Confounding factors in using upward feedback to assess the quality of medical training: a systematic review

Anli Yue Zhou1*, Paul Baker2
1Royal Bolton Hospital, Lancashire, United Kingdom; 2North Western Deanery, Manchester, United Kingdom

Abstract

Purpose: Upward feedback is becoming more widely used in medical training as a means of quality control. Multiple biases exist, thus the accuracy of upward feedback is debatable. This study aims to identify factors that could influence upward feedback, especially in medical training. Methods: A systematic review using a structured search strategy was performed. Thirty-five databases were searched. Results were reviewed and relevant abstracts were shortlisted. All studies in English, both medical and non-medical literature, were included. A simple pro-forma was used initially to identify the pertinent areas of upward feedback, so that a focused pro-forma could be designed for data extraction. Results: A total of 204 articles were reviewed. Most studies on upward feedback bias were evaluative studies and only covered Kirkpatrick level 1-reaction. Most studies evaluated trainers or training, were used for formative purposes and presented quantitative data. Accountability and confidentiality were the most common overt biases, whereas method of feedback was the most commonly implied bias within articles. Conclusion: Although different types of bias do exist, upward feedback does have a role in evaluating medical training. Accountability and confidentiality were the most common biases. Further research is required to evaluate which types of bias are associated with specific survey characteristics and which are potentially modifiable.

Key Words: Bias; Confidentiality; Feedback; Quality control; Social responsibility

Introduction

Multiple methods of feedback exist, which include downward feedback, upward feedback, peer feedback and self-evaluation. The most commonly known form of feedback is downward appraisal, where the supervisor gives feedback to the subordinate [1]. However, upward feedback, where the feedback is given from the subordinate to the supervisor is becoming more recognized and adopted, especially in the private sector. It has been reported that over 90% of fortune 100 companies in the United States participate in some form of upward feedback [1]. The role of upward feedback has also been widely acknowledged within the educational sector as well, where students give feedback to their lecturers [2-7]. Within medical training, the General Medical Council (GMC) in the United Kingdom has adopted upward feedback to monitor teaching performance for quality control purposes [8]. Although upward feedback has been advocated by the GMC, it is not immune from bias and there has been much debate about the accuracy of upward feedback [9-17]. This systemic review has been prompted by the increasing significant role of upward feedback as medical training becomes more closely regulated. Bias present within upward feedback could potentially skew feedback on medical training and this review aims to identify these factors.
METHODS

Search strategy

In order to obtain a comprehensive overview of the literature in upward feedback, a total of 35 databases were searched (Embase, Medline, PsychINFO, Cochrane and EBM Reviews, Allied and Complementary Medicine, CAB and ATLA Religion Database, Econ lit, GeoBase, Global Health, Health and Psychosocial Instruments, HMIC Health and Management, Index to Foreign Legal Periodicals, International Pharmaceutical Abstracts, Maternity and Infant Care, The Philosopher’s Index, Social Policy And Practice, Zoological Records, BNI, CINAHL, Health Business Elite, ERIC, British Educational Index, ASSIA, Web of Knowledge, Social Care Online, Sage Full Text Journals, IBBS, National Research Register Archive, Proquest, Wiley Online Library, Taylor and Francis, Engineering Village, Scopus, Science Direct, PubMed). A stratified search involving multiple keywords was used (Fig. 1).

Searches were initially done to search within all fields. If more than 1,000 results were returned, then the search would be repeated to search within keywords, then abstract and then within the title in order to narrow down results to less than 1,000 articles. Search results of less than 1,000 articles were reviewed by reading the abstract; relevant abstracts were then shortlisted. If no abstract was available, but the title appeared relevant, this would also be temporarily shortlisted until further information could be obtained from the full article. Further references were found by reviewing the reference bibliography of the shortlisted articles.

Inclusion and exclusion criteria

Both medical and non-medical articles written in English were included. No time limit was set. Books were excluded from the search.

Data management techniques

A proforma was developed to allow efficient and relevant data extraction. This included: study method (e.g., observational or review article), profession, type of participant, geographical location, purpose of feedback (e.g., summative or formative), feedback subject (e.g., trainer, training or environment), qualitative/quantitative feedback, the use of controls and type of intervention involved (e.g., counseling, timing of feedback), type of feedback used (e.g., paper survey, semi-structured interviews), quality of questions (e.g., closed, open), duration of study, number of participants, response rates, types of bias present (overt and implied), Kirkpatrick level [18] and whether outcomes were addressed.

RESULTS

Literature search and selection

A total of 8,914 potential articles were found using the search strategy (Fig. 1), in which 291 articles were shortlisted. The shortlisted articles were then subsequently pooled together and duplicates were removed. A total of 169 articles were shortlisted after the removal of duplicates. By reviewing the reference bibliography of the shortlisted articles, a further 70 articles were shortlisted. A total of 239 references were shortlisted. After reviewing the articles 35 articles were excluded from further analysis. This was due to: 10 articles were not relevant to the objective, 1 reference was a book, complete versions were not obtainable for 21 references, 2 references were not written in English and 1 reference was a duplicate of another shortlisted reference but was under a different title. This lead to a total of 204 articles being analyzed, all of which are presented in Table 1.

Fig. 1. Search strategy of related papers for systemic review.
Demographics

More than 50% of the references were related to the medical profession (n = 109). Other professions that have commonly utilized upward feedback include teaching and education (n = 39), nursing (n = 22) and management (n = 18). The majority of references included postgraduate participants (n = 106). Thirteen references included both undergraduate and postgraduate participants. A large proportion of references were from North America (Fig. 2).

Types of studies and feedback

Studies were categorized according to the definitions in Table 2. Most references were evaluation studies (n = 176) and most studies were done for formative purposes (n = 172). A large majority of studies were quantitative (n = 152) and high proportion of studies used paper surveys as a means of evaluating upward feedback (n = 124). Most studies (n = 162) only covered Kirkpatrick level 1, reaction. The median response rate was 76%, the median number of participants was 198 and the median duration of the study was 6 months. Only 1/3 of references addressed the outcomes of their study by develop-

---

**Table 1. Summary of all the references shortlisted and analysed in this systematic review**

<table>
<thead>
<tr>
<th>Type of participant</th>
<th>Medical References</th>
<th>Non-medical References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate</td>
<td>Langenfeld et al. [50], Rabow et al. [84], Brasher et al. [100], Metcalfe and Matharu [106], Blue et al. [117], Iqbal and Khizar [119], Solomon et al. [127], Johnson and Chen [129], Windish et al. [130], Ramsey et al. [135], Tochel et al. [144], Fallon et al. [152], Stritter et al. [153], Shellenberger and Mahan [154], Cohen et al. [155], Dalmans et al. [156], Donnelly and Wooliscroft [157], Irby and Rakeshaw [158], Parikh et al. [159], Wilson [160], De et al. [161], Duffield and Spencer [162], Tiberius et al. [163], Gil et al. [164], Pfeiffer and Peterson [165]</td>
<td>Alissa and Sulikeman [9], Bernardin [19], Crittenden and Norr [20], Adams and Umbach [21], Wolbring [22], Remedios and Lieberman [24], Chen and Hoshower [25], Worthington [26], Kember and Wong [27], Marsh [28], Marsh [29], Marsh and Roche [30], Rowden and Carlson [31], Goos et al. [32], Davies et al. [33], Blackhart et al. [34], Dvinelli and Higbee [35], Burdalls and Bardo [36], Theall and Franklin [38], Feldman [39], Sojka et al. [40], Berk [41], Greenland and Gillmore [42], Gigliotti and Buchtel [43], Doyle and Crichton [44], Aleamoni [46], Kember and Leung [52], Roch and McNall [67], Atwater et al. [73], Redman and McEbee [74], Some [76], Henderson et al. [77], Brugnoli et al. [78], Middley [80], Per Palmgren [82], Olson et al. [87], Braine and Parnell [88], Perli and Brugnoli [89], Heffeman et al. [90], Kelly [91], El Ansari and Oskrochi [102], Berber [110], Robbins and DeNissi [134], Govaerts et al. [146], Surratt and Desselle [149], Cardy and Dobbins [166], Henzi et al. [167], Parker and Carlisle [168], Cooke et al. [169], Myall et al. [170]</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>Archer et al. [10], Barrow and Baker [12], Coats and Burd [13], Anah et al. [47], Schneider et al. [51], Scott et al. [53], Ahearn et al. [55], Gravagubins and Scott [63], Owen [64], Fliander [65], Ricucci et al. [66], Kolank et al. [85], Smith et al. [86], Ranse and Grealish [92], O’Connor et al. [95], Luks et al. [96], Turnball et al. [97], Biller et al. [98], Carpenter et al. [99], Busari et al. [101], Basu et al. [103], Whang et al. [104], Devlin et al. [105], Barrett et al. [107], Seiner et al. [108], Getz and Evens [109], Girard et al. [111], Antiel et al. [112], Lin et al. [113], Ratana-wongsa et al. [114], Thangaratnam et al. [115], Kanashirou et al. [116], Watling et al. [118], Watling et al. [120], Peacock et al. [122], Conigliaro et al. [123], Dech et al. [124], Yarris et al. [125], Sargeant et al. [126], Sender Lieberman et al. [128], Tortolani et al. [131], O’Brien et al. [132], Claridge et al. [133], Hayward et al. [136], Sargeant et al. [137], Peace et al. [141], Ryland et al. [142], Rose et al. [145], Bing-you et al. [151], Kyer et al. [171], Hrisos et al. [172], Beckman et al. [173], Mattern et al. [174], Kendrick et al. [175], Keitz et al. [176], Moalem et al. [177], Sargeant et al. [178], Schuh et al. [179], Vasudev et al. [180], Ellodt [181], Harrison and Allen [182], Dola et al. [183], Cohn et al. [184], Fisher et al. [185], Pankhania et al. [186], Welch et al. [187], Geyser et al. [188], Mailoux [189], Buschbacher and Braddon [190], Cooke and Hutchinson [191], Holland et al. [192], Sabey and Harris [193], Nettleton et al. [194], Chamberlain and Nisker [195], Verhulst and Distleberger [196], Guyatt et al. [197], Barclay et al. [198]</td>
<td>McCarthy and Garavan [1], Hall et al. [14], Glassie et al. [15], Smith and Fortunato [16], Kudisch et al. [17], Mullen and Talant-Ruulens [37], Tews and Tracey [48], Tews and Tracey [49], Smith et al. [54], Antonini and Park [56], Tsui and Barry [57], Ryan et al. [59], Antonioni et al. [61], Goodwin and Yeo [62], Antonioni et al. [68], Bennehaven and Fedor [69], Westman and Rosse [70], Brugnoli et al. [71], Reid and Levy [72], Redman and Snafe [75], Cohain [81], Rak-konen et al. [83], Beecroft et al. [93], Sit et al. [94], Brett and Atwater [138], Barclay et al. [139], Tourish and Robson [148], Dibbyoe and de Pontbriand [199], Coppel et al. [200], Bratt and Feizer [201], Smithher and Walker [202], Becker et al. [203]</td>
</tr>
<tr>
<td>Both undergraduate and postgraduate</td>
<td>Gross et al. [23], Schum et al. [45], Albanese et al. [58], Evan et al. [60], Cannon et al. [121], Irby [140], Watling and Linjard [143], Williams et al. [147], McLeod et al. [204], Bennett et al. [205]</td>
<td>Iilgen et al. [150], Henzi et al. [206], Henzi et al. [207]</td>
</tr>
</tbody>
</table>
**Fig. 2.** Geographical locations of studies in the targeted papers for systematic review.

**Table 2.** Summary of categories used within the focused proforma

<table>
<thead>
<tr>
<th>Proforma categories</th>
<th>Further information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Number</td>
<td>Each article was allocated a number to allow easy identification.</td>
</tr>
<tr>
<td>2. Study method</td>
<td>What type of study was it?</td>
</tr>
<tr>
<td>3. Profession</td>
<td>What profession were the participants?</td>
</tr>
<tr>
<td>4. Type of participant</td>
<td>Undergraduate or postgraduate or both?</td>
</tr>
<tr>
<td>5. Geographical location</td>
<td>Which continent was the article from?</td>
</tr>
<tr>
<td>6. Purpose of study</td>
<td>Was the study for summative (for promotional/reward purposes) or formative (for improvement/development) purposes?</td>
</tr>
<tr>
<td>7. Feedback subject</td>
<td>Feedback on training, trainer or learning environment?</td>
</tr>
<tr>
<td>8. Quality of feedback</td>
<td>Quantitative or qualitative?</td>
</tr>
<tr>
<td>9a. Were controls used?</td>
<td>Controls may be used to compare the efficacy of different interventions.</td>
</tr>
<tr>
<td>9b. Type of interventions</td>
<td></td>
</tr>
<tr>
<td>10a. Type of evaluation</td>
<td>What type of feedback method was used? e.g., paper survey, focus groups</td>
</tr>
<tr>
<td>10b. Quality of questions</td>
<td>What types of questions were used? e.g., closed, open mixture</td>
</tr>
<tr>
<td>11. Duration of study</td>
<td>Measured in months</td>
</tr>
<tr>
<td>12. Number of participants</td>
<td>Total number of participants giving upward feedback</td>
</tr>
<tr>
<td>13. Response rates</td>
<td>Measured in percentages</td>
</tr>
<tr>
<td>14. Types of bias</td>
<td>Split into implied and overt:</td>
</tr>
<tr>
<td></td>
<td>Overt bias would be explicitly mentioned by the authors within the study.</td>
</tr>
<tr>
<td></td>
<td>Implied bias would be identified by the reviewer as potential bias but was not mentioned within the study.</td>
</tr>
<tr>
<td>15. Action plans</td>
<td>Did the authors address the outcomes/consequences of the article? Was an action plan devised to address this?</td>
</tr>
<tr>
<td>16. Kirkpatrick levels</td>
<td>Which level? [18]</td>
</tr>
<tr>
<td></td>
<td>(1) Reaction: What do the raters think about their trainer/training/environment?</td>
</tr>
<tr>
<td></td>
<td>(2) Learning: Was the ratee able to learn from this feedback? This can be identified through mechanisms such as feedback reports, receiving results.</td>
</tr>
<tr>
<td></td>
<td>(3) Behavior: Did the ratee change their behavior due to this feedback? This can be reflected in repeat ratings.</td>
</tr>
<tr>
<td></td>
<td>(4) Results: Was there any improvement in teaching after receiving the feedback? Did others benefit from this improvement? For example, did exam rates improve? Did this change improve company profits?</td>
</tr>
</tbody>
</table>

Fig. 3. Type of interventions used in control studies.

**Types of bias**

Types of bias data separated into implied and overt bias. Implied bias involves factors that potentially could affect the upward feedback process but was not explicitly acknowledged within the article. Overt bias included factors affecting the upward feedback process that were mentioned within the article. A summary of the different types of bias found in this systematic review can be found in Table 3. Accountability and confidentiality were the most common biases recognized within references. On the other hand, the method of feedback, which involves the type of survey, the location, the use and methodology of reminders and the duration, were most commonly implied within articles but not explicitly acknowledged (Table 4).
DISCUSSION

This review shows that multiple sources of bias, in the important task of using feedback in the assessment of training quality, are already described.

Feedback philosophy

Although there has been extensive research on upward feedback within an undergraduate classroom setting [2-7,9,19-46], the high proportion of references related to the medical profession and to postgraduate participants confirms the popularity of upward feedback in postgraduate medical training. The majority used surveys for formative purposes, which can provide the trainer/teacher with guidance on their current performance. The lack of studies for summative purposes could be due to raters tending to be over-lenient when upward feedback was for administrative purposes [14,17,39]. However, in contrast, Smith and Fortunato [16] found that rating purpose did not affect intentions to provide honest ratings since raters

| Table 3. Different types of bias identified within the systematic review |
|-----------------------------|---------------------------------|
| **Type of bias** | **Further information** |
| 1. Affect/leader-member relationship | Defines the relationship between ratee and rater [57,134]. The bias of liking someone may lead to potentially inaccurate ratings. |
| 2. Motivation | Low response rates may not be representative of the sampled population. This could potentially be due to lack of motivation. Prior interests, including prior subject interest [4,30] could also affect participation and enthusiasm. For example, did students volunteer themselves to enter into the study? A response rate of 60% or more is perceived as an acceptable level [208]. Articles that explicitly mention rater motivations, enthusiasm or prior subject interests were also included. |
| 3. Fear and retaliation, career progression | The fear that honest ratings could lead to retaliation and affect career progression, could potentially affect upward feedback outcomes [12]. |
| 4. Self efficacy, lack of understanding/knowledge of upward feedback, role appropriateness | Do raters feel they are suitable/appropriate/confidence to rate their superiors [11,17]? |
| 5. Cynicism and trust, perceived usefulness | Raters may not feel their voice will be heard and may be skeptical that changes will be made according to their feedback [16]. |
| 6. Ingratiation, yea saying, leniency, reward anticipation/incentives | Raters may rate leniently as a means of showing ingratiating or to receive reward in return [11]. |
| 7. Method of feedback | This includes how survey was implemented e.g. paper, online, the location of survey implementation [115], whether any reminders and method of reminders [55]. Also included whether the survey was done over a period of time or only used 1 day/session [115]. |
| 8. Voluntary/compulsory | All members had to participate or could choose not to participate. |
| 9. Frequency/timing, opportunity to observe | The timing of the survey: Was it done straight after rotation, or done many months after rotation, or done in the middle of the rotation [201]. |
| 10. Cultural/gender | Cultural differences may affect survey accuracy [78,119]. Gender could affect survey differences e.g., nursing where the survey population is predominantly female [83]. |
| 11. Halo effect | Raters have a tendency to give similar ratings to all aspects of a survey [11,57]. Raters are not able to differentiate between different traits. |
| 13. Survey fatigue | If there are multiple surveys to complete in the study or if the survey was very long, then this could affect survey accuracy. |
| 14. Survey purpose | Was the survey for administrative or developmental purposes [11,41]? Why was the survey done? |
| 15. Others | Potential biases that could also potentially affect bias but not mentioned above. e.g., recall bias [201]. |

| Table 4. Summary of types of upward feedback bias identified |
|-----------------------------|---|---|
| **Type of feedback bias** | **Implied** | **Overt** |
| Affect, leader-member relationship | 76 | 39 |
| Motivation | 42 | 14 |
| Fear and retaliation | 31 | 32 |
| Self efficacy, lack of understanding/knowledge of upward feedback, role appropriateness | 56 | 28 |
| Cynicism and trust, perceived usefulness | 67 | 32 |
| Accountability and confidentiality | 54 | 117 |
| Ingratiation, yea saying, leniency, reward anticipation/incentives | 30 | 52 |
| Method of feedback | 104 | 39 |
| Voluntary/compulsory | 35 | 102 |
| Frequency/timing, opportunity to observe | 37 | 31 |
| Cultural or gender bias | 68 | 23 |
| Halo effect | 8 | 10 |
| End aversion/extreme response | 14 | 5 |
| Survey fatigue | 50 | 8 |
| Survey purpose | 66 | 37 |
| Others | 13 | 11 |
may use the purpose as a tool to retaliate and reward their supervisors. Upward feedback could potentially be used as a tool to develop clinical trainers and to give guidance to clinical educators on their own career plans [47]. However, the effectiveness of upward feedback could be confounded by multiple factors, which will be discussed below. Most studies only evaluated Kirkpatrick level 1–reaction, which mostly involved surveying subordinate’s views on certain topics. Only 10 studies covered Kirkpatrick level 4–outcomes [1, 4, 5, 38, 44, 48–52]. The majority of studies did not address the consequences or results of the study. This could be because it may be difficult to develop specific action plans based on Kirkpatrick level-one evidence. Furthermore, very few studies specifically compare the different factors or their effect on feedback quality.

Study administration

Upward feedback usually involves subordinates to appraise their superiors or training, hence it is not surprising that the majority of studies were evaluation studies. Only one study was a randomized controlled trial that stratified participants into 3 groups (online survey, simultaneous paper and online survey, sequential online and paper survey) [53]. This study found that the sequential survey method, in which online and paper surveys were administered at different times, gave the highest response rate but increased costs [53]. The small number of studies involving controls could be due to time and financial constraints. Controlled trials of educational interventions are rare, but more studies may need to include controls if we are to assess the efficacy of the different interventions. Without evidence for the effectiveness of interventions, it may be difficult for trainers to accept upward feedback from their subordinates. Tews and Tracey [49] showed that managers who participated either in self-coaching courses or upward feedback intervention, improved interpersonal scores compared to controls. Managers who participated in the upward feedback training scored higher overall [49]. This could due to the fact that upward feedback, if utilized appropriately, can facilitate information sharing, act as a refresher in order to avoid complacency and promote further development of skills [48]. Another form of support in upward feedback was the use of feedback reports, as demonstrated in Smither et al. [54]’s study. Feedback reports enabled managers to improve their managerial skills and also encouraged communication with their subordinates. However, adequate support with regular formal feedback in order to facilitate the process [48], may be difficult to orchestrate in medical training where clinical educators work shift patterns. Moreover, the costs of facilitating upward feedback support may be quite high.

It is only in recent years as the internet has become widely accessible that online surveys have become more commonly utilized, hence why paper surveys were still the most commonly used form of feedback method within this review. Online surveys are cheaper and easier to administer in comparison to paper surveys and allow people to do the survey at a time that is convenient for them [55]. Scott et al. [53]’s study showed that although doctors in training did not give the highest response rates overall, trainee doctors gave the highest response rate when the survey was online. This may suggest the increasing role of online surveys in the newer generation of doctors. Furthermore, using online surveys to monitor training and trainers could allow the data to be more representative of the population of doctors in training.

Human factors in upward feedback bias

Affect describes the feeling of liking someone [56, 57]. It has been thought that affect can lead to leniency because it can prevent one’s ability to objectively and rationally evaluate someone [58]. Al-issa found that students gave higher ratings to teachers who they got along with [9]. Moreover, Antonioni and Park showed that the leniency was more profound in both peer and upward feedback compared to downward feedback [56], suggesting that affect may play a role in both peer and upward feedback. In contrast, a study by Ryan et al. [59] found that recipients of feedback were more likely to accept feedback from those who they are already acquainted to and this finding was confirmed in another study [60]. This could suggest that supervisors may be more accepting of honest feedback and this may encourage subordinates who have a positive relationship with their supervisors to give honest feedback.

Antonioni [61] found that participants who were not anonymous when they gave upward feedback did give higher ratings compared to anonymous participants. Furthermore, fewer participants stayed in the study after finding out they were in the group which could be identified [61]. However, this study was implemented within an insurance company where upward feedback could potentially be for used for summative purposes. This could lead to greater inflation in order to minimize the negative consequences. In contrast, upward feedback in medical training is more likely to be for formative purposes in order to further develop the clinical educator. Many studies have allowed upward feedback response to be confidential due to potential rating inflation [3, 4, 7, 12-15, 17, 22-24, 26, 28, 34-39, 43-45, 47-50, 52-55, 57, 58, 61-142], hence accountability and confidentiality was the most commonly acknowledged type of bias found within this systematic review. In contrast, Roch and McNall [67] that investigated whether anonymity affected ratings found that students who were not anonymous actually gave lower ratings compared to anonymous raters. Non-anonymous raters may feel more pressure to give high quality ratings [67]. So, there still may be a role for surveys in which sub-
ordinaries may be accountable for their ratings. Furthermore, supervisors seem to be more accepting of accountable surveys [61]. Unfortunately, in potentially negative situations, anonymity seems likely to be the best policy.

Reward anticipation could be related to evaluation inflation. Previous studies have found that course grades can significantly predict student ratings [7,9], but the causation is unclear. Marsh and Roche [30] found that giving high grades were not related to higher student evaluation, but instead a lot of the variation within student evaluations could be accounted for by prior subject interest, higher and challenging workloads and learning. Furthermore, Abrami et al. [6] found that student grades were unlikely to have an effect on student ratings. The relationship of reward and ratings has been inconsistent and can be subjected to interpretation, hence the need for further research in this area.

Even if confidentiality concerns are addressed, this may still affect participation due to fear and retaliation [10,12,15,61,62,132]. The miscorrelation of self-perception and upward feedback results could affect acceptability and credibility of upward feedback since it threatens self-esteem [143]. Multiple factors can affect people’s receptivity to feedback, this includes their motivation, fear and expectations [60]. However, if feedback is delivered appropriately and is perceived as valuable, then this can minimise the risk of negative emotions and dismissal of the feedback [60]. This is likely to require specialist input e.g., counselling which may have extra cost implications.

A lack of trust and cynicism was not an uncommon finding in both medical [45,53,55,58,137,142,144-147] and non-medical feedback [5,9,15-17,21,26,38-40,52,61,67,70,71,75,81,82,91,148-150]. If there is discrepancy between self-ratings and upward feedback ratings [128,145], there is a possibility that the recipient may not find the feedback credible. Also poorly designed surveys that may lack useful feedback can lead to reluctance to change. Even trainees question the credibility of some of the feedback provided by their supervisors [151], hence it is likely that supervisors may do the same to feedback from trainees. Moreover, upward feedback, especially in an undergraduate setting has been compared to ‘popularity contests’. Aleamoni [46]’s review article demonstrated that evidence supports the fact that students are able to judge the effectiveness of teaching. However, attitudes are harder to modify and this misperception may still lead to faculty being more resistant to change. This resistance could in turn affect raters’ enthusiasm, especially if previous experiences of upward feedback lead to no improvement.

**Limitations**

Although a comprehensive search was done, however, this may not be representative of all the data available on upward feedback. Also, a total of 35 articles shortlisted in the systematic review were not included in the results. There could potentially be other types of bias present in literature that was not reviewed within this systematic review. Moreover, we have identified a number of different biases that are involved in upward feedback, however we have not investigated how these biases can be minimised. Further research will be required in order to determine whether these biases are interrelated and if it is possible to minimise the effects of different biases, especially human factors.

**CONCLUSION**

Upward feedback is a multidimensional form of feedback that can lead to improvement if facilitated and implemented appropriately. This systematic review has shown that multiple different types of bias can exist within upward feedback. The established literature acknowledges and suggests likely causes of bias, without thoroughly investigating their effect on feedback quality. This highlights the importance for managers of training to consider important factors such as survey method and intended uses when designing and interpreting feedback. Currently, a mixed approach with triangulation of methods seems to be the best way to evaluate medical training. Further research is required in order to evaluate which types of bias are associated with specific survey characteristics and which factors are potentially modifiable.

**ORCID:** Anli Yue Zhou: http://orcid.org/0000-0003-2402-8289; Paul Baker: http://orcid.org/0000-0003-1493-9660.

**CONFLICT OF INTEREST**

No potential conflict of interest relevant to this article was reported.

**SUPPLEMENTARY MATERIAL**

Audio recording of the abstract.

**REFERENCES**


59. Ryan AM, Brutus S, Greguras GJ, Hakel MD. Receptivity to assessment-based feedback for management development. J Ma-


188-198411000-00003


