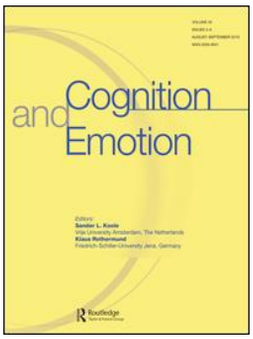

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: www.tandfonline.com/journals/pcem20

How to bring the news ... peak-end effects in children's affective responses to peer assessments of their social behavior

Vincent Hoogerheide, Marleen Vink, Bridgid Finn, An K. Raes & Fred Paas

To cite this article: Vincent Hoogerheide, Marleen Vink, Bridgid Finn, An K. Raes & Fred Paas (2018) How to bring the news ... peak-end effects in children's affective responses to peer assessments of their social behavior, Cognition and Emotion, 32:5, 1114-1121, DOI: 10.1080/02699931.2017.1362375

To link to this article: https://doi.org/10.1080/02699931.2017.1362375

Published online: 02 Aug 2017.

COGNITION AND EMOTION
2018, VOL. 32, NO. 5, 1114–1121
https://doi.org/10.1080/02699931.2017.1362375

Routledge
Taylor & Francis Group

BRIEF ARTICLE

How to bring the news … peak-end effects in children's affective responses to peer assessments of their social behavior

Vincent Hoogerheide[a,b], Marleen Vink[b], Bridgid Finn[c], An K. Raes[b] and Fred Paas[b,d]

[a]Department of Education, Utrecht University, Utrecht, the Netherlands; [b]Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands; [c]Educational Testing Service, Princeton, NJ, United States; [d]Early Start Research Institute, University of Wollongong, Wollongong, Australia

ABSTRACT

The retrospective evaluation of an event tends to be based on how the experience felt during the most intense moment and the last moment. Two experiments tested whether this so-called peak-end effect influences how primary school students are affected by peer assessments. In both experiments, children (ages 7–12) assessed two classmates on their behaviour in school and then received two manipulated assessments. In Experiment 1 ($N = 30$), one assessment consisted of four negative ratings and the other of four negative ratings with an extra moderately negative rating added to the end. In Experiment 2 ($N = 44$), one assessment consisted of four positive ratings, and the other added an extra moderately positive rating to the end. Consistent with the peak-end effect, the extended assessment in Experiment 1 and the short assessment in Experiment 2 were remembered as more pleasant and less difficult to deal with, which shaped children's peer assessment preferences and prospective choices of which assessment to repeat. These findings indicate that the process of peer assessment can be improved by ending the feedback with the most positive part of the assessment.

ARTICLE HISTORY
Received 12 January 2016
Revised 20 July 2017
Accepted 24 July 2017

KEYWORDS
Peak-end rule; peer assessment; memory; remembered utility; hedonic evaluation

The frequency with which students' knowledge, skills, and behaviors are assessed in schools has been increasing in many countries over the past decades (Abrams, Pedulla, & Madaus, 2003). One popular method of assessment is peer assessment, which was defined by Topping (1998) as "an arrangement in which individuals consider the amount, level, value, worth, quality or success of the products or outcomes of learning of peers of similar status" (p. 250). Peer assessment can be used for students of all ages, can be applied in a formative and summative manner, and has been associated with a wide range of benefits (Dochy, Segers, & Sluijsmans, 1999; Sluijsmans, Brand-Gruwel, & Van Merriënboer, 2002).

There are caveats to peer assessment however, with arguably the most important one being the negative affect that can result from providing peers with feedback and receiving feedback from peers – the core components of peer assessment – because both processes can be difficult and stressful for students to deal with (Hanrahan & Isaacs, 2001; Stagg, 2016). In prior research, unpleasant experiences with peer assessment have been linked to a decrease in enjoyment, motivation, well-being, and to damaging the relationship between students (Levine, 2008; Levine, Kelly, Karakoc, & Haidet, 2007). And because task enjoyment is an important component of motivation and linked to performance (e.g. Elliot & Harackiewicz, 1994), the low levels of enjoyment and motivation during subsequent peer assessments would likely affect the quality of the feedback and the learning from the feedback. Against this background, Stagg (2016) recently initiated a call for more research into this "dark side" of peer assessment.

CONTACT Vincent Hoogerheide v.hoogerheide@uu.nl

This necessity for more research is probably most needed for primary schools students, because they likely struggle the most with the process of peer assessment. Not only do primary school students lack experience with generating constructive feedback and receiving feedback, they can be extra sensitive to the opinion of others, especially their peers (Jacobs, Bleeker, & Constantino, 2003). Moreover, whereas in higher education peer assessment almost solely concerns academic competencies (e.g. math, writing), in primary school peer assessment is a commonly used tool to train and assess students' social competencies (Warnes, Sheridan, Geske, & Warnes, 2005). The underlying reason for this is that developing children's social competencies is an important aim of primary school, and those who lack these vital social skills are "often are rejected by their peers, have trouble interacting with their teachers and families, and have emotional difficulties" (Warnes et al., 2005, p. 173).

In sum, peer assessment has an important role in educational practice, but students, particularly younger children, may be negatively affected by its core components. Therefore, it is an important question for researchers and educators if there are methods that can help prevent peer assessment from becoming an unpleasant experience or even help make the process of peer assessment more enjoyable.

One possible answer to this question arises from research on the retrospective pain or pleasure associated with past experiences (i.e. remembered utility; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). It has been shown that when people form retrospective evaluations of an experience, they tend to neglect the length of an experience (i.e. duration neglect; Fredrickson & Kahneman, 1993). Instead, our memory is predominantly based on how we felt during the most intense moment (the peak) and when the experience ended. A classic study of Kahneman et al. (1993) provides an example of this so-called *peak-end effect*. Students were exposed to two painful cold water episodes, a short episode during which students kept one hand immersed in 14°C cold water for 60 s (the peak), and an extended episode that started identically (with the same peak) but added an interval of 30 s to the end during which the cold water gradually rose one degree in temperature. The extra interval extended the duration of the pain experienced, making it a more unpleasant experience as a whole. However, the extended experience ended relatively more pleasantly, which evoked a biased *memory* of the experience. Participants remembered the extended episode as less painful and as less difficult to cope with, which led a significant majority of participants to prefer to repeat the extended episode for their next trial.

Peak-end effects have also been demonstrated with pleasant experiences. Diener, Wirtz, and Oishi (2001) investigated how people evaluated various short stories about fictitious people's lives and found that people preferred stories that presented a shorter life that was very pleasant overall (i.e. the peak) over a longer life that was also very pleasant (the same peak) but in which there were mildly pleasant (rather than very pleasant) years experienced at the end. Similarly, Do, Rupert, and Wolford (2008) found that children were more pleased after receiving a highly enjoyable Hershey chocolate bar alone than when they received the same chocolate bar followed by a mildly enjoyable piece of bubble gum. Although it would have been more sensible for children to prefer two treats over one because receiving the gum added pleasure to the *experience*, they preferred to receive only a chocolate bar because their retrospective evaluations were most sensitive to the end of the experience, causing the entire experience to be *remembered* as being less pleasurable.

In sum, it has been shown in a multitude of situations that retrospective evaluations of unpleasant and pleasant experiences rely more heavily on the peak and end moments of the experience (e.g. Chajut, Caspi, Chen, Hod, & Ariely, 2014; Redelmeier, Katz, & Kahneman, 2003), probably because both the temporal integration and retrieval of experiences are prone to error (Miron-Shatz, 2009). These findings beg the question whether the same mechanism applies to contexts relevant for learning. Recent studies showed that this can indeed be the case for students studying foreign vocabulary for a test (Finn, 2010; Hoogerheide & Paas, 2012) and for students taking a difficult math test (Finn & Miele, 2016).

The present study

Two experiments addressed the question of whether primary school students' retrospective evaluations of unpleasant (Experiment 1) and pleasant peer assessments (Experiment 2) would be sensitive to the end of the assessment, and if so, whether this memory bias would translate to students' (future) preferences and choices. In both experiments, primary school

students first assessed two classmates on their social behavior and then "received their results". While students were led to believe that the results originated from the assessments filled in by two of their peers, in actuality the ratings were manipulated in accordance with the paradigm of Kahneman et al. (1993), meaning that we matched the peak at the start of the experience and varied the end of the experience. In Experiment 1, one assessment consisted of four negative ratings ("short assessment") and one of four negative ratings followed by one moderately negative rating ("extended assessment"). In Experiment 2, the negative ratings were replaced with positive ones to investigate effects of receiving positive assessments (i.e. short assessment: four positive ratings; extended assessment: four positive ratings followed by one moderately positive rating).

We chose to focus on primary school students because, as described, this group likely struggles most with peer assessment. Moreover, whereas the vast majority of studies investigated peak-end effects with adolescents or adults, only Hoogerheide and Paas (2012; studying vocabulary) and Do et al. (2008; receiving treats) investigated if children's memories of experiences are also vulnerable to the peak-end effect bias, but in these two studies, the task was not particularly emotionally salient. With adults, the "most impactful" peak-end effects have been found with (highly) emotional experiences such as giving birth (Chajut et al., 2014). Children, however, remember emotional salient events different than adults do (Leippe, Romanczyk, & Manion, 1991), and therefore the (strength of the) peak-end effect may apply differently for them. We chose a social behavior peer assessment because it represents an educationally relevant and emotionally salient situation for children.

Based on the findings reviewed above, we hypothesised that students would retrospectively evaluate the extended list in Experiment 1 (unpleasant assessment) and the short list in Experiment 2 (pleasant assessment) as more pleasant and less difficult to deal with than the short and extended list, respectively, because these assessments end with the most pleasant feedback. This, in turn, could lead to a preference for repeating that assessment immediately after and the next day. We also explored whether students' assessment preferences would influence choices outside the assessment context by asking which classmate (A or B) they would rather collaborate with on a task. To examine if the short and extended assessments really differed in the affect they induced, we asked participants immediately after each assessment (i.e. experienced utility) how pleasant the assessment was for them and to which degree they would want to receive the assessment again (i.e. the assessment supposedly provided by that peer). Lastly, to explore whether students were aware of the differential duration of the assessments, they were asked to indicate which assessment was longer.

Experiment 1

Method

Participants and design

Experiment 1 was conducted in a primary school in the Netherlands. The parents or caretakers of 55 students were contacted, of whom 38 gave their written informed consent. Four children were excluded based on their scores on an anxiety and/or depression questionnaire (see below), and four children were ill at the time of experiment. The final sample consisted of 30 students ($M_{age} = 9.90$, $SD = 1.81$; 13 boys). Using a within-subjects design, all students received a short peer assessment and an extended peer assessment, the order of which was counterbalanced. In the short peer assessment, students were led to believe that a peer student had given them four negative ratings. In the extended peer assessment, students were led to believe that a peer had given them four negative ratings followed by one moderately negative rating. The study was approved by the Ethics Committee of the Institute of Psychology of the Erasmus University Rotterdam.

Materials

Anxiety and depression questionnaire

To ensure that no student would be harmed by receiving the manipulated peer assessments, all students first completed the Child Depression Inventory (CDI; Timbremont & Braet, 2002) and the Spence Children's Anxiety Scale (SCAS; Spence, 1998). The CDI is a commonly used self-report measure of depressive symptoms for children. It consists of 27 items that measure cognitive, affective, and behavioural symptoms of depression. Each item is scored on a three point scale (0 = absence of symptom, 1 = mild symptom, 2 = clear symptom) and the total score is achieved by summing the obtained scores on the 27 items. Secondly, the SCAS measures the frequency to which young people experience symptoms of

different forms of anxiety. It consists of 45 items that ask for an indication of the extent to which each item occurs in daily life on a four point scale (0 = never occurs, 1 = sometimes occurs, 2 = often occurs, 3 = always occurs). The total SCAS score is achieved by summing 40 of the 45 items (5 items serve as a control).

Peer assessment

A three-page booklet was created for the peer assessments. The first page consisted of an example question: "How well do you think that … can share with other children in the class?". The second and third page consisted of the short and extended peer assessment. The order of these peer assessments was counterbalanced. The names of the two classmates that students would (anonymously) assess were written at the top of the page. Five questions were used for the peer assessments, such as: "How well do you think that [name] communicates with other children in the class?" and "How well do you think that [name] follows the rules?". Beneath each question, the words "insufficient", "insufficient to sufficient", "sufficient", "sufficient to good", and "good" were written. The extended assessment presented all five questions. One question was randomly omitted for the short assessment. The order of the questions was counterbalanced for both assessments.

Receiving the peer assessments

Two A4-papers were created that mimicked the short and extended peer assessments as if they were filled in by two students. In the short assessment, all four questions were marked with "insufficient", while for the extended assessment, the first four questions were marked with "insufficient" and the last item with "insufficient to sufficient". To get an indication of how students experienced these assessments (i.e. experienced utility), an extra A4-paper was created containing two questions, namely "How pleasant was it to receive this assessment?" and "To which degree would you like to receive this assessment again?". Both questions could be answered on a VAS-scale of 0 (not at all/entire unpleasant) to 100 (very much so/entirely pleasant).

Filler task

To ensure that the final questionnaire was completed using information in long-term memory, a filler task was created consisting of mathematical equations.

Retrospective evaluations

Another A4 paper was developed to measure students' remembered utility and prospective behaviour. The questions were: "Which assessment was the least pleasant to receive?", "Which assessment was the most difficult to deal with?", "Which assessment was longer?", "If you would get a third assessment, which assessment would you prefer to receive?", "If you would get another assessment tomorrow, which one would you prefer to receive?", and "with which classmate would you rather work together?". These questions could be answered by marking either the box "assessment 1" or "assessment 2", with the exception of the last question, which could be answered with the classmate that provided assessment 1 or 2. We recoded students' answer to the remembered utility and prospective behaviour questions to a preference for either the "short assessment" or the "extended assessment".

Procedure

Prior to the experiment, a letter was sent to the caretakers of all students from one primary school, containing basic information about the experiment and a consent form. Students who had not received consent, worked on their regular course work during the two experimental sessions.

The first session took place in the students' classroom with two experimenters and a teacher present. Before students completed the CDI and the SCAS, one experimenter explained the nature of the questionnaires and provided examples of how the questions could be answered. Afterwards, one experimenter scored the questionnaires. Four children scored above the normal range (slightly elevated or higher) on one or both questionnaires, and were therefore provided with a filler task during the second session.

One week later the second session took place. Two experimenters and a teacher were present, and the teacher was instructed to be extra attentive to the well-being of the students. At the start of the session, students were informed that the purpose of the experiment was to examine how children assess their classmates' social behaviour. Students were told that they would each rate two classmates on their social behaviour, and that an experimenter would individually and anonymously inform them of how they had been rated later on. Next, the peer

assessment booklet was handed out, and the exemplar item on the first page was discussed. Special attention was given to the ratings "insufficient to sufficient" and "sufficient to good" to ensure that all children were aware of what these ratings reflected moderately negative and moderately positive, respectively. Students then completed both assessments. When done, students continued their regular schoolwork in silence.

Half an hour later, students visited the experimenter in a separate room one by one. The experimenter verbally presented the manipulated peer assessments with a 2 s pause between each item. The order of the assessments was counterbalanced. After each assessment, participants filled in the two corresponding VAS-scales. Before completing the final questionnaire, the experimenter verbally presented the filler task. Students could write down the answers. After completing the final questionnaire, the experimenter informed the students that they had not been presented with the real peer assessments, but manipulated ones, and then each student returned to their own classroom to continue working on regular school work, under strict instructions not to talk to their classmates. After all the students in a class had received the assessments, the class was debriefed collectively. During this debriefing, the experimenter explicitly stated that the students had received manipulated assessments, and that "the actual assessments were very positive". Moreover, the experimenters explained the reason behind the experiment, and students were given an opportunity to ask questions.

Results

We first conducted two independent samples *t*-tests to investigate whether the order of the assessments had an influence on how students experienced receiving their manipulated peer assessments. The analyses showed no difference between those who first received the short assessment and those who first received the extended assessment on students' experienced utility of the short assessment and extended assessment, *t*s < 1. Therefore, the order of the assessments was not taken into account in subsequent analyses.

Experienced utility

One participant was excluded from these analyses because his evaluations of the extended peer assessment were severe outliers (+3 standard deviations). Excluding this participant did not affect any of the analyses in terms of accepting/rejecting the null-hypothesis. A paired-samples *t*-test showed that children experienced the extended peer assessment ($M = 19.31$, $SD = 14.28$) as significantly more pleasant than the short peer assessment ($M = 10.40$, $SD = 9.19$), $t(28) = 3.70$, $p < .001$, $d = 0.742$. Note that both assessments were experienced very negatively. The extent to which the participants wanted to receive the assessment again was also significantly higher for the extended ($M = 13.74$, $SD = 15.17$) than the short peer assessment ($M = 7.31$, $SD = 9.06$), $t(29) = 3.78$, $p < .001$, $d = 0.515$.

Retrospective evaluations

Binomial tests were used to investigate whether the outcomes for the remembered utility and prospective social behaviour measures differed significantly from the 0.50 that would be expected by chance. Firstly, participants were not aware that the extended assessment had a longer duration than the short assessment (18/30, $p = .362$). As hypothesised, the short assessment was remembered as less pleasurable (25/30, $p < .001$) and more difficult to deal with (26/30, $p < .001$).

With regards to prospective behaviour, most participants preferred to receive the extended assessment again as a third assessment (27/30, $p < .001$) and the day after (25/30, $p < 0.01$). Interestingly, most children also indicated that they would rather work together with the classmate who had filled in the short assessment (26/30, $p < .001$).

Experiment 2

Method

Participants and design

Experiment 2 was conducted in a primary school in the Netherlands. Of the 46 parents or caretakers that were contacted, 45 provided parental consent. None of the students showed elevated anxiety scores, but one student was ill at the time of the experiment, leaving a final sample of 44 children ($M_{age} = 9.50$, $SD = 1.47$; 23 boys). Again, a within-subjects design was used in which students were exposed to a short (four positive ratings) and an extended peer assessment (four positive ratings followed by one moderately positive rating), the order of which was counterbalanced.

Materials and procedure

The materials and procedure in Experiment 2 were identical to Experiment 1 with two exceptions, namely the CDI was not administered and the children received positive instead of negative peer assessments.

Results

Independent samples *t*-tests showed no difference between those who first received the short assessment and those who first received the extended assessment on students' experienced utility of the short assessment ($t < 1$) and extended assessment ($t(42) = 1.80$, $p = .079$), so subsequent analyses did not take the order of the assessments into account.

Experienced utility

A paired-samples *t*-test showed that the short peer assessment ($M = 94.91$, $SD = 6.53$) was evaluated as significantly more pleasant than the extended peer assessment ($M = 90.93$, $SD = 7.93$), $t(43) = 3.70$, $p < .001$, $d = 0.548$. Note that both assessments were experienced as very pleasant. The extent to which students indicated that they wanted to receive the assessment again was also significantly higher for the short ($M = 93.09$, $SD = 10.01$) than the extended peer assessment ($M = 81.11$, $SD = 14.36$), $t(43) = 3.23$, $p = .002$, $d = 0.968$.

Retrospective evaluations

Binomial tests showed that students were aware that the extended peer assessment had been longer than the short peer assessment (35/44, $p < .001$). The extended peer assessment was remembered as less pleasurable (40/44, $p < .001$) and more difficult to deal with (37/44, $p < .001$).

With regards to prospective behaviour, most participants preferred to receive the short assessment again as a third assessment (42/44, $p < .001$) and the day after (33/44, $p < .001$). Most children indicated a preference for working together with the classmate who had filled in the short assessment (37/44, $p < .001$).

Discussion

This study aimed to investigate whether primary school children's experiences with unpleasant and pleasant peer assessments of their social behaviour, and memories thereof, are sensitive to ending with the most pleasant part of the assessment. We kept the peak identical across the assessments and manipulated the endings of the experience (cf. Kahneman et al., 1993). Results showed that, as intended, the positive assessments (Experiment 2) were evaluated as very pleasant and the negative assessments (Experiment 1) as very unpleasant. As for the main research questions, participants' answers on the experienced utility measures showed clearly that they differentially enjoyed the short and extended peer assessments within both experiments. This differentiation was in line with what would be expected based on the peak-end effect: In Experiment 1, students preferred an unpleasant peer assessment to be extended with a moderately unpleasant part (i.e. extended assessment), and in Experiment 2, a pleasant peer assessment that was *not* extended with a moderately pleasant part (i.e. short assessment). This pattern was also reflected in the retrospective evaluation measures, as the large majority of students remembered this assessment as more pleasant and less difficult to deal with in both experiments, which, in turn, led to a preference for that assessment as the third assessment and the day after. Moreover, most children indicated a preference for working together on a task with the classmate who had filled in the preferred assessment.

But what can we conclude from these findings? On a theoretical level, our findings indicate that the peak-end effect likely plays a role in how children experience and remember peer assessments as well as their subsequent preferences. Importantly, the peer assessments, especially the unpleasant ones, comprised a very salient and emotional experience. It seems that, similar to adults (cf. Chajut et al., 2014; Redelmeier et al., 2003), children are very sensitive to the ending of emotionally salient experiences. Interestingly, participants were aware that the extended assessment had a longer duration than the short assessment in Experiment 2, but not in Experiment 1. These conflicting results are likely a result of the higher levels of negative affect that the unpleasant peer assessments induced, and suggest that, perhaps, the efficacy of duration neglect increases as participants get more distracted, in this case by their emotions.

Our findings are also relevant for educational practice. Our results suggest that the process of peer assessment and in particular receiving the results can be made more pleasant for students by ending with the best part of the assessment, which might

positively affect students' motivation, task engagement, learning, and the relationship between students. Thus, if students were to receive a negative assessment, it is probably advised to structure the feedback in such a way that the end comprises the most pleasant part of the negative feedback. Similar, if students were to receive a positive assessment, it is probably advised not to end with the most unpleasant part, but to end on a high note. Teachers could take this into account when giving the feedback, or students could be instructed to adhere to these principles.

A limitation of our study is that we only varied the ending of the peer assessments but not the peak. Although this is a limitation of many peak-end effect studies, including the original experiment of Kahneman et al. (1993), future research is recommended to vary the peak alongside the ending. Varying the location of the peak would have the added benefit of enabling a test of an alternative explanation for our findings, namely that children based their evaluations of the peer assessments on an average score of the peer ratings. Additionally, because being able to contrast two experiences to each other might affect how peaks and endings shape our memory (cf. Hoogerheide & Paas, 2012), it would be good to replicate these effects using a between-subject design and to uncover the information that underlies participants' perceptions of the experience. Another interesting avenue for future research would be to examine whether the same effects would be found with academic competencies, and whether, as has been shown with peer assessments of academic tasks (e.g. Strijbos, Narciss, & Dünnebier, 2010), perceptions of peer competence would also affect the processing and perceptions of the feedback given during peer assessments of social behaviour.

To conclude, this study indicates that children's memory of emotional experiences is very sensitive to endings and that the process of peer assessments can be improved by ending with the most positive aspect of the feedback. This study thereby contributes to a growing body of literature exploring practical applications of the peak-end effect in contexts relevant for learning (Finn, 2010; Finn & Miele, 2016; Hoogerheide & Paas, 2012).

Acknowledgments

The authors would like to thank Carla Stolk-Vink and Marian van Rees-Vermeulen for facilitating this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory Into Practice*, *42*, 18–29. doi:10.1207/s15430421tip4201

Chajut, E., Caspi, A., Chen, R., Hod, M., & Ariely, D. (2014). In pain thou shalt bring forth children: The peak-and-end rule in recall of labor pain. *Psychological Science*, *25*, 2266–2271. doi:10.1177/0956797614551004

Diener, E., Wirtz, D., & Oishi, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science*, *12*, 124–128.

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*, 331–350. doi:10.1080/03075079912331379935

Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*, 96–98.

Elliot, J., & Harackiewicz, J. M. (1994). Goal setting achievement orientation and intrinsic motivation a mediational analysis. *Journal of Personality and Social Psychology*, *66*, 968–980.

Finn, B. (2010). Ending on a high note: Adding a better end to effortful study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1548–1553.

Finn, B., & Miele, D. B. (2016). Hitting a high note on math tests: Remembered success influences test preferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*, 45–55.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development*, *20*, 53–70. doi:10.1080/07294360123776

Hoogerheide, V., & Paas, F. (2012). Remembered utility of unpleasant and pleasant learning experiences: Is all well that ends well? *Applied Cognitive Psychology*, *26*, 887–894. doi:10.1002/acp.2890

Jacobs, J. E., Bleeker, M. M., & Constantino, M. J. (2003). The self-system during childhood and adolescence: Development, influences, and implications. *Journal of Psychotherapy Integration*, *13*, 33–65. doi:10.1037/1053-0479.13.1.33

Kahneman, D., Fredrickson, D. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401–405.

Leippe, M. R., Romanczyk, A., & Manion, A. P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *Journal of Applied Psychology*, *76*, 367–379. doi:10.1037/0021-9010.76.3.367

Levine, R. E. (2008). Peer assessment in team-based learning. In L. K. Michaelsen, D. X. Parmalee, & K. K. McMahon (Eds.), *Team-based learning for health professions education* (pp. 103–116). Sterling: Stylus.

Levine, R. E., Kelly, P. A., Karakoc, T., & Haidet, P. (2007). Peer evaluation in a clinical clerkship: Students' attitudes, experiences, and correlations with traditional assessments. *Academic Psychiatry*, *31*, 19–24.

Miron-Shatz, T. (2009). Evaluating multiepisode events: Boundary conditions for the peak-end rule. *Emotion (Washington, D.C.)*, *9*, 206–213. doi:10.1037/a0015295

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, *104*, 187–194.

Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education*, *27*, 443–454. doi:10.1080/0260293022000009311

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, *36*, 545–566.

Stagg, K. (2016). Impact of peer assessment on student well-being. *Medical Teacher*, *38*, 212. doi:10.3109/0142159X.2015.1056127

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, *20*, 291–303. doi:10.1016/j.learninstruc.2009.08.008

Timbremont, B., & Braet, C. (2002). *Children's depression inventory: Nederlandstalige versie*[Children's Depression Inventory: Dutch version]. Lisse: Swets & Zeitlinger.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*, 249–276.

Warnes, E. D., Sheridan, S. M., Geske, J., & Warnes, W. A. (2005). A contextual approach to the assessment of social skills: Identifying meaningful behaviors for social competence. *Psychology in the Schools*, *42*, 173–187. doi:10.1002/pits.20052