The Impact of Disciplinary Background and Teaching Experience on the Use of Evaluative Language in Teacher Feedback

Guangwei Hu\textsuperscript{a} (corresponding author)
Nanyang Technological University

Li Lin Choo\textsuperscript{b}
Raffles Girls’ School

\textsuperscript{a} English Language and Literature
National Institute of Education
Nanyang Technological University
1 Nanyang Walk, Singapore 637616
Tel: (65) 6790 3484; Fax: (65) 6896 9149
E-mail: guangwei.hu@nie.edu.sg

\textsuperscript{b} Raffles Girls’ School
20 Anderson Road, Singapore 259978
E-mail: lilin.choo@rgs.edu.sg
The Impact of Disciplinary Background and Teaching Experience on the Use of Evaluative Language in Teacher Feedback

Abstract
This study was designed to examine secondary teachers’ use of evaluative language resources in their qualitative written feedback on student work and factors shaping the deployment of such resources. Drawing on appraisal theory as an analytic framework for the language of evaluation, the study analyzed 84 teachers’ evaluative reports on their students’ research projects. The teachers’ use of several types of evaluative language was found to differ along the lines of disciplinary background and teaching experience. These results can be explained by disciplinarily valued dispositions, epistemological beliefs, knowledge-making practices, and varying knowledge of (in)effective feedback resulting from cumulative teaching experience. Implications are derived from these findings for developing teachers’ competence in using evaluative language effectively to provide feedback that can support and maximize students’ learning.

Keywords: appraisal theory; disciplinary background; evaluative language; teacher feedback; teaching experience

Introduction
The potential of feedback to direct and facilitate learning and development is well recognized in education (Adcroft, 2011; Guskey & Bailey, 2010; Hattie & Timperley, 2007) and other fields (Ashford, Blatt, & Walle, 2003; Finkelstein & Fishbach, 2012; Kluger & DeNisi, 1996). Recent reform initiatives in many educational systems have given much prominence to student-centered instruction and called for alternative assessments of students’ learning that can yield qualitative teacher feedback in addition to, or instead of, test scores (Evans, 2013; Hu, 2007; Hu & McKay, 2012). Teacher feedback is pivotal because it can (de)motivate students in subtle or direct ways, scaffold their learning efforts (Orrell, 2006; Prince, Handley, & Millar, 2011; Värlander, 2008), improve their performance (Shute, 2008), and provide the ‘information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is
domain knowledge, metacognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies’ (Butler & Winne, 1995, p.275). Previous research examined various aspects of teacher feedback, for example, types of feedback, informational content of feedback, effectiveness of feedback, timing of feedback, and the relationship of feedback effectiveness to a variety of learner characteristics and task variables (for comprehensive reviews, see Butler & Winne, 1995; Evans, 2013; Hattie & Timperley, 2007; Shute, 2008). Although this extensive body of research has greatly expanded our understanding of how teacher feedback can impact on learning, insufficient research attention has been paid to the medium of teacher feedback, that is, the language used to communicate teacher feedback (Brinko, 1993). Given the inseparability of form and meaning (Halliday, 1994), there is good reason to examine whether and how the language resources deployed to convey feedback mediate its effectiveness. In this article, we report on a study designed to fill this gap by investigating factors that shape teachers’ use of evaluative meanings in their feedback on student work. We also discuss the implications of our findings for developing teachers’ competence in using evaluative language effectively to provide feedback that can support and maximize students’ learning.

**Previous research on teacher feedback**

The growing literature on feedback indicates that giving effective feedback is a central concern for teachers of all subjects (Hyland, 1998). Although teachers’ written feedback has the potential to live well beyond their time with students, it may also be ignored or misunderstood by the latter (Gibbs & Simpson, 2004; Prince et al., 2011). Previous research has found that a number of features are characteristic of effective teacher feedback. First, teacher feedback can direct and facilitate learning if its content focuses on tasks, processes, and actions under students’ control rather than themselves or their personal characteristics (Brinko, 1993; Evans, 2013; Kluger & DeNisi, 1996), provides information on (in)correct responses instead of mere teacher praise (Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008), suggests strategies for improvement (Hattie & Timperley, 2007), is goal-directed (Shute, 2008) and informative about success criteria (Gibbs & Simpson, 2004), facilitates a deep understanding (Shute, 2008), enhances self-regulation (Butler & Winne, 1995; Evans, 2013), encourages students to embrace rather than avoid challenging learning tasks (Hattie & Timperley, 2007), and reinforces positive self-perceptions (Hyatt, 2005; Price et al.,
Second, feedback is likely to be effective when it is provided in such a manner that it is dialogic (Evans, 2013; Hyatt, 2005), personalized (Prince et al., 2011; Sutton, 2012), balanced in positivity and negativity (Brinko, 1993), detailed and specific (Evans, 2013; Orrell, 2006; Shute, 2008), clear and accurate (Dowden, Pittaway, Yost, & McCarthy, 2013; Hattie & Timperley, 2007), and devoid of derogatory or inflammatory language (Brinko, 1993). Third, feedback has a better chance of being acted on if it comes from a teacher who is credible (Brinko, 1993), cognizant of student strengths (Dowden et al., 2013; Orrell, 2006), supportive and non-judgmental (Brinko, 1993; Shute, 2008), empathetic and responsive to students’ affective needs (Dowden et al., 2013; Värlander, 2008), and enjoys a positive student–teacher relationship (Dowden et al., 2013). Finally, feedback is likely to improve performance, knowledge, and learning when it is directed to students who commit themselves to learning (Hattie & Timperley, 2007), are attentive and capable of self-regulation (Butler & Winne, 1995; Hattie & Timperley, 2007), receive the feedback mindfully (Shute, 2008), and act on it (Evans, 2013; Price et al., 2011).

The aforementioned features of effective teacher feedback are likely to be mediated by many factors (Evans, 2013). One such factor is a teacher’s disciplinary background (Bernstein, 1999). As Siskin (1994) pointed out, ‘secondary teachers both describe and demonstrate the distinctive vocabularies, logics, and concerns of their subject specialties in subject-specific ways’ (p.152). In other words, there are ‘distinctive subject subcultures … characterized by differing beliefs, norms, and practices’ (Grossman & Stodolsky, 1995, p.5). Such subject subcultures tend to divide along the line of “hard” and “soft” disciplines, namely, the natural sciences vs. the humanities and the social sciences (Becher & Trowler, 2001; Neumann, Parry, & Becher, 2002). Disciplinary language and epistemology mediate how teachers conceptualize teaching and learning, and their understanding of knowledge and learning is anchored in their understanding of subject matter. For example, Brandes and Seixas (1998) found that school teachers and university academics working together on an interdisciplinary curriculum integration project tended to frame their understandings and proposals around their disciplinary/subject knowledge. Grossman and Stodolsky (1995) also found that the high school teachers in their study had discipline-specific conceptions of teaching content, instructional beliefs, and frames of reference. In a study of Norwegian secondary teachers responding to samples of
student writing in interdisciplinary staff development projects, Øgreid and Hertzberg (2009) observed that the teachers had different, discipline-specific conceptions of what constituted effective argumentation. In a similar study, Helstad and Lund (2012) reported that an interdisciplinary team of teachers at a Norwegian upper secondary school perceived and negotiated writing issues in student texts (e.g., peer collaboration, potential plagiarism, and use of digital technologies) in different ways that could be linked to discipline-specific genres and subject cultures. Of more direct relevance to the focus of the present study, Biberman-Shalev, Sabbagh, Resh, and Kramarski (2011), McMillan (2001), and Resh (2009) examined disciplinary influences on secondary teachers’ grading criteria (i.e., relative importance given to effort and academic performance) and found that teachers from different subject departments tended to attach varying importance to the criteria in their grading. Finally, in a study focused on feedback practices of university teachers from two disciplines, Fernández-Toro, Truman, and Walker (2013) noted that language teachers focused more on skills development and provided explanations in their comments on students’ assignments, whereas technology teachers tended to focus on the content of assignments and provide corrections. The studies reviewed above pointed to disciplinary background as an important influence on teachers’ pedagogical thinking and practice. With the exception of Fernández-Toro et al., however, they did not examine potential disciplinary effects on qualitative teacher feedback per se.

Another potential mediator of the characteristics of effective teacher feedback is teaching experience. Extant research has uncovered important differences between beginning and experienced teachers in pedagogical practices, conceptions of subject matter, knowledge of students, instructional planning, interaction with students, and monitoring of their progress. Schempp, Tan, Manross, and Fincher (1998), for example, found that experienced school teachers differed from beginning teachers in their assessment/attribution of students’ learning difficulties and in their ability to make ongoing reflective appraisals of student competence/progress and design learning activities accordingly. Similarly, Lunenberg and Korthagen (2009) pointed out that teaching experience contributes to the development of practical wisdom, namely, ‘the sensitivity for and awareness of the essentials of a particular practice situation that shape our perception of this situation, and help us find possible courses of action’ (p.227). This was borne out by Baldwin and Blackburn’s (1981) study of university
faculty at different stages of their academic career which evidenced ‘continuing faculty evolution … in areas directly related to their performance as teachers and scholars’ (p.610). Lee and Luft (2008) also found that experienced secondary science teachers developed their pedagogical content knowledge over time through extensive classroom experience, suggesting differences between experienced and beginning teachers in an area of knowledge central to teaching and feedback on student progress. In a recent study of newly qualified upper secondary teachers in Norway, Ulvik and Langørgen (2012) found that beginning teachers were perceived by their mentors and themselves to differ from experienced teachers in several important respects, including currency of knowledge, point of view, digital competence, and understanding of students. Of greater relevance to our study, Williams (2005) investigated university academics’ understanding of assessment tasks and uncovered clear differences between junior and long-serving university teachers in their interpretations of common verbs used in assessment tasks. Like disciplinary background, however, teaching experience has been under-researched as an influence on teachers’ feedback.

A further gap in the relevant literature is the paucity of research on how language choices, especially the deployment of evaluative language resources, may mediate the content of teacher feedback. Two decades ago, Brinko (1993, p.588) called for systematic research attention to ‘the impact of language on the effectiveness of feedback: the structure of sentences, the choice of words, the framework within which problems are approached, how problems are named.’ More recently, Higgins, Hartley, and Skelton (2001, p.269) pointed out that language is central to teacher feedback as ‘a unique form of communication’ and ‘a communication process.’ In a similar vein, Prince et al. (2011, p.879) argued that teacher feedback should be reconceptualized ‘as a long-term dialogic process in which all parties are engaged.’ Such engagement necessarily involves language use at its core. As they noted, well-intentioned teachers may ‘unwittingly employ language and tone that undermines’ (p.881) the intended purpose of their feedback. Despite the recognition of the need to examine language use in teacher feedback, little empirical research has focused on teachers’ use of attitudinal or evaluative language, though there is some research (e.g., Sutton, 2012) on academic language as a barrier to students’ uptake of teacher feedback. In one of the few extant studies, Hyatt (2005) examined the linguistic features (e.g., use of imperatives, pronouns, passives, and obligating
modalities such as should and must) and rhetorical functions of a corpus of lecturer comments on Master’s level assignments. In another study, Spinks (1998) considered how lexico-grammatical choices in feedback played a role in inducting university students into a disciplinary discourse community. Notably, both studies were conducted in university settings. Our extensive literature search has not located any empirical study of school teachers’ language use in feedback, let alone empirical investigations into the impact of disciplinary background and teaching experience on language use in teacher feedback.

In response to the research gaps identified above, this study focuses on linguistic choices of teacher feedback, specifically the use of evaluative language in secondary school teachers’ written comments on students’ project work. Following Martin and White (2005, p.1), we define evaluative language as various linguistic resources that are used to mark attitudes, for example, to ‘approve and disapprove, enthuse and abhor, applaud and criticize.’ In other words, it comprises ‘linguistic mechanisms for the sharing of emotions, tastes and normative assessments’ (Martin & White, 2005, p.1). Evaluative language merits research attention because it can shape the focus of teacher feedback, mediate the manner in which feedback is communicated intersubjectively, construe teachers’ attitudes and judgments, and affect students’ reception of and engagement with teacher feedback. It also bears on what Hargreaves (2000), in an interview-based study of school teachers, referred to as ‘emotional geographies of schooling,’ viz. ‘the spatial and experiential patterns of closeness and/or distance in human interactions and relationships that help create, configure and color the feelings and emotions we experience about ourselves, our world and each other’ (p.815). Thus, our focus on evaluative language allows teacher feedback to be understood as a social and socializing process that involves attitudes, beliefs, motivations, and values (Adcroft, 2011). The study aims to investigate whether and how teachers’ disciplinary background and teaching experience may impact on their use of evaluative language in their written feedback. The following research questions have been formulated to guide the study:

1) Are there differences/similarities in the use of evaluative language between secondary teachers with a hard disciplinary background and their counterparts with a soft disciplinary background?
2) Are there differences/similarities in the use of evaluative language resources between beginning and experienced secondary teachers?

Methodology

Participants

Participants in this study were 84 teachers (male = 27; female = 57) from a prestigious secondary school (students aged 12-16) in Singapore. These participants were sampled according to two criteria: teaching experience and disciplinary background. In line with previous research (e.g., Fantilli & McDougall, 2009; Fenwick, 2011), the beginning teachers (n = 33) had no more than three years of teaching experience, whereas the experienced teachers (n = 51) had at least five years of teaching experience. In operationalizing disciplinary background, we followed Bernstein’s distinction between disciplines of ‘hierarchical knowledge structures’ and ‘horizontal knowledge structures’ (1999, p.162). Hierarchical knowledge structures, typically found within the natural sciences, view scientific knowledge as cumulative, empirically grounded, value-free, and constitutive of general theories which can be verified through invariant procedures and consensual criteria. By contrast, horizontal knowledge structures, characteristic of the humanities and social sciences, see knowledge as reiterative, contextual, value-laden, argumentative, and ‘consist[ing] of a series of specialised languages with specialised modes of interrogation and criteria’ (p.162). Thus, disciplines or subjects with the same type of knowledge structure share similar conceptions of the nature of knowledge and knowledge-making practices. The distinction between hierarchical and horizontal knowledge structures corresponds with the aforementioned distinction between hard and soft disciplines (Hu & Cao, 2015; Hu & Wang, 2014). To examine the impact of disciplinary background on the use of evaluative resources in teacher feedback, we included 40 teachers from hard subject departments and 44 teachers from soft ones. The teachers with a hard disciplinary background came from departments of Math (n = 12), Science (i.e., Biology, Chemistry, and Physics; n = 23), and Physical Education (n = 5), whereas the teachers with a soft disciplinary background were from departments of Aesthetics (i.e., Home Economics and Art; n = 4), English Language and Literature (n = 18), Humanities (i.e., History and Geography; n = 10), Languages (n = 7), and Philosophy (n = 5). Table 1 presents demographic information on the participant teachers.
Table 1 Participants’ teaching experience (in years) by discipline

<table>
<thead>
<tr>
<th>Teaching experience</th>
<th>Hard discipline</th>
<th>Soft discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Beginning teachers</td>
<td>15</td>
<td>1.53</td>
</tr>
<tr>
<td>Experienced teachers</td>
<td>25</td>
<td>13.52</td>
</tr>
</tbody>
</table>

Data

The data for this study consisted of 84 evaluative reports (written in English) on students’ research projects. Each participant teacher contributed one report. The research projects were conducted as part of the Research Studies (RS) component of the school’s curriculum. Each RS project was a research effort undertaken by a small team of students to study a subject-based topic that resulted in a research paper and an oral presentation. A project team was mentored by a project advisor from one of the subject departments. The project advisor was responsible for advising the team on their learning process, the methodology adopted in the research project, the written research report, and the final presentation on the project. At the end of the project, the project advisor wrote an evaluative report which contained both grades for and qualitative feedback on different aspects of the team’s performance and product. The evaluative report was completed on a prescribed template which listed the different aspects of the RS project to be evaluated and the evaluative criteria to be used. Table 2 summarizes this information. All the project advisors were trained to use the template at benchmarking sessions so that they could develop a clear and common understanding of the standards expected of RS projects and ensure parity of grading across the various subject departments.

A composite score for each RS project was computed from the grades awarded for the criteria listed in Table 2, and a 2 (hard vs. soft disciplines) x 2 (beginning vs. experienced teachers) analysis of variance (ANOVA) was run on the composite grade scores. The ANOVA failed to find a significant main effect for disciplinary background, $F(1, 80) = 2.62, p = .11, \eta^2_p = .03$. Nor was there a significant main effect for teaching experience, $F(1, 80) = 1.44, p = .23, \eta^2_p = .02$. Furthermore, the interaction between disciplinary background and teaching experience was not significant, $F(1, 80) = 0.70, p = .41, \eta^2_p = .01$. These results ruled out discrepancy in the quality of the RS projects as an explanation for possible differences in the project advisors’ use of evaluative language in their reports.
Table 3 summarizes descriptive statistics for the data-set.

### Table 2 Assessment criteria for RS projects

<table>
<thead>
<tr>
<th>Product criteria</th>
<th>Process criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Analysis</strong></td>
<td><strong>1. Abilities</strong></td>
</tr>
<tr>
<td>a. Reasoning</td>
<td>a. Formulation of research question</td>
</tr>
<tr>
<td>b. critical thinking</td>
<td>b. Research methodology</td>
</tr>
<tr>
<td>c. Synthesis</td>
<td>c. Knowledge mapping</td>
</tr>
<tr>
<td>d. Creative thinking</td>
<td>d. Information processing</td>
</tr>
<tr>
<td>e. Logical induction</td>
<td>e. Resourcefulness</td>
</tr>
<tr>
<td>f. Reflection</td>
<td>f. Time management</td>
</tr>
<tr>
<td><strong>2. Evidence</strong></td>
<td><strong>2. Attitudes</strong></td>
</tr>
<tr>
<td>a. Breadth and depth</td>
<td>a. Task commitment</td>
</tr>
<tr>
<td>b. Organization: sequence/transition</td>
<td>b. Effort/perseverance</td>
</tr>
<tr>
<td>c. Originality</td>
<td>c. Pride in work/attention to detail</td>
</tr>
<tr>
<td><strong>3. Presentation</strong></td>
<td><strong>3. Collaboration</strong></td>
</tr>
<tr>
<td>a. Language use: appropriateness of vocabulary, style &amp; tone</td>
<td>a. Contribution</td>
</tr>
<tr>
<td>b. Audience awareness</td>
<td>b. Interpersonal skills</td>
</tr>
<tr>
<td>c. Neatness/visual appeal</td>
<td>c. Responsibility</td>
</tr>
<tr>
<td>d. Clarity &amp; effectiveness of presentation</td>
<td>d. Accountability</td>
</tr>
<tr>
<td></td>
<td>e. Cooperation</td>
</tr>
</tbody>
</table>

### Table 3 Descriptive statistics by disciplinary background and teaching experience

<table>
<thead>
<tr>
<th>Teaching experience</th>
<th>Hard discipline</th>
<th>Soft discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Words</td>
</tr>
<tr>
<td>Beginning teachers</td>
<td>15</td>
<td>3,424</td>
</tr>
<tr>
<td>Experienced teachers</td>
<td>25</td>
<td>5,411</td>
</tr>
</tbody>
</table>

### Analytic framework

To address our research questions, we draw on appraisal theory as our analytic framework. Appraisal theory, most thoroughly explicated in Martin and White (2005), is grounded in systemic functional linguistics (Halliday, 1994) and is concerned with the language of evaluation, that is, various linguistic resources for construing interpersonal meaning. Such linguistic resources can be used to mark ‘the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate’ (p.1). Appraisal theory divides evaluative language resources into three semantic domains: attitude, engagement, and gradation. Attitude is concerned with language resources that are used to construe ‘emotional reactions,
judgements of behaviour and evaluation of things’ (p.35; e.g., I am disappointed; He has reacted disconcertingly; The report is well organized). Engagement comprises language resources for dialogistic positioning, that is, resources such as concession, modality, polarity, and projection that source propositions (e.g., according to Halliday) and position voices with respect to certain viewpoints or opinions in discourse (e.g., I can’t agree more with his argument). Gradation deals with language resources for scaling the interpersonal force of an evaluation (e.g., a highly contentious view) or modulating the sharpness of boundaries between categories (e.g., a real friend). Because our study focuses exclusively on resources for construing teachers’ attitudes in their feedback, the rest of this section outlines only the attitude component of appraisal theory. Where available, examples are taken from the corpus of teacher feedback constructed for this study to illustrate the use of various evaluative language resources to construe stances. These examples are tagged with codes, for example, H/E10 and S/B4, where the first letter stands for a teacher’s disciplinary background (H = hard discipline; S = soft discipline), the second letter refers to his/her level of teaching experience (B = beginning teacher; E = experienced teacher), and the number distinguishes him/her from other teachers in the group.

In appraisal theory, attitude consists of three sub-systems: affect, judgment, and appreciation. Figure 1 presents the three sub-systems with their various categories. Affect is concerned with evaluative resources for registering positive and negative emotional reactions. Four broad categories of emotions are distinguished. Dis/inclination communicates a desire for something (Example 1) or a mental process of avoidance (Example 2). Un/happiness covers positive emotions of cheer and affection (Example 3) as well as negative emotional reactions such as misery and antipathy (Example 4). In/security is comprised of positive feelings of confidence, peace, and trust (Example 5) as well as negative emotions of anxiety, disquiet, and surprise (Example 6), that is, ‘emotions concerned with ecosocial well-being’ (Martin & White, 2005, p.49). Finally, as exemplified in Examples 7 and 8, dis/satisfaction is concerned with ‘feelings of achievement and frustration in relation to the activities we are engaged in, including our roles as both participants and spectators’ (p.50).

1) I’d like to see them more focused. (H/E24)

2) She is wary of potential setbacks.
3) [He is] an optimist…. (S/E16)

4) The group felt dejected because of the harsh criticism.

5) The team members are comfortable with their working relations.

6) He is constantly anxious in social situations.

7) Team’s effort was laudable. (S/B10)

8) The presence of many sentences which contained typographical errors was disturbing. (S/B10)

Figure 1 Categories of attitude (based on Martin & White, 2005)

The sub-system of judgment is concerned with attitudes toward human behavior and dispositions, ‘which we admire or criticise, praise or condemn’ (Martin & White, 2005, p.42), and encompasses five types of evaluation in relation to institutionalized norms: normality, capacity, tenacity, veracity,
and propriety. Normality covers positive (Example 9) and negative (Example 10) judgments of how customary/special someone or his/her behavior is. Capacity deals with how capable/competent (Example 11) or incapable/incompetent (Example 12) someone is. Tenacity involves positive (Example 13) and negative (Example 14) evaluations of how dependable/resolute/persevering a person’s behavior or psychological disposition is. Veracity encompasses positive (Example 15) and negative (Example 16) judgments of how truthful/honest someone or his/her behavior is. Propriety covers positive (Example 17) and negative (Example 18) judgments of morality/legality, that is, how ethical or reproachable someone’s behavior is.

9) [The group] did not wish to be conventional. (H/E6)

10) Expectedly, the team’s diverse personalities, talents, abilities and personal commitments clashed and quite often, this impeded progress. (H/E4)

11) [The team] was able to think outside the box. (S/E16)

12) … the project … could have been better showcased with good time management. (S/B3)

13) Nevertheless, the team learnt to work beyond these limitations and persevered together as best as they could to rise to the challenge. (H/E4)

14) The group members could have, however, been more independent and taken more initiative during the whole process. (S/E21)

15) The team’s reflections on the product, the processes and their roles were honest. (H/B7)

16) His apparent naivety is a deceitful disguise.

17) The group was consistently punctual in submitting their work. (S/E20)

18) [The report] demonstrated a lack of carefulness/lack of thoughtfulness on the project team’s part. (S/B10)

The last sub-system of attitude – appreciation – comprises positive and negative evaluations of objects, products, processes, states of affairs, and entities rather than human behavior. It can be further categorized into reaction, composition, and valuation. Reaction encompasses positive (Example 19) and negative (Example 20) assessments of the impact or quality of an object, product, process, etc. Composition involves positive (Example 21) and negative (Example 22) assessments of an artifact, object, product or process in terms of makeup, balance, and complexity. Finally, valuation deals with
positive and negative evaluations of processes, products, and artifacts in terms of their worth and in
relation to current social values, as demonstrated by Examples 23 and 24.

19) The portfolio was visually interesting. (S/E1)

20) [Something could have been done] so that it would have been easier for the audience to
distinguish between the two clearly. (H/E9)

21) Main ideas were communicated in a logical sequence. (H/E6)

22) The project report was made up of disjointed parts. (S/B10)

23) Some appropriate graphics and illustrations enhanced the quality of presentation. (S/E22)

24) The project did not contribute as much to the existing stock of knowledge. (H/B7)

Data coding and analysis

The 84 evaluative reports from the participant teachers were imported into the UAM CorpusTool
(version 2.8.7), a free software program for annotating text corpora at multiple levels (O’Donnell,
2011). They were then coded manually by the second author using the analytic framework of attitude
and its sub-systems presented earlier. In coding the reports for their use of evaluative language
resources, the context of use as well as the object of appraisal was taken into account. As pointed out
by Martin and White (2005, p.52), ‘when it comes to language use in context, it is often the case that a
given lexical item will vary its attitudinal meaning according to that context.’ To establish coding
reliability, a graduate student who was familiar with appraisal theory and used it in his own
dissertation study was involved in coding a portion of the data. He was given a training session and
coded four reports for practice. Discrepancies between his and the second author’s coding results were
resolved through discussion, and their understandings of the coding scheme were standardized. The
two coders then coded six randomly selected reports (i.e., 7% of the data) independently, and inter-
coder agreement was assessed with Cohen’s kappa. The kappa statistic obtained was .63, indicating
good inter-coder reliability according to Landis and Koch’s (1977) criteria.

To determine possible effects of disciplinary background and teaching experience on language
use in teacher feedback, a series of 2 (hard vs. soft disciplinary background) x 2 (beginning vs.
experienced teachers) ANOVAs were run on the incidence of the various types of evaluation
represented in the attitude component of appraisal theory. To control for the varying lengths of the
teachers’ reports, the raw frequencies were normalized and are reported as frequencies per 1,000 words. Because the frequencies for several types of evaluation were too low to allow reliable statistical analyses, it was decided that an ANOVA would be run only when at least one group mean exceeded 1.00. Consequently, the following types of evaluation were analyzed statistically: (a) satisfaction, (b) positive/negative capacity, (c) positive/negative tenacity, (d) positive/negative reaction, (e) positive/negative composition, and (f) positive/negative valuation. The statistical analyses were conducted with IBM-SPSS (21.0), and the alpha was set at .05 (2-tailed) for all the statistical tests.

Findings

Table 4 presents descriptive statistics for all types of evaluation on which ANOVAs were run. Two patterns stand out. First, frequencies of use vary greatly across the various types of evaluation represented in Figure 1. While evaluative resources for affect were rarely used, those for judgment were used very frequently. Second, the standard deviations are relatively large, indicating considerable intra-group variation. The following presentation on the results of the quantitative analyses is organized around the sub-systems of attitude, namely, affect, judgment, and appreciation.

Table 4 Descriptive statistics (per 1,000 words)

<table>
<thead>
<tr>
<th>Sub-system</th>
<th>Category</th>
<th>Hard discipline</th>
<th></th>
<th></th>
<th></th>
<th>Soft discipline</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Beginning</td>
<td>Experienced</td>
<td></td>
<td></td>
<td>Beginning</td>
<td>Experienced</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Affect</td>
<td>Satisfaction +</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.88</td>
<td>2.95</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Capacity −</td>
<td>10.52</td>
<td>6.99</td>
<td>6.59</td>
<td>6.06</td>
<td>8.58</td>
<td>4.66</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td>Tenacity +</td>
<td>21.77</td>
<td>12.44</td>
<td>20.82</td>
<td>13.31</td>
<td>15.68</td>
<td>12.16</td>
<td>16.06</td>
</tr>
<tr>
<td></td>
<td>Tenacity −</td>
<td>3.14</td>
<td>3.36</td>
<td>6.58</td>
<td>7.16</td>
<td>7.46</td>
<td>6.97</td>
<td>5.62</td>
</tr>
<tr>
<td>Appreciation</td>
<td>Reaction +</td>
<td>6.93</td>
<td>4.80</td>
<td>8.94</td>
<td>5.24</td>
<td>7.54</td>
<td>5.88</td>
<td>7.35</td>
</tr>
<tr>
<td></td>
<td>Reaction −</td>
<td>0.23</td>
<td>0.87</td>
<td>0.52</td>
<td>1.86</td>
<td>1.12</td>
<td>2.70</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Composition +</td>
<td>11.90</td>
<td>9.05</td>
<td>16.00</td>
<td>10.99</td>
<td>9.54</td>
<td>11.74</td>
<td>12.83</td>
</tr>
<tr>
<td></td>
<td>Composition −</td>
<td>4.48</td>
<td>4.94</td>
<td>3.95</td>
<td>6.18</td>
<td>8.15</td>
<td>9.33</td>
<td>5.19</td>
</tr>
<tr>
<td></td>
<td>Valuation +</td>
<td>3.76</td>
<td>4.15</td>
<td>9.63</td>
<td>6.93</td>
<td>6.31</td>
<td>7.44</td>
<td>9.16</td>
</tr>
<tr>
<td></td>
<td>Valuation −</td>
<td>3.14</td>
<td>5.82</td>
<td>2.73</td>
<td>3.86</td>
<td>3.28</td>
<td>4.38</td>
<td>4.79</td>
</tr>
</tbody>
</table>

Note. + = positive; − = negative
Affect

Of all types of evaluation in the affect sub-system, only instances of satisfaction occurred with some frequency to warrant the use of a statistical test. The ANOVA found a significant main effect of disciplinary background, $F(1, 80) = 10.66, p = .002, \eta_p^2 = .118$, with the teachers from the soft disciplines ($M = 0.96, SD = 2.19$) using more instances of satisfaction than the teachers from the hard disciplines ($M = 0, SD = 0$). The obtained effect size indicated that disciplinary background accounted for nearly 12% of the variance in the use of satisfaction, exceeding the criterion suggested by Cohen (1988) for a medium effect (i.e., $\eta_p^2 = .059$) and approaching the criterial value for a large effect (i.e., $\eta_p^2 = .138$). The ANOVA also found a significant main effect of teaching experience, $F(1, 80) = 5.24, p = .025, \eta_p^2 = .061$, with the beginning teachers ($M = 1.02, SD = 2.35$) expressing satisfaction more frequently than the experienced teachers ($M = 0.17, SD = 0.85$). There was also a significant interaction between disciplinary background and teaching experience, $F(1, 80) = 5.24, p = .025, \eta_p^2 = .061$. As can be seen from Figure 2, the teachers from the hard disciplines abstained from using satisfaction as an evaluative resource, whereas the beginning teachers ($M = 1.88, SD = 2.95$) indicated satisfaction more frequently than the experienced teachers ($M = 0.33, SD = 1.18$) within the soft disciplines.

![Figure 2](image-url)  
*Figure 2* Mean frequency of satisfaction by disciplinary background and teaching experience
Judgment

The ANOVA run on positive capacity revealed no significant main effect of disciplinary background, $F(1, 80) = 2.54, p = .115, \eta_p^2 = .031$. The frequencies of positive capacity were not markedly different in the evaluative reports written by the teachers from the soft disciplines ($M = 21.81, SD = 11.66$) and by their counterparts from the hard disciplines ($M = 26.95, SD = 14.99$). Nor was a significant main effect identified for teaching experience, $F(1, 80) = 0.18, p = .675, \eta_p^2 = .002$. The beginning teachers ($M = 23.43, SD = 11.11$) marked positive capacity as frequently as the experienced teachers ($M = 24.79, SD = 14.95$). Furthermore, no significant interaction was found between disciplinary background and teaching experience, $F(1, 80) = .19, p = .666, \eta_p^2 = .002$.

The ANOVA for negative capacity detected no significant main effect of disciplinary background, $F(1, 80) = 0.17, p = .678, \eta_p^2 = .002$, indicating a similar incidence of negative capacity in the evaluative reports of the teachers from the soft disciplines ($M = 7.89, SD = 5.46$) and those of their counterparts from the hard disciplines ($M = 8.06, SD = 6.62$). However, the ANOVA yielded a main effect of teaching experience approaching statistical significance, $F(1, 80) = 3.65, p = .059, \eta_p^2 = .044$, with the beginning teachers ($M = 9.46, SD = 5.82$) using negative capacity more frequently than the experienced teachers ($M = 7.01, SD = 5.98$). There was no significant interaction between disciplinary background and teaching experience, $F(1, 80) = 1.08, p = .303, \eta_p^2 = .013$, indicating that differences in the incidence of negative capacity between the beginning and experienced teachers remained consistent across the disciplinary backgrounds.

The ANOVA run on positive tenacity located a significant main effect of disciplinary background, $F(1, 80) = 4.03, p = .048, \eta_p^2 = .048$, with the teachers from the hard disciplines ($M = 21.17, SD = 12.84$) using positive tenacity markedly more frequently than their counterparts from the soft disciplines ($M = 15.91, SD = 11.00$). In contrast, no significant main effect of teaching experience was located, $F(1, 80) = 0.01, p = .916, \eta_p^2 = .000$. The occurrence of positive tenacity was as frequent in the beginning teachers’ evaluative reports ($M = 18.45, SD = 12.48$) as in the experienced teachers’ reports ($M = 18.39, SD = 12.02$). No significant interaction was found between disciplinary background and teaching experience, $F(1, 80) = 0.06, p = .806, \eta_p^2 = .001$, indicating that cross-disciplinary differences remained consistent across the levels of teaching experience.
The ANOVA for negative tenacity did not yield a significant main effect of disciplinary background, $F(1, 80) = 1.48, p = .227, \eta^2_p = .018$. The teachers from the soft disciplines ($M = 6.38, SD = 6.28$) and their colleagues from the hard disciplines ($M = 5.29, SD = 6.20$) used negative tenacity equally infrequently, when compared with positive tenacity. The ANOVA did not detect a significant main effect of teaching experience either, $F(1, 80) = 0.34, p = .564, \eta^2_p = .004$, showing a similar incidence of negative tenacity in the evaluative reports written by the beginning teachers ($M = 5.45, SD = 5.96$) and those penned by the experienced teachers ($M = 6.09, SD = 6.44$). Furthermore, no significant interaction was found between disciplinary background and teaching experience, $F(1, 80) = 3.64, p = .060, \eta^2_p = .044$.

**Appreciation**

The ANOVA for positive reaction revealed no significant main effect of disciplinary background, $F(1, 80) = 0.13, p = .719, \eta^2_p = .002$, indicating that the frequencies of positive reaction were comparable for the teachers from the soft disciplines ($M = 7.43, SD = 6.80$) and those from the hard disciplines ($M = 8.18, SD = 5.11$). The main effect of teaching experience was non-significant, $F(1, 80) = 0.44, p = .508, \eta^2_p = .005$, revealing that the beginning teachers ($M = 7.26, SD = 5.34$) and the experienced teachers ($M = 8.13, SD = 6.46$) were similar in their frequency of use of positive reaction. The disciplinary background/teaching experience interaction was also non-significant, $F(1, 80) = 0.65, p = .423, \eta^2_p = .008$.

As for negative reaction, the ANOVA found neither a significant main effect of disciplinary background, $F(1, 80) = 0.54, p = .465, \eta^2_p = .007$, nor a significant main effect of teaching experience, $F(1, 80) = 0.66, p = .419, \eta^2_p = .008$. In other words, the incidence of negative reaction did not vary significantly between the teachers from the hard disciplines ($M = 0.41, SD = 1.55$) and those from the soft disciplines ($M = 0.57, SD = 1.92$) or between the beginning teachers ($M = 0.71, SD = 2.10$) and the experienced teachers ($M = 0.35, SD = 1.48$). Finally, the interaction between disciplinary background and teaching experience was non-significant, $F(1, 80) = 2.44, p = .122, \eta^2_p = .030$.

The ANOVA run on positive composition detected no significant main effect of disciplinary background, $F(1, 80) = 1.38, p = .244, \eta^2_p = .017$, showing that the teachers from the hard disciplines ($M = 14.46, SD = 10.38$) were comparable to the teachers from the soft disciplines ($M = 11.48, SD = 10.12$).
10.70) in their use of positive composition. Nor was a significant main effect found for teaching experience, $F(1, 80) = 2.46, p = .121, \eta^2_p = .030$. The incidence of positive composition was similar in the evaluative reports of the beginning teachers ($M = 10.61, \text{SD} = 10.51$) and those of their more experienced colleagues ($M = 14.38, \text{SD} = 10.48$). The ANOVA did not find a significant interaction between disciplinary background and teaching experience, $F(1, 80) = 0.03, p = .866, \eta^2_p = .000$.

The ANOVA for negative composition did not find a significant main effect of disciplinary background, $F(1, 80) = 2.62, p = .109, \eta^2_p = .032$. There was no statistically reliable difference in the incidence of negative composition between the teachers from the soft disciplines ($M = 6.40, \text{SD} = 7.60$) and their counterparts from the hard disciplines ($M = 4.15, \text{SD} = 5.69$). The main effect of teaching experience was also non-significant, $F(1, 80) = 1.33, p = .253, \eta^2_p = .016$, indicating that the frequency of use did not vary markedly between the beginning teachers ($M = 6.48, \text{SD} = 7.77$) and the experienced teachers ($M = 4.58, \text{SD} = 6.08$). No significant interaction was found between disciplinary background and teaching experience, $F(1, 80) = 0.64, p = .425, \eta^2_p = .008$.

As regards positive valuation, the ANOVA yielded no significant main effect of disciplinary background, $F(1, 80) = 0.51, p = .478, \eta^2_p = .006$, indicating similar frequencies of positive valuation between the teachers from the hard disciplines ($M = 7.43, \text{SD} = 6.63$) and those from the soft disciplines ($M = 7.99, \text{SD} = 6.95$). However, teaching experience was found to have a significant main effect, $F(1, 80) = 8.96, p = .004, \eta^2_p = .101$, with the experienced teachers ($M = 9.39, \text{SD} = 6.64$) using positive valuation more frequently than the beginning teachers ($M = 5.15, \text{SD} = 6.21$). There was no significant interaction between disciplinary background and teaching experience, $F(1, 80) = 1.07, p = .304, \eta^2_p = .013$, indicating that the difference between the beginning and experienced teachers remained consistent across the hard and soft disciplines.

Finally, the ANOVA for negative valuation found no significant main effect of disciplinary background, $F(1, 80) = 0.80, p = .373, \eta^2_p = .010$, revealing comparable frequencies of negative valuation between the teachers from the soft disciplines ($M = 4.17, \text{SD} = 6.08$) and their colleagues from the hard disciplines ($M = 2.89, \text{SD} = 4.62$). Nor was the main effect of teaching experience significant, $F(1, 80) = 0.21, p = .651, \eta^2_p = .003$. The beginning teachers ($M = 3.21, \text{SD} = 5.00$) were similar to their more experienced colleagues ($M = 3.78, \text{SD} = 5.75$) in the use of negative valuation.
The interaction between disciplinary background and teaching experience was also non-significant, 
\[ F(1, 80) = 0.61, p = .436, \eta^2_p = .008. \]

**Discussion**

The results presented in the previous section indicated both differences and similarities in the use of various types of evaluation in the evaluative reports written by teachers from different disciplinary backgrounds and with varying teaching experience. The teachers from the hard and the soft disciplines were similar in the use of positive/negative capacity, negative tenacity, positive/negative reaction, positive/negative composition, and positive/negative valuation. However, the teachers from the hard discipline made no use of satisfaction but more frequent use of positive tenacity than their counterparts from the soft disciplines. As regards the comparisons based on teaching experience, the beginning teachers did not differ from their more experienced colleagues in the use of positive capacity, positive/negative tenacity, positive/negative reaction, positive/negative composition, or negative valuation. However, they used satisfaction and negative capacity more frequently than the experienced teachers, whereas the reverse pattern was found for positive valuation. In addition, disciplinary background and teaching experience were found to interact in the use of satisfaction. These results are discussed below in relation to our research questions presented earlier.

**Does disciplinary background have an influence on teachers’ use of evaluative language?**

The observed differences in the use of evaluative language between teachers from different disciplinary backgrounds are consistent with the distinction between hard disciplines (i.e., hierarchical knowledge structures) and soft ones (i.e., horizontal knowledge structures) established by previous research into disciplinarity (e.g., Becher & Trowler, 2001; Bernstein, 1999; Jones, 2011; Stodolsky & Grossman, 1995) and could be explained in terms of disciplinary characteristics concerning core knowledge, epistemology, valued dispositions, and knowledge-making practice. As Becher and Trowler (2001, p.36) point out, soft disciplines such as humanities and pure social sciences are ‘concerned with particulars, qualities, complication,’ tend to be ‘personal, value-laden,’ and aim to develop personal understanding and interpretation. Consequently, teachers from soft disciplines are likely to communicate their personal reaction or interpretation by revealing their satisfaction or lack of it with students’ subject-based project work (see Examples 7 and 8 presented earlier). Hard
disciplines, by contrast, are ‘concerned with universals, quantities, simplification,’ strive to be ‘impersonal, value-free,’ and seek discovery and explanation. Given such disciplinary characteristics, it is natural for teachers from hard disciplines to be impersonal in their feedback and focus on content of work, hard knowledge involved, and criteria of performance rather than personal reactions that find expression in evaluative language resources for marking satisfaction or dissatisfaction. This finding is consistent with previous research into teacher student communication that highlighted a transmission perspective in hard disciplines such as physics and chemistry, as compared to an interpretation perspective in the soft disciplines (Barnes & Shemilt, 1974; Grossman & Stodolsky, 1995). It is also broadly consistent with research that has found teachers’ views and assessments of student writing to be influenced by discipline-specific conceptualizations of effective academic writing (e.g., Helstad & Lund, 2012; Lea & Street, 2000; Øgreid & Hertzberg, 2009).

Similarly, the more frequent use of evaluative language to mark positive tenacity in the feedback provided by the teachers from the hard disciplines, as illustrated by Example 13, could also be accounted for in terms of disciplinary practices, epistemology, and values. According to Neumann et al. (2002), hard disciplines are typified as having knowledge communities that are competitive yet collaborative. Knowledge-making in these disciplines involve effort, perseverance, and meticulousness, all of which are manifestations of positive tenacity. By contrast, research activities in soft disciplines ‘are less competitive and less demanding in commitment’ (p.411) but place a premium on lateral thinking, fluent expression, critical facility, and originality of opinion. These disciplinary practices and values have also made their way into the classroom (Orrell, 2006). As pointed out by Nilson (2010), students in hard disciplines:

are supposed to apply hard facts and reliable data to a problem-solving situation, consider possible outcomes, hypothesize the most reasonable prediction, perform a tightly controlled experiment to test the hypothesis, measure the results meticulously, and come to probable, carefully qualified conclusions based on the resulting evidence. (p.225)

Such learning tasks require students to be careful, meticulous, thorough, resolute, and persevering. Consequently, it is understandable that teachers in hard disciplines value positive tenacity in their evaluation of student work more than their colleagues in soft disciplines who place a higher value on
Does teaching experience have an influence on teachers’ use of evaluative language?

Apart from disciplinary influences, the results reported earlier also suggest the impact of teaching experience on teachers’ use of evaluative language in their feedback on student work. There are at least two ways in which teaching experience could have contributed to the greater use of satisfaction by the beginning teachers (see Examples 7 and 8), as compared to the experienced teachers. First, the beginning teachers and their more experienced colleagues were likely to be attuned to the normative emotional geographies of secondary teaching to different extents. In his interview study of teachers’ emotional geographies reviewed earlier, Hargreaves (2000) noted that compared with primary teaching, ‘secondary teaching is characterized by greater professional and physical distance leading teachers to treat emotions as intrusions in the classroom’ (p.811). He found that the secondary teachers in his sample seemed ‘more concerned to fend off and manage negative emotion that threatens to intrude from the outside, rather than develop positive emotions in their own right’ (p.823). It is plausible to suggest that the experienced teachers in our study were more cognizant of such normative patterns of emotionality than the beginning teachers and consequently were less likely to openly show their affective response, for example, their satisfaction or dissatisfaction with student work.

Another possible explanation of the difference in the use of satisfaction between the beginning and experienced teachers may lie in varying knowledge of ineffective feedback as a result of cumulative teaching experience. Several state-of-the-art reviews and meta-analyses (Black & Wiliam, 1998; Evans, 2013; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008) of research on feedback have found that positive feedback and rewards, especially teacher praise, have little effect on students’ learning and performance. Notably, positive feedback and teacher praise overlap conceptually with what is characterized as satisfaction in our study. Factors potentially contributing to the ineffectiveness of such feedback are various, including a perceived lack of learning- or task-related information in such feedback (Hattie & Timperley, 2007), the deflection of attention away
from the learning task to the self (Kluger & DeNisi, 1996; Shute, 2008), recipients’ counterproductive self-evaluations of their own ability (Hattie & Timperley, 2007), inflated perceptions of success (Black & Wiliam, 1998), and avoidance of challenging tasks that may pose risks to the self (Hattie & Timperley, 2007). Contrary to previous beliefs, there is some research evidence that ‘the most effective teachers actually praise less than average’ (Black & Wiliam, 1998, p.49). Like those effective teachers, it can be argued, the experienced teachers in our study had experiential knowledge that feedback expressing teacher satisfaction was ineffective in improving student learning; consequently, they expressed satisfaction less frequently than their inexperienced colleagues, who might see expression of teacher satisfaction as a useful means of motivating their students.

Varying knowledge of what would constitute ineffective feedback could also explain why the experienced teachers as a group used negative capacity less frequently than the beginning teachers. Like other types of judgment, negative capacity construes ‘attitudes to people and the way they behave’ (Martin & White, 2005, p.52). More specifically, it conveys negative social esteem and concerns how incapable, incompetent, unproductive, or unsuccessful someone is (see Example 14). Thus, feedback that contains negative capacity draws attention to the self and threatens self-esteem. Feedback research has consistently demonstrated the ineffectiveness and even the detriment of feedback that is uninformative about the task (Hattie & Timperley, 2007), directs attention to the self (Black & Wiliam, 1998), invokes negative emotional responses (Brinko, 1993; Evans, 2013), attributes failure to ability (Kluger & DeNisi, 1996), demotivates students (Gibbs & Simpson, 2004; Shute, 2008), undermines self-efficacy (Kluger & DeNisi, 1996; Shute, 2008), and leads to ‘self-handicapping, learned hopelessness, or social comparison’ (Hattie & Timperley, 2007, p.97). It is very likely that the experienced teachers had learned from their extensive teaching experience and contact with students that feedback encoding negative capacity may be subject to all these problems and do little to enhance student learning. As a result, they expressed negative capacity less frequently than their less experienced colleagues. This explanation is consistent with Schempp et al.’s (1998) finding that teaching experience helps in appraising students’ progress and attributing their learning difficulties. It is also supported by the observed tendency of the experienced teachers to couple negative capacity with positive tenacity to mitigate the former’s negative impact, as exemplified
The group seemed a little unsure about the focus of the project initially [negative capacity], although they exhibited commendable resourcefulness getting the school estate management to work with them [positive tenacity]. (H/E25)

Finally, the observed difference in the use of positive valuation between the beginning teachers and their experienced colleagues could plausibly be attributed to varying knowledge about effective feedback that accumulates as a function of teaching experience. As summarized in the earlier review of the extant research, there is ample evidence that feedback is likely to be effective or acted on when it addresses tasks, processes, performances, and products that are under students’ control (Brinko, 1993; Evans, 2013; Hattie & Timperley, 2007), provides information on success criteria (Gibbs & Simpson, 2004; Shute, 2008), and encourages self-regulation (Butler & Winne, 1995; Evans, 2013). These are precisely what positive valuation is concerned with (see Example 23). As a type of appreciation, positive valuation encodes esthetics, values, and criteria upheld in a given field and drawn on for ‘construing our evaluations of ‘things’, especially things we make and performances we give’ (Martin & White, 2005, p.56). Evaluating processes and products (in this case, students’ research projects and written reports) against disciplinarily and educationally valued criteria and qualities requires skill, pedagogical knowledge, and internalization of institutional and disciplinary values. As suggested by previous research (e.g., Douglas, 2009; Lee & Luft; 2008; Lunenberg & Korthagen, 2009), these are typically areas where beginning teachers have much to learn from experienced teachers. The latter are likely to have a more confident knowledge of what is deemed worthwhile in their institutional and disciplinary contexts and, consequently, are more adept at using positive valuation to socialize their students into communal norms, values, and practices. As Wassermann (2004) aptly puts it, ‘in the hands of a skilled teacher, evaluative practices may inspire pupils to new heights of performance and guide them toward greater mastery of skills and love of subject’ (p.151).

Implications

Several implications can be derived from the findings reported above. To begin with, given the centrality of feedback in learning endeavors and the mediating role of evaluative resources in
feedback, teacher education programs need to give explicit and focused attention to the linguistic
dimension of teacher feedback and foster student teachers’ literacy in evaluative language. This
pedagogical attention can take the form of direct instruction in and guided exploration of various
evaluative language resources (e.g., those comprehensively canvassed by Martin and White, 2005) in
professional communication skills courses for student teachers that aim to enhance their knowledge of
and expand their repertoires of lexico-grammatical resources for sharing feelings, stances, and values
effectively. Second, the findings of our study indicate that teacher feedback is a carrier of
disciplinarity (see also Gibbs & Simpson, 2004) and that diverse epistemological assumptions of
different subjects underlie and are reflected in language use in teacher feedback. This suggests that
explicit work on evaluative language in generic communication skills courses, as outlined above,
should be complemented and extended by discipline-based inquiry into the use of evaluative language
in teacher feedback that subject-matter courses are well positioned to provide. Such inquiry should
aim to raise student teachers’ awareness of how discipline-specific orientations toward teaching and
learning may result in differing use of language resources in teacher–student interactions, how
evaluative language resources can be used effectively to address subject-specific concerns, and how
discipline-appropriate use of language in teacher feedback serves to induct students into disciplinary
norms, core knowledge, valued dispositions, and discipline-specific meaning-making practice. Third,
student teachers need to be quipped, through the joint pedagogical activities of communication skills
and subject-matter courses, with useful and practical strategies for engaging in their own investigation
into the disciplinarily, experientially, and socioculturally shaped nature of evaluative language use in
teacher feedback. Such strategies can be fostered by engaging student teachers in various data-driven
learning tasks, for example, a school-based project that involves interviewing students to understand
how evaluative language may mediate their responses to teacher feedback, a textual project that
requires an appraisal theory-informed analysis of a corpus of oral/written feedback collected from
experienced teachers to find out how evaluative language resources can be deployed strategically to
enhance students’ reception of and engagement with teacher feedback, and a comparative project that
collects effective teacher feedback from different disciplines to identify cross-disciplinary similarities
and differences in the use of evaluative language. In conclusion, research-informed and well-
coordinated pedagogical efforts are needed to help student teachers make sense of the role of evaluative language in teacher feedback and develop their competence in using various evaluative resources appropriately to provide effective feedback that can support and maximize student learning.

References


Lawrence Erlbaum.


