Online Appendix to "Blind Disclosure"

Aaron Kolb Marilyn Pease Daniel W. Sacks Joshua Quick*

December 11, 2021

1 Grades, final scores, and interim scores

1.1 Concordance between Canvas scores and final grades

We investigate the concordance between Canvas scores and final grades in Online Appendix Figure 1. Canvas and final grades are closely but not perfectly aligned. Overall they agree about three-quarters of the time. However there is considerable heterogeneity across courses. In about 1,400 of the 3,200 courses in our full sample, there is perfect agreement, in the sense that Canvas and registrar grades match exactly for each student in the course. However it is clear that one reason Canvas and registrar grades can disagree is that Canvas scores are rounded up; for example more than half of scores in $[89.9, 90)$ are awarded an A-. Our analysis sample therefore focuses on the 2,466 courses in which the Canvas and Registrar grades never diverge by more than one notch, e.g., B to B+.

1.2 May 2 scores, final scores, and signals

In our main analysis, we impute students' signal from their predicted final score, given their may 2 score, with a class-specific prediction function. Here we show why it is important to allow for course-specific heterogeneity.

We begin by showing that the raw May 2 score is a poor proxy for students' signals, because final scores tend to be higher on average than May 2 scores. We illustrate this drift in Online Appendix Figure 3, which plots the distribution of course-specific prediction errors, defined as the difference between course-level average scores on May 2 and at the end of the semester. There is a mass at zero, but also a long right tail. The mass at zero indicates that for many courses the May 2 scores are accurate on average, but the right tail indicates that for many other courses, May 2 scores are too low.

To establish the validity of our measure of students' signals, we show in Online Appendix Table 5 the R^2 from a regression of students' disclosure decision on their signal, as well as other variables for comparison. Our main signal measure strongly predicts disclosure, much more so than student GPA or the raw May 2 score. A downside of our signal measure is that it may contain too much

*Kolb: Indiana University, kolba@indiana.edu; Pease: Indiana University, marpease@iu.edu; Sacks: Indiana University, dansacks@indiana.edu; Quick: Indiana University, jdquick@iu.edu.

information; it might be unrealistic to assume that students know the exact βs for their courses. On the other hand it also misses some relevant information, as students have private information that may be relevant for forecasting their final performance, for example their overall aptitude. We therefore consider robustness to an alternative proxy which regresses final score on May 2 score as well as student incoming GPA (also with a class-specific coefficient), to capture some student-specific information. Alternatively, we can use students' actual final canvas grades as proxies for their signals. This proxy clearly contains too much information, but if students have information unobservable to us, then the final grade may be a better proxy for their signal than is their May 2 score. In practice we find that our results are not sensitive to the specific signal proxy that we use.

1.3 Correlates of course-level uncertainty

We measure course level uncertainty as the root mean squared prediction error of final scores given May 2 scores. Courses will have greater uncertainty to the extent that May 2 grades are less predictive of final grades, which reflects the notion of uncertainty in the model. Onnline Appendix Figure 4 shows the distribution of RMSE across courses. There is wide variation and two clear modes. In some of our analyses we will divide up the sample into above- and below-median RMSE courses.

We investigate the correlates of course-level uncertainty in Online Appendix Tables 2 and 3. Uncertainty is greater for courses with lower incoming student GPA, and it varies systematically across schools, with higher uncertainty in the College of Arts and Sciences, and lower uncertainty in the professional schools. However overall we find few strong predictors of $RMSE_c$; 85 percent of the variation in $RMSE_c$ remains unexplained.

To better understand the difference between high- and low- uncertainty courses, in Online Appendix Table 4 we report the average variance of final scores and interim scores (averaging at the course level), separately for above- and below-median $RMSE$ courses, along with average values of β_{0c} and β_{1c}. High uncertainty classes have higher variances of final scores, and lower values of β_{1c}, unsurprisingly. We also see, however, that high-uncertainty classes have higher values of β_{0c}, and on net they have larger "grade bumps," meaning larger changes from May 2 scores to final scores. Overall, grades are lower in high uncertainty classes, as we show in Online Appendix Figure 5. To avoid picking up the potentially confounding influence of either the lower grade level or the larger grade bumps, in our main regressions we will control for both expected grade and the average grade bump.

Table 1: Sample restrictions and sample sizes

Restriction	# Observations (1)	# Students (2)	# Courses (3)
Initial sample	107,552	29,328	3,211
Non-missing GPA	107,530	29,322	3,211
Non-missing Canvas scores (full sample)	106,877	29,315	3,191
Canvas-registrar grades agree	75,396	28,260	2,466
At least 30 students (Analysis sample)	46,515	23,717	778

Notes: Table reports the number of observations (student-courses), number of students, and number of students, as we impose increasingly stringent sample restrictions. The initial sample consists of all undergraduates enrolled in courses at Indiana University, Bloomington, in Spring 2020, with standard final grades (A-F or P), in full-semester or second-half courses. The restriction that Canvas and registrar grades agree means that, for each class, all Canvas grades exactly match registrar grades, or disagree by one notch. We use the "analysis sample" for all analyses involving students' signals.

Table 2: Determinants of course-level RMSE

	(1)	(2)	(3)	(4)	(5)
200-level	0.02			0.16	0.30
	(0.23)			(0.23)	(0.23)
300-level	-0.59			0.13	0.31
	(0.22)			(0.23)	(0.23)
400-level	-0.84			0.19	0.57
	(0.26)			(0.29)	(0.28)
Average incoming GPA		-3.03		-2.94	-2.67
		(0.39)		(0.46)	(0.47)
Square root class size			0.15	0.10	0.09
			(0.04)	(0.04)	(0.04)
R^2	0.02	0.08	0.02	0.09	0.15
# Classes	778	778	778	778	778
School dummies?	No	No	No	No	Yes

Notes: Table reports coefficients from a regression of course-level RMSE on the indicated controls. 100-level courses are the omitted category in columns (1), (4), and (5). The coefficients on the school dummies are in Column (2) of Online Appendix Table 3. The sample is the analysis sample, described in the notes to Table 1 in the main paper. Robust standard errors in parentheses.

Table 3: Course-level RMSE, by school

	(1)	(2)
Nursing	-3.18	-2.28
	(0.37)	(0.37)
Art	-3.29	-3.06
	(0.27)	(0.34)
Education	-1.67	-1.59
	(1.18)	(1.40)
Medicine	-1.45	-0.73
	(0.78)	(0.58)
Business	-1.36	-1.09
	(0.20)	(0.21)
Public Health	-1.32	-1.21
	(0.25)	(0.25)
Media	-1.41	-1.40
	(0.34)	(0.34)
Policy	-1.01	-1.13
	(0.24)	(0.24)
Music	-0.56	-0.42
	(0.55)	(0.54)
Informatics	-0.98	-0.95
	(0.52)	(0.53)
International	0.30	0.22
	(0.52)	(0.45)
Joint p-value for schools	< 0.001	< 0.001
R^2	0.10	0.15
# Classes	778	778
Other controls dummies?	No	Yes

Notes: Table reports coefficients from a regression of course-level RMSE on a set of school dummies, omitting the College of Arts and Sciences. The coefficients on the additional controls are reported in column (5) of Online Appendix Table 2. The sample is the analysis sample, described in the notes to Table 1 in the main paper. Robust standard errors in parentheses.

Table 4: Analysis of variance and mapping from interim to final grades, by course uncertainty

Course RMSE	Low	High
Variance, final score	58.4	117.6
Variance, interim score	72.5	155.3
MSE	4.8	32.5
β_{0c}	18.8	27.8
β_{1c}	0.84	0.78
Grade bump	4.4	8.2

Notes: Table reports the average value of the indicated course-level characteristic, averaging over courses with below- or above-median RMSE. MSE, $\hat{\beta}_{0c}$, and $\hat{\beta}_{1c}$ are the mean squared error and estimated coefficients from equation (8). "Grade bump" is the average difference in scores between May 2 and the end of the semester. The sample is the analysis sample, described in the notes to Table 1 in the main paper.

Table 5: Predictive power of signals for disclosure decision

Predictor:	May 2 Signal (1)	GPA (2)	Raw May 2 (3)	Homogeneous May 2 (4)	Final Grade (5)
R^2	0.265	0.029	0.096	0.097	0.386
% Correct	0.878	0.862	0.862	0.862	0.897
Model parameters	11	11	11	11	11
Sample size	46,515	46,515	46,515	46,515	46,515

Notes: For each column, we estimate a linear probability model of disclosure on dummy variables for each value of the indicated variable (except column (3), where we dummies for binned values for GPA grouped to grade bins (e.g. 3.3). We report the R^2 and the fraction correctly classified. We say we have correctly classified an observation if $disclose_{ic} = 1$ and the predicted probability is above 0.5, or $disclose_{ic} = 0$ and the predicted probability is less than 0.5. The May 2 signal is our main signal measure. "Raw May 2" does not adjust for drift between May 2 and final scores. Homogeneous May 2 adjusts for drifts but without class-specific coefficients. Final grade uses the Canvas final grade as the signal. The sample is the analysis sample, described in the notes to Table 1 in the main paper.

2 Robustness and heterogeneity of empirical results

2.1 Robustness

We investigate the robustness of our findings to alternative definitions of uncertainty, alternative ways of measuring students' signals, and alternative samples.

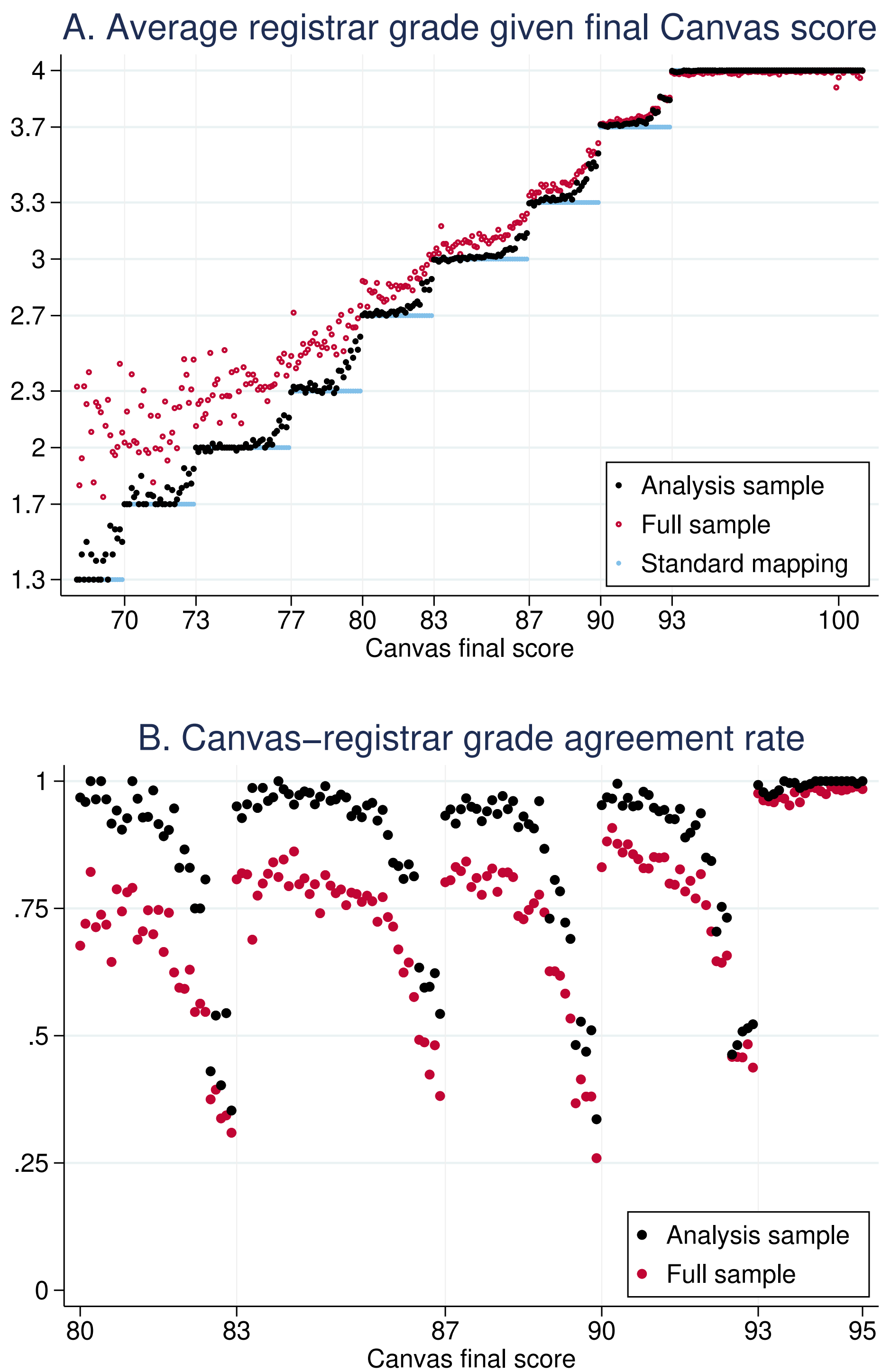
We show robustness to alternative uncertainty measures in Online Appendix Table 6. Column (1) reports our baseline specification, which measures uncertainty with the course-specific RMSE of May 2 grade as a predictor for final grade. In columns (2) and (3) we use an alternative measure which is based on the proportion of grades which change. Specifically in column (2) our measure of uncertainty is the proportion of students whose final grade differs from their May 2 grade by at least one grade notch, and in column (3) the measure is the proportion that differ by at least two notches. These measures address the concern that our underlying uncertainty measure reflects uncertainty in the continuous grade, but letter grade uncertainty may be especially important for students. In column (4) we measure uncertainty more coarsely, with an indicator for "above-median RMSE" (defined at the course level), as in Figure 7. These alternative measures imply that eliminating uncertainty would increase disclosure 2-4 percentage points (for a single student, holding fixed beliefs), similar to our baseline estimate. In column (5) we measure uncertainty by the course-specific correlation between interim and final scores (scaled by -1 so that higher values indicate more uncertainty). We continue to find a significant and negative effect of uncertainty on disclosure.

In columns (6)-(8) we measure uncertainty as the RMSE from a course-specific regression of final score on interim score (as in our baseline approach), but we allow for a nonlinear relationship between final and interim scores. Specifically, we estimate piece-wise linear models, with two, three, or four line segments. The break points are placed at the median for the two-line model, at each tercile for the three-line model, and each quartile for the four-line model. We find this alternative measure of uncertainty makes little difference to our estimates. This is important because moral hazard can create a piece-wise linear relationship between interim and final grades, and failing to account for it could bias our estimates.

We show robustness to alternative signal measures in Online Appendix Table 7. In our baseline specification in column (1), the signal measure is a set of dummy variables for the predicted final letter grade, given May 2 grade, with course-specific coefficients. Discretizing the predicted grade and turning it into a set of dummy variables may exacerbate measurement error problems (as the signal itself is measured with error). In columns (2) and (3) we drop the dummy variables for predicted grade and instead include linear or cubic terms in predicted final grade. A separate issue is that our signal misses some information available to students. We enrich the information set in column (2). To do so, we modify the signal measure so that each students' prediction depends on their incoming GPA as well as their May 2 grade.[1] When we work with this alternative signal, we also use as an RMSE measure the RMSE from the regression of final grade on May 2 grade and student GPA. In column

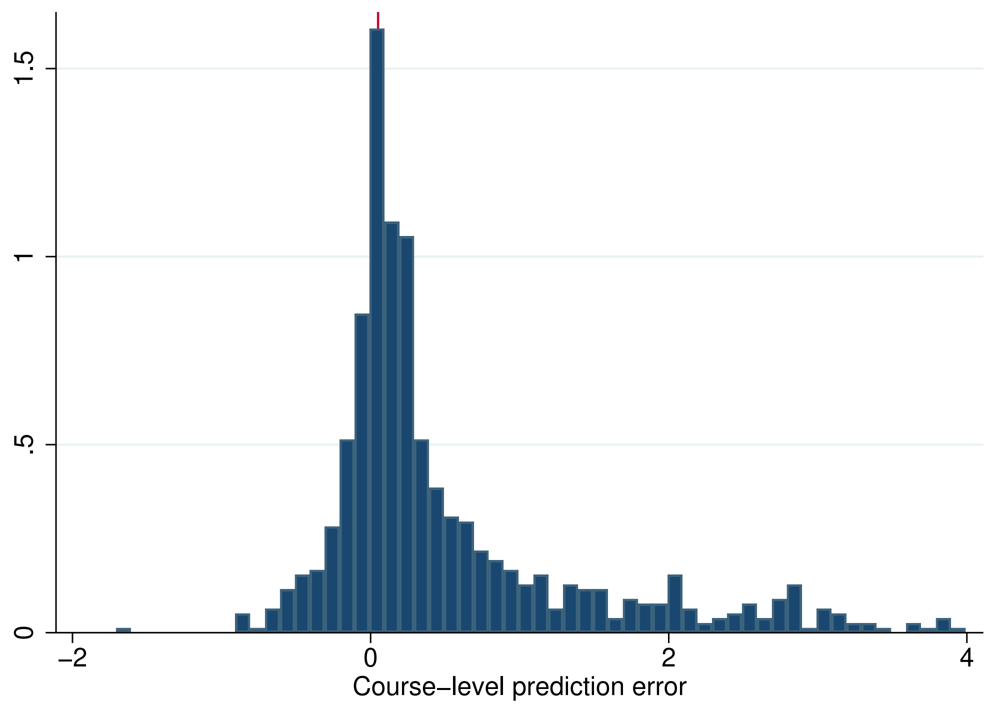
[1]That is, we regress final canvas grade on May 2 grade and student-level incoming GPA, with course-specific coefficients.

Figure 1: Alignment between Canvas and registrar grades



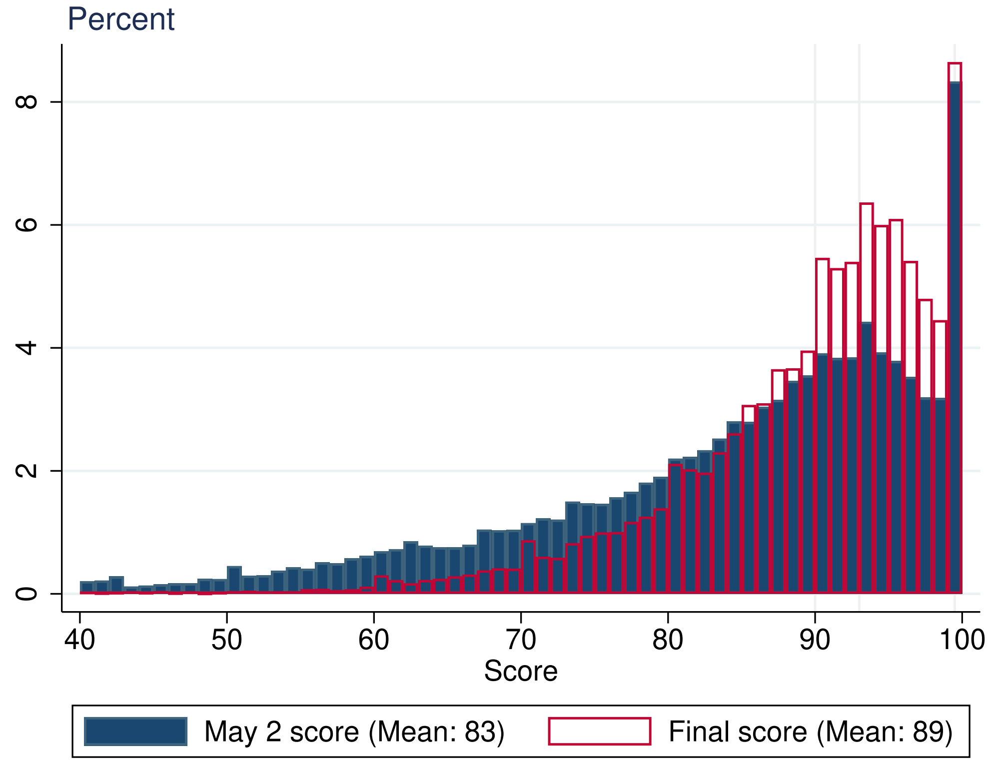
Notes: Top panel shows the average registrar grade (on a 0-4) scale for the full sample and the analysis sample, given Canvas final score, as well as the "standard mapping." The bottom panel shows the the fraction of registrar grades that equal the standard mapping from the Canvas final score, at each 0.1 point interval. The analysis and full samples are defined in the notes to Table 1 in the main paper.

Figure 2: Distribution of prediction errors across courses



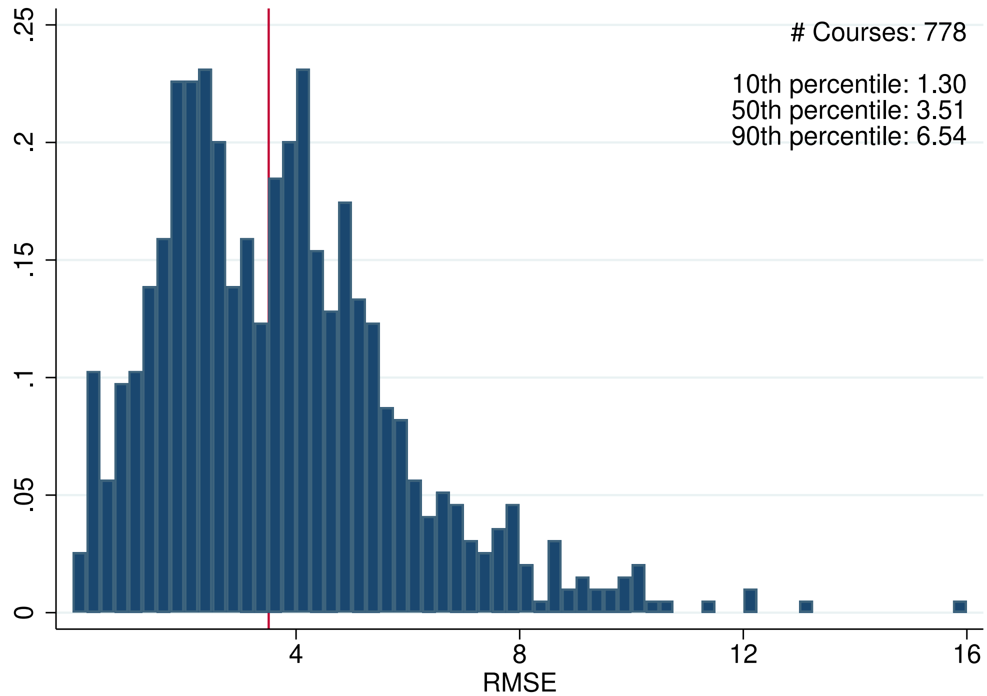
Notes: Figure plots the distribution of average prediction errors across courses. The prediction error is the final grade less the May 2 grade. The vertical line is at 0. If the signal were unbiased for all classes, we would expect a symmetric distribution around 0. The sample is the analysis sample as defined in the notes to Table 1 in the main paper.

Figure 3: Distribution of May 2 and final Canvas scores



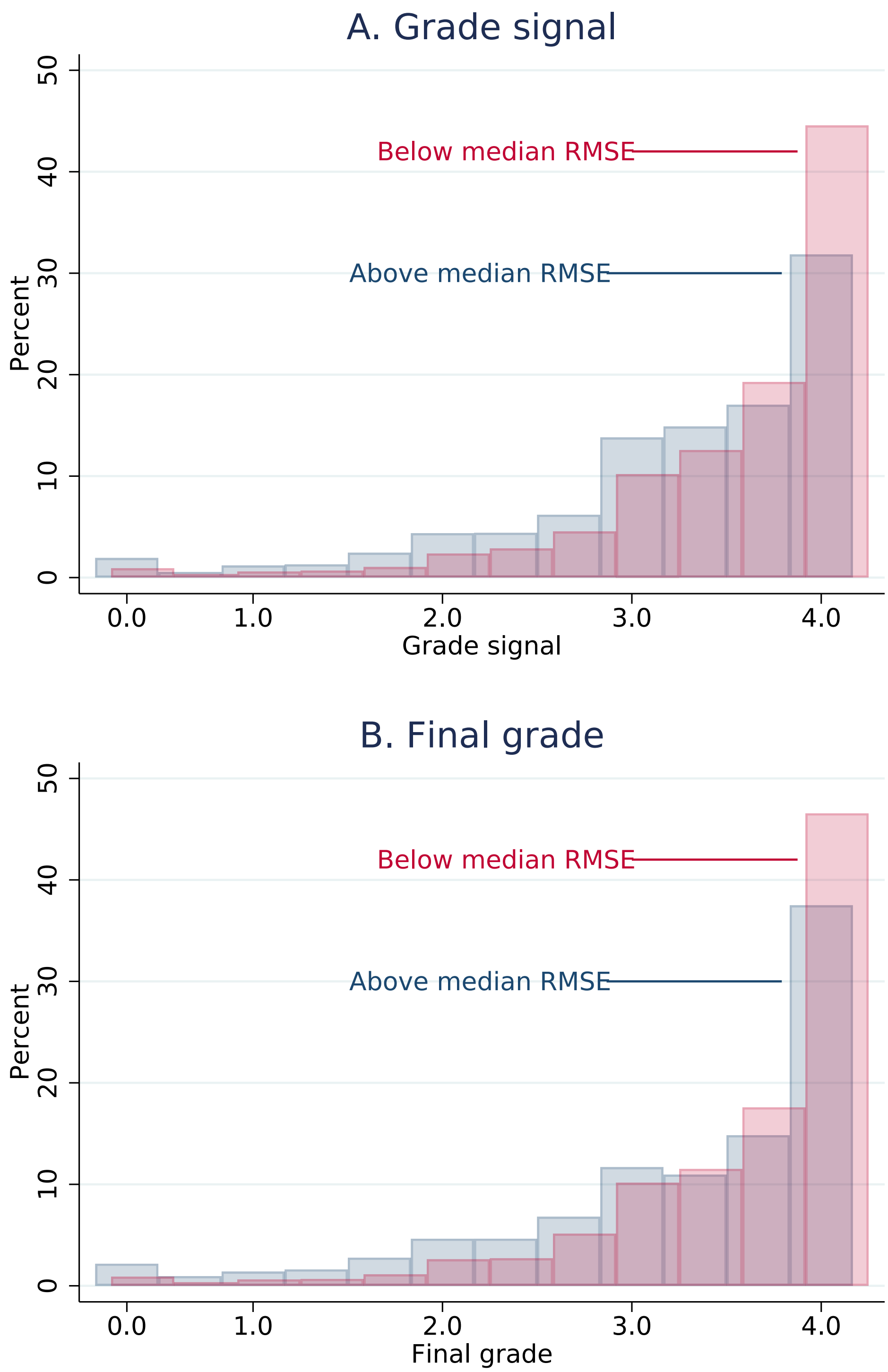
Notes: Figure plots the distribution of final and May 2 Canvas scores in the analysis sample. The vertical lines are at 93 and 93 percent. The sample is the analysis sample as defined in the notes to Table 1 in the main paper.

Figure 4: Distribution of course-level RMSE



Notes: Figure plots the distribution across courses of RMSE, the root mean squared prediction error from a regression of final Canvas grade on May 2 course grade, with course-specific coefficients. The sample is the analysis sample as defined in the notes to Table 1 in the main paper.

Figure 5: Grade distributions by grade uncertainty



Notes: Figure plots the distribution of grade signals (panel A, based on May 2 Canvas scores), and final grades (panel B, based on final Canvas scores), separately for courses with above and below median RMSE, the root mean squared prediction error from a regression of final Canvas grade on May 2 course grade, with course-specific coefficients. The sample is the analysis sample as defined in the notes to Table 1 in the main paper.

(5) we consider an alternative proxy for students' signal: their actual final canvas grade, entered as a set of dummy variables. This proxy is surely imperfect but it may be better than our predicted final grade measure (as a proxy for students' signals) because students may have private information about their final grade. We find a slightly smaller coefficient on uncertainty here. In column (6) we continue to use the final grade as a proxy for students' signals, but we round up final grades above .5 to the next letter grade (so an 89.5 is a B+ in column (5) but an A- in column (6)). These alternative ways of handling students' signals measures have very little effect on our estimates.

As a final set of checks, we show in Online Appendix Table 8 robustness to alternative samples. Our baseline sample, in column (1), is restricted to classes with at least 30 students, in which canvas and registrar grades agree to within one notch. In column (2), we further restrict the sample to classes in which there is no evidence of "bunching" in the grade distribution. Specifically, we exclude classes in which at least 1 percent of students have a final Canvas grade that is an exact multiple of 10. In classes with this exact bunching, our signal and precision measures may be less reliable. In column (3) we exclude classes in which a majority of grades are an A. For these classes there may be little uncertainty (despite our measure) as students might be able to forecast their final grade. In column (4) we expand the sample to include classes of all sizes (dropping the requirement that classes contain at least 30 students). In column (5) we tighten this condition to require an exact match for all students, and in column (6) we instead require that either Canvas-Registrar grades match exactly, or that rounded-up Canvas grades match exactly.[2] In column 7 we exclude students who received an F, because we do not observe their disclosure decisions. Finally, in columns (8) we exclude courses in schools which had P/F deadlines after the start of the finals period (the College of Arts and Sciences, the Business School, and the Media School). Excluding these courses cuts our sample by two-thirds but has little effect on our estimates, probably because even in these schools most courses did not release final grades until after the deadline, and students in practice made their P/F decisions well before the deadline.

Overall our estimates are not sensitive to the exact sample inclusion criteria. We acknowledge that our estimates would change substantially if we included classes in which the Canvas and Registrar grades disagreed. However in these courses our signal and uncertainty measures are both likely highly error-prone, and so it is not particularly concerning to us that our results would change if we included these courses.

2.2 Heterogeneity

In Online Appendix Table 9 we investigate heterogeneous effects of uncertainty across three dimensions. In column (1) we look at heterogeneous effects by course number, splitting on courses with numbers less than 300. This split is informative because letter grades in upper level courses might carry particular weight, and cause risk seeking rather than risk averse behavior. For example, achieving a high-enough grade in in organic chemistry might be especially important for medical school, and likewise for upper

[2] So for example if a student has an 89.5 on Canvas and an A- registrar grade, we consider it a match in this specification, but if she has an 89.4 and an A-, we do not.

level finance classes and investment banking recruiting. We find significant effects of uncertainty in both upper and lower level courses, but larger effects in upper level classes, showing that our results are not driven by a particular course level.

In columns (2) and (3) we split on grade. In column (2) we split on students' incoming GPA, allowing for separate effects for students with GPA's above 2.0 and below. This threshold is important because some programs require students to maintain a GPA of at least 2.0, and this GPA cutoff might, again, generate risk-seeking behavior for students below the cutoff. We find however that uncertainty reduces disclosure for both high and low GPA students. In column (3) we instead split on the signal, looking at observations with an interim grade of above or below the median, 3.3. Splitting on the signal is useful because because students with a signal above it are very unlikely to receive a low grade: fewer than 1 percent receive a grade of 2.7 or lower. This is important for two reasons. First, for these students there is no risk of failing, and we do not observe disclosure decisions for failing grades. Second, some prerequisite courses (e.g., calculus) require a minimum grade of 1.7 or 2.0, and these minimums could induce risk-seeking behavior for students with low signals. We see in column (3) that both high- and low-signal students respond similarly to uncertainty. Thus uncertainty is associated with disclosure for a wide array of students and courses.

Table 6: Robustness of effect of uncertainty on disclosure to alternative measures of uncertainty

Measure	RMSE	Pr(Switch $\geq$1)	Pr(Switch $\geq$2)	High RMSE	Negative correlation	RMSE 2 segments	RMSE 3 segments	RMSE 4 segments
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Uncertainty	-0.007	-0.098	-0.163	-0.039	-0.042	-0.009	-0.009	-0.009
	(0.001)	(0.013)	(0.025)	(0.005)	(0.014)	(0.001)	(0.001)	(0.001)
R^2	0.644	0.644	0.645	0.644	0.643	0.644	0.644	0.644
# Observations	37,210	37,210	37,210	37,210	37,107	37,210	37,210	37,210
# Students	14,412	14,412	14,412	14,412	14,386	14,412	14,412	14,412
# Courses	778	778	778	778	776	778	778	778
Mean uncertainty	3.914	0.436	0.143	0.554	0.174	3.805	3.768	3.755

Notes: Table reports coefficients from a linear probability model of disclosure on the indicated uncertainty measure, as well as controls for a set of dummy variables for each grade signal (A, A-, B+, etc.); dummy variables for course level (200/300/400); class size; average incoming GPA of students enrolled in the class, average grade bump, and student fixed effects. Pr(Switch $\geq$1) is the course-specific probability the interim and final grade differ by one or more notches, and Pr(Switch $\geq$2) is the probability they differ by two or more notches. High RMSE is defined as above-median RMSE. Negative correlation is -1 times the class-specific correlation between interim and final scores. In the final three columns, uncertainty is the RMSE from a piece-wise linear model of the relationship between interim and final scores, with the indicated number of line segments, with breaks at the median/tercile/quartile. Sample consists of full-term or second half courses, with at least 30 students in which Canvas and registrar grades always agree to within 1 notch. Robust standard errors, clustered on course, in parentheses.

Table 7: Robustness of effect of uncertainty on disclosure to alternative signal definitions

Signal	Baseline (1)	Linear (2)	Cubic (3)	+GPA (4)	Final (5)	Round Up (6)
RMSE	-0.007	-0.010	-0.008	-0.007	-0.005	-0.005
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
R^2	0.644	0.604	0.619	0.647	0.716	0.717
# Observations	37,210	37,210	37,210	37,210	37,210	37,210
# Students	14,412	14,412	14,412	14,412	14,412	14,412
# Courses	778	778	778	778	778	778

Notes: Table reports coefficients from a linear probability model of disclosure on course RMSE and the indicated signal measure, as well as dummy variables for course level (200/300/400); class size; average incoming GPA of students enrolled in the class, average grade bump, and student fixed effects. The baseline signal measure is a set of dummy variables for predicted final grade given May 2 grade. Column (2) controls linearly for predicted final grade, and column (3) controls for a cubic in predicted final grade. Column (2) uses student GPA as well as May 2 grade to predict final grade. Column (5) controls for dummy variables for final grade rather than predicted final grade. Column (6) controls for dummy variables for final grade, with grades ending in .5 and higher rounded up. Sample consists of full-term or second half courses, with at least 30 students in which Canvas and registrar grades always agree to within 1 notch. Robust standard errors, clustered on course, in parentheses.

Table 8: Robustness of effect of uncertainty on disclosure to alternative samples

Sample	Baseline (1)	No bunch (2)	Minority A (3)	Small (4)	Match Exact (5)	Match Round (6)	No F (7)	Early Deadlines (8)
RMSE	-0.007	-0.008	-0.006	-0.005	-0.007	-0.007	-0.008	-0.008
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
R^2	0.644	0.666	0.677	0.605	0.646	0.639	0.660	0.637
# Observations	37,210	26,733	19,905	70,436	9,033	17,626	36,552	13,110
# Students	14,412	10,944	8,438	23,346	4,090	7,577	14,173	5,246
# Courses	778	634	497	2,432	330	487	778	340

Notes: Table reports coefficients from a linear probability model of disclosure on course RMSE, as well as controls for a set of dummy variables for each grade signal (A, A-, B+, etc.); dummy variables for course level (200/300/400); class size; average incoming GPA of students enrolled in the class, average grade bump, and student fixed effects. The baseline sample consists of full-term or second half courses, with at least 30 students in which Canvas and registrar grades always agree to within 1 notch. "No bunch" excludes classes where more than 1% of grades are exact multiples of 10. "Minority A" excludes classes where a majority of students earned an A. "Small" includes classes of all sizes. "Match exact" excludes classes where Canvas and Registrar grades disagree at all. "Match Round" is limited to classes with at least 30 students in which all Canvas and Registrar grades either match exactly, or match after rounding up Canvas grades ending in .5 and higher. "No F" excludes students with a final grade of F. "Early deadlines" excludes courses in three schools with late P/F selection deadlines, (the College of Arts and Sciences, the Media School, and the Business School). Robust standard errors, clustered on course, in parentheses.

Table 9: Heterogeneous effects of uncertainty

Split on	Course number $<$ 300 (1)	GPA$\leq$ 2.0 (2)	Signal $\leq$ 3.3 (3)
RMSE main effect	-0.009	-0.006	-0.008
	(0.002)	(0.001)	(0.001)
RMSE-interaction	0.002	-0.011	0.002
	(0.003)	(0.003)	(0.003)
Sum of main effect and interaction	-0.007	-0.016	-0.006
	(0.002)	(0.003)	(0.002)
R^2	0.644	0.644	0.644
# Observations	37,210	37,210	37,210
# Students	14,412	14,412	14,412
# Courses	778	778	778

Notes: Table reports coefficients from a linear probability model of disclosure on course RMSE ("main effect"), as well as an interaction between course RMSE and the indicated splitting variable ("RMSE-interaction"), and the implied overall effect of RMSE for split group (i.e. the sum of the main effect and the interaction). Additional controls include a set of dummy variables for each grade signal (A, A-, B+, etc.); dummy variables for course level (200/300/400); class size; average incoming GPA of students enrolled in the class, and student fixed effects. The sample consists of full-term or second half courses, with at least 30 students in which Canvas and registrar grades always agree to within 1 notch. Robust standard errors, clustered on course, in parentheses.