Supplemental Materials

Methodological Details Omitted from Manuscript

**Pilot Study Information**

 A pilot study involving 30 participants was conducted to (1) establish goals for the relative condition that were comparable to those of the nominal condition, (2) examine whether the manipulation of feedback valence and FIT was effective, and (3) investigate whether the false monetary bonus was sufficient to motivate respondents to set goals above the minimum goal level option.

**Sampling Methodology**

Participants were sampled from Amazon’s Mechanical Turk and were screened to include only those who are 18 years of age and above, native English speakers since the task directions and survey was in English, and currently live in the United States or Canada. Participants were compensated $1.60 for completion of the tasks, which were completed along with our Likert measures via a single survey. This amount was derived based on the United States minimum wage of $7.25 per hour and an estimated 12 to 15 minutes to complete the study. An invitation was posted on MTurk, and the survey was open until 800 completed survey responses were obtained without violating the veracity and manipulation checks. For testing hypotheses, we sought to gather responses from a minimum sample size of approximately 100 per condition. This sample size was chosen in drawing on advice offered by Hair, Black, Babin, and Anderson (2010), who noted that when a structural equation model (SEM) has five or fewer latent constructs, at least four items per construct, and items load strongly onto respective constructs, 100 individuals are sufficient.

**Data Exclusion Rules**

**Participants.** For our main analysis, we screened out error outliers that were due to inattentive responding using three approaches (Aguinis, Gottfredson, & Joo, 2013) including (1) veracity-check items (e.g., “I do not speak a word of English;” Meade & Craig, 2012), (2) feedback-recall questions (i.e., individuals were asked to recall what kind of feedback message they received), and (3) consistent-response pattern plus failure to provide RGB codes. Three items were randomly placed throughout the survey as veracity checks for inattentive responders. Regarding veracity check items, the specific items that we used included “I am answering this questionnaire using an electronic device,” “I have been to every country in the world,” and “I do not understand a word of English.” Previous research has demonstrated that these items effectively flag inattentive participants (Meade & Craig, 2012). Additionally, a bolded statement was included in the informed consent form to make participants aware that safeguards are embedded in the survey to detect those who make insufficient effort while responding, such as not paying attention to the instructions or not reading the questions. Respondents who incorrectly responded to any one of the three veracity checks received a message indicating that they will be exited from the survey without payment due to inattentive responding. This message was displayed immediately following their incorrect response and was followed by the final page of the survey.

The survey link was opened 1,780 times; however, there were only a total of 903 participants who completed the survey and passed the inattentive responding checks and whose data were analyzed further. Third, we examined the data from participants whose patterns of responding showed the same response option for every question (with the exception of the inattentive-responding checks) and who also failed to provide a single RGB code. Two responses met these criteria for removal from the analysis. An additional 87 responses were deleted for failure to complete the survey. In total, 103 responses were eliminated from the final dataset of submitted surveys leaving 800 remaining for analysis. Generalized Cook’s distance led us to identify and screen out seven more influential outliers (see Aguinis et al., 2013).

**Measures.** Two measures were also gathered but excluded for reasons explained below. First, the Reversal Theory State Measure – Bundled Version (RTSM; Desselles, Murphy, & Theys, 2014) was administered to assess respondents’ psychological state. A single item on this measure reflects the participant’s presence in the telic or paratelic state. The item asks respondents to choose one of two groups of statements (“bundles”) that describe what is motivating them at a particular moment. For example, the participant would choose between the telic bundle containing “accomplish something for the future,” “do something serious,” and “do something crucial” and the paratelic bundle containing “enjoy myself at the moment,” “do something playful,” and “do something of no great concern.” We chose to exclude this measures from our analysis because of issues related to the unreliability of the single item approach.

Second, the desire to respond to feedback was measured with four items: “After reading the feedback, I am looking forward to improving on the next trial,” “I think that the feedback I received will help me to do better next time,” “After seeing my feedback, I have some ideas about how to improve,” and “I have no intention of using the feedback to guide my performance on the next task.” Response options were on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The fourth item was reverse-coded. This measure has demonstrated satisfactory construct validity through confirmatory factor analysis as well as high reliability (see Waples, 2015). Though we initially posited this factor as indirectly linking affect to goal setting, we chose to omit this measure from our report for the sake of brevity as our conclusions remain unchanged. We can make this analysis available upon request.

**Data Analysis Details**

After data screening, we tested whether our data violated the assumption of multivariate normality utilizing Mardia’s Multivariate Normality (MVN) test and Henze-Zirkler’s MVN test. These tests indicated that the multivariate normality assumption did not hold (*p* < .001). Therefore, we opted for diagonally weighted least squares (DWLS) estimation, which is more optimal for our Likert data (see Finch & French, 2015). Convergent-discriminant validity for the Likert-type measures of our model was examined using confirmatory factor analysis (CFA) via the approach established by Henseler, Ringle, and Sarstedt (2015). After establishing our measurement model, we then included our other measures (e.g., goal setting, performance, conditions, the interaction between feedback accuracy and information type) and used structural equation modeling (SEM) to test our proposed model and hypotheses.

**Procedural Details Regarding the Operationalization of Goal Setting**

Participants had some discretion over the goal level they set, with the restriction that the goal set is above a predefined minimum. The minimum goal for the relative condition was 40%, and the minimum goal for the nominal condition was calculated based on how many university RGB codes were identified by the pilot participants at the 40th percentile. The participants set their goals by selecting a choice out of a set of options in the survey with the lowest option being the minimum goal. In the goal-setting instructions, participants were told that they must select at least a minimum goal in order to receive the baseline payment for participation in the study. The reasoning behind the minimum goal communicated to the respondents was that a minimum goal discourages participants from getting the baseline payment without effort.

**Characteristics of feedback messages.** Table 1 contains the feedback messages provided to participants in each condition.

Table 1

*Feedback Messages Provided to Each Condition*

|  |  |
| --- | --- |
| **Condition** | **Feedback Message Shown** |
| **Nominal-Positive** | “Good job. You have reached your goal. You correctly identified the primary color RGB numbers for (your set/#) universities.” |
| **Nominal-Negative** | “Unfortunately, you failed to reach your goal. You failed to identify the primary color RGB numbers for (your set/#) universities.” |
| **Relative-Positive** | “Good job. You have reached your goal. You performed better than X% of the participants.” |
| **Relative-Negative** | “Unfortunately, you failed to reach your goal. You failed to perform better than X% of the participants.” |

The percentage of participants who selected each goal level at time 1 and time 2 is described by tables 2 and 3, which reveal a fairly uniform distribution of chosen goal levels. Table 4 reveals how goal levels were equated across nominal and relative goal setting conditions.

Table 2

*Percentage of Participants Who Selected Each Goal 1 Level*

|  |  |
| --- | --- |
| Goal 1 Level | % Chosen |
| 1 | 20.0 |
| 2 | 10.0 |
| 3 | 16.7 |
| 4 | 10.0 |
| 5 | 23.3 |
| 6 | 20.0 |

Table 3

*Percentage of Participants Who Selected Each Goal 2 Level*

|  |  |
| --- | --- |
| Goal 2 Level | % Chosen |
| 1 | 23.3 |
| 2 | 20.0 |
| 3 | 20.0 |
| 4 | 6.7 |
| 5 | 13.3 |
| 6 | 16.7 |

Table 4

*Number of RGB Code Pairs Found at Each Performance Percentile*

|  |  |
| --- | --- |
| Percentiles | Number found |
| 1% | 0 |
| 20% | 1 |
| 40% | 2 |
| 60% | 2 |
| 80% | 3 |
| 100% | 5 |

Lastly, since individuals were randomly assigned to one of eight conditions based on the pairing of (i) nominal or relative feedback, (ii) positive or negative feedback, and (iii) accurate or inaccurate feedback, we are making the distribution available for review. Notably, accuracy and valence were associated, reflecting to some extent the relative difficulty of our task, which made for an uneven distribution of participants across the eight conditions. However, it should be kept in mind that our interest was in examining the interaction between feedback information type and feedback accuracy (which was adequately sampled), not a three-way interaction (which was not adequately sampled).

*Table 5*

*n per cell breakdown*

|  |  |  |  |
| --- | --- | --- | --- |
| Feedback Information Type | Feedback Valence | Feedback Accuracy | *n* |
| Relative Feedback | Negative Valence | Inaccurate Feedback | 41 |
| Relative Feedback | Negative Valence | Accurate Feedback | 157 |
| Relative Feedback | Positive Valence | Inaccurate Feedback | 167 |
| Relative Feedback | Positive Valence | Accurate Feedback | 34 |
| Nominal Feedback | Negative Valence | Inaccurate Feedback | 33 |
| Nominal Feedback | Negative Valence | Accurate Feedback | 171 |
| Nominal Feedback | Positive Valence | Inaccurate Feedback | 161 |
| Nominal Feedback | Positive Valence | Accurate Feedback | 29 |
| Feedback Information Type | Feedback Accuracy | *n* |  |
| Relative Feedback | Inaccurate Feedback | 208 |  |
| Relative Feedback | Accurate Feedback | 191 |  |
| Nominal Feedback | Inaccurate Feedback | 194 |  |
| Nominal Feedback | Accurate Feedback | 200 |  |

**Deviations from Pre-Registration**

Though this study was pre-registered as part of a conventional dissertation, we should note that when our model was first tested, residualized gain scores were examined (i.e., goal change and performance change were modeled as causal variables), which as we learned was fraught with unreliability (for a review, see Bezruczko, Fatani, & Magari, 2016). Hence, we modeled goal setting and performance at time 1 as causal factors of goal setting and performance at time 2, respectively, which allowed us to model change.

**References**

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*(2), 270-301.

Bezruczko, N., Fatani, S. S., & Magari, N. (2016). Three tales of change: Ordinal scores, residualized gains, and rasch logits—when are they interchangeable? *SAGE Open*, *6*(3), 215824401665990. <https://doi.org/10.1177/2158244016659905>

Desselles, M. L., Murphy, S. L., & Theys, E. R. (2014). The development of the reversal theory state measure. *Journal of Motivation, Emotion, and Personality, 2*(1), 10-21.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. New York, NY: Taylor & Frances/Routledge.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455.

Waples, C. J. (2015). *Receptivity to feedback: An investigation of the influence of feedback sign, feedback specificity, and goal orientation.* (Unpublished doctoral dissertation). Kansas State University, Manhattan, KS.