Analysis of New Zealand primary and secondary student peer- and self-assessment comments: Applying Hattie and Timperley's feedback model

Lois R. Harris
Central Queensland University
Gavin T. L. Brown & Jennifer A. Harnett
The University of Auckland

Abstract

Peer- and self-assessment (PASA) can lead to increased student self-regulation and achievement. However, few studies have examined the content of the feedback students in primary and secondary schools provide themselves and their peers. This study used Hattie and Timperley's (2007) task, process, self-regulation, and self feedback categories to (1) examine the content of a sample of naturally occurring student written PASA comments (n = 471 utterances) and (2) explore the feasibility of using this model with student generated feedback. Students provided primarily task feedback to both themselves and their peers, with self-regulation feedback only found in self-assessment. Students in higher grades tended to provide more task and process feedback, while giving less self feedback during self-assessment and more during peer assessment. More refinement of the model is recommended for both research and professional development purpose to better capture the quality and complexity of student-led feedback comments.

Keywords: Formative assessment; self-assessment, peer-assessment, feedback, content analysis, Hattie and Timperley feedback model

Correspondence should be addressed to Dr Lois Harris, Central Queensland University, Australia or by email to l.irvin@cqu.edu.au

Accepted for publication; recommended citation

Harris, L. R., Brown, G. T. L., & Harnett, J. (2014, accepted). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: Applying Hattie & Timperley's feedback model. *Assessment in Education: Principles, Policy and Practice.*

Introduction

The global Assessment for Learning movement (Assessment Reform Group, 2002; Berry, 2011; Black & Wiliam, 1998) has encouraged educators to foreground the role of assessment in improving teaching and learning, prioritizing this above its traditional accountability functions. Within Assessment for Learning, there is a strong emphasis placed on students' active involvement in assessment, especially through processes like peer- and self-assessment (e.g., Black, Harrison, Lee, Marshall, & Wiliam, 2003; Leahy, Lyon, Thompson, & Wiliam, 2005). Peer- and self-assessment (PASA) allows students to participate in the assessment and evaluation process, which is an important part of the self-reflection phase of self-regulation within the cyclical model of self-regulated learning (Zimmerman, 2008). Zimmerman's (2008) model of self-regulation begins with a forethought phase where students analyse the task and activate self-motivational beliefs (i.e., 'where am I going?'). Subsequently, during the performance phase, students practice self-control and self-observation (i.e., 'how am I going?'). In the final self-reflection phase, students self-evaluate their work (i.e., 'where to next?').

PASA also gives students access to additional sources of feedback. Such feedback has the potential to be more accessible, timely, and plentiful than teacher generated feedback given normal student to teacher ratios (Hattie, 2009). Additionally, participating in student-led assessment practices may give students a greater sense of ownership in this process and encourage them to better understand the criteria describing quality work (Andrade, 2010), with peer assessment also potentially improving communication and collaboration (Topping, 2013).

However, while peer- and self-assessment have potential benefits, some issues are raised about their utility in helping students make progress in their learning. While some students and teachers alike question whether students, as novices, can consistently provide accurate and appropriate feedback that will lead to improvement (Cowie, 2005, 2009; Ross, 2006; Brown & Harris, 2013; Harris & Brown, 2013; Peterson & Irving, 2008; Ross, Rolheiser, & Hogaboam-Gray, 1998; Noonan & Duncan, 2005; Lee, 2007), at least within New Zealand, students do appear to accept these practices as a legitimate form of feedback (Harris, Brown, & Harnett, 2014). However, there has been little examination of the actual content of student generated feedback and how its messages compare to those created by teachers to determine how useful such feedback may be for improving achievement.

There are numerous models of feedback available in the literature (e.g., Shute, 2008; Tunstall & Gipps, 1996; Kluger & DeNisi, 1996); while it would be highly desirable to exemplify how multiple models could be applied to peer- and self-assessment data, that goal was outside of the scope of this paper. Hattie and Timperley's (2007) systematic meta-analysis of 12 previous meta-analyses incorporated 196 studies and nearly 7000 effect sizes and concluded that feedback had a powerful effect on learning outcomes. In addition to the sheer scale of Hattie and Timperley's (2007) analysis, their model was selected for this study as it explicitly includes the self and peers as legitimate sources of feedback. Hence, within this paper, feedback is defined as:

> … information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding. (Hattie & Timperley, 2007, p. 81)

Hattie and Timperley (2007) identify four levels of feedback [i.e., task (i.e., whether work was correct or incorrect, descriptive comments about the substance of the work), process (i.e., comments about the processes or strategies underpinning the task), self-regulation (i.e., reminders to students about strategies they can use to improve their own work), and self (i.e., non-specific praise and comments about effort)]. From their review, they demonstrated that task, process, and self-regulation feedback all contribute to learning outcome gains, while self

feedback generally does not (Hattie & Timperley, 2007). Their explanation is that ego-enhancement and ego-protective feedback do not guide the student as to what needs to be done next to revise their understandings or improve their work.

This study sought to determine whether Hattie and Timperley's (2007) four types of feedback could be found in naturally occurring examples of New Zealand upper primary and lower secondary school student peer- and self-assessment comments and identify whether the model required any adjustments or clarifications for use with student-led feedback. This paper begins by examining what is already known about the content of student peer- and self-assessment comments. It then introduces the New Zealand context where the study took place before sharing the methods and results of the study and discussing its implications.

Examining the content of peer- and self-assessment

While peer- and self-assessment are distinct processes, they are examined together here as both are examples of student-led feedback, and, thus, have different dynamics from teacher-student feedback practices. In PASA, students are expected to take on the complex role of assessor, develop their own understanding of what counts as quality work, and navigate social-psychological factors in their self and peer relations (van Gennip, Segers, & Tillema, 2009). While there has been some exploration of PASA content in higher education settings (e.g., Topping, Smith, Swanson, & Elliot, 2000; Dochy, Segers, & Sluijsmans, 1999; Miller, 2003; Taras, 2003), few studies have examined the content of student-generated feedback within compulsory school settings. Findings from higher education studies are unlikely to generalize to compulsory school settings because higher education students are much older, have far more expertise in the subject domains they are studying, and are completing more complex academic tasks.

Examining the accuracy of PASA feedback has been a major focus of K-12 studies conducted, with issues found relating to the validity of student-generated feedback in relation to 'expert' feedback (e.g., Brown & Harris, 2013; Topping, 2013). For example, Sung, Chang, Chang, and Yu (2010) found that while low achieving students overestimated the quality of their work in self-assessment comments, high achievers underestimated its worth and students were less accurate in their assessments of group work than of individual work. Student self-assessments have also been found to become more accurate with increased age and experience (Blatchford, 1997), with young children often believing that their effort and cooperation should be considered when determining their result (Higgins, Harris, & Kuehn, 1994). Chang, Tseng, Chou, and Chen (2011) found high school students' peer-assessments using rubrics within a computer course were generally inaccurate and that student ratings did not align well with peer or teacher judgements. However, increasing the number of raters can potentially improve the validity and reliability of peer assessment judgements (Sung, Chang, Chang, and Yu, 2010).

While PASA practices are generally used in low stakes situations within compulsory education, there are still consequences for under or overestimating the quality of one's work or that of a peer (e.g., choosing not to enrol in future subjects due to perceived lack of ability, failing to invest enough time in a task because you perceive or are told you have already mastered it). Asking students to justify their comments may help as Gielen, Peeters, Dochy, Onghena, and Struyven (2010) found this led to improvement, although expecting students to reflect on this feedback did not better performance.

Additionally, even accurate feedback fails to be useful if it does not help students identify the major areas for improvement within the piece of work (e.g., persuasive essay comments which focus on spelling, while ignoring major issues with its thesis statement and supporting arguments). Low-level feedback or comments not relating to the success criteria for the task do not provide the learner with ideas about how to improve his or her performance in a substantive way. There is evidence that without training, students are

unlikely to create comments well-aligned with the goals of the task. For example, while the 7th grade students in Tsivitanidou, Zacharia, and Hovardas' (2011) study were able to create comments including grades, positive and negative judgements about the work, and suggestions for improvement without explicit training, the assessment criteria that the students developed to guide their judgements were highly varied, and their criteria for quality were highly correlated with their criteria for quantity, elaboration, and organization. Dutch studies (van Zundert, Könings, Sluijsmans, & van Merriënboer, 2012; van Zundert, Sluijsmans, Könings, & van Merriënboer, 2011) have also shown that the cognitive load and complexity of the task matter to the effectiveness and utility of the peer-assessment.

While there are clearly challenges in getting students to create effective PASA feedback, research does suggest that training and practice can lead to improvements (e.g., Gan, 2011; McDonald, 2009; McDonald & Boud, 2003; Nicolaidou, 2013), although superficial training is unlikely to generate positive effects (Andrade & Boulay, 2003). For example, Gan and Hattie (2014) found that using Hattie and Timperley's (2007) framework moved evaluations of peer feedback from dichotomous (accurate vs inaccurate) to a progressive perspective that examined the complexity of comments. With training, the secondary chemistry students in their study were able to create more self-regulation focused comments, although instances of such comments remained quite low. Nicolauidou (2011) found that 4th grade students were able to create more complex corrective peer comments on writing tasks with continued practice through an e-portfolio system. While the capability of students to offer sophisticated feedback seems to depend largely upon expertise and ability in the domain (Boud & Falchikov, 1989; Falchikov & Goldfinch, 2000), training and practice can potentially help improve the utility of their comments.

Hence, it becomes important to have a model that can be used to evaluate the levels of feedback students are generating. Hattie and Timperley's (2007) model appears promising to use by itself or in conjunction with other frameworks to help explain to students the different levels of feedback they could provide. It would certainly help move students' focus towards assignment content type instead of measures like accuracy and quantity used in other studies (e.g., Nicolauidou, 2011). Using this model would also enable educational stakeholders to better examine the presence or role of any kind of self-feedback (i.e., general praise, comments about effort/enjoyment, behaviour), which has been largely ignored. Because interpersonal variables (e.g., psychological safety, value diversity, interdependence, trust, friendship) are present within PASA practices (van Gennip. Segers, Tillema, 2009, 2010; Harris & Brown, 2013; Panadero, Romero, & Strijbos, 2013), more work is needed to establish how self-feedback may be used for non-learning purposes (i.e., building relationships, qualifying negative evaluation). While work by Gan (2011) and Gan and Hattie (2014) utilized this model successfully to examine peer-assessment content generated by upper secondary chemistry students, it has yet to be applied to PASA feedback in other subject areas or from younger students, a gap this study hopes to help fill.

New Zealand Context

New Zealand is a favourable context for examining the quality of student generated feedback. The nation's official policy on assessment privileges Assessment for Learning as the dominant approach (Ministry of Education, 2010). The policy explicitly recommends student-led assessment practices:

> Students should be encouraged and supported to be involved in all aspects of their learning including setting goals, developing success criteria and exemplars, self- and peer assessment, reflecting on their learning, identifying what they are doing well and why, and considering what they need to do next to further their learning. (Ministry of Education, 2010, p. 25)

The New Zealand assessment policy has for more than two decades expected teachers to practice interactive assessment for learning in classrooms and permits high levels of student involvement in PASA practices (Crooks, 2010). Hence, given the favourable environment, it would be expected that students would be able to produce constructive feedback beneficial for learning.

This policy priority is reflected in the beliefs of New Zealand teachers who have strongly endorsed the notion that the primary purpose of feedback is to improve learning growth rather than promote student well-being (Brown, Harris, & Harnett, 2012). The self-report survey findings were corroborated in an examination of New Zealand teachers' written feedback practices (Harris, Harnett, & Brown, 2013), which found task rather than self feedback dominated, although instances of self-regulation feedback were rare. Hence, while individual teachers have their own feedback styles, in the main it can be expected that New Zealand students do receive learning-orientated feedback at task, process, or self-regulation levels.

Nonetheless, survey studies of New Zealand secondary school students reported, in general, that they did not associate self- and peer-assessment with assessment (Brown, Irving, Peterson, & Hirschfeld, 2009) and such informal, interactive assessment practices were not associated with increased academic performance (Brown, Peterson, & Irving, 2009). However, in both studies, the informal, interactive assessment practices (including PASA) were positively associated with the notion that assessment improves classroom climate, suggesting that interpersonal relationships are a salient aspect of classroom assessment practices. While New Zealand students generally endorse PASA as potentially helpful for learning (Harris, Brown, & Harnett, 2014; Cowie, 2009), qualitative studies suggested that some students still have doubts as to the veracity and helpfulness of peer and self-feedback (Cowie, 2005; Harris & Brown, 2013; Peterson & Irving, 2008). It remains unknown the exact extent to which students mimic teacher feedback types in their own peer- and self-feedback.

Method

Participants

This study collected data from Auckland teachers and students already participating in the Measuring Teachers' Assessment Practices (MTAP) Project studies around assessment and feedback. Participating teachers had previously completed a survey questionnaire about their conceptions of assessment and/or feedback (Brown & Harris, 2009; Brown, Harris, & Harnett, 2012) and had indicated on this initial survey that they might be willing to participate further in the project. Potential participants were invited via email and all volunteers were accepted into the study. Hence, data were provided by motivated volunteer teachers who may not be representative of normal classroom practices. However, given the range of desirable and undesirable practices observed in three of the volunteer teacher classrooms (Harris & Brown, 2013), it is difficult to conclude that these participating teachers were necessarily exemplary. Data were naturally occurring feedback statements students provided to themselves or classmates. These were collected from students in the classrooms of 11 teachers of Years 5-10 (students nominally aged 10-14). See Table 1 for demographic details relating to the classroom teachers who supplied examples of PASA comments their students had generated.

Table 1- Participating teacher demographic data

Pseudonym	Sex	Ethnicity	School level	Experience	School Socio-Economic Status	Feedback subject
Celia	Female	NZE	Primary (Year 5)	5-10 years	High	English
Rebecca	Female	NZE	Primary (Year 6)	>10 years	High	English
Isabel	Female	NZE	Primary (Year 6/7)	2-5 years	Mid	English
Danielle	Female	Indian	Primary (Year 7)	>10 years	Low	Mathematics
Jennifer	Female	NZE	Primary (Year 7)	>10 years	High	English
Phyllis	Female	NZE	Primary (Year 7)	5-10 years	High	English
Wendy	Female	Asian	Primary (Year 7/8)	>10 years	Low	English
Elsa	Female	NZE	Primary (Year 8)	2-5 years	High	English and Mathematics
Stephanie	Female	NZE	Primary (Year 8)	>10 years	High	English and Mathematics
Teresa	Female	NZE	Secondary (Year 9, 10)	>10 years	High	English
Sylvia	Female	NZE	Secondary (Year 10)	>10 years	Mid	English

Notes. NZE=New Zealand European

All teachers were women and 64% were employed in high socio-economic schools, with the balance evenly split between middle and low SES schools. Two teachers taught Year 5 or 6 classes, five taught Years 7 or 8, two taught composite classes (i.e., Year 6/7 and Year 7/8), and two taught secondary English. Seven teachers had more than 10 years of experience, two had between 5 and 10 years, and two had between 2 and 5 years. Nine teachers were New Zealand European (i.e., white majority) and two described themselves as having an Asian background. Given the anonymity of data collection processes, no demographic information about the students who generated the actual feedback was available.

Data Collection

Teachers, with the consent of their schools and adhering to the ethical guidelines set for the project, allowed the first author to photograph or photocopy PASA statements written in a convenience sample of students' work. Student work was obtained from student notebooks or written assignments that the teacher had on hand. Note that these student notebooks were not the structured 'workbooks' seen commonly in American schooling; these were used to record responses to whatever activities the teacher assigned, making them a repository for daily class work. These examples were obtained while visiting the classrooms for an observation or to conduct an interview focusing on the teacher's classroom assessment

and feedback practices. During these visits, teachers were asked to provide the researcher with some examples of the kinds of feedback students in their class receive. As these data were naturally occurring, the quantity of examples was based entirely on teacher willingness to share instances of feedback, including PASA, with the researcher as examples of their formative assessment practices.

Of those who shared examples of PASA, most provided both peer- and self-assessment examples for the researcher, but two teachers showed only peer-assessment and four provided examples of only self-assessment. A total of 74 self-assessment examples from nine classes produced 289 utterances (i.e., meaningful sentences/sentence fragments), 39% of which were from one secondary school English class. Only 32 peer-assessment examples were found from seven classes producing 182 utterances, 70% of which were from the same secondary school English class that produced the most self-assessment utterances. These data are not representative of the feedback New Zealand students generate because of the mechanism by which they were collected, characterized by convenience sampling. Nonetheless, given the relatively large number of PASA examples, there is sufficient data to test Hattie and Timperley's (2007) typology as a protocol for classifying student written feedback and to allow preliminary conclusions about the content of student feedback practices.

Data analysis

The study used a priori content analysis (Silverman, 2006) to classify instances of PASA feedback statements into the Hattie and Timperley (2007) task, process, self-regulation, and self feedback categories. The thematic content of each utterance was determined. For example, the statement 'I think I could have worked better, but I did figure out how to find the difference between times' (Stephanie's student) was divided into two utterances as the first part of the statement relates to the student's perception of his or her own effort (self feedback), while the second half of the statement was a positive evaluation of their work (task feedback).

The first and third authors discussed the framework and conceptualized what these four categories might look like within the context of student PASA comments. A codebook was developed and the second author provided feedback on the subcategories and examples. The first and third authors coded the data independently, then reconciled differences, drawing on the second author to help classify statements where agreement had not been reached. Tables 2 and 3 provide examples of data categorized into each classification; Table 2 includes comments students wrote to themselves, while Table 3 contains comments written by an assessing peer.

While the secondary school students' comments were longer and more developed, most primary students' comments were brief (i.e., 1-2 sentences or sentence fragments). This meant, since it was not possible to clarify intent from the students themselves, some utterances were vague and difficult to classify. This was most noticeable within self Feedback, as there appears to be an element of self-regulation in giving oneself any kind of comment. For example, behaviour related comments (e.g., 'I think it could have been a lot better if I work with someone else because my buddy just went of [sic] and played') may have been self-protecting excuse-making or could have been genuine reflections on how to self-regulate behaviour for improved learning. While the inter-rater reliability was K=.62 (Cohen, 1992), which is sufficiently beyond chance to allow interpretation of results, it is lower than ideal and highlights the complexity encountered in classifying student feedback statements without access to student intentions.

Table 2- Examples of student self-assessment feedback

Task Feedback	Process Feedback	Self-regulation Feedback	Self Feedback
I did well with sequencing, even though I got 5 wrong. (Isabel's student)	To do more research and have more verbs and adjectives and adverbs to make it more interesting. (Celia's student)	I thought I would get 15/17. (Isabel's student)	I talked a lot on this second session and it was fun! (Isabel's student)
Use more speech marks. (Jennifer's student)	I didn't do the korus of the symbol because they use to [sic] much space. I made the silver fern bigger to make it the main image. (Isabel's student)	At first I wasn't sure how to do it at first when explained but then I got it. (Elsa's student)	My favourite part was making the title page. (Stephanie's student)
I think it was a bit short. And could have used some quotes. (Elsa's student)	But if I did it again I would get more information. (Stephanie's student)	I think we were really successful today because we worked as a team. (Stephanie's student)	I think it could have been a lot better if I work with someone else because my buddy just went of [sic] and played. (Rebecca's student)
I focused on one point and concluded well. (Sylvia's student)	Next time to improve my story I think I could make it more interesting. (Jennifer's student)	Tomorrow I will be working on learning about angles. (Stephanie's student)	I thought that I put a lot of effort into this letter. (Phyllis's student)
I am good at understanding a topic using multiple texts. (Sylvia's student)	I need to work on finding information in texts and understanding what I am reading. (Sylvia's student)	If I could improve on anything for next time I would use my time more wisely. (Stephanie's student)	It came out good. (Isabel's student)
My presentation was good and appealing. (Stephanie's student)	I need to explain in detail more of the comparison I need to go into a lot more detail and connect all the ideas into one. (Sylvia's student)	What do you think? (Isabel's student) I did better than last year. (Sylvia's student)	To do my work next time. (Sylvia's student)

Table 3- Examples of student peer-assessment feedback

Task Feedback	Process Feedback	Self-regulation Feedback	Self Feedback
I really like how your handwriting is nice and neat. (Isabel's student)	I think you need to use more vocabulary in your story and maybe read it yourself so that it makes sense. (Jennifer's student)	No examples found	Well done. (Isabel's student) Keep up the good work! (Wendy's student)
I think it was good how you used lots of insteting [sic] vocabulary. (Jennifer's student)	More details/description is needed to go further. (Sylvia's student)		Otherwise it's good. (Wendy's student)
Your story has really nice descriptive words. (Wendy's student)	How? Quote would be good or example of this feeling. (Sylvia's student)		Brilliant ☺ (Danielle's student) I get the impression that it was rushed? (Sylvia's student)
Good proof reading and writing of words and sentences. (Wendy's student)	Proof read is always good. (Sylvia's student)		Sorry if this sound harsh. (Sylvia's student)
These topic sentences are all reasonably clear about what the paragraph is talking about. (Teresa's student)	Next time is to use more multisensory. (Jennifer's student)		(Please don't hate me). (Sylvia's student)
I like your questions and answers. They are very informative. (Phyllis's student)	And use more speech maybe. (Jennifer's student)		Overall, very, very good (better than you said it was. You're to [sic] modest!) (Sylvia's student)

Note. Self Feedback contains statements written by an assessing peer which were aligned with the Self feedback category.

Results

All categories of feedback were found in both types of assessment, except for self-regulation feedback, which was absent in peer-assessment (Table 4). The proportional distribution of feedback types by assessment method was not equivalent ($\chi^2_{(3)}=15.9$, $p<.001$). Peer-assessment had a much higher percentage of task feedback than self-assessment, but this type of feedback predominated both conditions. Consistent with the self-regulatory nature of self-assessment, a small proportion of self-regulation feedback was found only in self-assessment; previous studies have noted teachers also seldom use this feedback type (Harris, Harnett, & Brown, 2013).

Table 4. Frequencies of Feedback utterances by assessment method

	Self-assessment		Peer assessment	
Feedback	*N*	%	*N*	%
Task	144	50%	130	71%
Process	59	20%	16	9%
Self-regulation	22	8%	0	0%
Self	64	22%	36	20%

Within the self-feedback type, different emphases were noted according to assessment method. Within peer-feedback, general praise dominated; whereas, in self-assessment, students' self-feedback focused more on effort, perceived task difficulty (e.g., this was hard or easy), interest or enjoyment, and general classroom behaviour. Some student comments were more focused on their own effort and general behaviour or interest than on the actual qualities of the task that might lead to improved performance (refer to Tables 2 and 3 for examples). Hence, ego protective factors clearly played a role in this PASA feedback sample.

The correlation between school level and type of feedback within each assessment method, including statistical significance of values and the power of the study, was found (Table 5). Power estimates were calculated post hoc using observed *N* and *r* values with G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to determine the probability of correctly rejecting the null hypothesis. Four of the comparisons failed to meet the conventional power threshold of .80, so it was decided not to attempt an interpretation of their relationships to school level. Nevertheless, all the values, except for self-regulation, were statistically significant. Hence, for four values (i.e., task and process in self-assessment and task and self in peer assessment), relatively strong confidence in the relationships is possible. Positive values in the table indicate generally a positive relationship between increased school year and type of feedback. Increased level of schooling was especially correlated with increased use of task feedback. The positive correlation for process feedback in self-assessment is consistent with increasing age and experience being associated with a greater tendency to focus on learning-productive feedback (Brown & Harris, 2013). The positive association between age and self feedback within peer assessment seems to reflect a greater concern for interpersonal relations, also consistent with previous studies (Brown & Harris, 2013).

Table 5. Correlations between School Level and Feedback Type Frequency

	Method of assessment		Power Estimates	
Feedback	Self	Peer	Self	Peer
Task	0.61***	0.51***	1.00	.99
Process	0.61***	0.54*	.99	.61
Self-regulation	0.29^{ns}	na	.26	.00
Self	-0.27*	0.48**	.59	.86

Note. Coding for School Year levels: 5-6 = 1; 7-8 = 2; 9-10 = 3; all values are Pearson product moment correlations; positive values indicate that as school level increases so does the type of feedback; na=not applicable since no instances of self-regulation were found within peer-assessment; * = $p<.05$; ** $p<.01$; *** = $p<.001$; ns = not significant.

Discussion

This sample of New Zealand students in Years 5-10 tended to give predominantly task and process oriented feedback. It appeared that task feedback usually related to an evaluation of the learning, answering Hattie and Timperley's (2007) 'how am I going?' feedback question; whereas, process feedback often provided suggestions for improvement, more closely aligning to the 'where to next?' question. While self-regulation feedback comments occasionally referred to some kind of learning goal ('where am I going?'), most were also providing comments which evaluated progress or discussed next steps. Hence, if student self-regulation is an intended outcome from peer- and self-assessment practices, it is important for teachers to remember that this activity generally requires self-reflection in response to goals for learning (Zimmerman, 2008). Thus, classroom use of PASA is unlikely to generate self-regulation feedback in the absence of clear goals for learning. Teachers may, perhaps rightly, encourage students to provide themselves and each other with task and process feedback, with the expectation that students will incorporate such feedback into their own self-regulated learning processes.

Self-regulatory feedback was absent in peer-assessment modes and only rarely found in self-assessment. The absence of self-regulation feedback in the peer-assessment mode within this study aligns with Gan's (2011) findings that self-regulation feedback was extremely rare, even in upper secondary school students' peer-assessment comments, despite the training provided in that study. Hence, while it is possible that students in this study may actually creating more of these comments than the data would suggest, it is fairly certain that self-regulation feedback is less prevalent than other feedback levels.

Clearly, there is a need to better understand why self-regulation comments are relatively rare, with multiple potential hypotheses explaining their scarcity. First, teachers seldom provide this kind of feedback in their written comments (Harris, Harnett, & Brown, 2013), so students whose feedback practices mimic their teachers' feedback would also be unlikely to create self-regulation comments. Getting students to create useful self-regulation feedback statements is clearly something educators implementing PASA in compulsory classrooms can explicitly teach to their students. This is consistent with the strong role teachers need to play in setting up and guiding PASA practices and criteria despite the increased student responsibilities (Ploegh, Tillema, & Segers, 2009).

Second, students, as relative novices, may have difficulty in constructing such feedback. Perhaps it is unrealistic to expect peers to be able to create self-regulation level comments, a role that teachers themselves find difficult. In order to produce such feedback, the feedback giver must not only identify areas for improvement within the work, but must also avoid providing answers, instead directing the feedback recipient to rules, strategies, or behaviours previously learned that the learner can employ to improve his or her own work in

the future. This may be a set of skills beyond the scope of most students in compulsory schooling.

Third, the rarity of self-regulation feedback in these comments may also reflect how students logically enact such evaluation processes. As our sample was comprised of written comments, only what students actually wrote down about their work for a particular audience was captured. Thus, feedback in this format is unlikely to accurately reflect everything students think about the quality of their own or their peers' work. It is quite possible that students may have self-regulatory feedback in mind (e.g., 'My essay improved because I followed the writing process we have learned'), but do not see the need to write these kinds of observations down as they already have access to their own thoughts (Andrade, 2010). Moreover, writing self-oriented feedback comments that a teacher will see is a somewhat unnatural process. Cowie (2009) has shown that students can be distressed by compulsion to disclose their self-feedback to teachers; after all, such comments could be very private. Furthermore, when students know teachers will be reading their feedback, students may feel a need to provide the kind of comments they believe teachers want (i.e., social desirability) (Harris & Brown, 2013), which may decrease honesty in responding.

Also, within the data, there were correlations between age and experience and feedback levels. Within this sample, the effects of schooling experience or age seem apparent within self-assessment, with older students providing more self-regulation and less self feedback. The tendency for increased self-feedback in peer-assessment as students progress through school points out the potential importance of interpersonal relationships as a mediator when providing feedback to social peers. Nonetheless, despite the obtained power, it is possible, given the restricted sample of teachers, that the current results reflect teacher practices, rather than a more general pattern of changing feedback comments as level of schooling increases. Hence, future research needs to include a broader range and number of teachers to eliminate this confounding explanation. While the current sample is too small to be conclusive about this relationship, it certainly merits study in a larger, more representative sample. While the instructional activities contextualizing these PASA feedback samples were unable to be examined within this particular study given its data collection method, it would be useful for future studies to continue Gan's (2011) work examining the effects of particular instruction or supports (e.g., rubrics, exemplars, graphic organisers) on PASA content within diverse grade levels and content areas.

The study has shown that although Hattie and Timperley's (2007) four levels of feedback can be applied to student-generated PASA comments, there are still some challenges in using the model. Interpreting comments without access to students' intentions is difficult. Self feedback seems to be a catch-all category for all ego or relationship-oriented statements. While Hattie and Timperley's (2007) systematic review of literature suggested that self-feedback does not help improve academic performance, there may be some types of self feedback which interact with other factors to indirectly affect achievement. For example, while many studies identify that non-specific praise is ineffective (e.g., Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman, 2014; Burnett & Mandel, 2010), there is more debate about the potential role that effort feedback may have in relation to student motivation and persistence (e.g., Lam, Yim, & Ng, 2008), as such feedback may encourage students to view learning as an internal and controllable process. For the purposes of teaching and learning, it seems helpful to divide self feedback into subcategories (e.g., general praise, comments about effort/perceived task difficulty, behaviour comments, general non-task related comments, comments designed to help maintain social relationships) so teachers and students can more clearly identify the kinds of statements that fall into this category and are unlikely to improve learning, at least directly.

Within research contexts, it may also sometimes be helpful to further divide self feedback into more specific categories. For example, many self comments appeared to be written with specific social-psychological purposes in mind (e.g., to get the teacher to think you tried or to soften the blow of negative evaluations to a peer). While research suggests these may not actually benefit learning (Hattie & Timperley, 2007), it is necessary to examine if they have any positive classroom functions (e.g., enhanced motivation or improved relationships). Including sub-categories within self feedback would help enable researchers to potentially correlate these types of statements with particular outcomes of interest (e.g., student motivation, engagement, academic achievement scores) to better understand the roles that such statements play within the learning environment.

Further clarification may also be helpful with the task level of feedback. If assessing the quality of the feedback is a goal of content analysis, within task feedback, it would seem worthwhile to distinguish between comments about surface features (e.g., spelling, calculation, grammar) and deeper features (e.g., rhetorical organization, adoption of the correct procedure) inherent in a task. Clearly, a classificatory taxonomy (e.g., the Structure of Observed Learning Outcomes—Biggs & Collis, 1982) that allows separation of demand into surface and deep would be needed to establish the relative cognitive complexity of feedback content. It seems likely that task feedback consistently pitched at the surface level will not allow the learner to progress towards achieving more complex learning goals.

Conclusions

This study has built on work begun by Gan (2011) to examine the utility of using Hattie and Timperley's (2007) model as a means of classifying the content of peer- and self-assessment comments of upper primary and lower secondary students. It has demonstrated that their four levels of feedback can be found in upper primary and lower secondary student self-assessment feedback and has corroborated Gan's (2011) findings that the model works relatively effectively to classify peer-assessment comments. While it may be useful to create further distinctions at the task and self level, overall, the model was able to be used effectively to classify these data, suggesting it would be potentially useful within other studies seeking to examine PASA content. As Gan and Hattie (2014) found that upper secondary students were able to produce a higher percentage of self-regulation comments to peers after being taught about Hattie and Timperley's (2007) feedback types and provided with a graphic organiser to scaffold their thinking, this type of training may be a positive way forward to help address the predominance of task feedback found in this study. However, the fact that the students' comments were dominated by task rather than self feedback is positive and suggests that the majority of their PASA feedback is learning rather than well-being orientated.

However, Hattie and Timperley's (2007) model is only one which could be potentially used to classify feedback data (e.g., Shute, 2007, Tunstall & Gipps, 1996). Future studies should investigate the benefits and limitations of using alternative models to classify PASA data, allowing researchers and educators to select the model most suitable for their purposes.

While instructive, this study's data set is too small and unrepresentative to allow generalizations, a challenge to be undertaken in larger future studies. The study has established the relative utility of using this particular model to classify data and has provided some early insight into the content of student generated feedback comments. As data collected in this study were naturally occurring, it is less likely to be contrived, but a larger and more systematically collected sample is required before any generalizable comments can be made about patterns or trends in New Zealand PASA feedback practices. However, this study has helped validate a useful coding mechanism which, with modifications, has the

potential to help shed light on the quality and complexity of student PASA comments within global educational settings.

References

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of Formative Assessment* (pp. 90-105). New York: Routledge.

Andrade, H. G., & Boulay, B. A. (2003). Role of Rubric-Referenced Self-Assessment in Learning to Write. *The Journal of Educational Research, 97*(1), 21-34.

Assessment Reform Group. (2002). Assessment for learning: 10 principles. Retrieved December 9, 2009, from http://www.assessment-reform-group.org/CIE3.PDF

Berry, R. (2011). Assessment reforms around the world. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 89-102). Dordrecht, NL: Springer.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome).* New York: Academic Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Blatchford, P. (1997). Students' self-assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology, 17*(3), 345-359. doi: http://dx.doi.org/10.1080/0144341970170308

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education, 18*, 529-549.

Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation, 6*(12), 68-91.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 367-393). Los Angeles: SAGE.

Brown, G. T. L., Harris, L. R., & Harnett, J. A. (2012). Teacher beliefs about feedback within an Assessment for Learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education, 28*(7), 968-978. doi: 10.1016/j.tate.2012.05.003

Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction, 19*(2), 97-111. doi: 10.1016/j.learninstruc.2008.02.003

Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Beliefs that make a difference: Adaptive and maladaptive self-regulation in students' conceptions of assessment. In D. M. McInerney, G. T. L. Brown & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning.* (pp. 159-186). Charlotte, NC US: Information Age Publishing.

Brummelman, E., Thomaes, S., Orobio de Castro, B., Overbeek, G., & Bushman, B. J. (2014). "That's Not Just Beautiful—That's Incredibly Beautiful!": The Adverse Impact of Inflated Praise on Children With Low Self-Esteem. *Psychological Science*. doi: 10.1177/0956797613514251

Burnett, P. C., & Mandel, V. (2010). Praise and Feedback in the Primary Classroom: Teachers' and Students' Perspectives. *Australian Journal of Educational & Developmental Psychology, 10*, 145-154.

Chang, C. C., Tseng, K. H., Chou, P. N., & Chen, Y. H. (2011). Reliability and validity of Web-based portfolio peer assessment: A case study for a senior high school's students taking computer course. *Computers & Education, 57*(1), 1306-1316. doi: http://dx.doi.org/10.1016/j.compedu.2011.01.014

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cowie, B. (2005). Pupil commentary on assessment for learning. *Curriculum Journal, 16*(2), 137-151.

Cowie, B. (2009). My teacher and my friends help me learn: Student perspectives and experiences of classroom assessment. In D. M. McInerney, G. T. L. Brown & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about Assessment for Learning* (pp. 85-105). Charlotte, NC: Information Age Publishing.

Crooks, T. J. (2010). Classroom assessment in policy context (New Zealand). In B. McGraw, P. Peterson & E. L. Baker (Eds.), *The international encyclopedia of education* (3rd ed., pp. 443-448). Oxford, UK: Elsevier.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Gan, M. (2011). *The effects of prompts and explicit coaching on peer feedback quality.* PhD, University of Auckland, Auckland.

Gan, M. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 1-18. doi: 10.1007/s11251-014-9319-4

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304-315.

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*(0), 101-111. doi: http://dx.doi.org/10.1016/j.tate.2013.07.008

Harris, L., Brown, G. L., & Harnett, J. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability*, 1-27. doi: 10.1007/s11092-013-9187-5

Harris, L. R., Harnett, J. A., & Brown, G. T. L. (2013, April 27-May 1). *Exploring the content of teachers' feedback: What are teachers actually providing to students?* Paper presented at the American Educational Research Association Annual Meeting, San Francisco.

Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses in education*. London: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Higgins, K. M., Harris, N. A., & Kuehn, L. L. (1994). Placing assessment into the hands of young children: A study of student-generated criteria and self-assessment. *Educational Assessment, 2*(4), 309-324. doi: http://dx.doi.org/10.1207/s15326977ea0204_3

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Lam, S.-f., Yim, P.-s., & Ng, Y.-l. (2008). Is effort praise motivational? The role of beliefs in the effort-ability relationship. *Contemporary Educational Psychology, 33*(4), 694-710.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment minute by minute, day by day. *Educational Leadership, 63*(3), 18-24.

Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing, 12*(3), 180-198.

McDonald, B. (2009). Exploring academic achievement in males trained in self-assessment skills. *Education, 37*(2), 145-157. doi: http://dx.doi.org/10.1080/03004270802069244

McDonald, B., & Boud, D. (2003). The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*(2), 209-220. doi: http://dx.doi.org/10.1080/0969594032000121289

Miller, P. J. (2003). The Effect of Scoring Criteria Specificity on Peer and Self-assessment. *Assessment & Evaluation in Higher Education, 28*(4), 383-394. doi: 10.1080/0260293032000066218

Ministry of Education. (2010). Ministry of Education Position Paper: Assessment [Schooling Sector]: Ko te Wharangi Takotoranga Arunga, a te Tauhuhu o te Matauranga, te matekitenga. Retrieved from http://www.minedu.govt.nz/theMinistry/PublicationsAndResources/AssessmentPositionPaper.aspx

Nicolaidou, I. (2013). E-portfolios supporting primary students' writing performance and peer feedback. *Computers and Education, 68*, 404-415. doi: 10.1016/j.compedu.2013.06.004

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research, and Evaluation, 10*(17), 1-8.

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

Peterson, E.R., & Irving, S.E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction, 18*(3), 238-250.

Ploegh, K., Tillema, H. H., & Segers, M. S. R. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation, 35*(2-3), 102-109.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation, 11*(10), 1-13.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998). Skills training versus action research in-service: impact on student attitudes to self-evaluation. *Teaching and Teacher Education, 14*(5), 463-477.

Shute, V. (2008). Focus on formative feedback *Review of Educational Research, 78*(1), 153-189.

Silverman, D. (2006). *Interpreting qualitative data* (3rd ed.). London: Sage

Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence, 33*(1), 135-145. doi:10.1016/j.adolescence.2009.04.004

Taras, M. (2003). To Feedback or Not to Feedback in Student Self-assessment. *Assessment & Evaluation in Higher Education, 28*(5), 549-565. doi: 10.1080/02602930301678

Topping, K. (2013). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *Sage Handbook of Research on Classroom Assessment* (pp. 395-412). Los Angeles: SAGE.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education, 25*(2), 149-169. doi:10.1080/713611428

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*(4), 506-519. doi:10.1016/j.learninstruc.2010.08.002

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389-404.

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41-54.

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280-290.

van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies*, 1-17. doi:10.1080/03055698.2012.654920

van Zundert, M. J., Sluijsmans, D. M. A., Könings, K. D., & van Merriënboer, J. J. G. (2011). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology, 32*(1), 127-145. doi:10.1080/01443410.2011.626122