Empirical Support for Changes to Mosaic by ACT: Social Emotional Learning Assessment Individual Student Reports

Kate E. Walton, PhD, Dana Murano, PhD, & Jeremy Burrus, PhD

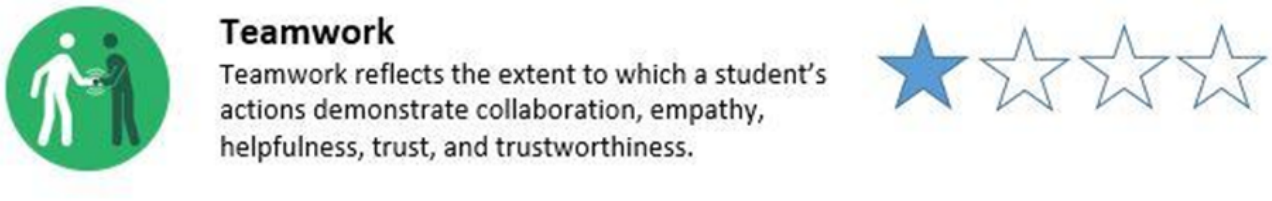
Overview of the Current Study

Listening to and implementing user feedback is an essential component of product development; however, such feedback cannot be followed blindly but instead must be considered in conjunction with empirical data. We received feedback from teachers and administrators on the prior version of the individual student report from the assessment component of Mosaic™ by ACT®: Social Emotional Learning (previously called ACT® Tessera®). In light of this feedback, we sought additional data from students to ascertain the extent to which the teachers' and administrators' concerns were warranted and to guide our decision making for future report edits. Overall, we found support for the use of reports with ratings rather than reports with skills rank ordered, and we found support for a rating scheme coupled with descriptive labels. These findings informed product changes to assessment student reports.

Background

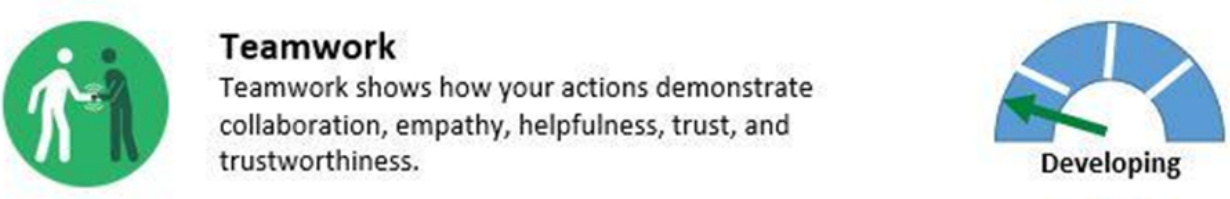
On prior individual student reports used with the assessment, students received a 1–4 star rating (with 1 being the lowest possible score) on each of the five social and emotional skills (SE skills), an example of which can be seen in Figure 1.

Figure 1. An Example of a Level 1 Score on the Teamwork Skill With the Star Rating Scheme



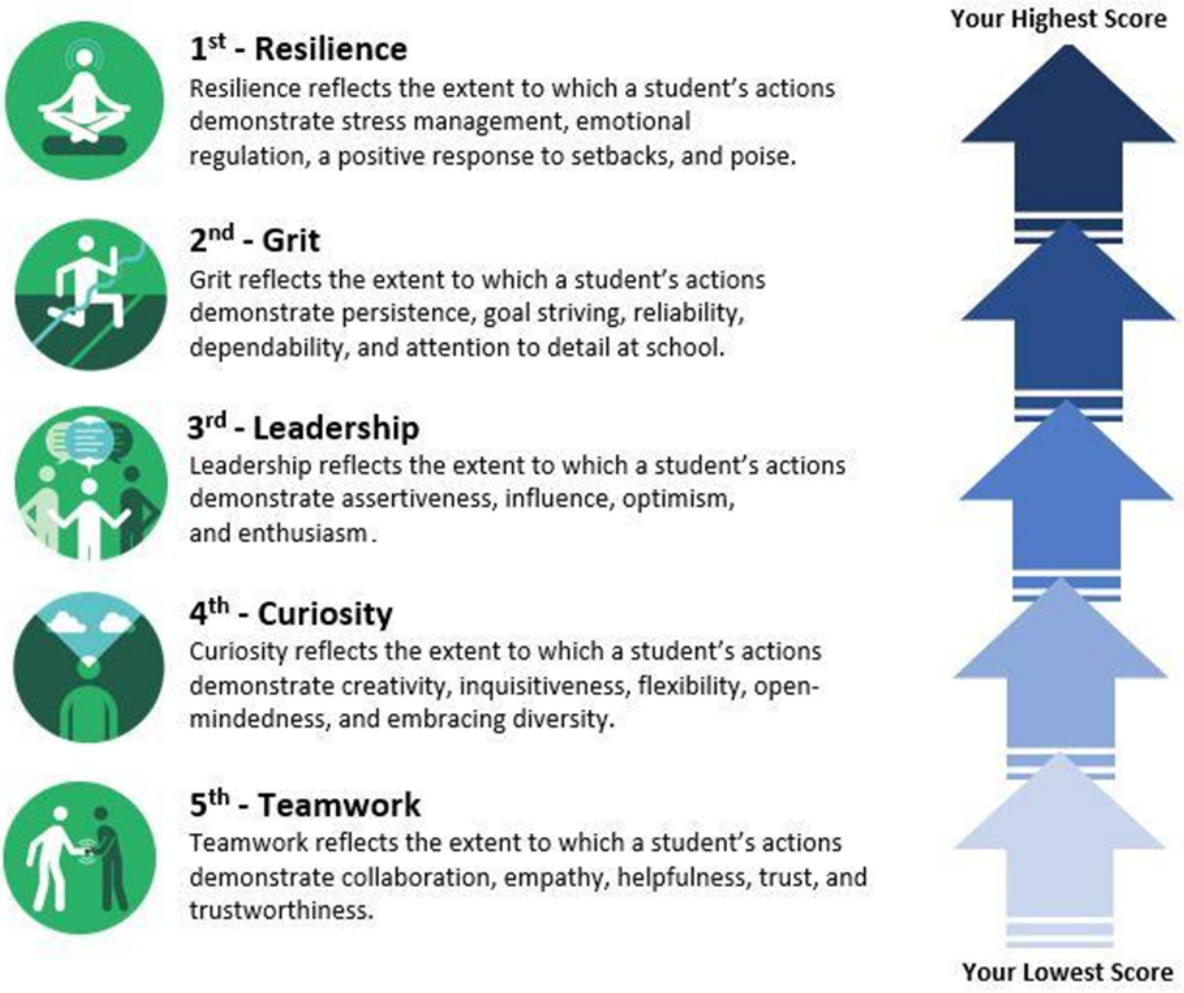
Some teachers and administrators voiced concern about this approach. One concern raised was the effect this way of communicating assessment scores may have on students' feelings and motivation levels. For example, students receiving all low scores may feel bad about themselves and/or feel discouraged. Alternatively, students receiving all high scores may feel little motivation for continued skill development. Therefore, we considered alternative reporting approaches, and we sought feedback from the end user, the student. One alternative we evaluated was a different scheme (i.e., gas gauge) with descriptive terms, an example of which can be seen in Figure 2. Notice that the lowest of the four possible levels is described as **developing**. Two, three, and four stars were replaced with equivalent levels in the gas gauge, and these levels were described as **approaching, demonstrating**, and **mastering**, respectively. The descriptors should offer encouragement for all students; for example, **developing** does not necessarily have a negative implication. Also note that one's gauge can never be completely empty or full, suggesting there is always room for development.

Figure 2. An Example of a Level 1 Score on the Teamwork Skill With the Gas Gauge Rating Scheme



In addition to comparing these two ratings-type schemes, we also evaluated effects of and student preference for ratings vs. a ranking of the skills. A ranking scheme, which is depicted in Figure 3, would eliminate the risk of a student receiving all low scores. In the example in Figure 3, the student received the lowest score on Teamwork, relative to the other four skills. All five skills were presented in order of strongest (1st) to weakest (5th).

Figure 3. An Example of a Student's Teamwork Level With the Ranking Scheme



Research Questions

In the current study, we specifically addressed the following research questions:

1. Does feedback on student performance (i.e., negative, mixed, positive) affect students' sense of pride and motivation for development?
    - Does this effect depend on whether the report consists of stars vs. gas gauges?
2. Does type of report (i.e., star scores reported for all skills vs. gas gauge scores reported for all skills vs. rank ordering of skills) affect students' sense of pride and motivation for development?
3. Which type of report (i.e., star scores reported for all skills vs. gas gauge scores reported for all skills vs. rank ordering of skills) do students find most informative and/or prefer?

Method

Participants were students who took the ACT on the October 2019 national testing date. After taking the ACT, they were randomly selected to receive an invitation to complete the survey. They were randomly assigned to one of two groups. Both groups received sample reports like those described above and were asked to imagine the reports contained their own scores on an assessment of five SE skills.

Group 1 participants, with whom we addressed the first research question, were randomly assigned to receive reports with star scores (n = 240) or gas gauge scores (n = 214; between-subjects factor). All participants received those reports with three different levels of feedback—worst possible scores on all skills, mixed scores on skills, best possible scores on all skills (within-subjects factor). After viewing "their" reports, they responded to five-point Likert-type items assessing likelihood of having a sense of pride (three items; e.g., *Feel as though you have a lot to be proud of*) and feeling motivated for SE skill development (six items; e.g., *Feel motivated to try to improve your social and emotional skills*). Internal consistency reliability estimates for both scales reached .90.

Group 2, with whom we addressed the second and third research questions, completed the same scales as those in Group 1. The internal consistency reliability estimates for sense of pride and motivation for development were .75 and .84, respectively. All participants (N = 273) received three different report types (i.e., star scores reported for all skills vs. gas gauge scores reported for all skills vs. rank ordering of skills). The level of feedback was mixed (i.e., not all negative or positive), and the scores for each skill were held constant, resulting in the same top-rated skill on each report type. In addition, these participants were asked, of the three report types, which was easiest to understand, which was most informative, and which they preferred.

Results

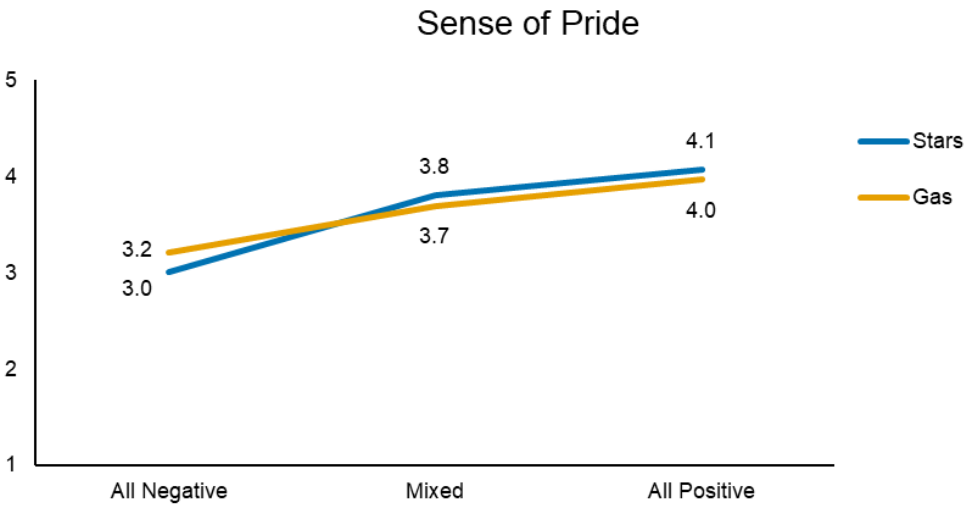
Effects of Nature of Feedback and Report Type on Students' Sense of Pride and Motivation for Development

A repeated measures MANOVA was carried out to test for main effects of level of feedback and reporting scheme (stars vs. gas gauges) and the interaction between the two. There was no significant main effect of stars vs. gas gauges (Wilk's Λ = 1.00, F(2, 451) = .04, p = .96). There was a significant main effect for level of scores (Wilk's Λ = .70, F(4, 449) = 48.06, p < .01), and there was a significant interaction (Wilk's Λ = .97, F(4, 449) = 3.71, p < .01). The within-subjects main effects and interactions are described below.

Sense of Pride: Results for sense of pride can be seen in Figure 4. Sense of pride significantly differed across the three levels of feedback (all negative M = 3.11; mixed M = 3.75; all positive M = 4.02; η_p^2 = .25, which is a large effect). Note that a score of 3 indicates a tendency to respond with "may or may not," and a score of 4 indicates a tendency to respond with "likely." For example, students who received all positive ratings said they are likely to recognize their strengths, and students who received all negative ratings said they may or may not recognize their strengths.

The interaction between report type and level of feedback was statistically significant, yet the effect size was small (η_p^2 = .01). The effect of score level was greater in the stars condition than in the gas gauge condition.

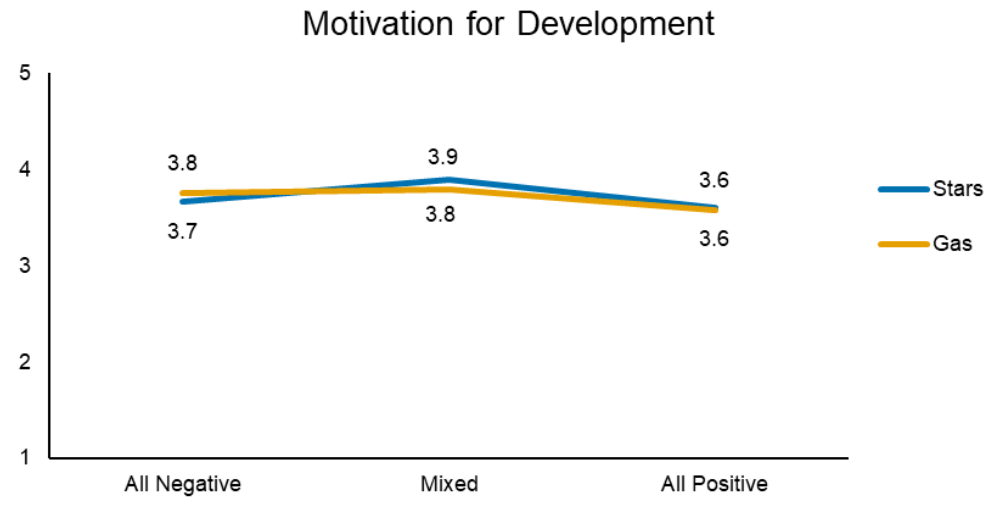
Figure 4. Mean Scores for Sense of Pride Across Conditions



Motivation for Development: Results for motivation for development can be seen in Figure 5. Motivation for SE skill development significantly differed across the three levels of feedback (all negative *M* = 3.72; mixed *M* = 3.83; all positive *M* = 3.59; η_p^2 = .04, which is a small-medium effect). Students with mixed feedback were most likely to feel motivated to improve their SE skills.

The interaction between report type and level of feedback was statistically significant, yet the effect size was small (η_p^2 = .01). The effect of score level was greater in the star condition than in the gas gauge condition. When receiving all negative scores, motivation was stronger in the gas condition, but with mixed scores, motivation was higher in the stars condition.

Figure 5. Mean Scores for Motivation for Development Across Conditions



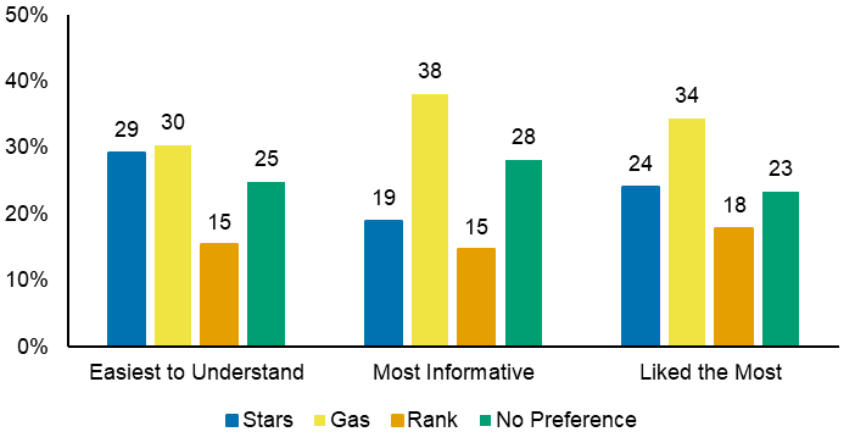
Effects of Report Type on Students' Sense of Pride and Motivation for Development

A repeated measures MANOVA was carried out to test for the main effect of reporting scheme (stars vs. gas gauges vs. rank order). The effect was not statistically significant (Wilk's $\Lambda = .97$, $F(4, 269) = 2.14$, $p = .08$).

Students' Subjective Preference for Report Type

In general, students found the rank report the most difficult to understand, and they preferred the gas gauges, which they found to be most informative (see Figure 6).

Figure 6. Students' Impressions of Various Report Types



Conclusions

In Group 1, the type of rating report had no effect on students' feelings, but the nature of that feedback did. The better the scores, the more pride students felt, and students with mixed feedback felt the greatest motivation to work on their SE skill development. Clearly, it would be advantageous to foster positive feelings and motivation for development in all students, and we sought to determine whether this would be accomplished with reports rank ordering the skills in a relative fashion vs. providing absolute ratings on the skills. In Group 2, the type of report had no effect on students' feelings. The type of report did, however, yield differences in terms of students' subjective experience; specifically, they found rank-order reports to be the most difficult to understand and found gas gauges with descriptive labels to be the most informative and thus the most preferred. Therefore, we concluded that the gas gauge reporting scheme was the best option of the three we evaluated.

In light of some critical user feedback we received, we took a systematic, empirical approach to evaluate current and alternative reporting strategies. This study represents an example of being both responsive (to customer needs) and responsible (making data-informed decisions) in making assessment-related revisions to create a more user-friendly and actionable experience. The most recently released version of the assessment makes use of the reporting scheme supported by our research, which, according to school personnel, has been well received by students.

Kate E. Walton, PhD

Kate Walton is a principal research scientist in ACT's Center for Social, Emotional, and Academic Learning. Prior to joining ACT in 2017, she was Associate Professor of Psychology at St. John's University. She is interested in social and emotional skill assessment and development.

Dana Murano, PhD

Dana Murano is a research scientist in ACT's Center for Social, Emotional, and Academic Learning. Her research focuses on the development and assessment of social and emotional skills in students.

Jeremy Burrus, PhD

Jeremy Burrus is the senior director of ACT's Center for Social, Emotional, and Academic Learning (SEAL). Before coming to ACT, he was a principal research scientist at ProExam's Center for Innovative Assessments, and prior to that he was a research scientist at Educational Testing Service. He graduated with a PhD in Social Psychology from the University of Illinois, Urbana-Champaign in 2006, and was a post-doctoral research scholar at Columbia Business School in New York City from 2006-2008.

Acknowledgements

We thank Alex Casillas and Jill McVey for their contributions to this work.