# The signal quality of grades across academic fields

James Thomas*

August 15, 2018

**Abstract**

Grades are an important mechanism for college students to learn about their academic abilities; however, there is limited evidence on which grades reveal abilities most efficiently. This paper estimates a Bayesian model of correlated learning to measure the overall signal quality of grades across academic fields. Grades in one academic field signal ability in all fields, allowing me to measure both "own-field" signal quality and "spillover" signal quality. I estimate this model using transcript data from Duke University and an adaptation of the Expectation-Maximization algorithm. Estimates reveal a clear division between information rich science, engineering, and economics grades and less informative humanities and social science grades. I find information spillovers across academic fields are substantial; in many specifications, spillovers are so powerful that precise science, engineering, and economics grades are more informative about humanities and social science abilities than humanities and social science grades. A supplemental analysis suggests grade compression reduces the signal quality of all grades but cannot explain the differences in signal quality across academic fields.

## 1 Introduction

Students arrive on university campuses with limited information about their academic abilities, making it difficult to navigate academic life. Previous literature identifies earned grades as an important signal of academic abilities to students, yet there is little evidence on which grades reveal abilities most efficiently.[1] Grading methods and standards vary widely across

*Yale University; james.r.thomas@yale.edu. I am especially grateful to Peter Arcidiacono, V. Joseph Hotz, Robert Garlick, Hugh Macartney, and Arnaud Maurel for ongoing advice on this project and to Ken Spenner for providing access to the Campus Life and Learning data. This paper has also benefited from useful feedback from Thomas Ahn, Jared Ashworth, Esteban Aucejo, Hanming Fang, Jon James, Elena Pastorino, Tyler Ransom, Ed Vytlacil, anonymous referees, and participants in the Duke University Labor Lunch Seminar. All remaining errors are my own.

[1] Arcidiacono (2004), Stinebrickner and Stinebrickner (2012; 2014a; 2014b), Zafar (2011). See Altonji, Blom, and Meghir (2012) for a review.

academic fields (Johnson, 2003), which suggests the information quality of grades may vary across fields as well.

It is important to know which fields offer the most informative grades because missing information about academic abilities can be quite burdensome. Missing information about academic abilities may lead to delayed graduation, delayed specialization, or specialization mismatch—whereby the chosen field differs from the field that would be chosen if full information were available.[2] This is especially important for underprivileged students. These students are generally less informed when they arrive on campus (Avery and Hoxby, 2014) and are more likely to face financial constraints which make delaying graduation infeasible (Manski, 1992).

This paper compares the information quality of grades across academic fields. I find evidence of sizable differences: natural science, engineering, and economics courses are significantly more informative than humanities and social science courses. In a preferred specification, one engineering grade has the same average information content as three humanities grades.[3]

I measure differences in the information quality of grades across academic fields using transcript data from Duke University and a Bayesian model of correlated learning. My approach uses unexplained variance in earned grades within and across students to infer the information quality of different grades. In high information content fields, the unexplained variance in grades earned by a single student will be low relative to the unexplained population variance in earned grades. For such a field, a student can form relatively accurate and precise beliefs about how her unobserved ability in this field compares to others after receiving just a few grades. Conversely, in low information content fields, a single student's unexplained variance in grades will closely resemble the population unexplained variance in grades. This implies grades in this field tell students very little about their relative unobserved ability.

The correlated learning feature allows grades in each academic field to inform students

[2]See Bound, Lovenheim, and Turner (2012) on delayed graduation. See Bordon and Fu (2015) and Malamud (2010, 2011) on the relative importance of delayed specialization and specialization mismatch. See Allen and van der Velden (2001) and Nordin, Persson, and Rooth (2010) on the effects of specialization mismatch in the labor market.

[3]I measure information content by how much a single grade reduces uncertainty in a student's beliefs about her initially unobserved abilities where uncertainty is defined as variance in beliefs. In a preferred specification, one engineering grade reduces uncertainty in beliefs about humanities ability by 24%, social science ability by 20%, natural science ability by 36%, engineering ability by 37%, and economics ability by 33%. Three humanities grades reduce uncertainty in humanities ability by 42%, social science ability by 31%, natural science ability by 22%, engineering ability by 28%, and economics ability by 25%. This corresponds to a 30% average reduction from the engineering grade and a 29% average reduction from the three humanities grades.

about their unobserved abilities in all fields. These information spillovers are identified by the within-student correlation in earned grade residuals across academic fields. Intuitively, if students who outperform expectations in field $k$ generally also outperform expectations in $k'$, students who outperform in either one of these fields will infer that they have high unobserved abilities in both fields. Allowing for these information spillovers across fields turns out to be quite important. Even with rich controls, I find high within-student correlation in grade residuals for all pairs of academic fields. In some specifications, this produces information spillovers which are so strong that high information content engineering signals reveal more about a student's unobserved ability in humanities than humanities signals themselves.

In this framework, it is important to distinguish what components of a student's abilities are known to them when they initially enroll and what components are initially unknown and revealed through grades. Students' initial information may derive from performance in high school, scores on admissions tests, knowledge of how individuals with similar demographics generally perform, or other factors. Because I cannot observe exactly what information students use to form initial beliefs about their abilities, I conduct my analysis under various assumptions about student information sets. First, I repeat my analysis including different sets of observed covariates in student information sets. Although the unexplained variance in grades clearly decreases when information sets are richer, the main finding that natural science, engineering, and economics grades are relatively more informative than humanities and social science grades holds in all specifications.

Second, I also report results which allow for unobserved heterogeneity in both the mean and variance in initial beliefs about unobserved abilities. This allows observationally equivalent students to differ in their initial expected performance and in their uncertainty—measured by variance—in their initial expectations. This implies that a student with higher initial expected ability may perceive a B+ to be worse performance than expected, leading to a downward revision in beliefs, while an observationally equivalent student with lower initial expectations may consider a B+ to be positive information worthy of an upward revision in beliefs. Furthermore, it implies that students with less certain initial beliefs will update their beliefs more in response to grade information than observationally equivalent students who are more certain about their initial beliefs. Results suggest some students arrive with higher and more certain initial expectations about their performance. These students come from more advantaged backgrounds, suggesting pre-college experiences play an important role in determining initial beliefs and highlighting the need to allow for unobserved heterogeneity in initial beliefs.

Another challenge for this study is that results can be quite sensitive to how individual courses are aggregated into academic fields. To see why, suppose there are two distinct subfields of natural science—physical sciences and life sciences—and that grades in each subfield are perfect signals of unobserved ability for that subfield but not for the other. A specification which treats physical sciences and life sciences as separate fields will find that there is no within-student unexplained variance in grades earned within either subfield and will thus conclude that grades in either science are perfectly informative for that field. However, a specification which aggregates both sciences into a single field will find within-student unexplained variance in grades in this aggregate field and will thus conclude that science grades are less than perfectly informative.

To assess robustness to aggregation misspecification, I report results using various definitions of fields. Results with the least aggregated definition reveal some interesting heterogeneity: grades in the foreign language sub-field of humanities are comparable in informativeness to grades in science, engineering and economics; however, aside from that exception, the main finding that humanities and social science grades are less informative than those in science, engineering, and economics holds across specifications.

Why are science, engineering, and economics—henceforth, SEE—grades more informative than humanities and social science—henceforth, HuSS—grades? One can think of each instructor's grading policy as a sequence of mappings: First, there is a mapping from a set of academic performances—responses to test questions, essays, etc.—into a cardinal ranking of students. Second, there is a mapping from the cardinal ranking of students to letter grades with associated grade point values.[4] One reason why SEE grades might be more informative than HuSS grades is that there is less subjectivity in the first mapping from academic performances to cardinal rankings in SEE courses. SEE courses typically assess student performance using multiple-choice or short-answer format questions with well-defined correct answers, while HuSS courses typically assess student performance with papers or essay questions where the grader's judgement plays a more prominent role. Greater subjectivity in the mapping from performances to cardinal rankings adds noise which could explain why SEE grades are more informative than HuSS grades.

Another reason why SEE grades might be more informative than HuSS grades is that SEE instructors generally employ more of the grade distribution in the second mapping from cardinal rankings of students to letter grades. In my data, 94.2% of HuSS grades are a B- or above compared to 82.1% of SEE grades, and 35.3% of HuSS grades are A (the maximum) compared to 29.7% of SEE grades. By compressing cardinal rankings of students into a smaller number of grade categories, HuSS instructors limit the capacity of their grades to distinguish students and thereby convey information (Mansfield, 2001). As such, greater

[4]Possible grades and corresponding grade point values are A (4.0), A- (3.7), B+ (3.3), B (3.0), B- (2.7), C+ (2.3), C (2.0), C- (1.7), D+ (1.3), D (1.0), F (0.0).

grade compression in HuSS courses could explain why SEE grades are more informative than HuSS grades.[5]

To provide some suggestive evidence on the role of grade compression in reducing the signal quality of grades, I repeat my analysis on a restricted sample which removes courses with severely compressed grade distributions. This increases estimates of the signal quality of all grades but does not reduce the gap in information quality between HuSS and SEE grades. This suggests that differences in the signal quality of HuSS and SEE grades primarily result from the larger role of grader judgement in the evaluation of student performances in HuSS courses.

My analysis suggests there are a number of policies universities could consider to improve information diffusion. First, additional graders could be assigned to HuSS courses so that multiple graders can evaluate the same responses to limit the role of individual graders' idiosyncratic tastes. Second grading policies could be introduced to encourage instructors to reduce compression and employ more of the grade distribution. This may not reduce the signal quality gap between HuSS and SEE courses, but it should increase the signal quality of all grades. Finally, curricula could be modified to reduce the number of poor signal quality courses students are obligated to take. At Duke University, students must take 13-15 humanities or social science courses but only 6-8 natural science, engineering or economics courses to satisfy graduation requirements.[6] While there are many justifications for requiring certain courses, my results suggest these requirements oppose information diffusion.

This paper relates to a long and rich literature on human capital and revelation of missing information.[7] Broadly speaking, two approaches are commonly employed: Some—including this paper—use data on observed choices and outcomes together with assumptions regarding how agents process information to estimate models of information revelation. In the

[5]Compressed distributions could also arise because true academic performances are more homogenous in the population. That is, if two instructors employ identical mappings but one observes more homogenous performances, then this instructor will assign more similar grades. Even with observed data on true academic performances, it would be impossible to compare the distributions of academic performances across fields. Moreover, as long as there is sufficient variation in academic performances to map these performances into a cardinal ranking, this cardinal ranking can be mapped into either very compressed or very dispersed grades. As such, compressed grades distributions ultimately reflect grading policies chosen by instructors.

[6]Requirements are: arts, literatures, and performance (2 courses); civilizations (2 courses); natural sciences (2 courses); quantitative studies (2 courses); social sciences (2 courses, includes economics); cross-cultural inquiry (2 courses); ethical inquiry (2 courses); science, technology and society (2 courses); foreign language (3 courses); writing (3 courses) ("Degree Requirements," 2013).

[7]Notable contributions not mentioned in the body include but are not limited to Jovanovic (1979), Miller (1984), Shaw (1987), McCall (1990), Farber and Gibbons (1996), Jovanovic and Nyarko (1997), Neal (1999), Altonji and Pierret (2001), Gibbons et al. (2005), Altonji (2005), Pavan (2011), Antonovics and Golan (2012), and Papageorgiou (2014). There also exists a rich literature examining the role of missing information in consumer decisions (Ackerberg, 2003; Crawford and Shum, 2005; Cai, Chen, and Fang, 2009) and health behaviors (Downs, Loewenstein, and Wisdom, 2009; Dupas, 2011; 2014; Jalan and Somanathan, 2008).

context of students learning about academic abilities through grades, prominent examples of this approach include Arcidiacono (2004) and Arcidiacono, Aucejo, Maurel, and Ransom (2016). Typically, this involves making the "rational expectations" assumption that initial beliefs about unobserved characteristics are based on the population distributions of those characteristics and the assumption that agents update beliefs according to Bayes' rule. An alternative strategy elicits beliefs about missing information directly from agents and analyzes how these beliefs change in response to signals. In the context of students learning about academic abilities through grades, prominent examples of this approach include Stinebricker and Stinebricker (2012; 2014b) and Zafar (2011).

Both approaches have relative advantages and disadvantages and thus complement each other nicely. The main advantage of using elicited beliefs data is that researchers can avoid making strong assumptions about how agents process information. This is especially relevant in the context of students learning about academic abilities through grades as some evidence suggests students are initially over-optimistic about their academic performance—violating the rational expectations assumption—and that some students attribute too little of poor performance to permanent factors in a way that is inconsistent with Bayes' rule (Stinebrickner and Stinebrickner, 2012).

The main advantage of the model-based approach is that it is difficult to measure uncertainty in expected performance and it is especially difficult to decompose this uncertainty into a component due to grading noise and into a component due to uncertainty in beliefs about abilities using survey questions alone. The Berea Panel Survey—by far the most extensive survey of elicited beliefs from students—includes questions asking students to attribute their better (worse) than expected grade point average to having higher (lower) ability than expected, being more (less) prepared than expected, studying more (less) than expected, or having good (bad) luck. This question has useful content and is the basis for the conclusion that some students attribute too little of poor performance to permanent factors; however, it yields only a rough measure of the signal quality of grades and may contain substantial measurement error (Stinebrickner and Stinebrickner, 2012). Without panels with reliable decompositions of uncertainty in expected performance, one cannot analyze which grades reduce uncertainty most efficiently or how grades in one field reduce uncertainty in beliefs about abilities in another.

Since my main objective is to measure the signal quality of grades across academic fields, I use a model-based approach while admitting the known limitations of assuming rational expectations and Bayesian updating. Relative to the existing literature on students learning about academic abilities through observed grades, my paper is the first analysis of the differences in the signal quality of grades across academic fields.

The remainder of the paper is organized as follows: Section 2 introduces the Duke University transcript data. Section 3 provides descriptive statistics on grade distributions and course choices. Section 4 presents the correlated learning model which I use to develop measures of signal quality across academic fields. Section 5 describes the adaptation of the EM algorithm I use to estimate the correlated learning model. Section 6 reports main results comparing the signal quality of grades across academic fields and supplemental results analyzing the relationship between grade compression and signal quality. Section 7 concludes.

## 2 Data University Transcript Data

For my empirical analysis, I employ administrative data from Duke University which include full academic transcripts and information used to make admissions decisions. These data are available for students who participated in the Campus Life and Learning (CLL) Survey which followed sub-samples of the 2001 and 2002 entering undergraduate cohorts at Duke University.[8]

These data are well suited for my analysis for two important reasons: First, full transcript data allow me to assign grade signals to academic fields. This is very important for comparing the signal quality of grades across academic fields. Previous literature often uses aggregate signals such as semester average GPA.[9] Without additional data, these aggregate signals cannot be assigned to academic fields, making it difficult to evaluate heterogeneity in signal quality. Second, full transcript data allows me to observe the same student taking many classes in multiple fields. Repeated observations of the same student in the same field are important for separating unobserved student abilities from grading noise; furthermore, outcomes for the same student across multiple fields are necessary for measuring information spillovers across academic fields.

The CLL Survey originally contacted 1,536 students in the 2001 and 2002 entering undergraduate cohorts at Duke University. Of these students, 1,132 gave consent to have their confidential records used for research purposes. Arcidiacono, Aucejo, and Spenner (2012) show non-respondents have lower SAT scores, have better educated parents, are more likely to be from private schools, and have slightly lower grade point averages than respondents. However, they note the differences are quite small and thus conclude non-response bias is minimal. To account for this non-response bias (and for the stratified sampling of the CLL), I use survey weights based on race and cohort to improve the representativeness of the sample.

After removing independent study courses and courses not taken for a grade, the dataset

[8]For detailed reports on the CLL data, see Bryant, Spenner, and Martin (2006, 2007).

[9]Arcidiacono (2004), Arcidiacono, Aucejo, Maurel, and Ransom (2016), Zafar (2011).

contains 37,432 course observations. For most of the results, courses are labeled as either humanities, social science (excluding economics), natural science, engineering, and economics using Duke's internal definition for these fields. The 37,432 course observations include 12,131 from humanities, 10,375 from social science, 10,638 from natural science, 1,622 from engineering, and 2,666 from economics.

To assess robustness to these field definitions, I also report results of a less aggregated specification which disaggregates humanities into arts (1,854 observations), foreign languages (3,740 observations), and other humanities (6,537 observations) and natural sciences into life sciences (3,312 observations) and physical sciences (7,326 observations). Additionally, I also report results of a more aggregated specification which pools humanities and social sciences into a single field and pools natural science, engineering, and economics into a single field.

While the main goal of this article is to compare signal quality across academic fields, my analysis also allows signal quality to vary by course level. This allows me to pay particularly close attention to the signal quality of grades across academic fields in introductory courses. These courses are taken when students are least informed and before crucial specialization decisions are made. As such, signal quality in these introductory courses is especially important.

To this end, I construct a measure of course difficulty level based on course number. I label all courses with a course number less than 100 as introductory and all courses with a course number equal to or greater than 100 as advanced. With this definition, 78.3% of classes taken during the freshman academic year are introductory, 42.8% of classes taken during the sophomore academic year are introductory, 18.9% of classes taken during the junior academic year are introductory, and 12.6% of the classes taken during the senior academic year are introductory.

# 3 Descriptive Statistics

Before proceeding to a model-based analysis of the signal quality of grades across academic fields I present descriptive statistics on grade distributions and course selection behavior. First, Panel A of Table 1 summarizes the distributions of earned grades by academic field and level. Notice that for introductory courses, grades are generally lower and more dispersed in science, engineering, and economics (henceforth, SEE) and higher and more compressed in humanities and social science (henceforth, HuSS). For introductory courses, humanities grades are the highest and most compressed—93.2% of grades are B- or above and 37.7% of grades are an A—while economics are the lowest and least compressed—75.2% are B- or above and 22.7% of grades are an A. Introductory engineering courses are a bit unusual in that they have a large share of A's (33.6%) but also a fairly large share of low grades (17.2% are C+ or below). Introductory social science courses have only 25.5% A's—which is comparable to science and economics and substantially lower than engineering—but still 90.7% of grades are B- or above which is substantially higher than all SEE fields.

Within all fields, grades are higher and more compressed in advanced courses relative to introductory courses. This is especially true for SEE courses, where the share of grades which are B- or above increases from 75.4% in introductory courses to 86.9% in advanced courses. However, despite the larger changes for SEE fields, grades in advanced courses remain lower and more dispersed in SEE courses and higher and more homogenous in HuSS courses. This illustrates that in both introductory and advanced courses, HuSS grades employ less of the grade distribution than SEE grades. By compressing students into a smaller number of grade categories, HuSS instructors may limit the capacity of their grades to distinguish students and thus reduce the information content of their grades.

Although SEE grades are more dispersed than HuSS grades, this alone does not necessarily imply they are more informative. In particular, variation in grades can arise either because different students have different average grades within a field or because individual students earn very different grades across multiple courses within a field. If most of the variance in grades comes from differences in average grades across students, then one earned grade gives a student an accurate and precise estimate of how she should expect to perform relative to other students. Conversely, if most of the variance in grades comes from differences in earned grades within students, then one earned grade tells a student very little about how she compares to others.

To provide suggestive evidence on the relative importance of within and across student variance, Panel B of Table 1 reports total variance in earned grades by academic field and level and residual variance by field and level after differencing out student-field-level fixed effects. Specifically, let $g_{itklc}$ represent a grade earned by student $i$ in course $c$ taken in semester $t$ which belongs to field $k$ and level $l$ and additively decompose earned grades into a student-field-level fixed effect $\alpha_{ikl}$ and a residual $\eta_{itklc}$ as follows:

$$g_{itklc} = \alpha_{ikl} + \eta_{itklc} \tag{1}$$

Total variance is variance in $g_{itklc}$ within each field-level pair. Residual variance is variance in $\eta_{itklc}$ within each field-level pair and is reported as a share of total variance.[10]

[10]Note that variance in OLS residuals $\hat{\eta}_{itklc}$ will always understate true variance in $\eta_{itklc}$ since OLS estimates of $\alpha_{ikl}$ are chosen to minimize the sum of squared residuals. As such, one should cautiously compare the ratio of residual variance to total variance across academic fields and levels rather than interpreting the the ratio of residual variance to total variance directly.

For introductory engineering and economics courses, a smaller share of total variance comes from within students (31.1% and 20.9% respectively) compared to a larger share for introductory humanities courses (54.9%). This suggests there are important differences in the informativeness of grades across academic fields in introductory courses. While the distribution of introductory humanities grades is tighter, the same student earns grades spanning a large section of the distribution in different introductory humanities courses. Conversely, introductory engineering and economics grades are more dispersed, but the same student usually earns similar grades in different introductory engineering and economics courses. Introductory engineering and economics grades thus appear to reveal the relative performance of students more clearly than introductory humanities grades.

In advanced courses, within student variance represents a similar share of total variance across all academic fields, suggesting that information differences across fields may be more important in introductory courses than in advanced courses. Introductory courses are generally taken when students are least informed and before crucial specialization decisions are made, implying that differences in information quality in introductory courses are especially salient.

Descriptive statistics in Table 1 suggest there may be important differences in the signal quality of grades across academic fields; however, these descriptive statistics may be confounded by non-random selection of students into courses. In the following sections, I present a framework for adjusting for selection on observed and unobserved student characteristics and discuss measures of the signal quality of grades across academic fields which account for this non-random selection.

Before introducing the framework for adjusting for selection, Table 2 provides brief descriptive statistics on how observed and initially unobserved student characteristics affect course choices. First, treating each instance of a student taking a course as an observation, Table 2 reports average composite SAT scores conditional on the academic field and level of the course. For comparison, Table 2 also reports average composite SAT scores conditional on only academic level. For introductory courses, statistics suggest that on average students in engineering and economics courses have significantly higher SAT scores and students in social science courses have significantly lower SAT scores than the general population of students taking introductory courses. Selection on observed characteristics is even stronger in advanced courses where science, engineering, and economics averages are all significantly higher than the general population and humanities and social science averages are significantly lower than the general population. This illustrates that non-random selection on observed student aptitude is present in introductory courses and grows stronger in advanced courses with higher aptitude students being more likely to choose science, engineering, and economics courses.

In addition to sorting into courses based on baseline aptitude and other observed measures, students may also be selecting into courses based on information which is revealed through earned grades. To provide descriptive evidence on this type of selection, Table 2 also assesses how grading residuals for introductory courses completed in freshmen and sophomore years relate to advanced courses chosen in junior and senior years. Specifically, I begin by estimating the following grade production function with rich covariates using only observations of introductory courses completed in freshmen and sophomore years:

$$g_{itklc} = X_{itk}\theta_{kl} + \eta_{itklc} \tag{2}$$

Observed covariates included in $X_{itk}$ include a quadratic in own-field experience, indicators for contemporaneous course load, indicators for race and gender, math and verbal SAT scores, and five measures of application quality. Grade production parameters $\theta_{kl}$ are allowed to vary by the field $k$ and level $l$ of the course to allow for substantial heterogeneity in the grade production process.

Next, I extract residuals from Equation (2) and average these for each student to obtain an estimate of the extent to which every student either out-performed or under-performed predictions based on her observed characteristics in introductory courses in freshmen and sophomore years. Finally, treating each instance of a student taking an advanced course in junior or senior years as an observation, I compute averages of these aggregate introductory residuals conditional on academic field and for all advanced observations for comparison. The goal is to analyze whether the unexplained component of performance in introductory courses as underclassmen is related to advanced course choices as upperclassmen.

Estimates suggest that on average students who take advanced science and economics courses as upperclassmen significantly outperformed predictions based on their observed characteristics in introductory courses as underclassmen. Conversely, students who take advanced social sciences courses as upperclassmen significantly underperformed in introductory courses as underclassmen relative to their observed characteristics. This suggests that in addition to sorting on observed characteristics such as SAT scores, student choices are also influenced by initially unobserved characteristics which are revealed by grades. The following sections present a framework for adjusting for selection on both observed and unobserved student characteristics and discuss measures of the signal quality of grades across academic fields which account for this selection.

# 4 Model of Ability Revelation

In this section, I present a model of the ability revelation process. I model how students use grade signals from each academic field to update their beliefs about their unobserved abilities in all academic fields, allowing for both "own-field" ability revelation and "spillover" ability revelation to other fields. I also model how students select courses given observed characteristics and unobserved beliefs about abilities.

I begin by introducing the primitives of my model; next, I describe how grades depend on observed and unobserved characteristics; following this, I show how grade residuals are used to update beliefs about abilities; next, I model how these beliefs and other observed characteristics affect course choices; and finally, I show how model parameters are identified from the data and discuss threats to identification.

## 4.1 Primitives

Index students by $i = 1, 2, \ldots, N$, academic semesters by $t = 1, 2, \ldots, T$, and courses within semester $t$ by $c = 1, 2, \ldots, C_{it}$. All courses are categorized into exactly one academic field $k \in \{1, \ldots, K\}$ and difficulty level $l \in \{1, \ldots, L\}$.[11] In my most general empirical application, I allow important student parameters to vary across student "types" $\tau = 1, \ldots, \mathscr{T}$ which are known by students but unobserved by the econometrician (Heckman and Singer, 1984).

## 4.2 Grade Production

Earned grades play a crucial role in this model as they are the sole mechanism through which students learn about their unobserved abilities. I assume the grade $g_{itklc}$ earned by student $i$ in a field $k$ level $l$ course $c$ taken in semester $t$ depends on observed covariates $X_{itkl}$, unobserved ability in field $k$ $\alpha_{ik}$, and grading noise $\eta_{itklc}$ according to:

$$g_{itklc} = X_{itkl}\theta_k + \alpha_{ik} + \eta_{itklc} \tag{3}$$

I assume that realizations of $\alpha_{ik}$ and $\eta_{itklc}$ are not separately observed by students and that grading noise realizations are drawn according to: $\eta_{itklc} \overset{iid}{\sim} N\left(0,\, \sigma^2_{kl}\right)$. The heteroskedasticity of grading noise by academic field and course level is one mechanism by which information quality is allowed to vary across academic fields.

Grading noise $\eta_{itklc}$ is meant to capture idiosyncratic student-course specific grading factors such as the effects of a grader's unobserved mood and preferences on her evaluation

[11] In the main empirical analysis, academic fields are humanities, social sciences, natural sciences, engineering, and economics and difficulty levels are introductory and advanced.

of a student's work. This cannot include factors which are shared with other courses in the same field—such as being high ability in a particular sub-field—or factors which are shared with other courses in the same semester—such as getting sick during finals week. Furthermore, it cannot include factors which students separately observe such as knowledge that an instructor is a particularly harsh grader or knowledge that a student did not spend her usual amount of time studying for a course. These assumptions are important for treating grades as noisy signals of unobserved abilities.

Let $\alpha_i = \begin{bmatrix}\alpha_{i1} & \cdots & \alpha_{iK}\end{bmatrix}'$ represent a vector containing all field specific unobserved abilities for student $i$. I assume individuals draw vectors of abilities from type specific multivariate normal distributions according to $\alpha_i \sim N\left(\gamma_\tau, \Delta_\tau\right)$. When students enroll as freshmen, they observe the distribution parameters $\gamma_\tau$ and $\Delta_\tau$ but do not observe their realization $\alpha_i$. The rational expectations assumption then implies that the distribution parameters $\gamma_\tau$ and $\Delta_\tau$ characterize the distribution of initial prior beliefs about abilities as well as the distribution of abilities themselves. As such, larger variance elements in $\Delta_\tau$ imply both more dispersion in unobserved abilities and more uncertainty in initial prior beliefs about abilities. More uncertain prior beliefs experience larger reductions in uncertainty from grade signals; as such, differences in the variance elements of $\Delta_\tau$ across fields provide another important mechanism by which information quality is allowed to vary across academic fields.

In a setting without information spillovers, the variance elements of $\Delta_\tau$ and the grading noise variance parameters $\sigma^2_{kl}$ can be used to construct signal-to-noise ratios which measure the own-field signal quality of grades. The signal-to-noise ratio for a field $k$ level $l$ grade signal for type $\tau$ students is given by:

$$SNR_{\tau kl} = \frac{\Delta_\tau\left(k,k\right)}{\Delta_\tau\left(k,k\right) + \sigma^2_{kl}} \tag{4}$$

where $\Delta_\tau\left(k,k\right)$ is the $k^{th}$ diagonal element of $\Delta_\tau$.

If $\Delta_\tau\left(k,k\right)$ is large relative to $\sigma^2_{kl}$, then there is high variance in unobserved abilities in field $k$—and thus high uncertainty in initial beliefs about abilities in field $k$—but very little noise in grade signals. In this scenario, $SNR_{\tau kl}$ approaches one implying grade signals approach perfect informativeness. Intuitively, this scenario implies one earned grade reveals exactly where a student falls in a very dispersed population and thus is highly informative. Conversely, if $\Delta_\tau\left(k,k\right)$ is small relative to $\sigma^2_{kl}$, then the population distribution of unobserved abilities in field $k$ is relatively homogenous—implying there is relatively little uncertainty in initial beliefs about abilities in field $k$—and only noisy signals are available. In this scenario, $SNR_{\tau kl}$ approaches zero, implying grade signals have little value.

In addition to allowing for own-field ability revelation, I also allow for information spillovers across academic fields by permitting $\Delta_\tau$ to be non-diagonal. Off-diagonal elements of $\Delta_\tau$ measure the covariance in unobserved abilities across academic fields and thus indicate the extent to which students with high unobserved abilities in one field are more likely to be high (or low) ability in other fields. As such, off-diagonal elements of $\Delta_\tau$ allow students to update their beliefs about their unobserved abilities in field $k$ after receiving grades in field $k'$. This is the third important mechanism by which information quality is allowed to vary across academic fields; noisy signals in a field that is highly correlated with others may yield more total information than precise signals in a field that is less similar to other fields.

The linear structure of Equation (3) additively decomposes earned grades into an observed (both by the student and the econometrician) component $X_{itkl}\theta_k$, an initially unobserved (both to the student and the econometrician) component $\alpha_{ik}$ which is revealed to students over time, and an idiosyncratic noise term $\eta_{itklc}$. The interpretation of $\alpha_{ik}$ changes dramatically when different variables are included in $X_{itkl}$: a rich specification assumes students are very knowledgeable about the grade production process while a parsimonious specification assumes they know very little. In the empirical analysis, I consider alternative choices for $X_{itkl}$ which range from very parsimonious to very rich. This allows me to compare the signal quality of grades across academic fields under varying assumptions about student information sets.

## 4.3 Ability Revelation

While the realization of $\alpha_i$ is unknown, students can use earned grades (and their knowledge of the grade production process) to refine their beliefs about this variable. The central focus of this paper is to compare the effectiveness of grades from different academic fields in improving the precision of these beliefs.

Let $b_{itk}$ represent student $i$'s beliefs about her unobserved ability in field $k$ prior to semester $t$ and collect all field specific beliefs into a vector $b_{it}$. When a student initially enrolls, she knows $\alpha_i \sim N(\gamma_\tau, \Delta_\tau)$ but does not observe her specific realization $\alpha_i$. The rational expectations assumption then implies that her initial prior beliefs are characterized by $b_{i1} \sim N(\gamma_\tau, \Delta_\tau)$

When a student earns a grade, she extracts the grade residual:

$$\begin{aligned} z_{itklc} &= g_{itklc} - X_{itkl}\theta_k \\ &= \alpha_{ik} + \eta_{itklc} \end{aligned} \tag{5}$$

As mentioned previously, an important assumption is that students can never separate specific signals $z_{itklc}$ into $\alpha_{ik}$ and $\eta_{itklc}$. If they could, an individual grade would be a perfect signal of $\alpha_{ik}$, and there would be no notion of signal quality. Under this assumption, $z_{itklc} \sim N\left(\alpha_{ik},\, \sigma^2_{kl}\right)$ from the student's perspective implying $z_{itckl}$ can be used as a noisy signal of field $k$ ability $\alpha_{ik}$. If unobserved abilities in field $k$ are correlated with unobserved abilities in other fields, this signal can also be used to learn about unobserved abilities in other fields.

Define the vector of average signals received prior to semester $t$ as

$$\bar{z}_{it}\left(k\right) = \begin{cases} \frac{1}{n_{itk}} \sum_{t'=1}^{t-1} \sum_{c=1}^{C_{it'}} z_{it'klc} & n_{itk} > 0 \\ \gamma_\tau & n_{itk} = 0 \end{cases} \tag{6}$$

where $n_{itk}$ is the number of field $k$ classes taken prior to semester $t$. Students combine this vector of average signals with initial prior beliefs to form posterior beliefs. The distribution of posterior beliefs is given by $b_{it} \sim N\left(\mu_{it}, \delta_{it}\right)$ (Degroot, 1970) where:

$$\mu_{it} = \gamma_\tau + \left(\Delta_\tau^{-1} + \sum_{l=1}^{L} D_{il}\Phi_l^{-1}\right)^{-1} \left(\sum_{l=1}^{L} D_{il}\Phi_l^{-1}\right)\left(\bar{z}_{it} - \gamma_\tau\right) \tag{7}$$

$$\delta_{it} = \left(\Delta_\tau^{-1} + \sum_{l=1}^{L} D_{il}\Phi_l^{-1}\right)^{-1} \tag{8}$$

where $\Phi_l$ is a diagonal $K \times K$ matrix of grading noise variances defined as,

$$\Phi_l\left(k,\, k'\right) = \begin{cases} \sigma^2_{kl} & k = k' \\ 0 & k \neq k' \end{cases}$$

and $D_{itl}$ is a diagonal $K \times K$ matrix of courses completed defined as,

$$D_{itl}\left(k,\, k'\right) = \begin{cases} n_{itkl} & k = k' \\ 0 & k \neq k' \end{cases}$$

The mean belief vector $\mu_{it}$ represents student $i$'s "best guesses" about her abilities at the beginning of period $t$. These best guesses are a weighted average of initial prior expectations $\gamma_\tau$ and new information from grade signals $\bar{z}_{it} - \gamma_\tau$, where the weights depend on the number and quality of signals received and the uncertainty in initial prior beliefs.

The diagonal variance elements of the belief covariance matrix $\delta_{it}$ measure the uncertainty in student $i$'s best guesses about her abilities. As such, these diagonal elements provide a complete and intuitive measure of how informed a student is prior to semester $t$. Large values indicate substantial uncertainty about future performance, while small values suggest students know their abilities well. Equation (8) shows that reductions in this uncertainty depend only on the variance-covariance matrix $\Delta_\tau$ and the grading noise variance terms $\sigma^2_{kl}$. As such, it is these parameters which measure the signal quality of grades and are thus the primary focus of this paper.

Notice that if $\Delta_\tau$ were diagonal, Equations (7) and (8) would imply that mean beliefs about unobserved abilities in field $k$ and uncertainty in beliefs about unobserved abilities in field $k$ would not update in response to grade signals from field $k'$. A non-diagonal $\Delta_\tau$ allows for information spillovers across academic fields allowing for a more complete analysis of signal quality across academic fields.

These formulas also demonstrate the important role of initial belief parameters $\gamma_\tau$ and $\Delta_\tau$ in the information revelation process. Students with different initial beliefs process grade signals very differently, possibly resulting in differences in the relative information quality of grades across academic fields. In some specifications, I let $\gamma_\tau$ and $\Delta_\tau$ vary by unobserved type, which allows unobserved characteristics to influence students' initial expected abilities and uncertainty in these initial expectations.

Finally, notice that when a student is forming beliefs prior to semester $t$, she gives the same weight to grade signals received in the most recent semester $t-1$ and grade signals received in the most distant semester 1. This property arises from the fundamental assumption that $\alpha_{ik}$ is fixed across semesters, which implies that old signals are equivalent to new signals for informing beliefs about $\alpha_{ik}$.

## 4.4 Course Choices

As I detail in Section 5, if the distribution parameters $\gamma_\tau$ and $\Delta_\tau$ vary by unobserved student type, data on observed student choices can help identify student types as long as there is a framework for capturing the relationship between unobserved type and observed choices. To provide such a framework, I develop a flexible model of course choices which allows unobserved type to influence choices through its affect on beliefs about unobserved abilities. The goal of this model is not to provide choice parameters with economic meaning but rather to approximate the relationship between unobserved type and other state variables and observed course choices. For this reason, I employ a flexible quasi-structural form following Bernal and Keane (2010).

I assume student $i$'s utility from choosing a field $k$ level $l$ course in semester $t$ flexibly depends on student $i$'s expected grade in a field $k$ level $l$ course $\tilde{g}_{itkl}$, the number of introductory and advanced courses completed in this field $n_{itk1}$ and $n_{itk2}$, and field-level intercepts which vary based on whether the student is an upperclassmen or underclassmen according to:[12]

$$U_{itkl} \quad = \quad \omega_{1kl}\tilde{g}_{itkl} + \sum_{l'=1}^{L} \omega_{2kl}^{l'} n_{itkl'} + \omega_{3klp(t)} + \epsilon_{itkl} \tag{9}$$

where expected grades given student information sets at the beginning of semester $t$ are given by:

$$\begin{aligned} \tilde{g}_{itkl} &= \mathbb{E}\left[g_{itckl} \mid X_{itk}, \theta_k, \mu_{it}, \delta_{it}\right] \\ &= X_{itkl}\theta_k + \mu_{itk} \end{aligned} \tag{10}$$

where $\mu_{itk}$ is the $k^{th}$ element of the vector of mean beliefs at the beginning of semester $t$ $\mu_{it}$.

From Equation (7), note that mean beliefs $\mu_{it}$ depend on the type specific parameters characterizing initial prior beliefs $\gamma_\tau$ and $\Delta_\tau$. This allows unobserved student type to influence course choices and thus provides variation for identifying unobserved student types from observed course choices. This variation will be especially useful in earlier semesters when beliefs are highly influenced by type specific initial prior beliefs and less informed by grade realizations. To be flexible in how current beliefs about expected grades affect course choices, I allow the importance of expected grades to vary by field and level so that students can place more value on higher marks in courses where they may have more future value.

The model also includes past experience and student level to flexibly account for path dependence and other dynamic considerations in course choices. The term $\sum_{l'=1}^{L} \omega_{2kl}^{l'} n_{itkl'}$ lets the importance of past coursework in own-field introductory and advanced courses vary by field and level to flexibly account for pre-requisites and allow for path dependence in course choices. Furthermore, the flexible intercepts $\omega_{3klp(t)}$ allow the general attractiveness of courses in each field-level pair to vary based on whether the student is an upperclassmen (senior or junior) or underclassmen (freshmen or sophomore).

While there are limits to this quasi-structural approach, it is worth noting that this specification can capture important economic behaviors such as strategic human capital accumulation and strategic information acquisition in a reduced form sense. For example, if past coursework affects students' capacity to succeed in subsequent courses, forward-looking students who value grades will strategically acquire human capital by giving more value to courses which provide higher human capital returns in subsequent semesters. While the

[12] Upperclassmen/underclassmen status is indicated by $p(t)$. Terms prior to the fall semester of junior year are assigned to underclassmen while terms after and including the fall semester of junior year are labeled upperclassmen.

quasi-structural approach does not explicitly model this strategic accumulation of human capital, it can capture this behavior in a reduced form sense. The flexible intercepts $\omega_{3klp(t)}$ can capture differences in the value of human capital across academic fields and course levels. Furthermore, the fact that these intercepts also vary by whether the student is an upperclassmen and the flexible way in which own-field experience enters into utility gives the specification generous scope for capturing diminishing returns to human capital.

In addition to strategic human capital accumulation, students may also strategically acquire information about unobserved abilities. If students are forward-looking and value expected grades, they will favor high information courses since richer information allows them to make more informed choices in subsequent periods. Importantly, the relative value of information quality likely diminishes as students progress through college both because marginal reductions in uncertainty decline as additional grade signals are received and because information about abilities is probably most valuable in early semesters before specialization decisions are made.

Once again, although the quasi-structural approach does not explicitly model strategic information acquisition, the flexible intercepts $\omega_{3klp(t)}$ can capture preferences for information quality which capture this behavior in a reduced form sense. These intercepts vary across academic field $k$ and course level $l$, allowing students to give higher value to field-level combinations which provide more information. Furthermore, the intercepts vary by whether the student is an upperclassmen or underclassmen, allowing the relative value of information to decay in later years. This illustrates that although this specification does not explicitly model strategic dynamic behaviors, certain behaviors crucial to this setting are nested inside the quasi-structural form.

## 4.5 Identification

As noted in Subsection 4.3, the variance covariance matrix $\Delta_\tau$ and the grading noise variance terms $\sigma^2_{kl}$ determine how much grading signals reduce uncertainty in beliefs and are thus the main focus of my analysis of the signal quality of grades across academic fields. In this subsection, I briefly discuss the intuition underlying how these and other parameters are identified from observed grades and course choices under various assumptions. For ease of exposition, I consider the simple case in which there are no observed characteristics $X_{itkl}$. It is straightforward to extend this argument to a general case with both fixed and time varying observed student characteristics.[13]

[13]Effects of time varying student characteristics are identified by within student variation in earned grades as in standard fixed effect models. Effects of fixed student characteristics are identified by projecting student-field fixed effects on fixed observed student characteristics as in Berry, Levinsohn, and Pakes (1995)

### 4.5.1 Without Unobserved Heterogeneity

First, consider a simplified case with no type specific unobserved heterogeneity in ability distributions so that all students draw their ability vectors according to $\alpha_i \sim N(\gamma, \Delta)$. With panel data in which every student takes multiple courses in field $k$, each student's unobserved ability $\alpha_{ik}$ is identified directly from that student's expected grades in field $k$. In notation,

$$\alpha_{ik} = \mathbb{E}\left[g_{itklc} \,|\, i, k\right] \tag{11}$$

where conditioning on $i$ and $k$ means restricting to observations of student $i$ in field $k$. Once $\alpha_{ik}$ is identified, one can extract grading noise realizations $\eta_{itklc}$ and identify $\sigma_{kl}^2$ from the field-level specific variance in these realizations. Finally, with panel data in which every student takes multiple courses in all fields, one can collect $\alpha_{ik}$ into ability vectors $\alpha_i$ and identify $\gamma$ and $\Delta$ from the first and second moments of the distribution of $\alpha_i$ in the student population.

The main assumption underlying this argument is that students are not sorting into courses based on time and course specific idiosyncratic grading factors $\eta_{itklc}$. In notation,

$$\mathbb{E}\left[\eta_{itklc} \,|\, i, k\right] = 0 \tag{12}$$

Two violations of this assumption seem plausible: First, students may sort into courses within fields in which they believe they have especially high ability. To assess robustness to this threat, I report results under varying definitions of fields. Sorting within fields should be less of a problem with finer definitions of fields; as such, comparing results while varying the granularity of field definitions gives suggestive evidence about the importance of within field sorting. Second, students may choose more courses in field $k$ in semesters where they expect to perform especially well in field $k$ courses. To address this, I estimate specifications which include categorical variables for course load and a quadratic in field specific experience in $X_{itkl}$ to capture the effects of semester specific factors which may make students more likely to choose courses in a certain field.

Notice that model parameters are still identified even if students sort across fields based on beliefs about fixed unobserved ability $\alpha_i$ as long as every student takes at least one course in each academic field. Taking expectations first within students and then across students ensures that the intensive margin selection of how many courses to take in a field does not give excessive weight to students who choose more courses in a field because they believe they have higher ability in that field.

While this argument demonstrates that parameters are identified in theory, the estimator implied by this argument has important limitations. The estimator implied by this argument is to estimate $\alpha_{ik}$ with student $i$'s average grade in field $k$ courses and to use these to subsequently estimate $\sigma^2_{kl}$, $\gamma$, and $\Delta$. This estimation strategy has two important limitations: First, even with full student transcripts the number of grade observations in each student-field bin is relatively small.[14] This implies that student-field averages will be noisy estimates of $\alpha_{ik}$ due to sampling error. This sampling error leads to overstating the distribution variance terms $\Delta(k,k)$—since variance in estimates of $\alpha_{ik}$ contains both true variance and variance in sampling error—and understating the noise variance terms $\sigma^2_{kl}$—since by construction student-field averages minimize within-student sum of squared grading residuals. Second, 24.9% of student-field bins contain zero courses, implying one could not even obtain a noisy estimate of $\alpha_{ik}$ for certain student-field pairs. If the choice of whether to take any courses in a field is non-random, this extensive margin selection may lead to bias in estimates of $\gamma$.

The estimator I employ uses the Bayesian structure of my model to address the issues caused by small and empty student-field cells; however, the general intuition of using within-student variation in grades and the population distribution of estimates of $\alpha_i$ to estimate parameters of interest carries forward. Specifically, let $\xi_i = \mu_{i(T+1)}$ represent student $i$'s mean belief vector after receiving all grade information and let $\Upsilon_i = \delta_{i(T+1)}$ represent the variance-covariance matrix characterizing uncertainty in student $i$'s beliefs after receiving all grade information. In addition to characterizing student $i$'s terminal beliefs about her abilities, the parameters $\xi_i$ and $\Upsilon_i$ also represent the econometrician's best guesses for $\alpha_i$ and the variance-covariance in these guesses under the Bayesian structure of my model. This implies $\xi_i$ can be used instead of student-field averages as an estimate of $\alpha_i$ to recover $\sigma^2_{kl}$, $\gamma$, and $\Delta$ following steps similar to those described previously.

The Bayesian structure addresses the issues caused by small and empty student-field cells in two ways: First, the correlated learning structure implies that grades earned in one field inform the entire vector of terminal means $\xi_i$. Intuitively, this means that guesses for $\alpha_{ik}$ for a student who takes few courses in field $k$ will be mostly informed by that student's performance in other fields and the relationship between abilities in those fields and field $k$. This provides substantially more variation for forming best guesses of field specific abilities. Second, the Bayesian structure implies that uncertainty in best guesses are characterized by the terminal variance-covariance matrix $\Upsilon_i$. This allows one to adjust for the error associated with using guesses for $\alpha_i$ rather than true abilities when recovering $\sigma^2_{kl}$ and $\Delta$.

From Equations (7) and (8), note that the parameters characterizing terminal beliefs $\xi_i$ and $\Upsilon_i$, which are used to recover $\gamma$, $\sigma^2_{kl}$ and $\Delta$, themselves depend on $\gamma$, $\sigma^2_{kl}$ and $\Delta$. The EM algorithm presented in the upcoming section describes an iterative method for breaking

[14] On average, there are 6.55 observations in a student-field bin.

this circular relationship.

### 4.5.2 With Unobserved Heterogeneity

I now consider a general case in which students draw their ability vectors from distributions which depend on their type $\tau$ where type is known by students but not by the econometrician. As in the case of no unobserved heterogeneity, student specific ability vectors $\alpha_i$ are identified directly from student-field specific expected grades, and $\sigma^2_{kl}$ are identified from field-level specific variances in grading noise realizations.

If student type $\tau$ were observed by the econometrician, one could identify $\gamma_\tau$ and $\Delta_\tau$ from the first and second moments of the distribution of $\alpha_i$ for the subset of the students who are type $\tau$. This suggests all parameters can be recovered given a method for classifying students into unobserved types.

My method for classifying students into unobserved types uses the structure of grade production and course choices and panels of earned grades and course choices to classify students based on persistent similarities in course choices and grade outcomes.[15] The method classifies students probabilistically rather than deterministically; as such, the goal is to compute the probability that student $i$ is type $\tau$ denoted by $q_{i\tau}$. To do this, I evaluate how consistent each student's choices and grades are with the choices and grades of a student whose abilities were drawn from the type $\tau$ distribution relative to other types. For observed grades, this amounts to assessing how likely one is to draw the estimate of $\alpha_i$ from $N(\gamma_\tau, \Delta_\tau)$ relative to other type specific distributions. Intuitively, a student who consistently performs well in all fields probably has high values in $\alpha_i$ and thus is more likely to have drawn her ability vector from a distribution with higher mean values.

For observed course choices, the method relies on the rational expectations assumption and the structure of utility in Equation (9) which assumes that expected grades influence course choices. The rational expectations assumption implies that the distribution $N(\gamma_\tau, \Delta_\tau)$ characterizes both the distribution from which ability vectors are drawn and the distribution which characterizes initial beliefs about abilities for type $\tau$ students. This implies that type $\tau$ students will choose different courses than type $\tau'$ students in expectation because their differing initial prior beliefs lead them to have differing expected grades according to Equations (7) and (10). This will be especially true in earlier semesters when few grade signals have been received and beliefs depend more on type specific initial prior beliefs. Intuitively, a student who consistently chooses natural science courses in early semesters is

[15]When observed student characteristics $X_{itkl}$ are included in Equation (3), the method classifies students based on persistent similarities in course choices and grade residuals which are not explained by observed student characteristics.

more likely to have initial prior beliefs characterized by a distribution with a relatively higher mean in science courses.

Once I have recovered the type probabilities $q_{i\tau}$, I can recover $\gamma_\tau$ and $\Delta_\tau$ from first and second weighted moments of the distribution of $\xi_{i\tau}$ which give more weight to students who are more likely to be type $\tau$ by using type probabilities $q_{i\tau}$ as weights. Once again, there is a circular relationship in this argument: recovering $\gamma_\tau$ and $\Delta_\tau$ requires type probabilities $q_{i\tau}$ and terminal belief parameters $\xi_{i\tau}$ and $\Upsilon_{i\tau}$ but constructing these probabilities and parameters requires estimates of $\gamma_\tau$, $\sigma^2_{kl}$ and $\Delta_\tau$. As before, the EM algorithm presented in the upcoming section describes an iterative method for breaking this circular relationship.

# 5 Expectation-Maximization Estimation Procedure

This section describes the adaptation of the Expectation-Maximization (EM) algorithm I use to estimate my ability revelation model. The EM Algorithm is an iterative approach to solve for the parameter values which maximize a likelihood function (Dempster, Laird, and Rubin, 1977). It is useful for cases in which unobserved variables make it computationally difficult to numerically maximize the likelihood function all at once. In this setting, the unobserved variables are unobserved student type $\tau$ and unobserved abilities $\alpha_i$. This adaptation of the EM algorithm is similar to the one used in James (2012).

I begin this section by deriving the full likelihood function which includes both the likelihood of observing course choices and the likelihood of observing grades given these choices. Next I provide a general overview of the EM algorithm. After this, I describe how my adaptation of the EM algorithm solves for the parameter values which maximize this likelihood function in the simplified case in which unobserved ability distribution parameters $\gamma$ and $\Delta$ are the same for all students. I later relax this assumption and describe the algorithm in a general case where $\gamma_\tau$ and $\Delta_\tau$ vary by unobserved student type $\tau$.

## 5.1 Likelihood Function

Each individual's likelihood contribution depends on observed course choices and observed grades in chosen courses. To build the full likelihood function, I begin by presenting grade likelihood contributions conditional on ability $\alpha_i$ and choice likelihood contributions conditional on unobserved student type $\tau$. I derive the full likelihood function by combining these conditional grade and choice contributions and integrating over unobserved ability and student type.

### 5.1.1 Grade Likelihood Conditional on Ability

The likelihood of observing grade $g_{itklc}$ given ability $\alpha_i$ and parameter values $\theta_k$ and $\sigma_{kl}$ is:

$$l\left(g_{itklc}\,|\,\alpha_i,\theta_k,\sigma_{kl}\right)=\frac{1}{\sqrt{2\pi\sigma_{kl}^2}}\exp\left(-\frac{\left(g_{itklc}-X_{itkl}\theta_k-\alpha_{ik}\right)^2}{2\sigma_{kl}^2}\right) \tag{13}$$

where the structure arises from the assumption that grading noise is normally distributed.

Because transitory grading noise $\eta_{itklc}$ is independent across courses, the likelihood of observing the grade vector $\mathbf{g_i}$ conditional on ability $\alpha_i$ and parameter values $\theta_k$ and $\sigma_{kl}$ is given by:

$$\mathscr{L}_i^g\left(\mathbf{g_i}\,|\,\alpha_i,\theta_k,\sigma_{kl}\right)=\prod_{t=1}^{T}\prod_{c=1}^{C_{it}}\prod_{k=1}^{K}\prod_{l=1}^{L} l\left(g_{itklc}\,|\,\alpha_i,\theta_k,\sigma_{kl}\right)^{(d_{itc}=\{k,l\})} \tag{14}$$

This represents individual $i$'s grade likelihood contribution conditional on her ability $\alpha_i$.

### 5.1.2 Choice Likelihood Conditional on Student Type

To develop choice likelihood contributions conditional on unobserved type, I begin by assuming preference shocks $\epsilon_{itklc}$ are iid type 1 extreme value. The probability that individual $i$ chooses a field $k$ level $l$ course in semester $t$ given student type, grade parameter values $\gamma_\tau$, $\theta_k$, $\sigma_{kl}$, $\Delta_\tau$, and choice parameter values $\omega$ is then given by:

$$\Pr\left(d_{itc}=\{k,l\}\mid\tau,\gamma_\tau,\theta_k,\sigma_{kl},\Delta_\tau,\omega\right)=\frac{\exp\left(u_{itkl}\right)}{\sum_{k'=1}^{K}\sum_{l'=1}^{L}\exp\left(u_{itk'l'}\right)} \tag{15}$$

where $u_{itkl}$ is the deterministic component of choice utility presented in subsection 4.4:

$$u_{itkl}=\omega_{1kl}\tilde{g}_{itkl}+\sum_{l'=1}^{L}\omega_{2kl}^{l'}n_{itkl'}+\omega_{3klp(t)} \tag{16}$$

After conditioning on unobserved student type, past grades, and observed characteristics, choices are independent of $\alpha_i$. This is because course utility depends on expected grades $\tilde{g}_{it\tau kl}$ at the beginning of period $t$, and these expectations are formed using current beliefs about $\alpha_i$ rather than true $\alpha_i$. Following Equation (7), beliefs about $\alpha_i$ are a function of type specific initial prior beliefs, model parameters, and grade residuals. As such, unobserved student type, past grades, observed characteristics, and model parameters contain all information necessary to capture how beliefs about unobserved $\alpha_i$ influence course choices (James, 2012).

Because preference shocks are independent across courses, the likelihood of observing the choice vector $\mathbf{d_i}$ conditional on type, grade parameter values $\gamma_\tau$, $\theta_k$, $\sigma_{kl}$, $\Delta_\tau$, and choice parameter values $\omega$ is given by:

$$\mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \gamma_\tau, \theta_k, \sigma_{kl}, \Delta_\tau, \omega\right) = \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} \prod_{k=1}^{K} \prod_{l=1}^{L} \Pr\left(d_{itc} = \{k, l\} \,|\, \tau, \gamma_\tau, \theta_k, \sigma_{kl}, \Delta_\tau, \omega\right)^{(d_{itc}=\{k,l\})} \tag{17}$$

#### 5.1.3 Full Likelihood

At this point, I have constructed grade likelihood contributions conditional on unobserved ability $\alpha_i$ and choice likelihood contributions conditional on unobserved student type. To form individual likelihood contributions, I combine grade likelihood contributions conditional on ability and choice likelihood contributions conditional on type and integrate over unobserved type and unobserved abilities conditional on type. Because choice likelihood contributions do not depend on unobserved abilities, the integration over unobserved abilities conditional on type only pertains to the grade component of the likelihood. This yields the following individual likelihood contributions:

$$\mathscr{L}_i = \sum_{\tau=1}^{\mathscr{T}} \pi_\tau \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \gamma_\tau, \theta_k, \sigma_{kl}, \Delta_\tau, \omega\right) \left[\int_{\alpha_i} \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma_\tau, \Delta_\tau\right) d\alpha_i\right] \tag{18}$$

where $\pi_\tau$ is the unconditional probability of being type $\tau$ and $f(\cdot)$ is the $K$-dimensional normal PDF.

The full log-likelihood is then given by:

$$\ln \mathscr{L} = \sum_{i=1}^{N} \ln \left\{ \sum_{\tau=}^{\mathscr{T}} \pi_\tau \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \gamma_\tau, \theta_k, \sigma_{kl}, \Delta_\tau, \omega\right) \left[\int_{\alpha_i} \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma_\tau, \Delta_\tau\right) d\alpha_i\right] \right\} \tag{19}$$

## 5.2 EM Algorithm: Overview

The $K$-dimensional integration of $\alpha_i$ for every student $i$ as each possible unobserved type $\tau$ makes it computationally burdensome to calculate the log-likelihood function value at specific parameter values. This means numerical maximization techniques that rely on repeated evaluations of the maximand will be unfeasibly slow. The EM algorithm is a useful method for maximizing likelihood functions, such as this one, which contain a large number of unobserved variables. Arcidiacono and Jones (2003) show the algorithm can be adapted to restore additive separability within the likelihood function, making it possible to solve for maximizing parameter values sequentially rather than jointly. This greatly reduces computational burden at the cost of modest efficiency losses.

In general, the EM algorithm involves iteratively repeating two steps: In the E step, observed data and guesses for parameter values are used to estimate probability distributions for each individual's unobserved variables. Intuitively, these probability distributions are estimated by evaluating how likely the individual's observed data are to have occurred for all possible values of the unobserved variables. For unobserved student types $\tau$, these distributions are fully defined by the type probabilities $q_{i\tau}$. For unobserved student abilities, these distributions are characterized by the mean $\xi_i$ and the variance-covariance matrix $\Upsilon_i$. These probability distributions are then used to construct an expected likelihood function given these distributions. In the M step, this expected likelihood function is maximized with respect to all parameter values. The E step is then repeated using these updated parameter estimates, and the algorithm iterates until sequential parameter estimates become arbitrarily close.

## 5.3 EM Algorithm: Homogenous Distributions of Unobserved Ability

I now describe my adaptation of the EM algorithm in the special case where all students draw unobserved abilities as $\alpha_i \sim N(\gamma, \Delta)$. This is equivalent to assuming all students enter college with the same beliefs about the unobserved components of their academic abilities.

### 5.3.1 Likelihood Function: Homogenous Distributions of Unobserved Ability

Without unobserved variation in initial beliefs, the full log likelihood is given by:

$$\ln \mathscr{L} = \sum_{i=1}^{N} \ln \mathscr{L}_i^c \left(\mathbf{d_i} \mid \gamma, \theta_k, \sigma_{kl}, \Delta, \omega\right) + \sum_{i=1}^{N} \ln \left[\int_{\alpha_i} \mathscr{L}_i^g \left(\mathbf{g_i} \mid \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \mid \gamma, \Delta\right) d\alpha_i\right] \quad (20)$$

While it is less efficient, the additive separability allows the grade parameters $\gamma$, $\theta_k$, $\sigma_{kl}$, and $\Delta$ to be consistently estimated by maximizing the grade component of the likelihood only. This obviates the need to model course choices and reduces the computational burden of the estimation routine. However, maximizing the grade component of the likelihood still requires $K$-dimensional integration of $\alpha_i$ for every student, which is computationally challenging. I use an adaptation of the EM algorithm to avoid this computational burden.

### 5.3.2 EM Algorithm: Homogenous Distributions of Unobserved Ability

In this case, the unobserved variable confounding estimation is $\alpha_i$. As such, the E step involves estimating the distribution of $\alpha_i$ given observed grades and parameter estimates from the previous iteration. As discussed previously, the Bayesian structure implies that the most informed distribution of $\alpha_i$ given parameter estimates from a previous iteration can be derived by updating iteration $m$ estimates of the initial prior distribution $\alpha_i \sim N\left(\gamma^m,\, \Delta^m\right)$ with grade signals from every grade earned by student $i$. This results in iteration $m$ terminal posterior distributions $\alpha_i \sim N\left(\xi_i^m,\, \Upsilon_i^m\right)$ where $\xi_i^m$ and $\Upsilon_i^m$ are equivalent to $\mu_{i(T+1)}$ and $\delta_{i(T+1)}$ in Equations (7) and (8) given iteration $m$ parameter estimates. Intuitively, this step involves using the Bayesian structure to form guesses of each student's ability vector and the uncertainty in these guesses given iteration $m$ estimates of model parameters.

Next, I use these iteration $m$ individual specific distributions to form the following expected log-likelihood function conditional on individual specific distribution parameters $\xi_i^m$ and $\Upsilon_i^m$:

$$Q_{hom}\left(\Theta_g \,|\, \Theta_g^m\right) = \sum_{i=1}^{N} \int_{\alpha_i} \ln \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \xi_i^m, \Upsilon_i^m\right) d\alpha_i \tag{21}$$

The M step maximizes $Q_{hom}\left(\Theta_g \,|\, \Theta_g^m\right)$ over the subset of parameters $\Theta_g = \{\gamma, \Delta, \theta_k, \sigma_{kl}\}$ taking iteration $m$ estimates as given.

All components of the solution to $\text{argmax}_{\Theta_g} Q_{hom}\left(\Theta_g \,|\, \Theta_g^m\right)$ have closed form expressions with relatively intuitive interpretations: First, iteration $m$ estimates of $\theta_k$ come from field specific regressions of $g_{itklc} - \xi_{ik}^m$ on $X_{itkl}$. Second, iteration $m$ estimates of $\sigma_{kl}^2$ come from the residuals of this regression using the measure of uncertainty in best guesses $\Upsilon_i^m$ to account for the additional noise generated by using $\xi_{ik}^m$ in the regression rather than $\alpha_{ik}$. Third, the iteration $m$ estimate of $\gamma$ is simply the average of best guesses $\xi_i^m$ across all students. Finally, the iteration $m$ estimate of $\Delta$ comes from the variance-covariance in the population distribution of $\xi_i^m$, once again using $\Upsilon_i^m$ to account for the additional noise generated by using $\xi_i^m$ rather than $\alpha_k$. Additional details and exact formulas are presented in Appendix A.

The algorithm is then repeated using the solution to $\text{argmax}_{\Theta_g} Q_{hom}\left(\Theta_g \,|\, \Theta_g^m\right)$ as iteration $m+1$ parameter estimates. Final estimates are obtained when the algorithm converges such that $\Theta_g^\star = \text{argmax}_{\Theta_g} Q_{hom}\left(\Theta_g \,|\, \Theta_g^\star\right)$. Intuitively, convergence means that the estimates of model parameters which are consistent with $\xi_i^m$ and $\Upsilon_i^m$ characterizing terminal beliefs are the same as the values of model parameters which imply that $\xi_i^m$ and $\Upsilon_i^m$ describe terminal beliefs. When this criterion is satisfied, the circular relationship between parameter estimates and terminal beliefs holds.

### 5.4 EM Algorithm: Heterogeneous Distributions of Unobserved Ability

I now describe my adaptation of the EM algorithm in the general case in which a student's unobserved type $\tau$ determines the distribution her unobserved abilities are drawn from. I allow both the distribution mean $\gamma_\tau$ and the variance-covariance matrix $\Delta_\tau$ to vary by unobserved type. This allows characteristics which are not observed by the econometrician to influence students' initial expected abilities and uncertainty in these initial expectations.

In this general setting, I solve for the parameters which maximize the full log-likelihood using a nested EM approach (James, 2012). The "outer" EM component deals with unobserved student types and allows the full likelihood to be additively separated into choice and grade contributions. The "inner" EM component addresses the $K$-dimensional integration over unobserved abilities. The inner EM is analogous to the EM algorithm used in the simplified case of homogenous distributions of unobserved ability described in Section 5.3.

The E step of the outer EM algorithm involves estimating the probability that each student is each possible type conditional on observed data and guesses for parameter values. Intuitively, these type probabilities are derived by comparing the likelihood of observing individual $i$'s actual grades and choices if she were type $\tau$ to the likelihood of observing her grades and choices if she were other types. Denote the $m$ iteration estimates of these probabilities by $q_{i\tau}^m$. The exact formula for estimating $q_{i\tau}^m$ is given in Appendix A.

These conditional type probabilities are then used to form the following expected log-likelihood function conditional on type probabilities $q_{i\tau}^m$:

$$\tilde{Q}\left(\Theta \,|\, \Theta^m\right) = \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \left\{ \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \Theta\right) \left[ \int_{\alpha_i} \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma_\tau^m, \Delta_\tau^m\right) d\alpha_i \right] \right\} \tag{22}$$

Notice this function can be additively separated into:

$$\begin{aligned} \tilde{Q}\left(\Theta \,|\, \Theta^m\right) &= \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \Theta\right) \\ &\quad + \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \int_{\alpha_i} \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma_\tau^m, \Delta_\tau^m\right) d\alpha_i \end{aligned} \tag{23}$$

The M step of the outer EM algorithm involves maximizing $\tilde{Q}\left(\Theta \,|\, \Theta^m\right)$ over parameters $\Theta = \{\gamma_\tau, \Delta_\tau, \theta_k, \sigma_{kl}, \omega\}$ taking $q_{i\tau}^m$ as given. While the log operator can be used to greatly simplify the choice contribution to the expected likelihood, the $K$-dimensional integration over unobserved abilities still makes it challenging to numerically maximize $\tilde{Q}\left(\Theta \,|\, \Theta^m\right)$. Including the inner EM component addresses this issue.

To include the inner EM component, the E step is expanded to include estimates of type specific probability distributions for each individual's unobserved ability vector $\alpha_i$ conditional on all earned grades. Similar to the case of homogenous distributions of unobserved ability, the distributions are estimated by updating type specific prior distributions $\alpha_i \sim N(\gamma_\tau, \Delta_\tau)$ with signals from all grades earned by individual $i$ to form individual-type specific posterior distributions $N(\xi^m_{i\tau}, \Upsilon^m_{i\tau})$. These estimated type specific distributions for $\alpha_i$ are then combined with conditional type probabilities $q^m_{i\tau}$ to form the following expected log-likelihood function conditional on estimates of ability distributions and type probabilities:

$$\begin{aligned} Q_{het}(\Theta \mid \Theta^m) &= \sum_{i=1}^{N}\sum_{\tau=1}^{\mathscr{T}} q^m_{i\tau} \ln \mathscr{L}^c_i(\mathbf{d_i} \mid \tau, \gamma_\tau, \theta_k, \sigma_{kl}, \Delta_\tau, \omega) \\ &\quad + \sum_{i=1}^{N}\sum_{\tau=1}^{\mathscr{T}} q^m_{i\tau} \int_{\alpha_i} \ln \mathscr{L}^g_i(\mathbf{g_i} \mid \alpha_i, \theta_k, \sigma_{kl})\, f(\alpha_i \mid \xi^m_{i\tau}, \Upsilon^m_{i\tau})\, d\alpha_i \end{aligned} \tag{24}$$

Once again, the M step entails maximizing $Q_{het}(\Theta \mid \Theta^m)$ over parameters $\Theta = \{\gamma_\tau, \Delta_\tau, \theta_k, \sigma_{kl}, \omega\}$ taking $q^m_{i\tau}$, $\xi^m_{i\tau}$, and $\Upsilon^m_{i\tau}$ as given.

With the exception of choice parameters $\omega$, all components of the solution to $\text{argmax}_\Theta Q_{het}(\Theta \mid \Theta^m)$ have closed form expressions with relatively intuitive interpretations. The solutions are similar to those in the case with no unobserved heterogeneity with the added complication that the distribution parameters characterizing terminal beliefs $\xi_{i\tau}$ and $\Upsilon_{i\tau}$ are now type specific. Iteration $m$ estimates of $\theta_k$ come from field specific regressions of $g_{itklc} - \tilde{\alpha}^m_{ik}$ on $X_{itkl}$ where $\tilde{\alpha}^m_{ik} = \sum_{\tau=1}^{\mathscr{T}} q^m_{i\tau}\xi^m_{i\tau}$ is a weighted average of type specific best guesses for ability. Iteration $m$ estimates of $\sigma^2_{kl}$ come from residuals of this regression with adjustments which now account for the use of $\tilde{\alpha}^m_{ik}$ instead of $\alpha_{ik}$. Iteration $m$ estimates of $\gamma_\tau$ come from a weighted average of $\xi^m_{i\tau}$, which uses $q^m_{i\tau}$ as weights to give additional weight to students who are more likely to be type $\tau$. Finally, iteration $m$ estimates of $\Delta_\tau$ come from the analogous weighted variance-covariance in the population distribution of $\xi^m_{i\tau}$, once again using $\Upsilon^m_{i\tau}$ to account for the additional noise generated by using $\xi^m_{i\tau}$ rather than $\alpha_i$. Once again, additional details and exact formulas are presented in Appendix A.

As before, the algorithm is repeated using the solution to $\text{argmax}_\Theta Q_{het}(\Theta \mid \Theta^m)$ as iteration $m+1$ parameter estimates, and final estimates are obtained when the algorithm converges such that $\Theta^\star = \text{argmax}_\Theta Q_{het}(\Theta \mid \Theta^\star)$. Once again, convergence means that the estimates of model parameters which are consistent with $\xi^m_{i\tau}$ and $\Upsilon^m_{i\tau}$ characterizing type specific terminal beliefs and $q_{i\tau}$ characterizing type probabilities are the same as the values of model parameters which imply that $\xi^m_{i\tau}$ and $\Upsilon^m_{i\tau}$ describe type specific terminal beliefs and

$q_{i\tau}$ describe type probabilities.

# 6 Results

In this section, I use estimates of my correlated learning model to compare expected grades and signal quality across academic fields.[16] My main finding is that grades in natural science, engineering, and economics—henceforth, SEE—are lower and more informative than grades in humanities and social sciences—henceforth, HuSS. I begin by presenting results which assume there is no unobserved heterogeneity in the distribution parameters $\gamma$ and $\Delta$. I later relax this assumption by allowing for type specific unobserved heterogeneity in $\gamma$ and $\Delta$ with two unobserved types.[17]

As I discussed previously, information parameters are sensitive to which variables are included as observed covariates in the grade production function. To address sensitivity to this empirical choice, I report results for four different specifications: specification 1 only controls for whether the course is introductory or advanced; specification 2 also includes a quadratic for number of courses taken in this field previously and dummy variables for contemporaneous course load; specification 3 adds race and gender dummy variables; and specification 4 adds math and verbal SAT scores and five measures of the student's application quality.[18] Results which allow for unobserved heterogeneity in $\gamma$ and $\Delta$ also address sensitivity to this choice by including unobserved student type in student information sets. Although the unexplained variance in grades clearly decreases when information sets are richer, the main finding that grades in SEE are lower and more informative than grades in HuSS holds across specifications.

Furthermore, as discussed previously, this analysis can be quite sensitive to how individual courses are aggregated into academic fields. To address sensitivity to this choice, I report results using various definitions of fields. First, I use a more aggregated definition with only two fields: SEE and HuSS. Second, I use a less aggregated definition which disaggregates humanities into arts, foreign languages, and other humanities and disaggregates natural

[16]Point estimates and bootstrapped standard errors (100 iterations) for structural grade production parameters for main specifications are reported in an Online Appendix.

[17]Supplemental analyses available upon request suggest that allowing for more than two unobserved types (and including the richest set of controls) explains almost all across-student variance in grades for one (non-trivial) unobserved student type. As such, if student information sets include the richest controls and unobserved discrete types with more than two types, then some students are fully informed about their academic abilities before enrolling at Duke implying there is no need for these students to learn about their academic abilities through grades.

[18]Admissions officers at Duke grade applications along five dimensions: academic achievement, high school curriculum difficulty, essay quality, personal qualities, and letters of recommendation. Grading scale is from 1 to 5; average grade is 3.88, standard deviation is .81.

science into physical science and life science. Results with the least aggregated definition reveal some interesting heterogeneity: grades in the foreign languages subfield of humanities are comparable in informativeness to grades in science, engineering, and economics; however, aside from that exception, the main finding that SEE grades are more informative than HuSS grades holds across specifications.

To conclude, I provide suggestive evidence on the role of grade compression in impeding information diffusion by reporting results which remove courses with more compressed grade distributions. Removing courses with more compressed grade distributions substantially increases estimates of the informativeness of all grades; however, it does not reduce differences in the relative informativeness of SEE and HuSS grades. This suggests that grade compression impedes information diffusion but cannot explain why SEE grades are more informative than HuSS grades.

## 6.1 Homogenous Ability Distributions

### 6.1.1 Expected Grades by Academic Field and Course Level

While it is not the main objective of this paper, estimates of model parameters can also be used to compare expected grades across fields for a representative student. Table 1 documented large differences in grade distributions across fields; however, these differences could reflect both differences in grading practices and differences in the observed characteristics and unobserved abilities of students who choose courses in each field. Under the identifying assumption that students are not sorting into courses within fields based on time and course specific idiosyncratic grading factors, estimates of the model parameters $\theta_{kl}$ and $\gamma_k$ can be used to construct expected grades for a representative student which are not confounded by non-random selection on observed characteristics or unobserved beliefs about abilities.

Table 3 reports expected grades across academic fields and course levels for a representative student for four possible specifications.[19] The results imply a student with specified characteristics expects to earn substantially higher grades in humanities and social science (HuSS) courses relative to science, engineering, and economics (SEE) courses. In introductory courses, the difference between harshest grading economics and most lenient humanities ranges from .41 to .50 grade points.[20] In advanced courses, the difference between harshest grading science and most lenient humanities ranges from .32 to .43 grade points. This suggests that differences in grade distributions reported in Table 1 cannot be explained by

[19] When covariates are included, representative student is white, male, taking four classes, has no prior experience for introductory courses and five courses of own category experience for advanced courses (0 and 5 are the respective modes), and has average SAT scores and application quality measures.

[20] For reference, this is greater than the difference between a B+ and an A- (.4 grade points).

sorting on observed characteristics or unobserved beliefs about abilities.

### 6.1.2 Decomposition of Variance in Grades

To begin the analysis of the signal quality of grades across academic fields, Table 4 reports variances in grades across academic fields and course levels, a decomposition of these variances into variance from observed covariates, variance from unobserved ability, and variance from idiosyncratic grading noise, and estimates of the correlation matrix corresponding to the variance-covariance matrix $\Delta$.[21] These results are produced using estimates of the richest specification of my correlated learning model, which includes controls for experience, course load, gender, race, SAT scores, and application quality measures. Because the correlated learning model controls for non-random selection on observed characteristics and unobserved beliefs about abilities, the results represent variance in the entire student population rather than variance in a selected population.

The results show there is substantially higher variance in SEE grades relative to HuSS grades for both introductory and advanced courses. For introductory courses, the most variable field is science, which has an interquartile range of 1.09 grade points, and the least variable field is social science, which has an interquartile range of .80 grade points. For advanced courses, the most variable field is once again science, which has an interquartile range of 1.05 grade points, and the least variable field is humanities, which has an interquartile range of .72 grade points. This mirrors the finding in Table 1 that HuSS grades are more compressed than SEE grades and implies that greater compression of HuSS grades cannot be explained by selection on observed covariates or unobserved beliefs about abilities.

The decomposition analysis shows idiosyncratic grading noise contributes the majority of variance (52.0% - 74.6%), unobserved ability contributes the second most variance (18.0% - 31.5%), and observed covariates contribute the least variance (7.5% - 16.4%). This has several important implications: First, it shows that even with a rich set of controlling covariates, there is substantial information to be revealed by grades. The interquartile ranges for field specific unobserved abilities are .34 grade points for humanities, .36 grade points for social

[21] To facilitate a clear comparison across fields and levels, the variance figures exclude variance in grades which arises because students have different experience or course loads. Let $\tilde{X}_{itkl}$ represent observed covariates other than experience and course load, and let $X_{itkl} \backslash \tilde{X}_{itkl}$ represent these experience and course load covariates. If $X_{itkl} \backslash \tilde{X}_{itkl}$ are fixed, variance in $g_{itklc}$ can be additively decomposed as:

$$\text{Var}\left(g_{itklc}\right) = \text{Var}\left(\tilde{X}_{itkl}\tilde{\theta}_k\right) + \Delta\left(k,k\right) + \sigma_{kl}^2$$

where $\tilde{\theta}_k$ are the coefficients corresponding to $\tilde{X}_{itkl}$. The first term represents variance due to differences in observed covariates, the second term represents variance due to differences in unobserved abilities, and the third term represents variance due to idiosyncratic grading noise.

science, .59 grade points for science, .58 grade points for engineering, and .49 grade points for economics. This implies that an incoming student with a very rich information set still believes that with 50% probability her average grades in science courses will be outside a range of .59 grade points. This is substantial uncertainty which makes it difficult to make informed academic decisions about which major to choose and which courses to enroll in. This highlights the importance of grades as a mechanism for revealing this missing information. Second, it shows that even a rich set of covariates do a poor job of capturing persistent heterogeneity across students. As such, research which aims to assess the effects of academic choices on later outcomes should be careful to control for confounding selection on unobserved factors.

More importantly, the decomposition analysis also shows that variation in HuSS grades is more explained by grading noise and less explained by variation in observed characteristics or unobserved abilities relative to variation in SEE grades. In introductory courses, 71.5% - 74.6% of variation in HuSS grades is due to grading noise versus 53.1% - 66.7% for SEE grades; in advanced courses, 68% of variation in HuSS grades is due to grading noise versus 52% - 60.8% for SEE grades. This implies that although SEE grades are more dispersed, a specific student generally experiences only 52% - 66.7% of this variance throughout her time at Duke. Conversely, a specific student generally experiences 68% - 74.6% of the variance in HuSS grades throughout her time at Duke. This implies that a single SEE grade gives students more information about their relative abilities in SEE courses than HuSS grades do for abilities in HuSS grades.

While the decomposition analysis gives clear evidence on the own-field signal quality of grades, it does not account for potential information spillovers across fields. To see how grade signals in one field inform beliefs about abilities in other fields, the bottom panel of Table 4 reports estimates of the correlation matrix corresponding to the variance-covariance matrix $\Delta$.[22] The estimates show unobserved abilities are highly correlated across fields. The largest correlations are within HuSS and SEE, but substantial correlations also exist between HuSS and SEE abilities. Recall that these estimates are for the specification with the richest set of controls, implying that these correlations exist even after controlling for SAT scores, application quality measures, and other observed characteristics. The large correlations imply there will be substantial information spillovers across fields and demonstrate the need to account for these spillovers when assessing information quality.

[22]Specifically, it reports estimates of:

$$\mathscr{R} = \left(\text{diag}\left(\Delta\right)\right)^{-1/2} \Delta \left(\text{diag}\left(\Delta\right)\right)^{-1/2}$$

Bootstrapped standard errors (100 iterations) are available upon request.

### 6.1.3 Signal Quality by Academic Field

Table 4 showed that HuSS grades are noisier signals of own-field abilities than SEE grades. Additionally, it showed that unobserved abilities are highly correlated across fields, implying there will be substantial information spillovers across fields. To assess the signal quality of grades across academic fields in a way that incorporates information spillovers, Table 5 reports initial uncertainty in beliefs about abilities by field and the reductions in uncertainty which occur in each field after earning one grade in a specific field. Initial uncertainty is measured by the diagonal elements of $\Delta$, which represent each individual's variance in initial beliefs about $\alpha_i$. Reductions in uncertainty are measured by comparing variances in initial beliefs to variances in beliefs after one field $k$ grade signal is received. These variances in revised beliefs are given by the diagonal elements of $\delta\left(e_k\right)$ where:

$$\delta\left(e_k\right) = \left(\Delta^{-1} + \mathrm{diag}\left(e_k\right)\Phi_1^{-1}\right)^{-1} \tag{25}$$

where $\Phi_1$ is a $K \times K$ diagonal matrix containing noise variances in introductory classes:

$$\Phi_1\left(k,\, k'\right) = \begin{cases} \sigma_{k1}^2 & k = k' \\ 0 & k \neq k' \end{cases}$$

and $\mathrm{diag}\left(e_k\right)$ is a sparse $K \times K$ matrix defined by:

$$\mathrm{diag}\left(e_k\right)\left(k_1,\, k_2\right) = \begin{cases} 1 & k = k_1 = k_2 \\ 0 & k \neq k_1 \text{ or } k \neq k_2 \end{cases}$$

Notice variance in revised beliefs depend on signal parameters $\Delta$ and $\sigma_{kl}$ but do not depend on grades earned or individual characteristics. As such, the reduction from $\mathrm{diag}\left(\Delta\right)$ to $\mathrm{diag}\left(\delta\left(e_k\right)\right)$ provides an intuitive measure of field $k$ information quality which is relevant for the entire population of students.

The variance decomposition in Table 4 shows there is substantial information left to be revealed by grades; results in Table 5 suggest SEE grades reveal this missing information more efficiently than HuSS grades. Engineering consistently offers the largest average reduction in uncertainty while humanities offers the smallest average reduction. Moreover, engineering signals are so informative and abilities are so correlated across fields that engineering grades almost always reduce uncertainty in beliefs about abilities in other categories more than grades from these categories. One can see this by comparing the elements of the engineering column to the diagonal elements of each row in specification 4: One engineering grade reduces uncertainty in beliefs about humanities ability by 24% while one humanities grade only reduces uncertainty in beliefs about humanities ability by 19%; one engineering grade reduces uncertainty in beliefs about social science ability by 20% compared to 23%; for science the reduction is 36% relative to 33%; and for economics the reduction is 33% relative to 26%.[23] This paradoxical result shows the importance of allowing for spillovers across fields; these spillovers are so powerful that students can learn more about their unobserved abilities in one field by taking a course in another field with more precise grade signals.

#### 6.1.4 Alternative Field Definitions

As mentioned previously, my results can be quite sensitive to how individual courses are aggregated into fields. This is because the method relies on decomposing unexplained variance in grades in a specific field into within-student and across-student contributions. If some field is really several distinct subfields which reward different talents and skills, this field will have high within-student variance not because its grades are noisy signals but because students have different abilities in the different subfields.

To assess sensitivity to how individual courses are aggregated into fields, Tables 6 and 7 reproduce the main results of Table 5 under alternative field definitions. Table 6 uses a more aggregated definition with only two fields: HuSS and SEE. Table 7 uses a finer definition with eight fields. With this definition, the largest field of humanities is disaggregated into three fields: general humanities (6,547 observations), foreign languages (3,740 observations), and visual and performing arts (1,854 observations); and the second largest field of natural sciences is disaggregated into two fields: life sciences (3,312 observations) and physical sciences (7,326 observations). For these tables, I report results only for the sparsest and richest specifications; results for the intermediate specifications are available upon request.

The results in Table 6 mirror the main finding from Table 5 that SEE grades are more informative than HuSS grades. With full controls, a single SEE grade reduces HuSS uncertainty by 16% and SEE uncertainty by 31%, while a single HuSS grade reduces HuSS uncertainty by 18% and SEE uncertainty by 10%. This illustrates that my main finding that SEE grades are more informative than HuSS grades holds under a more aggregated definition of fields.

The results in Table 7 tell a similar story but with slightly more nuance. The general humanities and visual and performing arts subfields remain less informative (average uncertainty reductions of 19% and 12% respectively with full controls), and life sciences and

[23] In Table 5, (*) indicates cases in which non-field signals are statistically more informative than own-field signals at 5% significance. For example: in Specification 1, the 35% reduction in humanities uncertainty after receiving one engineering grade is statistically greater than the 26% reduction from one humanities grade.

physical sciences remain more informative (average uncertainty reductions of 34% and 35% respectively with full controls); however, the foreign languages subfield looks closer to a SEE field in terms of signal quality. In the specification with full controls, one foreign languages grade yields an average uncertainty reduction of 35%, including a 54% reduction in foreign languages uncertainty, a 34% reduction in general humanities uncertainty, and a 39% reduction in engineering uncertainty. This is comparable to the signal quality of life sciences, physical sciences, and engineering grades (average uncertainty reductions of 34%, 35% and 37% respectively) and is substantially higher quality than economics grades (average uncertainty reduction of 29%).

The high signal quality of foreign language grades cannot be explained by lower grade compression as compression is only slightly lower in foreign languages than in other HuSS fields: 91.6% of foreign language grades are B- or above compared to 94.7% for non-foreign language HuSS grades, and 35.2% of foreign language grades are A compared to 35.4% of non-foreign language HuSS grades. Instead, the high signal quality of foreign language grades likely reflects the fact that student performance in these courses is typically evaluated with short-answer format questions with well-defined correct answers which limits the role of idiosyncratic grader preferences.

With the exception of foreign languages, the results in Table 7 show that SEE grades are more informative than HuSS grades. This implies that my main finding generally holds under a finer definition of fields as well. While the aggregation of courses into fields will always be somewhat arbitrary and one cannot repeat the analysis for all possible field definitions, I find it encouraging that the main result appears to generally hold under different definitions of fields.

## 6.2 Heterogeneous Ability Distributions

In this subsection, I present results which allow for type specific unobserved heterogeneity in the distribution parameters $\gamma_\tau$ and $\Delta_\tau$. These distribution parameters characterize initial prior beliefs about the unobserved component of grade production. In a Bayesian setting, initial prior beliefs inform all subsequent beliefs, making the information revelation process very sensitive to the distribution parameters $\gamma_\tau$ and $\Delta_\tau$. For example, larger values in $\gamma_\tau$ imply students initially expect to perform better than their observed characteristics suggest; grades which do not meet these high expectations result in downward revisions in beliefs. Additionally, smaller diagonal values in $\Delta_\tau$ imply students have more certain initial beliefs and will not update their beliefs as much when they receive grades. With two unobserved types I find one type has much lower initial expectations and higher uncertainty in these expectations. These students come from less advantaged backgrounds and experience fewer information spillovers across academic fields than their classmates.

### 6.2.1 Expected Grades and Type Characteristics

Panel A of Table 8 reports expected grades across academic field and levels conditional on observed covariates for each unobserved type.[24] Panel B of Table 8 reports type frequencies, average course choices by type, and average baseline characteristics by type. For brevity, I only report results for specification 1 and specification 4.[25] As in Table 3, students of both unobserved types expect to earn significantly lower grades in SEE courses relative to HuSS courses. This shows the differences in average grades between SEE and HuSS courses are not driven by sorting on initial beliefs about unobserved abilities.

The results also show 35-40% of students—labeled as Type 1—arrive at Duke with substantially lower expected grades in all five academic fields. The largest difference is in science courses, where Type 2 students expect to perform .63 - .79 grade points better than Type 1 students. Type 1 students also have larger differences between expected grades in SEE courses and expected grades in HuSS courses. In introductory courses, the difference between harshest grading economics and most lenient humanities is .57 - .62 grade points for Type 1 students relative to .33 - .39 grade points for Type 2 students. This shows Type 2 students initially expect to possess both an absolute advantage in all fields and a relative advantage in SEE courses when compared to Type 1 students.

To analyze the relationship between initial expectations and course choices, Panel B of Table 8 reports the average number of courses completed in each field for each unobserved type. Results show Type 2 students complete an average of 1.2 - 1.6 more science courses, .3 - .7 more engineering courses, and .4 - .6 more economics courses than Type 1 students. This suggests higher expected grades and an expected relative advantage in SEE courses is associated with taking more science and engineering courses. This is consistent with existing literature, which concludes expected grades largely determine whether a student pursues science or engineering (Arcidiacono, 2004; Arcidiacono, Aucejo, and Spenner, 2012; Stinebrickner and Stinebrickner, 2014a).

To understand what baseline covariates are associated with differences in initial expectations, Panel B of Table 8 also summarizes the demographic characteristics of students of each unobserved type. Results show Type 2 students are more likely to have attended private high schools (instead of public or religious high schools), are more likely to have at

[24] When covariates are included, representative student is white, male, taking four classes, has no prior experience for introductory courses and five courses of own field experience for advanced courses (0 and 5 are the respective modes), and has average SAT scores and application quality measures.

[25] Results for specifications 2 and 3 are available upon request.

least one parent with a professional degree (instead of a Bachelor's degree or some graduate coursework), and report more AP scores on average than Type 1 students.[26] This suggests higher socioeconomic status and more exposure to college level coursework are associated with higher initial expectations and an expected relative advantage in SEE courses.

### 6.2.2 Signal Quality by Academic Field

To understand how these differences in initial expectations relate to ability revelation, Tables 9 and 10 conduct the ability revelation analysis of Tables 5 separately for each unobserved type. This analysis reports the initial uncertainty in beliefs about unobserved abilities as well as the reductions in uncertainty following each possible grade signal. Table 9 reports results for the sparsest specification 1, while Table 10 reports results for the richest specification 4.[27]

The initial uncertainty results suggest low expectation type 1 students have substantially higher uncertainty in their initial expectations. In the richest specification, interquartile ranges for beliefs range from .42 to .64 grade points for type 1 students and from .14 to .35 grade points for type 2 students. One reason why low expectation students might also have high uncertainty is noise in the admissions process. Intuitively, lower ability students will only be admitted if they produce unusually strong applications, which is more likely to occur if there is more uncertainty about their abilities. Conversely, higher ability students will only be rejected if they produce unusually weak applications, which is more likely to occur if there is more uncertainty about their abilities. Appendix B formalizes this intuition with a simple model.

Alternatively, a negative relationship between expectations and uncertainty may exist because certain pre-college experiences both increase expectations and reduce uncertainty. Panel B of Table 8 shows high expectation and low uncertainty Type 2 students have higher socio-economic status and more exposure to college coursework during high school. This suggests having more educated parents and taking more AP courses may increase expected performance and reduce uncertainty about performance.

The reduction percentage results suggest there is substantial heterogeneity in how students process information. Low expectation type 1 students experience modest information spillovers between humanities and social science, social science and economics, and natu-

[26]AP courses are nationally standardized college level courses taken in high school. After taking an AP course, students have the option to take a standardized test of the material. Scores on these tests may be used to bolster college applications and earn college credit for the coursework. These are typically the most advanced courses taken by high school students, and they are more widely available at high schools with many high achieving students.

[27]Results for specifications 2 and 3 are available upon request

ral science and engineering, while high expectation type 2 students experience substantial spillovers between all fields. For type 1 students, modest spillovers imply own-field signals are always a much more efficient way to reveal abilities: In the richest specification, a humanities signal decreases humanities uncertainty by 26% versus 2% - 16% from other signals; social science grades yield an own-field reduction of 32% compared to 4% - 14%; natural science yields a 29% reduction relative to 2% - 31%; engineering yields a 42% reduction versus 7% - 22%; and economics yields a 27% reduction compared to 2% - 11%. Conversely, for type 2 students, spillovers are so strong that precise SEE signals always reveal more about HuSS abilities than HuSS signals. In the richest specification, an engineering grade reduces humanities uncertainty by 17% and social science uncertainty by 14%, whereas humanities and social science grades reduce these uncertainties by 4% and 6% respectively. Stars indicate that cross-field reductions are significantly greater than own-field reductions at 5% significance.

This finding paints an interesting picture of how students at Duke process grade information. A slight majority of Duke students expect to perform well and are fairly confident in these expectations. These students are mostly uncertain about factors which are general to the college experience such as their ability to adjust to new instruction formats or their ability to develop effective study habits at Duke. These students are more likely to have attended a private high school, are more likely to have highly educated parents, and have completed more advanced placement courses prior to arriving at Duke. These pre-college experiences may have revealed their aptitudes for different academic fields leaving only uncertainty about general aspects of the transition to college. For these students, precise grade signals from SEE classes reveal this general information most efficiently.

Slightly less than half of Duke students expect to perform worse but are more uncertain about these expectations. For these students, a large share of the uncertainty comes from factors which are specific to academic fields. These students must take courses in each academic field to reveal this field specific information.

## 6.3 Grade compression and signal quality

As discussed previously, compressing grades into a smaller number of discrete grade categories reduces an instructor's capacity to distinguish student performances and thus may reduce the signal quality of grades. Indeed, results discussed previously show that fields with more compressed grade distributions also have less informative grades. To provide additional suggestive evidence on the relationship between grade compression and signal quality, I estimate the correlated learning model on a restricted sample which excludes courses with severely compressed grade distributions. This allows me to analyze how much signal quality improves when compressed courses are removed and whether restricting to courses with less compressed distributions reduces the differences in signal quality across academic fields.

Table 11 contains details on the sample restriction. Before excluding severely compressed courses, I remove courses with too few observations to conclude whether they are severely compressed.[28] Removing these courses has a minor effect on introductory courses but requires dropping 20.0% of all advanced observations. Courses with few observations are more common in advanced courses both because class sizes are generally smaller in advanced courses and because many advanced courses only appear in one semester of my panel. After restricting to courses with at least 10 observations, my main restriction is to remove courses in which at least half of students earn an A- or above. This removes 57.1% of the remaining introductory HuSS observations, 62.2% of the remaining advanced HuSS observations, 18.4% of the remaining introductory SEE observations, and 33.0% of the remaining advanced SEE observations. This demonstrates that in addition to having more compressed aggregate grade distributions, HuSS fields also have disproportionately many severely compressed courses.

Results in Table 12 examine how excluding severely compressed courses affects estimates of the signal quality of grades. For simplicity, I report results which use the most aggregated definition of fields and which do not allow for type specific unobserved heterogeneity in $\gamma$ and $\Delta$. These results are directly comparable to the ones in Table 5 which estimate the same specifications on the unrestricted sample.

First, by comparing Tables 12 and 5, notice that estimates of initial uncertainty increase when small and compressed courses are excluded. For the specifications with full controls, excluding these courses increases the interquartile range for initial HuSS beliefs from .33 grade points to .39 grade points and increases the interquartile range for initial SEE beliefs from .56 to .62. These differences arise because removing compressed courses leads to more variance in average grades across students in the population. The rational expectations assumption asserts that initial beliefs about abilities are characterized by the population distribution of abilities; as such, increasing the variance in average grades increases initial uncertainty in beliefs.

Second, notice that removing small and compressed courses implies that individual courses yield larger reductions in uncertainty. For the specification with full controls, a single HuSS grade reduces HuSS uncertainty by 23% on the restricted sample relative to 18% on the full sample. Similarly, a single SEE grade reduces SEE uncertainty by 39% on

[28]Courses are defined by course number (e.g. ECON 101) and thus may appear in multiple semesters possibly being taught by different instructors. My criteria for "too few observations" is having fewer than 10 observations across semesters.

the restricted sample relative to 31% on the full sample. This provides suggestive evidence on the role of grade compression in reducing the signal quality of grades. Removing the most compressed courses significantly increases estimates of the reduction in uncertainty provided by a single grade and thus significantly increases estimates of the signal quality of grades.

Finally, note that while removing small and compressed courses does increase estimates of the signal quality of grades, it does not close the gap in information quality between HuSS and SEE grades. In the full sample results in Table 5, the average reduction in uncertainty from one SEE grade is 1.57 - 1.71 times greater than the average reduction in uncertainty from one HuSS grade. In the restricted sample results in Table 12, this ratio is virtually unchanged: the average reduction in uncertainty from one SEE grade is 1.52 - 1.74 times greater than the average reduction in uncertainty from one HuSS grade. This shows that restricting to a set of courses with less grade compression does not reduce the gap in information quality between HuSS and SEE grades.

This analysis has important limitations and thus should be interpreted as purely suggestive. Still, the results do suggest that although grade compression does reduce the signal quality of grades, it cannot explain the differences in signal quality between HuSS and SEE grades. This implies that the differences in signal quality between HuSS and SEE grades are primarily driven by the larger role of idiosyncratic grader judgement in the evaluation of student performances in HuSS courses.

# 7 Conclusion

This paper compares the signal quality of grades across academic fields by estimating a correlated learning model using transcript data from Duke University. The model allows for both "own-field" signal quality and "spillover" signal quality and controls for selection on unobserved beliefs about abilities. Estimates from specifications with no unobserved heterogeneity in distribution parameters suggest grades in science, engineering, and economics (SEE) courses are substantially more informative than grades in humanities and social science (HuSS) courses. In some of these specifications, cross-field spillovers are so powerful that grades in engineering courses reveal abilities in all other fields more efficiently than grades in these fields.

Estimates from specifications which allow for unobserved heterogeneity in distribution parameters tell a more nuanced story: Even with rich controls, some students initially expect to earn significantly lower grades but are less certain in these expectations. These students come from less advantaged backgrounds on average, suggesting that pre-college experiences may affect both initial expectations and uncertainty in these expectations. Furthermore, these students experience only modest information spillovers across fields, implying that they must complete courses in all fields to reveal all field specific abilities.

Finally, a supplemental analysis shows that while grade compression reduces the signal quality of all grades, it cannot explain differences in the signal quality of HuSS and SEE grades. This suggests that differences in signal quality across fields primarily result from the larger role of grader judgement in the evaluation of student performances in HuSS courses.

My analysis suggests there are a number of policies universities could consider to improve information diffusion. First, additional graders could be assigned to HuSS courses so that multiple graders can evaluate the same responses to limit the role of individual graders' idiosyncratic tastes. Second grading policies could be introduced to encourage instructors to reduce compression and employ more of the grade distribution. This may not reduce the signal quality gap between HuSS and SEE courses, but it should increase the signal quality of all grades. Finally, curricula could be modified to reduce the number of poor signal quality courses students are obligated to take. At Duke University, students must take 13-15 humanities or social science courses but only 6-8 natural science, engineering or economics courses to satisfy graduation requirements.[29] While there are many justifications for requiring certain courses, my results suggest these requirements oppose information diffusion.

One concern with my empirical analysis is it reports results which are only relevant to Duke University. This is only a concern if universities are heterogeneous. Moreover, if universities are heterogeneous, estimates which represent a large set of schools are not relevant to any specific school. Most policy decisions regarding curricula are made at the university level (or below). As such, the empirical results which should inform these policy decisions should come from analyzing the population of students at one university. This paper provides a theoretical framework and methodology for comparing the signal quality of grades across academic fields at a university and presents results of this analysis for one specific university. Policy decisions at other institutions should be informed by analyses which use data from these institutions.

While it is beyond the scope of this paper, the results also have important implications for employer-employee matching. Just as students use grade signals to inform specialization decisions, firms use grade signals to inform hiring decisions. If firms are risk averse (as in Greenwald and Stiglitz, 1990), they will favor students with a large number of SEE grades to those with mostly HuSS grades. This mechanism may partially explain the labor market premium for SEE majors documented in Altonji, Blom, and Meghir (2012). Further research

[29]Requirements are: arts, literatures, and performance (2 courses); civilizations (2 courses); natural sciences (2 courses); quantitative studies (2 courses); social sciences (2 courses, includes economics); cross-cultural inquiry (2 courses); ethical inquiry (2 courses); science, technology and society (2 courses); foreign language (3 courses); writing (3 courses) ("Degree Requirements," 2013).

may investigate how firms value expected ability by field and uncertainty in expectations by field.

# References


[1] Achen, A., Courant, P., 2009. “What are Grades Made Of?” Journal of Economic Perspectives 23(3), 77-92.

[2] Ackerberg, D., 2003. “Advertising, learning, and consumer choice in experience good markets: an empirical examination.” International Economic Review 44(3), 1007-1040.

[3] Allen, J., van der Velden, R., 2001. “Education mismatches versus skill mismatches: Effects on wages, job satisfaction and on-the-job search.” Oxford Economic papers 53(15), 434-452.

[4] Altonji, J. G., 2005. “Employer Learning, Statistical Discrimination and Occupational Attainment.” The American Economic Review Papers and Proceedings 95(2), 112-117.

[5] Altonji, J. G., Blom, E., Meghir, C., 2012. “Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers.” Annual Review of Economics 4, 185-223.

[6] Altonji, J. G., Pierret, C. R., 2001. ”Employer Learning and Statistical Discrimination.” The Quarterly Journal of Economics 116(1), 313-350.

[7] Antonovics, K., Golan, L., 2012. “Experimentation and Job Choice” Journal of Labor Economics 30(2), 333-366.

[8] Arcidiacono, P., 2004. “Ability sorting and the returns to college major.” Journal of Econometrics 121, 343-375.

[9] Arcidiacono, P., Aucejo, E., Maurel, A., Ransom, T., 2016. “College Attrition and the Dynamics of Information Revelation.” Working Paper.

[10] Arcidiacono, P., Aucejo, E., Fang, H., Spenner, K., 2011. ”Does Affirmative Action Lead to Mismatch? A New Test and Evidence.” Quantitative Economics 2(3), 303-333.

[11] Arcidiacono, P., Aucejo, E., Spenner, K., 2012. “What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice.” IZA Journal of Labor Economics 1(5).

[12] Arcidiacono, P., Hotz, V. J., Kang, S., 2012. “Modeling college major choices using elicited measures of expectations and counterfactuals.” Journal of Econometrics 166, 3-16.

[13] Arcidiacono, P., Jones, J. B., ”Finite Mixture Distributions, Sequential Likelihood, and the EM Algorithm.” Econometrica 71(3), 933-946.

[14] Arcidiacono, P., Sieg, H., Sloan, F., 2007. “Living Rationally Under the Volcano? An empirical analysis of heavy drinking and smoking.” International Economic Review 48(1), 37-65.

[15] Avery, C., Hoxby, C., 2013. “The Missing ’One-Offs’: The Hidden Supply of High-Achieving, Low-Income Students.” Brookings Papers on Economic Activity 1, 1-65.

[16] Bernal, R., Keane, M., 2010. “Quasi-structural estimation of a model of childcare choices and child cognitive ability production.” Journal of Econometrics 156, 164-189.

[17] Berry, S., Levinsohn, J., Pakes, A., 1995. “Automobile Prices in Market Equilibrium.” Econometrica 63(4), 841-890.

[18] Bordon, P., Fu, C., 2015. “College-Major Choice to College-Then-Major Choice.” Review of Economic Studies 82(4), 1247-1288.

[19] Bound, J., Lovenheim, M., Turner, S., 2012. “Increasing Time to Baccalaureate Degree in the United States.” Education Finance and Policy 7(4), 375-424.

[20] Bryant, A., Spenner, K., Martin, N., 2006. “The campus life and learning project: A report on the first two college years.” Duke University.

[21] Bryant, A., Spenner, K., Martin, N., 2007. “The campus life and learning project: A report on the college career.” Duke University.

[22] Cai, H., Chen, Y., Fang, H., 2009. “Observational Learning: Evidence from a Randomized Natural Field Experiment.” The American Economic Review 99(3), 864-882.

[23] Chan, W., Hao, L., Suen, W., 2007. “A Signalling Theory of Grade Inflation.” International Economic Review 48(3), 1065-1090.

[24] Crawford, G., Shum, M., 2005. “Uncertainty and learning in pharmaceutical demand.” Econometrica 73(4), 1137-1173.

[25] “Degree Requirements,” 2013. Retrieved from http://www.trinity.duke.edu/undergraduate/academic-policies/degree-requirements

[26] Degroot, M. H., 1970. "Optimal Statistical Decisions." McGraw Hill.

[27] Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society Series B (Methodological) 39(1), 1-38.

[28] Downs, J. S., Loewenstein, G., Wisdom, J., 2009. "Strategies for Promoting Healthier Food Choices." American Economic Review 99(2), 159–64.

[29] Dupas, P., 2011. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya." AEJ: Applied Economics 3(1), 1-36.

[30] Dupas, P., 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." Econometrica 82(1), 197-228.

[31] Farber, H. S., Gibbons, R., 1996. "Learning and Wage Dynamics." Quarterly Journal of Economics 111(4), 1007-1047.

[32] Gibbons, R., Katz, L. F., Lemieux, T., Parent, D., 2005. "Comparative advantage, learning, and sectoral wage determination." Journal of Labor Economics 23(4), 681-724.

[33] Greenwald, B., Stiglitz, J., 1990. "Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior." American Economic Review 80(2), 160-165.

[34] Heckman, J., Singer, B., 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." Econometrica 52, 271-320.

[35] Jalan, J., Somanathan, E., 2008. "The Importance of Being Informed: Experimental Evidence on Demand for Environmental Quality." Journal of Development Economics 87(1), 14–28.

[36] James, J., 2012. "Learning and Occupational Sorting." Federal Reserve Bank of Cleveland Working Paper 12-25.

[37] Johnson, V., 2003. "Grade Inflation: A Crisis in College Education." Springer-Verlag.

[38] Jovanovic, B., 1979. "Job Matching and the Theory of Turnover." Journal of Political Economy 87(5), 972-990.

[39] Jovanovic, B., Nyarko, Y., 1997. "Stepping-stone mobility." Carnegie-Rochester Conference Series on Public Policy 46, 289-325.

[40] Malamud, O., 2010. “Breadth versus depth: The timing of specialization in Higher Education.” Labour 24(4), 359-390.

[41] Malamud, O., 2011. “Discovering One’s Talent: Learning from Academic Specialization.” Industrial and Labor Relations Review 62(2), 375-405.

[42] Mansfield, H., 2001. “Grade Inflation: It’s Time to Face the Facts.” The Chronicle Review.

[43] Manski, C. F., 1992. “Parental income and college opportunity.” Democratic Study Center Report, Washington, DC.

[44] McCall, B. P., “Occupational matching: A test of sorts.” The Journal of Political Economy 98(1), 45-69.

[45] Miller, R., 1984. “Job matching and occupational choice.” The Journal of Political Economy 92(6), 1086-1120.

[46] Neal, D., 1999. “The complexity of job mobility among young men.” Journal of Labor Economics 17(2), 237-261.

[47] Nordin, M., Persson, I., Rooth, D., 2010. “Education–occupation mismatch: Is there an income penalty?” Economics of Education Review 29 (6), 1047–1059.

[48] Papageorgiou, T., 2014. “Learning Your Comparative Advantages.” Review of Economic Studies 81(3), 1263-1295.

[49] Pavan, R., 2011. “Career Choice and Wage Growth.” Journal of Labor Economics 29(3), 549-587.

[50] Shaw, K. L., 1987. “Occupational change, employer change, and the transferability of skills.” Southern Economic Journal 53(3), 702-719.

[51] Stinebrickner, T., Stinebrickner, R., 2012. “Learning about Academic Ability and the College Dropout Decision.” Journal of Labor Economics 30(4), 707-748.

[52] Stinebrickner, T., Stinebrickner, R., 2014a. “A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout.” Review of Economic Studies 81, 426-472.

[53] Stinebrickner, T., Stinebrickner, R., 2014b. “Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model.” Journal of Labor Economics 32(3), 601-644.

[54] Wikstrom, C., Wikstrom, M., 2005. "Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools." Economics of Education Review 24, 309-322.

[55] Zafar, B., 2011. "How Do College Students Form Expectations?" Journal of Labor Economics 29(2), 301-348.

# Appendix A: Estimation Details

This appendix contains explicit solutions to $\text{argmax}_{\Theta} Q_{het}(\Theta \mid \Theta^m)$ and expressions for the distribution variables $q_{i\tau}^m$, $\xi_{i\tau}^m$, and $\Upsilon_{i\tau}^m$ used to form the expected likelihood function in the iteration $m$ E step.[30] Formulas are presented in the general case in which there is type specific unobserved heterogeneity in the distribution parameters $\gamma_\tau$ and $\Delta_\tau$.

### Distribution variables used to form expected likelihood

**Unobserved Abilities** $\xi_{i\tau}^{m+1}$ and $\Upsilon_{i\tau}^{m+1}$ characterize the distribution of each individual's ability vector $\alpha_i$ conditional on student type, iteration $m$ parameter estimates, and all observed grades. These are derived by updating type specific initial prior distributions with all grade signals received by student $i$. Degroot (1970) shows the resulting posterior distribution is $N\left(\xi_{i\tau}^{m+1}, \Upsilon_{i\tau}^{m+1}\right)$ where:[31]

[30]For clarity, I present formulas which assume the observed sample is randomly drawn from the population. The formulas actually used in my estimation routine include sampling weights to correct for the stratified sampling of the CLL and to partially eliminate non-response bias.

[31] $\bar{z}_i$ is a vector of average grade signals received by individual $i$ given by:

$$\bar{z}_i(k) = \frac{\sum_{t=1}^{T}\sum_{c=1}^{C_{it}}(d_{ict}=k)\left(g_{itklc} - X_{itk}\theta_k^m\right)}{\sum_{t=1}^{T}\sum_{c=1}^{C_{it}}(d_{ict}=k)}$$

$\Phi_l$ is a $K \times K$ matrix defined as,

$$\Phi_l(k, k') = \begin{cases} \sigma_{kl}^2 & k = k' \\ 0 & k \neq k' \end{cases}$$

and $D_{il}$ is a $K \times K$ matrix defined as,

$$D_{il}(k, k') = \begin{cases} n_{ikl} & k = k' \\ 0 & k \neq k' \end{cases}$$

where $n_{ikl}$ is the total number of field $k$ level $l$ courses taken by individual $i$.

$$\xi_{i\tau}^{m+1} = \left( (\Delta_\tau^m)^{-1} + \sum_{l=1}^{L} D_{il} (\Phi_l^m)^{-1} \right)^{-1} \left( (\Delta_\tau^m)^{-1} \gamma_\tau^m + \sum_{l=1}^{L} D_{il} (\Phi_l^m)^{-1} \bar{z}_i \right) \tag{26}$$

$$\Upsilon_{i\tau}^{m+1} = \left( (\Delta_\tau^m)^{-1} + \sum_{l=1}^{L} D_{il} (\Phi_l^m)^{-1} \right)^{-1} \tag{27}$$

**Unobserved Types** $q_{i\tau}^{m+1}$ is the probability student $i$ is type $\tau$ conditional on iteration $m$ parameter estimates and observed grades and choices. These conditional probabilities can be expressed as the joint likelihood that individual $i$ is type $\tau$ and of observing individual $i$'s grades and choices divided by the marginal likelihood of observing individual $i$'s grades and choices.

$$q_{i\tau}^{m+1} = \frac{\pi_\tau^m \mathscr{L}_i^c (\mathbf{d_i} \mid \tau, \gamma_\tau^m, \theta_k^m, \sigma_{kl}^m, \Delta_\tau^m, \omega^m) \left[ \int_\alpha \mathscr{L}_i^g (\mathbf{g_i} \mid \alpha_i, \theta_k^m, \sigma_{kl}^m) f(\alpha_i \mid \gamma_\tau^m, \Delta_\tau^m) \right]}{\sum_{\tau=1}^{\mathscr{T}} \pi_\tau^m \mathscr{L}_i^c (\mathbf{d_i} \mid \tau, \gamma_\tau^m, \theta_k^m, \sigma_{kl}^m, \Delta_\tau^m, \omega^m) \left[ \int_\alpha \mathscr{L}_i^g (\mathbf{g_i} \mid \alpha_i, \theta_k^m, \sigma_{kl}^m) f(\alpha_i \mid \gamma_\tau^m, \Delta_\tau^m) \right]} \tag{28}$$

**Unconditional Type Probabilities** Unconditional type probabilities $\pi_\tau^m$ are updated by averaging conditional type probabilities over the population of students:

$$\pi_\tau^{m+1} = \frac{1}{N} \sum_{i=1}^{N} q_{i\tau}^{m+1} \tag{29}$$

**Solutions to** $\text{argmax}_\Theta Q_{het} (\Theta \mid \Theta^m)$

**Type Specific Ability Distributions** Ability vectors are drawn from the type specific distributions $N(\gamma_\tau, \Delta_\tau)$. These parameters can be directly recovered from individual-type specific distribution parameters $\xi_{i\tau}^m$ and $\Delta_{i\tau}^m$ and individual specific type probabilities $q_{i\tau}^m$. The closed form expressions are:

$$\gamma_\tau^m = \frac{\sum_{i=1}^{N} q_{i\tau}^m \xi_{i\tau}^m}{\sum_{i=1}^{N} q_{i\tau}^m} \tag{30}$$

$$\Delta_\tau^m = \frac{\sum_{i=1}^{N} q_{i\tau}^m \left( \Upsilon_{i\tau}^m + (\xi_{i\tau}^m)(\xi_{i\tau}^m)' \right)}{\sum_{i=1}^{N} q_{i\tau}^m} - \frac{\sum_{i=1}^{N} q_{i\tau}^m (\gamma_\tau^m)(\gamma_\tau^m)'}{\sum_{i=1}^{N} q_{i\tau}^m} \tag{31}$$

**Grade Production Parameters** If $\alpha_i$ were observed, grade production parameters $\theta_k$ and $\sigma_{kl}$ could be estimated by regressing $g_{itklc} - \alpha_{ik}$ on covariates $X_{itkl}$. Although $\alpha_i$ is not observed, individual-type specific ability means $\xi_{i\tau}^m$ and individual specific type probabilities $q_{i\tau}^m$ can be used to form best guesses of this variable given by $\tilde{\alpha}_i^m = \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \xi_{i\tau}^m$. Denote the errors in these best guesses by $v_i^m = \alpha_i - \tilde{\alpha}_i^m$. By construction, the information set used to construct $\alpha_i^m$ contains student $i$'s information set at the beginning of term $t$. As such, the residual $v_i^m$ is orthogonal to student information sets at the beginning of term $t$. This suggests the following regression can be used to estimate $\theta_k^m$:

$$g_{itklc} - \tilde{\alpha}_{ik}^m = X_{itkl}\theta_k + \tilde{\eta}_{itklc} \tag{32}$$

where the composite error term is given by $\tilde{\eta}_{itklc} = \upsilon_{ik}^m + \eta_{itklc}$.

Standard deviation in grading noise can be estimated using residuals from this regression with an adjustment for the extra noise introduced by $\upsilon_{ik}^m$. The closed form expression is:[32]

$$\left(\sigma_{kl}^{m+1}\right)^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T \sum_{c=1}^{C_{it}} (d_{ict} = \{k,l\}) \left\{ \left(g_{itklc} - X_{itkl}\theta_k^{m+1}\right)^2 - 2\left(g_{itklc} - X_{itkl}\theta_k^{m+1}\right)\mathbb{E}_m[\alpha_{ik}] + \mathbb{E}_m\left[\alpha_{ik}^2\right] \right\}}{\sum_{i=1}^N \sum_{t=1}^T \sum_{c=1}^{C_{it}} (d_{ict} = \{k,l\})} \tag{33}$$

**Choice Parameters** Choice parameters are estimated by numerically maximizing the choice contribution to the expected log likelihood taking iteration $m$ estimates of $\gamma_\tau$, $\Delta_\tau$, $\theta_k$, $\sigma_{kl}$ as given. Formally,

$$\omega^{m+1} = \text{argmax}_\omega \left\{ \sum_{i=1}^N \sum_{\tau=1}^{\mathscr{T}} \sum_{t=1}^T \sum_{c=1}^{C_{it}} q_{i\tau}^m \left[ (d_{itc} = \{k,l\}) u_{itkl} - \ln\left( \sum_{k'=1}^K \sum_{l'=1}^L \exp\left(u_{itk'l'}\right) \right) \right] \right\} \tag{34}$$

# Appendix B: Uncertain Admissions and Class Composition

This appendix presents a simple theoretical model demonstrating how uncertainty in admissions can generate a negative relationship between the means and variances of ability distributions. Suppose all students are endowed with a unidimensional ability vector $\alpha_i$ which is drawn from an individual specific distribution as: $\alpha_i \sim N(\gamma_i, \Delta_i)$. When universities evaluate applications for admissions, they do not observe $\alpha_i$ but instead observe an independent application draw $a_i$ which is also drawn as $a_i \sim N(\gamma_i, \Delta_i)$. This is equivalent to assuming universities observe noisy ability $\alpha_i + \eta_i$ where $\eta_i \sim N(0, \Delta_i)$.

Given the set of application signals $\{a_i\}$, universities choose an admission threshold $\underline{a}$ such that students with $a_i \geq \underline{a}$ are admitted and students with $a_i < \underline{a}$ are rejected. The probability of

[32] Where iteration $m$ expectations are given by:

$$\mathbb{E}_m[\alpha_i] = \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \xi_{i\tau}^m$$

$$\mathbb{E}_m[\alpha_i]^2 = \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \left( \Upsilon_{i\tau}^m(k,\, k) + \left(\xi_{i\tau}^m(k)\right)^2 \right)$$

admission given $\gamma_i$ and $\Delta_i$ is then:

$$\Pr\left(a_i \geq \underline{a} \,|\, \gamma_i, \Delta_i\right) = 1 - \Phi\left(\frac{\underline{a} - \gamma_i}{\sqrt{\Delta_i}}\right) \tag{35}$$

where $\Phi(\cdot)$ is the standard normal CDF.

This probability is increasing in $\gamma_i$; additionally, the probability is increasing in $\Delta_i$ for students with $\gamma_i < \underline{a}$ and decreasing in $\Delta_i$ for students with $\gamma_i > \underline{a}$. Intuitively, this second property occurs because lower ability students only gain admission if they produce an unusually impressive application which occurs more often when there is greater uncertainty. Conversely, higher ability students will only fail to gain admission if they produce an unusually poor application which occurs less often when there is less uncertainty.

As a result, admitted students with $\gamma_i < \underline{a}$ have selectively high $\Delta_i$ (and this selection strengthens as $\gamma_i$ gets further below $\underline{a}$). Additionally, admitted students with $\gamma_i > \underline{a}$ will have selectively low $\Delta_i$ (and this selection strengthens as $\gamma_i$ gets further above $\underline{a}$). This implies that if $\gamma_i$ and $\Delta_i$ are independent or inversely correlated in the general population then they will be inversely correlated in the selected population which gains admission. This is not necessarily true if $\gamma_i$ and $\Delta_i$ are positively correlated in the general population but it seems unlikely that more confident students are less certain.

If enrollment decisions conditional on admission are independent of $\gamma_i$ and $\Delta_i$ then the above statement applies to the selected population of enrolled students as well. This is consistent with my empirical finding: type 1 students have low expectations about their unobserved component of grade production and high uncertainty in these expectations while type 2 students have high expectations about their unobserved component of grade production and low uncertainty in these expectations.

Table 1: Grade Distributions and Total and Residual Variances

Panel A: Grade Distributions

| **Introductory Courses** | Humanities | Social Sciences | Sciences | Engineering | Economics |
|---|---|---|---|---|---|
| A | 37.65% | 25.54% | 24.53% | 33.61% | 22.72% |
| A- | 23.50% | 20.05% | 12.43% | 14.60% | 11.83% |
| B+ | 13.53% | 18.93% | 12.64% | 9.54% | 11.45% |
| B | 13.63% | 17.88% | 16.69% | 16.65% | 19.02% |
| B- | 4.86% | 8.33% | 8.12% | 8.39% | 10.17% |
| C+ or below | 6.82% | 9.27% | 25.60% | 17.20% | 24.81% |
| Observations | 5,648 | 2,244 | 4,774 | 642 | 996 |
| **Advanced Courses** | | | | | |
| A | 39.22% | 33.40% | 31.97% | 39.55% | 32.60% |
| A- | 26.66% | 25.16% | 15.15% | 17.86% | 22.35% |
| B+ | 16.47% | 18.00% | 13.32% | 12.76% | 15.28% |
| B | 9.57% | 12.97% | 15.92% | 13.15% | 14.19% |
| B- | 4.01% | 4.91% | 8.71% | 7.14% | 6.92% |
| C+ or below | 4.07% | 5.55% | 14.92% | 9.54% | 8.66% |
| Observations | 6,483 | 8,131 | 5,864 | 980 | 1,670 |

Panel B: Total and Residual Variances

| **Total Variance** | Humanities | Social Sciences | Sciences | Engineering | Economics |
|---|---|---|---|---|---|
| Introductory Courses | 0.3815 | 0.3883 | 0.6977 | 0.6525 | 0.6872 |
| Advanced Courses | 0.2800 | 0.3258 | 0.6029 | 0.4540 | 0.4074 |
| **Residual Variance (% of Total)** | | | | | |
| Introductory Courses | 54.9% | 39.0% | 41.9% | 31.1% | 20.9% |
| Advanced Courses | 56.8% | 58.0% | 49.9% | 58.3% | 54.6% |

Notes: Estimated with transcript data for respondents to the Campus Life and Learning Survey at Duke University. Total Variance is overall variance in earned grade points within each field-level. Residual variance is variance within each field-level after differencing out student-field-level fixed effects and is reported as a percentage of total variance. Residual variance is meant to illustrate how much of the variance in grades within each field-level is experienced by an individual student.

Table 2: Student Characteristics by Field of Selection

| | | Field of Selection | | | | |
|---|---|---|---|---|---|---|
| **Composite SAT** | Uncond. | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Intro Courses | 1398.0 | 1395.6 | 1382.2*** | 1397.5 | 1451.8*** | 1410.8*** |
| | *1.07* | *1.72* | *2.71* | *1.87* | *3.47* | *3.70* |
| Adv Courses | 1407.0 | 1400.3*** | 1383.7*** | 1434.0*** | 1448.2*** | 1419.3*** |
| | *0.83* | *1.59* | *1.48* | *1.49* | *2.77* | *2.87* |
| **Intro Residuals** | | | | | | |
| Adv Courses | 0.0027 | -0.0075 | -0.0200*** | 0.0283*** | -0.0011 | 0.0616*** |
| | *0.0030* | *0.0061* | *0.0049* | *0.0062* | *0.0136* | *0.0109* |

Notes: Standard errors in italics. *** indicates statistically different from unconditional average at 1% significance. A student's Composite SAT score is her math SAT score plus her verbal SAT score. A student's "intro residual" is computed by first running field specific regressions of grade points on a quadratic in own-field experience, indicators for course load, race and gender indicators, math and verbal SAT scores, and five measures of application quality using only observations of freshmen and sophomores taking introductory courses. A student's intro residual is then the average of all of her residuals from these regressions. I then calculate the average intro residual across all observations of juniors and seniors taking advanced courses (column Uncond.) and the average residual for observations of juniors and seniors taking advanced courses in a specific field (Field of Selection columns). Statistics indicate that students who outperform rich observed covariates in introductory courses when they are freshmen and sophomores are more likely to choose advanced science and economics courses and less likely to choose advanced social science courses when they are juniors and seniors.

Table 3: Expected Grades by Academic Field and Course Level

| | (1) | | (2) | | (3) | | (4) | |
|---|---|---|---|---|---|---|---|---|
| | Intro | Adv | Intro | Adv | Intro | Adv | Intro | Adv |
| Humanities | 3.507 | 3.567 | 3.474 | 3.534 | 3.469 | 3.529 | 3.48 | 3.538 |
| | *0.014* | *0.011* | *0.015* | *0.013* | *0.024* | *0.022* | *0.023* | *0.022* |
| Soc Science | 3.369 | 3.527 | 3.34 | 3.512 | 3.347 | 3.517 | 3.358 | 3.522 |
| | *0.016* | *0.01* | *0.017* | *0.011* | *0.025* | *0.021* | *0.025* | *0.021* |
| Science | 3.095 | 3.169 | 3.059 | 3.105 | 3.169 | 3.214 | 3.176 | 3.211 |
| | *0.021* | *0.02* | *0.022* | *0.022* | *0.038* | *0.037* | *0.037* | *0.038* |
| Engineering | 3.14 | 3.342 | 3.165 | 3.32 | 3.215 | 3.377 | 3.256 | 3.417 |
| | *0.036* | *0.031* | *0.037* | *0.034* | *0.052* | *0.045* | *0.061* | *0.057* |
| Economics | 3.025 | 3.373 | 2.973 | 3.338 | 3.06 | 3.423 | 3.057 | 3.422 |
| | *0.029* | *0.022* | *0.031* | *0.032* | *0.04* | *0.043* | *0.042* | *0.052* |
| | | | | | | | | |
| Course Level | Yes | | Yes | | Yes | | Yes | |
| Experience | No | | Yes | | Yes | | Yes | |
| Course Load | No | | Yes | | Yes | | Yes | |
| Gender | No | | No | | Yes | | Yes | |
| Race | No | | No | | Yes | | Yes | |
| SAT Scores | No | | No | | No | | Yes | |
| App quality | No | | No | | No | | Yes | |
| | | | | | | | | |
| Course Obs | 37255 | | 37255 | | 37255 | | 37255 | |
| Student Obs | 1127 | | 1127 | | 1127 | | 1127 | |

Notes: Bootstrap standard errors (100 iterations) are in italics. Results are from estimates of a Bayesian correlated learning model which controls for selection on observed covariates and unobserved beliefs about abilities. When covariates are included, expected grades are for a student who is white, male, taking four classes, has no prior experience for introductory courses and five courses of own category experience for advanced courses (0 and 5 are the respective modes), and has average SAT scores and application quality measures.

Table 4: Variance in Grades and Correlation in Abilities

| Introductory Courses | Humanities | Social Science | Science | Engineering | Economics |
|---|---|---|---|---|---|
| Total Variance | 0.3562 | 0.3476 | 0.649 | 0.5985 | 0.5652 |
| Standard Error | *0.0144* | *0.0134* | *0.0161* | *0.0403* | *0.0293* |
| *Share of variance due to* | | | | | |
| Observed Covariates | 7.50% | 7.60% | 10.90% | 16.10% | 9.50% |
| Unobserved Ability | 18.00% | 20.90% | 29.10% | 30.80% | 23.80% |
| Grading Noise | 74.60% | 71.50% | 60.00% | 53.10% | 66.70% |

| Advanced Courses | Humanities | Social Science | Science | Engineering | Economics |
|---|---|---|---|---|---|
| Total Variance | 0.2842 | 0.3122 | 0.6101 | 0.585 | 0.4799 |
| Standard Error | *0.0124* | *0.0144* | *0.0189* | *0.0437* | *0.0272* |
| *Share of variance due to* | | | | | |
| Observed Covariates | 9.30% | 8.40% | 11.50% | 16.40% | 11.20% |
| Unobserved Ability | 22.50% | 23.30% | 31.00% | 31.50% | 28.00% |
| Grading Noise | 68.20% | 68.30% | 57.50% | 52.00% | 60.80% |

| Correlation Matrix | Humanities | Social Science | Science | Engineering | Economics |
|---|---|---|---|---|---|
| Humanities | 1.0000 | 0.8612 | 0.7278 | 0.8076 | 0.7699 |
| Social Science | | 1.0000 | 0.6808 | 0.7466 | 0.8622 |
| Science | | | 1.0000 | 0.9811 | 0.9252 |
| Engineering | | | | 1.0000 | 0.9366 |
| Economics | | | | | 1.0000 |

Notes: Results are from estimates of a Bayesian correlated learning model which controls for selection on observed covariates and unobserved beliefs about abilities. These results use estimates of specification 4 which includes controls for course load, own-field experience, gender, race, SAT scores, and five measures of application quality. Total variance is hypothetical variance which would arise if course selection were random and all students had the same experience and course loads. Share of variance due to observed covariates is the share of variance due to differences in race, gender, SAT scores, and application quality measures in the entire student population. Share of variance due to unobserved ability is share of variance due to differences in unobserved ability in the entire student population. Correlation matrix is the correlation matrix corresponding to the unobserved ability variance-covariance matrix $\Delta$.

Table 5: Learning through Grade Signals

Spec 1: No Covariates

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.095 | 26 | 22 | 25 | 35* | 21 |
| Social Science | 0.102 | 20 | 29 | 24 | 30 | 26 |
| Science | 0.277 | 16 | 17 | 41 | 44 | 30 |
| Engineering | 0.281 | 19 | 18 | 39 | 47 | 29 |
| Economics | 0.200 | 17 | 22 | 37 | 41* | 33 |
| Average | | 20 | 22 | 33 | 39 | 28 |

Spec 2: Course load and Experience

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.095 | 26 | 24 | 26 | 37* | 23 |
| Social Science | 0.106 | 21 | 30 | 23 | 31 | 27 |
| Science | 0.282 | 16 | 17 | 42 | 45 | 31 |
| Engineering | 0.283 | 21 | 20 | 40 | 47 | 31 |
| Economics | 0.204 | 18 | 23 | 38 | 42* | 34 |
| Average | | 20 | 23 | 34 | 40 | 29 |

Spec 3: + Gender and Race

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.085 | 24 | 22 | 23 | 32* | 20 |
| Social Science | 0.093 | 19 | 27 | 20 | 26 | 24 |
| Science | 0.235 | 15 | 15 | 38 | 41 | 28 |
| Engineering | 0.229 | 18 | 17 | 36 | 42 | 28 |
| Economics | 0.168 | 16 | 21 | 34 | 38* | 31 |
| Average | | 18 | 20 | 30 | 36 | 26 |

Spec 4: + SAT scores and Application Quality

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.063 | 19 | 17 | 17 | 24 | 15 |
| Social Science | 0.072 | 14 | 23 | 14 | 20 | 20 |
| Science | 0.187 | 10 | 10 | 33 | 36 | 23 |
| Engineering | 0.180 | 13 | 12 | 32 | 37 | 23 |
| Economics | 0.135 | 11 | 17 | 28 | 33 | 26 |
| Average | | 13 | 16 | 25 | 30 | 21 |

Notes: "Initial" is variance in initial beliefs by field. Reduction (%) is the percent reduction in the variance in beliefs about ability in the row field after receiving one grade signal from the column field. (*) indicates greater than diagonal element of same row at 5% significance. Specification 1 controls for course level. Specification 2 adds course load dummy variables and quadratic own-field experience. Specification 3 adds gender and race dummy variables. Specification 4 adds SAT scores and five measures of application quality.

Table 6: Learning through Grade Signals (Coarse Fields)

Spec 1: No Covariates

| | | Reduction (%) in uncertainty | |
|---|---|---|---|
| | Initial | HuSS | SEE |
| HuSS | 0.093 | 25 | 26 |
| SEE | 0.263 | 16 | 40 |
| Average | | 21 | 33 |

Spec 4: All Covariates

| | | Reduction (%) in uncertainty | |
|---|---|---|---|
| | Initial | HuSS | SEE |
| HuSS | 0.061 | 18 | 16 |
| SEE | 0.171 | 10 | 31 |
| Average | | 14 | 24 |

Notes: "Initial" is variance in initial beliefs by field. Reduction (%) is the percent reduction in the variance in beliefs about ability in the row field after receiving one grade signal from the column field. Specification 1 controls for course level. Specification 4 controls for course load, own-field experience, gender, race, SAT scores, and five measures of application quality. HuSS includes humanities and social sciences. SEE includes sciences, engineering, and economics

Table 7: Learning through Grade Signals (Fine Fields)

Spec 1: No Covariates

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Initial | Hum. | For. Lang. | Arts | Soc. Sci. | Life Sci. | Phys. Sci | Eng. | Econ. |
| Hum. | 0.082 | 29 | 28 | 19 | 24 | 30 | 25 | 32 | 23 |
| For. Lang. | 0.216 | 17 | 48 | 12 | 16 | 23 | 22 | 33 | 17 |
| Arts | 0.057 | 20 | 21 | 26 | 18 | 22 | 20 | 26 | 17 |
| Soc Sci | 0.102 | 24 | 26 | 16 | 29 | 30 | 23 | 30 | 26 |
| Life Sci. | 0.253 | 18 | 24 | 12 | 19 | 46 | 32 | 39 | 28 |
| Phys. Sci. | 0.338 | 15 | 23 | 11 | 14 | 31 | 47 | 44 | 30 |
| Eng. | 0.294 | 19 | 33 | 14 | 18 | 37 | 43 | 48 | 30 |
| Econ. | 0.206 | 19 | 25 | 13 | 22 | 37 | 41* | 42* | 34 |
| Average | | 20 | 29 | 15 | 20 | 32 | 32 | 37 | 26 |

Spec 4: All Covariates

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Initial | Hum. | For. Lang. | Arts | Soc. Sci. | Life Sci. | Phys. Sci | Eng. | Econ. |
| Hum. | 0.072 | 26 | 34* | 15 | 23 | 30 | 27 | 31 | 26 |
| For. Lang. | 0.273 | 16 | 54 | 8 | 16 | 28 | 28 | 34 | 23 |
| Arts | 0.043 | 19 | 20 | 21 | 17 | 22 | 21 | 22 | 18 |
| Soc Sci | 0.094 | 22 | 33 | 13 | 27 | 31 | 26 | 30 | 29 |
| Life Sci. | 0.253 | 18 | 33 | 10 | 19 | 45 | 38 | 41 | 33 |
| Phys. Sci. | 0.362 | 15 | 31 | 9 | 15 | 36 | 48 | 46 | 33 |
| Eng. | 0.288 | 17 | 39 | 10 | 17 | 39 | 46 | 47 | 34 |
| Econ. | 0.224 | 18 | 33 | 10 | 21 | 40 | 43* | 43* | 37 |
| Average | | 19 | 35 | 12 | 19 | 34 | 35 | 37 | 29 |

Notes: "Initial" is variance in initial beliefs by field. Reduction (%) is the percent reduction in the variance in beliefs about ability in the row field after receiving one grade signal from the column field. (*) indicates greater than diagonal element of same row at 5% significance. Specification 1 controls for course level. Specification 4 controls for course load, own-field experience, gender, race, SAT scores, and five measures of application quality.

Table 8: Expected Grades and Type Characteristics

Panel A: Expected Grades

| | Specification 1 | | | | Specification 4 | | | |
|---|---|---|---|---|---|---|---|---|
| | Type 1 | | Type 2 | | Type 1 | | Type 2 | |
| | Intro | Adv | Intro | Adv | Intro | Adv | Intro | Adv |
| Humanities | 3.233 | 3.290 | 3.688 | 3.745 | 3.237 | 3.263 | 3.582 | 3.608 |
| | *0.033* | *0.032* | *0.020* | *0.017* | *0.040* | *0.041* | *0.027* | *0.028* |
| Social Sciences | 3.137 | 3.291 | 3.526 | 3.680 | 3.148 | 3.245 | 3.452 | 3.549 |
| | *0.031* | *0.028* | *0.026* | *0.020* | *0.040* | *0.040* | *0.031* | *0.027* |
| Sciences | 2.619 | 2.683 | 3.409 | 3.473 | 2.743 | 2.734 | 3.375 | 3.367 |
| | *0.046* | *0.045* | *0.045* | *0.041* | *0.060* | *0.060* | *0.062* | *0.061* |
| Engineering | 2.669 | 2.857 | 3.429 | 3.617 | 2.824 | 3.056 | 3.415 | 3.647 |
| | *0.091* | *0.087* | *0.049* | *0.044* | *0.101* | *0.106* | *0.081* | *0.092* |
| Economics | 2.616 | 2.947 | 3.297 | 3.629 | 2.664 | 2.908 | 3.252 | 3.495 |
| | *0.055* | *0.057* | *0.042* | *0.036* | *0.069* | *0.076* | *0.058* | *0.060* |

| Panel B: Type Characteristics | | | | |
|---|---|---|---|---|
| % of Students | 40.1% | 59.9% | 35.3% | 64.7% |
| **Avg. Completed Courses** | | | | |
| Humanities | 11.08 | 10.02 | 11.18 | 10.05 |
| Social Sciences | 10.61 | 8.04 | 10.28 | 8.41 |
| Sciences | 8.32 | 9.93 | 8.54 | 9.69 |
| Engineering | 1.16 | 1.86 | 1.37 | 1.70 |
| Economics | 2.13 | 2.53 | 2.00 | 2.57 |
| **Characteristics** | | | | |
| HS: Religious | 14.1% | 9.3% | 13.5% | 10.0% |
| HS: Private | 16.8% | 22.8% | 16.3% | 22.7% |
| Parent Ed: College | 19.4% | 14.6% | 18.5% | 15.4% |
| Parent Ed: Some grad | 35.3% | 33.0% | 35.8% | 32.9% |
| Parent Ed: Professional | 32.8% | 45.7% | 34.3% | 44.0% |
| Avg. Number of AP scores | 3.65 | 5.12 | 3.86 | 4.90 |
| | | | | |
| Course Level | Yes | Yes | Yes | Yes |
| Exp. & Course Load | No | Yes | Yes | Yes |
| Gender & Race | No | No | Yes | Yes |
| SAT & App quality | No | No | No | Yes |
| | | | | |
| Course Obs | 37255 | 37255 | 37255 | 37255 |
| Student Obs | 1127 | 1127 | 1127 | 1127 |

Notes: Bootstrap standard errors (100 iterations) are in italics. Results are from estimates of a Bayesian correlated learning model which allows for unobserved heterogeneity in initial prior beliefs. When covariates are included, expected grades are for a student who is white, male, taking four classes, has no prior experience for introductory courses and five courses of own-field experience for advanced courses (0 and 5 are the respective modes), and has average SAT scores and application quality measures. In Panel B, omitted high school type is public and omitted parent's maximum education is less than college.

Table 9: Learning through Grade Signals (Specification 1)

Panel A: Type 1 Students

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.1049 | 28 | 14 | 4 | 12 | 4 |
| Social Sciences | 0.1251 | 17 | 33 | 7 | 12 | 13 |
| Sciences | 0.1947 | 5 | 7 | 33 | 24 | 14 |
| Engineering | 0.3357 | 22 | 19 | 37 | 51 | 23 |
| Economics | 0.1657 | 4 | 12 | 13 | 13 | 30 |
| Average | | 15 | 17 | 19 | 23 | 17 |

Panel B: Type 2 Students

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.0140 | 5 | 8* | 18* | 22* | 16* |
| Social Sciences | 0.0236 | 5 | 9 | 16* | 20* | 15* |
| Sciences | 0.0916 | 5 | 7 | 19 | 24 | 16 |
| Engineering | 0.0964 | 5 | 7 | 19 | 24 | 16 |
| Economics | 0.0748 | 5 | 8 | 18 | 23* | 16 |
| Average | | 5 | 8 | 19 | 24 | 16 |

Notes: "Initial" is variance in initial beliefs by field. Reduction (%) is the percent reduction in the variance in beliefs about ability in the row field after receiving one grade signal from the column field. (*) indicates greater than diagonal element of same row at 5% significance. Specification 1 controls for course level only.

Table 10: Learning through Grade Signals (Specification 4)

Panel A: Type 1 Students

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.0949 | 26 | 16 | 3 | 16 | 2 |
| Social Sciences | 0.1157 | 13 | 32 | 4 | 14 | 8 |
| Sciences | 0.1587 | 2 | 5 | 29 | 31 | 6 |
| Engineering | 0.2271 | 10 | 11 | 22 | 42 | 7 |
| Economics | 0.1309 | 2 | 10 | 7 | 11 | 27 |
| Average | | 11 | 15 | 13 | 23 | 10 |

Panel B: Type 2 Students

| | | Reduction (%) in uncertainty after one introductory grade (by field) | | | | |
|---|---|---|---|---|---|---|
| | Initial | Humanities | Social Sciences | Sciences | Engineering | Economics |
| Humanities | 0.0101 | 4 | 6* | 14* | 17* | 11* |
| Social Sciences | 0.0168 | 3 | 6 | 11* | 14* | 10 |
| Sciences | 0.0692 | 3 | 5 | 15 | 18* | 10 |
| Engineering | 0.0691 | 4 | 5 | 15 | 18 | 11 |
| Economics | 0.0437 | 4 | 6 | 14* | 18* | 11 |
| Average | | 4 | 6 | 15 | 18 | 11 |

Notes: "Initial" is variance in initial beliefs by field. Reduction (%) is the percent reduction in the variance in beliefs about ability in the row field after receiving one grade signal from the column field. (*) indicates greater than diagonal element of same row at 5% significance. Specification 4 controls for course load, own-field experience, gender, race, SAT scores, and five measures of application quality.

Table 11: Excluding Small and Compressed Courses

| | Full Sample | Excluding Small Courses only | % Reduction (1) to (2) | Restricted Sample | % Reduction (2) to (4) |
|---|---|---|---|---|---|
| Introductory | (1) | (2) | (3) | (4) | (5) |
| HuSS Obs. | 7892 | 7321 | 7.2% | 3144 | 57.1% |
| SEE Obs. | 6412 | 6271 | 2.2% | 5118 | 18.4% |
| Total Obs. | 14304 | 13592 | 5.0% | 8262 | 39.2% |
| | | | | | |
| Advanced | | | | | |
| HuSS Obs. | 14614 | 11059 | 24.3% | 4175 | 62.2% |
| SEE Obs. | 8514 | 7452 | 12.5% | 4996 | 33.0% |
| Total Obs. | 23128 | 18511 | 20.0% | 9171 | 50.5% |

Notes: Courses are defined by course number (e.g. ECON 101) and thus may appear in multiple semesters possibly being taught by different instructors. "Small courses" have fewer than 10 observations. "Restricted Sample" excludes small courses and courses in which more than half of student observations earn an A- or above.

Table 12: Learning through Grade Signals Excluding Small and Compressed Classes

Spec 1: No Covariates

| | | Reduction (%) in uncertainty | |
|---|---|---|---|
| | Initial | HuSS | SEE |
| HuSS | 0.138 | 33 | 37* |
| SEE | 0.338 | 24 | 50 |
| Average | | 29 | 44 |

Spec 4: All Covariates

| | | Reduction (%) in uncertainty | |
|---|---|---|---|
| | Initial | HuSS | SEE |
| HuSS | 0.084 | 23 | 26* |
| SEE | 0.212 | 15 | 39 |
| Average | | 19 | 33 |

Notes: "Initial" is variance in initial beliefs by field. Reduction (%) is the percent reduction in the variance in beliefs about ability in the row field after receiving one grade signal from the column field. Specification 1 controls for course level. Specification 4 controls for course load, own-field experience, gender, race, SAT scores, and five measures of application quality. Excludes courses in which at least 50% of observations earned at least an A- and courses with fewer than 10 observations.