Online Appendix for
The Gender Gap in Self-Promotion

Christine L. Exley

Judd B. Kessler

Table of Contents

A Additional Results

Table A.1: Performance and Absolute Performance Beliefs

DV:	Performance	Belief		Belief–Performance
	(1)	(2)	(3)	(4)
Female	0.60***	-2.29***	-2.19***	-2.88***
	(0.13)	(0.14)	(0.14)	(0.17)
Constant	9.34***	11.05***		1.71***
	(0.09)	(0.09)		(0.12)
N	3587	3587	3587	3587
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The SEs are robust. Results are from OLS regressions of the noted dependent variable (DV). *Performance* is the number of questions a participant answered correctly out of the 20 questions on the test. *Belief* is the number of questions a participant believes he or she answered correctly. *Belief–Performance* is the difference between these two variables, calculated for each participant. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all study versions from waves 1–5 involving the math and science test (i.e., all but the *Private (Verbal)* version).

Table A.2: Regression results on the role of providing information on absolute and relative performance

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: *Self-Promotion* Version				
Female	-11.75***	-0.55***	-14.09***	-14.29***
	(2.95)	(0.13)	(3.44)	(3.43)
Informed	-1.10	0.04	1.67	-0.04
	(1.36)	(0.07)	(1.50)	(1.51)
*Informed***Female*	3.80	0.11	2.15	1.76
	(2.37)	(0.11)	(2.44)	(2.39)
N	604	604	604	604
Panel 2: All Versions with Evaluations Before and After Being Informed				
Female	-13.89***	-0.67***	-17.17***	-16.15***
	(1.14)	(0.05)	(1.31)	(1.32)
Informed	-1.49***	0.00	0.32	-0.84
	(0.56)	(0.03)	(0.55)	(0.52)
*Informed***Female*	4.10***	0.21***	2.30***	1.59**
	(0.88)	(0.04)	(0.81)	(0.80)
N	4188	4188	4188	4188
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are clustered at subject-level. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Informed* is an indicator for the evaluation being provided after the participant is informed of their absolute and relative performance. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data in Panel 1 are from the *Self-Promotion* version. Data in Panel 2 are from all versions that elicit evaluations of math and science performance before and after participants are informed of their absolute and relative performance (i.e., all but the *Private (Immediately Informed)* version, *Private (Other-Evaluation)* version, and *Private (Verbal)* version). Each participant in these versions is in the data twice for each specification, once providing an evaluation before being informed and once providing an evaluation after being informed.

Table A.3: Regression results on the impact of promotion incentives from the *Self-Promotion* and *Private* versions

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-13.86***	-0.59***	-17.85***	-16.52***
	(2.82)	(0.13)	(3.36)	(3.45)
Self-Promotion	6.25**	0.26**	4.27	6.93**
	(2.72)	(0.13)	(3.35)	(3.30)
*Self-Promotion*Female*	1.66	-0.00	2.30	1.07
	(4.04)	(0.18)	(4.77)	(4.84)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-8.55***	-0.33**	-13.81***	-13.88***
	(2.79)	(0.14)	(3.40)	(3.41)
Self-Promotion	7.79***	0.34**	6.72**	9.00***
	(2.85)	(0.14)	(3.34)	(3.24)
*Self-Promotion*Female*	1.41	-0.09	2.08	1.31
	(3.93)	(0.18)	(4.74)	(4.70)
N	606	606	606	606
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Self-Promotion* is an indicator for the evaluation being from the *Self-Promotion* version. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from the *Self-Promotion* version and *Private* version run in wave 1, so participants were randomly assigned between these study versions.

Table A.4: Deservingness Measure

Female	-0.88
	(1.23)
N	2394
Performance FEs	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the deservingness measure, which ranges from 0–100, in response to the following question: "Out of a maximum amount of 100 cents, what amount of bonus payment, in cents, do you think you deserve for your performance on the test you took in part 1." *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all versions in which participants are asked about their own performance on the math and science test but do not have an opportunity to influence their payments through self-promotion (i.e., all but the *Self-Promotion* version, *Self-Promotion (Risky)* version, *Private (Other-Evaluation)* version, and *Private (Verbal)* version).

Table A.5: Robustness to excluding performance fixed effects

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-15.76***	-0.78***	-19.25***	-18.07***
	(1.14)	(0.05)	(1.30)	(1.32)
Constant	58.50***	3.76***	57.36***	61.39***
	(0.72)	(0.04)	(0.84)	(0.81)
N	2094	2094	2094	2094
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-11.16***	-0.56***	-16.78***	-16.92***
	(1.01)	(0.05)	(1.12)	(1.11)
Constant	57.86***	3.77***	58.51***	61.92***
	(0.62)	(0.03)	(0.68)	(0.66)
N	2990	2990	2990	2990
Performance FEs	No	No	No	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. Performance FEs are not included. Data are from all study versions involving evaluations of the participant's own math and science performance. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.6: Robustness to controlling for other demographic variables

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
Panel 1: Evaluations (before performance information is provided)				
Female	-12.70***	-0.61***	-15.95***	-14.82***
	(1.09)	(0.05)	(1.28)	(1.29)
Age	-0.30***	-0.01***	-0.32***	-0.24***
	(0.05)	(0.00)	(0.06)	(0.06)
Education (demeaned)	4.08***	0.21***	4.44***	4.90***
	(0.39)	(0.02)	(0.45)	(0.46)
Republican Leaning (demeaned)	0.12***	0.01***	0.10***	0.10***
	(0.02)	(0.00)	(0.02)	(0.02)
N	2092	2092	2092	2092
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-8.67***	-0.41***	-13.76***	-14.20***
	(0.90)	(0.04)	(1.05)	(1.04)
Age	-0.29***	-0.01***	-0.25***	-0.20***
	(0.04)	(0.00)	(0.05)	(0.05)
Education (demeaned)	3.38***	0.16***	4.22***	4.47***
	(0.33)	(0.02)	(0.38)	(0.37)
Republican Leaning (demeaned)	0.15***	0.01***	0.13***	0.11***
	(0.02)	(0.00)	(0.02)	(0.02)
N	2986	2986	2986	2986
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Education (demeaned)* is a number from 1 to 9 that corresponds with education level (where the least education is 1 and the most education is 9), demeaned by the average. *Republican Leaning (demeaned)* is a number from 0 to 100 that is the extent to which the participant indicated feeling favorably about the Republican party, demeaned by the average. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all study versions involving evaluations of the participant's own math and science performance but excludes the participants who selected "other" as their educational attainment. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.7: Robustness to excluding very low performers

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-13.50***	-0.62***	-17.38***	-16.38***
	(1.18)	(0.05)	(1.40)	(1.41)
N	1771	1771	1771	1771
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-8.72***	-0.38***	-14.63***	-15.04***
	(0.96)	(0.04)	(1.15)	(1.14)
N	2456	2456	2456	2456
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all study versions involving evaluations of the participant's own math and science performance, restricted to the set of participants who answered 6 or more questions correctly out of 20. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.8: Robustness to quantile regressions

Question:	Performance	Willingness-to-Apply	Success
	(1)	(2)	(3)
Panel 1: Evaluations (before information), 25th percentile			
Female	-18.00***	-25.00***	-30.00***
	(1.87)	(2.73)	(2.91)
N	2094	2094	2094
Panel 2: Informed Evaluations (after information), 25th percentile			
Female	-10.00***	-20.00***	-24.00***
	(1.50)	(2.13)	(2.09)
N	2990	2990	2990
Panel 3: Evaluations (before information), 50th percentile			
Female	-14.00***	-24.00***	-19.00***
	(2.18)	(2.61)	(2.40)
Constant	75.00***	65.00***	82.00***
N	2094	2094	2094
Panel 4: Informed Evaluations (after information), 50th percentile			
Female	-9.00***	-18.00***	-17.00***
	(1.13)	(1.94)	(1.81)
N	2990	2990	2990
Panel 5: Evaluations (before information), 75th percentile			
Female	-11.00***	-13.00***	-11.00***
	(0.99)	(1.47)	(1.65)
N	2094	2094	2094
Panel 6: Informed Evaluations (after information), 75th percentile			
Female	-6.00***	-11.00***	-10.00***
	(0.96)	(1.31)	(1.08)
N	2990	2990	2990
Performance FEs	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are from from 100 bootstrap replications. Results are from quantile regressions, estimated at the percentile noted in each panel, of the responses provided to the question noted in each column, as defined in the notes of Table 2. We do not run quantile regressions for the performance-bucket question elicited on six-point scale to avoid convergence issues given the discrete nature of this question and the inclusion of performance fixed effects. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all study versions involving evaluations of the participant's own math and science performance. Panels 1, 3, and 5 analyze evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panels 2, 4, and 6 analyze evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.9: Robustness to ordered probit regressions

Question:	Performance-Bucket	
	(1)	(2)
Panel 1: Evaluations (before performance information is provided)		
Female	-0.66***	-0.59***
	(0.05)	(0.05)
N	2094	2094
Panel 2: Evaluations (after performance information is provided)		
Female	-0.45***	-0.41***
	(0.04)	(0.04)
N	2990	2990
Performance FEs	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from ordered probit specifications of the responses provided to the performance-bucket question, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. We show results both with and without performance FEs due to concerns related to the inclusion of fixed effects in order probit specifications. Data are from all study versions involving evaluations of the participant's own math and science performance. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.10: Considering the relationship between performance and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-16.37***	-0.80***	-19.64***	-18.57***
	(1.18)	(0.06)	(1.33)	(1.35)
Performance (demeaned)	-0.54***	-0.06***	-0.42**	-0.27
	(0.17)	(0.01)	(0.19)	(0.18)
Performance (demeaned)	1.59***	0.09***	1.10***	1.13***
**Female*	(0.33)	(0.02)	(0.37)	(0.37)
N	2094	2094	2094	2094
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-12.26***	-0.59***	-17.44***	-17.81***
	(1.00)	(0.05)	(1.12)	(1.11)
Performance (demeaned)	0.61***	-0.01	0.18	0.58***
	(0.14)	(0.01)	(0.16)	(0.15)
Performance (demeaned)	2.05***	0.10***	1.55***	1.48***
**Female*	(0.28)	(0.01)	(0.31)	(0.30)
N	2990	2990	2990	2990
Performance FEs	No	No	No	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Performance (demeaned)* is the number of questions a participant answered correctly out of the 20 questions on the test, demeaned by the average performance. Data are from all study versions involving evaluations of the participant's own math and science performance. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.11: Considering the relationship between performance and evaluations when excluding very low performers

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-17.68***	-0.82***	-20.81***	-20.15***
	(1.40)	(0.06)	(1.57)	(1.61)
Performance (demeaned)	0.55**	0.01	0.69**	0.78***
	(0.25)	(0.01)	(0.29)	(0.29)
Performance (demeaned)	2.34***	0.11***	1.80***	1.98***
* *Female*	(0.41)	(0.02)	(0.48)	(0.48)
N	1771	1771	1771	1771
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-13.15***	-0.59***	-18.52***	-18.65***
	(1.23)	(0.06)	(1.34)	(1.35)
Performance (demeaned)	2.53***	0.10***	1.93***	2.33***
	(0.22)	(0.01)	(0.24)	(0.23)
Performance (demeaned)	2.71***	0.12***	2.27***	2.10***
**Female*	(0.34)	(0.02)	(0.40)	(0.39)
N	2456	2456	2456	2456
Performance FEs	No	No	No	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Performance (demeaned)* is the number of questions a participant answered correctly out of the 20 questions on the test, demeaned by the average performance. Data are from all study versions involving evaluations of the participant's own math and science performance, restricted to the set of participants who answered 6 or more questions correctly out of 20. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.12: Considering the relationship between other demographics and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
Panel 1: Evaluations (before performance information is provided)				
Female	-12.77***	-0.61***	-15.95***	-14.76***
	(1.09)	(0.05)	(1.29)	(1.29)
Age	-0.24***	-0.01***	-0.22**	-0.09
	(0.07)	(0.00)	(0.09)	(0.08)
Education (demeaned)	4.21***	0.21***	4.17***	4.44***
	(0.51)	(0.02)	(0.62)	(0.60)
Republican (demeaned)	0.16***	0.01***	0.16***	0.15***
	(0.03)	(0.00)	(0.03)	(0.03)
*Age*Female*	-0.11	-0.00	-0.21*	-0.29**
	(0.11)	(0.00)	(0.12)	(0.12)
*Education (demeaned)*Female*	-0.28	-0.02	0.57	1.01
	(0.78)	(0.03)	(0.91)	(0.92)
*Republican (demeaned)*Female*	-0.08**	-0.01***	-0.12***	-0.11**
	(0.04)	(0.00)	(0.05)	(0.05)
N	2092	2092	2092	2092
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-8.67***	-0.41***	-13.72***	-14.13***
	(0.90)	(0.04)	(1.05)	(1.04)
Age	-0.24***	-0.01***	-0.18***	-0.10
	(0.06)	(0.00)	(0.07)	(0.06)
Education (demeaned)	3.42***	0.17***	4.15***	4.40***
	(0.44)	(0.02)	(0.49)	(0.47)
Republican (demeaned)	0.22***	0.01***	0.18***	0.15***
	(0.02)	(0.00)	(0.02)	(0.02)
*Age*Female*	-0.11	-0.00	-0.16*	-0.23**
	(0.08)	(0.00)	(0.10)	(0.10)
*Education (demeaned)*Female*	-0.11	-0.04	0.13	0.15
	(0.66)	(0.03)	(0.77)	(0.76)
*Republican (demeaned)*Female*	-0.15***	-0.01***	-0.10***	-0.09**
	(0.03)	(0.00)	(0.04)	(0.04)
N	2986	2986	2986	2986
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Education (demeaned)* is a number from 1 to 9 that corresponds with education level (where the least education is 1 and the most education is 9), demeaned by the average. *Republican Leaning (demeaned)* is a number from 0 to 100 that is the extent to which the participant indicated feeling favorably about the Republican party, demeaned by the average. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all study versions involving evaluations of the participant's own math and science performance but excludes the participants who selected "other" as their educational attainment. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.13: Considering the relationship between beliefs and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-4.49***	-0.25***	-8.46***	-7.07***
	(0.84)	(0.04)	(1.14)	(1.11)
Belief (demeaned)	3.49***	0.17***	3.55***	3.52***
	(0.17)	(0.01)	(0.19)	(0.20)
*Belief (demeaned)*Female*	1.41***	0.05***	0.95***	1.15***
	(0.21)	(0.01)	(0.26)	(0.26)
N	2094	2094	2094	2094
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-4.01***	-0.21***	-8.49***	-9.18***
	(0.86)	(0.04)	(1.00)	(0.99)
Belief (demeaned)	2.25***	0.11***	2.55***	2.43***
	(0.14)	(0.01)	(0.15)	(0.16)
*Belief (demeaned)*Female*	0.73***	0.02	0.87***	0.89***
	(0.20)	(0.01)	(0.22)	(0.22)
N	2990	2990	2990	2990
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Belief (demeaned)* is the number of questions a participant believes he or she answered correctly, demeaned by the average belief. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Data are from all study versions involving evaluations of the participant's own math and science performance. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.14: Considering the relationship between beliefs relative to performance and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-4.49***	-0.26***	-8.36***	-6.91***
	(0.88)	(0.04)	(1.15)	(1.12)
Belief–Performance (demeaned)	3.87***	0.19***	3.75***	3.75***
	(0.15)	(0.01)	(0.18)	(0.18)
Belief–Performance (demeaned)	0.65***	0.01	0.56***	0.72***
**Female*	(0.18)	(0.01)	(0.21)	(0.22)
N	2094	2094	2094	2094
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-4.33***	-0.22***	-8.67***	-9.34***
	(0.89)	(0.04)	(1.01)	(1.01)
Belief–Performance (demeaned)	2.69***	0.13***	2.89***	2.76***
	(0.13)	(0.01)	(0.14)	(0.14)
Belief–Performance (demeaned)	-0.25	-0.03***	0.12	0.17
**Female*	(0.17)	(0.01)	(0.18)	(0.18)
N	2990	2990	2990	2990
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. *Belief–Performance (demeaned)* is the number of questions a participant believes he or she answered correctly minus the number of questions the participant actually answered correctly, demeaned by the average difference. Data are from all study versions involving evaluations of the participant's own math and science performance. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance (as in Panel 12 of Table 2). Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance (as in Panel 12 of Table 3).

Table A.15: Considering the relationship between general math and science beliefs and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-7.89***	-0.36***	-11.68***	-11.60***
	(2.44)	(0.10)	(3.06)	(2.82)
General Math Belief (demeaned)	6.80***	0.33***	9.61***	10.42***
	(1.17)	(0.05)	(1.42)	(1.17)
General Math Belief (demeaned)	0.76	-0.01	0.56	0.33
**Female*	(1.44)	(0.06)	(1.59)	(1.50)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-3.45*	-0.05	-5.97**	-7.30***
	(2.07)	(0.09)	(2.81)	(2.69)
General Math Belief (demeaned)	6.24***	0.30***	9.46***	9.65***
	(0.99)	(0.04)	(1.31)	(1.14)
General Math Belief (demeaned)	-0.46	-0.06	0.47	0.09
**Female*	(1.20)	(0.05)	(1.50)	(1.49)
N	294	294	294	294
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. *General Math Belief (demeaned)* is a participant's answer on a seven-point scale (where 1 is "strongly disagree" and 7 is "strongly agree" with the statement "In general, I perform well when asked questions that test my math and science skills"), demeaned by the average response. Data are from the *Private* version that was conducted in wave 5 when we added the general belief questions to the follow-up survey. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance. Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance.

Table A.16: Considering the relationship between general verbal beliefs and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-0.20	-0.18*	0.14	-2.23
	(2.13)	(0.10)	(2.76)	(2.57)
General Verbal Belief (demeaned)	8.40***	0.38***	11.37***	11.58***
	(1.24)	(0.05)	(1.28)	(1.20)
General Verbal Belief (demeaned)	-1.39	-0.09	-1.53	-1.91
**Female*	(1.50)	(0.07)	(1.52)	(1.46)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-1.92	-0.09	-2.96	-3.00
	(1.74)	(0.08)	(2.38)	(2.23)
General Verbal Belief (demeaned)	5.41***	0.29***	9.33***	9.54***
	(1.13)	(0.04)	(1.18)	(1.12)
General Verbal Belief (demeaned)	1.14	-0.05	0.24	0.06
**Female*	(1.37)	(0.06)	(1.39)	(1.38)
N	305	305	305	305
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. *General Verbal Belief (demeaned)* is a participant's answer on a seven-point scale (where 1 is "strongly disagree" and 7 is "strongly agree" with the statement "In general, I perform well when asked questions that test my verbal skills"), demeaned by the average response. Data are from the *Private (Verbal)* version that was conducted in wave 5 when we added the general belief questions to the follow-up survey. Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance. Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance.

Table A.17: Probability of being hired

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
Panel 1: Employers' hiring decisions				
Answer	0.01***	0.18***	0.01***	0.01***
	(0.00)	(0.01)	(0.00)	(0.00)
Constant	0.07**	-0.02	0.08***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)
N	1490	1788	1490	1490
Panel 2: Workers' expected probability of being hired				
Female	-0.09***	-0.11***	-0.12***	-0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
N	1192	1192	1192	1192
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Panel 1 presents results on the decisions made by the employers in the *Employer* version. The results are from a linear probability model of the likelihood that an employer indicates they will hire a worker in a decision, with SEs clustered by employer. *Answer* is the answer to the self-evaluation question they were asked to consider in that decision. Panel 2 presents results on the expected probability of a worker being hired in the *Self-Promotion* or *Self-Promotion (Risky)* version. The results are from a linear probability model of the expected probability of a worker being hired, with SEs clustered by worker. For each worker, their expected probability of being hired was calculated as the average probability of being hired when considering all employers who made hiring decisions in response to the answer on the self-evaluation they provided. *Female* is an indicator for the worker being female. Performance FEs are dummies for each possible worker performance out of the 20 questions on the test. The columns restrict to the data associated with the noted question, as defined in the notes of Table 2.

Table A.18: Wages

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
Panel 1: Employers' wage decisions				
Answer	0.21***	4.26***	0.22***	0.21***
	(0.02)	(0.27)	(0.02)	(0.02)
Constant	22.70***	18.95***	21.94***	22.76***
	(0.75)	(0.70)	(0.61)	(0.78)
N	1490	1788	1490	1490
Panel 2: Workers' expected wage				
Female	-1.77***	-2.13***	-2.16***	-1.89***
	(0.49)	(0.37)	(0.56)	(0.53)
N	1192	1192	1192	1192
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Panel 1 presents results on the decisions made by the employers in the *Employer* version. The results are from OLS regressions of the wage an employer chose in a decision with SEs clustered by employer. *Answer* is the answer to the self-evaluation question they were asked to consider in that decision. Panel 2 presents results on the expected wage of a worker in the *Self-Promotion* or *Self-Promotion (Risky)* version. The results are from an OLS of the expected wage of a worker, with SEs clustered by worker. For each worker, their expected wage was calculated as the average wage when considering all employers who made hiring decisions in response to the answer on the self-evaluation they provided. *Female* is an indicator for the worker being female. Performance FEs are dummies for each possible worker performance out of the 20 questions on the test. The columns restrict to the data associated with the noted question, as defined in the notes of Table 2.

Table A.19: Predictions about performance

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Predictions about women	-1.54***	-1.47***	-2.25***	-2.24***
	(0.17)	(0.17)	(0.18)	(0.19)
Female predictor	0.18	0.25	0.22	0.36
	(0.20)	(0.21)	(0.20)	(0.23)
Predictions about women	0.21	0.08	0.01	-0.21
**Female predictor*	(0.24)	(0.24)	(0.27)	(0.29)
Constant	11.98***	12.49***	12.40***	13.04***
	(0.14)	(0.15)	(0.14)	(0.16)
N	1198	1198	1198	1198

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are clustered at the participant level. Results are from OLS regressions of the predicted average performance (i.e., the average number of questions answered correctly by a set of female participants or a set of male participants) based on the gender's average response to the question noted in the column. (Average responses are from the *Self-Promotion* version after information about absolute and relative performance on the test has been provided.) *Predictions about women* is an indicator that the question elicited a prediction for the average performance of female workers. *Female predictor* is an indicator for the predictor being female. Data are from the study versions conducted in wave 5 when we added the prediction questions to the follow-up survey.

Table A.20: Among our youth sample: considering the relationship between performance and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-11.63***	-0.54***	-4.98***	-8.19***
	(0.45)	(0.02)	(0.58)	(0.54)
Performance (demeaned)	4.55***	0.24***	2.15***	3.97***
	(0.14)	(0.01)	(0.18)	(0.17)
*Performance (demeaned)*Female*	-0.61***	-0.05***	-0.24	-0.56**
	(0.21)	(0.01)	(0.26)	(0.24)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-7.20***	-0.29***	-3.73***	-6.11***
	(0.53)	(0.03)	(0.60)	(0.59)
Performance (demeaned)	4.99***	0.27***	2.66***	4.17***
	(0.18)	(0.01)	(0.19)	(0.18)
*Performance (demeaned)*Female*	-0.50**	-0.01	-0.63**	-0.76***
	(0.25)	(0.01)	(0.27)	(0.26)
N	10637	10637	10637	10637
Performance FEs	No	No	No	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the evaluation question noted in each column (additional details on the question wording can be found in Appendix D.8). *Female* is an indicator for the participant being female in the administrative data provided by Character Lab Research Network. *Performance (demeaned)* is the number of questions a participant answered correctly out of the 10 questions on the test, demeaned by the average performance. Data are from the study among youth (i.e., middle-school and high-school students). Panel 1 analyzes evaluations from before participants are informed of their absolute performance. Panel 2 analyzes evaluations from after participants are informed of their absolute performance.

Table A.21: Among our youth sample: considering the relationship between beliefs and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-3.83***	-0.14***	-0.93	-2.48***
	(0.36)	(0.02)	(0.57)	(0.51)
Belief (demeaned)	7.22***	0.37***	4.21***	5.41***
	(0.15)	(0.01)	(0.20)	(0.18)
*Belief (demeaned)*Female*	-0.06	-0.00	-0.70***	-0.32
	(0.18)	(0.01)	(0.25)	(0.23)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-3.15***	-0.10***	-0.64	-2.48***
	(0.51)	(0.03)	(0.60)	(0.59)
Belief (demeaned)	3.51***	0.16***	3.11***	3.34***
	(0.18)	(0.01)	(0.20)	(0.20)
*Belief (demeaned)*Female*	-0.26	-0.00	-0.57**	-0.39
	(0.22)	(0.01)	(0.26)	(0.26)
N	10637	10637	10637	10637
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the evaluation question noted in each column (additional details on the question wording can be found in Appendix D.8). *Female* is an indicator for the participant being female in the administrative data provided by Character Lab Research Network. *Belief (demeaned)* is the number of questions a participant believes he or she answered correctly out of the 10 questions on the test, demeaned by the average belief. Performance FEs are dummies for each possible performance out of the 10 questions on the test. Data are from the study among youth (i.e., middle-school and high-school students). Panel 1 analyzes evaluations from before participants are informed of their absolute performance. Panel 2 analyzes evaluations from after participants are informed of their absolute performance.

Table A.22: Among our youth sample: considering the relationship between beliefs relative to performance and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-3.85***	-0.14***	-0.94	-2.50***
	(0.35)	(0.02)	(0.57)	(0.51)
Belief–Performance (demeaned)	7.00***	0.35***	4.11***	5.21***
	(0.15)	(0.01)	(0.19)	(0.18)
Belief–Performance (demeaned)	0.34**	0.03***	-0.51**	0.07
**Female*	(0.17)	(0.01)	(0.24)	(0.22)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-3.16***	-0.10***	-0.67	-2.51***
	(0.51)	(0.03)	(0.60)	(0.59)
Belief–Performance (demeaned)	3.39***	0.16***	2.86***	3.08***
	(0.18)	(0.01)	(0.20)	(0.20)
Belief–Performance (demeaned)	-0.04	-0.01	-0.10	0.10
**Female*	(0.22)	(0.01)	(0.25)	(0.25)
N	10637	10637	10637	10637
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the evaluation question noted in each column (additional details on the question wording can be found in Appendix D.8). *Female* is an indicator for the participant being female in the administrative data provided by Character Lab Research Network. *Belief–Performance (demeaned)* is the number of questions a participant believes he or she answered correctly minus the number of questions the participant actually answered correctly, demeaned by the average difference. Performance FEs are dummies for each possible performance out of the 10 questions on the test. Data are from the study among youth (i.e., middle-school and high-school students). Panel 1 analyzes evaluations from before participants are informed of their absolute performance. Panel 2 analyzes evaluations from after participants are informed of their absolute performance.

Table A.23: Among our youth sample: considering the relationship between racial minority status and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-9.71***	-0.47***	-4.70***	-7.90***
	(0.72)	(0.04)	(0.97)	(0.87)
Racial Minority	-1.11*	-0.09***	3.74***	-0.20
	(0.65)	(0.03)	(0.89)	(0.78)
*Racial Minority*Female*	-2.39***	-0.08*	-0.38	-0.05
	(0.91)	(0.05)	(1.20)	(1.10)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-5.67***	-0.23***	-3.73***	-6.81***
	(0.87)	(0.05)	(1.01)	(0.97)
Racial Minority	-1.80**	-0.13***	2.50***	-0.48
	(0.81)	(0.04)	(0.93)	(0.89)
*Racial Minority*Female*	-1.48	-0.04	0.23	1.60
	(1.08)	(0.06)	(1.24)	(1.21)
N	10637	10637	10637	10637
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the evaluation question noted in each column (additional details on the question wording can be found in Appendix D.8). *Female* is an indicator for the participant being female in the administrative data provided by Character Lab Research Network. *Racial Minority* is an indicator that the participant is not classified as a non-Hispanic White in the administrative data. Performance FEs are dummies for each possible performance out of the 10 questions on the test. Data are from the study among youth (i.e., middle-school and high-school students). Panel 1 analyzes evaluations from before participants are informed of their absolute performance. Panel 2 analyzes evaluations from after participants are informed of their absolute performance.

Table A.24: Among our youth sample: considering the relationship between FRPL status and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-10.64***	-0.51***	-4.93***	-7.80***
	(0.56)	(0.03)	(0.74)	(0.67)
FRPL	-1.20*	-0.09***	1.16	-1.69**
	(0.67)	(0.03)	(0.88)	(0.80)
*FRPL*Female*	-1.68*	-0.03	-0.02	-0.38
	(0.93)	(0.05)	(1.19)	(1.11)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-6.91***	-0.27***	-3.94***	-5.97***
	(0.67)	(0.04)	(0.77)	(0.74)
FRPL	-1.02	-0.07	0.02	-1.37
	(0.81)	(0.04)	(0.91)	(0.89)
*FRPL*Female*	0.64	0.02	0.92	0.55
	(1.06)	(0.06)	(1.22)	(1.20)
N	10637	10637	10637	10637
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the evaluation question noted in each column (additional details on the question wording can be found in Appendix D.8). *Female* is an indicator for the participant being female in the administrative data provided by Character Lab Research Network. *FRPL* is an indicator for the participant qualifying for free and reduced-price lunch according to the administrative data. Performance FEs are dummies for each possible performance out of the 10 questions on the test. Data are from the study among youth (i.e., middle-school and high-school students). Panel 1 analyzes evaluations from before participants are informed of their absolute performance. Panel 2 analyzes evaluations from after participants are informed of their absolute performance.

Table A.25: Among our youth sample: considering the relationship between GPA and evaluations

Question:	Performance	Performance-Bucket	Willingness-to-Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before performance information is provided)				
Female	-11.87***	-0.54***	-5.36***	-8.92***
	(0.46)	(0.02)	(0.59)	(0.54)
GPA (demeaned)	0.16***	0.01***	0.11***	0.27***
	(0.03)	(0.00)	(0.04)	(0.04)
*GPA (demeaned)*Female*	0.02	-0.00	-0.00	0.04
	(0.04)	(0.00)	(0.05)	(0.05)
Panel 2: Informed Evaluations (after performance information is provided)				
Female	-6.80***	-0.27***	-4.19***	-6.63***
	(0.53)	(0.03)	(0.61)	(0.59)
GPA (demeaned)	0.09**	0.00*	0.20***	0.29***
	(0.04)	(0.00)	(0.04)	(0.04)
*GPA (demeaned)*Female*	-0.11**	-0.00*	-0.06	-0.07
	(0.05)	(0.00)	(0.05)	(0.05)
N	10618	10618	10618	10618
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the evaluation question noted in each column (additional details on the question wording can be found in Appendix D.8). *Female* is an indicator for the participant being female in the administrative data provided by Character Lab Research Network. *GPA (demeaned)* is administrative data on participants' "overall marking period GPA" that ranges from 35 to 102, demeaned by the average. Data are from the study among youth (i.e., middle-school and high-school students) excluding the youth for whom we do not have a GPA recorded. Panel 1 analyzes evaluations from before participants are informed of their absolute performance. Panel 2 analyzes evaluations from after participants are informed of their absolute performance.

B The *Free-Response* Versions

In February 2019, we recruited 399 participants on MTurk to complete either the *Free-Response Employer* version (n=198) or the *Free-Response Predictor* version (n=201) of our study. In July 2021, we recruited 201 participants on MTurk to complete the *Free-Response Coding* version. Each participant received a guaranteed completion fee, which equaled $1.50 for the study versions run in 2019 and $3 for the study version run in 2021. After participants completed all decisions of the study, they took a short follow-up survey that collected demographic information.

In the *Free-Response Employer* version, participants made 21 hiring decisions. In the *Free-Response Predictor* version, participants made 21 sets of predictions. In the *Free-Response Coding* version, participants made 21 coding decisions. Before making each decision or set of predictions, participants were provided with the text—but no other information—entered by a wave 1 participant to the free-response question: "Please describe how well you think you performed on the test that you took in part 1 and why." The free response either came from part 2 or part 3. Participants were randomly assigned these 21 free responses from the set of eligible free responses written by the participants from wave 1.[13]

Participants assigned to the *Free-Response Employer* version were asked whether they would like to hire the participant who provided that free response and, if so, how much to pay them. One of their decisions—out of the 21 decisions in the study—was selected to determine a possible bonus payment for them and for an associated "worker."[14] The payoffs resulting from the one randomly selected decision for these employers are the same as described in the *Employer* version.

Participants assigned to the *Free-Response Predictor* version were asked to predict whether the participant who wrote the free response was male or female and how many questions, out of 20, that participant answered correctly on the math and science test. The payoffs for predictors were determined as follows. One of the two predictions from one of the 21 sets was randomly selected. If the prediction was correct, the predictor received a bonus payment of 50 cents.

Participants assigned to the *Free-Response Coding* version were asked to indicate either "yes" or "no" to whether the participant who wrote the free response was engaging in self-promotion.

Relative to the *Employer* version discussed in the main text, there are three important differences when considering results from the *Free-Response* versions. First, since there is no objective way to rank free-response answers, we cannot examine how hiring decisions or predictions vary as the responses improve (as we did when examining, e.g., the impact of a one unit increase on a 0-to-100 scale in the *Employer*

[13]Not all of the free responses collected in wave 1 of the study were evaluated. A research assistant—blinded to participant gender and study version—deemed 130 of the 1800 free responses "ineligible" due to the answer not relating to the question asked or due to severe grammar and/or spelling issues that made an answer incomprehensible. Consequently, the participants were each randomly shown 21 free-responses from the set of 1670 eligible free responses. Finally, note that some eligible free-responses were never randomly selected to be shown to a participant.

[14]Each participant who completed the *Self-Promotion* or *Self-Promotion (Risky)* versions of our study was matched with an employer from the *Employer* version of our study and received corresponding payoffs from their employers' hiring decisions. By contrast, only select participants from the *Self-Promotion* and *Self-Promotion (Risky)* versions were matched with a participant from a *Free-Response Employer* version, and received corresponding payoffs, rather than everyone. Since we also wanted to collect data on the free responses from the *Private* version, participants in the *Free-Response Employer* version were (accurately) told that one of their decisions would be selected to count but *not* that one of their decisions would be randomly selected to count (as this would have required putting 0% weight on free responses from the *Private* version in the randomization). .

version). Second, while participants are not informed of the gender of the individual who answered the free-response question, they may be able to infer gender—to some degree—given how the free responses are written. Below, we test this hypothesis using data from the predictors. Third, given the large number of possible free responses, we are underpowered to consider the effect of specific free responses.

For these reasons, we favor the analysis of the quantitative responses to the self-evaluation questions presented in the main text to examine the gender gap in self-promotion. Here, however, we investigate the hiring decisions and predictions from the *Free-Response* versions to present several interesting (but inherently secondary) results. Given our power issues, we jointly analyze free responses from all three study versions run in wave 1.

Table B.1 presents results from regressions testing whether the gender of the free response author affects how responses are coded, predictions, and hiring decisions. Column (1) is estimated from ratings in the *Free-Response Coding* version. The negative coefficients on *Female* in column (1) show that participants are less likely (at least directionally) to indicate that female participants engage in self-promotion given their free responses. Columns (2) and (3) are estimated from predictions from the *Free-Response Predictor* version. The negative coefficients on *Female* in column (2) show that participants predict (at least directionally) lower scores when reading free responses authored by female participants. This evidence is consistent with our findings from the quantitative self-evaluation questions discussed in the main text—women appear to provide less favorable subjective evaluations of their performance, even in the free responses. The positive coefficients on *Female* in column (3) show that, even though predictors are not informed of the gender of the participant who authored the free response, evaluators can infer gender—to some degree—when viewing the responses. Predictors are significantly more likely to predict that a response was written by a female participant when it was indeed written by a female participant. Column (4) is estimated from hiring decisions from the *Free-Response Employer* version. Based on the free response answers, employers pay at least directionally less to female workers.

An important caveat to the analysis in the prior paragraph, however, is that since evaluators can infer the gender of the associated worker based off of the free responses, the predictions of performance and hiring decisions may be influenced by the perception of the gender of the free response author (e.g., predictors might expect women to perform worse than men; employers may want to pay women more than men based on social preferences, etc.), which makes it difficult to isolate the effect of the language used in the free response (i.e., the self-promotion). As mentioned in footnote 6 and in the main text of the paper, difficulties with using free responses, and other qualitative data, contribute to our decision to focus our analysis on the quantitative self-evaluation questions we explore in the main text of the paper.

Table B.1: Free Response Regressions

	Coded as Self-Promotion	Predicted Performance	Predicted Probability Female	Expected Wage
	(1)	(2)	(3)	(4)
Panel 1: Free responses (before performance information is provided)				
Female	-0.07***	-0.67***	0.08***	-1.44*
	(0.03)	(0.22)	(0.03)	(0.81)
N	764	749	749	743
Panel 2: Free responses (after performance information is provided)				
Female	-0.03	-0.35	0.09***	-0.66
	(0.03)	(0.23)	(0.03)	(1.04)
N	757	773	773	755
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the noted dependent variable (DV). *Coded as Self-Promotion* equals 1 if the predictor indicated that the the participant was engaging in self-promotion and 0 otherwise. *Predicted Performance* equals the predictor's guess of the number of questions the participant answered correctly based on the free response. *Predicted Probability Female* equals the probability that the predictor placed on the participant being female.*Expected Wage* equals the expected wage given to the participant by an employer. In cases where multiple participants responded to the same free response, we use the average decision (e.g., if a free response is predicted to be written by a female participant once and a male participant once, that participant is recorded as being predicted to be female with a 0.50 probability). *Female* is an indicator for the participant who wrote the free response being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test of the participant who wrote the free response. Data in Panel 1 are from free responses elicited before performance information is provided to participants, and data in Panel 2 are from free responses elicited after performance information is provided to participants. Neither predictors nor employers were provided with any information on participants aside from these free responses. Data are from all three study versions run in wave 1: the *Self-Promotion* version, the *Self-Promotion (Risky)* version, and the *Private* version.

C Methodological Note: The Role of Beliefs

To explore how the gender gap varies by beliefs about absolute performance, the specifications in Appendix Table A.13 add a linear control for participants' beliefs about their absolute performance on the test. The results in Panel 1 show that—holding performance (i.e., the number of questions they answered correctly) constant—a more optimistic belief about their absolute performance (i.e., the number of questions they believe they answered correctly) is associated with more favorable self-evaluations. This relationship is even stronger for women, suggesting that the gender gap is larger among those who were more pessimistic about their absolute performance and smaller among those who were more optimistic about it. We see similar results in Appendix Table A.14, which replaces the linear belief control with a linear control for the gap between a participant's belief and their actual performance.

Intriguingly, the results in Panel 2 of Appendix Tables A.13 and A.14 show that beliefs about absolute performance are *still* correlated with self-evaluations after participants have been informed about their absolute and relative performance on the test. That is, individuals who initially thought they answered fewer questions correctly on the test *still* evaluate their performance less favorably *even after they learn how many questions they answered correctly on the test*. Why could this be? One explanation is that there are certain types of individuals who view their performance in math and science more positively than others or view their performance more negatively than others. Such positive types could subjectively evaluate their performance more positively in self-evaluations *and* overestimate their absolute performance. Such negative types could subjectively evaluate their performance less positively in self-evaluations *and* underestimate their absolute performance. Because such a type is not caused by the belief about absolute performance (indeed the type could cause the belief), the subjective evaluations continue to be influenced by the type, even after individuals are perfectly informed of their absolute (and relative) performance.

To further explore the possibility that certain types of individuals systematically view their math and science performance less favorably than others, we added two questions to the follow-up survey in our fifth wave of data collection to measure broader beliefs about performance.

One question asked participants to indicate their agreement (on a seven-point Likert scale from "strongly disagree" to "strongly agree") with a statement that reads "In general, I perform well when asked questions that test my math and science skills." As shown in Appendix Table A.15, answers to this question are highly and positively predictive of subjective evaluations that relate to math and science skills in the *Private* version (and equally so for men and women). The other question asked participants to indicate their agreement (on the same scale) with a statement that reads "In general, I perform well when asked questions that test my verbal skills." As shown in Appendix Table A.16, answers to this question are also highly and positively predictive of subjective evaluations that relate to verbal skills in the *Private (Verbal)* version (and, again, equally so for both men and women).[15]

These results further suggest the possibility of positive and negative types noted above and is consistent with individuals allowing their general perception of their math and science skills (or their verbal skills) to influence their perceptions of their specific performance on the math and science test (or verbal test) they take in our experiment.

[15]If we simultaneously include both performance beliefs and these broader beliefs in a regression, both measures of beliefs are positive and statistically significant.

The presence of types like those posited above highlights why caution is warranted when trying to assess the role of beliefs about absolute performance in contributing to self-evaluations by statistically controlling for reported beliefs about absolute performance. Such results may be confounded by measurement error, omitted variable bias (which could be caused by the positive and negative types discussed above), or reverse causality. Indeed, absent the relevance of such confounds, one cannot explain why the reported beliefs about absolute performance remain statistically significant even after participants are perfectly informed of their absolute and relative performance.

Controlling for beliefs by design—by providing participants with precise information on their absolute and relative performance prior to eliciting their informed self-evaluations—allows us to avoid these potential confounds. Thus, it is interesting to note that the apparent relevance of beliefs about performance in explaining the gender gap in self-evaluations is dependent on whether we control for beliefs by design or instead control for beliefs statistically. This is shown most clearly by the results in Appendix Table C.1. Panel 1 presents results the gender gap in self-evaluations before performance information is provided. Panel 2 shows the gender gap in self-evaluations after participants are perfectly informed of their absolute and relative performance and thus after controlling for these beliefs by design. Panel 3 returns to analyzing the data before performance information is provided but now adds in a fixed effect for each reported belief about absolute performance and hence controls for beliefs statistically. While a comparison between Panels 1 and 2 makes clear that beliefs about absolute and relative performance explain the minority of the gender gap in self-evaluations, a comparison between Panels 1 and 3 would have instead suggested that beliefs about absolute performance alone explain the majority of the gender gap in self-evaluations. Thus, when a research question asks what role beliefs play in driving some outcome (i.e., rather than how beliefs update in response to information in which case focusing on measured beliefs is essential), it may be preferable to control for beliefs "by design" than to measure beliefs and control for them statistically.

Table C.1: Statistically controlling for beliefs versus controlling for beliefs by design

Question:	Performance	Performance-Bucket	Willingness to Apply	Success
	(1)	(2)	(3)	(4)
Panel 1: Evaluations (before information)				
Female	-13.83***	-0.67***	-17.28***	-16.12***
	(1.13)	(0.05)	(1.31)	(1.32)
Belief FEs	No	No	No	No
Panel 2: Informed Evaluations (after information)				
Female	-9.84***	-0.46***	-14.75***	-14.60***
	(1.09)	(0.05)	(1.29)	(1.29)
N	2094	2094	2094	2094
Belief FEs	No	No	No	No
Panel 3: Evaluations (before information) with belief controls				
Female	-4.45***	-0.24***	-8.39***	-6.88***
	(0.88)	(0.04)	(1.16)	(1.14)
Belief FEs	Yes	Yes	Yes	Yes
N	2094	2094	2094	2094
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the responses provided to the question noted in each column, as defined in the notes of Table 2. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance out of the 20 questions on the test. Belief FEs are dummies for each possible belief about how many questions the participant answered correctly out of the 20 questions on the test. Data are from all versions that elicit evaluations of math and science performance before and after participants are informed of their absolute and relative performance (i.e., all but the *Private (Immediately Informed)* version, *Private (Other-Evaluation)* version, and *Private (Verbal)* version). Panel 1 analyzes evaluations from before participants are informed of their absolute and relative performance, reproducing Panel 12 of Table 2. Panel 2 analyzes evaluations from after participants are informed of their absolute and relative performance from the same participants presented in Panel 1. Panel 3 analyzes evaluations from before participants are informed of their absolute and relative performance but adds Belief FEs to control for beliefs statistically.

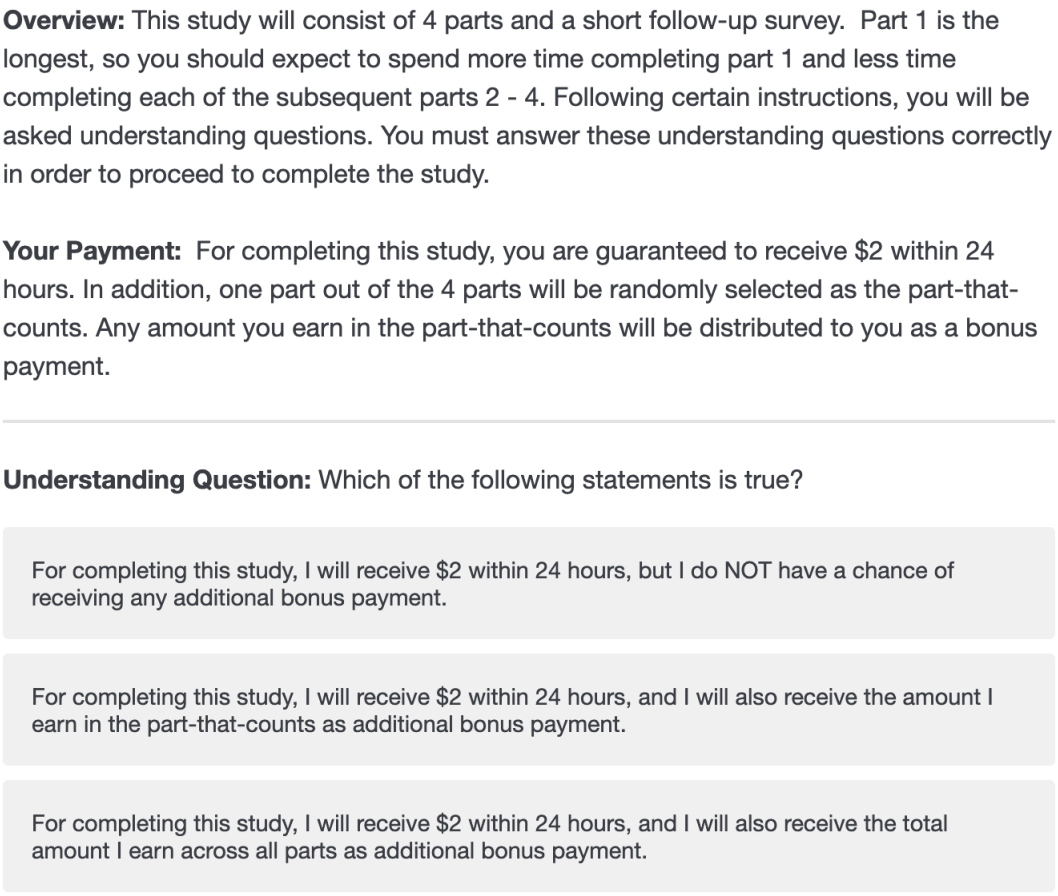
D Experimental Instructions

D.1 Instructions for *Self-Promotion* version

Prior to participating in the study, participants must correctly answer a captcha and consent to participate. At the end of the study, participants must complete a short follow-up survey to gather demographic information.

The study begins by informing each participant of the $2 study completion fee and of the opportunity to earn additional payment for themselves. Figure D.1 shows how this payment information is explained along with the understanding question that the participant must answer correctly to proceed.

Figure D.1: Payment Information



The instructions for part 1 are displayed in Figures D.2 and an example of an ASVAB question is displayed in Figure D.3 (note that the timer in that screenshot indicates the participant has 23 seconds left to answer the question although the timer starts at 30 seconds).

Figure D.2: Instructions for Part 1

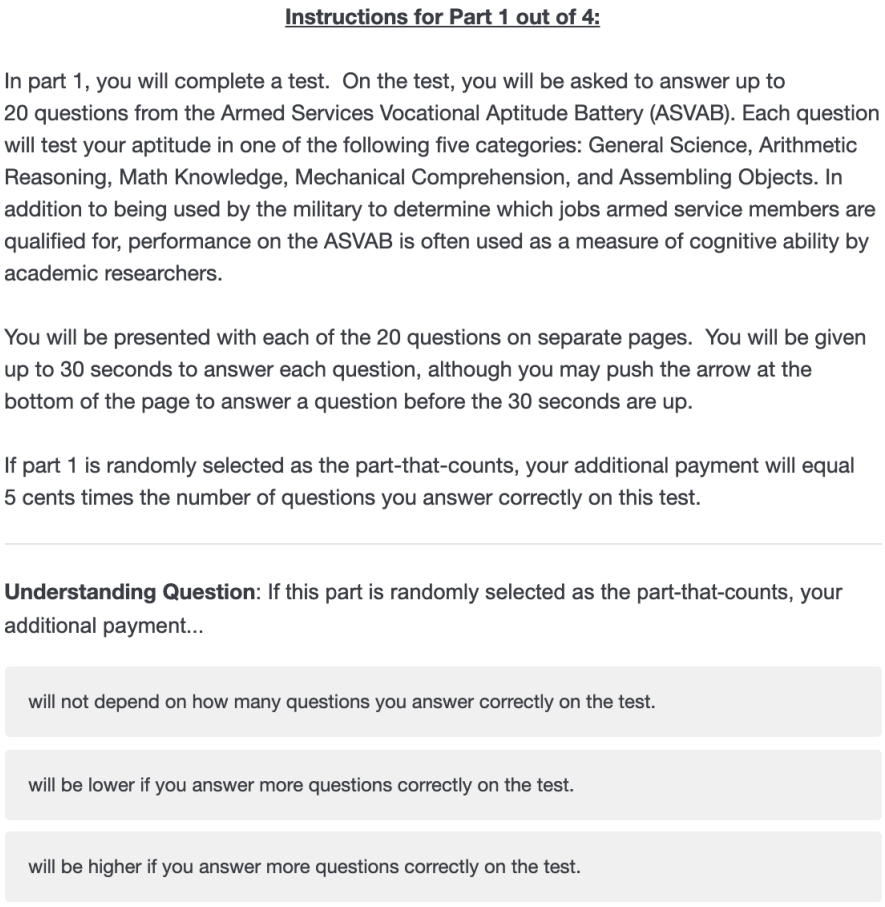
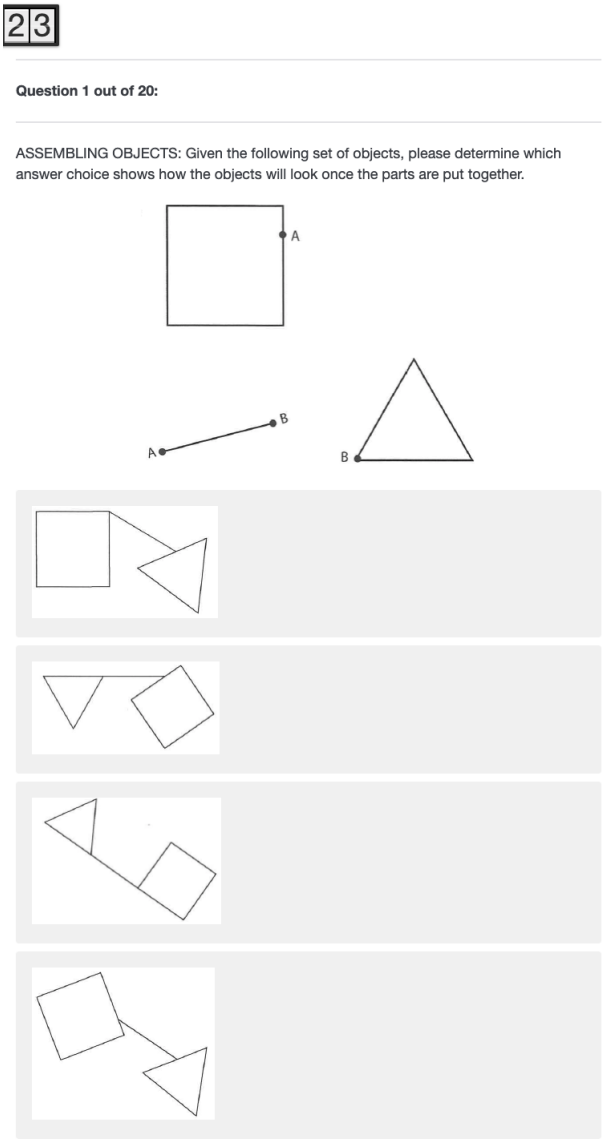
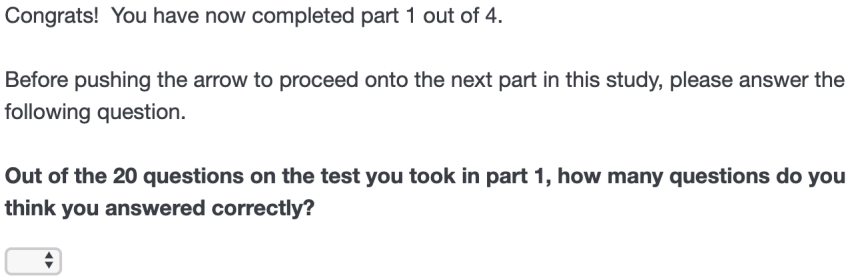


Figure D.3: Part 1: Example ASVAB question



After completing the ASVAB questions in part 1 but before proceeding to part 2, participants are asked about their absolute performance belief, as shown in Figure D.4.

Figure D.4: Absolute Performance Belief Question



Participants then receive instructions for part 2 (see Figure D.5), must correctly answer understanding questions about those instructions (see Figure D.6), and then are asked the self-evaluation questions (see Figure D.7).

Figure D.5: Part 2 Instructions

Instructions for Part 2 out of 4:

In part 2, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

One of your answers to these questions will be shown to "your part 2 employer," who will be another MTurk worker who completes a different version of this study. Your part 2 employer can decide whether to hire you and, if so, how much to pay you.

Prior to deciding whether to hire you and, if so, how much to pay you, your part 2 employer will NOT be informed of how many questions you answered correctly on the test in part 1.

If this part is randomly selected as the part-that-counts, the additional payment given to your part 2 employer and to you will be determined as follows:

- If your part 2 employer chooses NOT to hire you, your additional payment will equal 25 cents and your part 2 employer's additional payment will equal 100 cents.

- If your part 2 employer chooses to hire you, your additional payment will equal how much they choose to pay you, and your part 2 employer's additional payment will equal (i) 100 cents minus how much they choose to pay you, plus (ii) 5 cents times the number of questions you answered correctly on the test in part 1. Your part 2 employer can choose to pay you any amount between 25 cents and 100 cents.

Figure D.6: Part 2 Understanding Questions

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will equal 25 cents if you are not hired and the amount your part 2 employer chooses to pay you if you are hired.

Understanding Question: When deciding how much to pay you, your part 2 employer will only know...

how many questions you answered correctly on the test you took in part 1.

how you answer one of the questions -- on the next page -- about your performance on the test you took in part 1.

how you answer all of the questions -- on the next page -- about your performance on the test you took in part 1.

Figure D.7: Part 2 Self-Evaluation Questions

Now, please answer the five questions below to complete part 2. Note that, although the final three questions appear in the same block, they are three separate questions.

Please describe how well you think you performed on the test that you took in part 1 and why.

Please indicate how well you think you performed on the test you took in part 1.

Terrible	Very Poor	Neutral	Good	Very Good	Exceptional

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with each of the following statements:

Entirely Disagree	Strongly Disagree	Disagree	Somewhat Disagree	Neither Disagree Nor Agree	Somewhat Agree	Agree	Strongly Agree	Entirely Agree
0	10	20	30	40	50	60	70	80	90	100

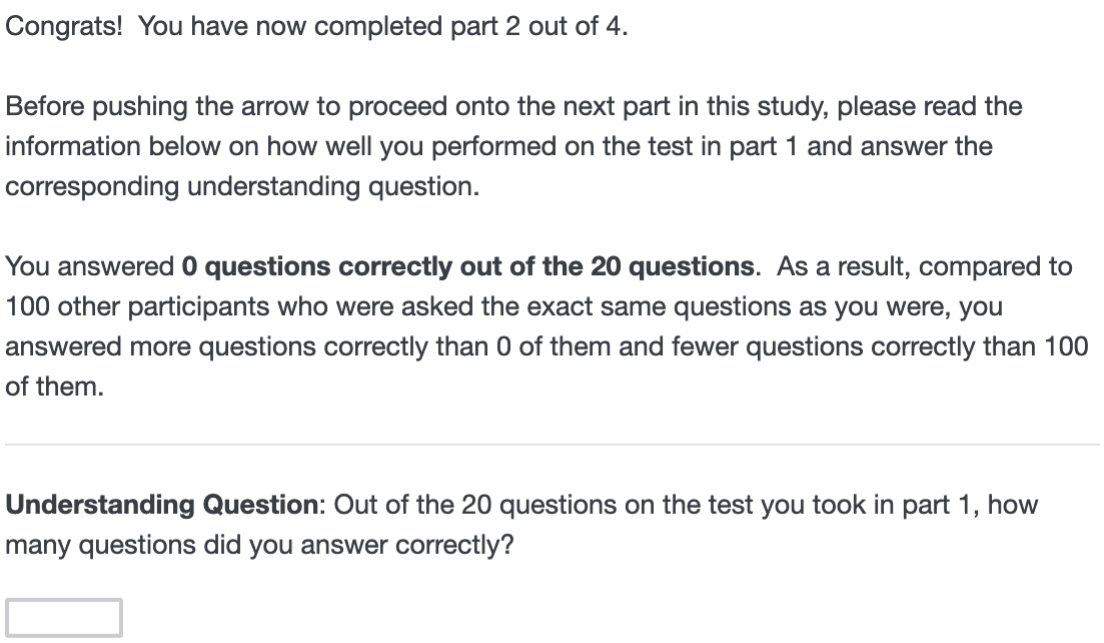
I performed well on the test I took in part 1.

I would apply for a job that required me to perform well on the test I took in part 1.

I would succeed in a job that required me to perform well on the test I took in part 1.

After completing part 2, participants are provided with perfect information on their absolute and relative performance and are required to correctly report back their absolute performance as shown in Figure D.8.

Figure D.8: Absolute and Relative Performance Information



In part 3, participants are provided with the same instructions (see Figure D.9), understanding questions (see Figure D.10), and self-evaluation questions (see Figure D.11) as they were in part 2.

Figure D.9: Part 3 Instructions

Instructions for Part 3 out of 4:

In part 3, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

One of your answers to these questions will be shown to "your part 3 employer," who will be another MTurk worker who completes a different version of this study. Your part 3 employer can decide whether to hire you and, if so, how much to pay you.

Prior to deciding whether to hire you and, if so, how much to pay you, your part 3 employer will NOT be informed of how many questions you answered correctly on the test in part 1 (even though you were informed of this information on the previous page).

If this part is randomly selected as the part-that-counts, the additional payment given to your part 3 employer and to you will be determined as follows:

- If your part 3 employer chooses NOT to hire you, your additional payment will equal 25 cents and your part 3 employer's additional payment will equal 100 cents.

- If your part 3 employer chooses to hire you, your additional payment will equal how much they choose to pay you, and your part 3 employer's additional payment will equal (i) 100 cents minus how much they choose to pay you, plus (ii) 5 cents times the number of questions you answered correctly on the test in part 1. Your part 3 employer can choose to pay you any amount between 25 cents and 100 cents.

Figure D.10: Part 3 Understanding Questions

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will equal 25 cents if you are not hired and the amount your part 3 employer chooses to pay you if you are hired.

Understanding Question: When deciding how much to pay you, your part 3 employer will only know...

how many questions you answered correctly on the test you took in part 1.

how you answer one of the questions -- on the next page -- about your performance on the test you took in part 1.

how you answer all of the questions -- on the next page -- about your performance on the test you took in part 1.

Figure D.11: Part 3 Self-Evaluation Questions

Now, please answer the five questions below to complete part 3. Note that, although the final three questions appear in the same block, they are three separate questions.

Please describe how well you think you performed on the test that you took in part 1 and why.

Please indicate how well you think you performed on the test you took in part 1.

Terrible	Very Poor	Neutral	Good	Very Good	Exceptional

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement:

Entirely Disagree		Strongly Disagree		Disagree		Somewhat Disagree		Neither Disagree Nor Agree		Somewhat Agree		Agree		Strongly Agree		Entirely Agree
0	10		20		30		40		50		60		70		80		90	100

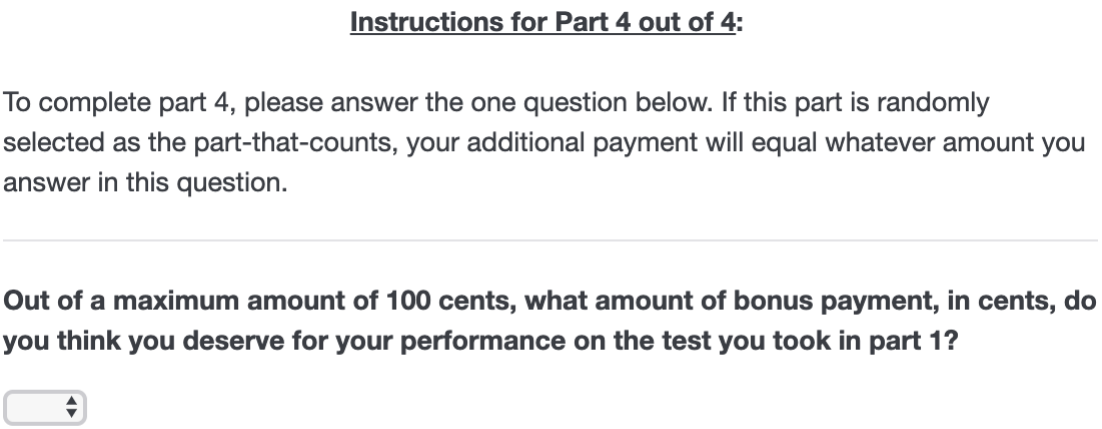
I performed well on the test I took in part 1.

I would apply for a job that required me to perform well on the test I took in part 1.

I would succeed in a job that required me to perform well on the test I took in part 1.

Finally, participants receive instructions about and are asked to answer the deservingness question in Part 4 (see Figure D.12). They then answer demographic questions, including the one that asks about their gender.

Figure D.12: Part 4 Instructions and Deservingness Question



D.2 Instructions for the *Self-Promotion (Risky)* version

The *Self-Promotion (Risky)* version of the study proceeds in the same manner as the *Self-Promotion* version of the study, except for the instructions about part 2 and part 3. Participants are informed that there is some chance that their employer will learn their actual performance. See Figures D.13 and D.14 for these instructions and the corresponding understanding questions, respectively.

Figure D.13: The *Self-Promotion (Risky)* version: Part 2 Instructions

Instructions for Part 2 out of 4:

In part 2, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

There is some chance that one of your answers to these questions will be shown to "your part 2 employer," who will be another MTurk worker who completes a different version of this study. Your part 2 employer can decide whether to hire you and, if so, how much to pay you.

Prior to deciding whether to hire you and, if so, how much to pay you, there is also some chance that your part 2 employer will be informed of how many questions you answered correctly on the test in part 1.

However, while your part 2 employer may learn one of your answers to the questions -- on the next page -- related to your performance on the test in part 1 and/or how many questions you answered correctly on the test in part 1, it is also possible that your part 2 employer will not learn any information related to your performance prior to deciding whether to hire you and, if so, how much to pay you.

If this part is randomly selected as the part-that-counts, the additional payment given to your part 2 employer and to you will be determined as follows:

- If your part 2 employer chooses NOT to hire you, your additional payment will equal 25 cents and your part 2 employer's additional payment will equal 100 cents.

- If your part 2 employer chooses to hire you, your additional payment will equal how much they choose to pay you, and your part 2 employer's additional payment will equal (i) 100 cents minus how much they choose to pay you, plus (ii) 5 cents times the number of questions you answered correctly on the test in part 1. Your part 2 employer can choose to pay you any amount between 25 cents and 100 cents.

Figure D.14: The *Self-Promotion (Risky)* version: Part 2 Understanding Questions

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will equal 25 cents if you are not hired and the amount your part 2 employer chooses to pay you if you are hired.

Understanding Question: When deciding how much to pay you, your part 2 employer will...

definitely know how many questions you answered correctly on the test you took in part 1.

definitely know how you answer all of the questions -- on the next page -- about your performance on the test you took in part 1.

will know nothing about your performance on the test in part 1, or instead will know one of your answers to the questions – on the next page -- related to your performance on the test in part 1 and/or how many questions you answered correctly on the test in part 1.

D.3 Instructions for the *Private* version

The *Private* version run in wave 1 proceeds in the same manner as the *Self-Promotion* version, except for the instructions about part 2 and part 3. Participants are simply informed that they will receive 25 cents regardless of how they answer the self-evaluation questions. See Figure D.15 for these instructions and the corresponding understanding question. The *Private* versions run in waves 2, 3, and 5 are identical to the *Private* version in the first wave, except for a slight formatting change in the part 2 and part 3 questions to allow for room to introduce the additional information in the *Private (Social Norms)* version. See Figure D.16 for the corresponding screenshot of the part 3 self-evaluation questions (and note that this is identical to how they appear in part 2).

Figure D.15: The *Private* version: Part 2 Instructions and Understanding Question

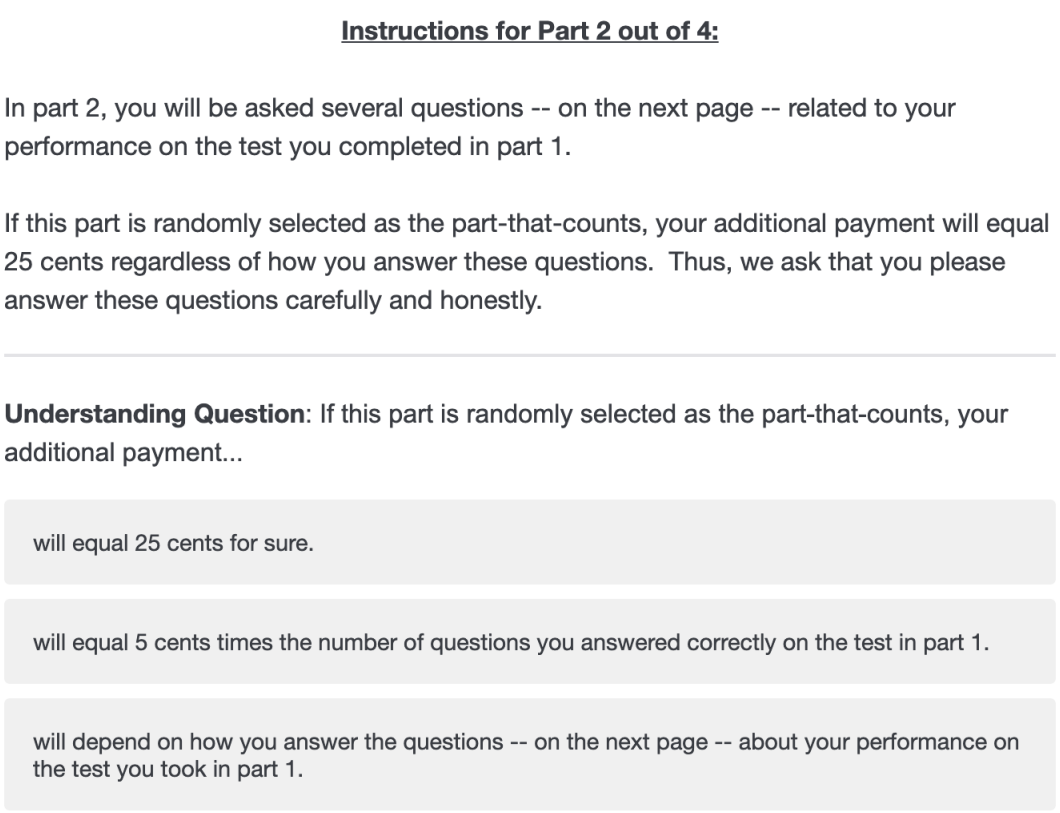
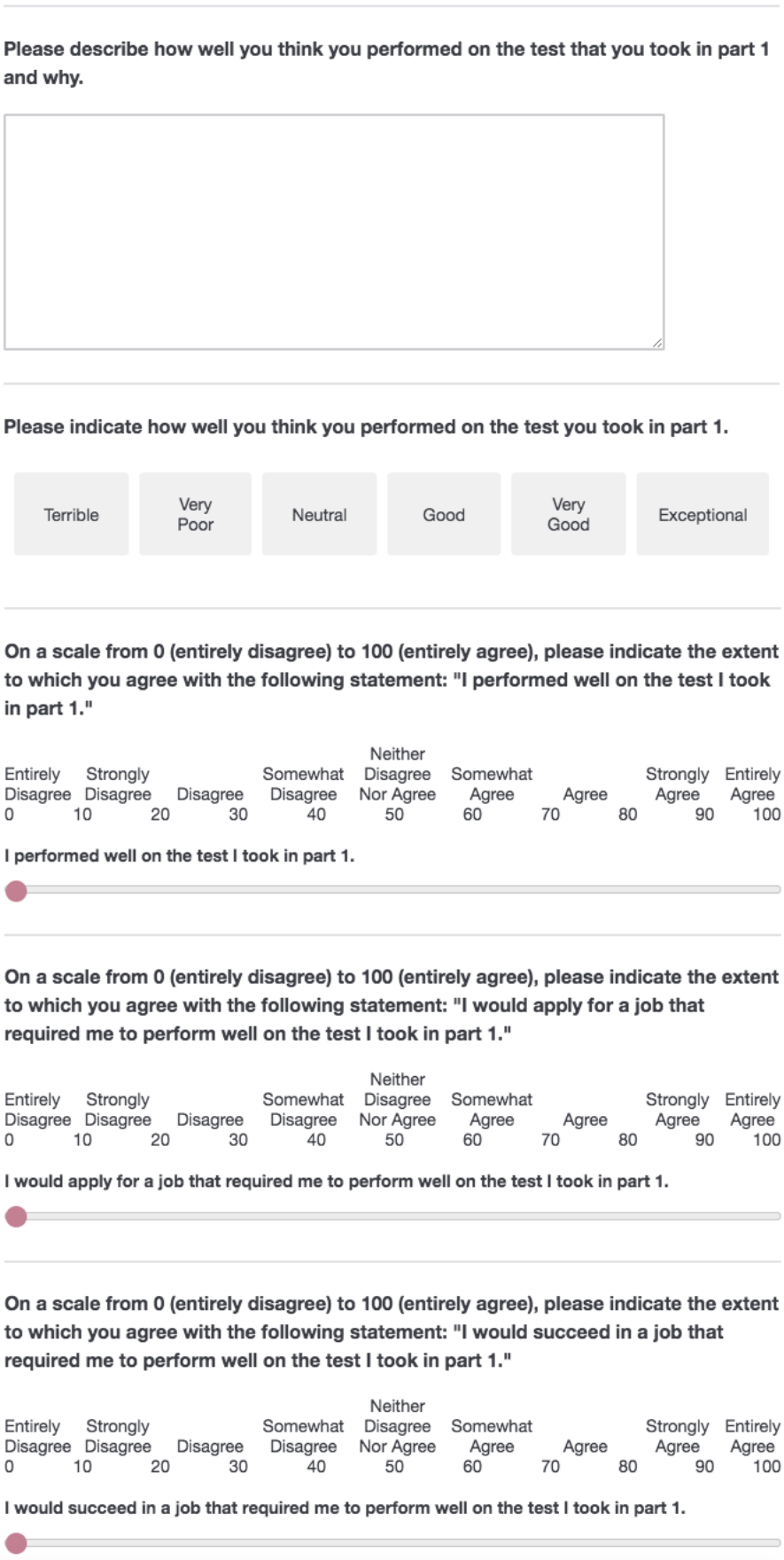


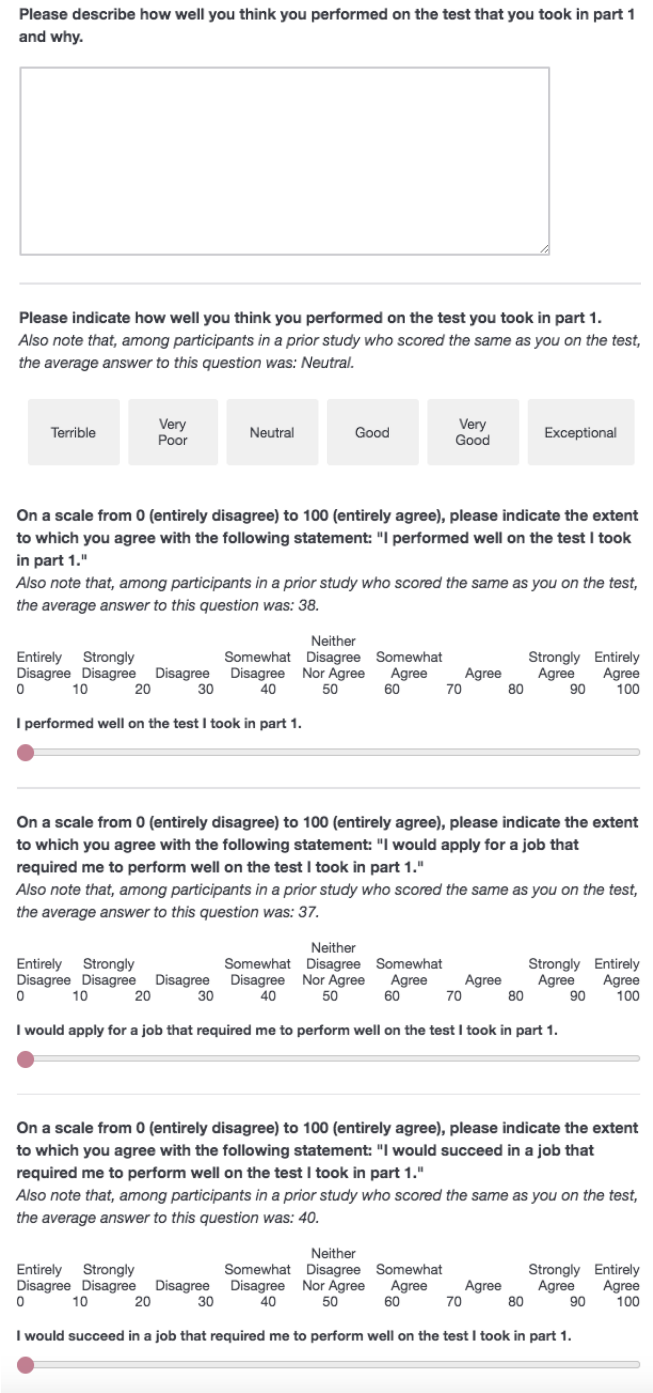
Figure D.16: The *Private* version: Part 3 Self-Evaluation Questions With a Slight Formatting Change



D.4 Instructions for the *Private (Social Norms)* version

The *Private (Social Norms)* version of the study proceeds in the same manner as the *Private* version of the study, except that, in part 3, additional information is provided on the average answer to each of the self-evaluation questions from prior participants with the same score as the participant. See Figure D.17 for the corresponding screenshot of the part 3 questions.

Figure D.17: The *Private (Social Norms)* version: Part 3 Self-Evaluation Questions for a Participant who Correctly Answered 10 out of 20 Questions



D.5 Instructions for the *Private (Immediately Informed)* version

The *Private (Immediately Informed)* version of the study proceeds in the same manner as the *Private* version of the study, except that participants learn their absolute and relative performance before answering any self-evaluation questions. That is, parts 3 and 4 in the *Private* version become parts 2 and 3 in this version so that the study proceeds as follows: participants complete the test in part 1, report their beliefs about their absolute performance on that test, are informed of their absolute and relative performance on that test, answer self-evaluation questions about that test in part 2, and answer the deservingness question in part 3.

D.6 Instructions for the *Private (Other-Evaluation)* version

The *Private (Other-Evaluation)* version proceeds in the same manner as the *Private (Immediately Informed)* version, except that participants are informed of the absolute and relative performance of another MTurk participant (see Figure D.18) and then are asked to provide informed other-evaluations about this other MTurk participant rather than themselves (see Figures D.19 and D.20).

Figure D.18: The *Private (Other-Evaluation)* version: Absolute and Relative Performance Information on Another MTurk Participant

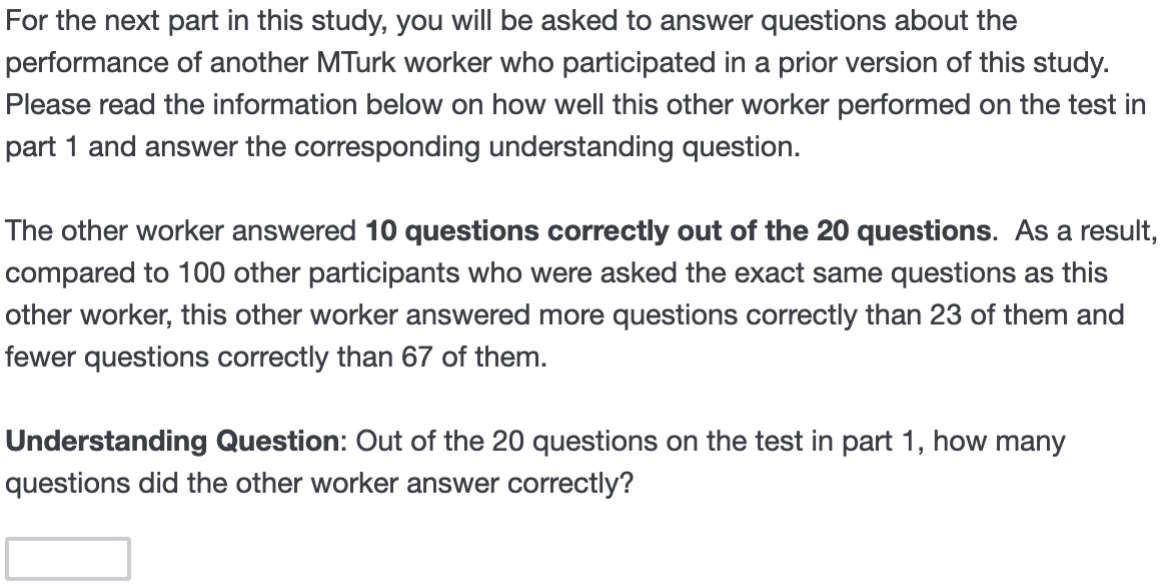


Figure D.19: The *Private (Other-Evaluation)* version: Part 2 Instructions and Understanding Questions

Instructions for Part 2 out of 3:

In part 2, you will be asked several questions -- on the next page -- related to the performance of the other worker, described on the previous page, on the test in part 1.

If this part is randomly selected as the part-that-counts, your additional payment will equal 25 cents regardless of how you answer these questions. Thus, we ask that you please answer these questions carefully and honestly.

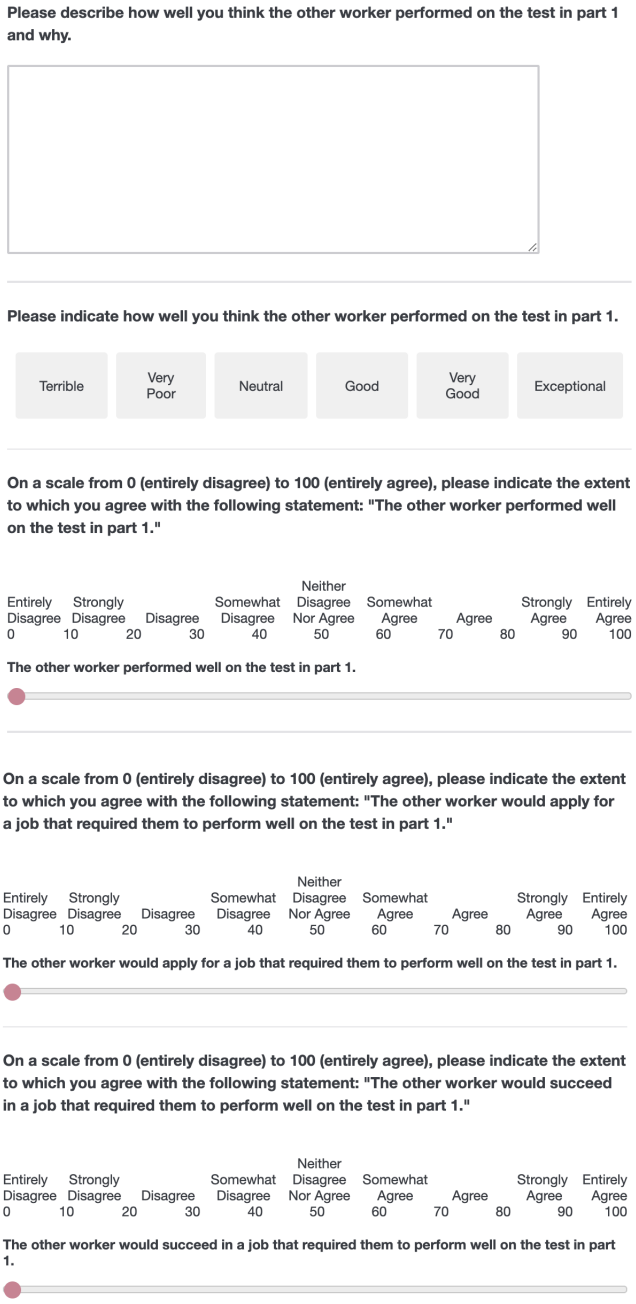
Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will depend on how you answer the questions -- on the next page -- about the performance of the other worker on the test in part 1.

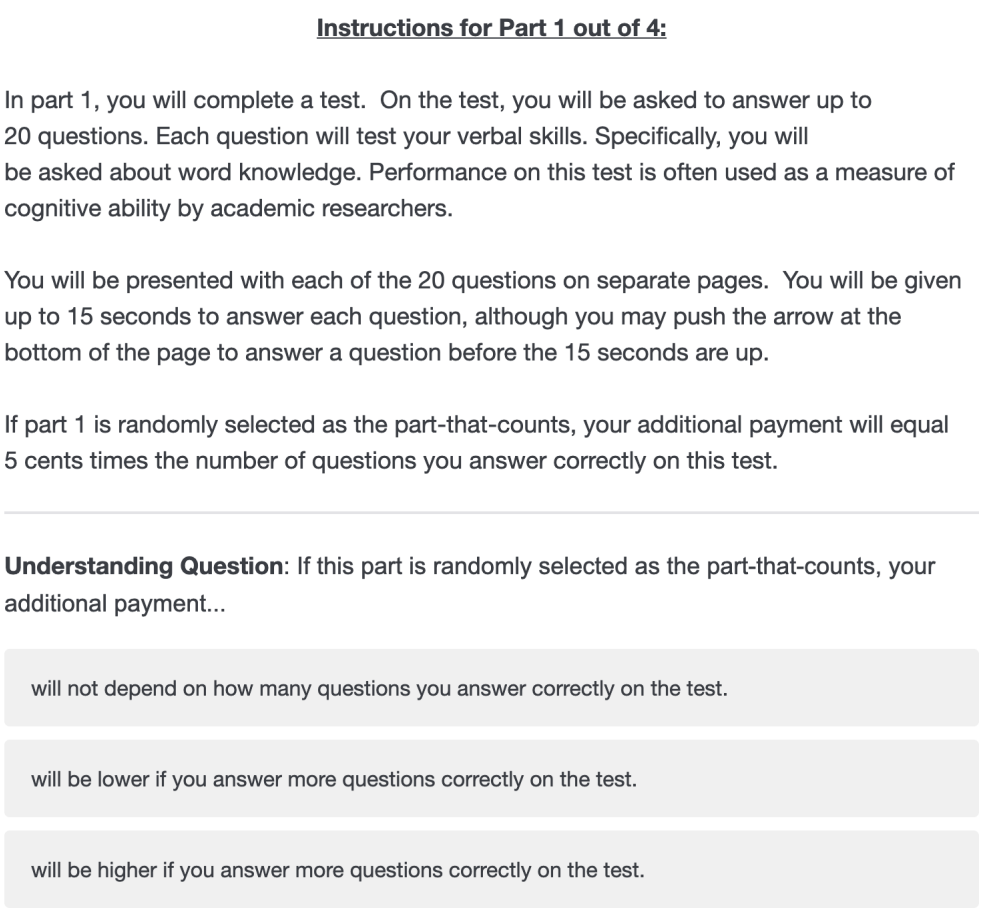
Figure D.20: The *Private (Other-Evaluation)* version: Part 2 Other-Evaluation Questions for Another Participant who Correctly Answered 10 out of 20 Questions



D.7 Instructions for the *Private (Verbal)* version

The *Private (Verbal)* version proceeds in the same manner as the *Private* version, except that the test that participants complete in part 1 asks them to answer 20 word knowledge questions rather than 20 math and science questions (see Figure D.21 for the instructions and Figure D.22 for an example question). In addition, there are two pages added to their follow-up survey that participants complete after they complete the other parts of the study.[16] As shown in Figure D.23, they learn (as a surprise) of the opportunity to earn additional bonus payment if they answer one of the eight prediction questions on the next two pages correctly. The order of the next two pages is randomly determined. On one of the pages, they are asked to answer four prediction questions about women (see Figure D.24). On the other page, they are asked to answer four prediction questions about men (see Figure D.25).

Figure D.21: Instructions for Part 1



[16] These same questions are also added to the *Private* version we ran in wave 5.

Figure D.22: Part 1: Example Verbal Question

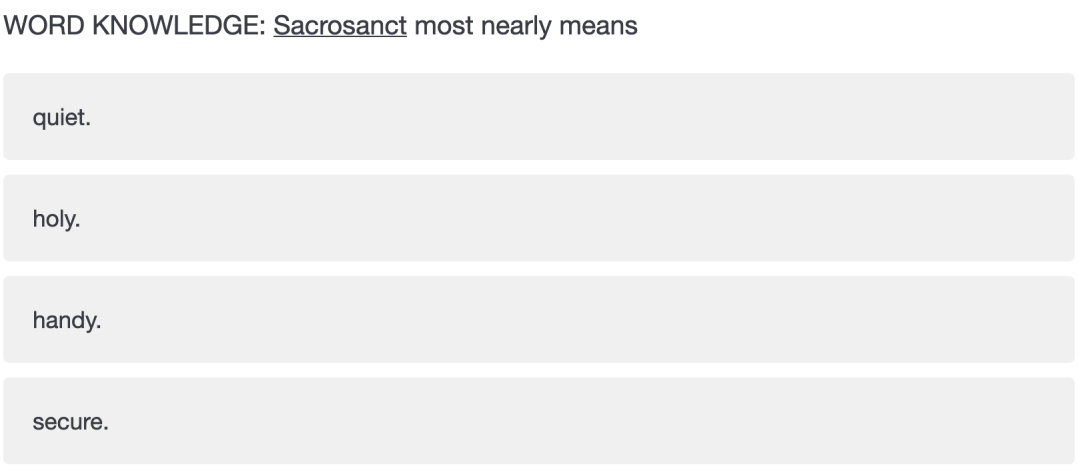


Figure D.23: Instructions for Predictions

On the remaining two pages of the follow-up survey, you will be asked 8 questions. One of these questions will be randomly selected as the question-that-counts. If your answer to the question-that-counts is correct, you will receive an additional bonus payment of $0.50. This bonus payment will be in addition to any bonus payment you earned from the part-that-counts. **Thus, please answer these questions carefully and honestly to maximize your chance of earning more bonus payment.**

In each question, you will be asked to guess the performance of other participants in a prior study. These other participants answered 20 questions on a test from from the Armed Services Vocational Aptitude Battery (ASVAB). Each question tested their aptitude in one of the following five categories: General Science, Arithmetic Reasoning, Math Knowledge, Mechanical Comprehension, and Assembling Objects.

After completing this test, these participants were asked questions about their performance on the test. They were told that their response to one of the questions would be shared with an "employer," who would be another participant who completed a different version of this study. They were also told that their employer would decide whether to hire them, and if so, how much to pay them. They were also told that if they were hired, their employer would earn more money if they answered more questions correctly on the test.

Figure D.24: Predictions about Women

Workers were asked to to indicate whether they thought their performance on the test was terrible, very poor, neutral, good, very good, or exceptional.

Among the set of **female workers** who indicated to their employers that their performance on the test was **neutral**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Workers were asked to indicate the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I performed well on the test I took."

Among a set of **female workers** whose average response to the question above was **48 out of 100, somewhat disagreeing with the statement that they performed well**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Workers were asked to indicate the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: I would apply for a job that required me to perform well on the test I took."

Among a set of **female workers** whose average response to the question above was **45 out of 100, somewhat disagreeing with the statement that they would apply for such a job**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Workers were asked to indicate the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: I would succeed in a job that required me to perform well on the test I took."

Among a set of **female workers** whose average response to the question above was **49 out of 100, somewhat disagreeing with the statement that they would succeed in such a job**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Figure D.25: Predictions about Men

Workers were asked to to indicate whether they thought their performance on the test was terrible, very poor, neutral, good, very good, or exceptional.

Among the set of **male workers** who indicated to their employers that their performance on the test was **good**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Workers were asked to indicate the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I performed well on the test I took."

Among a set of **male workers** whose average response to the question above was **59 out of 100, somewhat agreeing with the statement that they performed well**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Workers were asked to indicate the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: I would apply for a job that required me to perform well on the test I took."

Among a set of **male workers** whose average response to the question above was **60 out of 100, somewhat agreeing with the statement that they would apply for such a job**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

Workers were asked to indicate the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: I would succeed in a job that required me to perform well on the test I took."

Among a set of **male workers** whose average response to the question above was **65 out of 100, agreeing with the statement that they would succeed in such a job**, what is the average number of questions they got right on the test? (Please round to the nearest integer.)

D.8 Instructions for *Private* version run among youth

Prior to participating in the study, participants must correctly answer a captcha and consent to participate. At the end of the study, participants must complete a short follow-up survey to gather demographic information. Participants are recruited via the Character Lab Research Network and complete this study as part of the curriculum at school. There are no payments associated with this study.

The study begins by informing each participant about the test that they will take. The instructions for the test are displayed in Figure D.26 and an example of a question on the test is displayed as Figure D.27 (note that the timer in that screenshot indicates the participant has 24 seconds left to answer the question although the timer starts at 30 seconds).

Figure D.26: Instructions for the test

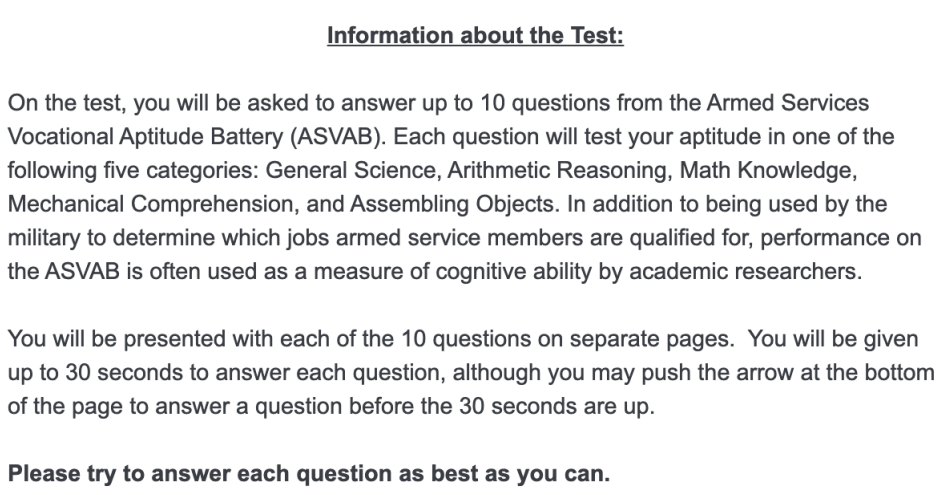
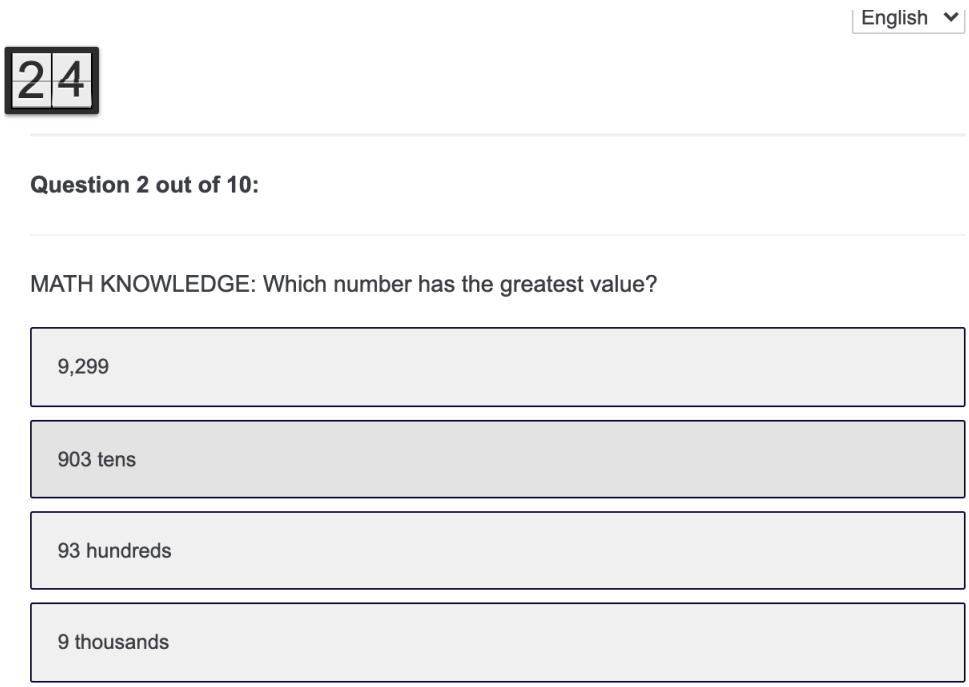
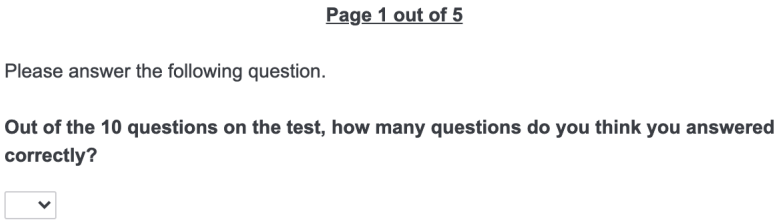


Figure D.27: Example question on the test



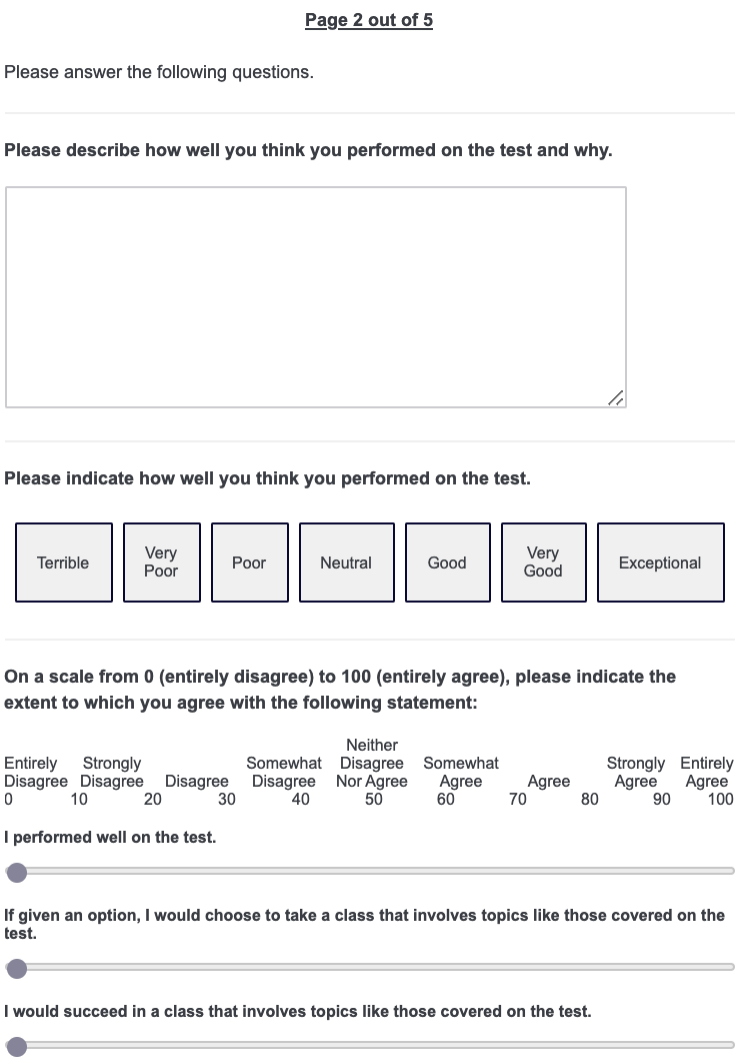
After completing the test, participants are asked to complete five additional pages of the study. On the first page, they are asked about their absolute performance belief, as shown in Figure D.28.

Figure D.28: Absolute Performance Belief Question



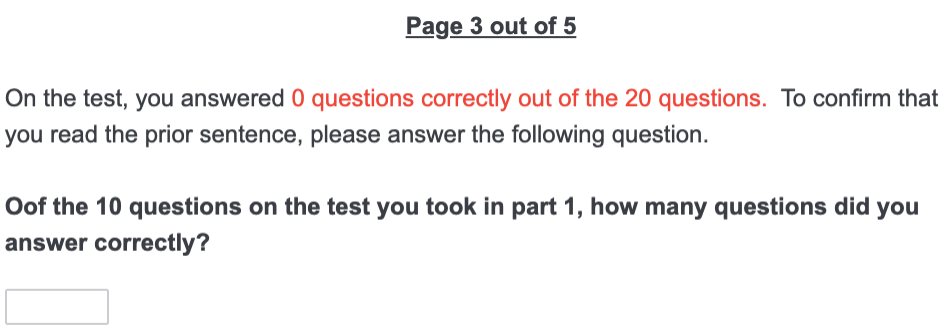
On the second page, they are asked the self-evaluation questions (see Figure D.29).

Figure D.29: Self-Evaluation Questions



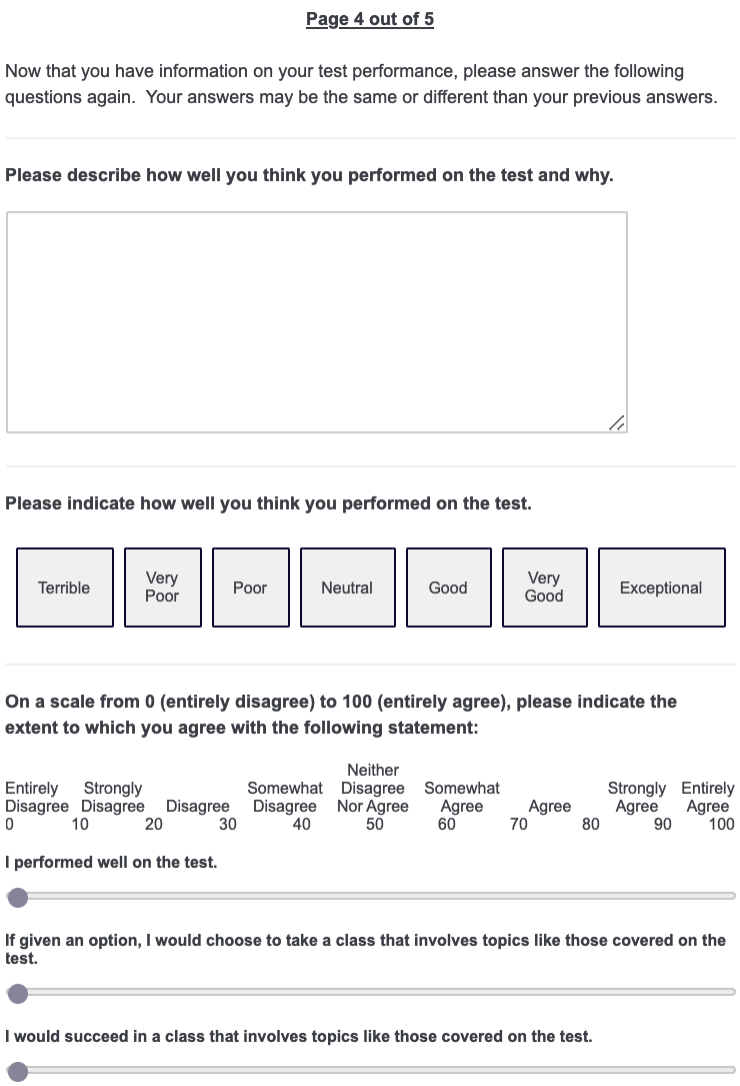
On the third page, participants are provided with perfect information on their absolute performance and are required to correctly report back their absolute performance as shown in Figure D.30.

Figure D.30: Absolute Performance Information



On the fourth page, they are asked the self-evaluation questions again (see Figure D.31). On the fifth page, they are asked for demographic information.

Figure D.31: Informed Self-Evaluation Questions



D.9 Instructions for *Employer* version

Prior to participating in the study, participants must correctly answer a captcha and consent to participate in the study. At the end of the study, participants must complete a short follow-up survey to gather demographic information.

The study begins by informing each participant of the $1.50 study completion fee and of the opportunity to earn additional payment. Figure D.32 shows how this payment information is explained. Figure D.33 shows the understanding questions that the participant must answer correctly to proceed.

Figure D.32: Payment Information

Overview:
This study will consist of 21 decisions and a short follow-up survey. For completing this study, you are guaranteed to receive $1.50 within 24 hours. In addition, any additional payment you earn will be distributed to you as a bonus payment.

The Workers:
In a prior study, MTurk workers completed a test. On the test, they were asked to answer up to 20 questions from the Armed Services Vocational Aptitude Battery (ASVAB). Each question tested their aptitude in one of the following five categories: General Science, Arithmetic Reasoning, Math Knowledge, Mechanical Comprehension, and Assembling Objects. In addition to being used by the military to determine which jobs armed service members are qualified for, performance on the ASVAB is often used as a measure of cognitive ability by academic researchers.

Your Decisions:
For each of the 21 decisions, you will be matched with one worker from the piror study. You then must decide whether to hire that worker, and if so, how much to pay that worker.

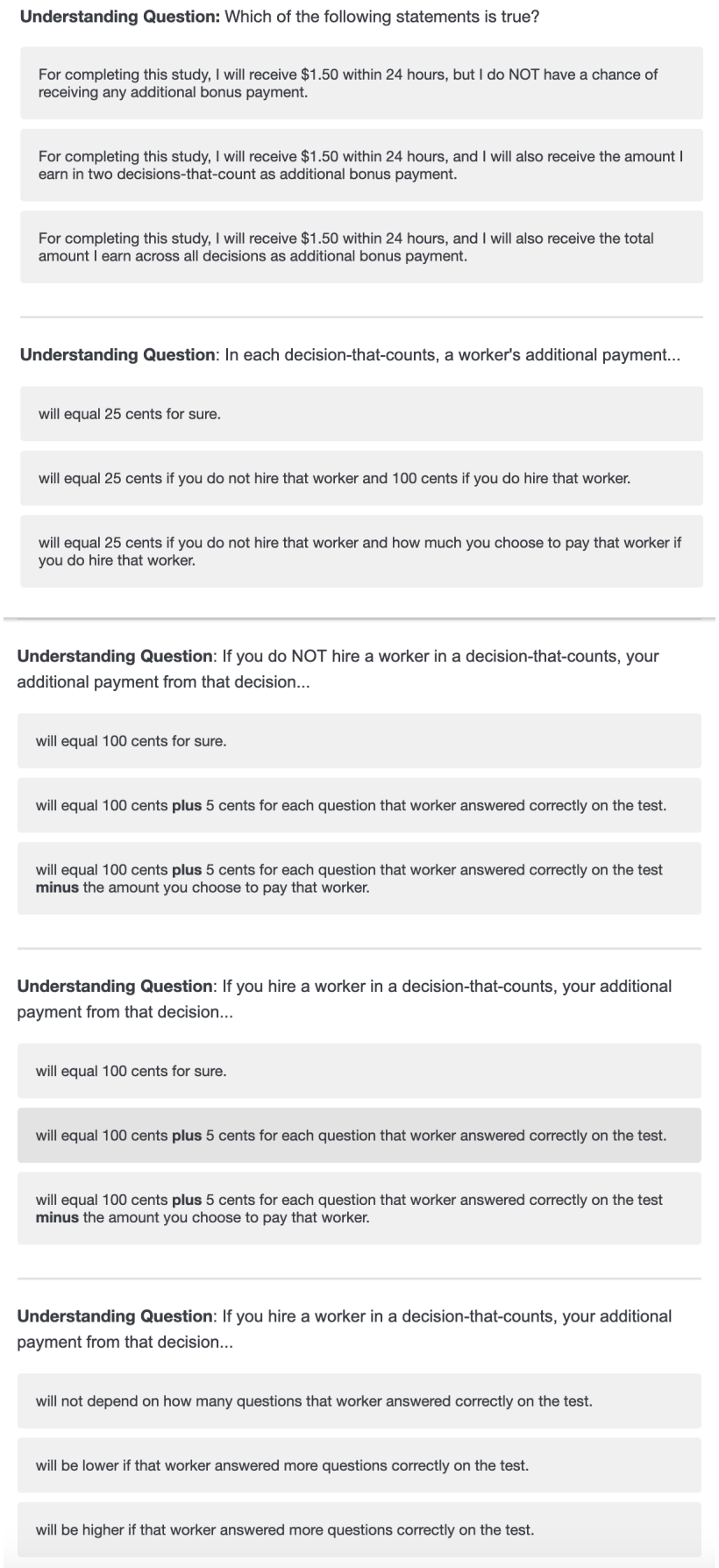
After you make all of your 21 decisions, two decisions will be selected as a decision-that-counts.

In each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. Your can choose to pay that worker any amount between 25 cents and 100 cents.

Figure D.33: Understanding Questions of Payment Information



The 21 decisions that employers face involve four blocks. Three blocks relate to the three evaluation questions that involve the 0-to-100 scale (i.e., the performance question, the willingness-to-apply question and the success question), and each of these blocks involves five decisions that correspond to five randomly selected evaluations (i.e., numbers from 0 to 100). Another block relates to the evaluation question involving a six point Likert-scale (i.e., the performance-bucket question), and this block involves six decisions that correspond to each of the six possible evaluations in that question. The order of these four blocks is randomized on the participant-level.

The instructions for, and examples of, decisions relating to the *performance* evaluations are displayed in Figures D.34 and D.35, respectively.

Figure D.34: Instructions for *Performance Evaluation* Decisions

Instructions for Decisions 1 - 5

In each decision below, you will learn how the worker in that decision answered a question in which they indicated the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I performed well on the test I took."

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. Your can choose to pay that worker any amount between 25 cents and 100 cents.

Figure D.35: *Performance Evaluation* Decisions

Decision 1 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 6, indicating strong disagreement with the following statement: "I performed well on the test I took." What would you like to do?

Decision 2 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 25, indicating disagreement with the following statement: "I performed well on the test I took." What would you like to do?

Decision 3 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 56, indicating neither much disagreement nor agreement with the following statement: "I performed well on the test I took." What would you like to do?

Decision 4 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 61, indicating agreement with the following statement: "I performed well on the test I took." What would you like to do?

Decision 5 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 93, indicating strong agreement with the following statement: "I performed well on the test I took." What would you like to do?

The instructions for, and examples of, decisions relating to the *performance-bucket* evaluations are displayed in Figures D.36 and D.37, respectively.

Figure D.36: Instructions for *Performance-Bucket Evaluation* Decisions

Instructions for Decisions 6 - 11

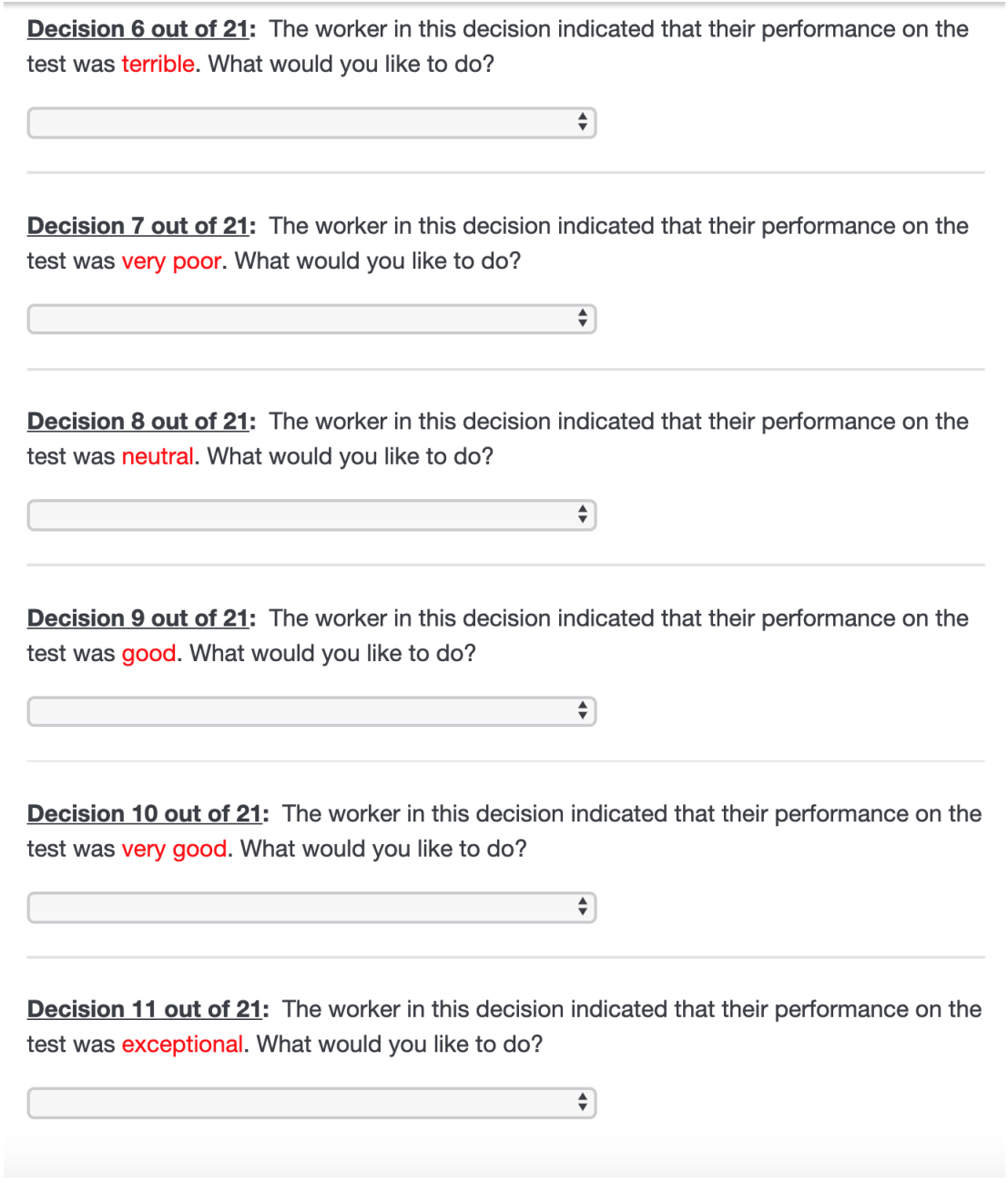
In each decision below, you will learn how the worker in that decision answered a question in which they indicated whether they thought their performance on the test was terrible, very poor, neutral, good, very good, or exceptional.

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

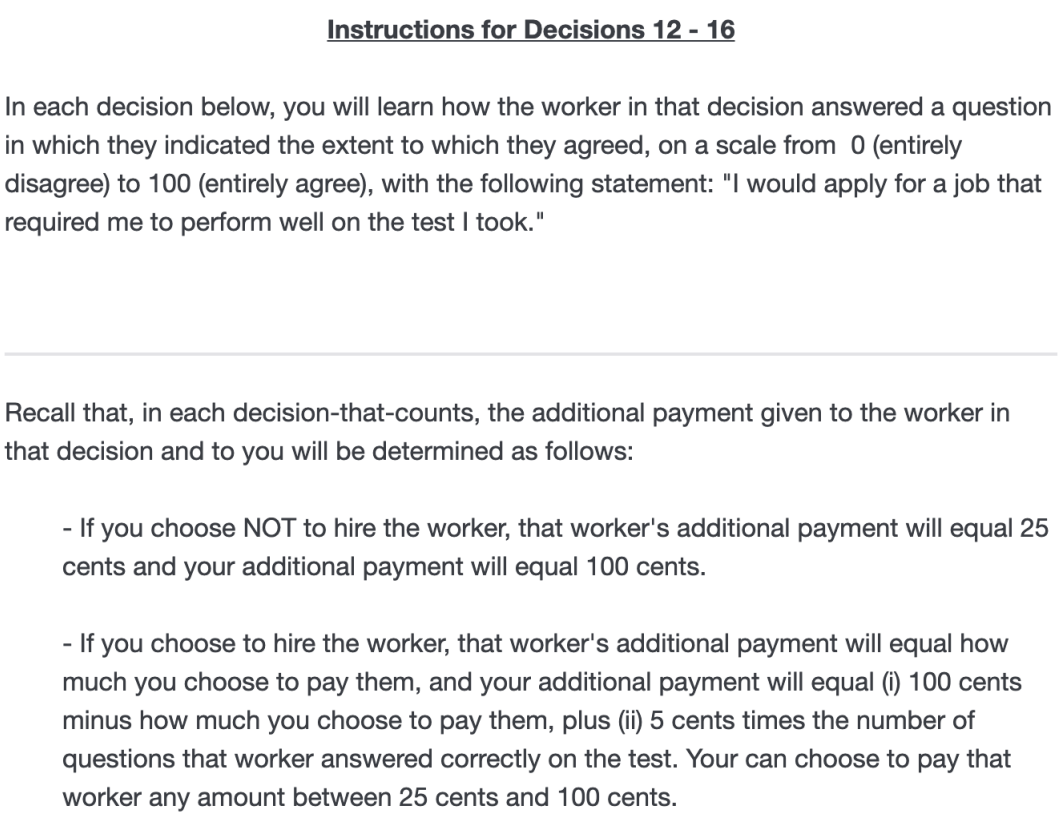
- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. Your can choose to pay that worker any amount between 25 cents and 100 cents.

Figure D.37: *Performance-Bucket Evaluation* Decisions



The instructions for, and examples of, decisions relating to the *willingness-to-apply* evaluations are displayed in Figures D.38 and D.39, respectively.

Figure D.38: Instructions for *Willingness To Apply Evaluation* Decisions



Instructions for Decisions 12 - 16

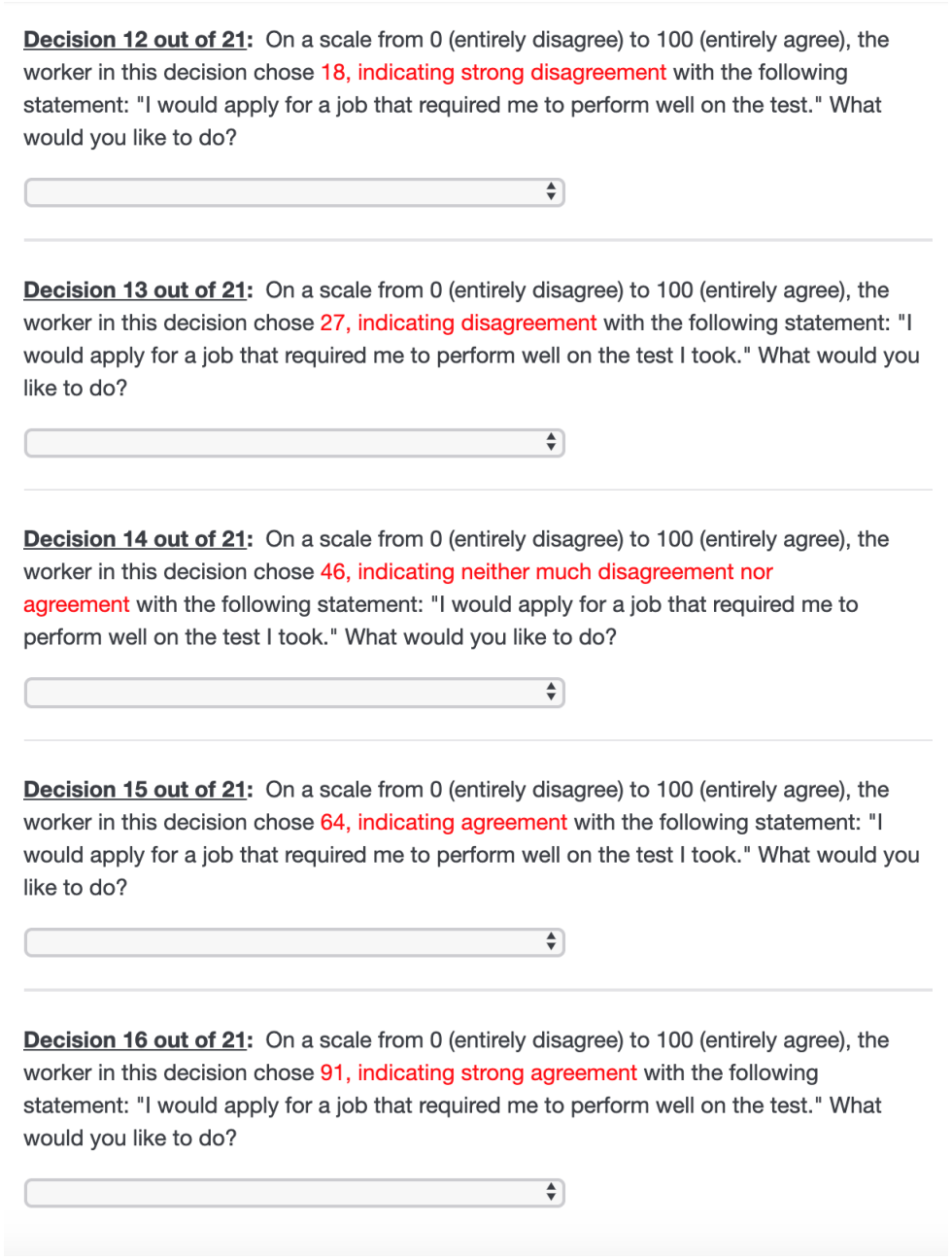
In each decision below, you will learn how the worker in that decision answered a question in which they indicated the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I would apply for a job that required me to perform well on the test I took."

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

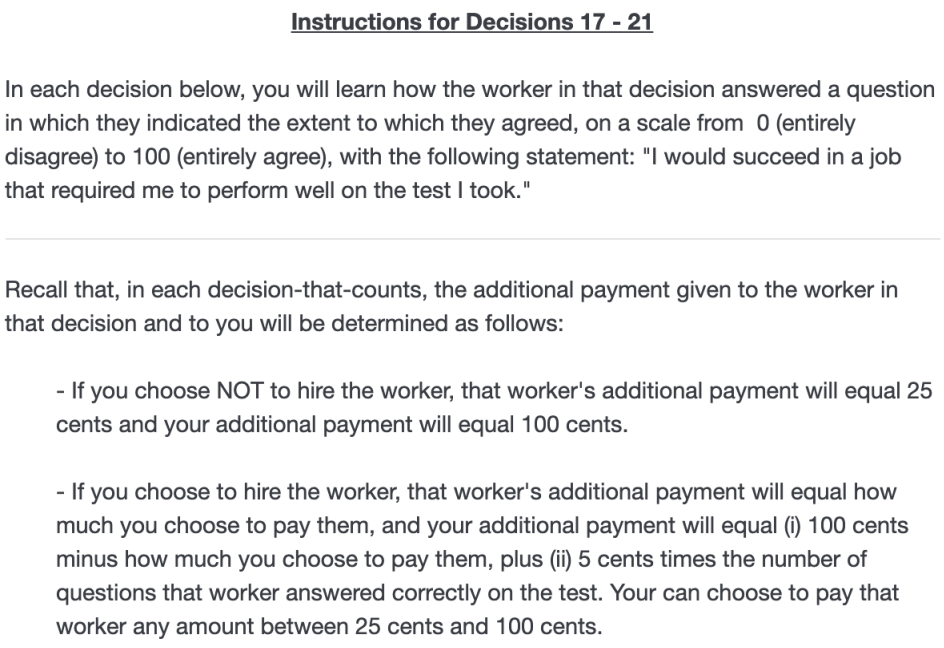
- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. Your can choose to pay that worker any amount between 25 cents and 100 cents.

Figure D.39: *Willingness To Apply Evaluation* Decisions



The instructions for, and examples of, decisions relating to the *success* evaluations are displayed in Figures D.40 and D.41, respectively.

Figure D.40: Instructions for *Success Evaluation* Decisions



Instructions for Decisions 17 - 21

In each decision below, you will learn how the worker in that decision answered a question in which they indicated the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I would succeed in a job that required me to perform well on the test I took."

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. Your can choose to pay that worker any amount between 25 cents and 100 cents.

Figure D.41: *Success Evaluation* Decisions

Decision 17 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 6, indicating strong disagreement with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 18 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 33, indicating disagreement with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 19 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 44, indicating neither much disagreement nor agreement with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 20 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 76, indicating agreement with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 21 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose 96, indicating strong agreement with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?