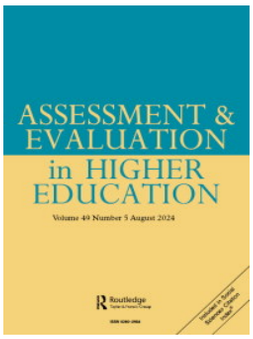

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/caeh20

What's the use of being nice? Characteristics of feedback comments that students intend to use in improving their work

David Playfoot, Ruth Horry & Aimee E. Pink

To cite this article: David Playfoot, Ruth Horry & Aimee E. Pink (08 Jul 2024): What's the use of being nice? Characteristics of feedback comments that students intend to use in improving their work, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2024.2373799

To link to this article: https://doi.org/10.1080/02602938.2024.2373799

Published online: 08 Jul 2024.

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
https://doi.org/10.1080/02602938.2024.2373799

Routledge
Taylor & Francis Group

What's the use of being nice? Characteristics of feedback comments that students intend to use in improving their work

David Playfoot, Ruth Horry and Aimee E. Pink

School of Psychology, Swansea University, Swansea, UK

ABSTRACT

Although teachers spend a lot of time striving to provide high-quality feedback, students do not always act upon it and may rate it as unhelpful. The current study attempted to determine the characteristics of feedback that influence students' intentions to use teachers' comments in their future work. Participants rated real feedback comments for clarity, tone and encouragement as well as their intention to use those comments if they had received them on their own work in three studies. Multiple regression analyses demonstrated that intention to use ratings were higher for comments which were 'nice' (supportive, encouraging, motivating, positive in tone) and 'usable' (clear, constructive, helpful). We argue that these data provide clear guidance for instructors as to how to approach the provision of written feedback on student work.

KEYWORDS

Feedback; intention to use; assessment

Feedback is an integral part of the learning process for students (Hyland 2013). Providing feedback on student work is time-consuming for teaching staff (Gibbs and Simpson 2004) and a lot of effort is expended in trying to provide high-quality feedback (e.g. Pitt and Norton 2017; Brooks et al. 2021). In spite of this, students in the UK rate feedback as one of the aspects of their university experience with which they are the least satisfied on the National Student Survey (Bell and Brooks 2017). It should be noted that the fact that students are the least satisfied with this aspect of their course does not indicate that the majority of students are *dissatisfied* just that satisfaction scores are lower than for other areas. Similar patterns are also seen in the Course Experience Questionnaire used to gauge student satisfaction in Australia (Quality Indicators for Learning and Teaching 2023). Research has also shown that students often do not act upon the feedback that they are given because they do not consider it to be useful (McGrath, Taylor, and Pychyl 2011). As a consequence, a key goal of research in recent years has been to examine what affects the likelihood that feedback comments are perceived positively by the students that receive them. The current paper outlines three studies which have attempted to determine the characteristics of comments that students believe that they would use to further develop their work.

The existing literature considers a variety of factors that may contribute to feedback being useful. There are already several excellent reviews and meta-analyses related to feedback practices (e.g. Wiliam 2018; Wisniewski, Zierer, and Hattie 2019; Van der Kleij and Lipnevich 2021; Winstone and Nash 2023). Despite this, there is still no clear consensus as to the characteristics of effective feedback – the effect sizes revealed by the meta-analyses are widely variable from

CONTACT David Playfoot d.r.playfoot@swansea.ac.uk School of Psychology, Swansea University, Swansea, SA2 8PP, UK.

paper to paper (Wiliam 2018). Part of the problem is likely to be that assessment and feedback practices are heavily constrained by the policies of the university in which a study is conducted, by the principles of assessment and programme design, and by individual differences between the students who receive the feedback (Price, Handley, and Millar 2011; Evans 2013; Ajjawi et al. 2022). This has led to claims that there can be no overarching 'gold standard' of feedback because the contextual factors are so influential (Krause-Jensen 2010). Nevertheless, we argue that there are likely to be underlying principles that apply to effective feedback; the *implementation* of these principles may be moderated by institutional influences or specific student populations, but they will provide a good foundation on which to build. In what follows, we outline the key characteristics of effective feedback identified in the literature and investigated in our own studies. To simplify our discussions, we will consider potential characteristics of feedback under umbrella categories, offering examples of the way that these categories have been operationalised in previous studies.

The first broad category can be referred to as 'usability' which subsumes feedback which is clear (Hattie and Timperley 2007; Ferguson 2011; Price, Handley, and Millar 2011; Li and De Luca 2012; Fong et al. 2018) and constructive (Lizzio and Wilson 2008; Dawson et al. 2019; Henderson et al. 2019). Ferguson's (2011) study, for example, surveyed 566 students at an Australian university and reported that good feedback should be clear and unambiguous as well as having an explicit connection to the marking criteria for the assignment. Dawson et al. (2019) study identified that clarity and constructiveness (and indeed 'usability') were key themes among their 400 student participants. This is not surprising – for feedback to be effective, it must be acted upon (Boud and Molloy 2013); and in order to act upon it, the student must understand what they are supposed to do, and it must be actionable (Ryan et al. 2021). In a similar vein, adopting the Transparency in Learning and Teaching framework (TILT e.g. Winkelmes 2023) has been shown to result in improvements across a variety of metrics of the student experience. TILT aims to make communication between students and teachers clear, and can be applied to all aspects of teaching practice (see https://tilthighered.com/tiltexamplesandresources for examples). Transparency as to why students are undertaking tasks and how their work is to be graded, has been shown to improve student perceptions of assignments and feedback, as well as improving the quality of the work (Winkelmes et al. 2015).

Both clarity and constructiveness were therefore characteristics of effective feedback that we considered in the current study. In addition, we included 'helpfulness' in our examination of feedback. This was motivated by the fact that large-scale student satisfaction metrics used to rank universities in the UK (the National Student Survey) ask final-year undergraduates to indicate whether the feedback that they received during their courses was helpful. Helpfulness could be considered as conceptually similar to constructiveness or usability, but to our knowledge there is no empirical evidence to support or refute this interpretation as it pertains to student perceptions of assignment feedback. We sought to gain this evidence as part of the current study.

A second umbrella category of feedback characteristics can be referred to as 'niceness'. A large body of research has identified that students prefer to receive feedback that is supportive (Xu and Carless 2017; Carless and Winstone 2023), encouraging (Abramowitz, O'Leary, and Rosén 1987; Lizzio and Wilson 2008), motivating (Henderson et al. 2019) and has a positive tone (Winstone et al. 2016; Dawson et al. 2019). Tone, in this case, refers to whether the feedback is framed in a positive or a negative manner. Dawson et al. (2019) demonstrated that feedback that appears overly critical can demotivate students and is unlikely to be used, while Winstone et al. (2016) reported that positively framed feedback is more likely to be acted upon. All of these characteristics of feedback reflect the fact that receiving comments on assignments can have an emotional impact on students (Weaver 2006; Parker and Winstone 2016). Ultimately, students will be less likely to engage with feedback that makes them feel demotivated (Ball et al. 2009) and thus the feedback will not achieve the desired effect.

In the current work, we asked participants to rate real feedback comments (received by other students in previous academic years) for clarity, constructiveness, helpfulness, encouragement, supportiveness, motivational value and tone to explore the interrelation of these characteristics. Previous studies have considered these factors but not all at the same time and hence it is possible that they are not distinct. Understanding the interrelations between these perceptions might help to understand some of the inconsistencies of previous findings and meta-analyses. Further, as outlined briefly above, there are multiple factors that may influence the effectiveness of a feedback comment. In all cases, though, the effectiveness of a feedback comment is contingent on the recipient engaging with the feedback and acting upon it. For this reason, we asked our participants to rate their 'intention to use' each of the feedback comments, imagining that they had received them on their own work. Previous studies have examined student preferences relating to feedback or changes in attainment following feedback (Winstone and Nash 2023). We argue that what students prefer is important information for instructors, but that preference does not guarantee that feedback will foster improvement in future assignments (Jonsson 2013). For example, feedback which is effusive is likely to be well-received but will not be likely to include the necessary information to allow the student to capitalise on what they did well or to correct what they did not (Holmes and Papageorgiou 2009). Thus, knowing the characteristics of feedback comments which are likely to be acted upon is key.

The current studies

We describe three studies in this paper, conducted in consecutive academic years, which aimed to examine the characteristics of feedback comments that students were likely to use in future assignments. These studies used the same stimuli but varied slightly in terms of the order in which ratings were given and the wording of the instructions. In the remainder of the introduction, we outline the rationale for the key differences in methodology, and for the outcome variables that were measured.

It should be noted that from here on in this paper we refer to 'useful' feedback rather than 'effective' feedback for two reasons. Firstly, effective feedback implies that it has an impact on the student who receives it, potentially changing the way that they complete their assignments and altering their grades. However, feedback may not always have a positive impact on the performance of a student – it is possible for feedback to be *useful* in that it confirms that the student should keep doing what they have been doing to maintain their current levels of performance i.e. there is not always a measurable *effect* of useful feedback comments. In other words, we consider that effective feedback is a subset of useful feedback. Secondly, in the studies we describe in this paper, participants rated comments that they did not receive on their own work and were not required to submit any sort of assignment to be graded after reading the feedback. Under these circumstances it is not possible for the effect of the feedback to be measured, but participants could suggest whether they would find a similar comment to be useful if they had received it (what we refer to as the 'intention to use'). Intention to use is the main outcome variable for the studies in this paper.

In our first study, we asked participants to rate feedback comments for their clarity, constructiveness, helpfulness, encouragement, supportiveness, motivational value and tone as if they had received them on their own summative assessments. We initially focused on summative feedback because a) it remains one of the most common feedback practices in higher education (Rand 2017), b) it is often considered as the primary means by which students may improve (Sadler 2010) and c) we used real feedback comments written by instructors at a UK university as our stimuli and these comments were given on summative work. As a consequence, the intention to use outcome variable is related to taking feedback from one assignment and applying it to another assignment. Participants were presented with a feedback comment and asked to rate it for all characteristics before moving on to the next comment.

In the second study, we made three methodological changes. We considered that the way in which we had presented the feedback rating scales might have increased the similarity between scores on different scales. In essence, by asking participants to rate all characteristics at the same time, it was possible that an intuitive 'I like this comment' appraisal would result in high scores for all characteristics. Therefore, we changed the presentation such that participants rated all comments for each characteristic in turn. We also provided a definition of what we meant by each characteristic. This was motivated by findings (e.g. Chanock 2000; Carless 2006) which have demonstrated that students and instructors do not always interpret feedback terms in the same way. The students in Chanock's (2000) study offered at least thirteen different definitions for the word 'analysis' in the context of essay feedback, some of which related to writing concisely, being original, providing facts or being relevant. Such a variety of interpretations might also be possible for terms like 'helpful' so we defined each characteristic before the participants rated the comments to ensure consistency across the participant sample. The final change for Study 2 was to add a second outcome variable. We explicitly asked participants to rate their intention to use each comment if they had received it on a draft which they could revise and resubmit (i.e. if it was a formative feedback comment). Adding this outcome measure would allow us to assess whether there was a substantive and systematic difference between what made a good feedback comment in the context of formative versus summative assessment.

In study 3, participants were again presented with definitions for the characteristics they were being asked to rate, and again we collected intention to use ratings for both formative and summative contexts. We made one methodological change versus study 2, however, which was to ask participants to rate all characteristics of a given comment at once before moving to the next comment. We made this change because a) we thought it was possible that ratings in study 2 might have been relativistic (i.e. that participants may have thought 'I'd find this comment more encouraging than the last one so I should rate it higher' rather than assessing each comment on its on merits) and b) so that we could disentangle any potential impact of the presentation method from the influence of providing the definition. In fact, the pattern of findings was the same in all three studies. The methodological alterations we have outlined above were made for sound reasons but appear to have had no influence on the responses of our participants. For the sake of simplicity, we will describe the method and results of all three studies together.

Studies 1-3 methods

Participants

We recruited participants for studies 1-3 from the undergraduate Psychology programmes at a UK university in the 19/20, 20/21, and 21/22 academic years, respectively. Participants were all students in their second year of the degree. For study 1, there were 158 participants (22 male, 135 female, 1 third gender) with a mean age of 20.36 years (SD = 3.60). In study 2, there were 116 participants (18 male, 98 female) with a mean age of 20.38 years (SD = 3.42). The third study recruited 103 participants (28 male, 70 female, 5 third gender) with a mean age of 20.52 years (SD = 2.82).

Design

All three studies used correlational designs. In each case, regression analysis was used to predict ratings of 'intention to use' a feedback comment when preparing a future assignment (i.e. using comments as summative feedback) from ratings of each of the following factors: clarity, constructiveness, helpfulness, encouragement, supportiveness, motivational value, and tone. In studies 2 and 3, we also included the 'intention to use' a feedback comment when redrafting the same

assignment (i.e. using comments as formative feedback) as an outcome variable in a separate set of analyses.

Materials

One hundred in-line feedback comments (i.e. annotations made directly on the text rather than as summary comments on the overall assignment) were selected from student assignments submitted to a UK university over the two academic years prior to data collection for study 1. The assignments from which the comments were drawn were 2000-word reports in the style of a short journal article, based on empirical data, written as part of a research methods class. The assignments had been marked by a team of staff, and the comments used as stimuli were not all written by a single marker. To select comments, assignments were randomly selected from among previous submissions and scanned for in-text feedback (i.e. comments presented as annotations within the text itself). Comments were selected if they addressed aspects of style or statistical principles that would be transferrable to another similar assignment, and were not more than three sentences in length. Comments which were specific to the topic of the particular report (e.g. comments related to misunderstanding of the theory being examined) were not eligible because they could not be easily transferred to future assignments. Once all the eligible comments had been extracted from a single student's submission, another assignment was randomly selected, and the comment extraction process began again. All comments were dissociated from the reports (i.e. the context in which the feedback was given), the identity of the students who submitted the work, and the identity of the marker, before being presented to the participants in study 1, 2, or 3. The same 100 comments were used in all three studies. Qualtrics (Qualtrics, Provo, UT) was used to present the rating scales.

Each comment was rated on a total of 8 dimensions (tone, helpfulness, constructiveness, supportiveness, clarity, encouragement, motivational value and intention to use). Ratings were given on 7-point semantic differential scales with antonyms (e.g. unhelpful – helpful; demotivating – motivating) at each end of the scale. Higher ratings indicated more positive appraisals. In each of these three studies, four stimulus lists were created such that each participant rated a subset of 40 out of the 100 comments. The lists were generated by assigning the extracted comments a number from 1 to 100, and then a random number generator was used to select five sets of 20 comments. The first set of 20 comments was designated as 'core' and was presented in all four stimulus lists. This would allow us to check that the instructions were being interpreted similarly by all participants – the ratings for the core comments ought to be roughly the same for all groups of participants. In addition to the core comments, each stimulus list contained one of the other randomly generated sets of feedback comments. Participants were randomly assigned to one of the four lists by Qualtrics once they had completed the consent form. All comments are available on the Open Science Framework repository (https://osf.io/wpuxj/?view_only=d8b2f02c41dc49cd96261197c8a7c8bb).

The instructions provided to participants in studies 1-3 were that they were to be presented with a series of feedback comments and that they were to imagine that they had received that feedback on their own work. In all three studies, the order of the presentation of the comments was in two blocks, although there was no break between them so participants would not have been aware of this. Block one presented the core comments in a random order per participant. The second block presented the additional twenty comments for that stimulus list, again in a random order per participant.

In Study 1, participants were presented with a feedback comment at the top of the screen and a list of characteristics to rate below. Once the participant had completed all eight ratings for a given comment (seven characteristics of the comment, plus 'intention to use'), they clicked 'next' and the second feedback comment appeared. This process was repeated until all 40 comments in their assigned list had been presented and rated. The order of the ratings was the same

for all participants and all comments. Definitions of feedback characteristics were not provided. In Study 2, participants were presented with a definition of the characteristic that they were going to rate at the top of the screen and their assigned list of 40 comments below. The definitions were generated by the first author of this paper and were based on dictionary definitions as applied to academic assignments (available at https://osf.io/wpuxj/?view_only=d8b2f02c41dc49cd96261197c8a7c8bb). Once they had rated all comments for a given characteristic, they clicked 'next' and the definition of the next characteristic was presented. In Study 3, participants were presented with one comment at a time at the top of the screen, and rating scales below, as in Study 1, but this time the definition of each characteristic was provided.

A final distinction between the three studies related to the outcome variables that were rated by participants. In all three studies, participants were asked to rate how likely they would be to use each comment as part of a future assignment. In studies 2 and 3 only, participants were also asked to rate how likely they would be to use each comment if they were given the opportunity to redraft the work and submit it again.

Results

In each of the three studies, the analysis process was the same: (1) we calculated average ratings for each characteristic for each comment, (2) we examined the correlations between the average ratings of the *core* comments provided by the four groups of participants to determine reliability, (3) we correlated the ratings of each characteristic with 'intention to use in future assignments' (and 'intention to use to redraft' in studies 2 and 3), and (4) we conducted regression analyses to determine which characteristics predicted the intention-to-use outcome variables. In all three studies, the pattern of findings was the same, and correlations between the ratings given for the same comments in each of the studies were high. Therefore, we combined participant ratings across all three studies and followed the same analysis steps. It is the analysis of the combined dataset that we report in full here (data for the individual studies, as well as the collated data, can be viewed here: https://osf.io/wpuxj/?view_only=d8b2f02c41dc49cd96261197c8a7c8bb). All analyses were conducted in JASP (version 0.17.2.1, JASP Team 2023).

Table 1 shows the mean and standard deviation of the average ratings for each characteristic across all 100 comments. It also contains a matrix of Pearson's r correlations between the variables. All pairwise correlations were significant at $p < .001$. However, the strength of the correlations varied considerably. The intercorrelations between clarity, helpfulness and constructiveness were all > 0.9; the intercorrelations between tone, supportiveness, encouragement and motivational value were also > .9. However, the correlations between variables across these clusters were smaller (though still strong according to Cohen 2013), ranging from 0.54 to .83. The

Table 1. Descriptive statistics and intercorrelations between variables.

			Pearson's r							
Variable	Mean rating	SD	Helpfulness	Construct	Clarity	Tone	Motivation	Support	Encourage	Future
Helpfulness	4.86	0.98	—							
Constructiveness	4.74	0.98	.98	—						
Clarity	5.02	0.95	.95	.90	—					
Tone	3.99	1.02	.63	.54	.64	—				
Motivation	4.17	1.00	.77	.70	.77	.97	—			
Supportiveness	4.33	1.01	.83	.78	.80	.93	.98	—		
Encouragement	4.29	0.98	.79	.72	.78	.96	.99	.98	—	
Future	4.92	0.86	.97	.94	.96	.63	.77	.81	.78	—
Redraft	5.36	0.76	.95	.95	.90	.45	.61	.68	.65	.94

Note: Ratings were collected on 7-point semantic differential scales. All correlations $p < .001$. Construct = constructiveness, Support = supportiveness, Encourage = encouragement, Future = intention to use as part of a future assignment, Redraft = intention to use when redrafting the same assignment.

strongest correlations between predictor and outcome variables included helpfulness ($r = 0.97$ with intention to use in future assignments, $r = 0.95$ with intention to use to redraft work).

As these correlations between predictor variables were likely to lead to multicollinearity problems if entered into a multiple regression analysis, we ran a Principal Components Analysis to identify the underlying structure of the data. However, this analysis (parallel analysis, oblimin rotation applied) revealed only a single component and would therefore be uninterpretable, at least from a pedagogical standpoint, if factor scores were generated for the regression. Therefore, we conducted a series of multiple regressions in which one of clarity/helpfulness/constructiveness was entered with one of tone/supportiveness/encouragement/motivational value until all possible pairs had been entered. The findings of the regressions predicting intention to use for a future assignment are presented in Table 2.

The two most successful models, according to the amount of variance accounted for as indicated by Adjusted R^2, contained helpfulness and tone or helpfulness and motivational value as predictors. Both of these models explained 93.8% of the variance in intention to use comments for future assignments, although helpfulness was the only significant predictor in its own right. All models (bar one) accounted for over 90% of the variance in the outcome measure. The beta coefficients for the predictor from the clarity/helpfulness/constructiveness group were substantially greater than those of the tone/supportiveness/encouragement/motivational value group in all analyses. This indicates that for students to intend to use feedback for future work, it is important that comments are usable, but the emotional component of the comments is not so integral.

Table 3 presents the findings of the multiple regressions that were conducted to predict the likelihood that participants would use feedback comments to assist in redrafting the same piece of work. These analyses are based on the data from studies 2 and 3 (combined $N = 219$). The predictors were entered following the same logic as in the analysis above. Again, all models were highly successful. The most successful model accounted for 93.8% of the variance as indicated by Adjusted R^2, and contained helpfulness and motivational value as predictors. Inspection of the beta coefficients shows that as helpfulness ratings increased, so too did the likelihood of use for redrafting work. In contrast to the previous analysis, motivational value was also a significant predictor in its own right. Interestingly, the beta coefficient here was negative – the more motivating a feedback comment was, the lower the rating for intention to use. This pattern was replicated in all models, wherein the beta coefficients for the predictor from the clarity/helpfulness/constructiveness group was positive and those of the tone/supportiveness/encouragement/motivational value group were negative. We argue that this may reflect that, in the context of redrafting work, a particularly positive comment will often indicate that that aspect of the assignment is already of a good standard and thus redrafting is not necessary.

Discussion

The studies described in this paper aimed to examine the characteristics of feedback comments that predicted that students would use them, and that students would consider them to be fair. Overall, we have shown that for students to rate feedback high on the intention to use scale, it should be clear, constructive and helpful. The emotional responses to a feedback comment, as reflected by ratings of supportiveness, encouragement, motivational value or general tone, were less important in predicting intentions to use.

Previous research has highlighted clarity and constructiveness as factors in effective feedback (Hattie and Timperley 2007; Lizzio and Wilson 2008; Ferguson 2011; Price, Handley, and Millar 2011; Li and De Luca 2012; Fong et al. 2018; Dawson et al. 2019; Henderson et al. 2019), as well as supportiveness and encouragement (Abramowitz, O'Leary, and Rosén 1987; Lizzio and Wilson 2008; Xu and Carless 2017; Carless and Winstone 2023). However, previous research has often been qualitative and/or based on what students and staff consider to be effective in general terms (e.g. Dawson et al. 2019). These approaches provide useful information but have not

Table 2. Standardized beta coefficients for regression models predicting intention to use feedback in future assignments.

	Model											
Predictors	1	2	3	4	5	6	7	8	9	10	11	12
Helpfulness	0.95***	0.96***	0.94***	0.92***								
Clarity					0.94***	0.86***	0.89***	0.90***				
Constructiveness									0.85***	0.79***	0.78***	0.78***
Tone	0.04				0.03				0.17***			
Supportiveness		0.01				0.13**				0.20***		
Encouragement			0.04				0.09				0.22***	
Motivation				0.06				0.08				0.23***
Adjusted R^2	.938***	.937***	.937***	.938***	.918***	.923***	.920***	.920***	.901***	.896***	.902***	.906***

Note: Motivation = motivational value. $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 3. Standardized beta coefficients for regression models predicting intention to use feedback in redrafting assignments.

	Model											
Predictors	1	2	3	4	5	6	7	8	9	10	11	12
Helpfulness	1.10***	1.24***	1.18***	1.18***								
Clarity					1.04***	0.98***	1.03***	1.07***				
Constructiveness									1.00***	1.06***	1.01***	1.02***
Tone	−0.24***				−0.22***				−0.09*			
Supportiveness		−0.35***				−0.09				−0.14**		
Encouragement			−0.29***				−0.16**				−0.09*	
Motivation				−0.30***				−0.21**				−0.10*
Adjusted R^2	.933***	.934***	.929***	.938**	.840***	.815***	.821***	.830**	.905***	.906***	.903***	.904***

Note: Motivation = motivational value. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

allowed for direct comparisons of the characteristics of feedback in order to determine which is the most important factor, or the potential overlap between terms. As an example, Dawson et al. (2019) asked an open question about what made feedback particularly effective. Their results grouped responses together in a logical way (e.g. nice/positive/constructive; supportive/encouraging/motivating – see their Table 5) but our own analysis has identified that 'constructive' is more closely aligned to clarity than it is to positivity. By getting our participants to rate real feedback comments for a variety of characteristics, we have been able to examine the interrelationship between them in ways that previous research has not. The correlations between clarity, constructiveness and helpfulness were all above $r = 0.90$ while the intercorrelations between encouragement, supportiveness, motivational value and tone were all above $r = .93$. This indicates that, although a variety of different terms have been used in the literature, students do not perceive them as distinct constructs – at least in relation to the comments rated in these studies. It is also interesting to note that in studies 2 and 3 we provided explicit definitions of the characteristics that participants were asked to rate but that the overall patterns of ratings did not differ from study 1 in which no definitions were provided. Therefore, the distinctions between characteristics (or lack thereof) are not attributable to participants misunderstanding what they are being asked. This is an important observation and provides additional guidance for feedback researchers as well as for teachers in their feedback practices – usability is more important than niceness. Our data also suggests that part of the variability in the findings of previous feedback studies may well be attributable to authors considering terms to be independent when they are not.

In studies 2 and 3 we asked our participants to rate the likelihood that they would use a particular feedback comment in a future assignment and the likelihood that they would use it to redraft the same piece of work. In doing so we were able to examine the characteristics that were influential in both formative and summative feedback. Our analyses showed that usability is important in both types of feedback. Niceness, however, is only a significant factor when the student would be given an opportunity to revise and resubmit their assignment and the relationship is negative – participants were less likely to use formative feedback comments that were nice. The absence of a significant effect of niceness in predicting feedback use for future assignments contradicts the reports of Dawson et al. (2019) and Winstone et al. (2016) but echoes those of Ferris (1997). The negative relationship in the analysis of formative feedback offers a potential explanation for these findings. Formative feedback comments which are particularly nice may be taken to indicate the aspects of the work that do not need to be changed before the final draft and are therefore directive. In summative feedback, nice comments will usually indicate that the learning outcome has been achieved, and this will not necessarily be transferable to a future assignment depending on the specificity of the comment. As our question did not indicate what the 'future assignment' would be, there was no guarantee that the same skill would be being assessed and hence the comment may not have been considered relevant. Nevertheless, it is evident from our data that usability is more important than niceness when writing feedback for students, at least as far as the intention to use outcome is concerned. It may be that the equivocal findings of meta-analyses in the area is due, in part, to a blurring of the distinction between formative and summative feedback in the participants in previous empirical studies or a lack of clarity from the authors of those studies that led to errors in the meta-analyses. It is also important for teachers – while it is important to provide feedback that is not overly negative, there is a possibility to be overly positive when writing comments. This will potentially result in students who are happy with the feedback that they receive, but who do not ultimately improve as a result. We considered whether the wording of a specific subset of the comments used in our studies might have led to the discrepancy between the findings of the formative and summative analyses. We do not think that is likely, because none of the comments that were included in our stimuli explicitly stated that no changes were needed, and in <10% of the comments was there a clear implication that the work could have been left entirely as it was. That

said, we acknowledge that presenting comments in isolation (i.e. separated from the work which elicited the feedback) could lead to ambiguity. For example, if a student had received several comments questioning the relevance of the points that they had made and then received 'very good point, this shows why the research is important', it might be more likely to elicit corrections to *other* parts of the work than it would if the same comment had been received in isolation. This might be more likely to influence ratings related to formative feedback than summative feedback, but it is an open empirical question that deserves investigation in future studies.

Another factor that might have influenced the results of the current studies is that the participants who rated feedback comments were not receiving the comments on their own work – the study was therefore somewhat artificial. This was a conscious decision in the design of the studies, but the potential implications deserve some consideration. The benefits of gathering ratings on hypothetical comments such as these are that the ratings are not coloured by the amount of time and effort that the student had expended on the work, or the advice that they had been given in the lead up to the assignment. It also allowed us to provide the same comments to a larger number of participants. In that regard, the findings of the current studies reflect a somewhat cleaner assessment of the wording of feedback than if the participants had rated comments on their own work. Of course, perceptions of *real* feedback would be influenced by precisely the factors that we sought to avoid. It would potentially be possible to explore whether students rated the same comments differently when attached to their own work by using the feedback we used as a bank of allowable comments during the marking process. However, this too would be artificial and would require that markers used comments that we would expect *not* to be well received or considered useful in order to confirm the findings of the current studies, when the students' academic development ought to be more important.

We acknowledge that being provided with nice feedback may be more important for some students than it is for others. Students who are high in neuroticism or anxiety, or low in self-efficacy, for example, might find receiving feedback that is not written in a positive tone detrimental to their wellbeing and this could lead to them disengaging from their studies entirely (Ball et al. 2009). We suggest that future research should examine the interplay of interindividual factors such as these with the characteristics of feedback comments.

In conclusion, the current studies have indicated that educators should prioritise usability (clarity and constructiveness) over niceness if they want their feedback to be used. We argue, however, that it appears straightforward to write comments that are considered to be nice simply by including 'good' within the message somewhere. Given that it is possible that some students may be detrimentally affected by perceptions of negativity in their feedback, it seems worth making feedback both nice *and* usable.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr David Playfoot is a Senior Lecturer in the School of Psychology at Swansea University. He has taught at university level for more than 10 years, and has applied his background in cognitive psychology of memory and language processing to studies of educational practice for most of that time.

Dr Ruth Horry is a Senior Lecturer in the School of Psychology at Swansea University. She has more than a decade of experience in teaching at undergraduate and postgraduate levels, and is interested in how cognitive processes including attention, memory and decision-making affect students' engagement with learning and feedback.

Dr Aimee E. Pink is an honorary research fellow at Swansea University and holds a joint appointment as a senior scientist at the Institute of High Performance Computing and Singapore Institute for Clinical science at the Agency

of Science, Technology and Research (A*STAR), Singapore. She is interested in understanding students' perceptions of the higher education environment including learning, feedback, and assessments.

ORCID

David Playfoot http://orcid.org/0000-0003-0855-334X
Ruth Horry http://orcid.org/0000-0003-3105-3781
Aimee E. Pink http://orcid.org/0000-0003-1516-7922

References

Abramowitz, A. J., S. G. O'Leary, and L. A. Rosén. 1987. "Reducing off-Task Behavior in the Classroom: A Comparison of Encouragement and Reprimands." *Journal of Abnormal Child Psychology* 15 (2): 153–163. doi:10.1007/BF00916345.

Ajjawi, R., F. Kent, J. Broadbent, J. Tai, M. Bearman, and D. Boud. 2022. "Feedback That Works: A Realist Review of Feedback Interventions for Written Tasks." *Studies in Higher Education* 47 (7): 1343–1356. doi:10.1080/03075079.2021.1894115.

Ball, E., H. Franks, J. Jenkins, M. McGrath, and J. Leigh. 2009. "Annotation is a Valuable Tool to Enhance Learning and Assessment in Student Essays." *Nurse Education Today* 29 (3): 284–291. doi:10.1016/j.nedt.2008.10.005.

Bell, A. R., and C. Brooks. 2017. "What Makes Students Satisfied? A Discussion and Analysis of the UK's National Student Survey." *Journal of Further and Higher Education* 42 (8): 1118–1142. doi:10.1080/0309877X.2017.1349886.

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. doi:10.1080/02602938.2012.691462.

Brooks, C., R. Burton, F. van der Kleij, A. Carroll, K. Olave, and J. Hattie. 2021. "From Fixing the Work to Improving the Learner: An Initial Evaluation of a Professional Learning Intervention Using a New Student-Centred Feedback Model." *Studies in Educational Evaluation* 68: 100943. doi:10.1016/j.tate.2021.103387.

Carless, D. 2006. "Differing Perceptions in the Feedback Process." *Studies in Higher Education* 31 (2): 219–233. doi:10.1080/03075070600572132.

Carless, D., and N. E. Winstone. 2023. "Teacher Feedback Literacy and Its Interplay with Student Feedback Literacy." *Teaching in Higher Education* 28 (1): 150–163. doi:10.1080/13562517.2020.1782372.

Chanock, K. 2000. "Comments on Essays: Do Students Understand What Tutors Write?" *Teaching in Higher Education* 5 (1): 95–105. doi:10.1080/135625100114984.

Cohen, J. 2013. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Cambridge: Academic press.

Dawson, P., M. Henderson, P. Mahoney, M. Phillips, T. Ryan, D. Boud, and E. Molloy. 2019. "What Makes for Effective Feedback: Staff and Student Perspectives." *Assessment & Evaluation in Higher Education* 44 (1): 25–36. doi:10.1080/02602938.2018.1467877.

Evans, C. 2013. "Making Sense of Assessment Feedback in Higher Education." *Review of Educational Research* 83 (1): 70–120. doi:10.3102/0034654312474350.

Ferguson, P. 2011. "Student Perceptions of Quality Feedback in Teacher Education." *Assessment & Evaluation in Higher Education* 36 (1): 51–62. doi:10.1080/02602930903197883.

Ferris, D. R. 1997. "The Influence of Teacher Commentary on Student Revision." *TESOL Quarterly* 31 (2): 315–339. doi:10.2307/3588049.

Fong, C. J., D. L. Schallert, K. M. Williams, Z. H. Williamson, J. R. Warner, S. Lin, and Y. W. Kim. 2018. "When Feedback Signals Failure but Offers Hope for Improvement: A Process Model of Constructive Criticism." *Thinking Skills and Creativity* 30: 42–53. doi:10.1016/j.tsc.2018.02.014.

Gibbs, G., and C. Simpson. 2004. "Conditions under Which Assessment Supports Students' Learning." *Learning and Teaching in Higher Education* (1), : 3–31.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.

Henderson, M., M. Phillips, T. Ryan, D. Boud, P. Dawson, E. Molloy, and P. Mahoney. 2019. "Conditions That Enable Effective Feedback." *Higher Education Research & Development* 38 (7): 1401–1416. doi:10.1080/07294360.2019.1657807.

Holmes, K., and G. Papageorgiou. 2009. "Good, Bad and Insufficient: Students' Expectations, Perceptions and Uses of Feedback." *The Journal of Hospitality Leisure Sport and Tourism* 8 (1): 85–96. doi:10.3794/johlste.81.183.

Hyland, K. 2013. "Student Perceptions of Hidden Messages in Teacher Written Feedback." *Studies in Educational Evaluation* 39 (3): 180–187. doi:10.1016/j.stueduc.2013.06.003.

JASP Team. 2023. JASP (Version 0.17.2) [Computer software].

Jonsson, A. 2013. "Facilitating Productive Use of Feedback in Higher Education." *Active Learning in Higher Education* 14 (1): 63–76. doi:10.1177/1469787412467125.

Krause-Jensen, J. 2010. "Seven Birds with One Magic Bullet: Designing Assessments That Encourage Student Participation." *Learning and Teaching* 3 (1): 51–68. doi:10.3167/latiss.2010.030104.

Li, J., and R. De Luca. 2012. "Review of Assessment Feedback." *Studies in Higher Education* 39 (2): 378–393. doi:10.1080/03075079.2012.709494.

Lizzio, A., and K. Wilson. 2008. "Feedback on Assessment: Students' Perceptions of Quality and Effectiveness." *Assessment & Evaluation in Higher Education* 33 (3): 263–275. doi:10.1080/02602930701292548.

McGrath, A. L., A. Taylor, and T. A. Pychyl. 2011. "Writing Helpful Feedback: The Influence of Feedback Type on Students' Perceptions and Writing Performance." *The Canadian Journal for the Scholarship of Teaching and Learning* 2 (2): 5. doi:10.5206/cjsotl-rcacea.2011.2.5.

Parker, M., and N. E. Winstone. 2016. "Students' Perceptions of Interventions for Supporting Their Engagement with Feedback." *Practitioner Research in Higher Education* 10 (1): 53–64.

Pitt, E., and L. Norton. 2017. "Now That's the Feedback I Want!' Students' Reactions to Feedback on Graded Work and What They Do with It." *Assessment & Evaluation in Higher Education* 42 (4): 499–516. doi:10.1080/02602938.2016.1142500.

Price, M., K. Handley, and J. Millar. 2011. "Feedback: Focusing Attention on Engagement." *Studies in Higher Education* 36 (8): 879–896. doi:10.1080/03075079.2010.483513.

Quality Indicators for Learning and Teaching. 2023. "2022 Student Experience Survey National Report." https://www.qilt.edu.au/surveys/student-experience-survey-(ses)

Rand, J. 2017. "Misunderstandings and Mismatches: The Collective Disillusionment of Written Summative Assessment Feedback." *Research in Education* 97 (1): 33–48. doi:10.1177/0034523717697519.

Ryan, T., M. Henderson, K. Ryan, and G. Kennedy. 2021. "Designing Learner-Centred Text-Based Feedback: A Rapid Review and Qualitative Synthesis." *Assessment & Evaluation in Higher Education* 46 (6): 894–912. doi:10.1080/02602938.2020.1828819.

Sadler, D. R. 2010. "Beyond Feedback: Developing Student Capability in Complex Appraisal." *Assessment & Evaluation in Higher Education* 35 (5): 535–550. doi:10.1080/02602930903541015.

Van der Kleij, F. M., and A. A. Lipnevich. 2021. "Student Perceptions of Feedback: A Critical Scoping Review and Call for Research." *Educational Assessment, Evaluation and Accountability* 33 (2): 345–373. doi:10.1007/s11092-020-09331-x.

Weaver, M. R. 2006. "Do Students Value Feedback? Student Perceptions of Tutors' Written Responses." *Assessment & Evaluation in Higher Education* 31 (3): 379–394. doi:10.1080/02602930500353061.

Wiliam, D. 2018. "Feedback: At the Heart of—but Definitely Not All of—Formative Assessment." In *The Cambridge Handbook of Instructional Feedback*, edited by A. A. Lipnevich & J. K. Smith, 3–28. Cambridge University Press. doi:10.1017/9781316832134.003.

Winkelmes, M.-A., D. Copeland, E. Jorgensen, A. Sloat, A. Smedley, P. Pizor, K. Johnson, and S. Jalene. 2015. "Benefits (Some Unexpected) of Transparently Designed Assignments." *The National Teaching & Learning Forum* 24 (4): 4–7. doi:10.1002/ntlf.30029.

Winkelmes, M. A. 2023. "Introduction to Transparency in Learning and Teaching." *Perspectives In Learning* 20 (1): 4–12. https://csuepress.columbusstate.edu/pil/vol20/iss1/2

Winstone, N. E., R. A. Nash, J. Rowntree, and R. Menezes. 2016. "What Do Students Want Most from Written Feedback Information? Distinguishing Necessities from Luxuries Using a Budgeting Methodology." *Assessment & Evaluation in Higher Education* 41 (8): 1237–1253. doi:10.1080/02602938.2015.1075956.

Winstone, N. E., and R. A. Nash. 2023. "Toward a Cohesive Psychological Science of Effective Feedback." *Educational Psychologist* 58 (3): 111–129. doi:10.1080/00461520.2023.2224444.

Wisniewski, B., K. Zierer, and J. Hattie. 2019. "The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research." *Frontiers in Psychology* 10: 3087. doi:10.3389/fpsyg.2019.03087.

Xu, Y., and D. Carless. 2017. "Only True Friends Could Be Cruelly Honest': Cognitive Scaffolding and Social-Affective Support in Teacher Feedback Literacy." *Assessment & Evaluation in Higher Education* 42 (7): 1082–1094. doi:10.1080/02602938.2016.1226759.