

Reflective self-evaluation of interviews

**Qualitative analysis of qualitative evaluation: An exploratory examination of  
investigative interviewers' reflections on their performance**

**Dr Andrew Griffiths, Visiting Fellow, University of Derby, UK**

**Dr David Walsh, Associate Professor, University of Derby, UK<sup>1</sup>**

---

<sup>1</sup> Contact email address [d.walsh@derby.ac.uk](mailto:d.walsh@derby.ac.uk)

**Qualitative analysis of qualitative evaluation: An exploratory examination of  
investigative interviewers' reflections on their performance**

**Abstract**

Self-evaluation of interviews conducted by law enforcement professionals is a principal feature of a prescribed interview framework in England and Wales, underpinning their practice development. However, self-evaluation has been found in prior research to be neglected. Building on our recent study (which found that interviewers regularly over-rated themselves, when compared to our independent ratings), the same interviewers assessed their interview skills by way of completing an extensive reflective log. We found that those we regarded as skilled in our prior study tended to be more accurate in identifying their strengths and areas for improvement, while planning to correct such shortfalls in their future practice. On the other hand, those we had earlier rated as least skilled tended to be much less reflective, being both descriptive and inaccurate in their understanding of key interview tasks. They also remained inaccurate concerning their own interview skills, failing to be prospective in planning to improve their skills. As such, while reflective logs appear to be, for skilled interviewers, both a prompt for accurate self-assessment and a catalyst for planning further professional development, we also caution that such tools need further refinement to achieve the same goals for those either less reflective or less skilled.

## **Qualitative analysis of qualitative evaluation: An exploratory examination of investigative interviewers' reflections on their performance**

The evaluation of interviewing performance, as a tool for professional development of interview skills, is a core element of the British PEACE model, being defined as central to the professional development of investigative interviewers (Shepherd & Griffiths, 2013). However, it was noted by these authors that little attention has been paid to the task of self-evaluation since the introduction of the model in the 1990s, largely (they argue) because it has been viewed as a complex task, existing within an expediency driven culture that prioritises process and outcome over the development of personal competence and skills.

There is a dearth of literature concerning how (and how effectively) law enforcement officers reflect and evaluate their own performance. A recent study, undertaken by Walsh, King, and Griffiths (2017), found that investigators regularly over-rated their own interview performance when compared to that of an independent expert. In that study, the participants assessed their own interview performance across 30 behavioural dimensions, using a five-point Likert scale. As their study required investigators to measure their performance on many behavioural dimensions, Walsh et al. reasoned that professionals (if they were being either honest or accurate) would, on occasions, have to attribute a low score, simply because interviewing is complex and no-one can perform skilfully all of the time. Since interviewing is one of the key (and frequently undertaken) tasks in criminal investigations, they continued, such truthful acknowledgement, when using the measurement scale, might well have been a difficult pill for them to swallow. The authors acknowledged that this methodological approach might have had the unintended consequence of both encouraging their participants to be less accurate about their performance, and also to overrate

themselves. This consequence was realised when the expert consistently awarded lower marks in his blind assessments of the same interviews. As such, the scale itself may have led to those less candid self-assessments. Nevertheless, given (a) the importance of investigative interviewing to an investigator's necessary repertoire of skills (Milne & Bull, 1999), and (b) the importance of evaluation within the PEACE model (Shepherd & Griffiths, 2013), it is imperative that an effective approach is found that encourages officers to be both reflective and also to enable them to accurately evaluate their performance to help them to progressively develop their interviewing skills. Further, quantitative field studies of investigators over several decades (e.g. Baldwin, 1993, Clarke & Milne, 2001, Griffiths, Milne & Cherryman, 2011, Walsh & Bull, 2012; 2015) have often found that interview performance is at unsatisfactory skill levels. As such, the present study was a sequel to the Walsh et al. study (2017), designed to examine whether a more reflective methodology would produce more accurate and insightful self-assessments.

### **The importance of reflection as a platform for skills development**

Moon (1999) describes reflection as a process that involves both thinking and learning either for personal future development or as a prompt for further reflection. Moon continues that reflection differentiates from mere description (which she argues relates to consideration, questioning, critique or developing meaning of an experience at only a superficial level). Reflection, in contrast, is defined by Moon as that which involves deeper levels of (i) analysis; (ii) attempts at behavioural explanation (that is characterised by "standing back", "mulling over", "self-questioning", and "critical awareness"); and (iii) acknowledgement that other prior experiences have interacted with, and accordingly influenced, the one being reflected upon. Such a process of

review, she concludes, prompts new ideas, reframing of perspectives or behavioural changes.

The benefits of reflection (as a partner to self-evaluation) have been well chronicled across many professional domains. For example, in the field of education, it has been found that reflection was the cornerstone of teachers' enhancement of the necessary skills that are core to their professional role (Baird, Fensham, Gunstone, & White, 1991). However, Argyris and Schön (1978) suggest that evaluating one's own professional performance is demanding, while Dunning, Johnson, Ehrlinger and Kruger (2003) have argued that inaccurate self-assessments should perhaps be expected when mediated by professionals who, say, either through inexperience and/or lack of expertise, display incompetence. This view would concur with Bloom's taxonomy of learning (1956), which maintains that there are six graduated domains of increasing difficulty related to the acquisition of cognitive skills, each one of which must be mastered before proceeding to the next. These are (i) knowledge; (ii) comprehension; (iii) application; (iv) analysis; (v) synthesis; and finally, (vi) evaluation. Bloom's model suggests that competent analysis and evaluation are not within the skills vocabulary of a practitioner until they are wholly competent in the knowledge underpinning the skill, mature comprehension of the same, and can apply the skill. Indeed, Kruger and Dunning (1999) found that improvements occurred among those responding positively to coaching in their metacognitive and logical skills. These findings highlight the importance of this particular set of cognitive skills in the act of reflection.

One of the earlier theoretical approaches to capture this concept was that of Kolb's experiential learning cycle (1984). The cycle, influenced by the even earlier work of Lewin (1951), commences with the concretisation of a specific event, which

Kolb ventured would prompt the learner to engage in a process that included self-critical feedback (by way of reflection), which in turn, would prompt future learning when similar circumstances were next encountered. In the context of the workplace, Schön (1983, p. 277) described this process as one of “reflection in action”. In brief, the knowledge and learning that is reckoned to emerge from such a cyclical exercise creates environments where “professional practice evolves from applying knowledge to practical situations” (Abbott, 1981, p. 826). Evaluation is an integral part of this process, and is multi-faceted. For example, feedback (as part of evaluation) can be further sub-divided into *outcome feedback* and *process feedback* (Hoffman et al., 2014), focusing on different aspects of performance. Hoffman et al. note the consensual view; that feedback is valuable to the acquisition of skills because it forms the basis of understanding (i) what needs to improve in the future; (ii) what steps are to be undertaken to trigger the desired improvement; and (iii) whether the undertaken remedial action actually led to improvements.

### **Evaluation as a platform for skills development**

The principles of Kolb’s model can be seen in the original PEACE training materials (Central Planning & Training Unit, 1992) where mock interviews and feedback sessions were included within the training model (Griffiths, 2008), and to an even greater degree in the later specialist training courses where an ability to both give feedback to colleagues and self-assess were specific competencies (Griffiths & Milne, 2006)., Given the general success of both these programs (Griffiths & Milne, 2006; McGurk, Carr, & McGurk, 1993) it might be wondered why the incorporation of Kolb’s model, or self-reflection in some other form, into real life PEACE practice, (and, indeed, wider law enforcement) has been less successful. Surveys of investigators since the turn of the century have regularly found inconsistent approaches to the task

of self- evaluation (where it is practised), and, as regularly, found that its practice not undertaken (Clarke & Milne, 2001; Walsh & Milne, 2007; Walsh & Bull, 2011).

Where self-evaluation has been practised, it is not uncommon to find that people tend to exaggerate their own capabilities. To the extent that such a phenomenon motivates individuals, while protecting their self-esteem, over-confidence might well be viewed as a good thing (Adams & Adams, 1960). Nevertheless, there are adverse impacts to this delusion, affecting law enforcement officers' beliefs as to how well they think they undertake key tasks, such as interviewing, and also acting as a hindrance to their improvement. That is, such wishful thinking may lead to over-optimism by professionals, leading to them deciding that there is no need to improve (Armor & Taylor, 2003).

Neutral of domain, it has been often found in the literature that the protection of self-esteem encourages people to over-assess their qualities, while overlooking their own shortcomings (e.g., Alicke & Govorun, 2005; Sedikides & Gregg, 2003; 2008). It has been argued that the motivation for this self-enhancement is a self-serving bias. Specifically, individuals undertake psychological re-constructions in attempts to assess their own attributes in a manner more advantageous to their own self-interests (Campbell & Sedikides, 1999; Sedikides, Campbell, Elliot, & Reeder, 2002). Indeed, Roy and Liersch (2013), in their study of self-assessment of driving skills, found a tendency for individuals to focus upon those skills in which they believed they possessed much ability (regardless of how vital these skills were to the task of driving). These authors also found a tendency for their participants to ignore some skills more central to the task (skills which third parties, familiar with the participants' driving ability, viewed as ones they did not undertake effectively). This led, in that study, to their

participants holding an idiosyncratic view, both of the required skills that defined good drivers, and of one's ability in undertaking them.

Roy and Liersch's findings correlate with those revealed in studies concerning the elicitation of professionals' beliefs with regard to investigative interviewing (e.g. Cherryman, 2000, Walsh & Milne, 2007; Walsh & Bull, 2011). In those studies, it was also found that there was no agreement as to an understood standard of interviewing excellence (and, implicitly, what poor interviewing involved). Bull and Cherryman (1996) also found that highly experienced detectives rated themselves as skilled, with none admitting to be 'poor'. (see also, Clarke & Milne, 2001; La Rooy, Lamb, & Memon, 2011 for similar findings).

As already noted, in our prequel study (Walsh et al., 2017), we found that investigators *consistently* exaggerated their own skills, when measured on a Likert scale. The participants rated themselves as more skilled in each of the 30 assessed behavioural dimensions (when compared with the blind ratings of an independent expert), with significant differences in twenty-seven of these dimensions. The findings of our prequel study provided strong evidence in support of the consensus in contemporary research literature with regards to 'what' is happening. However, in common with much quantitative research, our findings do not assist in explaining the 'why' of the phenomena. This conundrum led to the subsequent design of the present study, in order to gain a deeper understanding of both self-evaluation and reflection. The present study, therefore, focused on the content of reflection logs (completed by a representative sample of the participants from the earlier study, utilising a qualitative methodology), hypothesised to be better suited to deriving a deeper understanding from the text.

The use of both quantitative and qualitative methodology to explore interviewer performance has been successfully used previously. That is, having conducted quantitative research that provided strong evidence of skill enhancement attained by police officers that underwent specialist interview training (Griffiths & Milne, 2006; Griffiths, 2008), Griffiths and his associates subsequently utilised qualitative research methods in their next study (Griffiths et al., 2011). This later study explored the questioning strategies that were employed in interviews with suspects and witnesses that had been conducted by a smaller group drawn from the original cohort. The results of both studies, taken together, provided both statistical evidence that the questioning skills of the officers improved after training, and also a useful narrative about the complex nature of questioning, as understood by the participants. That is, Griffiths et al.'s (2006) first study provided the statistical evidence of the improvement, while their later study (Griffiths et al., 2011) provided evidence, not only concerning the reasons for such improvement, but also where further development of the training could be made. The present study was thus similarly designed with the aim of providing opportunity to enhance the results of the prequel study in a similar manner.

## **Method**

### **Participants and Procedure**

The participants in the present study (for which ethics authorisation had been given by the second author's home University) were the same 35 public sector fraud investigators from a small law enforcement agency in the UK, who had attended a three-day refresher interview-training course, and participated in the prequel study (Walsh et al., 2017). In that earlier study, the investigators interviewed mock suspects

concerning suspected crimes for a maximum of 45 minutes, based on real-life cases within their agency's jurisdiction (i.e. labour exploitation/modern slavery). All the investigators had earlier been trained in the PEACE model. In terms of their experience, twenty-five had conducted investigations for over ten years, with a further seven possessing between 5-10 years' experience, while of the remainder each had been serving as investigators for less than five years (and had each undertaken no more than 50 interviews). Of those with more than five years' experience, 27 stated that they had conducted more than 300 interviews, while five stated that they had conducted around 150-200 interviews. Twenty-five of the sample involved male investigators. The mean duration of the interviews in the present study was 37.03 minutes (SD = 7.62), in a range of 24-45 minutes.

Each of the investigators were given a copy of their video-recorded interview. They were then first asked to rate their performance, using a rating scale that had evolved from prior studies (e.g. Bull & Cherryman, 1996; Clarke & Milne, 2001; Griffiths, 2008; Walsh & Bull, 2010; Walsh & Milne, 2008, thus incorporating agreed interviewing skills, recommended in the PEACE framework). In sum, the scale possessed 30 dimensions of interview skills (measured on an ascending five-point Likert scale, where a score of '1' represented 'poor performance, '2' 'inadequate, and '3' being rated as satisfactory). A score of '4' would be applied to skilled performance levels, while a rating of '5' would be provided to highly skilled examples. Following other studies (e.g., Clarke & Milne, 2001; Griffiths & Milne, 2006; Walsh & Bull, 2010; Walsh & Bull, 2012; Walsh & Milne, 2008), a score of '3' on the scale would be the minimum viewed as acceptable. Once the investigators had independently undertaken their self-assessment task, they dispatched their assessments (within two weeks of the completion of the course by email to the second author, himself being highly

familiar with the rating scale and its definitions). An expert, possessing both 30 years investigative interviewing experience and a Masters' degree in criminal investigation, was recruited to undertake independent assessments of the sample. Having been trained by the second author in the rating scales, the expert (blind of the investigators' self-ratings) conducted his own assessments of the entire sample from a second copy of the video recordings. Before examining the self-ratings, thirty-three per cent (n =10) of the expert's ratings were subject to measures of reliability by the second author, generally finding a high level of concordance. That is, all scores were in excess of 0.80, using Cohen's kappa, except for two (*active listening* and *employing a logical interview structure*, where both scored 0.77).

Having undertaken the initial (numerical) self-assessments, investigators were then asked to complete a 2000-word reflection log in relation to aspects of their experience in undertaking the mock interview. The reflection log, based upon Gibbs' reflective cycle (1988), possessed six areas. These were (i) Description – *what happened*; (ii) Feelings – *what were you thinking and feeling*; (iii) Evaluation – *what was good and bad about the experience*; (iv) Analysis – *what sense can you make of the situation*; (v) Conclusion – *what else could you have done*; and (vi) Action Plan – *if it arose again what would you do*. The interviewers were given specific instructions to watch the recording of their interview independently, reflect upon the experience and then write a narrative addressing all those six areas. There was no specific word limit for each section. The logs were to be submitted no more than a month after their training had been undertaken. Twenty-seven completed reflection logs were submitted, and all were read in the initial phases of finalising the most appropriate method of analysis.

To better manage the large amount of data that was received, a two-factor purposive sampling rationale was designed to reduce the number of reflection logs analysed for this present study. Firstly, the expert's blind ratings from the initial study had identified a hierarchy of skill amongst the cohort (Walsh et al.,2017). Eight investigators had been rated in the earlier study as highly skilled (of which seven were males), 15 as skilled (of which seven were females) and four as low skilled (two males and females apiece). As such, it was important to distribute the data analysed evenly across the different skill levels to ensure that the results were not skewed by a disproportionate use of high or low skilled interviewers. As such, a sample of 12 reflection logs – evenly distributed across the three original skill levels was used in the present study. It is to be noted that definitions of interviewing skills pertain to their overall interview skill levels, not their skill levels relating to certain behavioural dimensions.

The final sample size of 12 reflection logs, involved all the four low skilled interviewers, as noted above, along with four highly skilled interviewers (three of which were males) and four skilled interviewers (two males and females). This purposive sampling process thus produced a cohort of seven males and five females, eleven of whom possessed over five years' experience as investigators (with four of that number possessing over ten years' experience).

In this exploratory study, the first author of the present study (who played no role in the assessments of the earlier study, though he was aware of those earlier ratings) analysed and coded the data contained in the essentially freehand reflection logs (which were repeatedly re-read). In order to identify themes, we undertook an adaptation of Interpretative Phenomenological Analysis known as 'thought listing procedure' (Petty & Cacioppo, 1981). A further variant of thought listing (known as

'Think Aloud') has been used previously to positive effect in qualitative analyses of interviewers' behaviour (Griffiths et al., 2011; Wright & Powell, 2006a; 2006b).

The present study's methodology also focused upon the data coding and analysis in areas of greatest interest within the overall content of the reflection logs. In the prequel study, a series of independent sample *t*-tests found that investigators' self-evaluation scores were higher than those of the expert along all 30 dimensions (with 27 of these found to be significantly different). As such, the present study restricted coding and analyses to those five behavioural dimensions with the greatest magnitude of *t*. These were when investigators provided (i) *a route map of the intended interview* (where  $t = 8.64$ ); (ii) *suspects with an opportunity to give an account* (5.65); (iii) *pauses and silences* (5.52); (iv) *periodic and final summaries* (5.76); and (v) *cognitive interview skills* (8.53). In addition, we reasoned that coding and analyses of the areas of most concordance between the expert and investigators might also assist in answering the research question (that is, those three behavioural dimensions where no significant differences were found prequel study; *preparation and planning*, covering the legal *points to prove*, and being *open minded*).

## Results

The results will be reported commencing with the three behavioural dimensions where (in our prequel study) there were no significant differences found between the assessments of both the expert and the investigator cohort in our prequel study, before proceeding to report our findings in the order of increasingly higher *t* values that were found in that study.

### Areas where no significant differences were found

#### *Preparation and Planning*

All 12 participants wrote about this element of the PEACE model (with the majority devoting extensive areas of their reflections to this dimension), emphasising not only its importance, but also its connection to overall interview performance. The reflections ranged from detailed and insightful entries, which included examples of what the participant did, accompanied with a rationale and the result, to basic statements that described following a process.

The highly skilled participants wrote about diverse aspects of the planning and preparation process, such as the dynamics between the lead and secondary interviewers, as well as the evidence and interview structure. There was a consensus from all participants that they were preparing the interview based on the information that was already in their possession, and that the information was both credible and unequivocal. None of the participants specified any contingency for the unexpected (as the training and guidance recommends that they should).

An example of a highly skilled participant writing about the preparation process comes from Interviewer 9;

*“We went through each piece of information and considered what area we would wish to discuss further with the interviewee. We then through each topic and considered what questions we may wish to ask”.*

Interviewer 14, another highly skilled participant was analytical about a negative element of their performance, writing, *“One area that was not discussed at the planning stage was the interaction between the interviewers, this would have an impact later on”*, then proceeding to admit that a lack of cohesion between lead and second interviewer led to incomplete coverage of key pieces of evidence. Interviewer 14 continued;

*“the roles and responsibilities of each interviewer should be agreed beforehand...this will avoid the possibility of the second interviewer interrupting or breaking planned silences...or pauses between questions”.*

Interviewer 15, another highly skilled participant made the connection between the planning process and events within the interview itself, writing;

*“I was not drawn into an argument when the roles of temporary workers were covered, the second interviewer also refused to be drawn and kept to the interview plan”.*

Interviewer 20 showed how the planning process had extended to consideration of question style, writing;

*“In my interview, I planned and prepared to use the funnel effect style of questions to achieve the best response. I would commence with a series of open questions that included TED questions then adopt a more probing style”.*

At the opposite end of the skills hierarchy, the lower skilled participants gave less comprehensive feedback, albeit technically accurate in relation to the PEACE model. Interviewer 10 wrote;

*“Planning and preparation for the interview was good, we identified objectives and selected topics to gain a detailed account’.* Interviewer 19 said *‘I adhered to my interview plan, covering overarching topics in a logical and sequential manner’.* Interviewer 8, *“we planned the questions we wished to ask, and the interview followed the plan”.*

Interviewer 1 reflected that

*“The interview moved through the topics as we had planned and was building to the challenge phase”*

It was noticeable from almost all of the reflections in this area that the participants' interview plans appeared quite rigid, as if there was a set pattern to follow. It was only Interviewer 14 (rated as skilled, as opposed to highly skilled), who introduced any notion of flexibility, writing;

*“Following a review of all available material, a written interview plan was produced with topic areas identified. I used this plan in interview but it was not intended to be rigid, and allowed for some fluency”*

#### *Coverage of the Legal Points to Prove*

This dimension received coverage in the reflection logs from all the high skilled and skilled participants, but with only one of the lower skilled participants including it within their reflections.

All participants, who included the dimension within their response, agreed that the interviews they conducted were only of value when seen in the context of the offences that were under investigation. Entries in this dimension were also heavily linked to both the *preparation and planning* and *open-minded* dimensions.

Highly skilled and skilled participants' reflections linked the concepts of planning and points to prove indicating a grasp of the approaching the interview with a view to progressing an investigation by supporting or negating the points to prove of a particular offence(s). The use of terms like 'defences' and 'mitigation' showed a balanced approach to the interviews, and an absence of guilt or confirmation bias from these participants (where the participant was conducting a suspect interview), or an objective view of the credibility of the information provided by the witness (where they were conducting a witness interview). For example, Interviewer 20 (a highly skilled interviewer) typified this professional approach, writing;

*“I undertook a review of the relevant points to prove for this offence, the available defences or mitigation that could be used by the suspect. I dealt with each topic separately to achieve all the elements to prove or disprove the level of involvement of this suspect”* (then listed those elements). Interviewer 12, also highly skilled, gave similar reflections to Interviewer 20;

*” I reviewed the relevant information and established points to prove, available defences or mitigation that could be used by the suspect”*

In contrast to the insight of the skilled and highly skilled participants, the only low skilled participant (Interviewer 10) to include this dimension, made rather more simplistic and basic comments;

*“We did establish the core points to prove the offence under the relevant law”*

That is, there was no mention of accommodating defences, mitigation or new information. The other low skilled participants wrote reflections that showed they had considered the information obtained as a result of the interview, but did not make the connection between that and the points to prove. Interviewer 8, for example, listed 12 questions that she required an answer to when describing her planning of the interview, and then listed the answer to each question within the reflection log, but did not articulate any assessment as to whether the answers received assisted in proving or disproving the allegation.

### *Open Mindedness*

All participants, regardless of their skill levels, made references to this dimension, with some more explicitly than others. Those more implicit in their comments appeared to demonstrate sentiments that might well lead to the avoidance of confirmation bias (if such feelings were replicated when actually undertaking

interviews with suspects). Interviewer 10, a low skilled interviewer summarised in one sentence the link between the three dimensions reported thus far writing;

*“Myself and X spent time planning and preparing for the interview, identifying what was required to prove or disprove the offence in order to obtain accurate and reliable information to ascertain the truth”.*

### **Areas of relatively moderate *t* values in the Prequel Study**

#### *Opportunity to Give an Account*

In this area, there was a marked difference in the reflections of the highly skilled and skilled participants when compared to the low skilled. Only one of the low skilled participants mentioned this dimension, whereas 90% (n=7) of the skilled and highly skilled participants made some reference to this fundamental reason for any interview with a suspect or witness. The reflections, however, were largely characterised by their apparent simplicity rather than being more insightful and revealing. For example, interviewer 9 stated;

*“I tried to allow the interviewee the opportunity to give me her account in her own words and uninterrupted by me”.*

Furthermore, there were only rare examples of the dimension being more deeply explored (that is, as part of rapport building in establishing the appropriate conditions to enable the interviewee to give maximum information disclosure). Interviewer 19, one of the low skilled participants, provided the most analytical point, wrote;

*“My failure to explain the reason for the interview was a fundamental flaw. The omission could have jeopardised the content of the entire interview. A clear*

*understanding by the interviewee of why they are there is imperative in ensuring co-operation, building rapport”.*

It was found more common, even from skilled interviewers, for them to provide comments associating this dimension with the objectives of the interview and the task orientated part of the process as opposed to facilitating effective conversation. For example, Interviewer 1 (rated as skilled) wrote;

*“it is vital that the reason for the interview is clearly explained, it is important to explain the objectives of the interview”.*

The same participant compounded this perspective adding, with respect to his future practice;

*“I will ensure that I provide a better explanation, I have been guilty of just referring to the offence without actually specifying what the details of the offence are”.*

### *Pauses and Silence*

Half of the participants (n=6) included reference to this dimension in their reflection logs, and the responses were distributed across the hierarchy of skill. Two of the lower skilled participants wrote comments concerning their own performance based on behaviour they also described in their reflection logs, whereas only one of the higher skilled participants mentioned the dimension. Three of the skilled participants reflected on this particular dimension. The content of some of the reflections revealed a distinct lack of knowledge concerning this dimension, arguably associated with the lack of reflections found when examining the *Cognitive Interview skills* behavioural dimension.

Interviewer 2 (skilled) observed that they made little or no use of *either pauses or silences*, noting that this would have allowed the interviewee time to think and

answer more fully. Interviewer 2 also reflected that this dimension required improvement in future interviews. Interviewer 14 linked their reflection to the planning stage, writing;

*“the planning did not include questioning strategies such as the use of silences to encourage the interviewee to reveal further information”.*

Interviewer 19, a lower skilled participant, demonstrated a fundamental lack of knowledge regarding the skill area, stating;

*“I have reviewed the interview and feel that silence would not have benefitted the outcome of the account provided by the interviewee”*

Also, adding;

*“Silence would ordinarily be used to reflect and revisit a question in their mind to which they have already stated a lie or partial truth”.*

Interviewer 28, also a low skilled participant, showed a more perceptive level of knowledge, stating;

*“I don’t feel that I used ‘appropriate silences’. I am fully aware of the benefit of them, I can sometimes be too keen to get to the next question”.*

This comment also gives insight to poor *listening* skills but the participant does not make this connection.

There were two particular comments that suggested a good understanding of the dimension and its tactical importance. Interviewer 12 (rated as skilled) wrote;

*“I believe I used silences and pauses effectively, in the first instance to allow the suspect to gather his thoughts, and also to give myself some breathing space to consider my next move”.*

The most comprehensive reflection was written by another skilled interviewer (#1), who wrote;

*“During the Account, Clarification and Challenge element it is important to use open ended prompts, active listening, and allowing the interviewee time to pause to search their memory.”*

### *Summarising*

The PEACE model recommends that interviewers summarise at the end of the first account and each topic. Despite this, however, reflections on this area were sparse amongst the reflection logs. Only 25% of the participants (n=3) included this dimension in their reflections. One of the lower skilled interviewers (#10) realised that she had not *summarised* at all during the interview. Interviewer 12 (a skilled interviewer) also realised that he had not *summarised*, whereas interviewer 1 (also a skilled interviewer) noted that he did *summarise* at the appropriate points. The only analytical critique with regard to this dimension came from Interviewer 14 (rated as highly skilled), but this was an abstract comment referencing an academic article that reinforces the importance of *summarising* within the model, as opposed to assessing her own use of the dimension.

### **Areas with the Largest *t* Values in the Prequel Study**

The last two dimensions that we analysed had both been found to have a large effect size in the prequel study. The analysis of these dimensions across reflection logs from all participants in this follow up study showed minimal consideration of these dimensions, as demonstrated by the paucity of comments relating to them. In relation to the first (that is, providing a *route map*) only one participant made any reference to this dimension. This was a skilled interviewer (#1) who stated that he omitted to utilise a *route map* at the start of his interview and also how this would have improved his

interview. In relation to the second dimension with largest  $t$  magnitude (*Cognitive Interview skills*), there were just two comments made. One made by a highly skilled interviewer, and the other by a low skilled interviewer. Skilled interviewer 15's reflection, however, was abstract and referred to the theoretical use of cognitive interview techniques as a means of increasing the amount of reliable information gleaned from an interviewee (rather than reflective comments concerning their own utilisation of the Cognitive Interview). The unskilled interviewer (#19) wrote (albeit inaccurately) that the circumstances of the interview conducted,

*“did not lend themselves to adopting a Cognitive Interview approach as this method of interview is more aligned to first hand witness events”*

## **Discussion**

This sequel study sought to ascertain whether narrative self-reflection logs provided more accurate and insightful feedback than numeric score evaluation sheets. The exploratory nature of the study required an innovative methodology dependent on two key factors; the self-reflection log itself and the analysis applied to the data within the logs. Both factors, could be viewed as potential limitations on the validity of the results, and so it is appropriate to address these concerns at the start of this discussion.

The reflection log was, effectively, a series of open questions asking participants to focus on broad areas of their experience (Gibbs, 1988). The design contrasts with a conventional questionnaire, or self-assessment, which contains a series of prescriptive or closed type questions offering a choice of answers or a numerical range for the participant to rate their experience and or ability. Appropriately for the subject matter of the study, the situation mirrors the difference between an

interviewer who uses a low number of open questions and one who asks a series of probing focused questions. Following this analogy, it is reasonable to assume, that the combination of open questions and the generous word limit would produce long replies (abundance of data), but that the replies (data) would be divergent on the basis that the reflection logs were completed independently by participants after their training course (and not prescriptive in asking set questions about specific behavioural dimensions).

The consequences of the reflection log design were that, while generally obeying the overall word limit, participants adopted different styles, which may reflect their personal preference of learning style when completing the reflection logs. Some opted not to follow the six-stage model (Gibbs, 1988), producing one long narrative, while others utilised part of the available space to either critique the logistics of the mock interview scenario, or describe the dynamics of the study group. Some participants did follow the six headings, but wrote extensively in the 'Description' area of the logs, and correspondingly, less in the Evaluation, Analysis, Conclusion and Action Plan areas (which were more important to the aims of the study). These actions might be interpreted either as ones relating to certain investigators failing to possess an inherent reflective learning style. Alternatively, it may well be that some participants found it less cognitively demanding to write descriptively rather than analytically. That is, even when investigators have much liberty to discuss in some depth their own performance, some are still unable to engage in reflection, stubbornly remaining in a descriptive mode. That is, they do not reflect or evaluate at anything other than what Moon (1999) would describe as a superficial level.

It could also be argued that the varying styles adopted by the participants was linked to their level of motivation in completing the task. One clue to answering this

question, might exist in the matter that they did comply with our request to complete them, taking likely several hours of their own time (when they not under any expectation or duress so to do), whereas around a quarter (presumably less motivated) of the original sample did not. Additionally, while we cannot overlook that issues of social desirability might have affected their responses, we assured the investigators that their reflective logs would not be shared with their employer, but only with the research team. Further, the candour (regardless of its accuracy) concerning criticisms of either their employer, the training products, or of wider criminal investigation processes, which were found in several of the logs, suggest that pleasing the research team was the least of their concerns.

Finally, the weaker logs may be a product of investigators being less articulate when undertaking such written exercises, rather than any lack of ability to reflect, or motivation. As such, it is suggested that future research examines both literacy and learning styles of participants. Nevertheless, given that many of the investigators possessed considerable experience, we might safely assume that they would be (i) capable of understanding and following the specific instructions given to them by the researchers, as well as (ii) being familiar with writing in-depth reports of their cases of similar length as the logs. As such, the task of writing the logs would be expected to be unlikely to have challenged them, even if they may have been less familiar with writing reflectively.

Therefore, overall, despite the potential limitations of the reflection logs, we argue that the design and manner in which the participants completed the reflection logs contributes to one of the major findings of the study, and is very helpful in answering the original research questions.

The second potential limitation concerned the analysis of eight of the 30 behavioural dimensions used as a basis for the coding and analysis, used in the earlier study (Walsh et al., 2017); that were themselves derived from previous research (e.g. Clarke & Milne, 2001; Griffiths, 2008; Walsh & Bull, 2010; Walsh & Milne, 2008). Using a rationale that not all behaviours in an investigative interview possess the same cognitive demands on the interviewer, Griffiths (2008) had further assigned all interview behaviours into 'simple' and 'complex' categories. The five areas, found in our prequel study, of most significant difference between the assessments of the investigators and the expert (namely *route map*, *Cognitive Interview skills*, *summarising*, *pauses and silence*, and *opportunity for suspect to give account*) would all be regarded as complex behaviours (Griffiths, 2008).

As such, it is argued, that the inability of lower skilled interviewers, found in the present study, to provide either accurate or insightful feedback concerning these complex behaviours is unsurprising. Further, the authors argue that these five particular dimensions were defined by the results of our earlier study and of greater interest *due* to the results of that research. In addition, the present study also incorporated analysis of three 'simple' behaviours (*preparation and planning*, *points to prove*, and *open mindedness*; Griffiths, 2008), where there were found no significant differences between the expert and participants in the earlier study (Walsh et al., 2017). Furthermore, the participants for the current study were purposively selected across the three categories within the skills hierarchy, and not just from the low skilled category. As such, the authors also contend that the choice of behaviours analysed is another strength of the study design as it covers both simple and complex behaviours, being balanced by the involvement of participants possessing a breadth of skill levels.

A further consideration in relation to the data analysis is that the first author of this study had to be aware of the skill levels of the interviewers (as rated in the prequel study), when both constructing the data sample and interpreting the results. As such, we cannot rule out the possibility that such existing knowledge may have skewed the present findings. However, both authors, having examined both sets of data contained in the present and prequel study, possess between them near 60 years of professional investigative experience, two PhDs, and have published around 30 of their research papers in peer reviewed journals concerning analyses of investigators when undertaking interview practice. While still not immune from potential biases of course, such extensive professional and academic experiences and expertise, should be considered as balance to such concerns.

Having addressed these important elements of the study design this remainder of the discussion will focus upon the results, arguing that the frequency, type and quality of reflections produced by the participants of this study are linked to their level of competence. We contend that these findings provide both accuracy and insight, believed superior to a simple numeric score sheet. Such numeric rating arguably provides accuracy, depending on who is assessing the interview (see Walsh et al., 2017), but is less effective in providing the necessary richness of insight, which was found in the present study.

The frequency of participants who wrote reflections for each of the six areas of the reflective cycle provides a major finding of this study. The number of participants writing reflectively decreased across the eight behavioural dimensions, inversely correlated to the increasing complexity of the dimensions. All twelve participants (100%) wrote reflections on both the (simple) *planning and preparation* and *open mindedness* dimensions (i.e. the behaviours where there had been no significant

difference between the scores of the participants and expert in the earlier study). With regard to the *points to prove* behavioural dimension (possessing a moderate  $t$  value in the prequel study), all highly skilled and skilled participants wrote reflections, while only one of the low skilled so contributed. In turn, eight participants mentioned *opportunity to give an account* (moderate  $t$  value) in their reflection logs – with, again, only one of the low skilled participants included in this figure. The number of participants making entries relating to *pauses and silence* (also moderate  $t$  value) decreased again to six, although in this dimension the participants who did mention the behaviour were evenly distributed across the skills range of scores.

Only three participants made specific mention of *summarising* (moderate  $t$  value). This downward trend was also found when we examined those dimensions where the largest  $t$  values were found in the prequel study. Here, only two interviewers mentioned *Cognitive Interview skills*, while just one made any mention of providing a *route map*. From the findings of both the prequel and the present study there appears an association between the number of occasions that participants made reference to a behavioural dimension in the present study and the more significant differences between interviewer self-ratings, and those independently rated by the expert (as found in the prequel study). That is, the *route map* behavioural dimension is *both* the area of most significant difference between the expert's view and participants' self-assessment in our prequel study, and *also* the behavioural dimension least mentioned by participants in the present study. These rare references to these three behavioural dimensions (i.e. *summarising*, *Cognitive Interview skills*, and *route map*), found in the present study, also reflect a pattern of their rare utilisation that was found in our prequel study (and, indeed, other studies of interview practice; e.g. Clarke & Milne, 2001; Walsh & Bull, 2010; 2012; 2015; Walsh & Milne, 2008).

The quality of the reflections within the self-reflection logs in the present study suggests that, among the simple behaviours, higher skilled participants appeared more analytical in their reflections than lower skilled participants (who were able to correctly identify elements of the process, but made fewer comments that could be interpreted as either evaluative or analytical). The lower skilled participants tended to describe events within the interview with less focus upon learning for future interviews. This finding indicates a level of competence commensurate with understanding the PEACE interview model (from a process perspective standpoint), but not one where they could convert their observations into tangible self-development strategies. In the more complex behaviours this picture changed, and the reflections of the higher skilled participants were also more functional, or absent completely. For example, in the *opportunity to give an account* dimension, only one participant linked the significance of the actual behaviour to rapport, a behavioural dimension acknowledged as a crucial element in investigative interviewing (Shepherd & Griffiths, 2013; St. Yves, 2014; Walsh & Bull, 2012).

It is also true, however, that there were exceptions to this overall picture. It was a low skilled participant who linked the behaviour above to the more important concept of rapport. This finding illustrates the point that, while the interviewers had been allocated to certain skill level categories (as a result of their overall ratings across all thirty behaviours in our prequel study), all interviewers will have particular individual skill strengths (Baldwin, 1993). The finding that almost all our participants failed to reflect on the *route map* and *Cognitive Interview skills* dimensions suggests that individuals tend to ignore important skills that they were rarely found (in the prequel study) to undertake. This finding was similar to that found in the study of driving

conducted by Roy and Liersch (2013), suggesting this is a persistent problem, neutral of domain.

Behaviours, defined by Griffiths (2008) as complex, were often ones that we found (in our prequel study) to possess those most significant differences between self and expert ratings. They were also ones (in the present study) rarely discussed in the reflective logs. This is an important finding concerning the role of reflection in the development of interview skills. Dunning et al. (2003) argued that inaccurate self-assessments were to be expected from the incompetent. However, although we tended to find only isolated inaccurate comments made by individual participants, the findings from this present study are argued to be more consistent with Bloom's taxonomy (1956), and Roy and Liersch (2013). That is, the lower skilled individuals appeared able to accurately comment on what they observed or felt, but lacked the analytical and evaluation skills required to independently formulate development strategies in order to write reflections aimed at developing their skills.

Bloom (1956), suggested that the ability to analyse and evaluate is not attainable until a learner both understands and can apply a cognitive skill competently. The findings of the present study suggest that lower skilled interviewers are less able than higher skilled interviewers to analyse and critique their own performance. However, we found that even higher skilled interviewers failed to reflect adequately when they examined their skills concerning the more complex behavioural dimensions. As such, this finding from the present study reinforces the argument (as expressed by Kruger & Dunning, 1999) that reflection is required to be combined with both feedback and coaching, both in relation to cognitive skills generally, and in regard to interview skills specifically (Griffiths, 2008, Walsh & Bull, 2011; Walsh & Milne, 2007). We argue that such approaches might well assist lower skilled interviewers in analysing and

evaluating their performance, and higher skilled interviews to maintain their development. Self-reflection appears inadequate as a development tool, unless it is combined with expert feedback (Hoffman et al., 2014; Shepherd & Griffiths, 2013). The role of the expert is to act as a coach, refraining from directive feedback, but rather guiding the interviewer to identify strengths and areas for improvement (Shepherd & Griffiths, 2013).

The present study also indicates that skills development in complex cognitive skills is more complicated than earlier suggested (e.g. Abbot 1981; Kolb, 1984; Schon,1983). This earlier literature tends to present skills development as a natural consequence of a cycle of action and reflection. In contrast, the present study strengthens the argument, put forward by Hoffman et al. (2014), that both the complexity of the skill involved, and the overall skill level of the practitioner, are relevant factors in this equation.

In conclusion, this present study sought to answer two questions. That is, firstly, whether the narrative reflection logs completed by the participants were ones that are more accurate, and (secondly) more insightful, as self-assessments of interviewer performance than earlier numerically-based assessments. In terms of defining whether the reflection logs are more accurate, the authors conclude that the logs do present a more faithful picture of the overall competence level of the participants. It also provides a portrayal as to whether their proficiency is either at the level of 'simple knowledge' or at a higher analytical level (Bloom, 1956). This more reflective form of self-assessment of professional competence is also, arguably, more insightful in terms of the individual's skills development than a numerical score of their own performance in a single interview scenario (real or simulated). As such, the authors contend that well designed narrative self-reflection logs are more valuable and superior to numeric

ratings, as one component of a coaching and development strategy to improve practitioners' skills.

We provided earlier our concerns regarding an organisational culture, driven by expediency rather than professionalism. We do not regress in our concerns, recognising that a comprehensive performance evaluation of a key skill for investigators (i.e. investigative interviewing) entails a more time-consuming approach by staff and their managers. In times when financial pressures may lead to possible sacrifice of quality, we need to counter such pressures if we are to be serious about increasing performance levels of investigators. It is important to ensure that considerable sums of money invested in initial training is not wasted later by a gradual and insufficiently managed deterioration of skills. We continue to witness a prevailing culture of reflection and evaluation in the fields of medicine and education, who each confront similar pressures upon resources as those that face law enforcement. Reflection and evaluation, after all, are the essential components of professionalism.

## References

- Abbott, A. (1981). Status and status strain in the professions. *American Journal of Sociology*, 86, 819-835
- Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, and J. I. Krueger (Eds.), *The self in social judgment* (pp. 85-106). Philadelphia: Psychology Press.
- Argyris, C., & Schön, D. (1978). *Organizational learning: A theory in action perspective*. New York: Addison-Wesley.
- Armor, D. A., & Taylor, S. E. (2003). The effects of mindset on behavior: Self-regulation in deliberative and implemental frames of mind. *Personality and Social Psychology Bulletin*, 29, 86-95.
- Baird, J. R., Fensham, P. J., Gunston, R. F., & White, R. T. (1991). The importance of reflection in improving science teaching and learning. *Journal of Research in Science Teaching*, 28, 163-182.
- Baldwin, J. (1993). Police interview techniques: Establishing truth or proof? *British Journal of Criminology*, 33, 325-351.
- Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay

- Campbell, K. W., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology*, 3, 23-43.
- Central Planning and Training Unit (1992). *A guide to interviewing*. Harrogate: Home Office.
- Cherryman, J. (2000) *Police investigative interviewing: Skill analysis and concordance of evaluations*. Unpublished PhD thesis. University of Portsmouth.
- Clarke, C., & Milne, R. (2001). *National evaluation of the PEACE investigative interviewing course*. Report no: PRAS/149. London: The Home Office.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–87. doi:10.1111/1467-8721.01235
- Gibbs, G. (1988). *Learning by Doing: A Guide to Teaching and Learning Methods*. Oxford: Oxford Further Education Unit
- Griffiths, A. (2008). *An examination into the efficacy of police advanced investigative interview training?* Unpublished PhD thesis. University of Portsmouth.
- Griffiths, A. & Milne, R. (2006). Will it all end in tiers? Police interviews with suspects in Britain. In T. Williamson, (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 167-189). Cullompton: Willan.
- Griffiths, A., Milne, R., & Cherryman, J. (2011). A question of control? the formulation of suspect and witness interview question strategies by advanced interviewers. *International Journal of Police Science & Management*, 13, 255-267. DOI: [10.1350/ijps.2011.13.3.219](https://doi.org/10.1350/ijps.2011.13.3.219)
- Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York: Psychology Press.
- Kolb, D.A. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134. doi:10.1037/0022-514.77.6.1121
- La Rooy, D., Lamb, M. E., & Memon, A. (2011). Forensic interviews with children in Scotland: A survey of interview practice among police. *Journal of Police and Criminal Psychology*, 26, 26–34. Retrieved from <http://dx.doi.org/10.1007/s11896-010-9072-9>
- Lewin K (1951) *Force Theory in Social Science*. New York: Harper and Row.
- McGurk, B. J., Carr, M. J., & McGurk D. (1993). *Investigative interviewing courses for police officers: An evaluation*. Police Research Group Paper No. 4, London: Home Office.
- Milne, R., & Bull, R. (1999). *Investigative interviewing*. Chichester: Wiley.
- Moon, J. (1999). *Reflection in learning and professional development*. Kogan Page: London.
- Petty, R.E. & Cacioppo, J.T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, Iowa: William C. Brown Co.
- Roy, M. M. & Liersch, M. J. (2013), I am a better driver than you think: examining self-enhancement for driving ability. *Journal of Applied Social Psychology*, 43, 1648–1659. doi:10.1111/jasp.12117
- Schön, D. (1983). *The Reflective Practitioner: How professionals think in action*. London: Temple Smith.

- Sedikides, C., & Gregg, A. P. (2003). Portraits of the self. In M. A. Hogg and J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 110-138). London: Sage Publications.
- Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, 3, 102-116.
- Sedikides, C., Campbell, W. K., Reeder, G., & Elliot, A. J. (2002). The self in relationships: Whether, how, and when close others put the self "in its place." In W. Stroebe and M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 12, pp. 237-265). Chichester: Wiley.
- Shepherd, E., & Griffiths, A. (2013). *Investigative interviewing*. Oxford: Oxford University Press.
- St-Yves, M. (2014). *Investigative interviewing: The essentials*. Toronto: Carswell.
- Walsh, D., & Bull, R. (2010). What really is effective in interviews with suspects? A study comparing interviewing skills against interviewing outcomes. *Legal and Criminological Psychology*, 15, 305– 321.  
[dx.doi.org/10.1348/135532509x463356](http://dx.doi.org/10.1348/135532509x463356)
- Walsh, D., & Bull, R. (2011). Benefit fraud investigative interviewing: A self-report study of investigation professionals' beliefs concerning practice. *Journal of Investigative Psychology and Offender Profiling*, 8, 131–148. Retrieved from <http://dx.doi.org/10.1002/jip.137>
- Walsh, D., & Bull, R. (2012). Examining rapport in investigative interviews with suspects: Does its building and maintenance work? *Journal of Police and Criminal Psychology*, 27, 73–84.
- Walsh, D., & Bull, R. (2015). The association between evidence disclosure strategies, questioning strategies, and interview outcomes. *Psychology, Crime and Law*, 21, 661–680. doi.org/10.1080/1068316x.2015.1028544
- Walsh, D., King, M., & Griffiths, A. (2017). Evaluating interviews which search for the truth with suspects: but are investigators' self-assessments of their own skills truthful ones? *Psychology, Crime, & Law*, 23, 647-665.  
DOI: 10.1080/1068316X.2017.1296149
- Walsh, D., & Milne, R. (2007). Perceptions of benefit fraud staff in the UK: Giving PEACE a chance?. *Public Administration*, 85, 525–540. doi:10.1111/j.1467-9299.2007.00645.x
- Walsh, D., & Milne, R. (2008). Keeping the PEACE? A study of investigative interviewing practices in the public sector. *Legal and Criminological Psychology*, 13, 39–57. Retrieved from <http://dx.doi.org/10.1348/135532506x157179>
- Wright, R. and Powell, M.B. 2006a. Investigative interviewers' perceptions of their difficulty in adhering to open-ended questions with child witnesses. *International Journal of Police Science & Management*, 8, 316–325.
- Wright, R. and Powell, M.B. 2006b. What makes a good investigative interviewer of children? A comparison of police officers' and experts' perceptions. *Policing: An International Journal of Police Strategies & Management*, 30, 21–31