Are We Collaborating Yet? Employee Assessment of Peer's Perceptions

Mr. Benjamin Heslop (corresponding author)

School of Medicine and Public Health, Faculty of Health and Medicine

University of Newcastle, Callaghan Campus

University Drive, Callaghan, NSW 2308, Australia

E-mail: benjamin.heslop@uon.edu.au

+61240420345

Dr. Elizabeth Stojanovski

School of Mathematical and Physical Sciences, Faculty of Science

University of Newcastle, Callaghan Campus

University Drive, Callaghan, NSW 2308, Australia

E-mail: elizabeth.stojanovski@newcastle.edu.au

Dr. Jonathan Paul

School of Medicine and Public Health, Faculty of Health and Medicine

University of Newcastle, Callaghan Campus

University Drive, Callaghan, NSW 2308, Australia

E-mail: jonathan.paul@newcastle.edu.au

Dr. Kylie Bailey

School of Medicine and Public Health, Faculty of Health and Medicine

University of Newcastle, Callaghan Campus

University Drive, Callaghan, NSW 2308, Australia

E-mail: kylie.bailey@newcastle.edu.au

Abstract

Employers rarely utilise their employees' capacity to assess the collegiality and productivity of their own work unit, yet they are determinants of employee retention and profitability. One reason is the lack of a reliable, valid survey instrument to measure collaboration viability (CoVi), which we postulate is the construct that employees use to implicitly assess their work unit. Inherent weaknesses of own-perception and peer-assessment instruments prevent them reliably measuring CoVi. A novel method overcoming respective deficiencies by combining the strengths of both approaches is proposed that we term peer's-perception. It is contended that such an instrument may be improved through formulation in accordance with a universal model of collaboration. The model chosen is PILAR as it encapsulates a variety of social and organisational psychology theories. Prospects, involved, liked, agency and respect constitute five Pillars of collaboration (Heslop, Bailey, et al., 2017). Based on this review, we propose a peer's-perception instrument (Pillar-PP) and that this instrument be formally evaluated.

Keywords: productivity, psychological safety, cohesion, collaboration, peer's-perception, PILAR

1. Measuring Collaboration Viability

Collaboration is a synergistic, voluntary combining of individual contributions for a collective goal (Hughes & Jones, 2011; Thomas et al., 2007). Viability of collaboration can be threatened, such as by interpersonal conflict, poor strategy, or selfish manoeuvring (Chen, Ünal, Leung, & Xin, 2016; Topping, 2010). Collaborators are instinctively wary of such behaviours, and monitor their threat to collective success. Upon sign of collaboration unviability, the member's first recourse is to limit their ongoing investment, such as dedicating less time or money. Once perceived risks outweigh the likelihood of success, leaving the collaboration may be the only option, knowing that doing so may cause sacrificing investments already made (Kendall & Salas, 2015).

It is desirable to ameliorate threats to collaboration before members curtail their participation, or withdraw completely, since either may become a self-fulfilling prophecy (Jehn & Mannix, 2001; Morgeson & DeRue, 2006). Reasons for collaboration breakdown highlighted by the literature are free-riders, interpersonal conflict and poor leadership (Kuhn, 2015), which may be ameliorated through retraining and changes to; roles, workflow processes, oversight systems, and delegation (Daley, 1992). To ensure that failing collaboration in the workplace may be identified and repaired, the present study seeks an instrument that can reliably measure *collaboration viability* (CoVi).

The standard method of measuring constructs reminiscent of CoVi is to survey group members for their perceptions of the group, which this article terms *own-perception* (Edmondson, 1999). A second approach is *peer-assessment*, which is rarely used to measure CoVi, but offers certain advantages over own-perception. We examine the weaknesses and

strengths of these two instruments (own-perception and peer-assessment) and then combine the two approaches to create a novel *peer's-perception* instrument.

1.1 Own-Perception Instruments

Three constructs are chosen for their relevance to CoVi: psychological safety; group harmony; and group cohesiveness. Psychological safety is the extent to which a team feels safe for interpersonal risk-taking, such as suggesting an idea or admitting a mistake (Edmondson, 1999). Group harmony is reflective of intra-group relationship quality, characterised by attribution of benign motives when disagreement is encountered, and an overall balance between individual needs and group unity (Chen et al., 2016). Cohesiveness refers to aspects of liking, task commitment and group pride (Lott & Lott, 1965). Each construct has accompanying instruments designed to assess a respondent's collaborative environment from their perception of its personal impact (Newman, Donohue, & Eva, 2017). Unfortunately, assessment of one's perceptions of the group is unconsciously influenced by various pro-self biases (Mathieu, Gilson, & Ruddy, 2006).

A bias towards self-enhancement tends to make perceptions of the group unduly negative because each respondent feels they are insufficiently celebrated by the group (Lönnqvist, Leikas, Verkasalo, & Paunonen, 2008). Secondly, a self-reporting bias encourages respondents to hide potentially embarrassing aspects of themselves. Hence, poor collaborators will instinctively seek to shift attention from themselves by emphasising fault elsewhere (Donaldson & Grant-Vallone, 2002). Thirdly, research on growth versus static mindsets shows that those with a static mindset tend to have an overly-pessimistic view of others' intentions (Chiu & Dweck, 1997). Those with a static mindset are not only likely to be poor collaborators, but also likely to blame others for problems they are experiencing, delivering an overly-negative view of CoVi (Zingoni & Corey, 2016). Due to these biases, there is a genuine risk of false positives when identifying poor CoVi using own-perception instruments (Hoenderdos, 2013).

1.2 Peer-Assessment Instruments

Peer-assessment instruments ask respondents to rate whether, or to what extent, a peer demonstrates certain behaviours or traits (Kusano, Conger, & Wright, 2015). While peer-assessment of employees has been undertaken for millennia (Wiese & Buckley, 1998), there remains uncertainty in terms of what qualities in peers to measure (Koh, Shibani, Tan, & Hong, 2016). In examining the literature, no peer-assessment instrument was found that had been used to measure CoVi, although numerous authors have anticipated an equivalent application (Koh, Hong, & Seah, 2014; Salas, Shuffler, Thayer, Bedwell, & Lazzara, 2015; The iTOFT Consortium, 2015).

1.3 Weaknesses of Peer-Assessment

Reliability and validity are metrics of instrument usefulness and accuracy (Cho, Schunn, & Wilson, 2006). A *reliable* instrument has consistent meaning across cohorts, time and in differing situations, implying that the instrument can be applied more universally (Messick, 1995). *Validity* occurs when an instrument is correctly measuring what it claims to measure.

Respondent disengagement is a major confounder in survey instruments, which can lead to inaccurate answers, a primary cause of which is poor question design (Hess & Stathopoulos, 2013), which can ultimately affect a measure's reliability and validity.

The respondent completes a peer-assessment survey for each peer, for example, up to five replicates for a team of six. Respondents that become mentally or emotionally exhausted from answering too many, or difficult, questions, may lose accuracy, and perhaps withdraw early. Yet, it is common practice in social science to sequence instruments, potentially accumulating hundreds of questions. The implicit presumption that instruments maintain their reliability and validity when part of a larger survey ignores potential disengagement due to accumulated effects of poor question design (Fowler, 1995). Specific weaknesses in instrument and question design that may cause respondents to disengage will now be examined, based upon the authors' previous research (Heslop, Stojanovski, Paul, Iverson, & Bailey, 2017).

1.3.1 Cognitive Dissonance

In the context of peer-assessment, cognitive dissonance is experienced when the respondent's emotional reaction to a question differs from the required answer (Claussen, 2004). For example, scoring a punctual, yet irritating, peer highly for attendance causes cognitive dissonance.

1.3.2 Opacity

When information regarding a peer is unavailable or unclear, the respondent perceives a question as opaque (Lord & Maher, 1990). Opacity depends primarily upon whether the respondent has observed peer behaviours relevant to the question. However, other factors, such as the respondent's social insight and empathy, their familiarity with the peer, and the peer's transparency and authenticity, can also affect responses (Wildman et al., 2012). For example, the question "do you trust your peer" assumes the ability to perceive others' hidden motivations (Bargh, Gollwitzer, & Oettingen, 2010). Own-perception instruments suffer less from opacity because the respondent's opinion is data for the construct being measured.

1.3.3 Varying Interpretation

When a question is poorly worded or conceptually unclear, respondents' interpretations will vary (van Zundert, Sluijsmans, & van Merrienboer, 2010). For example, brusque emails are preferred 'electronic communication' by some, whereas others expect a formal letter style. Own-perception may also suffer from varying interpretation, but since the instrument is applied only once, can average out misinterpretation with multiple, similar questions.

1.3.4 Cognitive Load

Cognitive load is incurred peer-assessment questions require substantial intellectual effort, for instance estimation of both quality and quantity of a peer's behaviour, followed by a comparison of their estimations against a numerical scale (Brody & Dietz, 1997). Repeated such comparisons within a single question are highly taxing because assessments are done in conscious memory against specific incidents recalled from long-term memory (O'Neill,

McLarnon, & Carswell, 2015). Peer-assessment surveys are typically applied to multiple peers, so shorter instruments that impose minimal cognitive load are preferred (Tziner & Kopelman, 2002).

1.3.5 Inter-Rater Bias

Inter-rater bias distorts the respondent's objective capacity to evaluate a peer. Potential causes of inter-rater bias are familiarity and liking, the latter of which can increase ratings of liked peers by 63%, relative to a neutral relationship (O'Neill et al., 2015; Viswesvaran, Schmidt, & Ones, 2005). Inter-rater bias can also prompt deliberate gaming of the instrument, for example, scoring a particularly disliked peer poorly on every question.

1.4 Remedies for Peer-Assessment Weaknesses

1.4.1 Score Matrix

Peer-assessment produces a matrix of scores that can be interrogated manually or via algorithms for inter-rater bias, or attempts to game the instrument (Propper & Wilson, 2003; Willey & Gardner, 2009).

1.4.2 Prior Instruction

Verbal instruction or a written rubric given to respondents prior to the survey can increase consistency of question interpretation (Cho et al., 2006; Kusano et al., 2015). Often a secondary aim of peer-assessment instruments, instruction has been shown to improve participant's enduring pro-social behaviours (Ceston, Levine, & Lane, 2008; Lievens, 2001). However, due to additionally imposed information and time, a rubric may also decrease engagement (Hastie, Fahy, & Parratt, 2014).

Rather than being provided within a separate document, rubric information may instead be pinned to each question, an approach used by a Behaviourally Anchored Rating Scale (BARS) that attaches descriptive *anchors* to answer levels (Kingstrom & Bass, 1981). Regarding the earlier example of electronic communication, one behaviour might be listed as an illustration of an 'unsatisfactory' level, while the other is 'above average.' Just as with a rubric, BARS information may cumulatively impose a burden, but its proximity to each question reduces the onus to refer to a separate document (The iTOFT Consortium, 2015).

2. Short, Universally-Valid Instrument

Fewer, clearer peer-assessment questions will minimise respondent disengagement, but a briefer instrument tends to be less universally reliable. One way to achieve both brevity and reliability is to craft questions directly against a universally-ideal peer. Such a model of the ideal collaborator is yet to be developed, and may indeed never exist since people suit different collaboration objectives, for instance requiring creativity or efficiency (Elfenbein, 2008).

In response to a lack of universal theory, questions may be rendered more general and abstract, such as "Is the peer a team player?" However, this supports varying interpretations, for example, is a pleasant and friendly peer who excessively seeks guidance on their tasks

from colleagues a team player? Weighing varying interpretations may then increase the respondent's cognitive load, and perhaps cause cognitive dissonance. Abstract questions also add to cognitive load by invoking a larger number of incidents from memory, which are weighed to arrive at a numerical answer. Simultaneously addressing tensions in peer-assessment design would be assisted by a universal model of the ideal collaborator, but in the meantime, existing instruments have made tradeoffs.

2.1 Examples of Peer-Assessment Tradeoffs

High reliability implies an instrument will be applicable to a wider range of contexts, for example, not-for-profit as well as for-profit organisations, or remote and in-person groups (Hastie et al., 2014). Yet there exists a trade-off between reliability and validity, since the greater detail and specificity that questions possess, the less often particular behaviours will be observed, reducing reliability (DeCoths, 1977). On the other hand, forming questions around behaviours that are more abstract invites varying interpretation and increases cognitive load, threatening validity (Tziner & Kopelman, 2002).

2.1.1 Behaviourally Anchored Rating Scales (BARS)

One method of obtaining greater validity is modifying questions to suit a context, which sacrifices reliability. For instance, BARS instruments are routinely modified after consultation with experts or representatives of the cohort (Debnath, 2015). Unfortunately, such a modified survey is not transportable to other cohorts (Ohland & Loughry, 2006), and for this reason BARS instruments are rarely recognised as broadly reliable (Panadero, Romero, & Strijbos, 2013).

2.1.2 Behavioural Observation Scales (BOS)

BOS allow greater context-specific modification, and ensuing loss of reliability, than BARS. Rather than the BARS approach of asking respondents to rate the subjective quality of a behaviour, BOS instruments ask how frequently a behaviour occurs (Konak, Magluilo, & Kulturel-Konak, 2016). BOS instruments therefore require an explicit, countable behaviour, for example, meeting attendance.

BOS instruments have less potential for cognitive dissonance, since a behaviour either does or does not occur. On the other hand, background behaviours often critical to collaboration, such as whether a peer is polite or respectful, cannot be counted. Not allowing these generic abstractions prevents a BOS instrument from becoming generalisable between different contexts, for example, a medical compared to a military setting.

2.1.3 CATME Instrument

Aiming to develop a validated instrument for peer-assessment of collaborative ability, the Comprehensive Assessment of Team Member Effectiveness (CATME) took a unique approach to developing its instrument. Initially, literature review of prosocial research generated hundreds of collaborative behaviours. Over 1000 students were then asked by CATME developers to rate each question for relevance, giving a short-list of 87, and subsequently five factors (Hughes & Jones, 2011, p. 59):

"(1) contributing to the team's work (for example, "Did a fair share of the team's work"), (2) interacting with teammates ("Communicated effectively"), (3) keeping the team on track ("Stayed aware of fellow team members' progress"), (4) expecting quality ("Expected the team to succeed"), and (5) having relevant knowledge, skills and abilities ("Had the skills and expertise to do excellent work")."

While CATME is ostensibly generally-applicable, it was initially too lengthy for peer-assessment within larger groups. CATME responded by offering an abbreviated instrument containing 33 questions (Hughes & Jones, 2011), which some believe remains too long (Ohland et al., 2012).

Existing peer-assessment instruments demonstrate a trade-off between length and universality. A short survey is unlikely to be universal, while a long survey may impose excessive cognitive load. On the other hand, allowing flexibility, as BARS does, risks questions that are opaque, or incur cognitive dissonance in the respondent. The BOS scale, while easier to adjust to context, must feature countable behaviours. Neither BOS nor BARS is likely to be reliable outside their original setting.

3. Novel Peer's-Perception Approach

A peer's-perception instrument asks the respondent to assess how each of their peers perceives the group, which has never to the author's knowledge, been suggested or implemented. Guidance is sought from own-perception and peer-assessment approaches, in order that their respective strengths may be utilised, and weaknesses overcome.

3.1 Addressing Weaknesses of Own-Perception through Peer's-Perception

An instrument that uses peer's-perception, rather than own-perception, may escape biases prompted by respondents' desire to guard their equanimity (Kim, Baek, & Kim, 2011). Self-enhancement bias is conceptually avoided since the respondent is asked to consider the perceptions of others, and not themselves. Unfortunately, instruments associated with established CoVi-related constructs, such as Edmondson's (1999) survey of psychological safety, are not suitable for assessment of peer's-perceptions.

Taking psychological safety as a representative example, one question asks, 'is it safe to take a risk on this team?' but when applied to peer's-perception, the question would become 'do you perceive that this peer feels safe to take a risk on this team?' When extended to a peer's perception, such a question may be too specific and technical to be evaluated without unrealistically close familiarity with the peer's intentions and perceptions. This is typical of own-perception questions that triangulate with specific questions concerning the same construct to overcome pro-self biases, which not only are too exact for peer's-perception, but which also extend the instrument's length.

3.2 Addressing Peer-Assessment Weaknesses through Peer's-Perception

Since evaluating a peer's perception is not an evaluation of the peer themselves, peer's-perception questions may be less likely to experience cognitive dissonance. Consider a peer whom the respondent feels has an incorrect perception of the group, whether unduly

positive or negative, and who is thus not respected. An accurate answer by the respondent is effectively a report to the survey provider of the extent of the peer's delusion. On the other hand, consider a peer who the respondent feels has an accurate view of their group, and is therefore respected. In either case, reporting the peer's view accurately is likely to synchronise with the respondent's inter-rater bias.

Even if the perceived accuracy of, and liking for, the peer be opposing, the respondent's ensuing cognitive dissonance will promote truth-telling since the outcome of lying is uncertain, yet it increases dissonance (Buss & Duntley, 2008). A peer who is disliked yet holds similar views to the respondent can either be scored to endorse the respondent's views, or else to contradict them, and be unconsciously painted by the respondent as deluded. By the same token, a liked peer who disagrees with the respondent can either have their views accurately represented, or have their theoretical reputation preserved by their views being moderated. It is considered that due to confusion from not knowing how deceit can strategically benefit the respondent nor their allies, respondents will simply choose to tell the truth, and hope the eventual survey results prove to be a validation of their own perceptions and inter-rater biases.

Humans predict the behaviour of peers by comprehending their information and emotional state through mirror neurons (Oberman, Pineda, & Ramachandran, 2007) and perspective-taking (Frith & Frith, 2006). A peer's-perception instrument will ideally replicate this monitoring, to not only lower cognitive load but also reduce the effect of varying respondent insight. Since a peer's-perception instrument will not be straightforward to game, a strategy of truth telling also lowers the respondent's cognitive load (Jacob, 2008).

3.3 Uncertainty in Achieving a Brief, Universal Instrument

If questions seek perceptions usually hidden, the resulting opacity may cause respondent disengagement due to frustration and uncertainty. In the case of poor familiarity between peer and respondent, the instrument is vulnerable to opacity, especially if combined with lack of respondent insight. As such, peer's-perception questions should aim for minimum opacity, even for those lacking familiarity and social insight. Abstract questions open to varying interpretation may also lead to respondent disengagement due to frustration and uncertainty.

Based upon this, we recommend that a peer's-perception instrument should be short and universal, without questions being opaque or open to interpretation. Just as with a peer-assessment instrument that is attempting to measure against the universally-ideal collaborator, a peer's-perception instrument measuring CoVi will be measuring against universally-ideal collaboration. This, however, requires a universal model of collaboration, which as per a universal model of the collaborator, is not present in the literature (Hackman & Katz, 2010; Salas et al., 2015).

3.4 Identifying a Universal Model of Collaboration

In seeking a short yet general instrument, an alternative approach to that taken by CATME is to formulate a small number of well-defined constructs directly from a model that has already categorised a wide range of relevant psychological literature. Should a model encapsulate a

significant proportion of group research, an ensuing instrument may conceivably be universal (Pounder, 2000). In the absence of an agreed universal model, the PILAR (*prospects*, *involved*, *liked*, *agency*, *respect*) model of collaboration is selected (Heslop, Bailey, et al., 2017). Its five perceptions of collaboration, or *Pillars*, encapsulate a wide variety of social psychology and organisational psychology (Table 1).

Table 1. Definition of each Pillar

Pillar perception	Total capital stock
prospects	Your opinion of whether the group will succeed, and if so, that you will receive your anticipated share of that success.
involved	Your willingness to cooperate with colleagues, either providing or receiving assistance; in the form of knowledge and physical aid
liked	Your sense of popularity and security based upon colleagues' warmth and affection towards you
agency	The permission you feel to suggest change to the group's norms, processes, task allocation and strategy
respect	Your opinion of each colleagues' task-relevant competence, and general trustworthiness

4. PILAR-Based Peer's-Perception Instrument

Contended to be a universal model of collaboration, PILAR informs design of a peer's-perception instrument, which we term *Pillar-PP*. PILAR contends that members of a viable collaboration will positively perceive each of its five Pillars. Further, each Pillar is not only perceived by each group member, but may be aggregated to represent a dimension of CoVi. For example, most members feeling disliked lowers aggregated *liked*, reducing CoVi along that dimension. When compared to own-perception constructs, low *agency* might indicate poor psychological safety; low *respect*, poor harmony; unfavourable *liked*, low cohesion; and low *prospects*, an overly-steep hierarchy. We now discuss how a PILAR-based peer's-perception instrument might specifically overcome weakness in peer-assessment instruments.

4.1 Addressing Weaknesses through Pillar-PP

In the previous section, we postulated the advantages a general peer's-perception instrument may have. We now consider specific advantages of an instrument based upon PILAR, in terms of addressing weaknesses still unresolved from the general discussion of peer's-perception.

4.1.1 Addressing Instrument Modification

Since PILAR is conceived as a universal model of collaboration, its components should be ubiquitous in any collaboration, and so an instrument resulting from directly from the Pillars should maintain its universality. Differing contexts may place stronger emphasis on certain perceptions, for instance, *prospects* of success are more immediately important when the group is actively competing, such as during a sporting contest. *prospects* become less important when the collaboration's objective is, for instance, personal development or happiness, close monitoring of which may not be constructive (Erber & Erber, 2000).

Although certain Pillars may become less relevant in differing contexts, they still may be interpreted at the individual level when applied to peer's perception. Taking the case above of personal development, a peer's perception of *prospects* may become construed as their resilience (Luthans, Avolio, Avery, & Norman, 2007).

4.1.2 Separate Ratings of Peer Reactions

We recommend not rating each reaction separately, because doing so would effectively be a peer-assessment instrument, and hence invoke its weaknesses. While such an instrument would still rely upon the PILAR model that we conjecture is universal, and hence have merit over existing peer-assessment instruments, it would still be quite long at 20 questions. It would also miss the critical benefit of Pillar-PP in giving respondents leeway to modify each perception question to their context, such as including reactions beyond those suggested, or even personal traits (Peeters, 2006). Socially-insightful respondents may disengage if the reactions do not apply, yet they are forced to answer each. By contrast, Pillar-PP reactions are only suggestions, and there may be others more relevant to the context. For instance, a respondent encountering an anxious, unhappy peer may wish to give a low score for *prospects*, even if the peer did not exhibit the four suggested reactions. In such a case, the instrument would prefer a low *prospects* score because a person experiencing negative affect has a similar impact upon the collaboration to a ordinarily-optimistic person with legitimate concerns (Kaplan, Bradley, Luchman, & Haynes, 2009). We contend that modification of suggested reactions is allowable because each perception is ostensibly universal, therefore intuitively-obvious and not prone to being incorrectly interpreted.

4.1.3 Addressing Inter-Rater Bias

As noted previously, we contend that peer's-perception would be less influenced by inter-rater bias. In the case of the *liked* Pillar, a peer's perception of their popularity is independent of the respondent's affect *for* the peer. For instance, a loving mother can recognise that her child has few friends. However, to further reduce inter-rate bias, we propose using a negative (reverse scoring) approach for questions (Table 2), with the intent of prompting recall of negative events that positive interpersonal bias might otherwise occlude (Kim et al., 2011). There is less concern for negative bias because humans tend to have clearer insight into disliked others, since negative stimuli is more impactful and therefore analysed (Smith, Cacioppo, Larsen, & Chartrand, 2003).

In rating the observed frequency of each Pillar (rarely, sometimes, often), the BOS method is utilised. For example, a respondent perceiving that a peer *often* perceives their colleagues are *not* worthy of *respect*.

Table 2. Pillar-PP questions (reverse scored)

Pillar	This colleague (often, sometimes or rarely) …
prospects	Is unhappy with the group's overall direction
involved	Avoids asking for help from, or helping, colleagues
liked	Feels unpopular and disliked
agency	Hesitates to give their true opinion about how the group runs
respect	Believes that colleagues can't, or won't, do a good job

Adjusting for varying familiarity is more difficult. One potential solution is, within a BOS scale, adding an option allowing the respondent to indicate that they do not feel sufficiently confident to assess how a certain peer perceives a specific Pillar. Should this selection be made routinely, analysis of the score matrix will indicate if the respondent has difficulty ascertaining specific Pillars, certain peers, or both. In the latter case, the respondent may not be willing to engage with the instrument, however, absent data is better than false data.

4.1.4 Addressing Varying Interpretation and Opacity

Emotions that assist ones' navigation of social situations are relatively consistent among individuals and cultures (Schwarz & Clore, 2003). Nevertheless, human insight varies, but a rubric may help clarify less-perceptive respondents on how to recognise emotion in their peers. To construct such a rubric, we seek guidance from the PILAR model that contends each Pillar affects the other via unconscious social-psychological reaction (Heslop, Bailey, et al., 2017). A BOS instrument embeds descriptions of behaviours, which in Pillar-PP align with the peer exhibiting four reactions related to each Pillar. Since questions are reverse-coded, if a reaction is observed often, the Pillar is likely to be low, and if the reaction is observed rarely, the Pillar is high (Table 3). Using these instructions, a less-aware respondent may induce the strength of a peer's perception of a Pillar based upon considering the four reactions (Greitemeyer, 2012). Additionally, when familiarity is a factor, we postulate that objective evaluation of familiar conspecifics is promoted by the BOS instructions, while recent acquaintances escape unfair presumption of guilt (Kendall & Salas, 2015).

Table 3. Behaviours related to each (low) Pillar often and rarely occurring in the observed peer, used for Pillar-PP.

Pillar	Behaviour indicating "often"	Behaviour indicating "rarely"
prospects	Wants the group to change direction	Wants the status quo to continue
	Has misgivings about their tasks	Is content with their assigned tasks
	Often blames people for problems	Rarely blames people for problems
	Feels colleagues are making mistakes	Thinks colleagues are performing well
involved	Confused or uncertain about colleagues' expertise	Certain of what colleagues are capable of
	Complains of rarely being asked to help or help provided	Confident that they will be asked to help, or help provided when required
	Doesn't check with others after having an idea	Always gets others' feedback on their ideas
	Unaware of the group's big picture strategy	Aware of the strategy and how their role fits in
liked	Resents being asked to help	Happy to help when asked
	Thinks others don't respect their abilities	Thinks their abilities are well-regarded
	Thinks other's will dismiss their ideas	Thinks others will listen to their ideas
	Doesn't care if the group fails	Cares a lot if the group succeeds
agency	Avoids causing disagreement	Willing to offer a different perspective
	Thinks colleagues won't change their minds	Confident that colleagues might change their minds

	Doesn't easily think laterally	Often has creative solutions
	Gives up on their ideas quickly	Persists in championing their ideas
respect	Avoids cooperating with colleagues due to a perception of their incompetence	Happy to cooperate with colleagues because they are competent
	Cross-checks and monitors colleague's work	Doesn't need to micromanage anyone
	Thinks the group wouldn't survive without them	Believes they are ultimately replaceable
	Doesn't care if they hurt colleague's feelings	At pains to avoid insulting people

Critically, Pillar reactions are separate social-psychological effects (Heslop, Bailey, et al., 2017), which avoids respondent confusion that may otherwise occur if multiple reactions were similar. Pillar reactions are also not exhaustive, yet including further reactions would unduly increase cognitive load while potentially losing universality. Only reactions between the five Pillars are considered to retain the universality of the entire PILAR model.

5. Pilot Testing Pillar-PP

Whether a PILAR-based peer's-perception instrument is a reliable and valid measure of CoVi remains to be tested. It is however recognised that the concept of peer's-perception is conjectural, and so is the PILAR model. Combining two sets of conjectures limits the claims that this article can make. The first step in obtaining evidence will be pilot testing Pillar-PP to ascertain whether respondents are comfortable assessing their peer's perceptions, preferably in an established group to reduce confounding from poor familiarity.

A challenge will be varying interpretation due to personality or cultural background (Mount, Barrick, & Stewart, 1998). To reduce this likelihood, we recommend that prior to application of the survey, respondents be instructed to avoid value judgements. For instance, those who culturally adhere to hierarchy are asked to be accurate rather than refuse to confer a junior colleague with *agency*, based upon perceived inappropriateness of their making suggestions to superiors (Sosik & Jung, 2002). Regarding personality, those with low self-efficacy may be less able to recognise competence in others for example, and so be less capable of recognising the *respect* perception in peers (Bandura, 1994).

Application of the Pillar-PP instrument is designed to give organisations early warning of threats to collaboration viability, allowing remedial actions to be taken. If a pilot test were successful, we would then recommend utilising Pillar-PP for ongoing monitoring within an organisation, for instance, applying the instrument every month. If Pillar-PP provides warning of poor organisation outcomes, such as employee absenteeism, retention and profitability, it will be useful for decision-makers. Unless regularly surveyed, a change in CoVi may not become apparent for months or years, for instance, until a report reveals loss of production, or employees formally complain about bullying.

6. Conclusion

The proposed Pillar-PP instrument is a short, five-item BOS instrument, intended to be feasible for regular deployment in workgroups. It uses a peer's-perception approach to counter pro-self and inter-rater biases that are otherwise detrimental in own-perception and

peer's-perception instruments. Pillar-PP is also designed to reduce respondent disengagement from opacity, cognitive dissonance and cognitive load sometimes found in peer-assessment instruments.

Early identification of poor CoVi maximises potential for rectification, thereby reducing damage to group productivity and member's mental health (Morgeson & DeRue, 2006). Postulated universality of PILAR (Heslop et al., 2016) extends to the instrument, conceptually allowing maintenance of reliability across varying contexts. Developing a valid and reliable measure of CoVi would be useful for detecting workplaces that suffer from poor collaboration, and our proposed Pillar-PP instrument is a contender. In utilising a universal model of collaboration, Pillar-PP adopts an approach that is the reverse of CATME's derivation.

References

Bandura. (1994). Self-efficacy. *Encyclopedia of Human Behavior*, *4*(1994), 1–65.

Bargh, J. A., Gollwitzer, P. M., & Oettingen, G. (2010). Motivation. In S. Fiske, D. T. Gilbert, & G. Lindzay (Eds.), *Handbook of Social Psychology* (5th ed., pp. 268–316). New York: Wiley. https://doi.org/10.1002/9780470561119.socpsy001008

Brody, C. J., & Dietz, J. (1997). On the Dimensionality of Two-Question Format Likert Attitude Scales. *Social Science Research*, *26*(26), 197–204. https://doi.org/10.1006/ssre.1997.0594

Buss, D. M., & Duntley, J. D. (2008). Adaptations for Exploitation. *Group Dynamics: Theory, Research, and Practice*, *12*(1), 53–62. https://doi.org/10.1037/1089-2699.12.1.53

Ceston, C. M., Levine, R. E., & Lane, D. R. (2008). Peer Assessment and Evaluation in Team-Based Learning. In *New Directions for Teaching and Learning* (pp. 69–78). Wiley Periodicals, Inc. https://doi.org/10.1002/tl.334

Chen, C. C., Ünal, A. F., Leung, K., & Xin, K. R. (2016). Group harmony in the workplace: Conception, measurement, and validation. *Asia Pacific Journal of Management*, *33*(4), 935–936. https://doi.org/10.1007/s10490-016-9461-4

Chiu, C., & Dweck, C. S. (1997). Lay Dispositionism and Implicit Theories of Personality. *Journal of Personality and Social Psychology*, *73*(I), 19–30. https://doi.org/10.1037/0022-3514.73.1.19

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891

Claussen, D. S. (2004). Cognitive Dissonance, Media Illiteracy, and Public Opinion on News Media. *American Behavioral Scientist*, *48*(2), 212–218. https://doi.org/10.1177/0002764204267265

Daley, D. M. (1992). Pay for Performance, Performance Appraisal, and Total Quality

Management. *Public Productivity & Management Review*, *16*(1), 39–51. https://doi.org/10.2307/3380804

Debnath, S. C. (2015). Fifty Years and Going Strong: What Makes Behaviorally Anchored Rating Scales So Perennial as an Appraisal Method? *International Journal of Business and Social Science*, *6*(2), 16–25.

DeCoths, T. A. (1977). An analysis of the external validity and applied relevance of three rating formats. *Organizational Behavior and Human Performance*, *19*(2), 247–266. https://doi.org/10.1016/0030-5073(77)90064-2

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-reported bias in organizational behavior research. *Journal Business and Psychology*, *17*(2), 245–261. https://doi.org/10.1023/A:1019637632584

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Elfenbein, H. A. (2008). Emotion in Organizations. In *The Academy of Management Annals* (pp. 315–386). University of California, Berkeley. https://doi.org/10.1080/078559812

Erber, R., & Erber, M. W. (2000). The Self-Regulation of Moods: Second Thoughts on the Importance of Happiness in Everyday Life. *Psychological Inquiry*, *11*(3), 142–148. https://doi.org/10.1207/S15327965PLI1103_02

Fowler, F. (1995). *Improving survey questions: Design and evaluation* (Applied Social Research Methods, Volume 38). Thousand Oaks, California: Sage.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–4. https://doi.org/10.1016/j.neuron.2006.05.001

Greitemeyer, T. (2012). Boosting One's Social Identity: Effects of Social Exclusion on Ethnocentrism. *Basic and Applied Social Psychology*, *34*(5), 410–416. https://doi.org/10.1080/01973533.2012.712013

Hackman, J., & Katz, N. (2010). Group behavior and performance. In *Handbook of social psychology* (pp. 1208–1251). Retrieved from https://doi.org/10.1002/9780470561119.socpsy002032

Hastie, C., Fahy, K., & Parratt, J. (2014). The development of a rubric for peer assessment of individual teamwork skills in undergraduate midwifery students. *Women and Birth*, *27*(3), 220–226. https://doi.org/10.1016/j.wombi.2014.06.003

Heslop, B., Bailey, K., Drew, A., Paul, J., Stojanovski, E., & Smith, R. (2017). Encapsulating Social Psychology: The PILAR Model of Collaboration. *Manuscript Submitted for Publication*.

Heslop, B., Bailey, K., Paul, J., Drew, A., & Smith, R. (2016). Collaboration Guidelines to Transform Culture. *Interdisciplinary Journal of Partnership Studies*, *3*(3).

Heslop, B., Stojanovski, E., Paul, J., Iverson, S., & Bailey, K. (2017). Respondent Disengagement in a Peer-Assessment Instrument Measuring Collaboration Viability. *Manuscript Submitted for Publication*.

Hess, S., & Stathopoulos, A. (2013). Linking response quality to survey engagement: a combined random scale and latent variable approach. *Journal of Choice Modelling*, *7*, 1–12. https://doi.org/10.1016/j.jocm.2013.03.005

Hoenderdos, J. W. (2013). *Towards an Observational Measure for Team Psychological Safety*. University of Twente. Retrieved from http://essay.utwente.nl/63309/

Hughes, R. L., & Jones, S. K. (2011). Developing and Assessing College Student Teamwork Skills. *New Directions for Institutional Research*, *149*, 53–64. https://doi.org/10.1002/ir.380

Jacob, P. (2008). What Do Mirror Neurons Contribute to Human Social Cognition? *Mind & Language*, *23*(2), 190–223. https://doi.org/10.1111/j.1468-0017.2007.00337.x

Jehn, K., & Mannix, E. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, *44*(2), 238–251. https://doi.org/10.2307/3069453

Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D. (2009). On the role of positive and negative affectivity in job performance: a meta-analytic investigation. *Journal of Applied Psychology*, *94*(1), 162–176. https://doi.org/10.1037/a0013115

Kendall, D. L., & Salas, E. (2015). Measuring Team Performance: Review of current methods and consideration of future needs. In *The Science and Simulation of Human Performance* (pp. 307–326). Online: Emerald Group Publishing Ltd. http://doi.org/10.1016/S1479-3601(04)05006-4

Kim, J., Baek, T. H., & Kim, D. (2011). Quality of Work and Team Spirit as Drivers of Student Peer Evaluation on Advertising Group Project Performance. *Journal of Advertising Education*, *15*(2), 14–24. Retrieved from http://eds.b.ebscohost.com.laureatech.idm.oclc.org/eds/pdfviewer/pdfviewer?sid=4891017d-0039-42ba-abba-ad9199198e2d@sessionmgr115&vid=2&hid=117

Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing Behaviorally Anchored Rating Scales (BARS) and other rating formats. *Personnel Psychology*, *34*(2), 263–289. https://doi.org/10.1111/j.1744-6570.1981.tb00942.x

Koh, E., Hong, H., & Seah, J. (2014). An analytic frame and multi-method approach to measure teamwork competency. *Proceedings - IEEE 14th International Conference on Advanced Learning Technologies, ICALT 2014*, 264–266. https://doi.org/10.1109/ICALT.2014.82

Koh, E., Shibani, A., Tan, J. P. L., & Hong, H. (2016). A pedagogical framework for learning analytics in collaborative inquiry tasks. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 74–83. https://doi.org/10.1145/2883851.2883914

Konak, A., Magluilo, S., & Kulturel-Konak, S. (2016). Behaviorally anchored rating scales for teamwork peer assessment. *ISEC 2016 - Proceedings of the 6th IEEE Integrated STEM Education Conference*, 168–172. https://doi.org/10.1109/ISECon.2016.7457525

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46–53. https://doi.org/10.3102/0013189X15569530

Kusano, S. M., Conger, A. J., & Wright, M. C. (2015). *Development and Assessment of Collaboration, Teamwork and Communication*. Michigan. Retrieved from http://crlt.umich.edu/engaged-learning/sites/default/files/white-papers/Collaboration_HI RES.pdf

Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *The Journal of Applied Psychology*, *86*(2), 255–264. https://doi.org/10.1037/0021-9010.86.2.255

Lönnqvist, J. E., Leikas, S., Verkasalo, M., & Paunonen, S. V. (2008). Does Self-Enhancement Have Implications for Adjustment? *Basic and Applied Social Psychology*, *30*(4), 377–386. https://doi.org/10.1080/01973530802502374

Lord, R. G., & Maher, K. J. (1990). Alternative Information-Processing Models and Their Implications for Theory, Research, and Practice. *The Academy of Management Review*, *15*(1), 9–28.

Lott, A., & Lott, B. (1965). Group cohesiveness as interpersonal attraction: a review of relationships with antecedent and consequent variables. *Psychological Bulletin*, *64*(4), 259–309. https://doi.org/10.1037/h0022386

Luthans, F., Avolio, B. J., Avery, J. B., & Norman, S. M. (2007). Positive psychological capital: Measurement and relationship with performance and satisfaction. *Personnel Psychology*, *60*, 541–572. https://doi.org/10.1111/j.1744-6570.2007.00083.x

Mathieu, J. E., Gilson, L. L., & Ruddy, T. M. (2006). Empowerment and team effectiveness: an empirical test of an integrated model. *The Journal of Applied Psychology*, *91*(1), 97–108. https://doi.org/10.1037/0021-9010.91.1.97

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Morgeson, F., & DeRue, D. (2006). Event criticality, urgency, and duration: Understanding how events disrupt teams and influence team leader intervention. *The Leadership Quarterly*. https://doi.org/10.1016/j.leaqua.2006.02.006

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-Factor Model of personality and Performance in Jobs Involving Interpersonal Interactions. *Human Performance*, *11*(2–3), 145–165. https://doi.org/10.1080/08959285.1998.9668029

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of

the literature. *Human Resource Management Review*, 1–15. https://doi.org/10.1016/j.hrmr.2017.01.001

O'Neill, T. A., McLarnon, M. J. W., & Carswell, J. J. (2015). Variance Components of Job Performance Ratings. *Human Performance*, *28*(1), 66–91. https://doi.org/10.1080/08959285.2014.974756

Oberman, L. M., Pineda, J. a, & Ramachandran, V. S. (2007). The human mirror neuron system: a link between action observation and social skills. *Social Cognitive and Affective Neuroscience*, *2*(1), 62–66. https://doi.org/10.1093/scan/nsl022

Ohland, M. W., & Loughry, M. L. (2006). Designing a Peer Evaluation Instrument that is Simple, Reliable, and Valid. In *National STEM Assessment Conference* (pp. 202–208).

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., … Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning and Education*, *11*(4), 609–630. https://doi.org/10.5465/amle.2010.0177

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

Peeters, M. A. G. (2006). The Big Five Personality Traits and Individual Satisfaction With the Team. *Small Group Research*, *37*(2), 187–211. https://doi.org/10.1177/1046496405285458

Pounder, J. S. (2000). Evaluating the Relevance of Quality to Institutional Performance Assessment in Higher Education. *Evaluation*, *6*(1), 66–78. https://doi.org/10.1177/13563890022209127

Propper, C., & Wilson, D. (2003). The Use and Usefulness of Performance Measures in the Public Sector. *CMPO Working Paper*, *May*(73). https://doi.org/10.1093/oxrep/19.2.250

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide. *Human Resource Management*, *54*(4), 599–622. https://doi.org/10.1002/hrm.21628

Schwarz, N., & Clore, G. (2003). Mood as Information: 20 Years Later. *Psychological Inquiry*, *14*(3), 296–303. https://doi.org/10.1207/S15327965PLI1403&4_20

Smith, N. K., Cacioppo, J. T., Larsen, J. T., & Chartrand, T. L. (2003). May I have your attention, please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia*, *41*(2), 171–183. https://doi.org/10.1016/S0028-3932(02)00147-1

Sosik, J. J., & Jung, D. I. (2002). Work-group characteristics and performance in collectivistic and individualistic cultures. *The Journal of Social Psychology*, *142*(1), 5–23. https://doi.org/10.1080/00224540209603881

The iTOFT Consortium. (2015). *Work-based assessment of teamwork: an interprofessional approach*. Canberra, Australia. Retrieved from http://derby.openrepository.com/derby/bitstream/10545/621207/1/ID12-2193_UQ_Thistlethwaite_Final+report_2015.pdf

Thomas, E. J., Taggart, B., Crandell, S., Lasky, R. E., Williams, A. L., Love, L. J., … Helmreich, R. L. (2007). Teaching teamwork during the Neonatal Resuscitation Program: a randomized trial. *Journal of Perinatology : Official Journal of the California Perinatal Association*, *27*(7), 409–414. https://doi.org/10.1038/sj.jp.7211771

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. https://doi.org/10.1016/j.learninstruc.2009.08.003

Tziner, A., & Kopelman, R. E. (2002). Is there a Preferred Performance Rating Format? A Non-psychometric Perspective. *Applied Psychology: An International Review*, *51*(3), 479–503. https://doi.org/10.1111/1464-0597.00104

van Zundert, M., Sluijsmans, D., & van Merrienboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is There a General Factor in Ratings of Job Performance? A Meta-Analytic Framework for Disentangling Substantive and Error Influences. *Journal of Applied Psychology*, *90*(1), 108–131. https://doi.org/10.1037/0021-9010.90.1.108

Wiese, D. S., & Buckley, M. R. (1998). The evolution of the performance appraisal process. *Journal of Management History*, *4*(3), 233–249. https://doi.org/10.1108/13552529810231003

Wildman, J. L., Shuffler, M. L., Lazzara, E. H., Fiore, S. M., Burke, C. S., Salas, E., & Garven, S. (2012). Trust Development in Swift Starting Action Teams: A Multilevel Framework. *Group & Organization Management*, *37*(2), 137–170. https://doi.org/10.1177/1059601111434202

Willey, K., & Gardner, A. (2009). Developing team skills with self- and peer assessment: Are benefits inversely related to team function? *Campus-Wide Information Systems*, *26*(5), 365–378. https://doi.org/10.1108/10650740911004796

Zingoni, M., & Corey, C. M. (2016). How Mindset Matters. *Journal of Personnel Psychology*, *Online pub*, 1–10. http://doi.org/10.1027/1866-5888/a000171

Copyright Disclaimer