Action research on the effect of descriptive and evaluative feedback order on student learning in a specialized mathematics and science secondary school

Hooi Ling Chua[1], Sheau Huey Lee[1] and Gavin W. Fulmer[2*]

* Correspondence:
gavin-fulmer@uiowa.edu
[2]The University of Iowa, N259
Lindquist Center, Iowa City, IA
52242, USA
Full list of author information is
available at the end of the article

Abstract

We investigated the effect of feedback order—teachers' written, descriptive comments followed by evaluative scores—on students' performance and learning for chemistry and mathematics in a Singapore mathematics and science specialized secondary school. This action research adopted an explanatory mixed-methods design with an intervention, with interviews, student survey, and assessments. The participants were 60 secondary school students: 33 students from secondary-one and 27 students from secondary-four. Repeated measures ANOVA results from the four-week study period indicated that there was no significant difference ($p > 0.10$) between the performance of the participants who received comments only (C group) and participants who received comments followed by evaluative scores (CS group) for both chemistry and mathematics, indicating there was no negative effect on receiving evaluative scores after the written descriptive comments. Qualitative findings indicated students could recognize the goals of the feedback and the score delay. The study shows that feedback order is important to consider in comparing effects of different forms of feedback, with implications for future research and for practice.

Keywords: Teacher comments, Teacher feedback, Feedback order, Classroom assessment, Action research, Science education, Mathematics education

Introduction & literature review

Effective feedback appears as one of the most powerful influences on learning, achievement, and teaching (Black and Wiliam, 1998a, 1998b; Hattie and Timperley, 2007; Tunstall and Gipps, 1996). Feedback can take on many forms including evaluative scores and descriptive comments, and is used extensively in learning science and mathematics as it is in other subjects (Sutton, 2010). The feedback model proposed by Hattie and Timperley (2007) is adopted as a theoretical framework for this study. Under this model, feedback serves as part of assessment for learning because it is focused on the students' receiving, understanding, and acting on the feedback in their learning. This is subtly different from the notion of formative assessment wherein assessment information is a guide to teachers in planning for instruction. According to this model, the main purpose of feedback is to reduce discrepancies between a current

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 2 of 22

understanding of a performance and a targeted goal, in order to enhance students' learning and performance (Hattie and Timperley, 2007). Therefore, an effective piece of feedback should provide explicit information to close this gap, and teachers must continue monitoring and evaluating students' understanding.

The literature has widely reported on the effectiveness of different modes of feedback, such as scores, grades, written comments, and comments plus grades (Black and Wiliam, 1998a, 1998b; Brookhart, 2011a, 2011b; Butler, 1988; Butler and Nisan, 1986; Crooks, 1988; Frisbie and Waltman, 1992; Guskey and Lee, 2013; Page, 1958). In this study, we focus explicitly on descriptive written comments with and without reporting of evaluative scores. Reviewing previous research shows no conclusive pattern of findings on the effects of grading and scoring for teaching and learning purposes, although a consensus does appear to be developing. On one hand, the conjectured benefits of reporting grades and scores include: providing information about students' achievement for self-evaluation; presenting evidence of students' presence or lack of effort and responsibility (ostensibly to bolster this effort or sense of responsibility); evaluating the effectiveness of instructional techniques; and providing an incentive or extrinsic motivation for students in their learning process (Frisbie and Waltman, 1992; Guskey and Bailey, 2001; Guskey and Lee, 2013).

On the other hand, and in contrast to these potential benefits of reporting students' grades and scores, research studies have explored the detrimental effects of overemphasis on grades (e.g., Crooks, 1988; Kohn, 1994). According to Butler (1988), a focus on grades and grades plus comments generally has an undermining effect on both the learning interest and performance of students. Providing grades and scores can be a main distracter for effective written feedback as students fail to read, digest, and act upon the descriptive written feedback that accompanies grades or scores (Sutton, 2010) because they focus directly on the performance result. Furthermore, other research on score reporting suggests that grades are not essential for instructional, teaching, and learning processes (Brookhart, 2011a, 2011b; Frisbie and Waltman, 1992; Guskey and Lee, 2013). In fact, it is found that without grades, students can and do learn equally well (Frisbie and Waltman, 1992). Guskey (2014) also suggested that grading and reporting grades are not essential to teaching instruction. He highlighted that the purpose of grading appeared to be more crucial in the grading process.

Compared to a grade or score alone, task-focused comments or descriptive feedback is found to be more effective in providing guidance (Hattie and Timperley, 2007) for learning improvement. It stimulates interest in the task itself and, thus, could achieve a significant improvement in test performance and attitudes amongst students (Butler, 1988; Black and Wiliam, 1998a, 1998b; Elawar and Corno, 1985; Kohn, 1994).

Task-focused comments benefitted both low- and high-achieving learners (Butler, 1988). After receiving comments, learners demonstrate higher interest and improved performance in related tasks (Butler, 1988). Many observations and findings on the effectiveness of feedback in the form of comments compared to grades in the learning process are widely reported in literature (Black and Wiliam, 1998a, 1998b; Butler, 1988; Butler and Nisan, 1986; Crooks, 1988; Kohn, 1994; Lipnevich and Smith, 2009; Page, 1958; Smith and Gorard, 2005). Additionally, the inclusion of the grade with descriptive comments also appears to undermine the positive effects of the comments on both learning interest and performance of students (Butler, 1988). Crooks (1988) suggested

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 3 of 22

some possible mechanisms by which reporting of grades could hurt students: using up classroom learning time that could be more beneficially spent; reduction in students' intrinsic motivation; increases in students' anxiety over scoring and evaluation; ability attributions for success and failure that could undermine student effort; lowered self-efficacy for learning among weaker students; reduced use and effectiveness of feedback to improve learning; and poorer social relationships among the students. Though context, cultural differences, and motivational and educational settings may influence these research findings, there appears to be a strong consensus that teachers should emphasize descriptive feedback for students rather than reporting grades or scores.

Despite the disputable effects of scores or grades upon learning, providing scores or grades is still a common practice in many schools throughout Singapore. This is inevitable and understandable as the reporting of grades or scores is the most direct way to reflect the academic performance of students in school. Numerous research studies reported that many students in Singapore experience high academic stress associated with expectations of parents, teachers, and the students themselves (Ho and Yip, 2003; Isralowitz and Ong, 1990; Joyce and Shirley, 2011). High academic expectations are communicated explicitly and implicitly by teachers in class or by parents' child-rearing attitudes (Ho and Kang, 1984). In such a climate, it may be impossible to completely forgo the reporting of grades or scores to indicate the academic performance of students in school. How then could schools and teachers achieve the beneficial effects of descriptive feedback without invoking the negative effects of over-reliance on scores as feedback? This article presents one attempt to answer this question in chemistry and mathematics classrooms.

Score reporting and feedback in the Singapore education context

As in other educational systems, especially in East Asia, Singapore teachers continue to practice the reporting of grades or scores alone, or providing grades and scores accompanied with comments. The existing discrepancy can be attributed to three factors: (i) teachers' understanding and beliefs in providing scores or grades to help students in learning instead of using the effective descriptive feedback; (ii) prevailing practices in using grades and marks as feedback are meant to fulfill the score and grade expectations of students and parents in terms of academic achievement; and (iii) teachers' teaching and administrative workloads in schools. We expand on these briefly below.

To further elaborate and explain the context, firstly, Singapore's meritocracy and its assessment system emphasize and promote individual academic performance and achievement in schools (Koh and Luke, 2009; Lim, 2013; Tan, 2011; Tan and Deneen, 2015). Hence, in order to facilitate the purpose of streaming and academic sorting among students in school, the academic performance and results are commonly revealed in the forms of grades and scores. For example, all primary six (grade 6) students are required to sit for the national examination (The Primary School Leaving Examination, PSLE) administered by the Ministry of Education. Based on the normed and academic T-scores in the PSLE, all students will be sorted and allocated to different academic tracks (such as 4 years' integrated programme, 5 years secondary schools or vocational schools) in the next secondary level of schooling system (Koh and Luke, 2009; Luke, Freebody et al., 2005). As a result, students' prior experiences in the

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 4 of 22

national examination (PSLE) and most of the scfhool exams in Singapore taken by all primary students may have rendered the students "grades and scores-dependent" expectations. Parents, students, and even teachers are conditioned to believe that the level of attainment, whether as a score or a grade, is predominantly considered a relative comparison to the performance of other students. These expectations would make it a challenge to relinquish the current feedback practices in the forms of grades and scores and to shift to written descriptive feedback. Teachers, students, and parents need to wean off this dependence on scores and grades and to focus on the more useful and specific descriptive feedback. Once again, the expectation of stakeholders and students for marks or grades appears to undermine the potential for learning.

Besides, the general practice for feedback is still very much based on norm-referenced forms (comparing the individual's performance with that of others) instead of focusing on self-referenced forms (comparing performance with other measures of the individual's ability) or of absolute achievement (comparing performance to a defined goal of mastery). The main implication is that low achieving students should be given opportunity to focus their attention on their own learning progress without scores (self-referenced feedback) rather than receiving normative feedback (Sutton, 2010).

Secondly, teachers are urged by a strong sense of accountability and obligation to communicate to both parents and students regarding the level of learning achievement attained in school. Teachers need to make parents and students satisfy their concerns for grades and scores' expectations in gauging the learning performance. Parents and even many teachers are unaware that research has demonstrated score reporting undermines the learning gains and test performance, especially for lower achievers (Butler, 1988), and that grading seems to emphasize more on "how well" than on "how to improve" (Ory and Ryan, 1993).

Thirdly, many teachers experience the demands of parents who are anxious about the "timeliness" of providing effective feedback (Sutton, 2010) for their classes. Teachers' practices and beliefs, coupled with heavy teaching and administrative workloads, form an inevitable impediment to providing descriptive feedback in formative assessments and to implement differentiated instructional approaches in class. The management of marking workload within a short timeframe is immense, particularly with teachers trying to mark everything themselves, which can cause marking backlogs that are difficult to be cleared (Sutton, 2010), adding more burden and demanding more time and greater commitment.

Research question

Given the stakeholders' expectations for score or grade reporting, and recognizing the research evidence for the benefits to achieving the desired learning goals that are best supported by descriptive feedback, we sought to give meaningful descriptive comments but find a way to report scores in a non-detrimental fashion. We do this by investigating the impact of feedback order on students' use of feedback information. There is relatively little research done to understand how learning is affected by providing two different modes of feedback in a sequential manner. As such, we created a time delay between the descriptive comments and the score reporting to examine how this affects students' learning.

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 5 of 22

Hence, we posed the following research question: *Is there any difference in students' learning between these two groups: (i) giving descriptive comments without revealing scores and (ii) giving descriptive comments with a delay in revealing scores?* By managing the feedback order, we hoped to maximize the benefits of descriptive comments so that students could focus on the self-correction and learning process by looking at the comments provided rather than the score. At the same time, the researchers also hoped to manage the score expectations from parents and students.

Method

We used an explanatory mixed-methods quasi-experiment design (Creswell, 2014, pp. 583) which began with an intervention to each group of participants: comments only (C group) or comments followed by evaluative scores (CS group). The study proceeded with a qualitative data collection to help understand the effects of each intervention. In the following sections, we presented our theoretical framework, the study procedures, measures, and analyses.

We adopt the model of feedback proposed by Hattie and Timperley (2007) as our theoretical framework. According to this model, the main purpose of feedback is to enhance students' learning and performance. Therefore, an effective piece of feedback should provide explicit information to close this gap, and teachers must continue monitoring and evaluating students' understanding.

Participants and setting

Participants were 60 secondary school students: 33 in a Secondary-One (equivalent to 7th grade in the U.S. system) chemistry class, with mean age of 13 years; and 27 in a Secondary-Four (equivalent to 10th grade in the U.S. system) mathematics class, with average age of 16 years. The participants include both Singaporean and non-Singaporean students, about 30% female, enrolled in an independent specialized mathematics and science high school. For enrollment into this specialized school, students sit for a DSA test (Direct School Admission Test) and must pass at the established benchmark to secure a place in the school. Hence, the participants were generally quite competitive and participative in terms of academic performance among their peers, because they were carefully selected before admission into this specialized school. As such, the participants were generally passionate learners particularly in science and mathematics subjects. This independent high school adopts a unique academic curriculum and modular credit system compared to other high schools. The participants are not required to sit for the national exams (i.e., Singapore's 'O' Level and 'A' Level). At the end of the 6 years' program, they will be conferred with a high school diploma certificate before embarking on their university life.

In our study, participants in each intact class underwent one intervention, either receiving comments only without evaluative scores (C group) or receiving comments followed by evaluative scores (CS group). Due to the practical limitations of the school context, a random assignment of participants is impossible. All classes were taught by the same teacher-researchers—the first or second authors on this article—who are subject-specialist teachers in chemistry and mathematics respectively. Both teachers have been teaching at this school for more than 6 years.

Instruments

Our instruments in this study are teacher-generated learning checks and quizzes, a post-intervention survey and open-ended interviews.

The learning checks in the chemistry unit addressed the 'O' Level learning concepts of chemical bonding (e.g., ionic, simple covalent, giant covalent, and metallic bonding), writing chemical formulae and names. For the mathematics unit, the learning checks covered the calculus topics in differentiation (e.g., concavity, critical points, local linear approximation) which were pitched at the 'A' Level standard. Each of the learning checks and the quiz consisted of core questions and challenging questions to stimulate students' thinking. The core questions addressed concepts taught in the lessons as a measure of students' learning. The challenging question was intended to encourage the participants to consciously increase their effort to tackle questions with higher-order thinking skills, rather than just doing more routine and similar questions (Kluger and DeNisi, 1996). After the quiz was administered, all participants completed a survey that addressed the participants' perceptions of their learning in the unit, the effectiveness of the learning check, and the role of the descriptive comments in their learning. Finally, a sample of 13 students was interviewed and they shared their experiences with the descriptive comments and scoring.

Procedures

Our research was conducted over a period of 4 weeks, selected because it suits the topics in the two subject areas. After these 4 weeks, the participants then proceeded to learn a new topic with different skills and difficulty levels. A new topic also required a different mark scheme and assessment criteria. Hence, we focused on just one particular topic in each subject throughout the 4-week period of our research study.

All participants from both CS and C groups took a series of three learning checks and a quiz. The checks and the quiz are pen-and-paper formative assessments consisting of questions on the content of units taught over the 4-week period of our research study. The duration of each learning check and quiz was between 20 to 30 min. Each question was assigned with a designated score. Teachers read and evaluated participants' learning check responses using a mark scheme, and wrote comments on each participant's paper.

The following research procedures were conducted:

1. The learning checks were returned to participants within a few days, as comment was more effective when provided sooner (Bangert-Drowns et al., 1991).
2. For each learning check, participants in the C group received written descriptive comments and time for self-correction, while participants in the CS group received written descriptive comments, time for self-correction, and then received scores *after* they had reviewed the descriptive comments.
3. At the end of the 4-week period, all participants took a quiz that consisted of core content and challenging questions learnt over the 4-week period as described above.
4. After completing the quiz, participants completed a survey about their experience with the learning checks, and a subsample of 13 participants was invited to complete an open-ended interview. Students were selected for interviews purposefully to include

      high- and low-achievers from both C and CS groups. All interviews were audiotaped by the researchers.

5. As this was an action research, with the first two authors serving the dual roles of teacher and researcher, the collection of data and analyses of the results were conducted with as neutral stance as possible.

6. The procedures and data were vetted and evaluated by two neutral evaluators who were not involved in teaching of the classes: the third author, a university faculty member who supervised the first two authors; and another humanities teacher with experience in action research in the same specialized school with more than 15 years of teaching experience.

The learning checks were marked and scored, but were not recorded for participants' semester grades. Unlike the learning checks, the quiz was scored and included in semester grades. Participants were aware of the grading for the learning checks and the quiz from the outset of the study. To encourage review of the comments, participants were provided with time in class to do self-correction on the returned learning checks. The hardcopy of the suggested answers was given to participants after they completed their self-correction. The availability of suggested answers was controlled in order for feedback to be effective (Kulhavy, 1977). No teacher coaching was provided during the self-correction process. At the end of the self-correction and review process, the participants from CS group received the scores on the learning checks. This delay in reporting the score was intended to prevent the participants from overlooking the comments, to help them to focus on the self-correction and learning process. The main purpose of feedback is to recognize and learn from mistakes, so the dedicated time for such self-correction action is important.

Descriptive comments

The provided comments were adopted from the Universal Intellectual Standards, which consists of seven components: accuracy, precision, relevance, depth, breadth, logic, and fairness (Elder and Paul, 2006). These standards were incorporated to stimulate the participants' thinking and reflection towards their own mistakes and to guide them to improve their work through a process of self-correction. Several types of questions were provided as part of the descriptive comments, including: questions used to probe participants' thinking; questions used to hold participants accountable for their own thinking; questions which, through consistent use by the teacher in the classroom, would be internalized by participants as reflection questions. Table 1 gives examples of some of the comments provided.

Grouping of participants

Both groups who received comments only (the C group) and comments followed by scores (the CS group) were categorized into high achiever (HA) and low achiever (LA) categories. The criterion used for these categories were based on their academic performances in chemistry and mathematics examinations in the previous semester. Table 2 shows the number of participants in each group and the subcategories.

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 8 of 22

Table 1 Framework for written descriptive feedback adopted in the study

Universal Intellectual Standards (UIS)	Descriptive comments will be provided in the form of critical thinking questions
1. Clarity	Could you elaborate further on that point?
	Could you illustrate more?
	Could you give me an example?
	Could you express this point in another way?
2. Accuracy	How could we check this? How could we find out if that is true?
3. Precision	Could you give more details?
	Could you be more specific?
4. Relevance	How is your answer/solution connected to the question?
5. Depth	How does your answer/solution address the complexities in the questions?
	How are you taking into account the problems in the question?
6. Breadth	Do we need to consider xxx?
	Is there another way to look at this question?
	What would your answer be if xxx?
7. Logic	Does this really make sense?
	How does this step follow?
	Before you implied this, and now you are making this claim, how can both be true?

Note. The levels and the questions used are adopted from Elder & Paul (2006)

Interview protocols

A total of 13 participants (Table 3) from both the C and CS group were selected for interviews. Fifteen open-ended interview questions were crafted. The interview data were coded for analysis. These interview questions focused on (i) the learning impact of learning checks' worksheets (with teacher's descriptive comments and participants' own corrections); (ii) the effect and preference of feedback order (comments followed by self-corrections, then suggested answers and lastly with scores or without scores); and (iii) the effectiveness of descriptive comments followed by suggested answers without teacher's verbal coaching.

As this was action research, the first two authors served as both teacher and researcher, so were consciously aware of these dual roles. The analysis and collection of data were conducted from a neutral stance to the extent possible. Interview and survey questions were crafted carefully and all data collated was evaluated equally. To avoid population definition errors, HA and LA participants from both CS and S groups were carefully identified and recruited for interviews to provide adequate representation. During the interview, the researchers were careful to avoid leading students to

Table 2 Number of LA and HA Participants for Chemistry and Mathematics

	Chemistry		Mathematics	
	C Group	CS Group	C Group	CS Group
Low Achiever (LA)	8	11	8	10
High Achiever (HA)	7	7	4	5
Total n	15	18	12	15

Table 3 Number of LA and HA Participants Involved in Interviews for Two Subjects

	C Group	CS Group
Low Achiever (LA)	3	4
High Achiever (HA)	3	3

particular statements, and use many open-ended questions with the goal of giving participants more opportunity to make themselves clear. These steps and the analyses undertaken (described below) are intended to represent students' ideas and experiences as authentically as possible, thus increasing the value of the results and reducing the risk of mistakenly interpreting students' ideas.
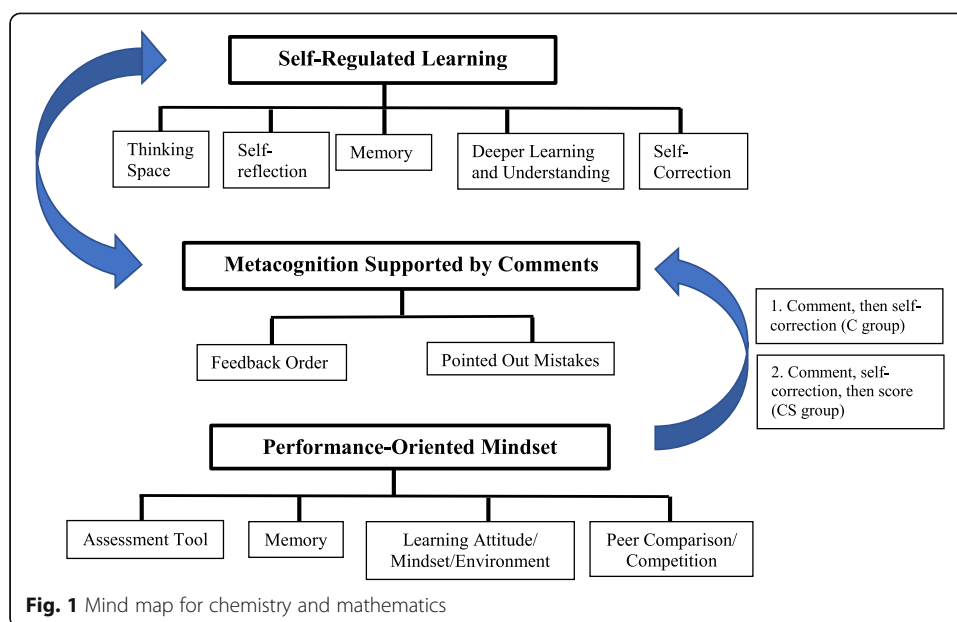
Analyses

For the quantitative findings, we used the score for each learning check (LC) and quiz to identify the differences in the participants' performance between the C and CS groups. A repeated measures analysis of variance (ANOVA) was adopted in our measurement. The group differences and changes over time were examined. The assumptions of ANOVA - sphericity and homogeneity of variance were also checked and confirmed. Mauchly's test of sphericity revealed non-significant results between the C and CS groups, whether in the subgroup of high or low achievers in mathematics (LA, $p = .702$; HA, $p = .977$) or chemistry subjects (LA, $p = .507$; HA, $p = .211$)—indicating that the sphericity assumption was met. Levene's test of equality of error variances showed non-significant results ($p > .10$) for all subgroups except for the high achievers in chemistry. The pattern of results in correcting for inequality of variances was consistent and the assumption of homogeneity of variance was met in this analysis.

For the survey questions, several steps of review process took place. For step 1: analyze how each survey questions and response options were comprehended/ understood by participants. Then, step 2: perform the survey task (includes data collection and collation, calculations, etc.) Lastly, step 3: match answer to an available response option. Overall, the survey results suggest that participants found the survey questions were short, clearly worded, and could be completed within a short time. Students' first impressions of the survey were favorable as it consisted of mostly short multiple choices questions and with an easy readable format.

For the interviews, audio tapes and notes were taken for each student interview. These notes and the transcriptions of the tapes were used in the analyses. The transcripts were coded following the steps recommended by Saldaña (2015). The text codes were then entered into spreadsheets. The transcribed data was coded using open codes descriptively so as to identify possible key ideas that were later arranged into categories. Categories were then grouped to form themes. A mind map was created to illustrate how the categories were related to each other and connected within the themes (Fig. 1).

The coding process was conducted in four stages. Stage 1 (Searching for Open Coding): The first two authors read through the entire interview data and searched for key ideas or open codes, which formed the basic units of the analysis. Stage 2 (Sorting out Key Ideas): All key ideas were arranged and sorted according to the interview questions respectively. Stage 3 (Grouping into Categories & Themes): All relevant key ideas were re-grouped and color highlighted into their respective categories in a table. Then, the

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 10 of 22

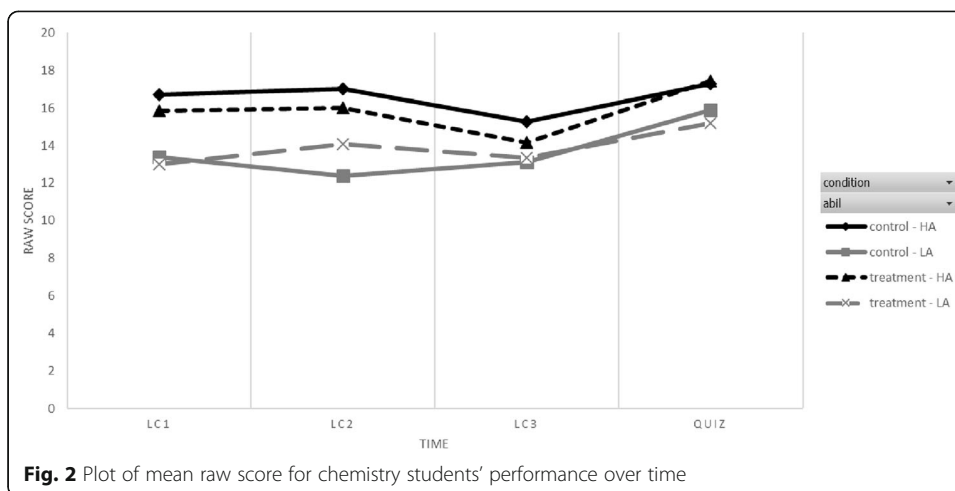**Fig. 1** Mind map for chemistry and mathematics

categories were analyzed and related to different themes in the mind map. Stage 4: The coding procedures and the data were then reviewed and findings clarified for the two neutral parties as described above in Procedures.

## Results

Group comparisons

For the chemistry participants, our repeated measures ANOVA showed that there was a significant main effect of time, $F(3, 87) = 9.067$, $p < 0.001$. However, there was no significant interaction of time with the experimental condition, $F(3, 87) = .385$, $p = .764$, or previous ability, $F(3, 87) = 1.715$, $p = .170$. These results demonstrated that the participants' performances varied for the LCs and quiz, and that the overall finding was consistent for all the participants. Note that the LCs were not cumulative in nature, hence the scores were not directly comparable across the four time points. There was a between-group difference for ability, $F(1, 29) = 11.137$, $p = .002$, showing that the participants of higher ability performed significantly better overall. There was no significant effect of experimental condition, $F(1, 29) = 1.936$, $p = .736$. This indicated that there was no difference for the CS group from the C group, meaning there was no negative effect of providing scores to the CS group after they received the feedback. Figure 2 shows the estimated marginal means of scores on the LCs and quiz for chemistry's participants in both the CS and C groups, and exhibited an overall better performance by the high ability (HA) participants. For chemistry, the challenging question was a compulsory question in LCs and it was an optional question in the graded quiz.

For the mathematics participants, the results were substantially similar to the chemistry group. The challenging question was an optional question in the quiz. This was to align with the format of modular assessment and the consensus among the relevant subject teachers. Not many participants attempted the challenging question in the quiz and hence the participants' ability to handle challenging question was not examined. As

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 11 of 22

Fig. 2 Plot of mean raw score for chemistry students' performance over time

such, we dropped the challenging question across all learning checks and quiz when we analyzed the participants' performance for mathematics. The repeated measures ANOVA showed a significant within-group effect of time, $F_{(3, 69)} = 7.359$, $p < .001$, but as with chemistry there was no significant interaction of time with the experimental condition, $F_{(3, 69)} = 1.250$, $p = .298$, or previous ability, $F_{(3, 69)} = .427$, $p = .734$. Figure 3 shows the estimated marginal means of the scores on the LCs and quiz for the mathematics participants. We also found no significant between-group difference for experimental condition, $F_{(1, 23)} = .495$, $p = .489$, although there was a difference by ability, $F_{(1, 29)} = 20.063$, $p < .001$. Similar with the chemistry groups, this indicates no negative effect of providing scores after the comments had been received. Figure 3 shows the estimated marginal means of the scores on the LCs and quiz for the mathematics participants in both the CS and C groups, and exhibited an overall better performance by the high ability participants.

To conclude, for both chemistry and mathematics subjects, there was no significant difference in the LCs or quiz scores for the C and CS groups. This demonstrates that the approach of providing descriptive comments followed by score did not significantly reduce participants' performance. This was the case regardless of participants' previous level of achievement.

Fig. 3 Plot of mean raw score for mathematics students' performance over time
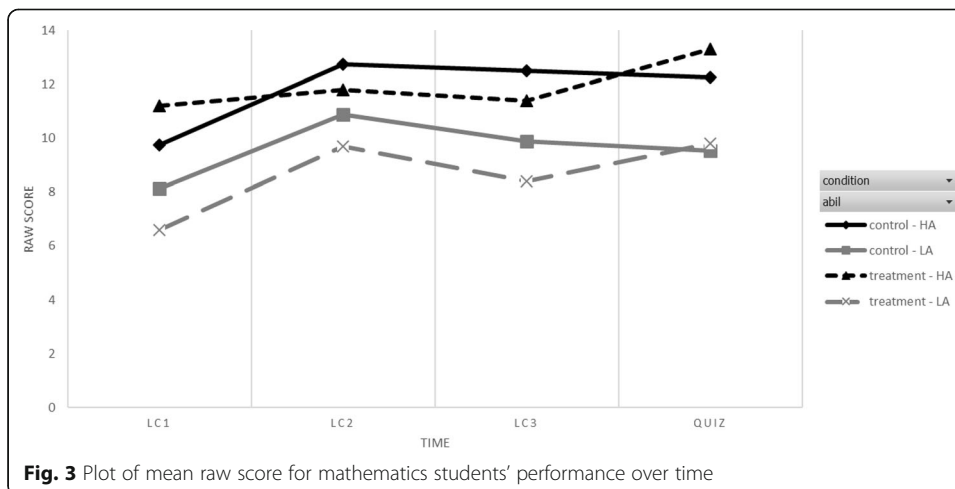
Survey

There were 60 participants involved in the survey and 13 participants selected in the interview sessions. The survey showed that 100% of the participants from the CS group agreed that the feedback order was important in their LCs' learning process (Table 4). Although, the majority of participants in both groups wanted to know the scores on the learning checks, the proportion was higher for the CS group (who received scores) compared to the C group (who did not receive scores) with a slightly lower interest in getting the scores.

Interviews

Our analysis of the interviews yielded several categories and themes (see Table 5) for interpreting the participants' experiences in the action research on feedback order. In both chemistry and mathematics students' interview responses, we observed the themes of metacognition and self-regulated learning. Many participants coherently pointed out that the most useful part in the feedback order was their metacognitive learning in the self-correction process, which was triggered by the teacher's comments. As a result, it enhanced their conceptual understanding in the subject. Identifying and rectifying own mistakes also aided in memory retention in their learning process. Completing the self-correction without teacher coaching would gradually equip them with a better self-directed learning skill in the long run. In chemistry, we also observed the theme of performance-oriented mindset, where the learning check was seen by some merely as a stepping-stone to another assessment rather than an opportunity for self-growth.

Performance-oriented mindset

The performance-oriented mindset is a theme about using the learning check (LC) and the feedback merely as a step toward performance on later assessments. This theme was surfaced in the secondary-one chemistry participants more so than the secondary-four mathematics participants. It could be explained by a greater maturity level among the higher year's mathematics participants who are more inclined towards a learning-oriented mindset.

Chemistry participants shared that they perceived LCs as a useful assessment tool to enhance their assessment performances in the graded tasks. The dialogues clearly showed that they related the academic benefits brought about by the learning checks with an improvement in their assessment performance. They shared several benefits of doing LCs, including exposure to various types of topical practice questions and

Table 4 Survey result for chemistry and mathematics

	Chemistry		Mathematics	
	C group (%)	CS group (%)	C group (%)	CS group (%)
I prefer my teacher to provide me with learning check	100	100	100	100
Written comments provided in learning checks were valuable	100	100	87	92
Scores provided on learning checks were valuable	NA	89	NA	52
Self-correction step helps me in learning process	87	100	78	92
The feedback order was important	100	100	83	100
The feedback order was appropriate	100	94	87	92
I would choose to know my score on my learning checks	73	89	61	70

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 13 of 22

Table 5 Coding Result for Chemistry and Mathematics

Subjects	Categories	Themes
Chemistry	1. Assessment Tool 2. Deeper Learning and Understanding 3. Feedback Order 4. Learning Mindset /Attitudes/ Environment 5. Memory 6. Peer Comparison and Competition 7. Pointed To Mistakes 8. Self-Correction 9. Self-Reflection 10. Thinking Space	1. Metacognition Supported by Comments 2. Self-Regulated Learning 3. Performance-Oriented Mindset
Mathematics	1. Deeper Learning and Understanding 2. Feedback Order 3. Pointed To Mistakes 4. Self-Correction 5. Thinking Space	1. Metacognition Supported by Comments 2. Self-Regulated Learning

improvement in learning retention, which lead to a better performance for upcoming assessments. In fact, the participants were required to go through a few steps in the self-correction process. The steps included reading the teacher's comments, re-analyzing the questions, identifying mistakes, improving their own answers and checking the model answers upon completed the correction. As a result of more time and greater effort in the learning cycles, it increased their learning retention and deepened the understanding of subject. Some students also commented that the score was a good reference for a peer comparison and to gauge their own performance, indicating that some desired receipt of the score for that comparative purpose rather than metacognitive or self-regulating (as discussed below). So, some students preferred to have both comments and score as their feedback.

The chemistry participants also shared their performance concerns on the negative effect of revealing score earlier or together with comments. They commented that knowing scores in advance may lead to a negative psychological mentality, such as a sense of complacency or disappointment, which subsequently affect their self-correction and learning attitude.

Metacognition supported by comments

Flavell (1979) defined "metacognition" as one's self-awareness and judgment of his or her cognitive processes and strategies. It does not only involve thinking but to think about one's own thinking. Perkins (1985) opined that students were able to reason with thinking better when they were probed with guidance. In the survey and interviews, our participants' statements addressed the usefulness of the learning checks as a way to stimulate metacognitive processes. It was prompted by descriptive comments in the form of critical thinking questions and the subsequent self-correction process. For example, a high achiever in mathematics from the C group commented as follows:

> "If you give the model answer first, we'll just copy and if it is self-correction, we make an effort to understand what we did wrong especially the comments. … I remember (and) I learn a lot better that I am asking question for trying to understand (Mathematics student 1)."

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 14 of 22

Similarly, a chemistry participant from the low-achiever group recognized that the descriptive comments helped him to reflect on mistakes:

> "If we think about it, we can identify where we went wrong… and it will lead us to redo that part…It gives us some thinking room, unless, we really don't know the answers. … We can think about how we think again, how we used to think…If we were given the direct answers straight away, it is a bit hard for us to think where we went wrong. Think deeper first and where you have made mistake and try to improve, before the suggested answers are given to you…(if not), you don't even need to think…(Chemistry student 1)."

These comments emphasized the usefulness of the descriptive comments as pushing to students to look back at their reasoning to find missteps and immerse themselves in the mistake-learning process, rather than simply providing a suggested answer for the participants to copy.

When feedback only indicates the correctness of a response, learning is lower than with more informed descriptive feedback (Bangert-Drowns et al., 1991). Participants learn more when they are provided with detailed comments on their answers rather than correct or wrong (Wiliam, 20 2010). One of our low-achiever in mathematics from the C group summed it up best by saying,

> "…It allows us to think rather than copying. If we just copy from answer, we don't learn much… If answer is provided before self-correction, this is just an example of copying from the board and not beneficial. It defeats the purpose of comments if answer is given before self-correction (Mathematics student 5)."

> Another mathematics low-achiever from the CS group added that,
> "Comments help to sort of get me thinking about the method and solution (Mathematics student 7)."

A high achiever's chemistry student from the CS group also agreed with the importance of thinking process, added that,

> "Thinking of our mistake (allow us) to re-learning and reviewing of the concept is very important (Chemistry student 2)."

The support of feedback and the space provided for self-correction allow students to recognize and learn from one's own mistakes contribute greatly towards an emotional satisfaction in the mistake-learning process. Hence, the self-correction not only supports meta-cognition but may further improve attitude or motivation for learning. This is evidently supported by a statement from a low achiever in chemistry from the CS group, who shared her feeling upon completing her correction,

> "The feeling after you manage to find out what's wrong…the feeling is very nice (Chemistry student 3)."

Self-regulated learning
The third theme of self-regulated learning was related to metacognitive awareness but distinguishable from it. The construct of self-regulation referred to the degree to which

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 15 of 22

students can regulate aspects of their thinking, motivation and behaviour during learning process (Pintrich and Zusho, 2002). The theme also reflected a strong initiative and willingness among the participants to take a greater ownership in self-learning and self-correction steps instead of waiting to be spoon-fed with teacher's suggested answers. A low achiever in chemistry from the C group explained that the self-correction process helped him to be an independent learner instead of relying on teachers.

> "Self-correction and self-checking allow you to think about it. You don't really depend on teacher to see where is the problem. Teacher will first at least let you try…let you do correction on your own using just some more clues…(Chemistry student 4)."

In fact, learning checks as a formative tool, which combined with comments and self-correction process empowers students as self-regulated learners.

Both interview and survey findings indicated that most of the participants would welcome their score results if the scores were provided. Nevertheless, they agreed that the score itself was not of utmost importance in the learning progress. This was clearly reflected in the following comments.

This is a comment made by a CS group low-achiever in mathematics:

> "In the end, I can roughly judge how much I know, how much I got right, how much I got wrong, what I got right, what I got wrong. I don't think that score is important (Mathematics student 7)."

> Two high achievers in chemistry also made a similar opinion and shared that, "Learning from mistake and trying to correct the mistakes, I think it is more important than knowing the scores (Chemistry student 2)."

> "The score doesn't signify anything. It doesn't matter about the score. It just matter about what mistake you make (Chemistry student 5)."

On the contrary, the score would become important to participants if the tasks were counted towards the grades—as it has a direct impact on the academic performance and achievement. It is shared by a low-achiever in mathematics from the CS group that,

> "…I find that when it comes to graded, I think we care a lot more for the score (Mathematics student 7)."

It was completely understandable when a high achiever in mathematics from the C group also shared her stressful feeling towards a graded task that,

> "I would probably feel stressed up and needs to try to get full mark and everything and would feel like graded assignment rather than just practicing questions. I think it is better if ungraded because it is just testing your understanding and we don't feel like competing with our friends to get high marks (Mathematics student 1)".

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 16 of 22

Nevertheless, the participants clearly recognized that targeting on learning goals will naturally provide a similar route to performance goals. This was reflected in a comment made by a chemistry participant from the CS group that,

> "The most important thing is that we learn the topic, we understand more. If we understand the topic well, the score will just come...(Chemistry student 6)."

By delaying the release of scores until the learning cycle in the self-correction process were completed, our findings showed no impact on participants' learning performances or negative effects on learning as a result of releasing scores. This could be attributed to the nature, notion, and objectives of learning check itself, which focused on the learning process, rather than the achievement of score. In the participants' perception, score was not the main focus in a non-graded task. As such, the participants had placed a premium on the learning or mastery goals as compared to the performance goals. Generally, the participants were not only aware of the value of comments, they also noticed the benefit of the delay in revealing scores until after self-correction and comments. As chemistry participants said,

> "If we get our score at first, then if we get a high score, let's not do correction! (Chemistry student 5)".

> "If they (students) score very badly, they will be depressed and would not concentrate on their mistakes, so (revealing) score at the last (step) is better...(Chemistry student 6)".

Therefore, the delay of releasing the score appeared to play a meaningful part in helping the participants to engage and concentrate in self-correction. Once the learning gains were achieved in the self-correction process, the score will no longer appear to be critical anymore.

However, an exceptional response came from a mathematics low-achiever in the C group who shared that he preferred to figure out his mistakes through self-correction without the help of teacher's descriptive comments. This participant said,

> "Not really I think. Because that mostly questions I already asked myself. Just mark it on then next time I know where I went wrong. Then during correction, I can see what I can do to make it correct (Mathematics student 6)".

For this participant, he was confident in his ability to correct his mistakes but yet, he did agree that the time provided for them to do self-correction in class enabled him to look for his mistakes. Generally, he felt that learning check is good for checking understanding but he would prefer a score rather than a teacher's comments.

Discussion

In this action research, our team of teachers worked across subject areas to study the possibility of achieving the benefits of descriptive feedback while still being able to reveal scores as expected by many students, parents, and others. Our study revealed

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 17 of 22

that there was no negative outcome for students in classes where a score was given but delayed after providing descriptive comments and a self-correction process.

Previous research found that over-emphasis on scores would lead to negative effects on learning (Crooks, 1988; Kohn, 1994). Similarly, Lipnevich and Smith's analysis (2009) contended that receiving a tentative grade depressed students' performance. A detailed descriptive feedback was found to be most effective when given alone, unaccompanied by grades or praise (Lipnevich and Smith, 2009). Butler and Nisan (1986) and Butler (1988) also posited that the effect of comments could be undermined by a negative motivational effect of providing grades and scores (Butler, 1988).

Why is giving the score not detrimental to participants in our study? Firstly, the negative effect of overemphasis on score was not manifested with the order of comments followed by revealing scores. The scores were released only after the participants completed the learning cycles. The feedback provided with a delay in revealing scores was crucial in enabling participants to rethink their mistakes and to stay focused in the self-correction process. As a result, this should enhance the learning impact and dilute the negative effect of scores substantially. The reflective and metacognitive steps are foregrounded for this task and attracted most of the participants' attention and effort. The increment in effort and time spent on the learning task itself then further off-set the negative effect of scores. If the evaluative feedback of scores was released earlier then this could have adversely highlighted the participants' relative standing compared to their peers and, thus, could have further created a performance orientation instead of learning orientation.

However, in our studies, a non-comparative and descriptive feedback was being released first and prioritized in the order. It channeled the focus on improving their answers, increased the likelihood of targeted towards learning orientation. This feedback order indirectly emphasized the use of feedback for self-learning, self-assessment and to encourage learning through corrections instead of learning to pass the scores.

Secondly, when it was a non-graded task, the participants were inclined to focus more on learning or mastery goals instead of performance-goals. In other words, score was perceived to be less important in a non-graded task. The learning objective was naturally placed on the process itself rather than obtaining high scores. In addition to the finding that our participants prefer to have written descriptive comments followed by scores, participants' comments showed that they were consciously aware of the benefits of mastering the concepts through the learning checks, which could help them perform well for later assessments.

Our study was consistent with the previous research works reported by Kulhavy (1977), Sadler (1989), and Wiliam (2011) on what is considered an effective feedback. The impact of mistake-correction was enhanced with a delay between comments and students' response (Kulhavy and Stock, 1989). They conjectured, and we observed here, that students' learning is enhanced if they are provided with time to correct their mistakes. Our findings were also consistent with Hattie and Timperley (2007) on the usefulness of mistake detection and self-feedback process. We see our findings corroborating an insight from Elshout-Mohr (1994), who suggested that providing correct answers was only useful for simple tasks or core questions but less useful for more complex tasks. Interview findings also pointed out that vaguely telling them that they need to improve or giving only score information turns out to be less helpful in the learning process.

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 18 of 22

Challenging questions encompassed the development of new learning capacities and strategies rather than practicing more similar questions (Elshout-Mohr, 1994). Hence, in order to solve a more complex and higher-order thinking question, participants would often require a dialogic feedback from teacher to improve their understanding (Elshout-Mohr, 1994).

Through useful self-identification and self-correction skills, learners could seek better ways to complete the task or solve the problems independently and to acquire a more in-depth understanding in the relevant topic. As a result, it further enhances their self-regulatory proficiencies (Hattie and Timperley, 2007). Once learners develop mistake detection skills, these skills could help guide them through a cycle of self-feedback independently. Our study was not long enough to follow students into the full development of mistake detection and self-feedback, but the early indications from interviews are promising.

Our study has shown that it may be possible to provide sores and grades to students but still to focus most students' attention on the feedback that will help them improve more. Our study suggested an alternative "middle way" that offers students the benefits from teachers' written comments while still releasing the numerical score result that is often requested by students, parents, and other stakeholders. More importantly, the negative effect of overemphasis on score was not observed in this study. This does not directly contradict previous work on the negative role of scores (Butler, 1988; Crooks, 1988; Kohn, 1994; Lipnevich and Smith, 2009), but rather provides a way to circumvent the negative effects by emphasizing first the descriptive feedback and learning through mistakes. This is consistent with previous work that showed how learning is enhanced with detailed comments rather than only marking for right/wrong response (Bangert-Drowns et al., 1991; Wiliam, 2010). Furthermore, as our interviews and survey findings showed, the feedback seemed to work for most of our students because it focused them on attending to their thinking (Flavell, 1979), taking ownership for their work (Pintrich and Zusho, 2002) and getting motivated to perform through the self-correction (Salomon and Globerson, 1987) rather than focusing on the relative score results.

In our study, participants demonstrated their strong willingness and enthusiasm in the self-correction process, which evidently showed that they focused more on learning goals. It is consistent with Brophy (1983)'s findings that learning or mastery goals are associated with "a strong learning motivation" whereby participants concentrate on mastering and understanding of content knowledge. Learning will be improved if participants are able to monitor and self-assess the quality of their work when they perform a task (Sadler, 1989). As such, it is important to develop students' self-regulation and independence in learning during the process of knowledge acquisition.

Critical thinking questions were adopted in the descriptive comments because all participants were perceived as able learners, who having a proactive rather than a passive role in generating and using feedback. These were closely aligned with Nicol and Macfarlane-Dick (2006) 's finding that feedback is effective only when it is utilised proactively by learners. As such, we posed critical thinking questions in comments to trigger metacognitive thinking through mistakes identification and self-correction, instead of spoon-fed them directly with model answers. Our study shows that giving descriptive feedback with clear indicative clues helped students recognize and learn from their own mistakes faster, and this contributed greatly towards emotional satisfaction and

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 19 of 22

self-achievement in the learning and self-correction process. Research finding by Butler and Nisan (1986) has further shown that intrinsic motivation in the learners depends on the interaction between the individual and the stimulus. In effect, descriptive feedback actually improves the participants' learning because it evokes their mindfulness and motivation in completing the learning check task. A similar finding was reported by Salomon and Globerson (1987).

Despite the overall promise of our findings, we did see some students whose interviews focused on the possible performance (more associated with achieving a score) rather than mastery (more associated with self growth). Participants' comments on this point were consistent with Lipnevich and Smith's (2009) analysis showing that receipt of a tentative grade depressed the performance of the learners. Their findings showed that the grades had potential negative influence on students' self-efficacy beliefs and elicited negative effect around the task. At the same time, grades also decreased the effect of detailed feedback (Lipnevich and Smith, 2009). On the other hand, the empirical base for the argument on the effect of comments with grade was not uniformly consistent. Conversely, Smith and Gorard (2005) reported that students received grades and comments outperformed those that received comments only.

Limitations

Our study was conducted in a Mathematics and Science specialized independent secondary school, which has a unique education setting. Different from most of the high schools in Singapore, this school has been provided autonomy and resources to craft its own curricula. Besides that, students are also able to skip the high stakes O and A-levels at the national exams. Hence, our findings may not translate directly to mainstream schools, and further study is needed to understand how the insights here could be extended to mainstream educational institutions in Singapore or international institutions.

The participants were divided into two different groups of C and CS groups. Each group was further separated into two subgroups of HA and LA to achieve a fair and parallel comparison between the participants with similar learning abilities. Therefore, it led to a rather limited sample size for these two sub-groups. Similarly, the research and data collection of scores were conducted over a rather short 4-week period to ensure a fair comparison for a particular selected topical subject within the experiment period. A different and new topic may introduce variation in difficulty level and different marking systems. In order to minimize the possible variations for the topical marking process such as mark scheme and assessment criteria in the given tasks, a similar topical subject was used throughout the study.

Another consideration for the future work is to understand the limitations on teachers' time to engage in the marking, commenting, and guidance to students. With heavy marking and teaching loads, as well as the pressure to return worksheets with timely feedback, writing descriptive comments on every student's worksheet seems a gargantuan task. Hence, further research is warranted to focus on strategies, feasibility, and effectiveness of providing descriptive comments in a simplified form to aid teachers in marking while still demonstrating student benefits. If some types of marking code were used, for example, students and teachers would need training on the code and its interpretation and application.

Conclusion

Our findings indicated that the approach of providing descriptive comments followed by score did not show any significant difference as compared to providing descriptive comments alone. No negative motivational effect was observed where a score was given but delayed after providing descriptive comments and a self-correction process. Our study proposed an alternative "middle way" that offers students the benefits from teachers' written comments while still releasing the numerical score result that is often requested by students, parents, and other stakeholders. With this sequential feedback order, the metacognitive and reflective steps involved in self-correction were emphasized in a learning task. The scores were released only after the participants completed the learning cycles.

These findings also further implied that the teacher may not need to spend additional time to compute scores on each learning worksheet. Computing score was more for academic sorting, recording and monitoring purposes. Rather, teacher's written comments which pinpointed what students' mistakes were and guided them on how to improve and overcome their own learning difficulties was proven to be a more useful and effective tool. The descriptive comments coupled with a self-correction process not only enhanced participants' metacognitive and reflective skills, but also improved their learning satisfaction and the retention of learning concepts.

In this study, the first two authors reflected on the lessons learned and implications of teaching from this study. As teachers, we may assume that students are able to understand answers immediately when solutions and corrective measures are directly proposed to them after marking their worksheets. In fact, when solutions are presented directly to students, there is lack of emphasis on students' mistake identification, rectification and re-thinking process in the learning cycle. It is important for teachers to take time and invite students to rethink upon their own mistakes; and engage in the process of making meaning and self-correcting these mistakes. A reflective learning space encourages students to analyze their mistakes, to enhance their metacognitive skills, and to gain a deeper understanding of the concepts through personal reflection.

In fact, an assessment provides an invaluable feedback to both teachers and students if they are reflective learners. Teachers could go beyond scores and conduct an in-depth diagnostic analysis on test results, in order to understand students' learning difficulties. Simply administering more assessments and recording scores, teachers actually miss out these critical learning points from assessment aspects. So, more carefully reviewing and giving comments on assessment can be extremely valuable. Teachers should proactively apply their metacognitive skills and engage in the process of rethinking and improving their assessments and teaching pedagogies. Making students' learning process more visible will help teacher to achieve self-feedback and self-improve on their own teaching practices and instructional strategies.

Besides gaining invaluable research experiences through this study, we were pleasantly surprised to find on this inquiry journey a greater and closer relationship between teacher-researchers and participants. We were heartening to observe that the participants embraced these learning check experiences with an attitude of gratitude and appreciation for the researchers' effort in crafting and enhancing their learning journey in the module.

Chua *et al. Asia-Pacific Science Education* (2017) 3:4

Page 21 of 22

Despite of contextual differences between a specialized and local mainstream schools, our findings on the feedback order that showed no negative effect in releasing scores to students and, yet enabling them to engage in self-learning process, may be used as a reference for instructional strategy in classroom learning for local schools.

Authors' contributions
The first two authors, HLC and SHL, both conducted the data collection in chemistry and mathematics classroom from their participant students, respectively. GWF provided advice on the entire research study. All authors contributed in the research design, drafted and approved the final manuscript.

Authors' information
Hooi Ling Chua is a teacher with more than 10 years teaching experiences in O-level Chemistry. She started her teaching career in an international high school in Malaysia for 4 years and followed by NUS High School of Mathematics and Science, Singapore for about 7 years. She completed her Master in Material Chemistry from Universiti Putra Malaysia in Malaysia and Master in Education (Curriculum and Teaching) from the National Institute of Education, Nanyang Technological University, Singapore.
Sheau Huey Lee is an experienced mathematics teacher cum assistant head of Mathematics Department at NUS High School of Mathematics and Science, Singapore. She has 14 years of teaching experiences in mathematics from secondary one level to pre-university level. She received her Master in Science (Mathematics) from National University of Singapore and Master in Education (Curriculum and Teaching) from the National Institute of Education, Nanyang Technological University, Singapore.
Gavin W. Fulmer is an Assistant Professor of Science Education at the University of Iowa. His research focuses on assessment and its implications in science education, including applications of Rasch measurement models and alignment analyses for assessments and standards. Previous affiliations include the National Institute of Education Singapore (NIE), U.S. National Science Foundation (NSF), and Westat. He holds a Ph.D. in Science Education from the University at Buffalo S.U.N.Y. and a B.S. in Mathematics and Physics from the University of Redlands.

Ethics approval and consent to participate
The study was reviewed and approved for ethics clearance by the internal committee as authorized by NTU-IRB for graduate student research projects. Parents could opt not to include their students in the study, and all student participants provided informed assent prior to data collection.

Competing interests
The authors declare that they have no competing interests to report.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]NUS High School of Mathematics and Science, 20 Clementi Avenue 1, Singapore 129957, Singapore. [2]The University of Iowa, N259 Lindquist Center, Iowa City, IA 52242, USA.

References
Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Am Educ Res Associat, 61*(2), 213–238.
Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.
Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan, 80*(2), 139–144 & 146-148.
Brookhart, S. M. (2011a). *Grading and learning: Practices that support student achievement*. Bloomington: Solution Tree Press.
Brookhart, S. M. (2011b). Starting the conversation about grading. *Educational Leadership, 69*(3), 10–14.
Brophy, J. E. (1983). Fostering student learning and motivation in the elementary school classroom. In S. Paris, G. Olson, & H. Stevenson (Eds.), *Learning and motivation in the classroom*. Hillsdale: Erlbaum.
Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British J Educ Psychol, 58*(1), 1–14.

Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *J Educ Psychol, 78*(3), 210–216.

Creswell, J. W. (2014). Mixed methods designs. In J. W. Creswell (Ed.), *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed., pp. 563–606). Harlow: Pearson Education Ltd..

Crooks, T. J. (1988). The impact of classroom evaluation on students. *Rev Educ Res, 5*, 438–481.

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework. *J Educ Psychol, 77*, 162–173.

Elder, L., & Paul, R. (2006). *Critical thinking: Concepts and tools*. Retrieved from https://www.criticalthinking.org/files/Concepts_Tools.pdf.

Elshout-Mohr, M. (1994). Feedback in self-instruction. *European Education, 26*(2), 58–73.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental enquiry. *American Psychologist, 34*(10), 906–911.

Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practices, 11*(3), 35–42.

Guskey, T. R. (2014). *On your mark: Challenging the conventions of grading and reporting*. Bloomington: Solution Tree.

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks: Corwin.

Guskey, T. R., & Lee, A. J. (2013). *Answers to essential questions about standards, assessments, grading, and reporting*. Thousand Oaks: Corwin.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Rev Educ Res, 77*(1), 81–112.

Ho, D. Y. F., & Kang, T. K. (1984). Intergenerational comparisons of child-rearing attitudes and practices in Hong Kong. *Dev Psychol, 20*, 1004–1016.

Ho, K. C., and Yip, J. (2003). *YOUTH.sg: The state of youth in Singapore*. Singapore: National Youth Council.

Isralowitz, R. E., & Ong, T. H. (1990). Singapore youth: The impact of social status on perceptions of adolescent problems. *Adolescence, 25*, 357–362.

Joyce, T. B., & Shirley, Y. (2011). Academic expectations as sources of stress in Asian students. *Soc Psychol Educ, 14*(3), 389–407.

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. *Psychol Bull, 119*(2), 254–284.

Koh, K., & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy and Practice, 16*(3), 291–318.

Kohn, A. (1994). Grading: The issue is not how but why. *Educational Leadership, 52*(2), 38–41.

Kulhavy, R. W. (1977). Feedback in written instruction. *Rev Educ Res, 47*(1), 211–232.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educ Psychol Rev, 1*(4), 279–308.

Lim, L. (2013). Meritocracy, elitism, and egalitarianism: A preliminary and provisional assessment of Singapore's primary education review. *Asia Pacific Journal of Education, 33*(1), 1–14.

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *J Exp Psychol: Applied, 15*(4), 319–333.

Luke, A., Freebody, P., Shun, L., & Gopinathan, S. (2005). Towards research-based innovation and reform: Singapore schooling in transition. *Asia Pacific Journal of Education, 25*(1), 5–28.

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

Ory, J., & Ryan, K. E. (1993). *Tips for improving testing and grading*. Newbury Park: Sage.

Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *J Educ Psychol, 49*, 173–181.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *J Educ Psychol, 77*(5), 562–571.

Pintrich, P. R., & Zusho, A. (2002). Student motivation and self-regulated learning in the college classroom. In J. C. Smart & W. G. Tierney (Eds.), *Higher education: Handbook of theory and research (vol. XVII)*. Agathon Press: New York.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Thousand Oaks: Sage.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *Int J Educ Res, 11*, 623–637.

Smith, E., & Gorard, S. (2005). They don't give us our marks: The role of formative feedback in student progress. *Assessment in Education Principles Policy & Practice, 12*, 21–38.

Sutton, R. (2010). Making formative assessment the way the school does business: The impact and implications of formative assessment for teachers, students and school leaders. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change, volume 23 of the series Springer international handbooks of education* (pp. 883–899). Netherlands: Springer.

Tan, K. H. K. (2011). Assessment for learning in Singapore - unpacking its meanings and identifying some areas for improvement. *Educ Res Pol Pract, 10*(2), 91–103.

Tan, K. H. K., & Deneen, C. C. (2015). Aligning and sustaining meritocracy, curriculum and assessment validity in Singapore. *Assessment Matters, 7*(1), 31–52.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educ Res J, 22*(4), 389–404.

Wiliam, D. (2010). *The role of formative assessment in effective learning environments, in the nature of learning: Using research to inspire practice*. Paris: OECD Publishing http://dx.doi.org/10.1787/9789264086487-8-en.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3–14.