Accepted version of Sridharan, B. and Boud, D. (accepted for publication 2 November 2018). The effects of peer judgements on teamwork and self-assessment ability in collaborative group work, *Assessment and Evaluation in Higher Education*, DOI: 10.1080/02602938.2018.1545898

The effects of peer judgements on teamwork and self-assessment ability in collaborative group work

Bhavani Sridharan

Deakin Business School, Deakin University, Geelong, Australia, bhavani.sridharan@deakin.edu.au, ORCID - 0000-0002-2217-949X

David Boud

Centre for Research in Assessment and Digital Learning, Deakin University, Geelong, Australia; Middlesex University, London, UK; Professor Emeritus, Faculty of Arts and Social Sciences, University of Technology Sydney, Australia; david.boud@deakin.edu.au, ORCID - 0000-0002-6883-2722

Abstract

A crucial determinant of the success or failure of collaborative group work is the effect of peer feedback interventions on learning. Research exploring such effects on developing soft skills is sparse. This study seeks to address whether peer feedback leads to enhanced teamwork behaviour and self-assessment ability, two skills highly sought after by employers. Specifically, this study examines the direct effect of formative performance rating and the mediating effect of praise and criticism in peer feedback messages on achievement in teamwork and self-assessment skills. The sample consists of quantitative and qualitative data from 98 students enrolled in business programmes using a particular form of collaborative group work. The paper finds a direct positive relationship between formative performance rating and summative self-assessment ability. It also finds that praise negatively mediates the relationship between formative performance rating and summative teamwork. Further analyses suggest that a significant proportion of comments provided is past- rather than future-oriented. Potential strategies to overcome the limitations of current practices are discussed.

Keywords: collaborative group work, teamwork behaviour, self-assessment ability, peer feedback literacy

Introduction

Collaborative group work, where small groups of people work together to achieve a shared goal, has long been recognised as a powerful approach to foster students' active/deeper engagement in developing cognitive and affective skills (Davies 2009). Collaborative group work, specifically self-managed work teams, have become a dominant practice in organisations to meet the dynamic demands of complex, agile and innovative projects (Magpili and Pazos 2017). Efficient completion and successful outcomes of these projects require two mutually interdependent skills: teamwork and self-management (Salas et al. 2015).

The importance of these skills is evident from the demand for explicit demonstration of development of these skills by universities. For instance, the Association to Advance Collegiate Schools of Business eligibility standards requires provision of evidence for development of teamwork and reflective thinking skills (AACSB 2017). However, studies have identified a skills gap in fulfilling employers' demands and called for business curriculum reforms to develop these skills among new graduates (Adrian 2017).

Undoubtedly collaborative group work provides an essential facilitating environment, but does not automatically lead to development of these skills, with problems including social-loafing and free riding (Davies 2009). This is particularly seen in learning team compared to work team environments, as group work in learning environments involves short-term and often transient collaboration. Additionally, transferability of skills is limited in learning teams, being influenced by the assessment task and composition of team members.

Self and peer judgement tasks are effective in addressing these problems, nonetheless students can fail to develop these skills due to lack of awareness and opportunity (Sridharan, Muttakin, and Mihret 2018). Previous research findings in this regard suggest students are reluctant to honestly assess their peers, especially when the mark is counted toward their final grade (Sridharan, Tai, and Boud 2018). Building upon these findings, this study investigates the effect of peer judgement on learning. Here, it is assumed that the influence of peer

judgment is through the feedback process on the premise that students can correct their behaviour only if they know something is wrong with it.

Against this backdrop, the power of feedback in enhancing learning is foregrounded in the literature (Hattie and Timperley 2007; Black and Wiliam 2010). Yet, scholars have expressed concerns about the effectiveness of various aspects of feedback in improving learning. Studies suggest that the feedback effect may vary depending on aspects including: the valence (praise or criticism); the focus (task, process, self and self-regulation); the orientation (past or future); the medium (written or verbal); learners' engagement; and learners' capability to recall and reuse information (Liu and Carless 2006; Hattie and Timperley 2007; Boud and Molloy 2013; Winstone et al. 2017; Nash et al. 2018).

There are two major gaps in the research on the influence of feedback on learning that this study seeks to address. First, empirical research on the effects of feedback on developing soft skills is limited or dated. For example, Gabelica et al. (2012) found only 8% of empirical articles (five out of 59) with specific focus on process feedback (against knowledge feedback) to teams with inconclusive results covering both organisational (80%) and higher education contexts (20%).

The second gap relates to underdeveloped empirical research examining indirect effects of peer judgements delivered via digital technology on students' soft skills development in collaborative group work setting, controlling for performance type and delivery mode. Kluger and DeNisi (1996) examined the effects of source (peer vs. superior) and sign (positive vs. negative) of feedback, however, self–assessment and indirect effects were not considered. Similarly, effects of self and peer rating and feedback on teamwork behaviour were explored using exposure and repeated exposure conditions (Dominick, Reilly, and McGourty 1997; Donia, O'Neill, and Brutus 2018), while feedback signs were not considered.

This study seeks to address these gaps by exploring whether or not self and peer judgement efforts positively influence learning using a particular case of collaborative group work. A conceptual model of feedback intervention is proposed and empirically tested by evaluating the direct and mediating effects on outcomes, controlling for types of performers (high, middle and low achievers), and mode of study (off-campus/on-campus). Accordingly, the research questions are:

Does students' performance rating by self and peer positively influence the outcome variables?

Does the quantum of praise or criticism mediate the relationship between the performance rating and the outcome variables?

Literature review

In recent years, feedback has been championed as a key factor in supporting improvement and progress in student learning. Feedback is defined as "a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies" (Carless and Boud 2018, 1315). Within a formative assessment framework, feedback agents include teachers, self, peers, coaches and books (Hattie and Timperley 2007). Amongst these, the most common in the literature are self, teachers, and peers. However, in the collaborative group work context self and peers are the pertinent sources, as teachers are normally absent from most interactions and students are well-positioned to comment on each other's interpersonal performance (Ohland et al. 2012). Self-agency involves students taking responsibility for their own learning and using learning opportunities for long-term growth in affective and cognitive skills (Barber et al. 2013). In the present study, this requires students judging their own work and comparing it with that of peers to generate feedback through self-reflection, through which students can become self-regulated learners (Nicol and Macfarlane-Dick 2006). Peer feedback is defined as "a communication process through which learners

enter into dialogues related to performance and standards" (Liu and Carless 2006, 2).

Feedback processes have the potential to: facilitate students to better self-assess; develop and showcase a range of soft skills; enhance student engagement and autonomy; facilitate students taking responsibility for their learning (Liu and Carless 2006); and develop self-assessment aspects of self-regulated learning (Nicol and Macfarlane-Dick 2006). Benefits of self and peer feedback interventions in comparison to teacher feedback include increased depth and breadth of information, and exposure to multiple perspectives (Donia, O'Neill, and Brutus 2018); and greater volume and immediacy (Topping 2009). However, it is not known whether current self and peer judgement practices are effective in instilling teamwork and self-assessment skills, particularly in collaborative group work contexts.

Conceptual model and hypothesis development

In this section, a conceptual model is proposed followed by the theoretical considerations supporting the relationships in the model. The model represents three key components: performance rating (self and peer rating), qualitative comments (praise and criticism) and outcome variables (teamwork behaviour and self-assessment ability) (Figure 1). The first two components are from formative assessment and the third component is from summative assessment. Recognising the positive influence of formative assessment on learning, the performance rating in our model is an input (antecedent) variable that influences both qualitative comments and outcome variables. The qualitative comments are mediating variables as they explain the indirect influence of performance rating on outcome variables. Post facto analysis of comments revealed the strong presence of praise and criticism from peers, accordingly the quantum of praise and criticism is considered. Study mode and type of performance are used as moderators to evaluate whether these aspects distort the effects on outcome variables. The following sections define the variables, discuss the literature around the relationships between them and identify hypotheses to be explored.

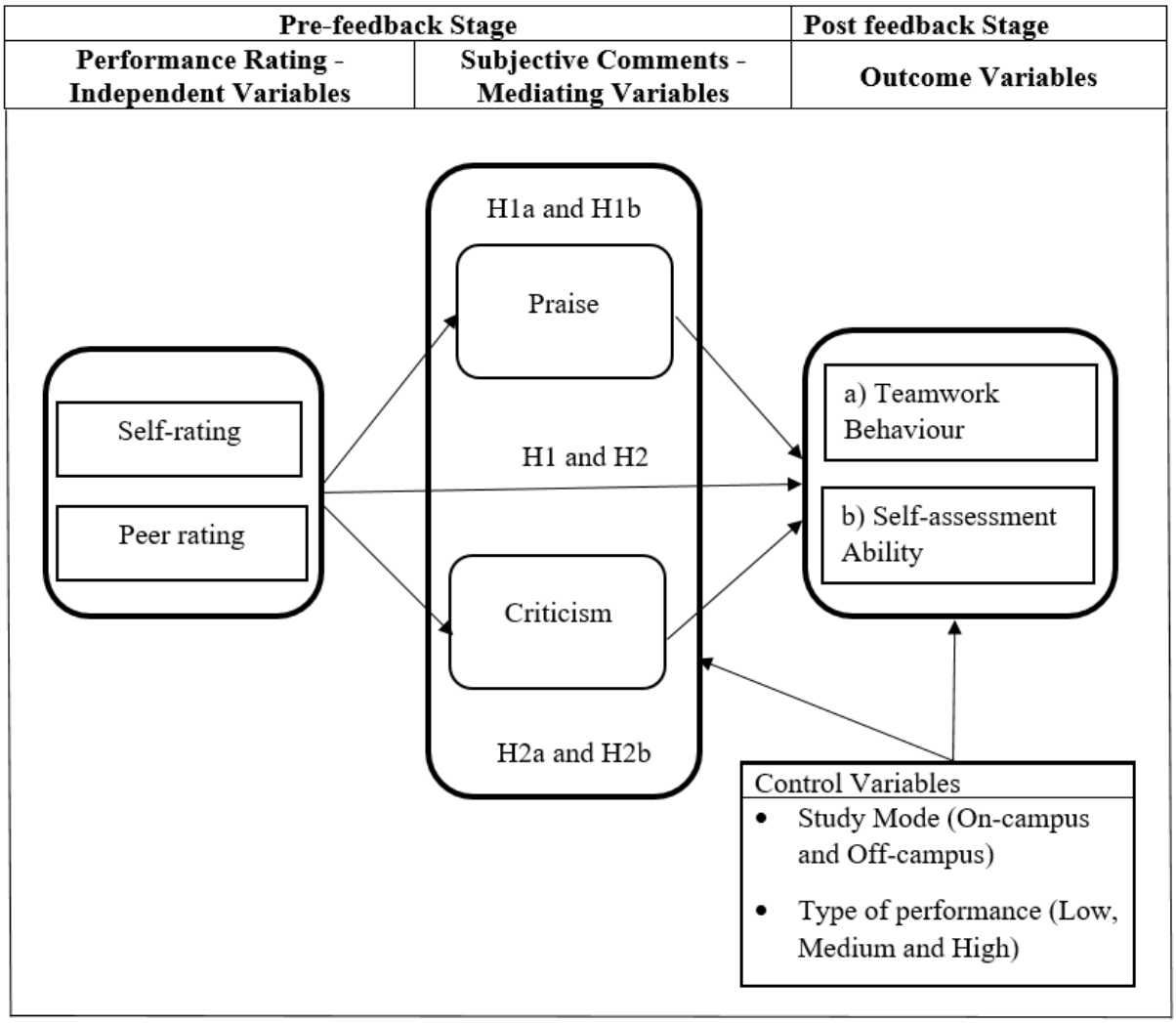

Figure 1. A conceptual model of feedback intervention

Performance rating and teamwork behaviour

First, the direct influence of *performance rating* on *teamwork behaviour* is considered. The *performance rating* is a quantitative evaluation of the extent to which an individual team member contributed towards the group project goals during formative regime. *Teamwork* refers to a small group working interdependently to accomplish a shared common goal (Varela and Mead 2018). Accomplishing the common goal requires shared behaviours, attitudes and cognizance (Salas et al. 2015). This entails a wide-range of skillsets such as interpersonal communication, decision-making, conflict resolution, empathy, mutual support and feedback

intervention.

Studies directly exploring the influence of *performance rating* on performance outcomes are limited. Improvements to interpersonal skills are found through exposure to assessment criteria and repeated exposure to self and peer ratings (Dominick, Reilly, and McGourty 1997; Donia, O'Neill, and Brutus 2018). Conversely, Kluger and DeNisi (1996) highlight the ineffectiveness of feedback intervention cues such as grades to influence performance. Nevertheless, scholars argue that self and peer rating *per se* is ineffective as a feedback input, as a grade contains little information, does not indicate what needs to be improved and lacks a future orientation (Hattie and Timperley 2007). This leads to the first hypothesis.

H1: Performance rating influences teamwork behaviour.

Performance rating and self-assessment ability

Second, the direct influence of *performance rating* on s*elf-assessment ability* is examined. *Self-assessment* requires students' judging their own contribution against explicitly selected criteria towards supporting long-term personal and professional development (Barber et al. 2013). In the collaborative group work context, self-assessment requires thinking about thinking, at both individual (through self-monitoring and self-regulation) and social levels (through shared cognition and co-regulation) (Cho and Kim 2013). The core to the development of self-assessment capability is evaluative judgement – the ability to provide a context-specific judgement of self (and peer) performance based on a predefined standard (Tai et al. 2017). The development of evaluative judgement is a pedagogic manifestation of self-assessment ability.

Studies explicitly exploring the effect of performance rating on the development of self-assessment ability, particularly in collaborative group work context, are unknown. Studies suggest that self-assessed grades are flawed as a measure of performance with significant gaps

between perception and reality (Dunning, Heath, and Suls 2004; Johnson and Molloy 2018). Two reasons attributed to such errors are incomplete knowledge of self-competency and overlooking relevant information (Dunning, Heath, and Suls 2004). Sitzmann et al. (2010) suggest that learning can be enhanced through opportunities to self-assess and receive feedback on both the accuracy of self-assessment and learner's performance.

The lack of research on the impact of performance rating on self-assessment underscores the need for further research. Accordingly, the second hypotheses proposed is:

H2: Performance rating influences self-assessment ability.

Performance rating and outcome variables via qualitative comments (mediating variable)

Third, the indirect influence of *performance rating* on *outcome variables* through the mediator variable (*qualitative comments)* is investigated. *Qualitative comments* refer to narrative peer remarks embedding elements including (a) valence (praise or criticism); (b) orientation (past or future); (c) level (individual); and (d) agent (self or peer). Valence refers to the overall tone of comments. Praise includes peer remarks that acknowledge students' past behaviour (eg. 'delivers work on time'), which may contain elements of future-oriented directive input (eg. 'needs to be a better listener'). Criticism refers to peer remarks that highlight areas of poor past work/behaviour (eg. 'did not complete work on time'), which may contain future-oriented directive input (eg. 'needs to engage more in meetings').

Theoretically, it has been suggested that qualitative comments accompanied by a grade can have a superior effect on learning outcomes (Hattie and Timperley 2007), yet how this intervenes in the relationship has not been explored. Methodologically, the choice of qualitative comments as a mediator variable satisfies the requirement of the causal order

including temporal precedence[a] (i.e. formative rating occurred before formative qualitative comments justifying their rating, which happened before summative rating). In this study, the relationship depicted by conceptual model unfolds in that sequence.

Evaluating the role of the mediation variable requires decomposing the relationship into two paths – from *performance rating* to *qualitative comments*; and from *qualitative comments* to *outcome variables* (see Figure 1). Studies explicating the first path (*performance rating* to *qualitative comments)* are sparse. Hattie and Timperley (2007) noted the tendency of teachers to give more praise to poor-performing students to enhance motivation. Cushing et al. (2011) found student anxiety in giving negative feedback to poor performing students for fear of being unkind.

Exploring the second path (*qualitative comments* to *outcome variables*), mixed findings exist. Hattie and Timperley (2007) found a low effect size for praise on achievement, however, they noted its effect depended on multiple factors. They revealed that feedback is less effective if the focus is on the person, and more effective if the focus is on the task or process. Medvedeff et al. (2008) emphasised the value of negative process feedback to improve performance and adjust behaviour. Similarly, Kluger and DeNisi (1996) found negative peer feedback to be more influential on teamwork performance. Van Dijk and Kluger (2011), among others, found mixed results indicating that positive or negative messages can increase or decrease performance depending on the task. Kohn (1999) posits the damaging effects particularly of praise on learning.

Overall, these results point to lack of clarity and empirical evidence around how the quantum of praise and criticism from peers indirectly influences the development of teamwork and self-assessment skills. Thus, the following hypotheses are proposed.

---

[a] In this study the relationship depicted by conceptual model unfolds in that sequence

H1a and H1b: Praise mediates the relationship between performance rating and outcome variables.

H2a and H2b: Criticism mediates the relationship between performance rating and outcome variables.

Methodology

Participants and material

The naturalistic and retrospective nature of the study design necessitated analysis of students' rating and feedback data collected from 98 students enrolled in two undergraduate and one postgraduate unit at an Australian university over two trimesters (2014 and 2015). 72% of students were undergraduate and 28% postgraduate. Most students (86%) were on-campus, while 14% were off-campus; whether or not these students had face-to-face meetings with team members was not known.

Group sizes ranged from three to five and group formation was through random allocation of students. Each group was assigned a unique real-life project and completed the tasks collaboratively over an 11-week period. The course content was the same during both offerings with minor differences in assessment tasks for undergraduate and postgraduate students.

Students completed three assessment tasks as part of the collaborative group project. These included (a) delivery of a written product as a group (80% weight) with four key deliverables at different stages, (b) individual submission of a reflective essay for undergraduate students, or a project presentation for postgraduate students (20%), and (c) mandatory anonymous completion of self and peer rating along with narrative comments in two iterations (no explicit weight) using the online SPARK^{PLUS} tool. Learning resources for developing teamwork skills were provided. Students had prior exposure to collaborative group

work in earlier units, though they were not exposed to self and peer assessment of teamwork or giving and receiving peer comments.

This study's focus is on the third task in which students complete two iterations of self and peer assessment with peer comments. This involves students rating themselves and their peers, and providing qualitative comments to each team member. The first iteration is formative with the aim of helping students improve their teamwork behaviour and self-assessment ability by taking cues from the self and peer ratings and comments. Students completed this cycle in week 4 after submitting the first two deliverables. During this cycle, students were given specific criteria to complete the self and peer assessment process. The results and peer comments were published to students soon after the submission deadline. The second iteration is summative as the results from this iteration were used to adjust the collaborative group work product mark, that is, the mark that was weighted as part of their overall grade for the unit. Here, the assessment is holistic whereby students give an overall rating and provide feedback. Students complete this cycle upon delivery of the final collaborative group work product. Again, results and qualitative comments were published after the deadline.

The study involved retrospective analysis of 523 instances of self and peer rating, and students' peer comments during both formative and summative assessment regimes. The two outcome variables are derived from the relative performance factor (RPF) (measuring teamwork behaviour) and the self and peer assessment (SAPA) factor (measuring self-assessment ability) from the summative regime. RPF is a proxy for teamwork behaviour that measures the level of individual student's final contributions in comparison to the entire team's contribution after taking corrective actions based on prior input. SAPA is a proxy for self-assessment ability that measures the level of self-perceived contribution in comparison to the entire team's perceived individual contributions. The calculation details of RPF and SAPA are given in Figure 2.

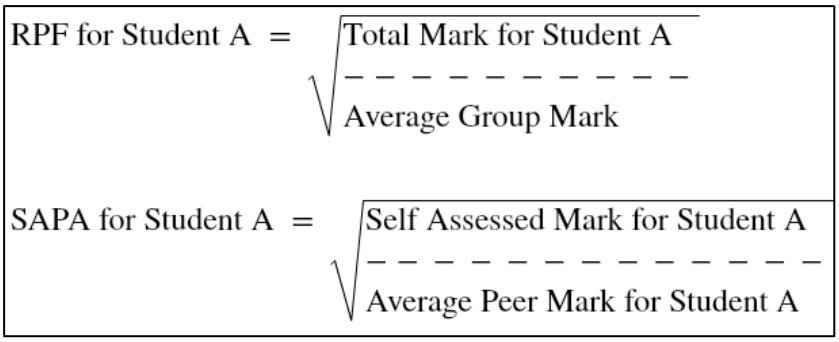

Figure 2. Example RPF and SAPA Calculations (Willey and Gardner 2009)

The performance rating is derived from self and peer rating of students for multiple criteria using a behaviourally anchored rating scale. The criteria against which students evaluated self and peers falls into three categories: ideas, tasks and collaboration. The rating scales include: 1=Never-NV; 2=Rarely-RY; 3=Sometimes-ST; 4=Often-OF; and 5=Almost always-AA (Figure 3). As part of the self and peer assessment process, students receive multiple types of input: (a) detailed criteria level and overall performance rating (both self and peer side by side); (b) RPF and SAPA factor scores; and qualitative comments. Figure 3 shows how students view these using the online SPARK^{PLUS} tool.

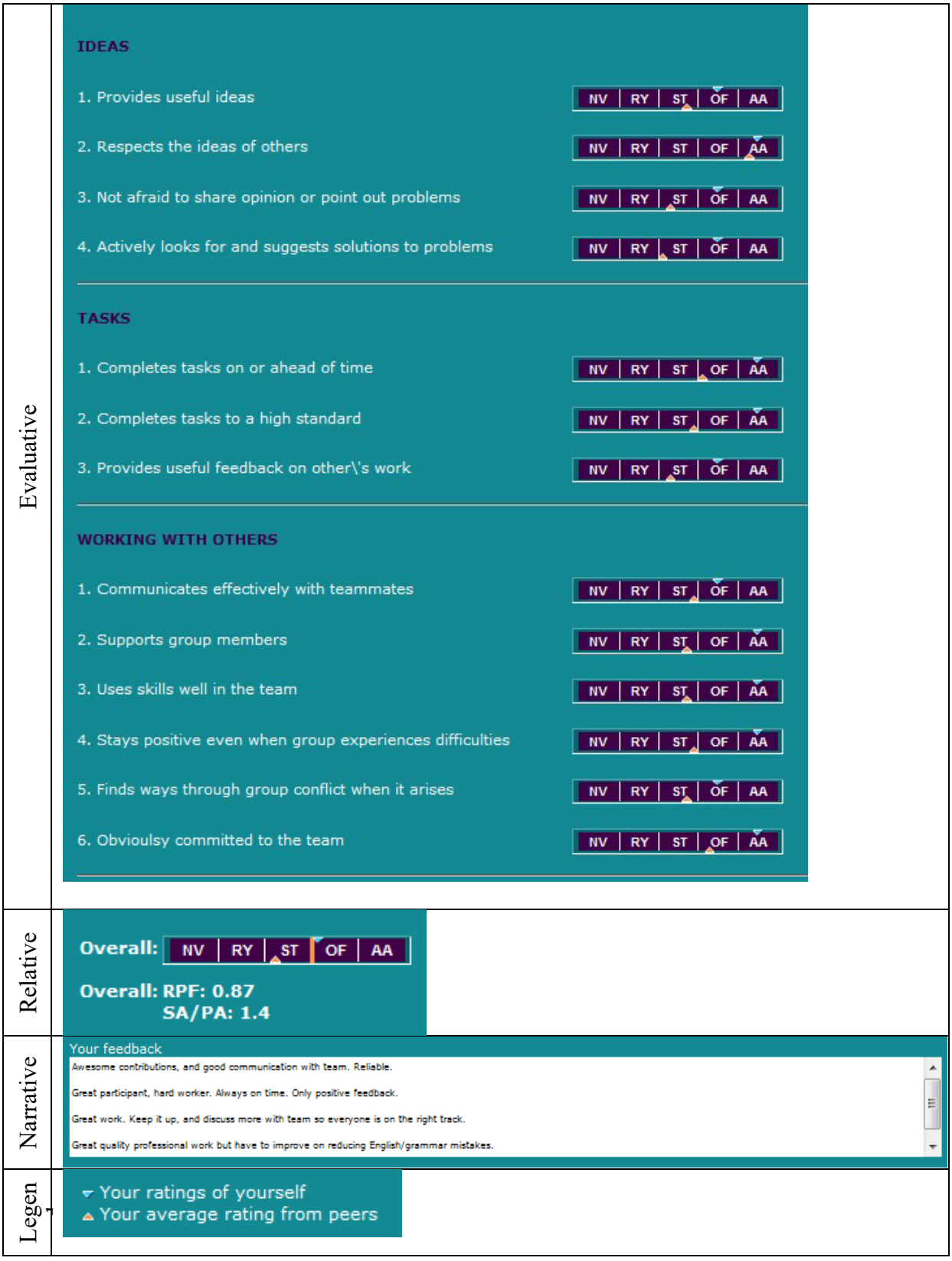

Figure 3. Typical student view of self and peer ratings, relative performance and peer comments

Data measurements

Measures are derived by computing the average/total scores of multiple criteria/assessors to arrive at the aggregate level measure, that is, a parcelling approach. Parcelling in multivariate analysis (eg. path analysis – a type of Structural Equal Modelling (SEM)) is common to address problems such as small sample size, non-normality and to develop a more parsimonious model (Little et al. 2002). The measurement details of the independent variable (performance rating), the mediator variables (quantum of praise and criticism), and the dependent variable (teamwork behaviour and self-assessment skills) are given in Table 1.

Table 1. Measurement details of variables

Constructs/Variables	**Assessment Regime**	**Is relative measure?**	**Agent**	**Data Manipulation technique**
Performance rating (Independent variable)	Formative,	no	Self and peer	Average of self and peer rating for multiple criteria from multiple peers
Quantum of praise (Mediator variable)	Formative	no	Peer	Frequency count of comments oriented towards praise
Quantum of Criticism (Mediator variable)	Formative	no	Peer	Frequency count of comments oriented towards criticism
Teamwork behaviour (Dependent variable)	Summative	Yes	Self and peer	System generated RPF for each student.
Self-assessment ability (Dependent variable)	Summative	yes	Self and peer	Inverse of the system generated of SAPA for each student

The qualitative comments from multiple peers during the formative regime were coded by two experienced researchers and counted to measure quantum of praise and criticism. The coding process involved double validation (two experts independently coding the comments) and conflict resolution of peer comments which entailed resolving any differences in coding between the two through a mutual consensus process. The system-generated RPF is used directly and SAPA is manipulated to ensure consistency in the direction of measure for *teamwork behaviour* and *self-assessment ability* respectively.

In addition to the above variables, the two *control* variables used were *mode of study* and *performance type*. Mode of study is acquired from the university's enrolment data with on-campus or off-campus as the two classifications. Performance type is derived by using the average peer assessment score accrued to individual students for multiple criteria from multiple group members and classified as high (>80%), average (between 60% and 80%) or low (below 60%).

Analysis method

The minimum sample size requirement, for conducting path analysis, of at least ten samples per variable (Hair et al. 2008) is satisfied with around 19 samples per variable. Skewness and kurtosis statistics are used for testing of non-violation of the normality assumption (threshold range of ±3). A two-step approach is used to transform non-normal variables (criticism and teamwork behaviour), requiring converting into percentile rank and then application of the inverse-normal transformation method (Templeton 2011).

SEM path analysis is used to examine how *performance rating* is related to outcome variables without and with parallel mediation measures. All hypotheses are tested using a bootstrapping (number of iterations =2000) procedure with a bias-corrected confidence interval (90%) to estimate the significance of mediation effects (Hayes and Preacher 2008). The presence or absence of a suppression effect is recognised by evaluating the signs of the direct and mediation effect. The opposite sign of the direct and mediator effects indicate the presence and the same sign indicates the absence of a suppression effect (MacKinnon, Krull, and Lockwood 2000). The overall model fitness is evaluated using recommended threshold values from three categories (absolute, incremental and parsimony) of the goodness of fit indices.

Results

Table 2 presents the means, standard deviations and correlation among the measurement variables. A moderate positive correlation is found between performance rating and self-assessment ability, and a negative correlation between performance rating and both praise and criticism. This implies that high performing students received fewer examples of both praise and criticism. Comparatively, low performing students received more praise and criticism. Similarly, self-assessment ability is negatively correlated with both praise and criticism. There was no significant correlation between teamwork behaviour and other variables.

Table 2. Correlation coefficients, mean and standard deviations

Variables	**1**	**2**	**3**	**4**	**5**
1. Performance rating	1				
2. Quantum of praise	-0.41**	1			
3. Quantum of criticism	-0.40**	0.15	1		
4. Teamwork behaviour	0.13	0.01	-0.13	1	
5. Self-assessment ability	0.35**	-0.27**	-0.26*	0.18	1
Mean	4.30	2.52	0.65	1.00	0.97
SD	0.69	1.40	0.75	0.05	0.07

Note: (n=98); *p<0.05l; **p<0.01

Path model without mediation

The path model excluding mediating variables answers the first research question by testing the hypotheses that *performance rating* would positively influence the final *teamwork behaviour* (H1) and *self-assessment ability* (H2). This model displays a good fit for the study's dataset with acceptable threshold values. The model accounts for 4% (weak effect) of the variance in teamwork behaviour and 17% (moderate effect) of the variance in self-assessment ability (Figure 4). The results reveal strong support for H2, with significant path coefficient values of 0.46 ($p < 0.01$) (Table 3). However, the results show the rejection of H1 (dotted line), implying that performance rating failed to influence teamwork behaviour.

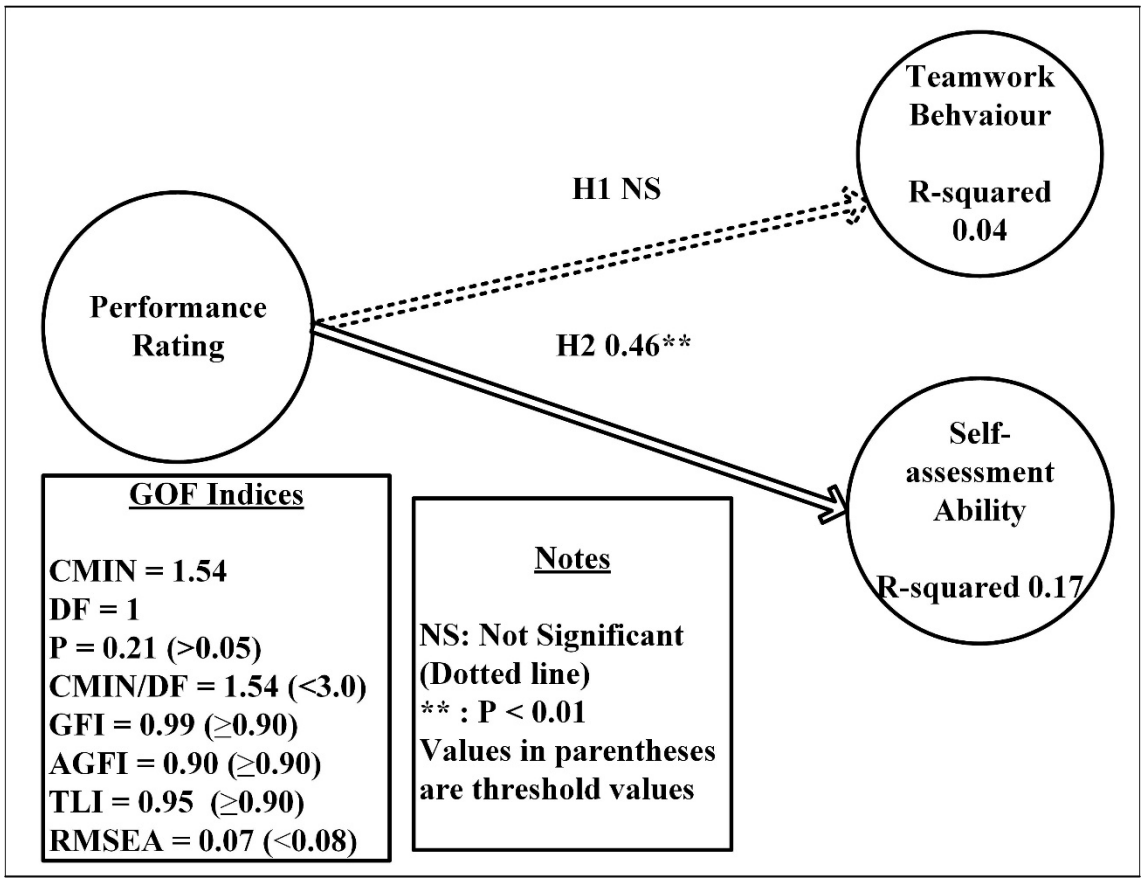

Figure 4. Path model relating performance rating and outcome variables

Table 3. Standardised path coefficients, standard errors and P values for the model without mediators

Hypotheses	**H1**			**H2**		
Variables	**Teamwork Behaviour**			**Self-assessment Ability**		
	Standardized β	SE	P value	Standardized β	SE	P value
Performance rating	0.186	0.015	0.347	**0.460****	0.018	0.012
Performance Type - High	-0.039	0.016	0.805	0.038	0.019	0.797
Performance Type - Low	0.014	0.023	0.201	0.168	0.019	0.921
Off-campus	0.146	0.015	0.153	**0.180†**	0.018	0.061
R square	0.036			0.171		
Effect size (Cohen's f^2)	0.037			0.206		
Hypotheses Result	**Not supported**			**Supported**		
Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001 Controlling for the mode of study and performance type						

To answer the first research question, performance rating positively influences self-assessment ability, however, it does not influence teamwork behaviour. The results of

including confounding variables implies that off-campus students' self-assessment ability is significantly above that of on-campus students. Further investigation is required to explore if this could be due to higher proportion of mature-age students off-campus. Conversely, there was no significant difference between different types of performers (high, medium and low) on both outcome variables.

Path model with mediation

The role of mediating variables, even in situations with lack of significant relationships between dependent and independent variables, is emphasised in the literature (MacKinnon, Krull, and Lockwood 2000) . Therefore, we proceeded with testing the combined effect of two mediators (praise and criticism) on the relationship between performance rating and two outcome variables. Overall goodness of fit results indicate good model fit with acceptable values. The results show that the model accounts for 17% of the variance in praise, 35% of the variance in criticism, 7% of the variance in teamwork behaviour and 21% of the variance in self-assessment ability (see Figure 5).

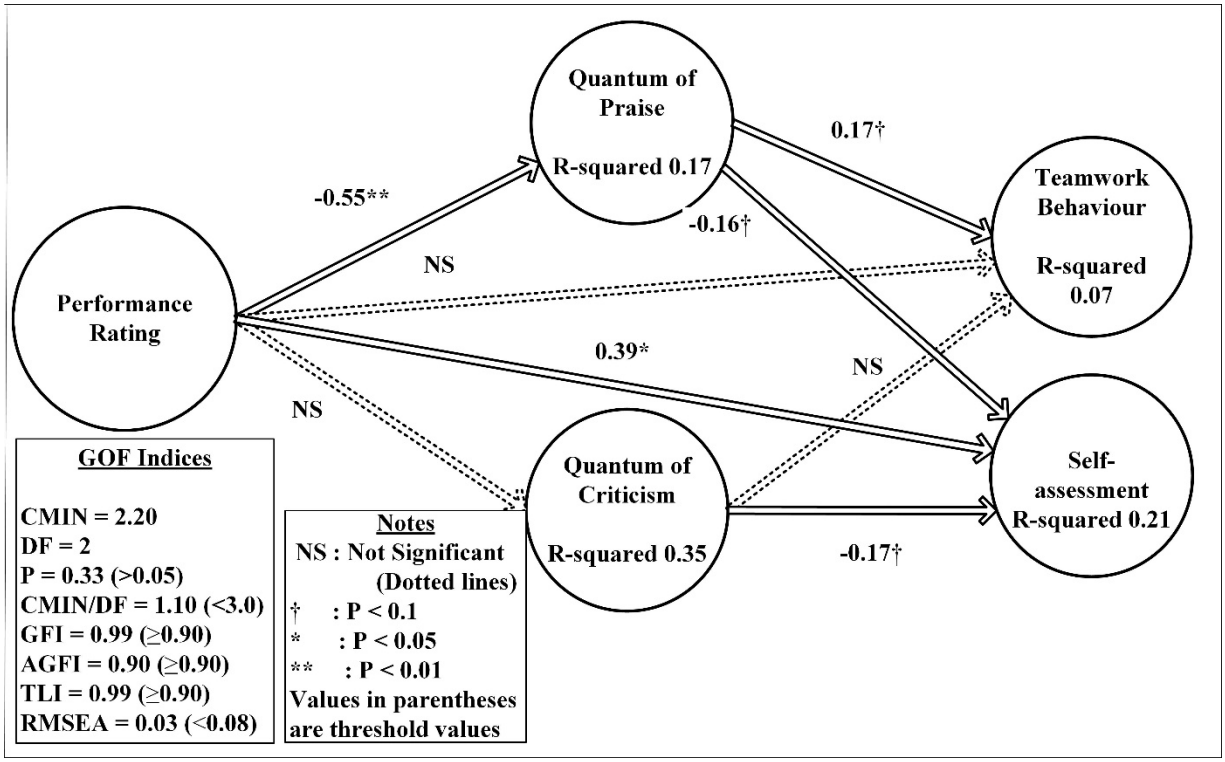

Figure 5. Path model relating performance rating and outcome variables with parallel mediators

Figure 5 shows the combined effect of both mediating variables (praise and criticism) in the relationship between performance rating and outcome variables. The dotted line represents non-significant results and the solid line represents significant results. The results suggest the combined effect of mediating variables on the relationship between performance rating on self-assessment ability is significant and positive (0.39 at 5% significance level). However, the combined mediation effect on teamwork behaviour is insignificant, indicating a lack of effect.

This study aims to explicate the role of each mediating variable on outcome variables. Therefore, the significance of direct and indirect (or mediating) effects is tested separately after controlling for each of the mediating variables (Table 4). *Praise* fully mediates the relationship between performance rating and teamwork behaviour with a significant indirect

effect (-0.09, CI:-0.038, -0.011). Yet, the negative coefficient value indicates the detrimental effect of *praise* on students' *teamwork behaviour*. Additionally, the opposite signs of direct and indirect effect caused weakened (and insignificant) total effect on teamwork behaviour - indicating the presence of a suppression effect. Contrarily, the mediation effect of *praise* on *self-assessment ability* is insignificant. Nevertheless, the positive sign of both indirect and direct effect resulted in higher levels of total (positive and significant) effect on self-assessment ability – suggesting the absence of a suppression effect.

Table 4. Total, Direct and indirect mediation effects, mediation type and suppression effect

Relationship paths	**Total effect**	**Direct effect**	**Indirect effect**	**CI - Lower**	**CI - Upper**	**Mediation Type**	**Suppression effect**
Performance rating→*Praise*→ Teamwork Behaviour	0.19 (NS)	0.28** (0.01)	-0.09† (0.06)	-0.04	-0.01	Full Mediation	Present
Performance rating→*Praise*→Self-assessment Ability	0.46** (0.01)	0.37*** (0.00)	0.08 (NS)	-0.02	0.30	No Mediation	Absent
Performance rating→*Criticism* → Teamwork Behaviour	0.26 (NS)	0.27*** (0.00)	-0.01 (NS)	-0.09	0.03	No Mediation	Present
Performance rating→*Criticism* → Self-assessment Ability	0.37* (0.05)	0.38*** (0.00)	-0.01 (NS)	-0.10	0.03	No Mediation	Present

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001; *NS* Not significant.
Controlling for the mode of study and performance type.

Criticism fails to mediate the relationship between performance rating and both outcome variables. This indicates that criticism neither improved *teamwork behaviour* nor *self-assessment ability*. In fact, the negative sign of the indirect effect reveals the presence of mild suppression effects resulting in a slightly reduced total effect, indicating that criticism was harmful to both outcome variables (see Table 4). The effect of mediating variables on self-assessment ability is positive and much stronger for low performing students compared to their cohorts.

To answer the second research question, quantum of praise negatively mediated the relationship between performance rating and teamwork behaviour, however, no mediation effect was found between performance rating and self-assessment ability.

Table 5 summarises the results of the hypotheses for models with and without mediation. The results without mediation reject H1 and accept H2 with a moderate relationship. The results with mediation provide support for H1a with weak negative relationship and the remaining hypotheses are rejected (H1b, H2a, and H2b rejected).

Table 5 Hypothesis Results

Hypothesis	**Effect**	**Standard Error**	**P Value**	**Hypothesis Supported?**
H1: Performance rating positively influences teamwork behaviour	0.19	0.02	0.35	No
H2: Performance rating positively influences self-assessment ability	**0.46****	0.02	0.01	Yes
H1a: Praise mediates the relationship between performance rating and teamwork behaviour	**-0.09†**	0.07	0.06	Yes (Negative)
H1b: Praise mediates the relationship between performance rating and self-assessment ability	0.08	0.09	0.14	No
H2a: Criticism mediates the relationship between performance rating and teamwork behaviour	-0.01	0.03	0.54	No
H2b: Criticism mediates the relationship between performance rating and self-assessment ability	-0.01	0.04	0.52	No

Note: : †p<0.1; **p<0.01.
Controlling for the mode of study and performance type.

As the study results indicate a detrimental effect of peer comments valence on teamwork behaviour and no effect on self-assessment ability, a further analysis was carried out to gain insight into the study results. The comments are further classified and coded as praise or criticism, future-oriented or past-oriented. Future-oriented comments provide areas for improvement which are absent in past-oriented comments.

Of the total 523 qualitative comments, the percentage of praise (65%) is much higher than criticism (17%). Likewise, past-oriented comments (71%) are more common than future-oriented (11%) comments (see Figure 6). Substantial invalid responses (18%) suggest poor student engagement. Possibly this explains the inefficacy of peer comments on improvements to outcome variables.

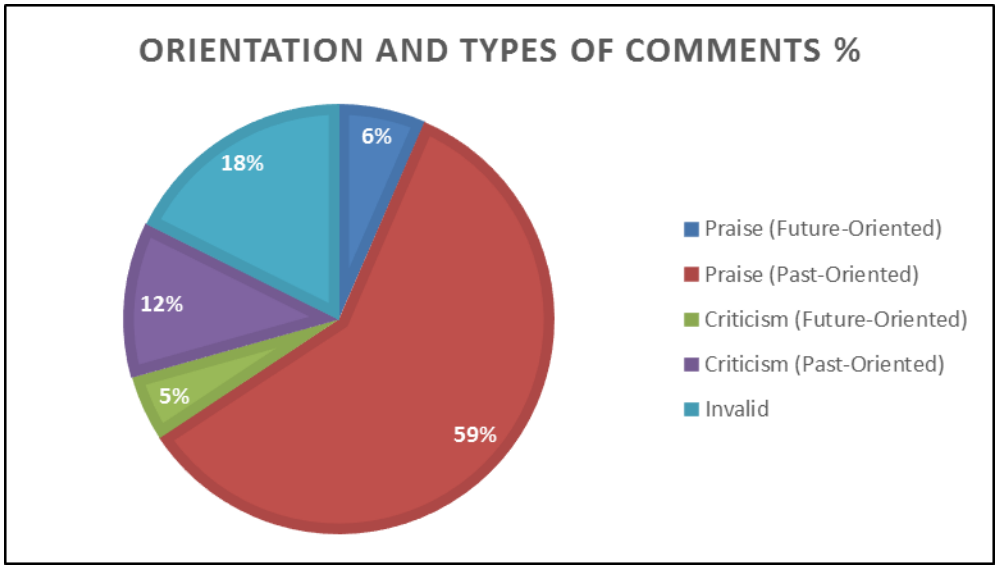

Figure 6. Orientation and types of peer comments

Discussion

This study investigated the role of formative self and peer judgement in influencing learning (outcome variables) in group tasks. To fulfil this objective, a conceptual parallel mediation model was developed and tested using SEM path models. The results suggest that the rating itself, even though past-oriented, is superior to qualitative comments (mostly past-oriented) in impacting on outcome variables.

The key findings from the quantitative analysis are: performance rating has a positive effect on improvements in self-assessment ability, but no effect on teamwork behaviour; praise has weak but significant negative mediation effect ($p<0.1$) on teamwork behaviour, but no effect on self-assessment ability; and criticism has no mediation effect on both outcome variables. Key findings from the qualitative analysis are: a high proportion of peer comments

are past-oriented evaluative comments, focus is mostly around self and process, and lack of engagement of peers. Qualitative comments, as represented in this study, apparently suppress or neutralise the positive effects of performance rating on both outcome variables.

The positive influence of the rating on outcomes is consistent with previous literature. For example, the act of generating self and peer ratings provides an opportunity to self-reflect and help establish normative standards and goal setting (Ohland et al. 2012). Likewise, Manzone et al. (2014) found the positive effect of quantitative grades on performance, although with short-term retention in a cognitive task.

With respect to *praise*, the finding is interesting due to the presence of a suppression effect. The positive direct effect of performance rating and the negative mediator effect of *praise* on *teamwork behaviour*, neutralised each other, resulting in the total effect being negligible (and insignificant). This is not surprising when a significant proportion of praise messages contained no direction and were self-focused. Johnson and Molloy (2018) report 'mealy mouthed' feedback is more common than critical feedback in education. This is consistent with findings that self-focused positive comments have the least effect on improvements to learning (Hattie and Timperley 2007). The findings are similar to the meta-analytic study (Kluger and DeNisi 1996), which found one-third of feedback interventions had a negative effect on performance. Comparably, Van Dijk and Kluger (2011) found that positive feedback (ie. praise) decreased motivation and performance. These findings align with Kohn's (1999) longstanding view that manipulating people by dangling carrots (praise) is harmful, and will lead to long-term failure.

In contrast, the mediation effect of *praise* on *self-assessment ability* is insignificant. Hattie and Timperley (2007) argued that students refrain from investing their effort in feedback if they perceive no favourable benefits. Again, as evidenced by the qualitative analysis, ineffective results make sense, as there is not much value addition evoked from praise on improvements in self-assessment ability.

The insignificant mediation effect of *criticism* reveals that it failed to influence both *teamwork behaviour* and *self-assessment ability*. In fact, the mediator effect is negative on both outcomes and weakened the positive significant direct effect of *performance rating* on both outcome variables. This suggests *criticism* worsened students' *teamwork behaviour* and *self-assessment ability*. Prior studies have however found that negative feedback can improve motivation and performance (Kluger and DeNisi 1996; Hattie and Timperley 2007; Van Dijk and Kluger 2011). For example, Van Dijk and Kluger (2011) found negative feedback is more effective when the task requires creativity. However, if the negative feedback demoralises and negatively impacts their self-image, students would not engage with such feedback (Hattie and Timperley 2007). Two potential explanations for the study's findings are, the agent involved is the peer (equals and not superior), and there is an absence of forward-looking peer critical comments. Overall, the results emphasise that even though rich information is associated with praise and criticism, the important requirement for an influence on learning is substantive content and not emotional tone.

These findings highlight how erroneous it is to assume that peer comments *per se* will impact learning. The fundamental problems identified are as follows: (a) poor quality of peer comments (mostly backward-looking); (b) lack of engagement in giving feedback messages with high invalid responses (18%); (c) possible perceived low value as it is coming from their counterpart, rather than from teachers (power imbalance); (d) lack of specifics for actioning peer comments; (e) possible defiance reaction to peer comments resulting in a behavioural confirmation effect, or self-fulfilling prophecy, or perceived patronising nature of peer comments; and (f) possible reticence of low performing students towards evaluating their peers.

It is unsurprising students have difficulty in providing, interpreting and acting on feedback. The process is even more intricate with respect to emotional (teamwork) and self-regulatory aspects of learning, and when the sources are self and peers. Three practical

strategies are proposed to deal with these challenges. First, *scaffolding the curriculum* by the integration of initiatives to develop feedback literacy into the curriculum throughout the teaching and learning cycle (Carless and Boud, 2018). Second, *capacity building* exercises to provide multiple opportunities for students' skills to develop in giving, receiving, acting on feedback. Third, embedding *forward* (eg. feedback on peer feedback) *and reverse feedback* (eg. teachers taking cues from students feedback abilities) processes to augment peer engagement and quality of peer feedback by scaffolding feedback evaluation tasks. We suggest that rather than using students' taken-for-granted views about peer feedback activities, investigations be based on interventions which involve training, exposure and experience in giving, receiving and acting on feedback, utilising a strong pedagogical base.

This study advances feedback intervention theory by uncovering boundary conditions for success, by identifying the presence of a suppression effect in the feedback domain. It provides empirical evidence demonstrating that feedback literacies are essential for realising the benefits of peer interventions. In a practical sense, the findings are of value to higher education practitioners, who may be oblivious to the possibly detrimental effects of peer comments on learning in unprepared students. From a methodological perspective, the parallel mediation model provides empirical evidence identifying the hidden negative effects of praise, and uncovers the presence of suppression effects on learning of both praise and criticism.

A number of limitations must, however, be noted. Because of the small sample size, aggregating of self and peer rating was required to retain the parsimony of the model, but separate measures need to be considered in the future. Replication with a larger sample is required to confirm the findings. Other factors such as diversity of students (domestic and international) may confound the effects of this study. The retrospective nature of the study resulted in the domination of praise and criticism in qualitative comments. Prospective longitudinal research is required to evaluate the long-term effect of feedback on outcome variables.

Conclusion

The study indicates that peer feedback interventions can fail, when feedback literacies of students are neglected before implementation of such practices; specifically, in the absence of appropriate training on how to give, receive and act on peer comments. Surprisingly, the use of ratings alone appears to be superior to qualitative comments. Regarding valance, the results suggest that praise is detrimental and criticisms are ineffective on outcome variables, as substantive peer comments were static and past-oriented. This study demonstrates that the common practice of using peer comments with no preparation, which follows the 'cart before horse' approach (implementation before training), are not only ineffective but also can be detrimental owing to the resulting suppression effect of qualitative peer comments on outcome variables. A fundamental transformation is thus required in peer feedback practices by scaffolding them into a cycle of teaching and learning.

Notes on Contributors

Bhavani Sridharan is a Lecturer in Learning Innovations with the Faculty of Business and Law, Deakin University, Melbourne, and a Research Fellow at the Centre for Research in Assessment and Digital Learning (CRADLE). She earned her Masters' degree from the London School of Economics, University of London and PhD from RMIT University.

David Boud is Alfred Deakin Professor and the Director of the Centre for Research in Assessment and Digital Learning, Deakin University, Melbourne. He is also emeritus professor in the Faculty of Arts and Social Sciences, University of Technology Sydney and a research professor in the Middlesex University, London.

Funding

This work was supported by funding by the Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, [grant number 2016/01].

Acknowledgements

We would like to thank Margaret Bearman and Colin Higgins for their insightful feedback on an earlier version of the paper, Tarikul Islam for his expert methodological input, Jamie Mustard for his support,

Jade McKay for her research assistance and Andrew Torre for proof reading support in preparing this paper.

References

AACSB. 2017. *Eligibility procedures and accreditation standards for business accreditation*. Tampa, USA: AACSB

Adrian, M. 2017. "Determining the Skills Gap for New Hires in Management: Student Perceptions vs Employer Expectations." *International Journal for Innovation Education and Research* 5 (6):139-47.

Barber, J.P., P.M. King, B. Magolda, and M. B. 2013. "Long Strides on the Journey toward Self-Authorship: Substantial Developmental Shifts in College Students' Meaning Making." *The Journal of Higher Education* 84 (6):866-96.

Black, P., and D. Wiliam. 2010. "Inside the black box: Raising standards through classroom assessment." *Phi Delta Kappan* 92 (1):81-90.

Boud, D., and E. Molloy. 2013. "Rethinking models of feedback for learning: the challenge of design." *Assessment & Evaluation in Higher Education* 38 (6):698-712.

Carless, D., and D. Boud. 2018. "The development of student feedback literacy: enabling uptake of feedback." *Assessment & Evaluation in Higher Education*:1315-25.

Cho, M.-H., and B.J. Kim. 2013. "Students' self-regulation for interaction with others in online learning environments." *The Internet and Higher Education* 17:69-75.

Cushing, A., S. Abbott, D. Lothian, A. Hall, and O.M. Westwood. 2011. "Peer feedback as an aid to learning–What do we want? Feedback. When do we want it? Now!" *Medical teacher* 33 (2):e105-e12.

Davies, W.M. 2009. "Groupwork as a form of assessment: Common problems and recommended solutions." *Higher Education* 58 (4):563-84.

Dominick, P.G., R.R. Reilly, and J.W. McGourty. 1997. "The effects of peer feedback on team member behavior." *Group & Organization Management* 22 (4):508-20.

Donia, M.B.L., T.A. O'Neill, and S. Brutus. 2018. "The longitudinal effects of peer feedback in the development and transfer of student teamwork skills." *Learning and Individual Differences* 61:87-98.

Dunning, D., C. Heath, and J.M. Suls. 2004. "Flawed self-assessment: Implications for health, education, and the workplace." *Psychological science in the public interest* 5 (3):69-106.

Gabelica, C., P. Van den Bossche, M. Segers, and W. Gijselaers. 2012. "Feedback, a powerful lever in teams: A review." *Educational Research Review* 7 (2):123-44.

Hair, J.F., M.W. Celsi, D.J. Ortinau, and R.P. Bush. 2008. *Essentials of marketing research*. Ohio: McGraw-Hill.

Hattie, J., and H. Timperley. 2007. "The power of feedback." *Review of educational research* 77 (1):81-112.

Hayes, A.F., and K.J. Preacher. 2008. "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models." *Behavior research methods* 40 (3):879-91.

Johnson, C., and E. Molloy. 2018. "Building opportunities for evaluative judgement in feedback sessions." In *Developing Evaluative Judgement in higher education*, edited by D Boud, R Ajjawi, P Dawson and J Tai. London: Routledge.

Kluger, A.N., and A. DeNisi. 1996. "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory." *Psychological Bulletin* 119 (2):254-84.

Kohn, A. 1999. *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes*: Houghton Mifflin Harcourt.

Little, T.D., W.A. Cunningham, G. Shahar, and K.F. Widaman. 2002. "To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits." *Structural Equation Modeling: A Multidisciplinary Journal* 9 (2):151-73.

Liu, N.-F., and D. Carless. 2006. "Peer feedback: the learning element of peer assessment." *Teaching in Higher Education* 11 (3):279-90.

MacKinnon, D.P., J.L. Krull, and C.M. Lockwood. 2000. "Equivalence of the Mediation, Confounding and Suppression Effect." *Prevention Science* 1 (4):173-81.

Magpili, N.C., and P. Pazos. 2017. "Self-managing team performance: A systematic review of multilevel input factors." *Small Group Research* 49 (1):3-33.

Manzone, J., L. Tremblay, K.E. You-Ten, D. Desai, and R. Brydges. 2014. "Task-versus ego-oriented feedback delivered as numbers or comments during intubation training." *Medical education* 48 (4):430-40.

Medvedeff, M., J.B. Gregory, and P.E. Levy. 2008. "How Attributes of the Feedback Message affect Subsequent Feedback Seeking: The interactive effects of feedback sign and type." *Psychologica Belgica* 48 (2-3):109–25.

Nash, R., N. Winstone, S. Gregory, and E. Papps. 2018. "A memory advantage for past-oriented over future-oriented performance feedback." *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Nicol, D.J., and D. Macfarlane-Dick. 2006. "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice." *Studies in Higher Education* 31 (2):199-218.

Ohland, M.W., M.L. Loughry, D.J. Woehr, L.G. Bullard, R.M. Felder, C.J. Finelli, R.A. Layton, H.R. Pomeranz, and D.G. Schmucker. 2012. "'The comprehensive assessment of team member effectiveness: development of a behaviorally anchored rating scale for self- and peer evaluation." *Academy of Management Learning & Education* 11 (4):609-30.

Salas, E., M.L. Shuffler, A.L. Thayer, W.L. Bedwell, and E.H. Lazzara. 2015. "Understanding and improving teamwork in organizations: A scientifically based practical guide." *Human Resource Management* 54 (4):599-622.

Sitzmann, T., K. Ely, K.G. Brown, and K.N. Bauer. 2010. "Self-assessment of knowledge: A cognitive learning or affective measure?" *Academy of Management Learning & Education* 9 (2):169-91.

Sridharan, B., M.B. Muttakin, and D.G. Mihret. 2018. "Students' perceptions of peer assessment effectiveness: an explorative study." *Accounting Education*:1-27.

Sridharan, B., J. Tai, and D. Boud. 2018. "Does the use of summative peer assessment in collaborative group work inhibit good judgement? ." *Higher Education*.

Tai, J., R. Ajjawi, D. Boud, P. Dawson, and E. Panadero. 2017. "Developing evaluative judgement: enabling students to make decisions about the quality of work." *Higher Education*.

Templeton, G.F. 2011. "A two-step approach for transforming continuous variables to normal: implications and recommendations for IS research." *Communications of the Association for Information Systems* 28 (1):41-58.

Topping, K.J. 2009. "Peer assessment." *Theory into practice* 48 (1):20-7.

Van Dijk, D., and A.N. Kluger. 2011. "Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective." *Journal of Organizational Behavior* 32 (8):1084-105.

Varela, O., and E. Mead. 2018. "Teamwork skill assessment: Development of a measure for academia." *Journal of Education for Business*:1-11.

Willey, K., and A. Gardner. 2009. "Improving self-and peer assessment processes with technology." *Campus-Wide Information Systems* 26 (5):379-99.

Winstone, N.E., R.A. Nash, J. Rowntree, and M. Parker. 2017. "'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience." *Studies in Higher Education* 42 (11):2026-41.