進むためのボタンが表示された。

模擬面接の実施にあたり、参加者は4回面接群(n

= 11) と 8 回面接群 ($n = 10$) の 2 群にランダムに割り当てられた。なお、子どもの面接に関する研修の受講経験がある 6 名は、4 回面接群と 8 回面接群に 3 名ずつ割り当てた。4 回面接群では、参加者は 1 回目から 4 回目の各面接後に、事案の結末と質問方法に関するフィードバックを受け取った。事案の結末に関するフィードバックは、虐待ありの事例では、どのような犯行が行われていたかを説明する文章、虐待なしの事例では、虐待が疑われた状況で実際はどのようなことが起きていたかを説明する文章が、それぞれ 100 字前後で専用ソフトウェア上にテキスト形式で提示された。質問方法に関するフィードバックは、直前の模擬面接で使用した質問（望ましい質問 2 問と望ましくない質問 2 問の計 4 問）について、該当する質問タイプの説明およびそのタイプの質問を使用して子どもから得られる返答の信頼性に関する情報（e.g., 「この質問は『誘いかけ』と言って、子どもから自由な発話を引き出せる質問です。信頼できる返答が期待できます。」）が専用ソフトウェア上にテキスト形式で提示された。フィードバックに使用する質問は、それ以前のフィードバックで取り上げたものとは異なる質問タイプを優先的に選択した。なお、これらのフィードバックは、アバタートレーニングに関する先行研究と同様の手続きであった。8 回面接群では、参加者は 1 回目から 4 回目の面接を、フィードバックを受け取らずに実施し、5 回目から 8 回目の面接を、各面接後に事案の結末と質問方法に関するフィードバックを受け取りながら実施した。本研究が用いた 4 回面接群と 8 回面接群の手続きは、参加者である警察官が業務で子どもへの面接に関わる可能性があり、業務上の訓練として受講する上ですべての参加者へ同等の訓練を施すことが求められたことから、参加者が所属する機関と協議の上、すべての参加者が 4 回のフィードバックを受けられる条件を満たすものとして採用されたものである。そこで本研究では、4 回面接群における 1 回目から 4 回目および 8 回面接群における 5 回目から 8 回目の面接結果をフィードバック条件として、それぞれについて 8 回面接群における 1 回目から 4 回目の面接結果を統制条件とした比較を行った。

面接前の参加者には、アバターに対して自由に口頭で質問を行うこと、面接は虐待の調査に焦点を当てて行うこと、面接は被害事実を聴き出す段階から

始まること、面接は 10 分間であることが教示された。各模擬面接では、16 体のうち参加者がまだ面接を実施していないアバターをランダムに選択した。参加者は、アバターの背景シナリオを確認した後、10 分間の面接を実施した。

参加者が尋ねた質問の分類は、先行研究（e.g., Korkman et al. 2006）にしたがった類型（Table 1）を用いて 1 名のオペレータが実施した。関連情報（事案の解明に関連する情報）は、望ましい質問に対してのみ提供された。各アバターは関連情報を 9 つ持っており、1 回の返答で 1 つずつ提供された。関連情報は決められた順序で提供され、最後の 4 つに事案の解明に必要な情報が含まれていた。例えば、義理の父親から虐待を受けていたアバターの事例では、最後の 4 つの情報として「父親が秘密にしようと言ったこと」「父親と二人きりで動画を撮ったこと」「二人とも服を脱いで、父親が児童にキスをしたこと」「父親が児童の性器に口で触れたこと」が提示された。また、虐待なしの事例では、例えば、児童の父親が再婚した女性（児童の義母）について、児童と別居する実母が「面会時に児童から『義母にやりたくないことを無理にやらされている』と聞き、それは性的な虐待だと言っていた」と申告した事案で、最後の 4 つの情報として「義母が児童に対して弟の体が汚れたときの洗い方を教えたこと」「児童の家に、実父母が一緒に写った写真がたくさんあること」「実母が怖いのでやりたくなかったこと（この時は何をしたのかは明言しない）」「実母の写真をすべて捨てたこと」が提示された。こうした最後の 4 つの関連情報に事案の解明に必要な情報を含める構成は、望ましい質問の使用を訓練する観点から、望ましい質問をある程度の回数使用した段階で事案の解明に必要な情報が得られるようにするために設定された。

参加者が望ましい質問を尋ねたときに関連情報が提供される確率は、4 歳で 12.5%、6 歳で 25% に設定されていた。これらの数値は、先行研究が示した望ましい質問に対する子どもの反応に関する知見をもとに算出された（cf. Korkman et al., 2006; Lamb et al., 2018）。したがって、アバターの年齢によって関連情報を収集する難易度が異なる。本研究で面接に用いた 4 歳のアバターの比率は、統制群として扱った 8 回面接群（1-4 回目）が 42.5%、フィードバックを受け取る群として扱った、4 回面接群が 45.5%、8 回面

Table 1 各質問タイプの説明と例

質問タイプ	説明	質問の例
望ましい質問		
誘いかけ	誘導的でなく子どもから自由な発話を引き出す質問。	「最初から最後まで、あったことを全部教えて」「家族についてお話しして」
促し	子どもの返答に対して、さらなる発話を促す非誘導的な質問。発話内容の明確化を求める場合もある。	「その後何があったの」「続けて」「それから」
WH	子どもがすでに言及した内容について、Where (場所), Who (誰), What (何), How (どのように)を尋ねる質問。	「ママとどこに行ったの」「何のゲームをしたの」
望ましくない質問		
はい・いいえ質問	子どもがまだ言及していない事実に子どもの注意を向けさせるが、特定の種類の返答を暗示しないクローズド質問。	「パパと遊ぶの」
明確な誘導質問	子どもがまだ言及していない事実を仮定して、どのような返答を期待しているかを示す質問。	「パパが嫌なことをしたの」「パパは悪い人」
不明確な誘導質問	子どもがまだ言及していない事実を仮定せず、どのような返答を期待しているかを示す質問。	「お話しすることあるのは分かっているよ、お話しして」
繰り返し質問	望ましい質問または望ましくない質問について、2回以上連続で同じことを尋ねる質問。	
長すぎる質問	二つ以上の内容について一度に尋ねる質問。	「お父さんと一緒にどこにいたの、あとそこでどんなことしてたの」
あいまいな質問	子どもの認知レベルでは難しすぎる単語を含む、または文法的に不明確な質問。	「ママとパパの関係はどんな感じ」
選択式質問	特定の返答に子どもの注意を向けさせる、または選択肢から選ぶよう強制する質問。	「サッカーの練習は貫太君と行ったの、それとも三浦さん」
時間に関する質問	6歳以下の子どもでは発達が不十分な時間的な認知プロセスに依拠した質問。	「いつママは公園を離れたの」
感情に関する質問	子供がまだ言及していない感情の状態を仮定した質問。	「あなたが遭った悲しいことについて教えて」
想像力を求める質問	子どもの空想を促し、不正確な答えを生み出す可能性がある質問。	「あなたがパパだったらどうすると思う」

接群 (5-8 回目) が 57.5% であった。

関連情報以外の返答には、虐待と関係のない事実に関する非関連情報 (e.g.「絵を描くのが好きなこと」「父親や兄と一緒に住んでいること」) やその他の一般的な返答 (e.g., 尋ねられた事実を肯定する『うん』, 否定する『うーうん』, 『わかんない』) があり、例えば、6歳のアバターに対する誘いかけ質問については、非関連情報に 25%、その他の一般的な返答に 50% が割り当てられていた。また、面接のリアリティを高めるため、各アバターの非関連情報はキーワードごとにまとめてリスト化され、キーワードに応じた返答が返されるよう調整されていた。例えば参加者が「おじいちゃんといつも何をしていますか?」と質問した場合、キーワード「おじいちゃん」に該当する返答が選択された。

誤情報(アバターが保有する情報に矛盾する情報)

は、参加者が尋ねた望ましくない質問に対するアバターの返答が、アバターが保有する情報として事前定義された一連の返答と矛盾した場合に、誤情報と判定された。例えば面接対象のアバターが、誰かがアバターを叩いたという情報を持っていないにもかかわらず、面接者の「誰かに叩かれましたか」という質問に対して「うん」と肯定する返答をした場合に、誤情報が生み出されたと判定された。

なお、本研究は、法政大学文学部心理学科・心理学専攻倫理委員会 (No. 18-0048) の承認を得て実施された。

統計分析

仮説 1, 2 を検証するため、面接中の全質問(虐待に関わる質問および虐待と無関係の質問を含む)に占める望ましい質問の割合、関連情報の数、誤情報の数を従属変数として、8回面接群の1回目から4回

目(統制条件)と4回面接群の1回目から4回目(フィードバック条件)の面接結果を用いた2(フィードバックの有無, 参加者間)×4(面接回数, 参加者内)の混合要因の2要因分散分析, および8回面接群の1回目から4回目(統制条件)と5回目から8回目(フィードバック条件)の面接結果を用いた2(フィードバックの有無, 参加者内)×4(面接回数, 参加者内)の参加者内要因の2要因分散分析を実施した。加えて, 仮説3を検証するための計画された比較として, 各従属変数について, 面接回数ごとにフィードバック条件と統制条件を比較する多重比較及びフィードバック条件における面接回数間を比較する多重比較を, Holmの方法により実施した。

結 果

記述統計

1回目の面接における全参加者の平均値は, 4回面接群における望ましい質問の割合が49%, 関連情報の数が3.2 ($SD = 2.0$), 誤情報の数が2.9 ($SD = 3.7$)であり, 8回面接群における望ましい質問の割合が44%, 関連情報の数が2.7 ($SD = 2.0$), 誤情報の数が1.5 ($SD = 1.1$)であった。8回面接群における5回目の面接における平均値は, 望ましい質問の割合が41%, 関連情報の数が2.1 ($SD = 0.4$), 誤情報の数が2.1 ($SD = 0.6$)であった。したがって, フィードバック条件の4回面接群(1-4回目)と8回面接群(5-8回目)におけるフィードバックを受け取る前の発問スキルは, 統制条件の8回面接群(1-4回目)における1回目の発問スキルと類似した水準であった。

面接で使用された質問の数は, 8回面接群の8回の面接の平均が25.3問, 4回面接群の4回の面接の平均が27.8問であった。各群の面接回数ごとの推移(Figure 2)を見ると, 4回面接群のみ面接回数を重ねるごとに質問数がわずかに増加した(1回目: 27.1問, 4回目: 30.0問)が, 全体として明確な増加・減少傾向は認められなかった。

従属変数間の相関をTable 2に示す。望ましい質問の割合は, 関連情報の数と有意な正の相関がみられ, 誤情報の数と有意な負の相関が認められた。したがって本研究が用いた返答選択アルゴリズムは, より多くの望ましい質問の使用に対してより多くの関連情報を提供し, より多くの望ましくない質問の使用に対してより多くの誤情報を生み出すように機能

Figure 2 面接回数ごとの質問数

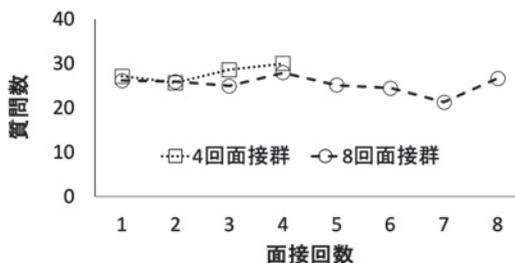


Table 2 各従属変数の平均, 標準偏差および相関係数

	M	SD	1	2
1. 望ましい質問の割合	61.7	25.6		
2. 関連情報の数	4.0	2.5	0.69***	
3. 誤情報の数	1.6	2.0	-0.47***	-0.36***

*** $p < .001$.

していた。

発問方法の改善

分散分析の結果をTable 3に示す。望ましい質問の割合については, 8回面接群と4回面接群の1回目から4回目の比較(8回面接群の1回目と4回面接群の1回目 [Figure 3のA], 8回面接群の2回目と4回面接群の2回目 [Figure 3のB], 8回面接群の3回目と4回面接群の3回目 [Figure 3のC], 8回面接群の4回目と4回面接群の4回目 [Figure 3のD]の比較: 以下, 8回 [1-4] vs. 4回 [1-4]とする), および8回面接群における1回目から4回目と5回目から8回目の比較(8回面接群の1回目と5回目 [Figure 3のE], 8回面接群の2回目と6回目 [Figure 3のF], 8回面接群の3回目と7回目 [Figure 3のG], 8回面接群の4回目と8回目 [Figure 3のH]の比較: 以下, 8回 [1-4] vs. 8回 [5-8]とする)のいずれにおいても, フィードバックおよび面接回数の主効果と, フィードバックと面接回数の交互作用が認められた。

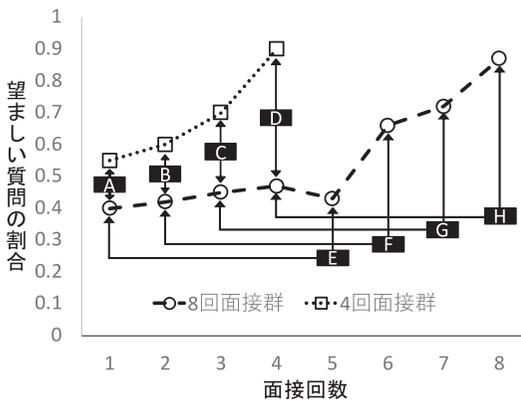
望ましい質問の割合は, フィードバックを受け取る面接を重ねるごとに増加する傾向が見られ (Figure 4A), 単純主効果検定の結果, 8回 [1-4] vs. 4回 [1-4] (2回目: $p = .006$, $d = 1.34$, 3回目: $p < .001$, $d = 1.80$, 4回目: $p < .001$, $d = 1.90$) と8回 [1-4] vs. 8回 [5-8] (2回目: $p = .009$, $d = 1.43$, 3

Table 3 分散分析の結果

	Effect	df _{Effect}	df _{Error}	F	p	η_p^2
8 回面接群 (1-4 回目) vs. 4 回面接群 (1-4 回目)						
望ましい質問の割合	F	1	19	11.72	.003**	.38
	面接回数	3	57	8.51	<.001***	.31
	F×面接回数	3	57	7.21	<.001***	.28
関連情報の数	F	1	19	11.23	.003**	.37
	面接回数	3	57	1.78	.161	.09
	F×面接回数	3	57	1.38	.258	.07
誤情報の数	F	1	19	.10	.754	.01
	面接回数	3	57	.707	.577	.02
	F×面接回数	3	57	2.25	.092+	.11
8 回面接群 (1-4 回目) vs. 8 回面接群 (5-8 回目)						
望ましい質問の割合	F	1	9	27.03	<.001***	.75
	面接回数	3	27	13.67	<.001***	.60
	F×面接回数	3	27	8.05	<.001***	.47
関連情報の数	F	1	9	9.20	.014*	.51
	面接回数	3	27	4.97	.007**	.36
	F×面接回数	3	27	2.85	.056+	.24
誤情報の数	F	1	9	5.65	.041*	.39
	面接回数	3	27	0.91	.451	.09
	F×面接回数	3	27	2.88	.054+	.24

*** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$.
 F: フィードバック

Figure 3 分散分析で行った比較の例



1 回目: $p = .004$, $d = 1.60$, 4 回目: $p < .001$, $d = 2.29$ のいずれにおいても 1 回目を除くすべての面接回数でフィードバックの有無による有意差がみられたとともに、面接回数を重ねるほど効果量が大きくなった。

また、フィードバック条件における望ましい質問の割合については、それぞれ 4 回面接群(1 回目と 2 回目: $p = .017$, $d = 1.10$, 1 回目と 3 回目: $p = .005$, $d = 1.70$, 1 回目と 4 回目: $p = .002$, $d = 1.75$) と 8

回面接群 (5 回目と 6 回目: $p = .022$, $d = 1.57$, 5 回目と 7 回目: $p = .003$, $d = 1.78$, 5 回目と 8 回目: $p < .001$, $d = 2.41$) で、1 回目 (8 回面接群では 5 回目) と他の面接回数との間に有意差が見られた。

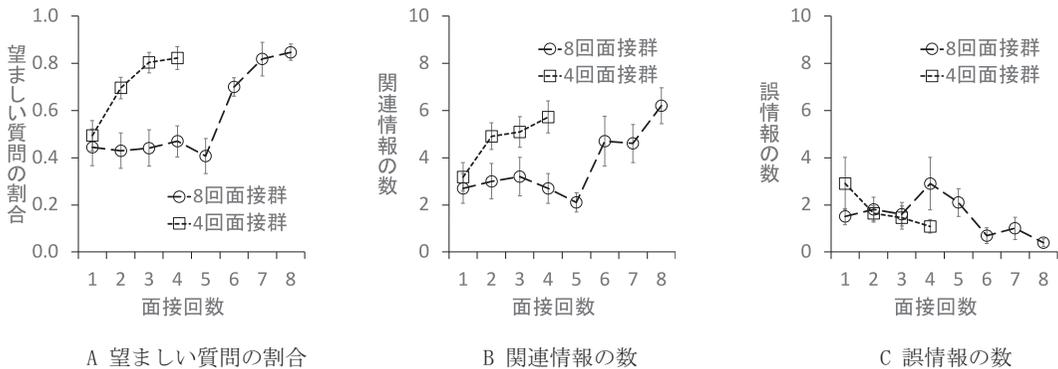
関連情報の数

関連情報の数については、フィードバックの主効果のみ 2 つの分散分析で有意となったが、面接回数的主効果は 8 回 [1-4] vs. 8 回 [5-8] のみ有意、フィードバックと面接回数の交互作用は 8 回 [1-4] vs. 8 回 [5-8] のみ有意傾向となった。

関連情報の数についても、望ましい質問の割合と同様にフィードバックを受け取る面接を重ねることで増加する傾向が見られ (Figure 4B), 単純主効果検定の結果, 8 回 [1-4] vs. 4 回 [1-4] ではフィードバックの有無による差が 2 回目 ($p = .053$, $d = .90$) と 3 回目 ($p = .083$, $d = .80$) で有意傾向, 4 回目 ($p = .004$, $d = 1.42$) で有意となった。8 回 [1-4] vs. 8 回 [5-8] では 4 回目 ($p < .004$, $d = 1.59$) のみフィードバックの有無による差が有意となった。

また、フィードバック条件における関連情報の数については、4 回面接群ではいずれの面接回数の組み合わせにおいても有意差は認められなかったが、8

Figure 4 各条件における面接回数ごとの望ましい質問の割合と得られた関連情報および誤情報の数



回面接群では5回目と7回目 ($p = .010, d = 1.24$), 5回目と8回目 ($p = .009, d = 2.13$) で有意差が見られた。

誤情報の数

誤情報の数については、フィードバックの主効果が8回 [1-4] vs. 8回 [5-8] のみ有意、面接回数の主効果は認められず、フィードバックと面接回数の交互作用はいずれも有意傾向となった。

誤情報の数についても、フィードバックを受け取る面接を重ねることでわずかに減少する傾向が見られたが (Figure 4C), 単純主効果検定の結果, 8回 [1-4] vs. 4回 [1-4] ではいずれの面接回数でもフィードバックの有無による差は有意ではなかった。8回 [1-4] vs. 8回 [5-8] ではフィードバックの有無による差が2回目 ($p = .057, d = .78$) と4回目 ($p = .050, d = .99$) で有意傾向となった。

また、フィードバック条件における誤情報の数については、4回面接群ではいずれの面接回数の組み合わせにおいても有意差は認められなかったが、8回面接群では5回目と8回目 ($p = .070, d = 1.25$) のみで有意傾向となった。

考 察

本研究では、警察官を対象としたアバタートレーニングについて、複数のフィードバックを行うことによる訓練効果を検討することを目的とした。結果から、フィードバックを受け取らなかった統制条件 (8回面接群の1回目から4回目) に比べて、フィードバック条件 (4回面接群の1回目から4回目および8回面接群の5回目から8回目) において、望まし

い質問の割合の増加が認められ、発問スキルの改善が認められた。したがって仮説1が支持された。

変数間の相関係数から、望ましい質問の割合の増加によってアバターからより多くの関連情報が引き出され、反対に誤情報が生み出された数は少なかったことがうかがえる。実際に、面接回数の主効果や交互作用についてはほとんどの分析で有意な結果は得られなかったものの、分散分析の多くで関連情報および誤情報の数についてもフィードバックの主効果がみられたことから、仮説2は部分的に支持された。

望ましい質問の割合に比べて面接で得た情報の数について仮説を明確に支持する結果が得られなかった理由としては、本研究のサンプルサイズが小さかったことに加えて、1つの望ましい質問の使用に対して事案に関する情報を得る確率が比較的少なく設定 (たとえば関連情報は4歳で12.5%, 6歳で25%) されていたために値の範囲が小さくなったこと、また、確率分布を介する返答選択アルゴリズムによって、面接ごとに同じ数の望ましい質問を使用しても得られる関連情報の数は異なる場合があることが背景として考えられる。特にサンプルサイズの問題については、今後より大きなサンプルを用いて複数のフィードバックを行う効果を確認する必要がある。

面接回数ごとのフィードバックの単純主効果は、望ましい質問の割合については2回目以降いずれも有意であり、面接回数を重ねるほど効果量が大きくなった。一方で、関連情報の数では4回目のみフィードバックの有無による差が有意となり、誤情報の数についてはいずれの面接回数についても有意な結果

は得られなかった。また、フィードバック条件における面接回数間の比較においても、望ましい質問の割合については1回目(8回面接群では5回目)と他の面接回数間で有意差が見られたが、関連情報の数では、8回面接群のみ5回目に比べて7回目と8回目が有意に大きく、誤情報の数ではいずれの比較についても有意な結果は得られなかった。したがって、主に望ましい質問の割合の結果から仮説3は部分的に支持された。

従属変数間で結果に違いが見られたことは、上述の分散分析の結果で述べた理由により説明されるものと考えられる。2回目の面接(i.e., 1回のフィードバックを受けた後の面接)で関連情報および誤情報の数に改善が見られないことは、本研究と類似したサンプルサイズ(各群13-15名)で警察官を含むサンプルに対するフィードバックの訓練効果を調べた先行研究(Haginoya et al., 2023)の結果と一致しており、本研究において4回目の面接で関連情報の数にフィードバックの有無による有意差がみられたことは、複数回のフィードバックを重ねることの重要性を示唆する結果と考えられる。

前述の通り、本研究は面接回数を重ねることで望ましい質問の割合が増加する傾向が見られた。この傾向は、日本(Haginoya et al., 2020)、欧州(Krause et al., 2017)、中国(Zhang et al., 2023)の大学生および公認心理師・臨床心理士(Haginoya et al., 2021)を対象としてフィードバックの訓練効果を示した研究でも一貫して示された傾向である。

また、本研究が4回目の面接において80%以上の望ましい質問の割合を示した一方で、フィードバックを用いた先行研究が示した4回目の面接における望ましい質問の割合は、大学生(Haginoya et al., 2020; Krause et al., 2017; Zhang et al., 2023)と公認心理師・臨床心理士(Haginoya et al., 2021)のいずれにおいても60-70%であった。本研究のサンプルである警察官は、いずれも子どもへの聴取を業務内で行う可能性がある職員であり、業務の一環として訓練を受講した。一方で、先行研究が対象とした大学生は現実の面接に関わることはなく、Haginoya et al. (2021)の公認心理師・臨床心理士についても犯罪被害者の子どもと関わる専門職のみが募集されたわけではなかった。これらを考慮すれば、学習への高い動機づけはより高い学習成果に結びつくことから

(Colquitt et al., 2000)、本研究のサンプルにおいては訓練に対する動機づけの高さがより大きな発問スキルの改善に繋がった可能性があると言えるだろう。

本研究が、70%以上が子どもの面接に関する研修の受講経験を有するサンプルにおいて、発問スキルの改善を示したことは重要である。この結果は、子どもへの面接に関する既存の研修を受講した被訓練者に対してもアバタートレーニングが活用できる可能性を示唆している。ただし、被訓練者が受講した研修の多くは、日本で広く用いられており比較的幅広い年齢の子どもを面接対象とするNICHDプロトコルに基づくものである。一方、アバタートレーニングは、発達の比較的早期段階にあり被暗示性も高い低年齢の子どもの面接場面を対象としており、またNICHDプロトコルと直接の関連がない学術研究についても質問分類やアルゴリズム構築に取り入れていることから、そうしたNICHDプロトコルと異なる要素が参加者にとって新たな学習要素となった可能性も考えられる。

アバタートレーニングによる訓練効果が示された一方で、本研究にはいくつかの限界があることに留意する必要がある。

例えば本研究では、アバターの年齢(i.e., アバターから情報を引き出す難易度)、背景シナリオから想定される虐待被害の深刻さといった要因について、面接するアバターをランダムに選択することで統制を試みたが、十分に統制されたとは言にくい。望ましい質問で得られる情報量が少なければその後の面接での質問パターンに影響する可能性があることから、今後の研究ではアバターの年齢を固定した条件でのデータ収集を行う必要もあるだろう。虐待被害の深刻さについては、面接者が背景シナリオから虐待の事実があると仮定して面接に臨んだ場合により多くの望ましくない質問を使用することが示されており(Zhang et al., 2022)、仮定された虐待の深刻さが面接者の質問パターンに影響する可能性が考えられる。本研究では、各アバターのシナリオがどの程度の被害の深刻さを参加者に想起させたかは測定しておらず、その面接への影響を評価することは難しいが、各シナリオから面接者が想定する被害の深刻さと面接への影響については今後検討する必要がある。

また、本研究では参加者について子供の面接に関

する研修の受講経験を収集したが、研修の内容(ロールプレイ等の模擬面接の有無など)や、実際の聴取に関わった経験と被聴取者の人数、捜査経験年数などの情報についても、訓練効果に関連する可能性がある。したがって、今後の研究はこれらの情報も収集し、結果への影響を考慮する必要があるだろう。

さらに、アバタートレーニングにより身につけられるのは誘いかけなどの望ましい質問を面接中により多く使うスキルのみであり、近年面接への有効性が示されている支援的発話 (Ahern et al., 2018; Blasbalg et al., 2018) などのより多様なスキルを身につける訓練への拡張が必要である。また、アバタートレーニングでは比較的低コストで実践的な演習訓練を受講できることから、訓練で向上した発問技術の経時による減衰傾向の確認、およびブースターセッションによる訓練効果の維持を、縦断的研究により検証することも重要である。面接技術の維持に必要な受講間隔や頻度が見出されれば、面接者訓練へ持続的に寄与する手法として確立していくことが期待できる。

引用文献

- Ahern, E. C., Hershkowitz, I., Lamb, M. E., Blasbalg, U., & Karni-Visel, Y. (2019). Examining reluctance and emotional support in forensic interviews with child victims of substantiated physical abuse. *Applied Developmental Science*, 23, 227-238. <https://doi.org/10.1080/10888691.2017.1387057>
- Allwood, C. M., Innes-Ker, Å. H., Homgren, J. & Fredin, G. (2008) Children's and adults' realism in their event-recall confidence in responses to free recall and focused questions. *Psychology, Crime and Law*, 14, 529-547. <https://doi.org/10.1080/10683160801961231>
- Blasbalg, U., Hershkowitz, I. & Karni-Visel, Y. (2018) Support, reluctance, and production in child abuse investigative interviews. *Psychology, Public Policy and Law*, 24, 518-527. <https://doi.org/10.1037/law0000183>
- Ceci, S. J. & Bruck, M. (1993) Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403-439. <https://doi.org/10.1037/0033-2909.113.3.403>
- Colquitt, J., LePine, J., & Noe, R. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707. <https://doi.org/10.1037//0021-9010.g5.5.678>
- Finnilä, K., Mahlberg, N., Santtila, P., Sandnabba, K. & Niemi, P. (2003) Validity of a test of children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *Journal of Experimental Child Psychology*, 85, 32-49. [https://doi.org/10.1016/S0022-0965\(03\)00025-0](https://doi.org/10.1016/S0022-0965(03)00025-0)
- Haginoya, S., Ibe, T., Yamamoto, S., Yoshimoto, N., Mizushi, H., & Santtila, P. (2023). AI avatar tells you what happened: The first test of using AI-operated children in simulated interviews to train investigative interviewers. *Frontiers in Psychology*, 14, 1-11. <https://doi.org/10.3389/fpsyg.2023.1133621>
- Haginoya, S., Yamamoto, S., & Santtila, P. (2021). The combination of feedback and modeling in online simulation training of child sexual abuse interviews improves interview quality in clinical psychologists. *Child Abuse & Neglect*, 115, 105013. <https://doi.org/10.1016/j.chiabu.2021.105013>
- Haginoya, S., Yamamoto, S., Pompèdda, F., Naka, M., Antfolk, J., & Santtila, P. (2020). Online Simulation Training of Child Sexual Abuse Interviews With Feedback Improves Interview Quality in Japanese University Students. *Frontiers in Psychology*, 11, 998. <https://doi.org/10.3389/fpsyg.2020.00998>
- Hershkowitz, I., Fisher, S., Lamb, M. E. & Horowitz, D. (2007) Improving credibility assessment in child sexual abuse allegations: The role of the NICHD investigative interview protocol. *Child Abuse and Neglect*, 31, 99-110. <https://doi.org/10.1016/j.chiabu.2006.09.005>
- 法務省 (2022) 代表者聴取の取組の実情 (法制審議会刑事法 [性犯罪関係] 部会第 5 回会議配布資料 10) Retrieved January 19, 2024, from <https://www.moj.go.jp/content/001367831.pdf>
- Korkman, J., Santtila, P., & Sandnabba, N. K. (2006). Dynamics of verbal interaction between interviewer and child in interviews with alleged victims of child sexual abuse. *Scandinavian Journal of Psychology*, 47, 109-119. <https://doi.org/10.1111/j.1467-9450.2006.00498.x>
- Krähenbühl, S., Blades, M., & Eiser, C. (2009). The effect of repeated questioning on children's accuracy and consistency in eyewitness testimony. *Legal and Criminological Psychology*, 14, 263-278. <https://doi.org/10.1348/135532508X398549>
- Krause, N., Pompèdda, F., Antfolk, J., Zappalà, A., Santtila, P. (2017) The Effects of Feedback and Re-

- reflection on the Questioning Style of Untrained Interviewers in Simulated Child Sexual Abuse Interviews. *Applied Cognitive Psychology*, 31, 187-198. <https://doi.org/10.1002/acp.3316>
- Lamb, M. E. (1995) The Investigation of Child Sexual Abuse. *Journal of Child Sexual Abuse*, 3, 93-106. https://doi.org/10.1300/J070v03n04_06
- Lamb, M. E., Brown, D. A., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2018). *Tell Me What Happened: Questioning children about abuse*. John Wiley & Sons.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., Horowitz, D. & Esplin, P. W. (2002) The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science*, 6, 114-125. https://doi.org/10.1207/S1532480XADS0603_2
- 仲真紀子 (2012) 面接のあり方が目撃した出来事に関する児童の報告と記憶に及ぼす影響 心理学研究, 83, 303-313. <https://doi.org/10.4992/jjpsy.83.303>
- 仲真紀子 (2015) 多専門連携による司法面接の実施を促進する研修プログラムの開発と実装 RISTEX 戦略的創造研究推進事業 (社会技術研究開発) 平成 27 年度研究開発実施報告書 科学技術振興機構 http://www.jst.go.jp/ristex/pdf/pp/JST_1115150_15666857_2016_naka_YR.pdf
- Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W. & Horowitz, D. (2000) Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse and Neglect*, 24, 733-752. [https://doi.org/10.1016/S0145-2134\(00\)00137-X](https://doi.org/10.1016/S0145-2134(00)00137-X)
- Pompedda, F., Zappalà, A. & Santtila, P. (2015) Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. *Psychology, Crime and Law*, 2, 28-52. <https://doi.org/10.1080/1068316X.2014.915323>
- Zhang, Y., Segal, A., Pompedda, F., Haginoya, S., & Santtila, P. (2022). Confirmation bias in simulated CSA interviews: How abuse assumption influences interviewing and decision-making processes? *Legal and Criminological Psychology*, 27, 314-328. <https://doi.org/10.1111/lcrp.12213>
- Zhang, Y., Li, S., Zhang, Y., Haginoya, S., & Santtila, P. O. (2023). Effects of combining feedback and hypothesis-testing on the quality of simulated child sexual abuse interviews with avatars among Chinese university students. *PLOS ONE*, 18, e0285114. <https://doi.org/10.1371/journal.pone.0285114>

(受稿: 2024.1.23; 受理: 2024.7.22)
