Studies in Educational Evaluation 76 (2023) 101236

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

The influence of self-assessment practices on online training for investigative interviewing skills☆

Zi Yan [a], Sonja Brubacher [b,*], David Boud [c,d,e], Martine Powell [b]

[a] Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, China
[b] Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Australia
[c] Centre for Research in Assessment and Digital Learning, Deakin University, Geelong, Australia
[d] Faculty of Arts and Social Sciences, University of Technology Sydney, Australia
[e] Centre for Research on Work and Learning, Middlesex University, London, UK

ARTICLE INFO

Keywords:
Self-assessment
Online training
Police education
Investigative interview skill
Personal attributes

ABSTRACT

This paper investigates how 192 trainees' self-assessment practices and personal attributes related to their success in an online interview training program. Performance was measured pre- and post-training by evaluating the use of open-ended questions and positive interviewing behaviors. Personal data and a pre-training survey on self-assessment practices (i.e., SEFI, *seeking external feedback through inquiry;* SEFM, *seeking external feedback through monitoring;* SIF, *seeking internal feedback*; and SR, *self-reflection*) were also collected. Trainees with higher levels of SEFI had larger learning gains from the training, whereas trainees with higher levels of SEFM or SIF had smaller learning gains. Trainees with lower interviewing skills before the training, and those from policing backgrounds, demonstrated larger learning gains. Younger trainees evidenced a larger increase in the use of open-ended questions, while older trainees showed greater changes in prevalence of positive interviewing behaviors. The implications of the findings for the development of training programs are discussed.

There is growing international awareness that training quality underlies the development and sustainability of many skills, such as investigative interviewing (Benson & Powell, 2015; Lamb, 2016). Training that consists of multiple sessions, delivered in a spaced and cumulative manner, has been shown to produce deep and enduring learning of this highly complex activity (Akca et al., 2021; Brubacher et al., 2021; Cederborg et al., 2021; Hershkowitz et al., 2017). On an individual level, however, not all trainee interviewers accrue the expected benefits even when the learning format is optimized. Studies about professional training (e.g., teacher training) have revealed that training outcomes may be related to individual characteristics (e.g., gender, age, and baseline performance prior to the training) (Gore et al., 2017; Makopoulou et al., 2021; Yoon & Kim, 2022). Similar studies are relatively scarce in the field of investigative interviewing training. Identifying some of these variables can provide valuable insight to organizations when deciding how best to support their trainees, or where to allocate training resources; to training developers with regards to creating an experience that is accessible for all learners; and to researchers invested in furthering the theoretical literature on adult professional education.

The present study addressed knowledge gaps by investigating how individuals' self-reported self-assessment practices and demographic factors relate to their training outcomes in an online training course that taught professionals to interview children and other vulnerable witnesses about abuse and maltreatment. The dominant measure of effective interviewing skill in the field of forensic interviewing has historically been open-ended question use (Brown & Lamb, 2015; Cederborg et al., 2021; Hershkowitz et al., 2017). Persistence in carefully crafted, non-leading open-ended questions is considered the gold standard in questioning but is a notoriously difficult skill to maintain (see Benson & Powell, 2015, for review). Effective open-ended questioning is supported by positive interviewing behaviors such as adopting an interviewee's terminology, delivering information about the interview purpose and process, and avoiding introducing leading information (Brubacher et al., 2021; Lawrie et al., 2021).

In the present research, we tested whether there were measurable

☆ This research was funded in part by an Australian Research Council Discovery grant DP 180100715 to MBP, DB, and SPB.

* Correspondence to: Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, 176 Messines Ridge Rd., Mt Gravatt, QLD 4122, Australia.

E-mail address: s.brubacher@griffith.edu.au (S. Brubacher).

https://doi.org/10.1016/j.stueduc.2023.101236
Received 11 July 2022; Received in revised form 21 November 2022; Accepted 10 January 2023

individual differences in trainee self-assessment approaches that could predict propensity to benefit from interviewer training (measured as increases in the proportional use of open-ended questions and number of positive interviewing behaviors). We further examined several demographic factors that could potentially impact training outcome variability.

1. The role of self-assessment practices in training

Self-assessment has attracted increasing interest in vocational and professional education (Kersh et al., 2011; Panadero et al., 2018). This is because both during and after professional training programs, trainees need to make continuing appraisals of their own work to identify their learning needs and make changes accordingly. Good self-assessment is thus crucial for trainees to maximize their learning gains from training programs because it helps clarify what they need to learn and how to learn effectively. To sustain their interviewing skills after training concludes, trainees require opportunities and the capacity to reflect on their performance and consolidate learning gains (Lamb, 2016).

Self-assessment is a key element of self-regulated learning, which is crucial in online training programs where independent learning is usually necessary (Cho & Shen, 2013; Liberty & DeSpain, 2020). Self-regulated learning refers to the process by which students plan, execute, monitor, and adapt their actions to attain their learning goals (Zimmerman, 2000). It is a proactive activity highlighting the learner's personal initiative, rather than a passive reaction (Zimmerman & Labuhn, 2012). Effective self-regulated learning is more likely to happen when learners have a sensible appraisal of their performance so that they can direct their further learning to critical aspects of their particular needs (Boud et al., 2013). Self-assessment is a fundamental process occurring at different phases of self-regulated learning (i.e., the preparatory, performance, and appraisal phases) for different purposes (Harris & Brown, 2018; Yan, 2020). In the preparatory phase, self-assessment facilitates locating available resources and setting reasonable learning goals; in the performance phase, self-assessment can be used to evaluate the effectiveness of learning strategies and monitor the learning process; in the appraisal phase, self-assessment helps reflect learners' strengths and weaknesses as well as the directions for future learning (Yan, 2020).

Research in self-assessment has shifted from grade-guessing to a focus on process. From a process perspective, self-assessment offers opportunities for learners to progressively improve their ability to evaluate their own work rather than requiring them to give a fixed grade at a given point in time (Boud et al., 2018). Scholars have recognized that the self-assessment process is complex and involves multiple steps (Andrade & Valtcheva, 2009; Sargeant et al., 2010). Yan and Brown (2017) proposed and empirically demonstrated a cyclical process model of self-assessment that incorporates three major actions: (1) determining assessment criteria, (2) self-directed feedback-seeking from external and internal sources, and (3) self-reflection. In the self-assessment process, learners first determine the criteria for self-assessment. They then set out to collect feedback on their own performance from external and/or internal sources. External feedback could be obtained through inquiry or monitoring approaches. Inquiry refers to asking relevant people (e.g., teachers and peers) for feedback information (i.e., seeking external feedback through inquiry; SEFI); monitoring refers to feedback from interactions with learning materials (e.g., prior assignments, extra work, learning guides) (i.e., seeking external feedback through monitoring; SEFM). Internal feedback comes from internal reactions (e.g., emotions, feelings, physical sensations, and internal states) triggered by their performance (i.e., seeking internal feedback; SIF). Based on information from these sources, learners then reflect on the quality of their learning, which results in the identification of their own strengths and weaknesses (i.e., self-reflection; SR). Following this process, learners make a judgement that is subject to continuous calibration.

Yan and Brown's (2017) model was adopted in the present study for two reasons. First, their model explicitly identifies the common actions learners engage in within the self-assessment process. It breaks the "black box" of the inner process of self-assessment and facilitates researching explicit self-assessment practices. Second, their model integrates feedback-seeking, a frequently neglected component, into the self-assessment process. In the investigative interviewing literature, it is understood that high-quality feedback information from trainers and colleagues is crucial for the effective generalization of the learning gains from training contexts to the working context (Lamb et al., 2002a). Experts in interview training (e.g., Lamb, 2016; Powell et al., 2008) have also highlighted the importance of feedback inputs from external parties, such as experts and peers, in maintaining the effects of training for investigative interviewers.

Although beneficial, ongoing external monitoring is resource intensive in terms of cost, time, and expertise to lead the sessions (Lamb et al., 2002a, 2002b; Price & Roberts, 2011). Thus, trainees should also be taught how to continuously reflect on their own performance, to consolidate and sustain their learning gains from training programs (Lamb, 2016), and their self-assessment practices might be an important determinant of the impact of the training on their skills. Studies in other professional training contexts (e.g., teacher training) have revealed that the impact of training programs is influenced by trainees' personal attributes (Makopoulou et al., 2021), but these links have not been tested in the context of learning investigative interviewing skills. In the next section, we briefly review what is known about interviewing skills training.

2. Interviewing skills and training

2.1. Effective interview skills

As noted earlier, the primary measure of good interviewing performance has been open-ended question use (Benson & Powell, 2015). Open-ended questions encourage elaborate recall but do not specify exactly what information an interviewee should report (e.g., "Tell me what happened"; Powell & Snow, 2007). Effective use of open-ended questions is supported by other interviewing behaviors including, but not limited to, employing backchannel utterances (e.g., "Uh-huh"), maintaining simple language, adopting a variety of question stems, and ensuring that questions are developmentally appropriate (e.g., Brown & Lamb, 2015; Brubacher et al., 2021). There is a greater opportunity for miscommunication and misinterpretations to occur when interviewing children (compared to adults) due to children's developing language and increased suggestibility (e.g., Poole, 2016). As such, the interviewing of children requires especially careful attention to questioning practices and a thorough understanding of evidence-based interview approaches (Brown & Lamb, 2015). This set of knowledge and skills requires highly specialized training.

2.2. Training programs for interviewers of children

Experts in child development, memory, and interviewer training now recognize that the traditional approach to interviewer training (with a group of learners in a classroom absorbing, in lockstep, information delivered in massed sessions over a few days) is ineffective (Lamb, 2016; Poole, 2016; Powell, 2013). Several training groups adapted their curricula to be congruent with principles of human learning and published results of evaluation studies showing sustained improvement in interviewer skills six months or more post-training (e.g., Benson & Powell, 2015; Brubacher et al., 2021; Cederborg et al., 2021). Specifically, these programs included a) spaced learning over time; b) interactive, ongoing, and varied opportunities to practice skills; and c) individualized expert feedback (Bjork & Bjork, 2011; Donovan & Radosevich, 1999; Kluger & DeNisi, 1996; Powell, 2013).

Of particular interest to the current study is an evaluated *online* training program that was delivered to police interviewers (Benson &

Powell, 2015), and later adapted and evaluated for different professional groups (Brubacher et al., 2021; Lawrie et al., 2021). The results showed that this type of program was valuable in interviewer training. Much of the learning is done independently, with periodic one-on-one sessions with a trainer via videochat. Online learning, where relational support is relatively less compared to face-to-face learning, is particularly likely to require individual self-regulation and reflection (Cho & Shen, 2013; Liberty & DeSpain, 2020; Yan & Carless, 2022), yet trainees will vary in the degree to which they demonstrate those behaviors. Even when a training program is designed in accordance with elements of human learning, and evaluations of the training program demonstrate overall success, not every individual improves. For example, in Brubacher, et al. (2021), 6% ($n = 3$) of trainees, and 9% ($n = 9$) in Cederborg et al. (2013), did not increase in skill from pre- to post-training. A study of self-reported professional development (optional training) by 136 employees from diverse professions revealed that individual characteristics of learning goal orientation, openness to experience, and conscientiousness were important to self-development when workplace support was low, but less so when it was high (Orvis & Leffler, 2011).

There are at least three important questions for any organization when deciding what training to offer their staff, with regards to the financial and temporal costs of training (O'Toole & Lawler, 2006): a) is the training program evidence-based, with demonstrated successful outcomes?; b) which individuals should receive training?; and c) how can trainees be best supported during training to maximally benefit? For training developers, understanding what accounts for individual variation in learning could potentially lead to the ability to tailor training to accommodate a wider variety of learners.

3. The role of demographic factors in training impact

Despite the increasing number of training programs on investigative interviewing skills, the relationship between trainees' characteristics and training effectiveness has not been well studied. This is a research gap given the identified association between personal attributes and the patterns of interviewing behaviors. For example, Lamb and Garretson (2003) found a gender effect on interviewing practices (unrelated to training): female interviewers asked boys more open-ended and suggestive questions than girls, whereas male interviewers questioned boys and girls in a similar way. There was no overall main effect of gender, however, on the type of questions interviewers asked. Thus, it is unclear whether gender differences in propensity to benefit from interviewer *training* would emerge.

The literature in other professional training contexts (e.g., teacher training) reported that training outcome is dependent on a range of individual factors, such as gender, age, and baseline performance prior to the training (Makopoulou et al., 2021). The findings regarding the gender effect on teacher training are mixed. Some studies found that female schoolteachers tended to report higher training needs (Estepp et al., 2014; Gokmenoglu et al., 2016), higher levels of involvement in training activities (Richter et al., 2011), and greater learning gains (Gore et al., 2017) compared with their male counterparts. In contrast, Yoon and Kim (2022) reported that male teachers were more likely to have higher involvement levels in training than female teachers. The authors speculated that the different findings across studies were due to the different samples used (e.g., the country the teachers came from and the subject areas the teachers teach). Age can also be a moderator of training impact and may interact with gender. In Makopoulou et al.'s (2021) study, the older group (age > 45) had more gains than younger age groups for females, whereas for males, the older group (age > 45) experienced less learning gains than younger groups. The occupational role is another demographic variable likely to influence training outcomes. Experiences and expectations associated with different occupational roles may influence trainees' perceptions of and attitudes towards investigative interviewing which, in turn, determines what and how much they can learn from the training.

4. The current study

Overall, from a theoretical perspective, it appears that trainees' self-assessment practices may influence the impact of training programs where independent learning is a crucial component (Cho & Shen, 2013; Liberty & DeSpain, 2020). No empirical study, however, has tested this argument. Prior research in professional training contexts, like teacher training, suggests that personal attributes such as gender and age could also influence the effectiveness of training programs (Gore et al., 2017; Makopoulou et al., 2021). More research is needed to understand the degree to which trainees' characteristics determine the impact of interviewing skill training programs. Thus, the current study aimed to address these gaps by examining the role of trainees' self-assessment practices and personal characteristics in determining the impact of an online training program on investigative interviewing skills. There were two specific research questions that guided the study.

RQ1. . How are trainees' self-assessment practices related to trainee outcomes in the online program?

RQ2. . How do trainees' characteristics (i.e., demographic factors such as gender, age, and occupation) relate to the impact of the online training program, in terms of trainees' outcomes?

5. Method

5.1. Participants

Participants were sampled from professional learners who accessed an online training program during 2020–2021 via one of five organizational sites ($n = 201$). The program was offered by researchers at Griffith University. To be eligible for inclusion in the present study, participants had to agree to the usage of their data for research purposes (all did) and they had to complete the pre- and post-training assessment. Trainees had up to 12 months to complete the self-paced course. There were 9 interviewers who did not complete the training entirely; thus, valid responses from 192 trainees were analyzed.

The sample comprised 117 females, 66 males, and 9 trainees without gender information. Trainees were distributed into four age groups: 20–29 ($n = 56$), 30–39 ($n = 67$), 40–49 ($n = 52$), and 50 or above ($n = 17$). Their professional backgrounds included policing ($n = 135$), child protection and social work ($n = 14$), forensic interviewing ($n = 28$), and others ($n = 15$). Ethics approval was given by Griffith University. All trainees initially read and signed informed consent statements. They were permitted to continue with the training even if they declined participation.

5.2. Research design and procedure

A pre-post single-group design was applied in an online learning environment. Learners enrolled in a 15-module, self-paced interview training course (approximately 45 h; see Benson & Powell, 2015 for an overview of content). Before the first module unlocked, trainees completed a short online survey to report their self-assessment practices in learning and they participated in a baseline 10-minute mock interview with an actor playing the role of a child interviewee. No feedback information was provided at this initial interview session. Actors had extensive experience playing the role of interviewees in mock interviews. They adhered to a set of guidelines that outlined the nature of the case and how to respond to various types of questions. Such guidelines have been shown to yield response consistency across trainers (Brubacher et al., 2021). This procedure (compared to using learners' colleagues as mock interviewees) produces superior and more sustainable outcomes in learners (Powell et al., 2008). Mock interviews have been used for training, assessment, and evaluation purposes in the field of investigative interviewing and their use is growing (Benson & Powell, 2015; Powell et al., 2022).

Table 1
Descriptive statistics and bivariate correlations.

	Mean[a]	SD	1	2	3	4	5	6	7	8
1. PropOpen_pre	0.29	0.22	-							
2. Pos12_pre	5.86	1.94	.65**	-						
3. PropOpen_post	0.82	0.14	.07	.02	-					
4. Pos12_post	8.88	2.06	.28**	.23**	.23**	-				
5. SEFI	0.30	1.16	.11	.09	.08	.07	-			
6. SEFM	0.19	1.00	-.10	-.03	-.14	-.31**	.29**			
7. SIF	-0.24	1.06	-.09	-.03	-.03	-.24**	.22**	.20**	-	
8. SR	1.16	1.29	.01	.02	-.10	-.20**	.51**	.23**	.16*	-

Notes:
SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry; SIF: seeking internal feedback; SR: self-reflection.
[a] The means for SEFM, SEFI, SIF, and SR are Rasch-calibrated measures in logits.
*$p < 0.5$, **$p < 0.01$.

After the baseline mock interview was complete, learners began the course. It consisted of readings, self-tests, films, narrated PowerPoints, workbook exercises and some assignments (see Benson & Powell, 2015, for an overview). The course was largely self-paced and asynchronous but had completion hurdles in ten modules that had to be passed (e.g., score over 80%) or graded (pass/fail) by a course trainer before the next module could be accessed. These hurdles included quizzes, short written assignments, and mock interviews for training purposes. Mock interviews used for training focused on a specific topic (e.g., practice interviewing about allegations of repeated abuse) and included discussion and verbal or written feedback information (Powell et al., 2022). At the conclusion of Module 15, learners participated in the post-training mock interview, which was used to determine the change in their interviewing skills. At the conclusion of the course, learners received a certificate once all activities were complete. No numerical or letter grade was assigned.

5.3. Measures

5.3.1. Interviewing skills

Research assistants coded baseline and post-training mock interviews for the types of questions asked and 12 interviewer behaviors.

Proportion of open-ended questions (PropOpen): Question types were coded as open-ended, wh-, or closed. Open-ended questions were defined as those that encouraged an elaborate response and did not dictate the content of the response (Powell & Snow, 2007). Examples of open-ended questions include, "Tell me everything you can remember," "What else happened?", and "What happened when [pre-disclosed detail]". Like open-ended questions, wh- questions queried recall memory, but they specified the desired response content. Wh- questions included who, when, where, why, how, and what questions that were narrow in scope (e.g., "What color was the car?"). Closed questions elicit information from recognition memory. They do not encourage elaborate responses and they implicitly or explicitly present the answer within the question (Powell & Snow, 2007). Closed questions included yes-no questions and multiple choice. Research assistants summed the total number of questions asked. Each type was divided by the total to obtain proportions; the proportion of open-ended questions asked was used as the dependent variable in analyses.

Positive interview behaviors (Pos12): The 12 interviewer behaviors and their coding definitions can be found in Appendix A. They were coded dichotomously as present/absent and summed to produce a maximum score of 12 (for a similar procedure, see Brubacher et al., 2021).

Question types and interviewer behaviors were coded by a post-doctoral research fellow with prior experience coding these variables on over 2000 interviews (mock interviews and field interviews concerning allegations of abuse), as part of her role as a professional trainer. A random reliability assessment on 15% of the present sample revealed percent agreement of over 95% on question types and Cohen's kappa of.76–1.00 for interviewer behaviors. The research fellow's codes were entered for analyses.

5.3.2. Self-assessment practice

Before the training began, all trainees reported their self-assessment practices in learning by responding to a short online survey (i.e., the Self-assessment Practice Scale, SaPS). The SaPS was initially developed in a school context (Yan, 2018) and then revised for use in professional training contexts (Yan et al., 2020). The scale contains 16 items grouped into four subscales, including (1) seeking external feedback through inquiry (SEFI; 3 items; $\alpha = .69$; e.g., I ask my teachers/trainers to give me feedback about my performance), (2) seeking external feedback through monitoring (SEFM; 5 items; $\alpha = .71$; e.g., I check whether I have mastered course content by doing extra work), (3) seeking internal feedback (SIF; 3 items; $\alpha = .54$; e.g., My gut feelings tell me whether my work is good or bad), and (4) self-reflection (SR; 5 items; $\alpha = .71$; e.g., Any areas I am unsure of after finishing my work, I go over again). All items are on a Likert-type scale (1-Strongly Disagree to 5-Strongly Agree).

5.4. Data analysis

To evaluate whether the training had the expected outcomes, pre- and post- training differences in the two measures of trainees' interviewing skills, i.e., PropOpen and Pos12, were examined using the Wilcoxon signed-rank test for paired samples. To address RQ1 and RQ2, hierarchical multiple regression analyses were carried out with the trainees' learning gains (i.e., post-training score minus pre-training score) on two measures of interviewing skills (i.e., PropOpen and Pos12) as dependent variables. Trainees' characteristics (gender, age, occupation), pre-training score of interviewing skills, and self-reported performance on the four self-assessment practices were used as predictors. Before the regression analyses, the psychometric properties of the SaPS used with the present sample were first examined using Rasch analysis (Rasch, 1960). Since the four subscales within the SaPS were theoretically related and all items share the same 5-point Likert type response scale, a multidimensional Rating Scale Rasch Model using ConQuest 2.0 (Wu et al., 2007) was applied to the data. In a multidimensional Rasch model, all subscales are calibrated simultaneously with their correlations taken into account so that the measurement precision on each subscale could be enhanced compared to a conventional unidimensional Rasch model (Adams et al., 1997). Multiple indices, including the Rasch reliability, the response scale functioning, and the item fit statistics, were utilized to examine the psychometric properties of the instrument. The Rasch-calibrated person measures of self-assessment practices were then used for hierarchical multiple regression analyses. This approach of data analysis has been widely used in empirical investigations (e.g., Coniam & Yan, 2016; Guo & Yan, 2019; Ma et al., 2017) due to the advantage of Rasch analysis in converting the ordinal rating data into interval data which permits subsequent

Table 2
Pre- and post-training difference in scores on two interviewing skill indicators.

	Mean	SD	$Z^{a}_{\text{Post-Pre}}$	P^{b}
PropOpen_pre	0.294	0.224	-11.906	.000
PropOpen_post	0.818	0.137		
Pos12_pre	5.865	1.937	-10.871	.000
Pos12_post	8.880	2.055		

Note. [a] Based on negative ranks. [b] Wilcoxon Signed Ranks Test.

interviewing behaviors. As the Kolmogorov-Smirnov test (K-S test) indicated that trainees' scores on the pre- and post-training mock interviews [PropOpen_pre (Z = 1.544, p = 0.017), PropOpen_post (Z = 1.989, p = 0.001), Pos12_pre (Z = 1.635, p = 0.010) and Pos12_post (Z = 1.932, p = 0.001)] were not normally distributed, we used the Wilcoxon signed-rank test for paired samples to calculate the pre- and post-training difference. The results showed that the training program was effective because trainees' scores on two interviewing skill indicators (i.e., PropOpen and Pos12) increased after the training, and the differences were statistically significant ($p < .01$, see Table 2).

Table 3
Summaries of hierarchical regression analyses for PropOpen.

Model	Variable	*B*	*SE B*	*STD β*	*t*	$R^2_{adjusted}$	F	*P*
1	Constant	.461	.047		9.700***	.172	10.442	.000
	Policing	.191	.039	.338	4.846***			
	Age group= 30–39	-.091	.044	-.169	-2.061*			
	Age group= 40–49	-.140	.047	-.248	-2.988**			
	Age group= >50	-.059	.066	-.068	-.898			
2	Constant	.760	.030		25.215***	.747	108.697	.000
	Policing	.090	.022	.159	4.016***			
	Age group= 30–39	-.036	.025	-.067	-1.461			
	Age group= 40–49	-.073	.026	-.129	-2.805**			
	Age group= >50	-.002	.036	-.002	-.057			
	PropOpen_pre	-.905	.045	-.786	-20.163***			
3	(Constant)	.763	.031		24.786***	.789	83.095	.000
	Policing	.097	.022	.171	4.329***			
	Age group= 30–39	-.044	.024	-.083	-1.824			
	Age group= 40–49	-.079	.026	-.140	-3.091**			
	Age group= >50	-.006	.036	-.007	-.171			
	PropOpen_pre	-.926	.044	-.804	-20.907***			
	SEFI	.023	.009	.107	2.716**			
	SEFM	-.025	.010	-.100	-2.542*			

Note:
Only significant predictors are shown in the table.
SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry.
Reference age group: 20–29.
* $p < .05$; * * $p < .01$; * ** $p < .001$.

parametric analysis (e.g. regression analysis) (Bond et al., 2020).

6. Results

We confirmed the instrument quality before using the data collected through it to address the core research questions. Although the revised SaPS has been previously validated in a similar context (see Yan et al., 2020 for details), the psychometric properties of the instrument were further examined with the present sample, from a Rasch measurement perspective, as a continuous validation process. The results showed that all items sufficiently fit the Rasch model. The fit statistics (both infit and outfit MNSQ) of all items were within the acceptable range (0.75–1.33), as suggested by Wilson (2005). The step calibrations (i.e., the measures of the thresholds between adjacent categories) increased monotonically from − 1.11, − 0.83, − 0.10, to 2.03 logits, indicating a well-functioning five-point response scale. The Rasch reliabilities for the four subscales (i.e., SEFI, SEFM, SIF, and SR) were 0.65, 0.75, 0.62, and 0.82, respectively. The reliabilities for subscales SEFI (i.e., *seeking external feedback through inquiry*) and SIF (i.e., *seeking internal feedback*) were marginal but still acceptable. No missing items were observed in trainees' responses to the SaPS items. Descriptive statistics and bivariate correlations are presented in Table 1.

6.1. Impact of the online training program

We first evaluated whether the training program had the expected outcomes with the present sample. Although the program has been evaluated previously (Benson & Powell, 2015; Lawrie et al., 2021), this specific sample had never been analyzed. We compared trainees' pre- and post-training performance on open-ended question use and positive

6.2. Predictors of the impact of the online training program

To examine the role of trainees' self-assessment practices and personal characteristics in predicting the impact of the training program, separate hierarchical regression analyses were carried out with the pre- and post-training differences in the two interviewing skills as dependent variables. In each analysis, gender, age, and professional background (policing or not)[1] were entered into the regression in Step 1, pre-training interviewing skills were entered in Step 2, and trainees' Rasch-calibrated measures on the four dimensions of self-assessment practices were entered in Step 3.

Table 3 presents the results of the hierarchical regression analysis for PropOpen. There are five significant predictors of the impact of the online training program on the proportion of open-ended question use. Trainees' pre-training interviewing performance was the strongest and negative predictor ($\beta = -.809$, $p < .001$) of the training impact. Not surprisingly, trainees with higher pre-training performance had smaller learning gains from the training. Policing was a positive predictor ($\beta = .164$, $p < .001$), indicating trainees from the policing background had larger learning gains than trainees from other professional backgrounds. All age groups had descriptively smaller learning gains than the 20–29 age group, but only the 40–49 age group reached the significance level

[1] As the majority of trainees were from policing background (69.2%), the professional background was categorized into a dichotous variable, i.e., policing and non-policing.

Table 4
Summaries of hierarchical regression analyses for Pos12.

Model	Variable	*B*	*SE B*	*STD β*	*t*	$R^2_{adjusted}$	F	*P*
1	Constant	1.981	.338		5.869***	.051	4.246	.006
	Age group= 30–39	1.245	.458	.237	2.720**			
	Age group= 40–49	1.385	.477	.251	2.901**			
	Age group= >50	1.843	.680	.215	2.710**			
2	Constant	6.457	.485		13.314***	.430	35.336	.000
	Age group= 30–39	1.478	.355	.281	4.160***			
	Age group= 40–49	1.861	.372	.337	4.998***			
	Age group= >50	1.978	.527	.230	3.752***			
	Pos12_pre	-.800	.073	-.618	-10.960***			
3	Constant	6.741	.459		14.681***	.504	27.437	.000
	Age group= 30–39	1.131	.338	.215	3.349**			
	Age group= 40–49	1.529	.353	.277	4.330***			
	Age group= >50	1.466	.505	.171	2.904**			
	Pos12_pre	-.820	.068	-.633	-11.989***			
	SEFI	.323	.120	.151	2.692**			
	SEFM	-.569	.140	-.230	-4.063***			
	SIF	-.372	.138	-.150	-2.693**			

Note:
Only significant predictors are shown in the table.
SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry; SIF: seeking internal feedback.
Reference age group: 20–29.
* $p < .05$; * * $p < .01$; * ** $p < .001$.

($\beta = -.142$, $p < .01$). Two of the four dimensions of self-assessment practices were significant predictors. Trainees with higher levels of SEFI (i.e., seeking external feedback through inquiry) demonstrated larger learning gains ($\beta = .106$, $p < .01$), while trainees with higher levels of SEFM (i.e., seeking external feedback through monitoring) tended to have smaller learning gains ($\beta = -.102$, $p < .05$). The interactions between predictors were also examined, but no interaction was significantly related to training impact.

The results of the hierarchical regression analysis for Pos12 are shown in Table 4. Five predictors were found significant. Similar to PropOpen, trainees' pre-training performance on positive interviewing behaviors was the strongest and negative predictor ($\beta = -.631$, $p < .001$). The higher the pre-training performance, the smaller the learning gains on the positive interview behaviors measure. All older age groups had larger learning gains than the 20–29 age group, and all differences were significant (β ranging from.125 to.276). Three dimensions of self-assessment practices were significant predictors. Trainees who performed better on SEFI (i.e., seeking external feedback through inquiry) had larger learning gains ($\beta = .146$, $p < .05$), whereas trainees who performed better on SEFM (i.e., seeking external feedback through monitoring) ($\beta = -.240$, $p < .001$) or SIF (i.e., seeking internal feedback) ($\beta = -.142$, $p < .05$) tended to have smaller learning gains. There was no significant interaction between predictors.

7. Discussion

The present research sought to identify how trainees' self-assessment practices and personal characteristics relate (if at all) to their propensity to benefit from a program requiring a high degree of self-regulated learning. Having confirmed that the program enhanced skills (both through prior evaluations with other samples and by evaluating learning changes in the present sample), we then took a closer look at how the trainees' demographic and self-assessment practices related to their training outcomes.

7.1. The relationship between trainees' self-assessment practices and training impact

The results showed that feedback-seeking components in trainees' self-assessment practices were significantly related to training impact. Interestingly, the influencing patterns of various self-assessment practices differed. For both interviewing skill indicators (i.e., the proportion of open-ended questions and positive interview behaviors), SEFI (i.e., *seeking external feedback through inquiry*) was significantly positively related to the training impact, whereas SEFM (i. e., *seeking external feedback through monitoring*) was significantly negatively related to training impact. SIF (i.e., *seeking internal feedback*) was negatively related to training impact, but only showed significance on positive behaviors (see Table 3 and 4). Overall, these findings demonstrated that the ways in which trainees self-assessed their performance had a crucial role in determining the outcome of training on their investigative interviewing skills. In particular, the feedback information sources in self-assessment (e.g., verbal feedback from course trainers rather than scores on course self-tests) mattered in terms of how much trainees learned from the training program.

The positive association of SEFI on training impact corroborates prior studies that highlight the importance of external feedback inputs in sustaining training impact. High-quality feedback information from external parties (e.g., trainers and peers) is crucial not only for sustaining the training effect (e.g., Lamb, 2016; Powell et al., 2008), but also for the effective generalization of learning gains to the working context for investigative interviewers (Lamb et al., 2002a). Trainees with higher levels of SEFI are more likely to seek information on their own performance from relevant persons. As such, these trainees can attain more feedback information to use during the training which, in turn, results in more learning gains. To some extent, this finding echoes the advocation for ongoing supervision and support from experts after the completion of the training (Lamb et al., 2002b; Smith, 2009) because continuous input from experts reinforces what they know and can do. Further, because the course included one-on-one mock interview sessions (a minimum of 6), the course itself afforded sufficient opportunities and a supportive learning environment for trainees to receive expert feedback inputs.

In contrast, SEFM was negatively associated with training impact. There is no direct reference in the professional training context, but this finding is different from what has been found in higher education. For example, Yan (2020) revealed that SEFM in the performance phase of self-regulated learning was the strongest, positive predictor of students' academic achievement. There are two speculative explanations. First, as learners with higher levels of SEFM rely on interactions with written learning materials (e.g., extra work, past assignments, reference guides or websites) to get information on their own performance, the

availability of the learning materials is crucial. That SEFM did not show a positive link with trainees' learning gains in the current study is possibly because the professional trainees had fewer written assignments/papers and more practical exercises compared to university student samples. Thus, the professional learners had fewer opportunities, and perhaps less need, to interact with written learning materials. Perhaps more relevant to the current sample, SEFM depends on trainees' own evaluation and interpretation of their performance. Trainees in the current course were only directed explicitly to reflect on their own performance one time, and were given limited guidance on how to do that. The feedback information obtained from SEFM may be viewed as more subjective compared with that generated from SEFI. Hence, trainees relying on SEFM can generate inaccurate judgments of their own performance (i.e., the extent to which they open-ended questions and positive interviewing behaviors, in this case) which, in turn, reduces the learning effectiveness (Boud et al., 2013).

SIF also had a negative relationship with training impact. Internal feedback originates from within the self. For example, the gut feelings or emotions triggered by learning performance may influence trainees' subsequent learning behaviors. Trainees with higher levels of SIF may be influenced by their internal/intuitive feedback, which might have a counter effect on learning interviewing skills that require interviewers to control emotions during interviews or unlearn habitual responses. Furthermore, without external verifications, self-assessment based on internal feedback is vulnerable to idiosyncratic heuristics and bias (Joughin et al., 2019), which may negatively impact learning outcomes.

Nevertheless, we should not conclude that SEFM or SIF are undesirable self-assessment practices in the professional training context. Instead, these findings suggest that there are some conditions to meet to make trainees' SEFM or SIF work better for learning purposes. For example, it would be useful to provide trainees with sufficient learning materials to facilitate their feedback-seeking from those materials. Training programs for interviewers that include self-evaluation should offer opportunities for expert trainers to assess the accuracy of the self-evaluations and discuss results with the learners. Trainees need expert guidance around whether their self-evaluations are correct, and if not, why so, to ensure the appropriate use of the self-generated feedback information for learning improvement. This recommendation is similar to that suggested in peer-to-peer interviewing evaluations as well (e.g., Brubacher, et al., 2021).

It is interesting that self-reflection (SR) was not significantly related to training impact. This finding indicates, for the current sample, that the self-reflection component in the self-assessment process did not determine how much the trainees could learn from the training program. More studies are necessary to determine whether this finding could be duplicated in other samples. Moreover, we should be cautious in comparing this finding with past studies that highlighted the crucial role of reflection in sustaining the training impact (e.g., Lamb et al., 2002b; Smith et al., 2009). The target of the reflection in those studies was trainees' own interviewing performance (i.e., the focus of the training), while self-reflection measured in this study referred to general learning performance, such as the effectiveness of learning strategies and assessment results. Although the transfer of reflection from the general learning scenarios to interviewing is possible, this study suggests that such a transfer might not be straightforward.

7.2. The relationship between trainees' characteristics and training impact

Some trainees' characteristics were significantly associated with training impact (see Table 3 and 4). Trainees' baseline performance before training was negatively associated with training impact for both the proportion of open-ended questions and positive interview behaviors. Trainees with higher pre-training performance tended to have smaller learning gains from the program. This ceiling effect is not surprising given that the training and the pre/post-tests focused on the same performance indicators (i.e., the proportion of open-ended questions and positive interview behaviors).

Trainee age was significantly related to training impact, but the patterns differed across indicators (see Table 3 and 4). For the proportion of open-ended questions, the 20–29-year group had higher learning gains than the other age groups (only the difference with 40–49 was significant). The youngest age group had significantly lower baseline performance on this measure than other groups, and it is reasonable to assume the youngest group had comparatively less interviewing experience. By extension, they may also have had fewer negative question habits to correct (Powell et al., 2014; e.g., asking a series of specific questions to obtain a statement). For the positive interview behaviors, there was no significant difference in the baseline among groups, but the 20–29-year group had lower learning gains than the other age groups (all differences were significant). It is possible that the older (more experienced) interviewers picked up on the positive behaviors quickly because the older interviewers were less focused than the younger interviewers on question content. Presumably, the older interviewers may be more comfortable with what information they need to find out, allowing them to focus on the interview style. But it is this same knowledge about what they need to find out (i.e., who, what, when, where, how) that impairs them from using open-ended questions because open-ended questions (e.g., "What happened?") are a less direct (albeit higher quality) way to obtain this detail (Powell et al., 2014).

The police trainees began with a lower baseline performance on both the proportion of open-ended questions and positive interview behaviors, compared to the trainees from a non-policing background. This may explain why police had significantly larger learning gains. However, only the difference in the proportion of open-ended questions asked was statistically significant. This is probably because the dominant focus of the training course was on increasing open-ended question use.

Gender had no significant relationship with training impact. This result is incongruent with some studies on teacher training that reported a significant difference across gender in terms of training behaviors and gains (e.g., Gore et al., 2017; Richter et al., 2011; Yoon & Kim, 2022). Conversely, the finding is consistent with the investigative interviewing literature in that gender differences tend not to be observed in interviewing performance, regardless of training or interviewing experience (e.g., Akca & Eastwood, 2021; Powell et al., 2014; Smith et al., 2009). One study found no absolute difference between male and female forensic interviewers in question types used but there was an interaction between interviewer and witness gender. Even here, however, the effects were moderate and attenuated when interviewers adhered to an evidence-based interview protocol (Lamb & Garretson, 2003). It is perhaps the formulaic nature of questioning in formal investigative interviews that minimizes any potential gender differences – both in the propensity to use various question types and to benefit from interviewer training.

7.3. Implications

The finding that trainees' self-assessment practices are related to training impact is novel in this professional training context and informs the development of training programs regarding how to make training suitable for trainees with various self-assessment dispositions.

First, the findings help address questions regarding the provision of training, i.e., which individuals should receive the training and how best to support trainees in their learning (O'Toole & Lawler, 2006). The links between trainees' self-assessment practices, personal attributes, and their learning gains offer insight to organizations that need to identify which individuals would benefit most from training, and how to tailor individual trainee support to maximize learning outcomes for all trainees.

Second, as different types of self-assessment practices have different relationships with training impact, it is essential for training organizers to conduct assessments to understand trainees' patterns of self-

assessment practices in the early stages of the training program. On the one hand, such an understanding can inform the tailoring of the training program to accommodate trainees' learning differences. On the other hand, it facilitates the integration of self-assessment interventions, through explicit or implicit instruction, which may enhance the training effectiveness.

Third, given that self-assessment, as a learning process, can be fostered or inhibited by the conditions in which learning takes place, the findings indicate that if training program design facilitates desirable *self-assessment* practices, this in turn may increase training impact. For instance, as SEFI is a constructive strategy to enhance trainees' learning gains, trainees may be encouraged to take the initiative to seek feedback by asking relevant persons. At the same time, the program needs to offer sufficient opportunities that can foster interactions among trainees, trainers, and their peers. Moreover, providing more learning materials (e.g., references, learning guides, and websites) might be beneficial to those trainees who rely on SEFM for their self-assessment.

7.4. Limitations

Several potential limitations of the study need to be considered. First, without a control group, it is difficult to establish a cause-and-effect relationship by isolating the effect of the intervention, i.e., the training. Although the single-group design with participants working as "self-control" is common in program evaluation studies, future studies should consider the inclusion of control groups when possible (e.g., Brubacher et al., 2021). Second, we did not assess participants' motivation levels at various time points during the training. It is possible that some trainees were not motivated to take the training, or they were not motivated at the beginning (i.e., at the pre-training mock interview) but became more motivated as training progressed. If this were the case, such motivation could have led to better performance on the post-training assessment.

With respect to self-assessment, such practices exclusively relied on trainees' self-reporting, which might be vulnerable to recall bias. It would be beneficial for future studies to use more behavior-based measures of self-assessment practices to have a more realistic understanding of trainees' self-assessment practices. In this study, self-assessment practices were reported by trainees immediately before the training. Although cognitive and metacognitive learning practices are likely to remain stable across a short period of time (e.g., one year) (Roick & Ringeisen, 2018), future studies can relate the training outcomes to trainees' self-assessment practices observed during the training. Finally, future work could re-evaluate the relationship of SEFM with training outcomes after course materials are modified to allow learners greater opportunities to check their learning against course materials. The course in the present study already includes self-tests, but it may benefit from including other types of materials that can facilitate self-assessment activities.

8. Conclusion

The current study examined the relationship between trainees' self-assessment practices and personal attributes on the impact of an online training program on investigative interviewing skills. The results showed that the feedback information sources in self-assessment do matter. Trainees who seek feedback information from others with expertise can potentially learn more from training programs than those who rely on course materials or internal sensations for feedback. Trainees' personal characteristics are also related to training impact. Trainees with lower pre-training performance, and those from policing backgrounds, demonstrated larger learning gains. Age was significantly related to training impact, but the patterns differed across indicators. Overall, the findings shed light on the development of training programs in terms of how to make training suitable for individuals who approach and self-assess their learning in different ways.

Table A1
Coding Definitions for Interviewer Behaviors.

Code	Behavior
Undesirable Behaviors	Coded as 0 if three or more instances of the behavior were counted; 1 otherwise
Changes terms[a]	Interviewer changed a term. For example: C: "He touched my minny" I: "Tell me about him touching your private."
Complex[a]	Interviewer avoided double negatives, embedded clauses, multiple questions.
Developmentally inappropriate[a]	Questions about time, duration, measurement, specific location, body parts, or other abstract concepts.
Jargon[a]	Interviewer avoided difficult words and other legal jargon.
"Can you"[a]	Interviewer avoided open questions beginning with the phrasing "Can you."
Desirable Behaviors	Coded as 0 if not identified; 1 otherwise
Narrative exhausted	A complete narrative account of the event is elicited that includes information about what led up to the offence, the offence itself, and what happened after the offence (to ensure no other offending occurred on that occasion).
Ground rules	A ground rule was delivered during the main interview phase where appropriate (e.g., "remember, I wasn't there, so I don't know what happened" or "thank you for letting me know I made a mistake. It's important to tell me if I get something wrong").
Frequency Question	The interviewer used a recommended question to ask about abuse frequency, or explicitly identified information that suggested the allegation was repeated (e.g., "You said 'he always does it when he babysits you.' Tell me about the last time he babysat you."
Transition prompt	Interviewer transitioned to the allegation phase using a variation of the question, "Tell me what you have come to talk about."
Verifies leading	Where the interviewer had to introduce a piece of information during transition to the main interview phase (e.g., "I heard you have a friend named Janet"), the interviewer verified it was true (i.e., "Do you have a friend named Janet?").
Effective minimal encouragers (ME)	Interviewer used "uh-huh," "Mmm-hmm," or similar, alone (not in combination with a question) to encourage further narrative at least three times during the mock interview.
Open question variety	Interviewer used at least four different open-ended question stems. No more than two identically phrased open-ended questions were asked in a row.

Appendix A

See appendix Table A1.

References

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23. https://doi.org/10.1177/0146621697211001

Akca, D., & Eastwood, J. (2021). The impact of individual differences on investigative interviewing performance: a test of the police interviewing competencies inventory and the five factor model. *Police Practice and Research, 22*(1), 1027–1045. https://doi.org/10.1080/15614263.2019.1644177

Akca, D., Larivière, C. D., & Eastwood, J. (2021). Assessing the efficacy of investigative interviewing training courses: A systematic review. *International Journal of Police Science & Management, 23*(1), 73–84. https://doi.org/10.1177/14613557211008470

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12–19. https://doi.org/10.1080/00405840802577544

Benson, M. S., & Powell, M. B. (2015). Evaluation of a comprehensive interactive training system for investigative interviewers of children. *Psychology, Public Policy and Law, 21*(3), 309–322. https://doi.org/10.1037/law0000052

Bjork, E.L., & Bjork, R.A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. Psychology and the real world: Essays illustrating fundamental contributions to society, 2, 59–68.

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (2018). *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*. Routledge.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education, 38*(8), 941–956. https://doi.org/10.1080/02602938.2013.769198

Brown, D. A., & Lamb, M. E. (2015). Can children be useful witnesses? It depends how they are questioned. *Child Development Perspectives, 9*(4), 250–255. https://doi.org/10.1111/cdep.12142

Brubacher, S. P., Powell, M. B., Steele, L. C., & Boud, D. (2021). The use of a guided peer review assessment for forensic interviewers. *Journal of Forensic Practice, 24*(1), 1–17. https://doi.org/10.1108/JFP-07-2021-0040

Brubacher, S. P., Shulman, E. P., Bearman, M. J., & Powell, M. B. (2021). Teaching child investigative interviewing skills: Long-term retention requires cumulative training. *Psychology, Public Policy, and Law, 28*(1), 123–136. https://doi.org/10.1037/law0000332

Cederborg, A. C., Alm, C., Lima da Silva Nises, D., & Lamb, M. E. (2013). Investigative interviewing of alleged child abuse victims: An evaluation of a new training programme for investigative interviewers. *Police Practice and Research, 14*(3), 242–254. https://doi.org/10.1080/15614263.2012.712292

Cederborg, A. C., Lindholm, T., Lamb, M., & Norrman, E. (2021). Evaluating the quality of investigative interviews conducted after the completion of a training program. *Investigative Interviewing. Research and Practice, 11*(1), 40–52.

Cho, M. H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education, 34*(3), 290–301. https://doi.org/10.1080/01587919.2013.835770

Coniam, D., & Yan, Z. (2016). A comparative picture of the ease of use and acceptance of onscreen marking by markers across subject areas. *British Journal of Educational Technology, 47*(6), 1151–1167. https://doi.org/10.1111/bjet.12294

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795. https://doi.org/10.1037/0021-9010.84.5.795

Estepp, C. M., Thoron, A. C., Roberts, T. G., & Dyer, J. E. (2014). Variations in professional development needs of Florida agricultural education teachers based on gender, school level, and experience. *Career and Technical Education Research, 39*(1), 23–36. https://doi.org/10.5328/cter39.1.23

Gokmenoglu, T., Clark, C. M., & Kiraz, E. (2016). Professional development needs of Turkish teachers in an era of national reforms. *Australian Journal of Teacher Education, 41*(1), 113–125.

Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: results from a randomised controlled trial of quality teaching rounds. *Teaching and Teacher Education, 68*, 99–113. https://doi.org/10.1016/j.tate.2017.08.007

Guo, W. Y., & Yan, Z. (2019). Formative and summative assessment in Hong Kong primary schools: Students' attitudes matter. Assessment in Education: Principles. *Policy & Practice, 26*(6), 675–699. https://doi.org/10.1080/0969594X.2019.1571993

Harris, L. R., & Brown, G. T. L. (2018). *Using self-assessment to improve student learning*. Routledge.

Hershkowitz, I., Ahern, E. C., Lamb, M. E., Blasbalg, U., Karni-Visel, Y., & Breitman, M. (2017). Changes in interviewers' use of supportive techniques during the revised protocol training. *Applied Cognitive Psychology, 31*, 340–350. https://doi.org/10.1002/acp.3333

Joughin, G., Boud, D., & Dawson, P. (2019). Threats to student evaluative judgement and their management. *Higher Education Research and Development, 38*(3), 537–549. https://doi.org/10.1080/07294360.2018.1544227

Kersh, N., Evans, K., Kontiainen, S., & Bailey, H. (2011). Use of conceptual models in self-evaluation of personal competences in learning and in planning for change. *International Journal of Training and Development, 15*(4), 290–305. https://doi.org/10.1111/j.1468-2419.2011.00387.x

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Lamb, M. E. (2016). Difficulties translating research on forensic interview practices to practitioners: Finding water, leading horses, but can we get them to drink? *American Psychology, 71*, 710–718. https://doi.org/10.1037/amp0000039

Lamb, M. E., & Garretson, M. E. (2003). The effects of interviewer gender and child gender on the informativeness of alleged child sexual abuse victims in forensic interviews. *Law and Human Behavior, 27*(2), 157–171. https://doi.org/10.1023/A:1022595129689

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Mitchell, S. (2002b). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science, 6*, 35–41. https://doi.org/10.1207/S1532480XADS0601_04.

Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., Horowitz, D., & Esplin, P. W. (2002a). The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science, 6*(3), 114–125. https://doi.org/10.1207/S1532480XADS0603_2

Lawrie, M., Brubacher, S. P., Earhart, B., Powell, M. B., Steele, L. C., & Boud, D. (2021). Testing the effectiveness of a blended vulnerable witness training for forensic interviewers. *Journal of Family Trauma, Child Custody & Child Development, 18*(3), 279–297. https://doi.org/10.1080/26904586.2021.1894303

Liberty, L. M., & DeSpain, S. N. (2020). Self-regulated learners: Strategies that promote self-regulated learning in online environments. *Handbook of Research on developing engaging online courses* (pp. 48–71). IGI Global.

Ma, S.-C., Wang, H.-H., & Chien, T.-W. (2017). Hospital nurses' attitudes, negative perceptions, and negative acts regarding workplace bullying. *Annals of General Psychiatry, 16*, 33. https://doi.org/10.1186/s12991-017-0156-0

Makopoulou, K., Neville, R. D., Ntoumanis, N., & Thomas, G. (2021). An investigation into the effects of short-course professional development on teachers' and teaching assistants' self-efficacy. *Professional Development in Education, 47*(5), 780–795.

O'Toole, J., & Lawler, E. E. (2006). *The new American workplace*. Palgrave MacMillan.

Orvis, K. A., & Leffler, G. P. (2011). Individual and contextual factors: An interactionist approach to understanding employee self-development. *Personality and Individual Differences, 51*(2), 172–177. https://doi.org/10.1016/j.paid.2011.03.038

Panadero, E., Garcia, D., & Fraile, J. (2018). Self-assessment for learning in vocational education and training. In S. McGrath, M. Mulder, J. Papier, & R. Suart (Eds.), *Handbook of vocational education and training: Developments in the changing world of work* (pp. 1–12). Springer.

Poole, D. A. (2016). *Interviewing children: The science of conversation in forensic contexts*. American Psychological Association.

Powell, M. B. (2013). An overview of current initiatives to improve child witness interviews about sexual abuse. *Current Issues in Criminal Justice, 25*(2), 711–720. https://doi.org/10.1080/10345329.2013.12035992

Powell, M. B., Brubacher, S. P., & Baugerud, G. A. (2022). An overview of mock interviews as a training tool for interviewers of children. *Child Abuse & Neglect, 129*, Article 105685. https://doi.org/10.1016/j.chiabu.2022.105685

Powell, M. B., Fisher, R. P., & Hughes-Scholes, C. H. (2008). The effect of intra- versus postinterview feedback during simulated practice interviews about child abuse. *Child Abuse & Neglect, 32*, 213–227. https://doi.org/10.1016/j.chiabu.2007.08.002

Powell, M. B., Hughes-Scholes, C. H., Smith, R., & Sharman, S. J. (2014). The relationship between investigative interviewing experience and open-ended question usage. *Police Practice and Research, 15*(4), 283–292. https://doi.org/10.1080/15614263.2012.704170

Powell, M. B., & Snow, P. C. (2007). Guide to questioning children during the free-narrative phase of an investigative interview. *Australian Psychologist, 42*, 57–65. https://doi.org/10.1080/00050060600976032

Price, H. L., & Roberts, K. P. (2011). The effects of an intensive training and feedback program on police and social workers' investigative interviews of children. *Canadian Journal of Behavioural Science/Revue canadienne des Sciences Délután Comportement, 43*(3), 235–244. https://doi.org/10.1037/a0022541

Rasch, G. (1960). Probabilistic models for some intelligence and achievement test. Copenhagen, Denmark: Danish Institute for Educational Research. Expanded ed. (1980). The University of Chicago Press.

Richter, D., Kunter, M., Klusmann, U., Lüdtke, O., & Baumert, J. (2011). Professional development across the teacher career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education, 26*, 116–126.

Roick, J., & Ringeisen, T. (2018). Students' math performance in higher education: Examining the role of self-regulated learning and self-efficacy. *Learning and Individual Differences, 65*, 148–158.

Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., Lockyer, J., Loney, E., Mann, K., & van der Vleuten, C. (2010). The processes and dimensions of informed self-assessment: A conceptual model. *Academic Medicine, 85*(7), 1212–1220.

Smith, R. M., Powell, M. B., & Lum, J. (2009). The relationship between job status, interviewing experience, gender, and police officers' adherence to open-ended questions. *Legal and Criminological Psychology, 14*(1), 51–63. https://doi.org/10.1348/135532507×262360

Wilson, M. (2005). Constructing measures: An item response modeling approach. *Erlbaum Associates*.

Wu, M.L., Adams, R.J., Wilson, M.R., Haldane, A.A. (2007). ACER ConQuest [Computer program and manual]. ACER Press.

Yan, Z. (2018). The Self-assessment Practice Scale (SaPS) for students: Development and psychometric studies. *The Asia-Pacific Education Researcher, 27*(2), 123–135. https://doi.org/10.1007/s40299-018-0371-8

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education, 45*(2), 224–238. https://doi.org/10.1080/02602938.2019.1629390

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education, 42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

Yan, Z., Brubacher, S., Boud, D., & Powell, M. (2020). Psychometric properties of the Self-assessment Practice Scale for professional training contexts: Evidence from confirmatory factor analysis and Rasch analysis. *International Journal of Training and Development, 24*(4), 357–373. https://doi.org/10.1111/ijtd.12201

Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: The enabling role of feedback literacy. *Assessment & Evaluation in Higher Education, 47*(7), 1116–1128. https://doi.org/10.1080/02602938.2021.2001431

Yoon, I., & Kim, M. (2022). Dynamic patterns of teachers' professional development participation and their relations with socio-demographic characteristics, teacher self-efficacy, and job satisfaction. *Teaching and Teacher Education, 109*, Article 103565.

Zimmerman, B. J. (2000). Attaining self-regulation: A social-cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.

Zimmerman, B. J., & Labuhn, A. S. (2012). Self-regulation of learning: Process approaches to personal development. In K. R. Harris, S. Graham, & T. Urdam (Eds.), *APA educational psychology handbook* (pp. 339–425). American Psychological Association.