REFLECTING ON EVIDENCE: LEADERS USE ACTION RESEARCH TO IMPROVE THEIR TEACHER PERFORMANCE REVIEWS

Eileen Piggot-Irvine
Royal Roads University

ABSTRACT
The paper reports on an action research (AR) project with six public high school leaders (reviewers) who volunteered to engage in an 18 month project to overcome their own defensiveness in addressing concerns with teachers (reviewees) whose performance they were evaluating. In the paper I outline how I acted as a coach in a long-term development approach where participant ownership of focus, data collection, analysis and interpretation was given highest priority. An exploration of the AR approach adopted, and the theory and strategies for addressing concerns is provided. The strategies may likely be a new, unique, contribution for many reviewers. A transcript of one reviewer-reviewee discussion sets the scene for an outline of reviewer tracking of their implementation strategies for improvement and subsequent evaluation. The final part of the paper covers a meta-level discussion of outcomes associated with the overall evaluation findings. Positive outcomes were shown for four of the six leaders for enhanced employment of strategies.

INTRODUCTION
In Canada, the approach adopted for performance review (also termed performance management, appraisal, and evaluation) varies from province to province, and within school districts in provinces. For example, in Ontario the teacher performance appraisal (TPA) system aims to have a fair, effective and consistent system with a public accountability component while promoting professional growth. An annual learning plan (ALP) is part of the growth intent. In the summative component, experienced teachers are assessed every five years against standards (Foundations of Professional Practice, Ontario
College of Teachers, 2010) outlined in five domains and 16 competency statements. In British Columbia (BC) a similar approach occurs with teacher evaluation (formal, summative, cyclical review) and teacher assessment (more formative, collaborative and may include self-assessment, goal setting and planning as well as external feedback). A minority of school districts in BC use a professional growth plan model as part of their assessment process, though such a model is not supported by the BC Teachers Federation (the union). Johnson (2012) described BC teacher evaluations a “hot-button topic” and suggested:

If you want to press a hot button inside or outside the hallowed halls of public education, teacher evaluation would be the one almost guaranteed to light up the "Emergency!" display on every teachers' union office desk.

Why is performance review such a contentious process? The answer is complex because there are personal, practical and political reasons for resistance. The purpose of this paper is not to examine the reasons, or performance review, per se. A comprehensive account of varied reasons for resistance to performance review can be found in previous publications (Piggot-Irvine, 2003a, 2003b, 2001). Rather, the focus in the paper is on just one potential underlying reason for such contention and resistance – reviewer defensiveness when they need to address concerns with reviewees. An action research (AR) project with six reviewers in school leadership roles was employed in an attempt to overcome such defensiveness.

In two decades of research on performance review and conducting over 100 principal performance reviews, I have experienced my own and other reviewers’ deep seated fear of engaging in conversations where concerns need to be raised: fear that often leads to highly avoiding and/or controlling defensive reviewer responses which appear to cripple the entire performance review system. Further, in my experience, many reviewers are often even more defensive about engaging in dialogue about their own defensiveness, or how to address that. In this AR project I needed to take care to ensure that my previous experiences did not lead me to influencing the way that participants drew their own insights and conclusions.

The paper begins with a summary of literature associated with effective performance review and the importance of addressing concerns, identification of defensiveness and ways to overcome it, and the AR methodology employed the 18 month project with a goal: To improve ability of performance reviewers to address concerns, areas for improvement, with staff. A transcription of discussion between a participant reviewer and reviewee introduces the AR project phases. Finally, the last sections of the paper cover conclusions, suggestions for further research associated with the project, and opportunity for dialogue about the paper.

**LITERATURE REVIEW**
Varied areas of literature were relevant to the project. The first is the topic of effective
performance review, including the importance of reviewers addressing concerns with reviewees. Theory of factors associated with defensive blocks to addressing concerns and then aspects of overcoming defensiveness with *productive* interactions is discussed in the next section.

**Effective Performance Review and Addressing Concerns Link**

Performance review is often maligned as a managerial imposition, or considered ineffectual for meeting its potential to improve teaching and learning (Barnett, 2006; Darling-Hammond, Amrein-Beardsley, Haertel & Rothstein, 2012). The malign has largely perpetuated due to lack of evidence showing improved teaching and learning outcomes (Donaldson & Donaldson, 2012; Forrester, 2011; Lissitz, 2012; Perillo, 2006; and Timperley, 1998). As Barnett suggested for the Ontario context:

The performance appraisal has the potential to be of significant value to teachers ... review of the related literature however indicated that outcomes from teacher evaluations are often not used to benefit practice. Current information on how today's teachers are implementing the results from the performance appraisal system to improve learning for their students is lacking. (p. ii)

Further, Down, Chadbourne and Hogan (2000) stated that:

... despite the official rhetoric of professional growth found in the policy, there is little evidence that performance management systems of this kind are effective in enhancing teachers’ learning and their capacity to improve their classroom practice ... (p. 213).

To extend the ineffectual reputation, Larsen's (2009) study in Ontario showed the majority of teachers (125 surveyed, 25 interviewed) considered the TPA system to have contributed to enhancing their levels of stress, self-doubt and anxiety.

In summary, although these authors characterize performance reviews negatively, the practice itself may not be fundamentally flawed. In Ontario there is evidence for teachers wanting such honesty and feedback for improvement (Barnett, 2006; Bolger & Vail, 2003). The reasons for reviewers providing limited or poor feedback for improvement are not entirely clear because little research has been conducted about this topic. However Maharaj (2013), also in Ontario, has drawn the following conclusion:

Almost all administrators reported rarely having to deal with teachers disagreeing with their assessment with 97% reporting that this happened only once in a while or never. ... it appears that the only time administrators had a teacher disagree with their assessment was when they were given an unsatisfactory rating, which does not appear to happen very often. ...some administrators reported receiving pressure from union officials when an unsatisfactory rating was given. ... it is very stressful and the federation can make your life just plain miserable” (p. 51).
The situation is troubling given the message from Bolger and Vail (2003) and Barnett (2006) that teachers want feedback. Such a situation has implications for school effectiveness overall, because if concerns are not discussed and improvement remains unaddressed there is reduced opportunity for enhanced teaching and learning.

Addressing concerns involves honest interactions beyond merely the sort of good listening and questioning skill strategies several early authors nominated as generally important for performance review (Edwards, 1992; Immegart, 1994; Marshall, 1995; Middlewood, 1997). The interactions needed should be situated at a deeper, non-defensive, level requiring what Goleman (1998) describes as high “self-awareness, self-regulation, motivation, empathy and social skills” or “emotional intelligence” (p. 94).

**The Honesty, Concern Addressing and Productive Reasoning Overlap**

In situations such as performance review where contention and threat are heightened, a set of automatic, highly conditioned, ineffective and *defensive* reasoning responses frequently dominate (Argyris, 1996, 2003; Dick & Dalmau, 1999; Piggot-Irvine & Cardno, 2005; Robinson, 1992; Senge, 1990). The two guiding values of defensiveness are winning (unilateral control) and avoidance (especially overemphasis on being nice when difficult messages need to be revealed). Neither value is likely to be associated with performance review effectiveness. Helping reviewers to overcome a defensive response and adopt strategies of openness when addressing concerns was the focus of this AR project. Strategies discussed may be relatively new knowledge contribution for performance reviews.

Openness and honesty are frequently associated with bilateral (shared rather than unilateral) control, mutual agreement and understanding about concerns, and employment of agreed resolution strategies. Such strategies have considerable overlap with seminal thinking about *productive* reasoning (Argyris, 1990, 1996; Argyris & Schön, 1996) based on the set of guiding values and key strategies summarised in Table 1.

Productive reasoning involves a sensitive balancing act between the two predominant values of advocacy and inquiry; two terms widely used in the literature on defensive and productive reasoning. An outcome should be a dialogue, or empirically informed “debate” (Robinson, 1992, p. 349), between reviewer and reviewee leading to a mutual understanding and agreement about issues (even if to disagree). The over-use or under-use of the values can result in reversion to either controlling or avoidance defensive strategies.

To expand on *advocacy*, Table 1 includes the statement “Invite challenge, evaluation, and public testing of those views”. The statement infers the tentative, almost hypothetical, expression of reviewer’s views and evidence about an issue in order to open up opportunity for the reviewee to challenge these views with their own views and evidence. In contrast, a reviewer’s presentation of their own evidence as indisputable is usually perceived as controlling. In advocacy, therefore, reviewer’s genuine intent to create reviewee response to their views is vital.
**Table 1: Productive Reasoning**

<table>
<thead>
<tr>
<th>Guiding values</th>
<th>Key strategies</th>
</tr>
</thead>
</table>
| **Increase Valid Information For All** *(Advocacy, i.e. advocating a position)* | * When working or talking with others:  
- share control by exposing rather than withholding key information; state position  
- share responsibility for goal achievement
* Disclose views, premises, and the evidence (hard data) or logic leading to those views
* Invite challenge, evaluation, and public testing of those views |
| **Enhance Freedom of Informed Choice (Inquiry)*              | * Treat views and reactions of self and others as hypotheses (rather than predetermined outcomes) to be tested
* Check to see how views have been understood and what views others hold; encourage and non-defensively receive others' views and disagreements without pre-judgement; check perceptions in ways which reveal implicit and explicit assumptions |
| **Gain Internal Commitment to Choice and to Monitoring (Bilateralism)* | * Seek bilateral solutions and joint responsibility for planning, implementing, and monitoring of achievement of goals
* Manage difficult emotional as a joint responsibility |

*Adapted from Piggot-Irvine (2012); Cardno (1994); Robinson, Absolum, Cardno and Steele (1990).*

*Inquiry*, as shown in Table 1, includes reviewer and reviewee checking and clarifying each other's views and evidence. Together with advocacy this step creates dialogue and dialogue, in turn, is designed to enhance ownership.

Presenting specific data is also particularly important in performance review. In Miller’s (2009) study with Ontario teachers, one of those interviewed said she “felt a little bit let down that he [the administrator] looked at the lesson overall instead of looking at the specifics of the lesson to be able to give [her] feedback.” (p.142). Schmoker (1999) summed up well avoidance of providing specific data as evidence, and its roots and consequences, in explaining:
Why do we avoid data? The reason is fear - of data’s capacity to reveal strength and weakness, failure and success. Education seems to maintain a tacit bargain among constituents at every level not to gather or use information that will reveal a clear need for improvement: where we need to do better, where we need to make changes. Data almost always point to action - they are the enemy of comfortable routines. By ignoring data, we promote inaction and inefficiency (p. 39).

The work of many previous authors (Argyris, 1990, 2003; Argyris & Schön, 1996; Cardno, 1994; Senge, Cambron-McCabe, Lucas, Smith, Dutton & Kleiner, 2000; and Robinson, 1992) helped inform the approach for enhancing productive reasoning employed in this project. My own research (Piggot-Irvine, 2001) had already revealed that short-term training (one to two days) was largely ineffective in helping reviewers to sustain non-defensive, open, approaches in performance review. The long-term project reported upon in this paper was designed to not only allow time for the “... complexity of the curriculum ... to be ... learnt, internalised and acted upon” (Cardno, 1994, p. 222), but was also associated with follow-up support and implementation. As I suggested several years ago, because avoidance is a conditioned response, learning to overcome such response is a lengthy and difficult task and “requires rethinking and altering our underlying value systems, and this involves changing many automatic, conditioned responses” (Piggot-Irvine, 1995, p. 140). For the latter reason the AR development project was 18 months in duration.

Within the 18 months deliberate spacing of events allowed participants to assimilate and internalise their new learning. The project was intended to produce double-loop learning outcomes (Argyris, 1996) where reviewers would be helped to examine and reflect upon their underlying governing values and strategies for resolving issues with reviewees.

Argyris’ (1990, p. 95) suggested four stages for learning productive reasoning guided my approach. I have described the stages (Piggot-Irvine, 2003b, 2012) as: first, issue mapping; second, diagnosis of the extent to which reviewers had created and maintained any defensive response; third, helping leaders to take productive reasoning from an espoused theory (a belief state) to a theory-in-use (implementation phase); and fourth, repeating the learning experience in multiple events. The AR methodology employed to structure the learning is described next.

**Methodology: AR approach for development**

The type of AR methodology employed for the project was similar to that which I have used in multiple development projects and is discussed in detail previously in publications (e.g. Piggot-Irvine, 2012; Piggot-Irvine & Doyle, 2010; Piggot-Irvine & Bartlett, 2008). In brief, AR has grown in popularity as a developmental research methodology with data collection (research) and change (action) elements. AR includes iterative planning, acting, reflecting and evaluating phases and has been described by Zuber-Skerritt (2012) as practical, democratic, participative and collaborative. Claims of personal, team, organizational, and
**Preparation:**
- Clarifying principles underpinning model
- Confirming commitment, roles, ethical issues
- Purpose and priority setting
- Planning for phases
- Negotiating logistics and resources

**Underpinnings:**
- Evidence based decision making
- Theoretically informed
- Clear planning
- Authentic collaboration

*Figure 1: Focused Action Research (FAR) Model*
community (and less frequently global) improvement, sustained change and transformation are often touted by authors such as Coghlan and Brannick (2004); Cardno (2003); Stringer (2007); and Wadsworth (2011), with some also emphasizing outcomes of empowerment or emancipation (e.g., McTaggart, 1991; Reason & Bradbury, 2001; Stringer, 2007). Regardless of intent, a degree of flexibility and open-endedness is evident in AR process and it is contextually specific rendering the findings non-generalizable.

The AR model I have developed, the Focused Action Research Model (FAR) as shown in Figure 1 (adapted from Piggot-Irvine, 2015), has several elements designed to enhance rigour. For example, the model incorporates: informed decision making with an emphasis on collection of data/evidence rather than unsubstantiated assumptions; becoming informed via examination of theory and previous research preceding implementation of action; planning as a component of each phase; and authentically collaborating with others participating in the project to ensure member checking, critique, peer review of all findings.

In the FAR model an initial preparatory phase focusing on project intent is followed by a reconnaissance, or informed current situation analysis, phase. Proceeding phases include: implementation of improvement; evaluation of achievement; creation of recommendations for further improvement; reporting of findings; and links established for further action associated with improvement. An iterative, or cycling and often overlapping, phased activity therefore occurs in the model, as shown in Table 2.

A key reason I chose AR as a methodology for this specific project with reviewers was because the approach offered a development context which could foster the already noted high ownership by the reviewers. Ownership was enhanced when reviewers responded to an invitation to participate (though as reported later in the participants section one reviewer was instructed to respond by his principal). Ownership enhancement also occurred through reviewers themselves collecting and analysing raw data from events such as videotaped classroom observations or audiotaped reviewer-reviewee conversations in the reconnaissance and evaluation phases. Further, although I held debriefing and clarifying dialogue sessions with each reviewer following analysis in these phases, the reviewers led those sessions by offering their own analysis of conversations. In dialogue sessions regularly conducted with the whole participant group, I strove to adopt a facilitative role designed to enhance their ownership of the session. In the reconnaissance and implementation phases I acted as a guide to helping them distil their own criteria for analysis (based on the theory linked to defensive and productive reasoning strategies) and facilitated training/development sessions for implementation of productive skills.

I have used my AR model in other school leadership projects focused on improvement of implementation of productive reasoning. In Piggot-Irvine et al. (2011), for example, the approach to development and outcomes were reported for seven principals. As co-authors of the latter paper, the principals themselves strongly expressed their thoughts on both the process of AR for improvement as well as the commitment they needed to make to the double-loop learning required for change to their practice. Further, in Piggot-Irvine and Doyle (2010) I jointly reported with another principal on a similar AR project for
improving implementation of productive reasoning with a senior leadership team in her school. Here, the focus of the team was on creating enhanced trust with teachers. The substantial self-report of enhanced trust by the senior leaders was triangulated with teacher feedback as well as an independent national audit team review. Though both of these latter projects involved improvement in productive reasoning implementation, and resulted in reported relational openness, neither project was focused within the complex and threatening performance review context that existed for the project reported upon in the following sections of this paper.

<table>
<thead>
<tr>
<th>Table 2: FAR Model Phases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preparatory</strong></td>
</tr>
<tr>
<td>clarifying principles underpinning model</td>
</tr>
<tr>
<td>confirming commitment, roles, ethical issues</td>
</tr>
<tr>
<td>setting purpose and priorities</td>
</tr>
<tr>
<td>planning for phases</td>
</tr>
<tr>
<td>negotiating logistics and resources</td>
</tr>
<tr>
<td><strong>Reconnaissance</strong></td>
</tr>
<tr>
<td>participant information gathering (literature, previous research etc.) on effective practice with the relevant topic(s) and distilling key points: becoming informed</td>
</tr>
<tr>
<td>drawing up criteria for analysing own practice from distilled key points collecting data/evidence on current situation in terms of practices (and gaps) associated with the goal drawing conclusions about where improvement is needed</td>
</tr>
<tr>
<td><strong>Implementation</strong></td>
</tr>
<tr>
<td>engaging in deeper development on areas for improvement drafting a plan for the improvement steps, with clear timelines and measurable outcomes carrying out the planned improvements recording reflections on implementation as change progresses</td>
</tr>
<tr>
<td><strong>Evaluation</strong></td>
</tr>
<tr>
<td>collecting data/evidence on achievement of improvement steps drawing conclusions about where further improvement is needed</td>
</tr>
<tr>
<td><strong>Recommendations and Reporting</strong></td>
</tr>
<tr>
<td>presenting and reporting findings drawing up recommendations for next steps</td>
</tr>
<tr>
<td><strong>Further Improvement</strong></td>
</tr>
<tr>
<td>planning for further improvement associated with next steps.</td>
</tr>
</tbody>
</table>
Participating in the project, I conducted several seminars in two public high schools on creating productive interactions when addressing concerns in performance review and then subsequently invited leaders who were performance reviewers to participate in a longer term AR development project. Six leaders (four Heads of Department, one Deputy Principal, and one Principal) responded to the invitation: five participants were male and one female. To anonymize respondents, participants as reviewers have been alphabetically labelled A1 to A6 and their reviewees as Ae1 to Ae6 i.e. A1 is the reviewer and Ae1 the associated reviewee.

As already noted, my role in the project was as a coach, developer and facilitator of the AR process: in every way possible I attempted to empower reviewers to be self-directed and have high ownership. I was acutely aware that despite my collaborative intentions I was also an outsider in the way Anderson and Herr (2005) noted as approximately a 4-5 (outsider in collaboration with insiders) on their continuum of positionality from 1 (insider) to 6 (outsider). My intent in the role with the FAR model was to help participants understand the AR process phases, to assist them in learning how to use theory for establishing criteria for analysis and understanding of defensive and productive strategies, to clarify how they would collect evidence and analyse and report on such evidence, and to coach them to work in authentic ways where dialogue and productive strategies underpinned their collaborative process (Piggot-Irvine, 2012).

The AR phases guided the project. The following sections outline activity in the phases.

The Beginning: Preparatory and Reconnaissance Phases
At the initial preparatory AR phase the participants met with me to: clarify principles underpinning the AR and its approach; confirm involvement; jointly clarify our roles; secure understanding of ethical issues; define the project purpose and goal; set up group and individual plans for project phases; and negotiate logistics and resources for time release for involvement.

A set of reconnaissance, or current situation analysis, phase activities quickly followed. I introduced participants to literature associated with effective performance review and the importance of addressing concerns as well as that of creating open, productive, non-defensive, interactions. The intent was for reviewers to be informed before starting any analysis. My emphasis in this activity was on coaching participants themselves to distill key points from theory. They also drew up their own criteria for addressing concerns from literature on defensive and productive strategies and used the criteria to establish an analysis protocol for both the reconnaissance and evaluation phase discussions with reviewees.

Next, each reviewer collected data/evidence on their current practice in addressing concerns with reviewees by individually audiotaping a performance review discussion. Immediately prior, and subsequent to, taping each reviewer outlined to me their intended discussion and then afterwards outlined how well they felt the discussion had been
conducted. Approximately a week later each reviewer received and analyzed the transcribed discussion using their pre-developed analysis protocol and I then met with them again to debrief their analysis.

The following transcript extract provides an example of the results of this process. Most importantly at our pre-meeting A1 had told me he had considerable concerns with the observation that he needed to discuss with Ae1.

A1: I was very impressed with the manner you had with the kids, a good manner. Others could benefit from seeing you. Work was pitched at just the right level ... if you want to discuss any of this just let me know and stop me. Everyone was doing something. You used good questioning. Spot on for the class.

Ae1: It’s good to get positive feedback. I don’t get it from all staff. I didn’t do anything special for you coming. I felt comfortable with you there. I used a variety of questioning techniques. I mainly used talk and whiteboard.

A1: I was impressed. You related the lesson to their own level ... it was concerned with math in building ... the money ... practical stuff.

Ae1: One thing I’m trying is to not give them the correct jargon at first. I use anecdotal expressions. Then later I introduce the mathematical terms. Then I insist on them using these when asking questions. They’re difficult to manage as a class. Some of them only got 23% last year.

A1: One thing needed (and I forgot to do this) was to compliment them before I left the room. I thought they did really well. The worst kid was really good, meek and mild.

Ae1: He’s not naughty...attention seeking ... he seems to work well with me ... responds well. I can empathise with these kids.

A1: Every teacher has a class we could put at the bottom of the pile.

The four pages of transcript continue in much the same way as the extract. A1’s initial response when we met straight after the discussion with Ae1 indicated he was partially aware of not being clear about his concerns but could not be specific beyond vague awareness. He conducted transcript analysis during the next week using the pre-developed protocol covering defensive and productive strategies and revealed: multiple positive statements usually without evidence or reasoning supporting the statements; only one negative statement in the discussion with no support of evidence or reasoning; limited inquiry or probing into Ae1’s perception of what had happened; frequent jumping to his own conclusion without her input; and offering his own solutions to Ae1’s questions but no examples of engaging her in deriving her own solutions. Overall, A1 summarized he showed little dialogue through a balance of advocacy and inquiry, and minimal use of
evidence. He concluded that the result was a discussion leading to limited bilateral solution generation and no planning for implementation of improvements.

Following discussion of his own analysis of defensive and productive interactions in our debriefing meeting A1 strongly committed to learning more open, productive, interactions and this commitment deepened when A1 engaged in further dialogue with the whole project team dialogue session at completion of the reconnaissance phase analyses. In that team dialogue session, as with all subsequent group sessions, participants began by engaging in dialogue about their own analyses then we focused on strengthening use of productive strategies. By the end of the second and final reconnaissance phase group meeting all but A2 had shared their individual analysis summaries and they had helped each other draw conclusions about improvement areas for the implementation phase.

Following this meeting, as with many earlier and subsequent interactions, I met confidentially with A2 and used every productive reasoning and counselling skill I could draw upon to try to understand his unresponsive approach. Without repeatedly returning to this issue of his response, I will quickly summarize here that it was only late in the project that he revealed he was instructed by his principal to attend and was angry about that. The revelation led to extended dialogue between the principal and myself to resolve a difficult situation that both myself and A2 had been put into given that the project was premised on voluntary participation. I note this issue to clarify why some results for A2 may indicate resistance and low commitment and I am not sure I would have felt any different if I had been in A2’s situation.

The Implementation Phase: Enhancing Ability to Confront Issues
Once project participants had drawn up key areas for improvement, we then moved into the implementation phase. A first step was to strengthen the group development of productive strategies. I facilitated two day experientially-based sessions using multiple well-researched and documented tools for revealing and resolving defensiveness; see Piggot-Irvine & Doyle, (2010) for a summary of the development tools and Piggot-Irvine, (2003b) for training approach. For example amongst tools used was the two column analysis (Senge, 1990) employed for reflecting on discrepancies between what was thought (recorded on the left side column of a sheet of paper) and what was actually said (recorded on the right side column) between reviewers and reviewees. Also utilized was practice and analysis of reviewer ability to: stay low on the ladder of inference (remain at a low level of assumption in conversation); move to double-loop learning (surfacing, examining and resolving underlying issues); reflect in (on the spot) and on (after the event) action; and engage in a dialogue process.

By the end of the development session participants were ready to tighten the improvement area focus. A1, for example, identified a focus on the two key areas of revealing specific evidence in the negative advocacy component of dialogue and enhancing hearing the reviewee perception. In essence, he wanted to enhance both advocacy and inquiry when discussing concerns. After clarifying the focus, participants drafted a detailed plan covering implementation of productive strategies linked with performance review events, and
scheduled regular meetings for group dialogue to debrief practice and analysis of discussion transcripts. In the subsequent six months participants carried out their implementation plans and recorded analyses and reflections progressively.

The next phase of development involved evaluating effectiveness of the implementation strategies.

**Evaluation Phase**
Evaluation occurred at three levels in the project. First, each reviewer progressively evaluated their improvement through analysis of transcripts of reviewer-reviewee discussions linked to performance review events.

Second, I also conducted an overall end of project evaluation (covering both process and facilitation) via use of questionnaires and interviews with each reviewer and their reviewee. The use of the multiple methods was designed to triangulate data (Cohen, Manion & Morrison, 2007; Denzin, 1997) in order to search for convergence of findings. Since I employed reviewer self-report questionnaires in this part of the evaluation, it was particularly important to triangulate questionnaire data to avoid what Bandura (1986) described as potential for bias resulting from reviewers inaccurately reporting on, and perhaps inflating positively, their interactions with reviewees. The latter could be considered a type of self-protective response (Watson, 1993) in evaluation of practice.

With reviewer permission, I conducted a third type of evaluation at a meta-level through recording of all of my meetings with reviewers and a combined analysis of all of their individual transcript analyses. In the following sections, my meta-level recording is drawn upon for reporting extent of implementation of key values and strategies of productive and defensive reasoning under the headings of advocacy, inquiry, and gaining internal commitment to improvement.

**Advocacy**
Meta-level examination of reviewer transcript analyses indicated the six reviewers had no trouble being positive with their reviewees. Being open about positive issues is easy: of more interest was whether reviewers advocated negative issues. Reviewers themselves were well aware of the emotional difficulty associated with such advocacy, as A1 revealed:

> To some extent only as truth hurts and we need to work together, we need to share resources, we socialise. By confronting the issue we may offend etc. etc.

The difficulty became evident early in the reconnaissance phase when four reviewers (A2, A3, A4, A5) had made conflicting statements about whether they had areas of concern with reviewees, as the following example statement from A5 shows:

> A5: No, no problems, but somewhere I want to somehow retrain Ae5’s manner.
There is a possibility that A5, like others making conflicting comments about concerns, may have unknowingly or knowingly defensively covered-up, minimalized and therefore avoided acknowledging the issues. The latter avoidance is in keeping with an observation previously reported more generally with reviewers (Piggot-Irvine, 2003a).

Despite potential initial defensiveness about acknowledging concerns or improvements needed, four reviewers (A1, A3, A4 and A6) showed strong improvement in employing advocacy with negative issues throughout the implementation phase. They became consistently more effective in both stating their concerns and also providing data to support concerns and associated reasoning/logic. The remaining two reviewers (A2 and A5) showed little evidence of employing advocacy associated with any concerns.

**Inquiry**

Inquiry (the second value of productive reasoning), with its implicit checking component, was also increasingly employed by all but two of reviewers (A2 and A5 again). It could be said that A2 was exceptionally honest given that his espousals of not inquiring matched his practice because from the outset he was extremely forthright in rejecting implementation of productive reasoning. A2 stated openly that dealing with concerns was irrelevant in his relationships with staff, and A2's practice seemed to indicate the same. Neither advocacy nor inquiry featured in his interactions with A5 and false reassurance alongside advice giving statements dominated.

With A5, inquiry took what looked like two counterproductive forms. A5 either used closed or rhetorical questions that outweighed all other aspects of productive reasoning implementation and resulted in an imbalanced and what seemed to be an avoiding performance review process, or A5 used little inquiry about A5's perceptions of issues. The latter was indicated through a limited level of probing to explore issues, as demonstrated in the following example:

A5: Okay, but you're okay about how to use that device and how it has to be recorded
A5: mmmm
A5: Well those are the key tasks and the job description as we have it.

**Gaining internal commitment to improvement**

Overall, for all reviewers less success appeared evident in terms of the AR project leading to implementation of the third value of productive reasoning i.e., the reviewer helping the reviewee to gain internal commitment to improvement through use of skills of bilateralism in solution generation, planning and monitoring. A1 had the biggest issue. A1’s tendency to provide solutions unilaterally remained unchanged, despite the transcript data indicating the tendency in initial stages of the project. The following example from towards the end of the implementation phase indicates provision of solutions:
All other reviewers also had issues with helping the reviewees find their own solutions, but to a much less extent than A1. With A3 and A4, the tendency to provide solutions reduced a little as the implementation phase progressed, but they still showed limited bilateral solution generation. Such a tendency to offering solutions is reported elsewhere associated with the teaching profession (Smith, Crutchfield & Culbreth, 2001).

A3 and A4 also did not jointly create strong reviewee plans for improvement despite their own development in the project itself included each reviewer creating clear and deep level planning showing intent to comprehensively examine and improve an issue (Piggot-Irvine, 2003a). Such planning was not apparent in five out of the six action plans developed by reviewers with reviewees. Only A6 showed evidence of mutual (bilateral) planning for improvement at a deep level. All other plans demonstrated a surface, or more superficial, approach to development which included such statements as attend a workshop. Further, only A3 implemented any monitoring of planning.

**Outcomes Overall**

The findings overall from the AR project suggest that five of the reviewers (A1, A3, A4, A5, A6) were open to learning productive reasoning skills. They did not, as Cardno (1994) suggested, “resist the unlearning of instinctive skills”, or “block the learning of new skills” (p. 223). Furthermore, the findings somewhat refute Cardno’s (1994) contention that few people employ productive reasoning in problem or improvement situations such as performance review. I would offer that the findings in this project show that many reviewers can develop such reasoning. I also add that with A5 specifically, though the learning may not be immediately obvious, an openness to learning might suggest that the productive skills could be evident in time.

An indicator of success of the project centred on whether the intended double-loop learning outcomes (Argyris, 1996) were produced i.e. whether the productive strategies employed helped reviewers to actually address problems. The findings showed that despite initial denial of problem existence by most reviewers, A1, A3, A4 and A6 were all insistent in self-reports that they addressed concerns and, further, that problems were resolved. Their reviewee feedback supported the self-report in the meta-level interviews I conducted.

In the case of A3 and Ae3, evidence from student work added support to A3’s assertion that a concern of lack of varied, creative and interactive teaching techniques had been addressed with Ae3. In the following meta-level interview I conducted with A3 and Ae3 there is verbal confirmation of the outcome (note I have anonymized the option chosen as ‘zzz’ because it was specific enough to identify the teacher):

*A1: Could you consider perhaps as an intermediate ....
And later
Well they can spend more time going through ....*
Ae3: I’m working on mixed ability strategies with more confidence.

A3: Yes, Ae3 definitely did achieve this. The students in Class A were questioned by me about whether Ae3 was using more interactive teaching. The students were particularly positive - especially the students involved in the activities. The students (every one of them) in the activities reported that they really enjoyed this and felt it had helped them more than formal work.

I suppose the performance review process enabled Ae3 and I to zero in on perceived weaknesses regarding creativity in approaches. We implemented a process to change this. But here goes one bit of evidence. In the end of year exam, almost all Ae3’s class chose the zzz option. There were about 10 options and they had to pick three and I thought not many kids would pick this option and almost all of A3’s guys chose to do the zzz option (the one A3 had diversified with).

**DISCUSSION OF FACTORS INFLUENCING SUCCESS IN IMPLEMENTATION**

Triangulated data collected in the project has led me to propose several success factors associated with the demanding development approach. I have labelled these as commitment and motivation, openness to learning and non-defensive honesty, consciousness-raising, narrowing the espousal-practice gap, extended support, and repeated experience. Each is discussed in the following sections.

**Commitment and Motivation**

I believe the success of the AR project, in terms of whether reviewers were helped to develop a productive process, primarily hinged on commitment of the reviewer. Four of the five reviewers (A1, A3, A4 and A6) who had total commitment to implementing productive reasoning skills also showed the greatest gains in terms of self-reported improved practice and their reviewees confirmed this. Commitment was not linked to my prompting – it seemed to be derived from an intrinsic motivation to do the very best for both themselves and their reviewees and student learning. The intrinsic element is in keeping with a conclusion drawn about motivation generally by both Dinham and Scott (1998), and myself (Piggot-Irvine, 2000), that teachers’ strongest commitments were altruism and personal growth values. The conclusion is also in keeping with the following drawn by Down et al. (2000):

The teachers in our study learned and changed, not because they were monitored and managed by others, but because they all shared one important attribute - a professional conscience that committed them to caring for their students and continually seeking better ways to help them achieve their potential (p. 219).

**Openness to Learning and Non-Defensive Honesty**

Openness to learning, and non-defensive honesty about performance, were also evident for A1, A3, A4, A5 and A6. They were all exceptionally honest and accurate in analysing productive reasoning practice, even when they knew their practice was initially defensive. Apart from the earlier mentioned initial denial of any improvement needed by these
reviewees, they did not try to subsequently cover-up their defensive responses. An anomalous situation was evident for A5 in terms of follow-through to implementation of the skills, but I have nothing but admiration for his openness to learning.

**Consciousness-Raising**
Consciousness-raising derived from transcript analysis was another factor I propose may have influenced reviewer success. Once these reviewers analysed their initial transcripts they all saw the extent of the gap between the theory and practice of productive reasoning implementation. I consider that change only occurred for four of the reviewers when they saw they had a problem or need for improvement in practice. In Down et al.’s (2000) terms this was when reviewers saw or felt “the need for help” (p. 219).

Consciousness-raising occurred most intensely for A1, A3, A4 and A6. A2 articulated an interest in the transcript findings, but did not record areas for improvement. A5 showed considerable interest and conducted detailed transcript analysis but in following sessions with Ae5 there was little evidence that his practice changed. I engaged in many dialogue sessions about the latter with A5 and can report that his enthusiasm to try never waned.

I believe the indisputable evidence of existing practice which the other four reviewers saw, in Dick and Dalmau’s (1999) terms, “generated the dissonance” (p. 5) so necessary to motivate improvement. The evidence starkly highlighted the gap between reviewers' espoused theories and theories-in-use (Argyris, 1986). The gap identification was a turning point in learning for A1, A3, A4 and A6, and one which was in keeping with Argyris’ (1990) suggestion of trainees needing to diagnose the extent to which they create and maintain problems.

**Narrowing the Espousal-Practice Gap**
The findings of this research showed that most reviewers (the anomaly was A5) who were committed and open went beyond just highlighting the espousal-practice gap to narrowing the gap. As the earlier summary of findings shows, four of the reviewers made considerable shifts to “to resolve the inconsistency” (Dick & Dalmau, 1999, p. 5) which the evidence of practice highlighted. They made a conscious effort to improve.

**Extended Support**
I believe another reason for enhanced success shown for four of the reviewers could have been due to their enthusiastic acceptance of the extended support I offered. The long-term characteristic of the project offered me an opportunity to provide intensive assistance for those reviewers who wanted it: assistance that needed to be challenging of trainee assumptions and actions on the one hand, and respectful of the fragile learner on the other. A critical feature of support was a requirement for a high trust and open relationship between myself and reviewers. I needed to model my espousals in a non-defensive approach if I wanted a parallel process to occur between reviewers and their reviewees.
Repeated Experience
In keeping with Argyris’ (1990) suggested fourth stage for training for productive reasoning, the project offered reviewers opportunity to repeat their learning experience with on-going performance review interactions during the year. The repeated experience extended and deepened learning opportunities by providing further examination of continuing practice.

CONCLUDING COMMENT
Findings from the AR project have provided some evidence to answer questions Nathan, Mohrman, and Milliman (1991) posed many years ago about whether “performance reviews actually change subsequent employee performance” (p. 252). Though discussion of performance reviews impacting on teaching is not the specific focus of this paper, the finding from Ae3 adds moderate evidence to work of the likes of Stiggins and Duke (1988) showing how performance review can have an influence on personnel development and teaching.

Whether the project will have made a sustained difference to the reviewers’ implementation of productive strategies is, as yet, unknown. Many years of training may be necessary to help reviewers to rethink and alter their underlying value systems - the systems which should be altered if the approach is to be genuinely implemented. With continued practice of the productive strategies I am hopeful the sustained success in implementation may result. In the meantime, however, the project has had an immediate impact on several of the reviewers, as Ae6 said about their reviewer A6:

Unbelievable advances/changes in just one year. Full credit to A6 preparedness to adapt. This was the main area needing development. Very few niggles this year.

At a personal level, after facilitating more than 50 development sessions on productive strategies, I continue to be humbled by the challenge of the difficult emotional and cognitive learning for participants. The project reported upon in this paper was probably the most challenging I have conducted mainly because, as noted earlier, the challenge is enhanced due to the performance review context itself being fraught with threat and complexity. At times, the latter created defensiveness for both myself and participants, particularly when data from transcripts necessitated heightened levels of openness to improvement in following implementation steps. I believe the AR approach with its emphasis on creating high ownership, dialogue, collaboration, transparency and openness, alongside evidence based improvement, has been an invaluable approach for reducing such defensiveness.

FURTHER STUDY AND FURTHER DIALOGUE
In this project, many questions remain unanswered and give rise to opportunities for further exploration. Included are questions associated with the AR process facilitation such as its relevance, effectiveness and areas for improvement, as well as questions about the longer term impact and transferability of both productive strategies and AR itself to varying contexts. Consequently there are multiple potential areas for further study arising
from the project. Just a few of those include: investigation of the value of expansion the AR development approach to a broader group of performance reviewers; examination of whether the strategies for improvement noted have been sustained for reviewers and reviewees in this project; and collection of data from a wider range of those directly impacted by the project.

A mixed methodology evaluation of several aspects of the project process, outcome and impact is currently planned for 2016 as part of an international *Evaluative Study of Action Research* (ESAR) with 16 AR projects. In the ESAR, in-depth case studies will include documentary analysis, surveys, interviews, and focus groups with all 16 project participants and all boundary partners i.e. those impacted by projects. As one of the 16 chosen, the project reported upon in this paper will be comprehensively evaluated for longer term impact and the findings reported in future papers.

Earlier I have noted that dialogue, or empirically informed “debate” (Robinson, 1992, p. 349) which includes a balance of advocacy and inquiry, is a desired feature of most AR. The participants in the AR project described in this paper have highlighted that it is a challenge to practice a non-defensive, dialogical and productive, approach when giving and receiving feedback. Like the participants, I find it easy to implement such an approach when in non-contentious situations but, again like the leaders in this project, I also struggle to maintain a consistent approach when challenged with uncomfortable feedback. As shown in this paper, a key to lowering defensiveness in contentious situations is to authentically create opportunities for enhanced practice. In keeping with such intent and with the iterative, ongoing, nature of AR, an aim of reporting on this somewhat contentious project in this paper is to invite such dialogue from others. I welcome your input and challenge.

**REFERENCES**


Dr. Eileen Piggot-Irvine is a Professor of Leadership at Royal Roads University (Victoria, Canada) and an Adjunct Professor at both Griffith University (Brisbane) and Unitec (Auckland). She was formerly Director of the New Zealand Action Research and Review Centre (NZARRC), Director of the New Zealand Principal and Leadership Centre (NZPLC) and senior lecturer at Massey University, and Head of the Education Management Centre, at Unitec. She established the New Zealand Action Research Network (NZARN) and was formerly Vice-President of the Action Learning Action Research Association (ALARA). Most of her contract work has been based on an action research model.