Compliers, defiers and the evaluation of information interventions

Marco Castillo* Mofioluwasademi Ige†‡

July 2023

Abstract

Information campaigns have the potential to influence decisions, but how should these campaigns be evaluated in the presence of heterogeneous responses? For example, information on the returns to education can lead to continuing and discontinuing education or entering and exiting a major. We make explicit the conditions for prior beliefs to identify heterogeneous responses to treatment. These conditions have testable implications. We use this approach to analyze an information provision RCT in Ibadan, Nigeria, implemented at a critical juncture when adolescents had to decide to stop or continue schooling. The average effect of the intervention is a 3.8 percentage point increase in the high school dropout rate one year later from a baseline of 9.4 percent. However, more than twice as many students changed their decisions due to the intervention, with one student deciding to continue schooling per two students deciding to discontinue schooling.

JEL classifications: C93, J24, D83

Keywords: human capital accumulation, costly information acquisition, randomized controlled trial

*Department of Economics, Texas A&M University, Melbourne Institute and IZA, e-mail: marco.castillo@tamu.edu

†Freddie Mac, e-mail: Mofioluwasademi_ige@freddiemac.com

‡We gratefully acknowledge the support of the state government of Oyo, Nigeria for this research. The project was pre-registered at the AEA RCT Registry (AEARCTR-0004839) and received human participants approval from Texas A&M University's IRB (IRB2018-0424D).

1 Introduction

At age 15 or earlier, about half of adolescents worldwide decide whether to continue their education and, if so, along which path (e.g., academic, vocational).[1] These decisions require careful consideration of available options since mistakes can be very costly (Lai, Sadoulet and de Janvry, 2009; Lucas and Mbiti, 2012). Existing evidence on the effectiveness of information provision in increasing levels of education is mixed.[2] However, an identification problem intrinsic to information interventions is that information can significantly affect some individuals without affecting average behavior. Indeed, the value of information is highest when a decision-maker is indifferent between alternatives (see de Lara and Gossner, 2020), meaning that information is likely to have nonuniform effects among those who value it most. The lack of identification can distort policy evaluation and theory testing. In this paper, we show how to use prior beliefs to identify treatment effects and how to test for the instrumental value of information. We show that without access to beliefs, the identification problem can be severe, and the interpretation of the willingness to pay for information might be biased.

Under random assignment, if beliefs are updated in the direction of signals, then prior beliefs can be used to identify nonuniform responses to information. Intuitively, if responses to information are nonuniform, we should observe distributional changes in prior beliefs across treatment conditions conditional on actual choices. For participants with ex-ante optimistic beliefs about the returns to education, we should observe relatively fewer of them continuing education in the treated group, when they receive "bad" news that the returns are lower than expected, than in the control group that does not receive news. Analogously, for participants with ex-ante pessimistic beliefs on the returns to education, we should observe relatively more of them deciding to continue education in the treated group, when they receive "good news," than in the control group.

In the language of the potential outcomes literature (Imbens and Angrist, 1994), having access to information on prior beliefs helps uncover both compliers and defiers. We can then determine if an information campaign's small effect is due to ineffectiveness or nonuniform response. The usefulness of this observation increases with the intervention's complexity

[1] See https://data.worldbank.org/indicator/SE.COM.DURS.

[2] Jensen (2010) finds that information about the returns of education increase schooling. Other studies providing information about several aspects of educational investment report null as well as positive and negative results (see Ajayi, Friedman and Lucas, 2020; Bergman, Denning and Manoli, 2019; Bettinger, Long, Oreopoulos and Sanbonmatsu, 2012; Busso, Dinkelman, Martinez and Romero, 2017; Goux, Gurgand and Maurin, 2017; Gurantz, Howell, Hurwitz, Larson, Pender and White, 2021; Hastings, Neilson and Zimmerman, 2015; Hoxby and Turner, 2015; Hyman, 2020; Loyalka, Liu, Song, Yi, Huang, Wei, Zhang, Shi, Chu and Rozelle, 2013).

and the options available to participants. This is particularly relevant in education, where individuals have multiple options and priors, and not all priors are necessarily updated. Since random treatment assignment identifies the marginal distributions of behavior across treatments, we can bound the maximum effect of an information intervention. For instance, Fréchet bounds can be used to derive upper and lower limits to the joint probability of outcomes. In the context of information interventions, those on the diagonal of the joint distribution of outcomes are the ones who do not react to information. This implies the model is testable; the measured effect of information cannot exceed these bounds. Experimental data can be consistent with uniform responses to information, non-uniform responses to information, or inconsistent. We discuss conditions for consistency to be expected.

A second issue is whether information interventions fail because the information provided has no instrumental value. This hypothesis is difficult to test because it requires knowing who would benefit from the information. Behavioral economics reminds us that the demand for information, and the reaction to it, might be distorted by agents' motivations (e.g., Golman, Loewenstein, Molnar and Saccardo, 2021), making it difficult to test this hypothesis. To address this challenge, we develop a test based on the intuition that willingness to pay (WTP) for information is a function of its anticipated influence on behavior. Blackwell (1951) offers the insight that data are valuable because they enable better decision-making. The same idea is discussed by Hirshleifer (1971) and de Lara and Gossner (2020), who propose that information is valuable insofar as it influences choices.[3]

In the binary choice case, e.g. continue education or not, we show that the derivative of the WTP for information with respect to relative expected payoffs equals the expected change in behavior (i.e., positive or negative). Testing this implication of rational information acquisition requires observing the WTP for information and the actual effect of the information on decisions. Note that to test if WTP is higher among those responsive to information, it is necessary first to identify who is affected by the intervention. If the effect of information is large but heterogeneous, assuming monotonicity could prevent testing this hypothesis. We implement the test by embedding an incentive-compatible measure of adolescents' WTP for information in an information-provision randomized control trial (RCT). Random assignments of information provide the needed counterfactual responses. The elicited WTP for information allows testing if those affected by the provided information are willing to pay more and by how much.

[3]We exploit the fact that the vector of choice probabilities compatible with rational choice coincides with the gradient of the ex-ante expected utility (Sørensen and Fosgerau, 2022; Chiong, Galichon and Shum, 2016; McFadden, 1978; Rust, 1994) to derive a variational representation of the WTP for information as a function of the anticipated changes in behavior it generates. The assumptions needed for this result are given in Section 2 and their relationship to previous results in the literature.

We use these ideas to analyze a field experiment in Ibadan, Nigeria, that randomized information about wages and college admission rates to over 3,600 14-year-old adolescents deciding whether to continue to senior high school. In Nigeria, as in other areas of the world, college admission is very selective, thus choosing to continue to senior high school in order to go to college can be risky if a student is unsuccessful.[4] We subsequently observe whether they did or did not continue to senior high school and which track they followed. As part of the study, we collected the adolescents' beliefs about earnings and their own future academic choices as well as their WTP for different amounts of information. For a subgroup of participants, these beliefs were collected twice, before and after randomization into groups that received information or not.

Participants updated their beliefs on earnings at levels found for adults (Fuster, Perez-Truglia, Wiederholt and Zafar, 2022; Hjort, Moreira, Rao and Santini, 2021). Similarly, the participants' beliefs about their future choices were consistent with partial sorting based on earnings (Arcidiacono, Hotz, Maurel and Romano, 2020; Wiswall and Zafar, 2021). Finally, using the framework developed by Wiswall and Zafar (2015a), and consistent with related studies (Delavande and Zafar, 2019; Haaland, Roth and Wohlfart, forthcoming; Wiswall and Zafar, 2018), we find that participants' choice elasticity with respect to earnings was relatively low (about 17 percent). These estimates are commensurate with observed field behavior. This suggests that our belief data is informative.

The information intervention led to an average 3.8 percentage point decrease in education continuation rates one year later, from a baseline dropout rate of 9.4 percent.[5] We find evidence against a uniform response to information. Firstly, under this hypothesis, we would expect the WTP of compliers in the Control and Treated conditions to be the same.[6] We strongly reject this hypothesis, thus casting doubt on a uniform effect of information. Secondly, we test for the hypothesis of a uniform response to information and find evidence against it. We use instrument validity tests (Kitagawa, 2015; Sun, 2022) to assess the statistical significance of the distributional changes in prior beliefs conditional on continuing education. We implement the test by constructing a measure of lifetime earnings from continuing and discontinuing education using elicited prior beliefs and measuring distributional changes conditional on choices. We estimate that the number of participants affected by information is more than twice as large as the intention-to-treat estimate suggests. For each

[4] According to Nigeria's Joint Admissions and Matriculation Board, one in three students who apply to college is admitted (see Section 3).

[5] Contrary to Jensen (2010), who found that students underestimated the returns of education, the participants in our study overestimated these returns.

[6] Under the assumption of monotonicity and random assignment, we can estimate the characteristics of compliers in both the control and treatment groups (e.g., Heckman and Pinto, 2018).

participant deciding to continue education due to the information provided, two students decided to discontinue it (4.9pp v 9.7pp).

The information intervention also affected the distribution of fields of study. Accounting for changes in fields of study, the intervention affected close to one in three students. We show these estimates can be recovered from distributional changes in prior beliefs conditional on educational choices by treatment condition. Finally, we use the proposed method to test if those affected by the intervention were ex-ante more willing to pay for information as theory predicts. We find that those who discontinue education due to the intervention valued information more, but those who continue education due to the intervention valued information less. This is consistent with some participants valuing information instrumentally, while other participants either avoided information or failed to anticipate their behavior. This pattern suggests caution in interpreting WTP for information as purely instrumental.

The paper's main contribution is determining the identifying power of prior beliefs to quantify non-uniform responses to information. We show that prior beliefs can be used under certain conditions to fully recover, or to bound, the joint distribution of choices across treatment conditions. Since the maximal effect of information on decisions is bounded, the validity of these conditions is potentially testable. The importance of prior beliefs in the analysis of information interventions is well-established. For instance, the analysis of the effect of information on posterior beliefs conditional on prior beliefs is a standard way to evaluate the validity of survey experiments (see Section 4.5 in Haaland et al., forthcoming, and the references therein). We build on those insights to derive identifying conditions of the effect of information on field outcomes.

That prior beliefs can affect the effectiveness of information interventions has also been recognized by several authors (e.g., Thornton, 2008; Hoxby and Turner, 2015; Jensen, 2010; Bursztyn, Gonzalez and Yanagizawa-Drott, 2020). Our work extends that approach and provides a way to assess its validity. To our knowledge, these insights have not been exploited to identify different types of marginal subjects. Importantly, we show an intuitive way to extend the approach to more complex information interventions. In the context of information campaigns in education, nonuniform responses are expected due to heterogeneity in beliefs and self-selection.[7] Heterogeneity in beliefs in the educational context has been reported before (Jensen, 2010; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015b) and is confirmed in our data. A complete analysis of information interventions, therefore, needs to invoke alternative identification assumptions (e.g. Dahl, Huber and Mellace, 2017; De Chaisemartin, 2017), collect additional data, or both.

[7]Initial beliefs affect treatment selection producing a non-separable treatment choice equation as discussed in Heckman, Urzua and Vytlacil (2006) section 7.

The second main contribution is empirical. We provide direct evidence that the effect of information can be significantly underestimated. We estimate that up to one in three students change their decisions due to the intervention. As we show in the paper, underestimation can be due to the assumption of uniform treatment response and the underestimation of the dimension of the relevant information set. Aggregation across dimensions leads to a loss of information. We also provide direct evidence for and against the instrumental value of information. As we discuss in the paper, assuming uniform response to treatment would have biased our test of theory. This implies that our approach affords to uncover violations of theory in different subpopulations. A researcher assuming uniformity will derive conclusions on a subpopulation, perhaps not representative, of those affected by the intervention (De Chaisemartin, 2017). Finally, our approach provides a new way to test for selection and evidence of its empirical relevance in education.

Our study speaks to the design and analysis of experiments. We find there is value in exploring the reaction to information across a wide spread of prior beliefs. The heterogeneous response to treatment suggests that one of the dimensions along which information interventions must be evaluated is their ability to affect choices. Our results advise waiting for additional information on life outcomes to test whether the intervention was beneficial or harmful. The study also highlights the importance of testing theoretical implications across different populations. While we expected symmetric results on the value of information for those continuing and discontinuing education, we found the opposite. This discovery would not be possible had we only implemented an intervention to change beliefs in one direction.

Our paper contributes to the growing literature on decision-making by minors (Brocas and Carrillo, 2021; Brocas, Carrillo, Combs and Kodaverdian, 2019; Castillo, Ferraro, Jordan and Petrie, 2011; Castillo, Jordan and Petrie, 2018, 2019; Harbaugh, Krause and Berry, 2001; Sutter, Kocher, Glaetzle-Ruetzler and Trautmann, 2013), which focuses on children's economic rationality and strategic sophistication and its relationship to life outcomes (see List, Petrie and Samek, forthcoming, for a review of this literature). The paper opens a new area of inquiry by looking directly at how minors deal with costly acquisition of information (see Caplin, 2016; Caplin, Csaba, Leahy and Nov, 2020). Our results are consistent with some minors rationally allocating their attention, but also with information avoidance among others (Golman et al., 2021)

Our paper contributes to the literature on the formation of human capital. We provide further evidence that minors are actors in their development (Del Boca, Flinn, Verriest and Wiswall, 2019). Embedding behavioral measures in an RCT enables us to show the impact of adolescents' information processing on outcomes. The finding are relevant to the literature on the measurement and importance of expected returns to education. That literature shows heterogeneous beliefs and preferences regarding career choices are important; however, it retains the assumption that such beliefs are exogenous.[8] Our research shows that these beliefs are not independent of the responsiveness to returns to education.

Recent work has explored the demand for potentially valuable information (e.g., Allcott and Kessler, 2019; Fuster et al., 2022, and references therein). While the idea that information is valuable to the extent that it can influence behavior is old, such a connection is rarely exploited. An important exception is Chassang, Padro i Miquel and Snowberg (2012) who show that the willingness to pay to be treated in an experiment can be used to estimate actual and perceived treatment effects. Berry, Fischer and Guiteras (2020) apply this to sanitation interventions. Our approach applies these ideas to interventions where the treatment being sold is information. Fuster et al. (2022) show that the WTP for information increases in experimental stakes.[9] Our paper uses WTP to test for information processing biases.

The rest of this paper is organized as follows. Section 2 derives the main theoretical results used in the paper. Section 3 describes the RCT. Section 4 presents the main results. Section 5 discusses the potential limitations of the study. Section 6 concludes the paper. A series of appendices provide additional results.

2 Theoretical framework

2.1 Decision framework

This section discusses the decision framework used in the analysis. Appendix A.1 provides a more general model.

Students choose from $K+1$ choices $0, 1, \ldots, K$. The utility they derive from choice k can be decomposed into two additive parts: $U_k + e_k$, where U_k represents the lifetime expected utility associated with k, while e_k represents idiosyncratic shocks across choices. A student chooses option k^* if $U_{k^*} + e_{k^*} \geq max_k\{U_k + e_k\}$. We assume that $U_0 = e_0 = 0$. We define the "social surplus function" (McFadden, 1981; Sørensen and Fosgerau, 2022), which is the expected utility obtained from the choice problem:

$$\mathcal{W}(U) = E[max_k\{U_k + e_k\} - max_k e_k | U] \tag{1}$$

[8] This body of literature is too large to summarize here. Relevant papers utilizing belief elicitation include Arcidiacono et al. (2020), Delavande and Zafar (2019), and Wiswall and Zafar (2015a).

[9] In Fuster et al. (2022), rewards are based on the accuracy of beliefs. In our study, we observe behavior after information is provided. Bronchetti, Kessler, Magenheim, Taubinsky and Zwick (2020) use the framework developed by Caplin et al. (2020) to derive a test of rational inattention. They find that participants undervalue reminders relative to the costly information acquisition benchmark.

where $U = (u_1, \ldots, u_K)$. We define the conditional choice correspondence, $\mathcal{P}(U)$, as the probability of choosing each option given U is consistent with maximization. $\mathcal{P}(U)$ is a correspondence because without further assumptions ties and different tie-breaking rules are possible. Sørensen and Fosgerau (2022) prove that $\mathcal{W}(U)$ is finite, convex and everywhere subdifferentiable, and its subdifferential coincides with the conditional choice probability correspondence, i.e. $\partial\mathcal{W}(\cdot|U) = \mathcal{P}(\cdot|U)$. Moreover, the conditional choice correspondence $\mathcal{P}(U)$ is cyclic monotone. Shi, Shum and Song (2018) show that if the distribution of e_k's is absolutely continuous with respect to the Lebesgue measure and independent of U, then $\mathcal{W}(U)$ is differentiable.

The implication of cyclic monotonicity can be seen most clearly in comparing the conditional choice probabilities given two possible vectors of expected utilities U and U'. In particular, cyclic monotonicity requires that, for all (U', U), we must have that $(\mathcal{P}(U') - \mathcal{P}(U))'(U' - U) \geq 0$.[10] In particular, if there are only two options $k = \{0, 1\}$ with $k = 0$ representing dropping out of school and $k = 1$ representing continuing education, cyclical monotonicity implies that if the relative utility gain from continuing education, U', is larger than U, then the observed probability of continuing education must weakly increase. The model's predictions depend on the assumptions of how information affects expected utilities. If we have a proxy of changes in expected utility, cyclical monotonicity can help identify the effects of information on behavior consistent with maximization. The following section discusses the needed assumptions to implement this approach using belief data as proxies of utilities.

This framework can be used to evaluate the value of information. For instance, let $U(s_j), j = 1, \ldots, M$ be the vector of expected utilities if signal s_j is received. Signal s_i is distributed according to a finite probability distribution π such that $E_\pi[U(s)] = U$.[11] For a given status quo U, we can define the willingness to pay for this information, WTP, as the solution to the following equation:

$$E_\pi[\mathcal{W}(U(s))] - \text{WTP} = \mathcal{W}(U) \tag{2}$$

The case in which $k \in \{0, 1\}$ illustrates the usefulness of this result. Differentiating equation (2), we obtain that $d\text{WTP}/dU = E_\pi[\mathcal{P}(U(s))] - \mathcal{P}(U)$, where U is the return to continuing education.[12] The slope of the WTP is positive for those who would ex-ante increase the chances of continuing education upon receipt of the information, and negative for

[10]This condition must hold for any selection $p \in \mathcal{P}(U)$.

[11]Information structures can be represented as a distribution over a set of posterior distributions that average to the prior distribution. Section A.1 discusses this approach in more detail.

[12]We assume here that $U(s) = U + \eta(s)$, where $\eta(s)$ are adjustments to priors given signal s.

those who would ex-ante decrease the chances of continuing education with information. The WTP is maximal among those who are ex-ante indifferent between the options and therefore have the most to gain from acquiring information (see de Lara and Gossner, 2020, for results in greater generality). This result is useful because it implies that if the information has instrumental value, then behavioral changes should be larger among those with higher WTP for information.

2.2 Identifying behavioral types

We build on the previous framework to develop a simple procedure to identify non-monotone responses to information. We discuss when prior beliefs can be used to identify who switched to continue education and who switched to discontinue education. The procedure is generalized to more complex information interventions.

We let $D_i = 1$ be the decision of individual i to continue education and $D_i = 0$ be the decision to discontinue education. $Y_{i,1}$ denotes i's prior belief in lifetime earnings of continuing education and $Y_{i,0}$ denotes i's prior belief in lifetime earnings of discontinuing education.

Following our experiment, we denote the updated beliefs after receiving a signal S by $Y_i' = (Y_{i,0}', Y_{i,1}')$. We assume that beliefs are not updated if no new information is available, i.e. $Y_{i,j}' = Y_{j,i}, j = 0, 1$. Finally, let Z_i equal 1 if i receives signal S and Z_i equal 0 if i receives no signal. We let variable U_i be a barrier to continuing education not observed by the researcher. We make the following assumption.

Assumption 1: (i) $sgn((Y_{i,1}' - Y_{i,0}') - (Y_{i,1} - Y_{i,0})) = sgn(S - Y_{i,1})$, (ii) $D_i = \mathbf{1}[\nu(Y_{i,1}' - Y_{i,0}') \geq U_i]$ for a non trivial increasing function $\nu(\cdot)$ of $Y_{i,1}' - Y_{i,0}'$, (iii) Z_i is jointly independent of $(Y_{i,0}, Y_{i,1}, U_i)$.

Assumption 1(i) states that beliefs on returns to education update in the direction of the signal. Expected returns to education are updated upwards if the prior belief of lifetime earnings of continuing education is below the signal. Expected returns to education are updated downwards if the prior belief of lifetime earnings of continuing education is above the signal. Assumption 1(i) is satisfied if prior beliefs and signals are distributed normal, and signals on returns to education do not affect the belief on earnings of discontinuing education too strongly. In this case, $(Y_{i,1}' - Y_{i,0}') - (Y_{i,1} - Y_{i,0}) = \frac{\sigma_1^2 - \sigma_{1,0}}{\sigma_1^2 + \sigma_\varepsilon^2}(S - Y_{i,1})$ and $sgn(Y_{i,1}' - Y_{i,1}) = sgn(S - Y_{i,1})$ if $\sigma_1^2 - \sigma_{1,0} > 0$, where σ_i^2 is the variance of Y_i, $\sigma_{1,0}$ their covariance, and σ_ε the variance of the signal. Bayesian updating and normality are not necessary conditions for Assumption 1 to hold (see Benjamin, 2019). Assumption 1(ii) states that students self-select into education based on their expected returns (see Willis and Rosen, 1979). Assumption 1(iii) states that priors and selection into education are independent of treatment assignment.

Proposition 1: Under Assumption 1, $Pr(D_i = 1, Y_{i,1}|Z_i = 1) \geq Pr(D_i = 1, Y_{i,1}|Z_i = 0)$ if $Y_{i,1} < S$ and $Pr(D_i = 1, Y_{i,1}|Z_i = 1) \leq Pr(D_i = 1, Y_{i,1}|Z_i = 0)$ if $Y_{i,1} > S$.

Proof: Fix prior $(Y_{i,0}, Y_{i,1})$. For those receiving the signal, i.e., $Z_i = 1$, Assumption 1(i) implies that $Y'_{i,1} - Y'_{i,0} \geq Y_{i,1} - Y_{i,0}$ if $Y_{i,1} < S$ and $Y'_{i,1} - Y'_{i,0} \leq Y_{i,1} - Y_{i,0}$ if $Y_{i,1} > S$. For those not receiving the signal, i.e., $Z_i = 0$, $Y'_{i,1} - Y'_{i,0} = Y_{i,1} - Y_{i,0}$. Assumption 1(ii) implies that $E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,0} = y_{i,0}, Y_{i,1} = y_{i,1}, Z_i = 1] \geq E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,0} = y_{i,0}, Y_{i,1} = y_{i,1}, Z_i = 0]$ if $y_{i,1} < S$, and $E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,0} = y_{i,0}, Y_{i,1} = y_{i,1}, Z_i = 1] \leq E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,0} = y_{i,0}, Y_{i,1} = y_{i,1}, Z_i = 0]$ if $y_{i,1} > S$. Assumption 1(iii) then implies that $E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,1} = y_{i,1}, Z_i = 1] \geq E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,1} = y_{i,1}, Z_i = 0]$ if $y_{i,1} < S$, and $E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,1} = y_{i,1}, Z_i = 1] \leq E[\mathbf{1}[\nu(Y'_{i,1} - Y'_{i,0}) \geq U_i]|Y_{i,1} = y_{i,1}, Z_i = 0]$ if $y_{i,1} > S$. To see this, note that the inequalities hold pointwise for all $Y_{i,0}$. Since the distribution of $Y_{i,0}$ conditional on $(Y_{i,1}, Z_i)$ is invariant due to random assignment, the result follows from the monotonicity of the integral operator. Those with pessimistic beliefs ($Y_{i,1} < S$) reconsider discontinuing education, and those with optimistic beliefs ($Y_{i,1} > S$) consider discontinuing education.

Before proceeding, we note that assumption 1(ii) is stronger than required for Proposition 1. The previous section shows that if preferences are separable, any conditional choice correspondence consistent with maximization will satisfy cyclic monotonicity. We do not need to assume a tie-breaking rule as in Assumption 1(ii). The crucial assumption is that students behave as if belief changes correspond to expected utility changes and beliefs update in the direction of the signal. In particular, suppose updated beliefs of earnings to continuing education are $Y'_1 = Y_1 + \theta_1(S - Y_1), \theta_1 \in (0, 1)$. If we let $U_1 = Y'_1$ and $U_0 = Y'_0 = Y_0$, cyclic monotonicity implies that $(Pr(D = 1|Y'_1, Y'_0) - Pr(D = 1|Y_1, Y_0))(\theta_1(S - Y_1)) \geq 0$. This is equivalent to Proposition 1 under the assumption of random information assignment. We should also point out that Proposition 1 does not imply that the sign of the treatment effect varies with Y_1. Some subjects might not react to information at all. Depending on the size of this group, estimated treatment effects might flip sign or not. In the presence of unresponsive subjects, Proposition 1 predicts a change in the magnitude of the effect only.

Proposition 1 is useful because it provides a method to estimate how many participants are affected by the information campaign. Let $q_d(Y_d) = f(Y_d, D = d|Z = 0)$ and $p_d(Y_d) = f(Y_d, D = d|Z = 1)$ be the joint densities of prior beliefs Y_d and $D = d$ given $Z = z$. Under Assumption 1, we can estimate the proportion of participants switching from discontinuing education to continuing education by $\int_{Y_1} max\{p_1(Y_1) - q_1(Y_1), 0\}dY_1$, and the proportion of participants switching from continuing education to discontinuing education

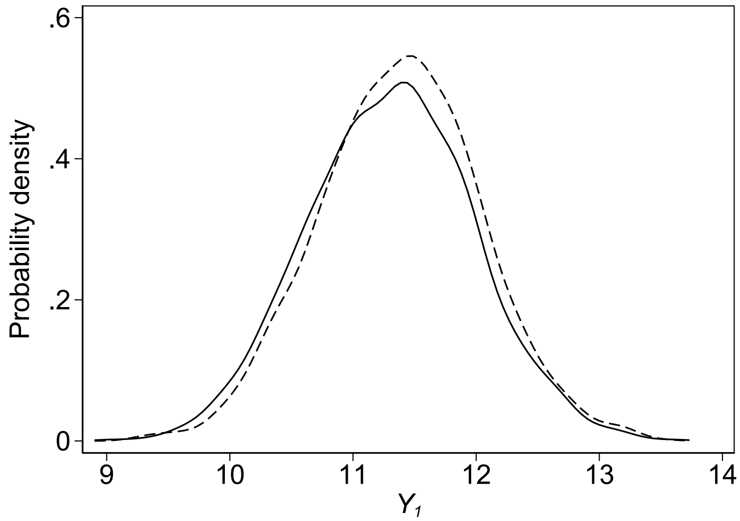


Figure 1: Simulated belief densities under Proposition 1. Prior beliefs on earnings of continuing education by treatment conditional on continuing education. Solid lines for treatment, dotted lines for control.

as $\int_{Y_1} max\{q_1(Y_1) - p_1(Y_1), 0\}dY_1$. The first expression is the number of participants continuing education, and the second is the number of participants discontinuing education.[13] Proposition 1 establishes when ignoring information about alternative options is possible. Since education decisions depend on several alternatives, this provides a lower bound to the effect of information on choices.

To illustrate the use of Proposition 1 to identify heterogeneous responses to information treatments, Figure 1 shows functions $q_1(Y_d)$ and $p_1(Y_d)$ using simulated data. We assume that $\bar{Y_1} = 11.34$, $\bar{Y_0} = 9.87$, $\rho(Y_0, Y_1) = 0.19$, $S = 10.8$, $\theta = 0.8$ and $Pr(D = 1|Y_1(S), Y_0) = (1 + exp(-1 - 5(Y_1' - Y_0')))^{-1}$.[14] Consistent with Proposition 1, $q_1(Y_1) \geq p_1(Y_1)$ if $Y_1 > S$ and $q_1(Y_1) \leq p_1(Y_1)$ if $Y_1 < S$. The average difference in dropout rates for those receiving signal S is 4 percentage points. However, the proportion deciding to stop education due to the signal is 9 percentage points and the proportion deciding to continue education due to the signal is 3 percentage points.

Multiple choices and signals. Information interventions might provide information on several educational alternatives. For instance, Jensen (2010) provided information on pri-

[13]Dahl et al. (2017) show that treatment effects, and characteristics, of compliers and defiers, can be identified if a condition called local monotonicity holds. Local monotonicity states that conditional on a potential outcome, compliers and defiers do not co-exist. The authors propose a method to identify different regions of the potential outcome distribution to estimate treatment effects on compliers and defiers and to characterize them. This condition is violated here since for each set of beliefs, individuals can choose to continue or discontinue education.

[14]The values chosen are based on the data we collected and describe in the next sections.

mary, secondary, and tertiary education, and our study provided information on alternative education tracks: Arts, Commerce, and Science. Information is likely to have a uniform effect if the signal of one of the alternative education options was higher than the subjects' beliefs and the signals of all other alternatives were lower than the subjects' beliefs. Information is likely to have a nonuniform effect whenever subjects have mixed signals. To generalize the approach, define $q_k(Y) = f(Y, D = k|Z = 0)$ and $p_k(Y) = f(Y, D = k|Z = 1)$ for Y the vector of prior beliefs for which a signal is provided. $q_k(Y)$ and $p_k(Y)$ are the joint density functions of prior beliefs given an educational choice k for the control and treatment group. We can estimate the proportion of participants switching to k by $\int_Y max\{p_k(Y) - q_k(Y), 0\}dY$, and the proportion of participants switching away from k as $\int_Y max\{q_k(Y) - p_k(Y), 0\}dY$.

We provide a set of assumptions under which the proposed procedure is valid. Let Y' be a vector of posterior beliefs, Y a vector of prior beliefs, S a set of signals, $D \in \{0, 1, \ldots, K\}$ a choice, and $\eta_i, i = 1, 2$ vectors of parameters. We assume that $\eta_i, i = 1, 2$ are measurable.

Assumption 2: (i) $Y', Y, S, D, \eta_1, \eta_2 \perp\!\!\!\perp Z$, (ii) $Pr(D = k|Y', Y, S, \eta_1, \eta_2) = Pr(D = k|Y', \eta_1)$, (iii) $Y' = b(Y, S, \eta_2) = \{b_k(Y, S, \eta_2)\}_k$ where $b_k(\cdot)$ are functions of (Y, S, η_2).

Assumption 2(i) says that the relationship between beliefs, signals, and decisions is independent of treatment assignment. Assumption 2(ii) is an exclusion restriction implying that treatment effects are mediated only by their effect on beliefs. Assumption 2(iii) says that knowledge of (Y, S, η_2) is enough to determine Y'.

Proposition 2: Under Assumption 2, the proportion of participants switching to k is $\int_Y max\{p_k(Y) - q_k(Y), 0\}dY$, and the proportion of participants switching away from k is $\int_Y max\{q_k(Y) - p_k(Y), 0\}dY$.

Proof: We have that $f(Y, D = k|Z = z) = Pr(D = k|Y, Z = z)f(Y|Z = z)$. We have that $Pr(D = k|Y', Y, S, \eta_1, \eta_2) = Pr(D = k|b(Y, S, \eta_2), \eta_1) = Pr(D = k|Y, Z = k, \eta_2, \eta_1)$. The first equality follows from assumptions 2(ii) and 2(iii). The second inequality follows from the fact that S is fixed for each $Z = z$. So, we have that $Pr(D = k|Y, Z = k) = \int_{\eta_1}\int_{\eta_2} Pr(D = k|Y, Z = k, \eta_2, \eta_1)dF_{\eta_1}dF_{\eta_1} = \int_{\eta_1}\int_{Y'=b(Y,S,\eta_2)} Pr(D = k|Y', \eta_1)dF_{\eta_1}dF_{\eta_1}$. Finally, under assumption 2(iii), $f(Y|Z = 1) = f(Y|Z = 0)$. This implies that $max\{f(Y, D = k|Z = 1) - f(Y, D = k|Z = 1), 0\}$ is the probability density of those switching into k and $max\{f(Y, D = k|Z = 0) - f(Y, D = k|Z = 1), 0\}$ is the probability density of those switching out of k.

Proposition 2 shows that distributional changes can be used to measure flows in and out of educational choices even if students are inconsistent with the behavioral model in Proposition 1 or Section 2. The assumptions do not impose rational information processing. It imposes that posterior beliefs capture all the information needed to decide and that the experiment influences choice only through beliefs. The behavioral model presented in Section

2 implies that cyclic monotonicity should hold for changes in the distribution of choices and beliefs conditional on prior beliefs (see Shi et al., 2018). Such a test requires making additional assumptions about how beliefs affect choices (e.g., linearity). Those assumptions in conjunction with Assumption 2 can be exploited to identify a semi-parametric model of educational choice.

The approach proposed here identifies the joint probability distribution of outcomes in the treatment and control group if subjects have at most three options. It partially identifies this joint probability if four or more options exist. Additional moment conditions for identification can be derived from joint events like $D \in \{k, j\}$. This implies that the model has empirical content. The model is rejected if no joint probability distribution satisfies these constraints. Some potential sources of failure can be tested. For instance, assumptions 1(ii) and 1(iii) can be directly tested if data on prior and posterior beliefs are collected. We discuss evidence consistent with these assumptions in the paper.

We provide an example illustrating this extension and adapt this approach in the results section. Suppose a student receives a signal S_1 of the return to continuing education and a signal S_0 of the return to discontinuing education. Suppose the updated beliefs are $Y_1' = Y_1 + \theta_1(S_1 - Y_1)$ and $Y_0' = Y_0 + \theta_0(S_0 - Y_0)$ for $\theta_i \in (0, 1), i = 0, 1$. The updated return to education is now $Y_1 - Y_0 + \theta_1(S_1 - Y_1) - \theta_0(S_0 - Y_0)$. Cyclic monotonicity requires that $(Pr(D = 1|Y_1', Y_0') - Pr(D = 1|Y_1, Y_0))(\theta_1(S_1 - Y_1) - \theta_0(S_0 - Y_0)) \geq 0$. This implies that those with $S_1 > Y_1$ and $S_0 < Y_0$ will increase the probability of continuing education and those with $S_1 < Y_1$ and $S_0 > Y_0$ will decrease the probability of continuing education. Students with other prior beliefs will have mixed signals and their decision will vary depending on the value of parameters. Identifying heterogeneous responses requires conditioning decisions on the joint distribution of prior beliefs in this case. Conditioning on both prior beliefs might be advisable if one expects beliefs to be updated by introspection or information outside the experiment.

We recognize that, in practice, the assumptions supporting Proposition 1 are likely to fail. There is heterogeneity in the exact prior belief at which participants take a signal to be good or bad news. For instance, men and women likely face different labor market conditions and might adjust signals to reflect the likelihood of employment. Participants might also make mistakes in reporting their own priors that are corrected under closer inspection.[15] The beliefs collected in experiments might be far from those used in practice and subject

[15] We can model belief revisions as the existence of *personal* signals that are not observed by the researcher. Subjects update these updated beliefs, making measured priors beliefs noisy measures of actual priors. For Proposition 1 to hold, we need these *personal* signals to be uncorrelated with the treatment, and for the signal provided by the experimenter to be stronger. We provide evidence of learning in the Results section. However, we cannot directly test if *personal* signals are affected by treatment assignment.

to measurement error. Finally, extant research, which we verify in our data, shows that non-pecuniary considerations are important in education and labor decisions. This implies that reactions to earning information might be muted or distorted. Ex-ante estimates of discounted utility streams using prior data can therefore be used to estimate responses to information based on utilities rather than priors on earnings. We use earning beliefs as a first step to illustrate the feasibility of the proposed approach and avoid making parametric assumptions. The method can be helpful if these assumptions are approximately correct. We implement the approach given these cautionary statements.

3 Experimental design and implementation

We provide background on the study setting and details about the data collection.

3.1 Background

The secondary education system in Nigeria is divided into junior high school and senior high school. After spending six years in primary school, students attend three years in junior high school and can spend three years in senior high school. At the end of junior high school, students have an opportunity to choose between an academic or a vocational curriculum. Students wishing to proceed with the academic option in senior high school have three curriculum options: arts, commercial/social science, and science. The vocational track also provides students with different areas for specialization. The diversity of the course curricula in senior high school gives students the opportunity to choose their future career paths. Students are exposed to both mainstream academic courses and vocational courses. At the end of junior high school, or grade 9, students take a statewide examination—the Basic Education Certificate Examination (BECE)—which allows the transition to the next level of schooling.

3.2 Overview

The present study recruited students in their last year of junior high school who had to decide whether to continue to senior high school, go to vocational school/take an apprenticeship, or drop out of school entirely. The study took place in the city of Ibadan, the capital of the state of Oyo, Nigeria and Nigeria's third most populous city (3.2 million). The study was conducted with the approval of the State of Oyo's Ministry of Education, Science, and Technology. The experiment had five stages: recruitment, baseline data collection, information provision in treated schools, collection of endline data, and collection of administrative data on educational choices.

In the first stage, schools were recruited and consent was obtained. Students were assigned to three experimental conditions: treatment, impure control and pure control. All students were asked to fill out a baseline survey that collected basic demographic information, attitudes towards schooling, a proposed curriculum/track choice, career aspirations, and participantive expectations. Students in the treatment group and impure control group were then asked to respond to three distinct information elicitation tasks. Following this, students in the treatment group were provided with information on average earnings for the different tracks and college admission probabilities. Then students in the treatment group and the impure control group were asked to respond to the same expectation questions a second time. This design follows Wiswall and Zafar (2015a)'s design for college choice. The survey ended by collecting information about (hypothetical) time and (paid) risk preferences.

3.3 Measures of expectations

The questions regarding self-beliefs were questions about educational outcomes, i.e., the students' chances of ending their education with junior high school, going to a vocational school or apprenticeship, dropping out of senior high school, finishing senior high school, dropping out of college, and finishing college together with a curriculum track. These questions also included predicted probabilities of working full-time at a job related to a specific major and earnings after finishing schooling, at ages 30 and 50. We also asked for their estimated probabilities of earning at least 50,000 Nigerian Naira (N50,000), N100,000, and N200,000. Similar questions were asked regarding their beliefs about the population, with reference to a typical student. Figure 2 provides the wording of one of the self-belief elicitation tasks regarding educational attainment.

What are the chances that you will	**Number**
go to art class?	
go to science class?	
go to commercial/social science class?	
go to vocational school after JSS3?	
drop out of school after JSS3?	
TOTAL: THE TOTAL SHOULD ADD UP TO 100	

Figure 2: Example of self-belief elicitation

The information treatment included statistics about the earnings and labor supply in Nigeria and population-level college acceptance rates and college choices. This information came from the Joint Admissions and Matriculation Board (JAMB) and Stutern (2018).[16] To the best of our knowledge, this is the most up-to-date information on admissions, graduation, and earnings for recent graduates in Nigeria.

Section 3.5 presents the information provided in the study. The instrument used in this study was a simplified version of the approach of Wiswall and Zafar (2015a), who used this kind of information to estimate human capital accumulation models; see Haaland et al. (forthcoming) for a review of the approach. Importantly for us, the belief information we collect allows testing if students update their beliefs when the information is provided. Verifying that the students update their beliefs is a necessary condition for the rational use of information. It is also a direct way to test the behavioral assumptions introduced in Section 2.2.

3.4 Elicitation of willingness to pay for information

We elicited the WTP for different pieces of information using a multiple price list (MPL) that is a discretization of the Becker-DeGroot-Marshack (BDM) incentive-compatible mechanism. The MPL included prices from N0 to N200 in increments of N25. Participants were asked to respond to 3 MPLs. The first one asked for their WTP for information on college admissions, the second asked for their WTP for information on wages by curriculum track, and the third MPL asked for their WTP for both pieces of information. This was done to test for differences in WTP for the different types of information and to check for adherence to the law of demand: more information should be valued (weakly) more.

To embed the WTP elicitation in the information RCT, we drew prices from the binary set $\{0, 250\}$. These draws were fixed at the classroom/school level in order to avoid the expected spillover effects if randomization were done at the individual level. As a consequence, either all of the students in a school were assigned to the information treatment group or they were all were assigned to a no-information group. Since we could not ask students to pay for information with their out-of-pocket money, we provided all the students in our study with N200 that they could use in different experimental tasks, including the WTP tasks. They were told that payments would be calculated based on their choices in one of the tasks chosen at random.[17] Providing participants with money is consistent with common

[16] A total of 5,219 Nigerian graduates who graduated during the years 2013–2017 completed the survey. The data collection took place between February 8 and May 15, 2018. The survey was hosted using Google Forms, and Stutern.com recruited respondents via email and social media sites. To account for graduates in marginalized locations, an offline version of the survey was conducted in five states (Edo, Enugu, Ibadan, Imo, and Kaduna).

[17] In practice, one of the three MPLs was chosen at random to determine the cost of information. If the price was \$0, we provided all pieces of information. We did this because we would not be able to detect the effects of different sets of information since the treatments were assigned at the school level to avoid

practice in experimental economics. Note, however, that this imposed a budget constraint on the participants. The amount that we provided to each student is about 1.1% of Nigeria's minimum wage and enough to cover a student lunch. Our study balances the need for the salience of payoffs and the risk associated with transferring money to minors.

Previous research using the BDM shows that both the distribution of prices (Mazar, Koszegi and Ariely, 2014) and the upper bound of the distribution of prices (Bohm, Linden and Sonnegard, 1997) can affect elicited valuations. To avoid these issues in our experiment, we indicated to participants that the prices could take values as low as N0 and higher than N200. However, we cannot test whether the set of prices that was provided altered the average valuation of information. Different approaches to eliciting valuations that would further minimize these issues (e.g., Allcott, Braghieri, Eichmeyer and Gentzkow, 2020; Mosquera, Odunowo, McNamara, Guo and Petrie, 2020) were not feasible because students needed to be provided with money to participate.

Whenever price randomization is implemented at the individual level, instead of at the cluster level as in our experiment, it is possible to disentangle the effect on behavior of the treatment status and the price paid (see Berry et al., 2020). That is, it allows to identify screening and sunk-cost effects separately. Sunk-cost bias is unlikely in our experiment since students paid a price of zero for information. Also, as we will show later, the behavior of those not receiving information is not affected by their WTP being elicited. We note that our study does not have treatment where information is given and WTP is not elicited. This treatment would allow to test if elicitation itself has an added effect on behavior beyond information. Given that Berry et al. (2020) find no evidence of sunk-cost effects in their experiment, we interpret our results as mainly identifying screening effects. Finally, students might value information even if it does not affect their behavior. This might be due to risk attitudes (e.g. Kreps and Porteus, 1978) or curiosity. Our results depend on the assumption that compliers with our treatment are not more intrinsically curious or differ in their preference for early resolution of uncertainty. While these alternative explanation are testable, they are beyond the scope of our study.

3.5 Randomization and implementation

Information intervention: The intervention provided information to students in randomly selected schools. Students in the treatment schools received information about average wages, the percent working full time, and the percent earning more than N50,000 and N100,000 for the different curricula/tracks, as well as the proportion of males (females) who applied and

potential contamination.

were admitted to college across the three tracks. Figure 3 shows how the information was presented; it was done in this way to make it easy to understand. We consulted with the State's Ministry of Education to ensure that the information was accurate.

In the state of Oyo, there are over 600 public secondary schools. The sample of schools includes 115 coeducational junior high schools out of a universe of 133 in Ibadan city. These schools are evenly distributed in four areas of the city across the 5 local government areas in the city. The subset of schools randomized into the study were visited by enumerators carrying official letters from the government, our IRB approval letter, a study overview/permission letter to obtain permission to visit the school, and a proposed date and time to visit.

The pre-registered study planned for 32 schools: 16 in the treatment group and 16 in the control group (planned number of observations = 5,200).[18] We divided the control group so that half would be asked the belief questions and the other half would not. This resulted in a pure control group, an impure control group, and a treatment group, which allowed us to test for rationality as well as whether asking belief questions to those who did not receive information would affect their behavior. We planned for 16 schools in the treatment group, 8 schools in the pure control group, and 8 schools in the impure control group. The intervention was implemented with students who would take their exams in June/July 2020 and needed to decide which track to choose by the beginning of senior high school in September 2020. The study was designed to detect a 5 percentage point change in dropout rates with a power of 0.8 at a 5% significance level.[19]

The first stage of the study was conducted between November 8 and December 3, 2019. To account for potential attrition of schools from the experiment, we decided to recruit additional schools. We retain the 36 schools that completed the data collection. There were 18 schools in the treatment group (N = 1,925), 6 in the impure control group (N = 658), and 12 in the pure control group (N = 1,054). The main deviations from the original protocol were a reduction in the time that schools allowed for implementing the study[20] and the delay in student examinations until August 2020 due to Covid-19. Figure 4 presents the design and implementation of the study in a graphic form.[21]

Table 1 presents basic statistics for the sample and its comparability across the three groups.[22] The average age of participants was 14 years. There were slightly fewer females

[18] AEA RCT Registry number AEARCTR-0004839.

[19] We used administrative data on dropout rates to calculate the intraclass correlation (0.02).

[20] We reclassified three schools from the impure control to the control group since belief data were not collected due to time constraints.

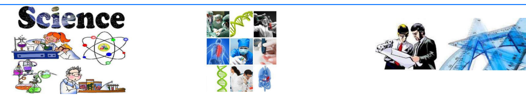
[21] We conducted a pilot test in 2018 to assess the feasibility of WTP elicitation techniques. We visited three schools for a total of 195 students. Two of these schools were single-sex and not included in this study. The WTP elicitations from the pilot are comparable to those in this study. Results are available from the authors upon request.

[22] All classrooms in the last year of junior high school of the participating schools were visited by enumer-

Admissions information

Arts

- 15% of girls applied to Arts in university
- 10% of boys applied to Arts in university
- Of the girls that applied, 32% got admission
- Of the boys that applied, 30% got admission



Science

- 53% of girls applied to science in university
- 60% of boys applied to science in university
- Of the girls that applied, 26% got admission
- Of the boys that applied, 29% got admission

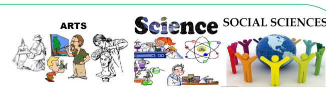
Social-sciences

- 33% of girls applied to social-science in university
- 30% of boys applied to social-science in university
- Of the girls that applied, 26% got admission
- Of the boys that applied, 27% got admission

Source: JAMB and CINFORES (2017)

Salary information

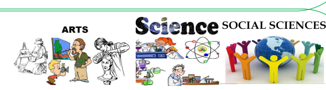
Female



This information is from a survey of university graduates in Nigeria. Among recent ***female*** graduates from a university with a Bachelor's degree in each of the above fields:

	Arts	Science	Social sciences
The percentage that are working full time is	**44.6%**	**45.6%**	**45.9%**
The average monthly salary of those that are working full time is	**₦59,158**	**₦68,740**	**₦64,696**
The percentage of those who are working full time that earn more than N50,000 monthly	**45.74%**	**56.2%**	**52.23%**
The % of those who are working full time that earn more than N100,000 monthly is	**5.26%**	**7.85%**	**4.86%**

Male



This information is from a survey of university graduates in Nigeria. Among recent ***male*** graduates who just graduated from a university with a Bachelor's degree in each of the above fields:

	Arts	Science	Social sciences
The percentage that are working full time is	**48.54%**	**50.4%**	**49.35%**
The average monthly salary of those that are working full time is	**₦53,100**	**₦77,849**	**₦71,924**
The % of those who are working full time that earn more than N50,000 monthly is	**44%**	**60%**	**53.16%**
The % of those who are working full time that earn more than N100,000 monthly is	**6%**	**25.4%**	**8.42%**

Among recent university graduates who received a Bachelor's degree in the above fields

	Arts	Science	Social sciences
The percentage of those who are women is	**65.52%**	**35.64%**	**56.62%**

Source: JAMB and CINFORES (2017)

Figure 3: Information provided

in the sample than males. About 40 percent declared themselves to be Christian and had roughly four siblings. Almost 80 percent of the students lived with both of their parents. Around 10 percent of the students declared that they had repeated at least one grade. The three groups are balanced in all of the variables we checked, except for the number of females. The pure control group had slightly more females than the information treatment group.

Information treatment (N = 1,925, 18 schools):

Background survey → Beliefs → WTP → Info → Beliefs → Preferences → Field outcomes

Impure control (N = 658, 6 schools):

Background survey → Beliefs → WTP → Beliefs → Preferences → Field outcomes

Pure control (N = 1,054, 12 schools):

Background survey → Beliefs → Preferences → Field outcomes

Figure 4: Study implementation

Notes: A subset of the expectation questions and preferences were collected from all participants. Preferences were collected at the end of the survey. Students were provided with N200 to be used in the willingness to pay (if applicable) and preference elicitations.

Variable	(1) Cntrl (0)	(2) Info (1)	(3) Imp. Cntrl (2)	(4) (1)-(0)	(5) (2)-(0)	(6) (1)-(2)
Age	13.978	14.018	14.189	-0.041	0.083	0.007
	(1.591)	(1.398)	(1.549)	(0.141)	(0.183)	(0.203)
Female	0.498	0.448	0.465	-0.037	-0.029	-0.029
	(0.500)	(0.497)	(0.499)	(0.028)	(0.050)	(0.039)
Mother's years of education	12.511	12.484	12.644	-0.079	0.443	-0.457
	(3.677)	(3.540)	(3.216)	(0.264)	(0.359)	(0.310)
Christian	0.394	0.364	0.394	-0.030	0.021	-0.052
	(0.489)	(0.481)	(0.489)	(0.058)	(0.097)	(0.074)
No. of siblings	3.985	3.892	3.930	-0.072	-0.075	-0.030
	(3.118)	(2.636)	(2.290)	(0.187)	(0.203)	(0.205)
Two-parent household	0.786	0.785	0.757	0.010	-0.014	0.007
	(0.411)	(0.411)	(0.429)	(0.019)	(0.023)	(0.026)
Grades	0.617	0.616	0.605	0.004	-0.005	0.001
	(0.144)	(0.141)	(0.135)	(0.011)	(0.016)	(0.013)
Repeated a grade	0.097	0.093	0.156	-0.027	0.057	-0.048
	(0.296)	(0.290)	(0.363)	(0.022)	(0.042)	(0.033)
Observations	1,054	1,925	658	3,637	1,528	2,767

Table 1: Characteristics by treatment group

Notes: Columns 1-3 show the average by treatment group. The numbers in parentheses are used to indicate pairwise comparisons between the groups in columns 4-6. Missing data items are replaced with the mean of the variable over the entire sample.

ators.

3.6 Outcomes collected

The study has two main outcome variables: self-beliefs about educational choices and actual educational choices. The self-beliefs were collected before and after the information intervention, and the educational outcomes were collected more than a year after the baseline survey. We present the outcomes below.

Probability of enrolling in school: During the baseline study, we asked students to report their estimated probability of enrolling in school beyond junior high school. We also asked for their estimated probability of choosing different senior high school and college tracks (arts, commerce, science, or vocational education) as well as dropping out of school after junior high school, during senior high school, and during college. These data were used to investigate the way students updated their self-beliefs upon receiving information.

Attendance/dropout rates: We obtained administrative data from schools on attendance and enrollment. In particular, we recorded whether a student took a junior high school exit exam, the grade obtained on the exam, and whether they registered for senior high school. For students to qualify for admission to a senior high school and higher education, nationwide examinations are held each year. Because exam scores determine a student's future educational choices, schools tend to stress memorization of facts rather than creative problem-solving. Students are required to pass at least six subjects to proceed to senior high school at the same institute or a different institute.

Curriculum choice: We obtained administrative data from schools on high school track choices. The curriculum tracks include arts, commerce, and science. All senior secondary students are required to study English, mathematics, one science, and one Nigerian language course.[23] The remaining participants are electives and are selected based on the students' interest in either the sciences, the social sciences, or the arts. We note that it is too early in the study to know the actual career paths taken by students who registered in senior high school. We are also not aware of any data showing a correlation between curriculum choices in senior high school and career paths.

The state of Oyo does not have a centralized system with all students' data. In order to minimize potential biases due to non-response, we visited each high school in the study to collect information on registration. This allowed us to cross-check whether students changed schools after junior high school as well as their decisions. We also conducted a phone survey for all students who were not found in any of the school records. This procedure allowed us to determine the outcomes for over 95% of the sample. We do not find significant differences in missing data across the treatments.

[23]Science is not required for non-science tracks.

4 Results

4.1 The use of information

In this section we describe how students use the provided information. We first discuss the extent to which students update their beliefs. Then we evaluate whether their career choices reflect selection based on earnings. Next, we discuss the effect of the provided information on beliefs about these career choices and the effect of the provided information on the decision to continue to senior high school. Finally, we use belief data to uncover heterogeneous responses to information and to characterize their information seeking behavior.

4.1.1 Information updating

We use a Bayesian learning model to evaluate information updating. In this model, a student has a normally distributed prior belief $prior \sim N(\mu_0, \sigma_0^2)$. The variance of the prior belief, σ_0^2, captures the uncertainty of this belief. A signal is drawn from the true distribution of the variable that is distributed: $signal \sim N(\mu, \sigma^2)$. A Bayesian agent will update her prior according to the following formula:

$$posterior = (1 - \theta)prior + \theta signal,$$

where $\theta = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}$. Rearranging terms, we obtain:

$$posterior - prior = \theta(signal - prior).$$

This provides an empirical framework for assessing the confidence students place on their beliefs. More confident students will update information less or will have lower values of θ.

Figure 5 shows the average prior by educational choice and age. Numbers 20, 30, and 50 refer to the age at which expected earnings are reported. For each educational choice, prior beliefs on earnings increase with age. Prior beliefs are also increasing in educational choice, earning beliefs are lowest for Junior High School and largest for those with a college degree in sciences. Arts, Commerce and Science refer to earnings to different college degrees. Figure 6 shows the density functions of the difference between the log of earning beliefs at 30 years of age and the log of the gender-specific signal provided. The densities correspond to the earning if a major in Arts, Science, or Commerce is followed. We see there is a high degree of heterogeneity in beliefs and a majority of participants over-estimating earnings. This confirms that the necessary condition for nonuniform response to treatment is satisfied in our setting.

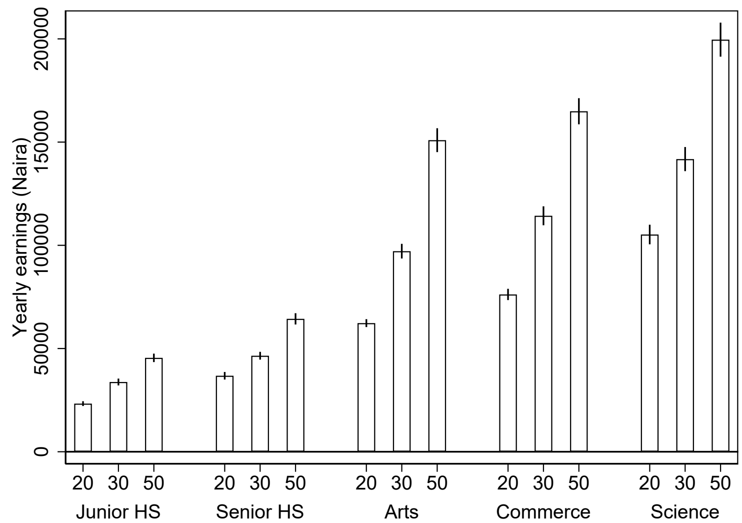


Figure 5: Prior beliefs by educational choice and age

Our design allows testing if information updating is due to new information that has been received or due to a reversion to the mean. For instance, students might correct reported expectations after noticing that they made a mistake or simply upon reflection. We can address this issue by comparing information updating between the information treatment and impure control conditions.

Table 2 presents estimates of a regression of the change in the log of earnings beliefs as a function of the difference between the log of the information provided on earnings and the log of the first set of earnings beliefs.[24] We interact this variable with an indicator of having received the information and estimate these regressions for the subset of beliefs closest to the data provided to students (i.e., 21 and 30 years of age). The parameters are identified because male and female students have different relevant information. The results are similar if the regression is pooled across all questions to provide additional variation. We observe that the net effect of information on beliefs ranges from 0.12 (Science at 30) to 0.24 (Commerce at 30). This is comparable to the estimates in Hjort et al. (2021), who found the effect of signals to be between 0.26 and 0.37 (see Table 3 of their paper).[25]

Table 3 reproduces this analysis on the set of probability beliefs. This provides a second test of the ability of students to process the information that is provided. We find that

[24]We winsorized the data at 1% to avoid extreme reports. The log of the earnings beliefs is close to a normal distribution. We did not collect information on beliefs of being admitted to college, so we cannot conduct a similar analysis for these beliefs.

[25]The estimates on information updating for those who did not receive any information suggest that the belief data in this population are measured with error. Fuster et al. (2022) also observe this phenomenon, although to a lesser extent, when analyzing the effect of information on beliefs about housing prices.

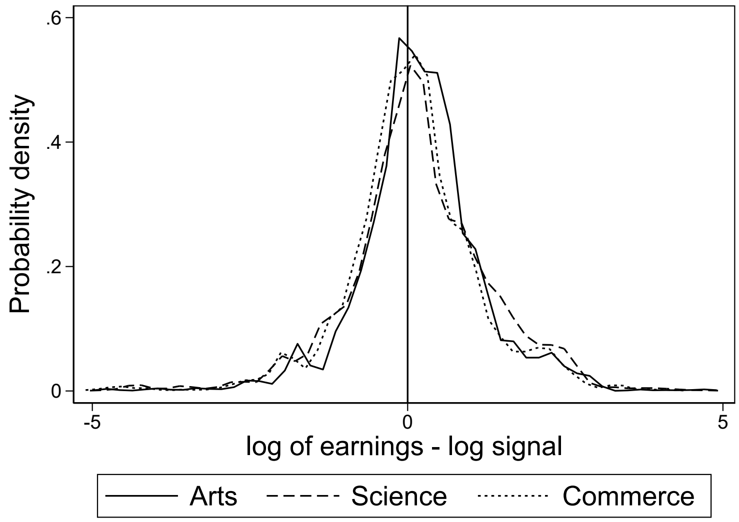


Figure 6: Distribution of earnings beliefs at 30 around signal

beliefs on the probability of earning a certain salary are less responsive to information.[26] The parameters associated with the signal are a fraction of those estimated for beliefs about earnings. We confirm that participants update information consistent with Assumption 1(i).

Table 2: Expectation updating

	Salary at 21			Salary at 30		
	Arts	Comm	Science	Arts	Comm	Science
(signal-prior) × Info treatment	0.212***	0.198***	0.239***	0.234***	0.245***	0.126***
	(0.046)	(0.045)	(0.045)	(0.043)	(0.046)	(0.043)
Observations	2207	2206	2187	2228	2226	2202
Adj R2	0.364	0.340	0.340	0.369	0.308	0.348

* p<0.10, ** p<0.05, *** p<0.010

Notes: Salary at 21 for Arts refers to the answer to the question: "Imagine that you enrolled in the arts track and studied one of the arts courses at university, and tell me how much you think you would be paid monthly if you have just graduated and now work full time?" The labels for the other columns follow the same pattern. The dependent variable is the difference between the belief elicited the second time and the belief elicited the first time. Beliefs are winsorized at 1% and expressed in logs. "Info treatment" equals 1 if the participant was provided with information, and 0 otherwise.

4.2 The effect of information

The analysis so far shows that information is used for updates and that career choices reflect earnings considerations. We now discuss the effect of information on perceived future

[26]This could partly be due to the fact that probability beliefs are not necessarily distributed normally, and therefore the learning model is inadequate for these data.

Table 3: Threshold probability updating

	Pr(Salary at 30<50,000)			Pr(Salary at 30<100,000)		
	Arts	Comm	Science	Arts	Comm	Science
(signal-prior) × Info treatment	0.146***	0.083*	0.065	0.180***	0.152***	0.205***
	(0.049)	(0.048)	(0.047)	(0.049)	(0.050)	(0.046)
Observations	2133	2115	2121	2115	2108	2105
Adj R2	0.324	0.345	0.369	0.391	0.364	0.334

* p<0.10, ** p<0.05, *** p<0.010

Notes: The first column for Arts refers to the answer to the question: "What is the percentage chance that if you were working full time you would earn at least N50,000 monthly if you graduated from the arts track at university and were 30 years old?" The labels of the other columns follow the same pattern. The dependent variable is the difference between the belief elicited the second time and the belief elicited the first time. "Info treatment" equals 1 if the participant was provided with information, and 0 otherwise.

decisions and actual decisions.[27]

To measure the effect of information on career choices we follow Wiswall and Zafar (2015a), who estimate changes in beliefs about career choices as a function of changes in relative earnings. Specifically, let $\pi_{k,i}$ be student i's belief that she will choose career k. We can define the log-odds of choosing k over $\tilde{k}$ as $ln\pi_{k,i} - ln\pi_{\tilde{k},i}$. Let $w_{k,i}$ be students i's belief about the earnings associated with career k. We can define the relative earnings with respect to option $\tilde{k}$ as $lnw_{k,i} - lnw_{\tilde{k},i}$.

A simple regression of $ln\pi_{k,i} - ln\pi_{\tilde{k},i}$ on $lnw_{k,i} - lnw_{\tilde{k},i}$ is likely to be biased, since career preferences and abilities are likely not independent. However, we can take advantage of the fact that the information intervention did change expectations to estimate the following regression:

$$(ln\pi'_{k,i} - ln\pi'_{\tilde{k},i}) - (ln\pi_{k,i} - ln\pi_{\tilde{k},i}) = \beta_0 + \beta_1[(lnw'_{k,i} - lnw'_{\tilde{k},i}) - (lnw_{k,i} - lnw_{\tilde{k},i})] + \nu_k + \Delta\epsilon_{k,i},$$

where $'$ indicates beliefs elicited after the information is provided, ν_k is a fixed effect for choice k, and $\Delta\epsilon_{k,i}$ captures changes that are uncorrelated with beliefs about earnings. The first-difference regression eliminates unobservable differences across students.

Table 4 provides estimates of this regression using data from the information treatment group.[28] We use beliefs about earnings at 30 years of age since these are closest to the information provided. We provide estimates using all of the participants in the information

[27] Section A.2.1 provides further evidence of the quality of the beliefs data following Arcidiacono et al. (2020).

[28] We use junior high school as the reference group and consider senior high school, arts, commerce, and science as alternatives. Wages are winsorized at 1% to eliminate extreme values. Beliefs about career choices are bounded between 0.01 and 0.99. The results are qualitatively similar if we use different bounding thresholds.

treatment group and only the participants in the information treatment group whose WTP was monotonic on the amount of information acquired. We find a moderate estimate of career choice elasticity. The elasticity is 0.06 for the overall population and 0.10 for the population with monotone preferences for information. Preferences for information are monotone if the WTP for information is (weakly) increasing on the amount of information. Section A.2 describe the data on WTP.

The estimates increase to 0.165 and 0.252 when we correct for measurement error.[29] For comparison, Wiswall and Zafar (2015a)'s estimate for a sample of New York University college students is 0.275 (see Table 6 in their paper). Given that the perceived returns of continuing education after information is obtained are about half of the initially perceived returns, these estimates suggest a 2.5 percentage point decrease in each educational choice other than junior high school if preferences for these alternatives are uniformly distributed.[30] Equivalently, they predict a 2.5 percentage point increase in senior high school dropout rates.

These results are consistent with Assumption 1(ii).

Table 4: Changes in self-beliefs about career choices as a result of information provision

	Accounts for measurement error?			
	No		Yes	
	All	Monotonic	All	Monotonic
Log of wage at 30 (rev.)	0.060**	0.101***	0.165*	0.252**
	(0.030)	(0.037)	(0.087)	(0.098)
Observations	4726	3227	4726	3227

* p<0.10, ** p<0.05, *** p<0.010

$^{+}$ clustered at the individual level

Notes: The dependent variable is the change in the log odds of choosing a career path between the first and second round of belief elicitation. The choices considered are senior high school, arts, commerce, and science. The baseline choice is junior high school. The explanatory variable is the change in relative earnings at 30 years of age between the first and second rounds of belief elicitation. The baseline comparison is junior high school. Dummies for each choice are included. The estimation excludes self-beliefs that add up to less than 95 or more than 105 in either the first or second round of belief elicitation. The results are less precise if these data are included, but the qualitative results are the same.

[29] To correct for measurement error, we calculate the reliability of changes in relative wages using data collected in the impure control group. Specifically, let x be the change in relative wages. We calculate the reliability coefficient as $\frac{Var(x|Control)}{Var(x|Information)}$, where $Var(x|Control)$ is the variance of x in the impure control condition and $Var(x|Information)$ is the variance of x in the information treatment condition.

[30] This calculation assumes that the pre-intervention probability of choosing to finish junior high school only is 10% and the probability for the other four alternatives is 22.5%.

4.3 Treatment effects

So far, the analysis indicates that students use the information provided in expected ways. Beliefs about earnings are updated if the information is provided, and beliefs about career paths are updated once the earnings beliefs are updated. Next, we look at the information intervention's effect on field outcomes.

Table 5 shows linear probability models for the effects of the group treatments on not pursuing senior high school. We find that students in the pure and impure control group are respectively 3.9 and 3.7 percentage points less likely not to pursue senior high school. This difference is not significant (p-value = 0.9442). The effect is 3.7 if we combine both control groups into one control group.[31] The estimated effect is similar if we account for non-responses.[32] This effect is within the 90 percent confidence interval of the predictions using the estimates in Table 4. The estimated effect of information on education is large. The percentage of students not continuing to senior high school is 9 percentage points in the control groups. We conclude that information significantly affects students' reported and actual decisions in our study. Importantly, estimates using the elicited beliefs and the field outcomes are compatible.

Columns (3) and (4) of Table 5 test whether the size of the information treatment effect varies with the WTP for information. The regressions use the WTP for both pieces of information as a moderator.[33] As theoretically predicted, those who were more willing to pay for information reacted to the information treatment more strongly. While the effect on the treated is 3.7 percentage points, the effect on those willing to pay N200 for both pieces of information is almost 10 percentage points. Column (4) in the Table shows that the result is robust to the inclusion of additional moderating variables. Table 5 allows us to calculate how much WTP is likely to increase due to an increase in expected behavior. We have that a 100% increase in WTP is associated with a 70% increase in the probability of discontinuing education. Inversely, a doubling of the probability of discontinuing education increases WTP for information by 143%.[34]

[31] We implement the randomization test proposed by Canay, Romano and Shaikh (2017) to account for possible biases due to a small number of clusters. The test requires that treatment effects can be estimated in each cluster, so we grouped adjacent clusters into one to implement the test. This reduced the number of clusters from 36 to 18. The estimated treatment effect is significant at the 10 percent level (p-value = 0.077). For robustness, we estimate the p-values for 1,000 random pairings of clusters. The average p-value is 0.077 (s.e. 0.033.)

[32] The estimated treatment effect is 0.0343 (s.e. 0.0215) using an inverse probability weighting correction. We predict attrition using age, sex, mother's years of schooling, Christianity, average grades, and indicators for two-parent households and having repeated a grade.

[33] Figure A.3 graphically shows the implied information treatment effects of the model (using estimates from Table 5, column (3)).

[34] We reject the hypothesis that the coefficient on the interaction term of information treatment and WTP

Table 5: Treatment effects and the value of information continuing to senior high school

	(1)	(2)	(3)	(4)
Impure control	0.038			
	(0.028)			
Pure control	0.037*			
	(0.020)			
Info treatment		-0.037*	0.031	0.003
		(0.020)	(0.033)	(0.061)
WTP/100			0.038***	0.049***
			(0.007)	(0.011)
WTP/100 × Info treatment			-0.063***	-0.075***
			(0.014)	(0.017)
Constant	0.869***	0.906***	0.863***	0.891***
	(0.017)	(0.010)	(0.026)	(0.031)
Observations	3473	3473	2279	2135

* p<0.10, ** p<0.05, *** p<0.010

Standard errors in parentheses. Clustered at the school level.

Notes: The dependent variable is equal to 1 if a student does not register for senior high school and 0 otherwise. The estimates are marginal effects obtained from a Probit regression. The second column is estimated using data from the information treatment and impure control groups (24 instead of 32 schools). The third column includes a dummy for female, two-parent household, being suspended, repeating a grade, averages grades, average investment in the paid lotteries and their interactions with the information treatment. The estimates including additional covariates are qualitatively similar.

4.4 Compliers and defiers

Following Imbens and Angrist (1994), we divide the population between always takers (AT), never takers (NT), compliers (C), and defiers (F). Always takers dropout of school in the treatment and control conditions. Never takers continue education in the treatment and control group. Since the treatment effect increased dropout rates, we assume compliers dropped out of school in the information treatment and remain in school in the control group. Defiers continue education in the information treatment and dropout of school in the control group.

We start by describing the characteristics of response types under the assumption of monotonicity (see Table A.3). The willingness to pay for information (WTP) of compliers is significantly different from those of always takers and never takers. Under monotonicity, it is possible to identify the characteristics of compliers both in the Control and in the Treated group (e.g. Heckman and Pinto, 2018). If the condition holds, we would expect that the characteristics of compliers will be similar in both groups due to random assignment. Table 6 provides such comparison. We observe that estimates of WTP are significantly higher for compliers in the Control group than in the Treated group. The significant difference in estimated WTP for compliers across treatment conditions, and the violation of the bounds of WTP for information, cast doubt on the assumption of monotonicity.

Table 6: Willingness to pay for information by compliers under monotonicity

	In Control	In Treatment	p-value	Combined
Admission & Earnings	345.80	132.71	0.01	239.26
Admission rates	292.33	132.21	0.02	212.27
Earnings	251.04	95.43	0.02	173.24

To fully assess if these results are due to a non-uniform response to information, we follow the procedure suggested in Section 2.2. We now describe how we construct variable Y_1, the prior beliefs on life earnings from continuing education. The experiment collected prior beliefs on the earnings for ages $k =$\{at graduation, at 30 years of age, at 50 years of age\} corresponding to dropping out of senior high school, finishing senior high school, and graduating with a degree in arts, science or commerce. We construct a measure of earnings corresponding to continuing education, Y_1, as the natural log of the arithmetic mean of the winsorized earnings beliefs of finishing senior high school, graduating with a degree in

is equal to 0.1 (p-value = <0.001).

arts, science or commerce at graduation, 30 years of age, and 50 years of age. While we collected data on the likelihood that each one of these alternative education paths is taken and the probability of full employment by educational choice, the data is incomplete. We use an average of all earnings beliefs as a way to reduce measurement error. As mentioned in Section 2.2 there is not an exact correspondence between the signal provided in the experiment and prior beliefs. We are therefore agnostic about the exact threshold used by each participant. However, Proposition 1 should hold upon aggregating across individuals unless the distribution of individual threshold beliefs is sufficiently distinct from the signals. This measure of earnings is balanced across treatments.[35]

Proposition 1 in Section 2.2 shows that under Assumption 1 either $Pr(D = 1, Y_1 \in A|Z = 1) \geq Pr(D = 1, Y_1 \in A|Z = 0)$ or $Pr(D = 1, Y_1 \in A|Z = 1) \leq Pr(D = 1, Y_1 \in A|Z = 0)$ might fail to hold for all $A \in B_{Y_1}$, where B_{Y_1} is a collection of Borel sets on the support of Y_1. We implement the tests proposed by Kitagawa (2015) and its modification proposed by Sun (2022).[36] We should remark that these tests are based on the joint hypothesis that monotonicity holds for those who continue to senior high school and those who are discontinuing education. We implement the tests in the subsample of those continuing education since Proposition 1 applies to this subsample only.[37]

We define $p_1(Y_1) = f(Y_1, D = 1|Z = 1)$ and $q_1(Y_1) = f(Y_1, D = 1|Z = 0)$ as in Section 2.2. Figure 7 shows these density probability functions for the decision to continue education for all participants with belief data. The solid lines represent densities for those assigned to the Information condition and the dotted lines represent densities for those assigned to the Control condition. Under Proposition 1, these densities can cross for those choosing to continue to senior high school.

Figure 7 shows a pattern predicted by Proposition 1. Table 7 provides a formal test for the failure of monotonicity as suggested by Kitagawa (2015) and the modification of the test suggested by Sun (2022).[38] Table 7 also shows test results for the sub-sample with WTP for information above N100. According to Section 2, the evidence against a uniform response to information must be stronger in this case. Kitagawa (2015) finds that the trimming parameter affects the test's power depending on whether monotonicity violations occur at the tails or in areas with higher density. In our sample, violations tend to occur when densities are larger which would favor using larger trimming parameters. Table 7 presents

[35]Neither differences in means (p-value = 0.172) nor distributions (p-value = 0.205) are significant.

[36]We thank Zhenting Sun for providing code to implement these tests.

[37]As discussed in Section 2.2, the procedure can be modified to include prior beliefs about discontinuing education. However, we do not have evidence that these beliefs are updated differently between treated and non-treated subjects. We will include results using both priors in the next sub-section.

[38]Sun (2022) suggests a modification of the test by Kitagawa (2015) that is more powerful.

results with a range of trimming parameters for completeness.

Evidence against monotonicity for the whole population is mixed. We can reject monotonicity for trimming parameters above $\xi \geq 0.22$, but not for $\xi < 0.22$. Evidence consistent with Proposition 1 is the clearest for those with WTP for information above N100. We can reject that densities dominate each other at conventional significance levels for all values of ξ. This is consistent with the theoretical prediction that those more likely to use information are willing to pay more for it.

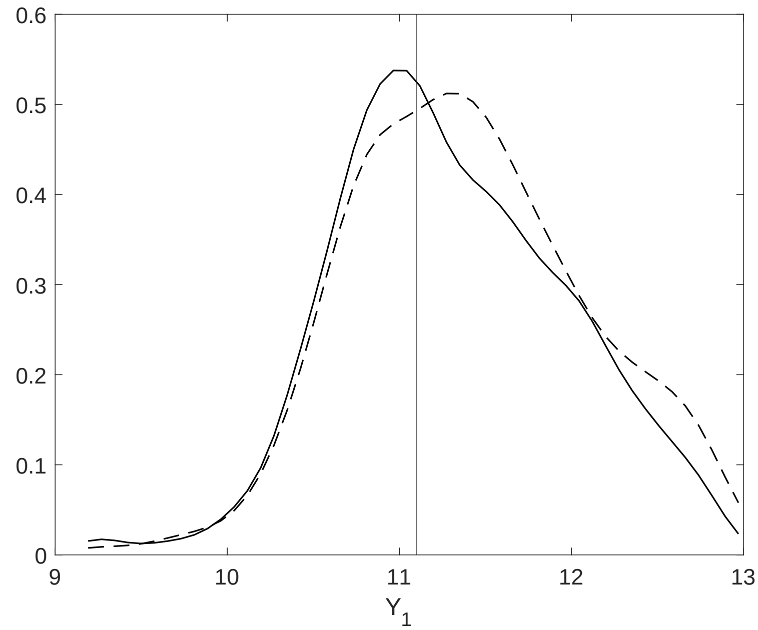


Figure 7: Kernel (Gaussian) density estimates. Prior beliefs on earnings of continuing education by treatment conditional on continuing education. Solid lines for treatment, dotted lines for control.

Table 7: Monotonicity test using baseline beliefs of those continuing education

	Trimming constant (ξ)									
	0.07	0.1	0.13	0.16	0.19	0.22	0.25	0.28	0.3	1
	Intervention increases education									
All (Kitagawa)	0.070	0.067	0.066	0.063	0.060	0.054	0.045	0.042	0.037	0.012
All (Sun)	0.072	0.069	0.068	0.065	0.062	0.055	0.046	0.042	0.037	0.012
WTP>100 (Kitagawa)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
WTP>100 (Sun)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Intervention decreases education									
All (Kitagawa)	0.346	0.216	0.134	0.148	0.134	0.098	0.085	0.072	0.068	0.021
All (Sun)	0.355	0.227	0.139	0.154	0.140	0.104	0.090	0.076	0.073	0.026
WTP>100 (Kitagawa)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
WTP>100 (Sun)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

4.5 Response types

Section 2.2 provides a method to estimate the proportion of compliers and defiers by inspecting distributional changes in prior beliefs among those continuing education in the treatment and control conditions. Table 8 shows those probabilities estimated for the whole population. We obtain estimates by first re-balancing the sample to satisfy the conditions in Proposition 1.[39] Table 8 presents the mean estimates and their standard deviations. The estimates of those switching in and out of education differ from zero. The top panel presents estimates using prior beliefs on earnings if continuing education. To bottom panel presents estimates using prior beliefs on earning if continuing and discontinuing education.[40] The measured effect of information on behavior is twice as large using both priors. Of each participant deciding to continue education, two decide to discontinue education. Using simple comparisons of means of treatment and control in information campaigns might be misleading. Table 8 shows the estimated effect of information is significantly below the Fréchet upper bound. That is, Proposition 1 is not rejected in our experiment.

Table 8: Distribution of behavioral types

	One dimensional				
	Switched in	Switched out	Out minus In	Out plus In	Fréchet UB
Mean	0.009	0.058	0.049	0.067	0.212
SE	0.004	0.006	0.008	0.006	0.008

	Two dimensional				
	Switched in	Switched out	Out minus In	Out plus In	Fréchet UB
Mean	0.049	0.097	0.049	0.146	0.212
SE	0.006	0.007	0.008	0.010	0.008

Table 9 investigates the distribution of behavioral types over prior beliefs. To do this, we divide the distribution of beliefs into four quadrants defined by the mid-point value of priors (τ_0 for discontinuing education and τ_1 for continuing education). For instance, $(Y_0 < \tau_0, Y_1 < \tau_1)$ is the quadrant of beliefs where both priors are below the mid-point and $(Y_0 > \tau_0, Y_1 > \tau_1)$ is the quadrant of beliefs where both priors are above the mid-point. Those who switched to education were more likely to have lower expectations of discontinuing education and higher

[39] We binned beliefs and sample from the treated group to match the proportions in the control group. We use samples for which we find no significant difference in the distribution of prior beliefs. We then obtain kernel density estimates for these samples using Matlab's mvksdensity command. The results use Silverman's rule for bandwiths and Epanechnikov kernel. The results are robust using alternative kernel methods.

[40] The method presented in Section 2.2 affords estimations using the full set of prior beliefs. We refrain from that approach due to the curse of dimensionality.

expectations of continuing education. These differences are not significantly different. Those who switched out of education were less likely to have lower expectations of discontinuing education and higher expectations of continuing education. These differences are significantly different. The table illustrates how prior information helps uncover flows in and out of education. The proposed method *let the data speak* in the sense that it does not impose a specific way to cut the data to detect treatment effects.

Table 9: Distribution of types on (Y_0, Y_1)

	(1) $Y_0 < \tau_0$ $Y_1 < \tau_1$	(2) $Y_0 < \tau_0$ $Y_1 > \tau_1$	(3) $Y_0 > \tau_0$ $Y_1 < \tau_1$	(4) $Y_0 > \tau_0$ $Y_1 > \tau_1$	(5) Total	(6) (1)+(2)	(7) (3)+(4)	(8) (1)+(3)	(9) (2)+(4)	(10) H0:(6)=(7)	(11) H0:(8)=(9)
Switched In:											
Mean	0.012	0.017	0.009	0.011	0.049	0.029	0.020	0.020	0.028	0.264	0.150
SD	0.003	0.004	0.002	0.004	0.006	0.005	0.005	0.004	0.004		
Switched Out:											
Mean	0.009	0.017	0.019	0.053	0.097	0.026	0.071	0.028	0.069	< 0.001	< 0.001
SD	0.002	0.003	0.003	0.005	0.007	0.004	0.006	0.004	0.006		

4.6 Willingness to pay for information

Table 10 investigates the willingness to pay for information of those switching decisions after receiving information. We take advantage of the fact that our estimations in the previous section allow us to detect the distribution of participants switching in and out of education based on their prior beliefs. The first column is the average WTP for information of those who continue education in the control group and have prior beliefs as those who switched to continue education. The second column is the average WTP for information on those who continue education in the treatment group and have prior beliefs as those who switched to continue education. The only difference between these two groups is the participants who switched to continuing education due to the intervention. The third column presents the implicit willingness to pay for information of those switching.[41] The second column is lower than the first, implying that those who switched behavior are less willing to pay for information. We do not impose boundary conditions in the estimation and obtain negative values. The fourth column presents the willingness to pay for information of those who continue education in the control group and have beliefs similar to those who discontinue education due to the intervention. The fifth column presents the willingness to pay for information of those who continue education in the treatment group and have beliefs similar to those who discontinue education due to the intervention. Since the difference between these two groups is those who switched behavior, we conclude that those who switched must

[41] We use the law of total probability to estimate this value.

be more willing to pay for information.

We find significant differences in willingness to pay for all pieces of information across groups. The evidence for sub-components of information is similar but much noisier. Importantly, we find a significant difference in willingness to pay for information between those who switched in and out of education due to the intervention. We note that this conclusion is not due to boundary violations. The willingness to pay for information for those who discontinue education is significantly larger than zero (and the mean), which would be an alternative estimate of the willingness to pay for information for those who switched to continue education. This is evidence consistent with information avoidance in some sub-populations and warrants caution in interpreting WTP solely as a measure of the instrumental value of information.

Table 10: WTP by prior beliefs, treatment, and response type

	AT	AT+In	Switched-In	AT+Out	AT	Switched-Out			
	(1) $Z=0$	(2) $Z=1$	(3)	(4) $Z=0$	(5) $Z=1$	(6)	(7) H0:(1)=(2)	(8) H0:(4)=(5)	(9) H0:(3)=(6)
Both:									
Mean	112.4	106.0	-97.9	113.5	101.8	212.7	0.083	< 0.001	< 0.001
SD	1.7	3.1	48.4	0.9	2.1	22.0			
Earnings:									
Mean	105.2	105.0	-35.4	110.3	102.3	178.6	0.947	0.001	0.001
SD	2.0	3.0	44.3	1.0	2.1	20.5			
Admissions:									
Mean	111.5	109.2	-84.7	115.2	104.1	209.0	0.492	< 0.001	< 0.001
SD	1.9	2.7	45.2	1.0	2.0	20.5			

4.7 Flows across eduational choices

The method proposed in Section 2.2 provides a way to partially identify flows in and out of several educational paths.[42] The information intervention likely altered beliefs of several options relative to discontinuing education and relative to each other. Since subjects with similar beliefs are expected to react similarly to information, we can assess flows across choices by measuring distributional changes in the joint distribution of beliefs across treatment and control groups.

To implement this approach, we estimate the joint density function of the priors on earning for each choice (Arts, Commerce, and Science) and earning for junior high school for those choosing each of these options. For example, we estimate the joint density function of the priors on earnings for Arts, prior for junior high school, and choosing Arts conditional on

[42]The flows are partially identified if the number of options exceeds 3.

being in the Control and Treated group.[43] We estimate the probability of moving into Arts due to the intervention as the integral of the difference between the density in the Treated and Control group truncated at zero. We estimate the probability of moving out of Arts due to the intervention as the integral of the difference between the density in the Control and Treat group truncated at zero. This procedure is repeated for Commerce and Science. We also estimate the flows in and out of senior high school by aggregating prior as in the previous section.[44] We remark that this approach does not double count movements in and out of educational paths since they are constructed using marginal densities on mutually exclusive choices. For instance, when we estimate the flows in and out of senior high school, we do not account for movements across fields of study. When we estimate the flows in and out of Arts, we account for flows into other fields and in and out of senior high school.

The estimates proposed in the previous paragraph are insufficient to identify the flows across educational paths. However, they impose bounds on the true distribution of flows across choices. To obtain interpretable and conservative results, we use linear programming to find the largest solution consistent with these flows and a joint probability distribution. We minimize the probability that the intervention generated flows across choices. Table 11 provides estimates using this approach. The last column is the marginal distribution of choices for the Control group and the third from the bottom row is the marginal distribution of choice for the Treatment group. We estimate ninety-five confidence intervals using Bootstrap. We do not estimate the movements across choices since they are not point identified. The estimates on the flows in and out of Junior High School are different from those in Table 8 because we are estimating the model on the subsample that has prior beliefs for all these options. For instance, we obtained that 4.0 percent (8.2-4.2) switched from junior high school into senior high school, and 8.8 percent (13.0-4.2) switched from senior high school into junior high school only. The intent-to-treat estimate is 3.8 (13.0-8.2).

To assess the impact of information on school choices, we sum the terms of the diagonal of the matrix. We can calculate the minimum effect of the intervention by estimating the maximum joint probability of not reacting to the policy using Fréchet bounds. The lower bound of the effect of the information intervention is 10.6 percent. Using the proposed method, we estimate this proportion at 36.2 ($\text{CI}_{95\%} = [26.1, 57.5]$). Given these findings, we might observe larger treatment effects on income in the future.

[43] That is, $f(Y_0, Y_{Arts}, Arts = 1|Z = 0)$ and $f(Y_0, Y_{Arts}, Arts = 1|Z = 1)$.

[44] We use Matlab mvksdensity to estimate these densities over a 50×50 grid. We use a Gaussian kernel with Silverman's rule of thumb bandwidth. We use the trapezoidal rule for numerical integration. We rebalance the sample before conducting the analysis to minimize spurious effects due to a lack of balance in small samples. We split the data in 5×5 bins and resample to obtain balance on these bins. We estimate confidence intervals using clustered bootstrap.

Table 11: Treatment effects on the distribution of educational choices

	Treated				
Control	Junior HS	Arts	Commerce	Science	Total
Junior HS	**0.042**				**0.082**
2.5%	0.002				0.052
97.5%	0.053				0.130
Arts		**0.295**			**0.387**
2.5%		0.207			0.348
97.5%		0.336			0.418
Commerce			**0.146**		**0.314**
2.5%			0.058		0.229
97.5%			0.234		0.371
Science				**0.152**	**0.217**
2.5%				0.036	0.160
97.5%				0.223	0.285
Total	**0.130**	**0.356**	**0.240**	**0.274**	**1.000**
2.5%	0.096	0.310	0.194	0.219	1.000
97.5%	0.161	0.414	0.292	0.322	1.000

5 Study limitations

Our study provides a limited amount of information pertinent to the decision to continue education, and it is possible that a larger set of variables might lead to larger or different changes in behavior. However, we have little knowledge of the kind of information adolescents' value. Our findings suggest that time spent searching could be used as a proxy for the value of information. The main analysis is restricted to one field outcome: discontinuing education. We also pre-registered other outcomes for the study, including senior high school graduation, choice of major, and post-secondary education. The framework developed in Section 2 and Section 2.2 can be extended to non-binary outcomes by acknowledging the possibility of mixed signals. We do not implement this approach presently because those outcomes are not yet available. Data collection did not include beliefs about the likelihood of being admitted to college. In retrospect, this is likely an important variable in the Nigerian case because of the limited supply of post-secondary education alternatives.

6 Conclusions

Despite large disparities in access to information on returns to education and the long-term consequences associated with poor decisions, the evidence on the effectiveness of information interventions in education is mixed. A possible interpretation of this mixed evidence is that information is of secondary importance in addressing gaps in human capital accumulation. This paper shows that standard analysis of information interventions likely underestimates their full effect. This advises caution in the analysis of cross-cut experimental designs with information intervention arms.

We develop a method to detect nonuniform responses to information and, therefore, the full effect of information on educational choices. In our study, we find that the conventional measure of average treatment effects underestimates the effect of information by a factor of two. For every three participants choosing to discontinue education due to treatment, one participant decided to continue education. At least one in twelve participants changed their decision due to the intervention. This is likely a lower bound since we show that the intervention affected choices across fields of study whose outcomes are yet to be observed. Information constraints are not trivial in this population. Standard experimental design advises sample sizes inversely related to expected effect sizes. If small effect sizes are due to nonuniform response to treatment, our results suggest that richer data might be a better alternative than more data.

References

Ajayi, K.F., Friedman, W.H., Lucas, A.M., 2020. When information is not enough: Evidence from a centralized school choice system.

Allcott, H., Braghieri, L., Eichmeyer, S., Gentzkow, M., 2020. The Welfare Effects of Social Media. American Economic Review 110, 629–676.

Allcott, H., Kessler, J.B., 2019. The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. American Economic Journal-Applied Economics 11, 236–276.

Arcidiacono, P., Hotz, V.J., Maurel, A., Romano, T., 2020. Ex Ante Returns and Occupational Choice. Journal of Political Economy 128, 4475–4522.

Benjamin, D.J., 2019. Errors in probabilistic reasoning and judgment biases, in: Bernheim, B., DellaVigna, S., Laibson, D. (Eds.), Handbook of Behavioral Econoics - Foundations and Applications 2. Handbooks in Economics, pp. 69–186.

Bergman, P., Denning, J.T., Manoli, D., 2019. Is Information Enough? The Effect of Information about Education Tax Benefits on Student Outcomes. Journal of Policy Analysis and Management 38, 706+.

Berry, J., Fischer, G., Guiteras, R.P., 2020. Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana. Journal of Political Economy 128, 1436–1473.

Bettinger, E.P., Long, B.T., Oreopoulos, P., Sanbonmatsu, L., 2012. The Role of Application Assistance and Information in College Decisions: Results from the H & R Block Fafsa Experiment. Quarterly Journal of Economics 127, 1205–1242.

Blackwell, D., 1951. Comparison of experiments, in: In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. volume Berkeley, CA: University of California Press of *Annual Review of Economics*, pp. 93–102.

Bohm, P., Linden, J., Sonnegard, J., 1997. Eliciting reservation prices: Becker-DeGroot-Marschak mechanisms vs. markets. Economic Journal 107, 1079–1089.

Brocas, I., Carrillo, J.D., 2021. Steps of Reasoning in Children and Adolescents. Journal of Political Economy 129, 2067–2111.

Brocas, I., Carrillo, J.D., Combs, T.D., Kodaverdian, N., 2019. The development of consistent decision-making across economic domains. Games and Economic Behavior 116, 217–240.

Bronchetti, E.T., Kessler, J.B., Magenheim, E.B., Taubinsky, D., Zwick, E., 2020. Is attention produced rationally?.

Bursztyn, L., Gonzalez, A.L., Yanagizawa-Drott, D., 2020. Misperceived social norms: Women working outside the home in saudi arabia. American Economic Review 110, 2997–3029.

Busso, M., Dinkelman, T., Martinez, C.A., Romero, D., 2017. The effects of financial aid and returns information in selective and less selective schools: Experimental evidence from Chile. Labour Economics 45, 79–91.

Cabrales, A., Gossner, O., Serrano, R., 2017. A normalized value for information purchases. Journal of Economic Theory 170, 266–288.

Canay, I.A., Romano, J.P., Shaikh, A.M., 2017. Randomization tests under an approximate symmetry assumption. Econometrica 85, 1013–1030.

Caplin, A., 2016. Measuring and Modeling Attention, in: Aghion, P and Rey, H (Ed.), Annual Review of Economics. volume 8 of *Annual Review of Economics*, pp. 379–403.

Caplin, A., Csaba, D., Leahy, J., Nov, O., 2020. Rational inattention, competitive supply, and psychometrics. Quarterly Journal of Economics 135, 1681 – 1724.

Caplin, A., Dean, M., 2015. Revealed Preference, Rational Inattention, and Costly Information Acquisition. American Economic Review 105, 2183–2203.

Castillo, M., Ferraro, P.J., Jordan, J.L., Petrie, R., 2011. The today and tomorrow of kids: Time preferences and educational outcomes of children. Journal of Public Economics 95, 1377–1385.

Castillo, M., Jordan, J.L., Petrie, R., 2018. Children's rationality, risk attitudes and field behavior. European Economic Review 102, 62–81.

Castillo, M., Jordan, J.L., Petrie, R., 2019. Discount Rates of Children and High School Graduation. Economic Journal 129, 1153–1181.

Chambers, C.P., Liu, C., Rehbeck, J., 2020. Costly information acquisition. Journal of Economic Theory 186.

Chassang, S., Padro i Miquel, G., Snowberg, E., 2012. Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. American Economic Review 102.

Chavas, J., 1993. Oon the demand for information. Economic Modelling 10, 398–407.

Chiong, K.X., Galichon, A., Shum, M., 2016. Duality in dynamic discrete-choice models. Quantitative Economics 7, 83–115.

Dahl, C., Huber, M., Mellace, G., 2017. It's never too late: A new look at local average treatment effects with or without defiers. Discussion Papers on Business and Economics No. 2/2017.

De Chaisemartin, C., 2017. Tolerating defiance? local average treatment effects without monotonicity. Quantitative Economics 8, 367–396.

Del Boca, D., Flinn, C.J., Verriest, E., Wiswall, M., 2019. Actors in the child development process.

Delavande, A., Zafar, B., 2019. University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints. Journal of Political Economy 127, 2343–2393.

Eeckhoudt, L., Godfroid, P., 2000. Risk aversion and the value of information. Journal of Economic Education 31, 382–388.

Frankel, A., Kamenica, E., 2019. Quantifying information and uncertainty. American Economic Review 109, 3650–3680.

Fuster, A., Perez-Truglia, R., Wiederholt, M., Zafar, B., 2022. Expectations with endogenous information acquisition: An experimental investigation. Review of Economic and Statistics 104, 1059–1078.

Golman, R., Loewenstein, G., Molnar, A., Saccardo, S., 2021. The demand for, and avoidance of, information. Management Science .

Gould, J., 1974. Risk, stochastic preferences, and value of information. Journal of Economic Theory 8, 64–84.

Goux, D., Gurgand, M., Maurin, E., 2017. Adjusting Your Dreams? High School Plans and Dropout Behaviour. Economic Journal 127, 1025–1046.

Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., White, B., 2021. A National-Level Informational Experiment to Promote Enrollment in Selective Colleges. Journal of Policy Analysis and Management 40, 453+.

Haaland, I., Roth, C., Wohlfart, J., forthcoming. Designing Information Provision Experiments. Journal of Economic Literature .

Harbaugh, W., Krause, K., Berry, T., 2001. GARP for kids: On the development of rational choice behavior. American Economic Review 91, 1539–1545.

Hastings, J., Neilson, C.A., Zimmerman, S.D., 2015. The effects of earnings disclosure on college enrollment decisions.

Heckman, J.J., Pinto, R., 2018. Unordered Monotonicity. Econometrica 86, 1–35.

Heckman, J.J., Urzua, S., Vytlacil, E., 2006. Understanding instrumental variables in models with essential heterogeneity. Review of Economic and Statistics 88, 389–432.

Hirshleifer, J., 1971. Private and social value of information and reward to inventive activity. American Economic Review 61, 561–574.

Hjort, J., Moreira, D., Rao, G., Santini, J.F., 2021. How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities. American Economic Review 111, 1442–1480.

Hoxby, C.M., Turner, S., 2015. What High-Achieving Low-Income Students Know About College. American Economic Review 105, 514–517.

Hyman, J., 2020. Can Light-Touch College-Going Interventions Make a Difference? Evidence from a Statewide Experiment in Michigan. Journal of Policy Analysis and Management 39, 159+.

Imbens, G., Angrist, J., 1994. Identification and Estimation of Local Average Treatment Effects. Econometrica 62, 467–475.

Imbens, G., Rubin, D., 1997. Estimating outcome distributions for compliers in instrumental variables models. Review of Economic Studies 64, 555–574.

Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. The Quarterly Journal of Economics 125, 515–548.

Kitagawa, T., 2015. A test for instrument validity. Econometrica 83, 2043–2063.

Kreps, D., Porteus, E., 1978. Temporal Resolution of Uncertainty and Dynamic Choice Theory. Econometrica 46, 185–200.

Lai, F., Sadoulet, E., de Janvry, A., 2009. The adverse effects of parents' school selection errors on academic achievement: Evidence from the Beijing open enrollment program. Economics of Education Review 28, 485–496.

de Lara, M., Gossner, O., 2020. Payoffs-beliefs duality and the value of information. SIAM Journal on Optmization 30, 464–489.

List, J., Petrie, R., Samek, A., forthcoming. How experiments with children inform economics. Journal of Economic Literature 135, 1681 – 1724.

Loyalka, P., Liu, C., Song, Y., Yi, H., Huang, X., Wei, J., Zhang, L., Shi, Y., Chu, J., Rozelle, S., 2013. Can information and counseling help students from poor rural areas go to high school? Evidence from China. Journal of Comparative Economics 41, 1012–1025.

Lucas, A.M., Mbiti, I.M., 2012. The Determinants and Consequences of School Choice Errors in Kenya. American Economic Review 102, 283–288.

Marbach, M., Hangartner, D., 2020. Profiling Compliers and Noncompliers for Instrumental-Variable Analysis. Political Analysis 28, 435–444.

Mazar, N., Koszegi, B., Ariely, D., 2014. True Context-dependent Preferences? The Causes of Market-dependent Valuations. Journal of Behavioral Decision Making 27, 200–208.

McFadden, D., 1978. Modelling the Choice of Residential Location.. Elgar Reference Collection. Pioneering Papers of the Nobel Memorial Laureates in Economics, vol. 3. Cheltenham, U.K. and Northampton, Mass.: Elgar, Unlisted. pp. 409 – 430.

McFadden, D., 1981. Econometric models of probabilistic choice, in: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press: Cambridge, MA., pp. 198–272.

Mosquera, R., Odunowo, M., McNamara, T., Guo, X., Petrie, R., 2020. The economic effects of Facebook. Experimental Economics 23, 575–602.

Rust, J., 1994. Structural estimation of Markov decision processes. Handbook of Econometrics 4, 3081–3143.

Shi, X., Shum, M., Song, W., 2018. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. Econometrica 86, 737–761.

Sørensen, J., Fosgerau, M., 2022. How McFadden met Rockafellar and learned to do more with less. Journal of Mathematical Economics 100.

Stinebrickner, R., Stinebrickner, T.R., 2014. A major in science? initial beliefs and final outcomes for college major and dropout. Review of Economic Studies 81, 426–472.

Stutern, 2018. Nigerian graduate report: 2018. https://www.kaggle.com/stutern/nigerian-gradaute-report-2018.

Sun, Z., 2022. Instrument validity for heterogeneous causal effects.

Sutter, M., Kocher, M.G., Glaetzle-Ruetzler, D., Trautmann, S.T., 2013. Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. American Economic Review 103, 510–531.

Thornton, R.L., 2008. The demand for, and impact of, learning hiv status. American Economic Review 98, 1829–1863.

Willis, R., Rosen, S., 1979. Education and self-selection. Journal of Political Economy 87, S7–S36.

Wiswall, M., Zafar, B., 2015a. Determinants of college major choice: Identification using an information experiment. The Review of Economic Studies 82, 791–824.

Wiswall, M., Zafar, B., 2015b. How do college students respond to public information about earnings? Journal of Human Capital 9, 117–169.

Wiswall, M., Zafar, B., 2018. Preference for the workplace, investment in human capital, and gender. Quarterly Journal of Economics 133, 457–507.

Wiswall, M., Zafar, B., 2021. Human capital investments and expectations about career and family. Journal of Political Economy 129, 1361–1424.

APPENDICES INTENDED FOR ONLINE PUBLICATION

A ADDITIONAL MATERIAL

A.1 The value of information

This section discusses the problem of costly information in more generality that in Section 2. The formalisms in this section are based on the presentation in de Lara and Gossner (2020) who analyzes the value of information in contexts like ours. To mimic the decision to continue to senior high school that we analyze in this paper, we concentrate on a two-action model. We represent the decision to continue to senior high school as action a being equal to 1 and the decision to not continue to senior high school as action a being equal to 0. Uncertainty about the returns to alternative career paths are represented as a finite set of possible states of the world $\Omega = \{\omega_1, \ldots, \omega_N\}$. Students are endowed with a utility function $u(a, \omega) = u(a) + e_a(\omega)$ that depends on the action a that is taken and the state of the world ω. The term $u(a)$ depends only on action a, and term $e_a(\omega)$ depends both on ω and option a. Decisions are made prior to the realization of ω, and students have a prior distribution $\mu \in \Gamma = \Delta(\Omega)$ over states of the world.

The set Π of information structures includes all of the functions $\pi : \Omega \rightarrow \Delta(\Gamma)$ that select measures with a finite support $\Gamma(\pi) \subset \Gamma$, and satisfy Bayes's law. Intuitively, an information structure provides a finite set of signals that agents can use to form posterior distributions of the state of the world, with different signals producing different posterior distributions. These posteriors are then used to decide which actions are optimal. The gross ex-ante payoff $G(\pi, u)$ for a student who uses information structure π optimally is:

$$G(\pi, u) = \sum_{\gamma \in \Gamma(\pi)} q(\gamma)[max_{a \in \{0,1\}} \sum_{\omega \in \Omega} \gamma(\omega)(u(a) + e_a(\omega))] \tag{1}$$

where $q(\gamma) = [\sum_{\omega \in \Omega} \mu(\omega)\pi(\gamma|\omega)]$ (see Caplin and Dean, 2015). The equation above says that, given a state of the world ω, a person receives a signal that allows them to form a posterior belief γ. This posterior beliefs is used to evaluate which action is optimal. The gross ex-ante payoff is the expected utility taking into account all the signals that are possible for each state of the world. More succinctly, information structures π can be represented as a distribution over a set of posterior distributions that average to the prior distribution.

Let $G(\mu, u)$ be the gross ex-ante payoff given prior μ. The WTP for information structure π is the number $W_\pi(u)$ such that:

$$G(\pi, u) - W_{\pi}(u) = G(\mu, u). \tag{2}$$

i.e., the difference between the gross ex-ante payoffs under π and μ. WTP is positive (see Chavas, 1993, for sufficient conditions for this result). de Lara and Gossner (2020) show that WTP is zero unless information structure π produces a signal at which optimal behavior under prior μ is no longer optimal. Moreover, WTP is increasing in the probability of receiving a signal leading to a change in behavior.[45]

Since the gross ex-ante payoff function is convex (see de Lara and Gossner, 2020), it is subdifferentiable. Sørensen and Fosgerau (2022) show that this subdifferential coincides with the choice probability correspondence, i.e. $\partial G(\pi, u) = Pr(a = 0|\pi, u)$, and satisfies cyclical monotonicity. If the gross ex-ante payoff function is differentiable we have that $\frac{\partial G(\pi,u)}{\partial u(0)} = Pr(a = 0|\pi, u)$, which an instance of the Williams-Daly-Zachery theorem for additive random utility (McFadden, 1978; Rust, 1994). Given (2), we conclude that:

$$dW_{\pi}(u)/du(i) = Pr(a = i|\pi, u) - Pr(a = i|\mu, u) = \Delta Pr(a = i|u), i = 0, 1. \tag{3}$$

at any point of differentiability.[46]

Chassang et al. (2012) were the first to point out that WTP to be treated can be used to elicit actual and perceived treatment effects. We note, however, that if preferences are not separable, equation (3) can over(under)estimate the relationship between expected changes in behavior ($\Delta Pr(a = 0|u)$) and WTP.[47,48] Non-separable preferences are a common assumption in the literature on the value of information (e.g. Gould, 1974; Chavas, 1993; Eeckhoudt and Godfroid, 2000; Cabrales, Gossner and Serrano, 2017). We present evidence that WTP is increasing in risk tolerance. Importantly, the qualitative result that WTP is positive only if it has an expected effect on behavior does not depend on the separability assumption.

We illustrate these results with an example (see Figure A.1). The example assumes that $u_0 = \Delta_0$ and $u_1 = 0$, where Δ_0 is the return to discontinuing senior high school. It also assumes that $e_0 = -4$ or $e_0 = 4$ with equal probability, and that e_1 is always equal to 0. That is, the are 2 states of the world that obtain with equal probability. In the absence of new information, students with $\Delta_0 < 0$ choose $a = 1$ (continue education) and those with

[45] These are Proposition 3.1 and Theorem 3.2 in de Lara and Gossner (2020)

[46] This equation holds for small changes in $\Delta Pr(a = i|u)$ which is our case. Closed-form solutions for function $W_{\pi}(u)$ require parametric assumptions as in (McFadden, 1978).

[47] Chambers, Liu and Rehbeck (2020) note that the prior and posterior distributions play a role similar to that of prices in consumer theory. Information structures are, by definition, mean preserving spreads of the prior distribution. We therefore expect that the certainty equivalent, and hence WTP, associated with an information structure is increasing in risk tolerance.

[48] The value of information itself depends on risk attitudes, however, incentive compatible measurement of this value of information is robust to risk attitudes.

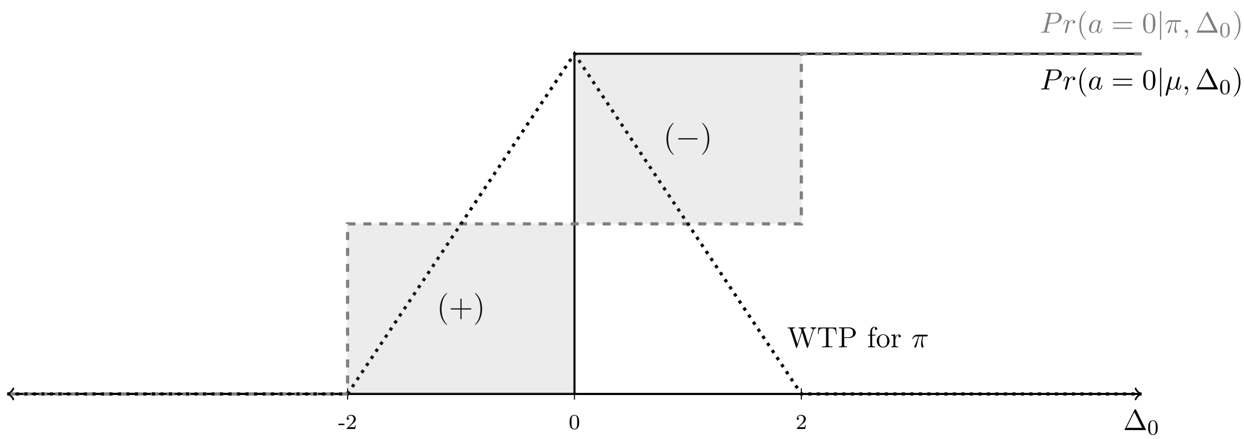


Figure A.1: WTP for information as a function of returns to discontinuing education

Notes: The x-axis represents the return to discontinuing education, Δ_0. The y-axis represents the probability of discontinuing education and the willingness to pay for information structure π. $Pr(a = 0|\mu, \Delta_0)$, black solid line, denotes the best-response correspondence as a function of Δ_0 given prior μ. $Pr(a = 0|\pi, \Delta_0)$, gray dashed line, denotes the best-response correspondence as a function of Δ_0 given information structure π. The willingness to pay for information structure π is represented by the dotted black line. The marginal WTP for information structure π equals $Pr(a = 0|\pi, \Delta_0) - Pr(a = 0|\mu, \Delta_0)$ whenever this is single valued. It increases in the area marked with $(+)$, when a student might switch to $a = 0$ if information structure π is available, and decreases in the area marked (-), when a student might switch to $a = 1$ if information structure π is available. WTP is maximal when an agent is indifferent between options and decreases as choices become more certain.

$\Delta_0 > 0$ choose $a = 0$ (discontinue education). Any choice is optimal if $\Delta_0 = 0$. This obtains because the expected payoff to $a = 0$ equals $\frac{1}{2}(\Delta_0 + 4) + \frac{1}{2}(\Delta_0 - 4) = \Delta_0$. The optimal choice is illustrated in Figure A.1 by a solid black line.

Consider now an information structure π that updates the probability that $e_0 = -4$ to $\frac{1}{4}$ with probability $\frac{1}{2}$ and that updates the probability that $e_0 = -4$ to $\frac{3}{4}$ with probability $\frac{1}{2}$. This is a valid information structure since the average of the posteriors equals the prior (i.e., $\frac{1}{2}$). A student in possession of this information will update her options accordingly. With probability $\frac{1}{2}$, the expected payoff to $a = 0$ equals $\frac{3}{4}(\Delta_0 + 4) + \frac{1}{4}(\Delta_0 - 4) = \Delta_0 + 2$, and with probability $\frac{1}{2}$, the expected payoff to $a = 0$ equals $\frac{1}{4}(\Delta_0 + 4) + \frac{3}{4}(\Delta_0 - 4) = \Delta_0 - 2$. A person with $\Delta_0 < 0$ will not change her behavior if the probability of $e_0 = -4$ increases, but could change her behavior if the probability of $e_0 = -4$ decreases. In particular, any student with $\Delta_0 \in (-2, 0)$ will switch from $a = 1$ to $a = 0$ when the positive signal is received. Since this switch in behavior happens with probability $\frac{1}{2}$, we obtain that the probability of choosing $a = 0$ increases to $\frac{1}{2}$ when information structure π is available. For those with $\Delta_0 < -2$, the increase in probability of $e_0 = 4$ is not large enough to change their decisions. Those with $\Delta_0 = -2$ are indifferent between $a = 0$ and $a = 1$ when good news are received, so any choice is optimal. Their probability of choosing $a = 0$ is therefore between 0 and $\frac{1}{2}$ (half of the time $Pr(a = 0) = 0$ and half of the time $Pr(a = 0) \in [0, 1]$). Similar arguments imply that students with $\Delta_0 \in (0, 2)$ will decrease their probability of choosing $a = 0$ from 1 to $\frac{1}{2}$, and those with $\Delta_0 = 2$ will choose $a = 0$ with probability between $\frac{1}{2}$ and 1. Finally, those with $\Delta_0 = 0$ will choose $a = 0$ when the signal is good and choose $a = 1$ when the signal is bad. They choose $a = 0$ with probability $\frac{1}{2}$. Figure A.1 represents these choices with a dashed gray line.

WTP for information depends on the optimal choices given information is accessed and the gains from these choices. Figure A.1 shows that those with $\Delta_0 \notin (-2, 2)$ do not change their decisions if information structure π is available. So, their WTP for information is 0. Following the results above, the expected payoff for those in $(-2, 0)$ is $\frac{1}{2}(\Delta_0 + 2) + \frac{1}{2} \times 0$ under information structure π and 0 (since $a = 1$) under prior μ. Their WTP for information structure π is therefore $\frac{1}{2}\Delta_0 + 1 - 0$ and it is increasing in Δ_0. The expected payoffs for those in $(0, 2)$ is $\frac{1}{2}(\Delta_0 + 2) + \frac{1}{2} \times 0$ under information structure π and Δ_0 (since $a = 0$) under prior μ. Their WTP for information structure π is therefore $\frac{1}{2}\Delta_0 + 1 - \Delta_0 = 1 - \frac{1}{2}\Delta_0$ which is decreasing in Δ_0. Figure A.1 represents WTP by a dotted black line. We confirm that WTP for π is largest when indifference between alternatives is the smallest ($\Delta_0 = 0$).

Figure A.1 makes clear that a change in beliefs is a necessary, but not sufficient, condition for a change in behavior (those with $\Delta_0 \notin (-2, 2)$ do not change behavior with new information). An information RCT might change the beliefs of two students in the same way, but it might change the behavior on only one person, both or neither one. Finally, the shape of $Pr(a = 0|\pi, \Delta u_0)$ reflects the number and distribution of signals. It will be a step function as long as the number of signals is finite.

Figure A.1 helps illustrate our main identification challenge when only actual behavior is available. Information structure π can both increase and decrease the proportion of those choosing $a = 0$ through the assessment of information relative to prior beliefs. This implies that we might fail to detect the effect of information on average behavior even when information has an effect on individual behavior. Changes in behavior provide a lower bound of the treatment effect of information on choices. An information intervention might fail because the information has no value, relative to prior beliefs, or because it is poorly targeted. The policy conclusions are different in either case.

How can we test if behavior is rational? Equation 3 tells us that WTP will be highest for those most likely to change their behavior. Absent knowledge of individual returns to education, our ability to test for rational use of information depends on the latent distribution of returns to education (Δ_0). If either most students are such that $\Delta_0 < 0$ or $\Delta_0 > 0$, we will be able to test that WTP is significantly correlated with $\Delta Pr(a = 0|\pi)$.[49] In this case, those

[49]The identification challenge persists in the case in which the outcome variable is not binary, e.g. when

who expect to change their behavior upon receiving information will be willing to pay for it. Moreover, for small changes in $\Delta Pr(a = 0|\pi)$, this relationship should be proportional.

If treatment assignment has a monotone effect on participants, it is possible to estimate the treatment effect on compliers and to estimate their ex-ante characteristics (e.g. Imbens and Angrist, 1994; Imbens and Rubin, 1997; Marbach and Hangartner, 2020). De Chaisemartin (2017) introduced an assumption, weaker than monotonicity, that applies to a subgroup of compliers.[50] Since theory predicts that those who are affected by the treatment should value information more, these results suggest a simple test: compliers should be willing to pay more for information. The discussion above identifies two potential challenges to this approach. First, information interventions are unlikely to produce monotone effects, and second, identification of response types require access to counterfactual outcomes. The next section proposes using belief data to improve identification of behavioral types.

A.2 The demand for information

This section describes the demand for information corresponding to the 2,583 participants in the information treatment and impure control conditions. Eighty-three percent of the answers to the MPLs have no switch-backs. This is comparable to the 95% consistency rate in Fuster et al. (2022) and the 98% consistency rate in Allcott et al. (2020), who both use only one MPL per participant. Eighty percent of the participants are consistent on each of the three MPLs. Preferences for information are monotonic if $WTP_{\text{college \& wage info}} \geq max\{WTP_{\text{college info}}, WTP_{\text{wage info}}\}$. Seventy percent of the participants satisfy monotonicity.

Figure A.2 presents the demand curves for the monotonic and non-monotonic participants. The x-axis in each panel gives the percent of participants whose WTP is less than the number on the y-axis. We define a participant's WTP as the maximum price for which a participant would be willing to purchase the information. In the case where a participant has a switch-back, we use the WTP that minimizes the absolute distance to actual choices weighted by the price at which decisions are reversed.

In order to make our results comparable to those reported for developed countries, here we calculate the WTP in US dollars. The average WTP is \$0.25, which is equivalent to 0.6% of the monthly minimum wage (\$43.4), or 0.2% of the monthly wage of liberal arts graduates (\$143.6). For comparison, Allcott and Kessler (2019) found that participants were willing to pay \$3 on average for home energy reports, and Fuster et al. (2022) found that people were

several career paths are available. The formula for the marginal WTP also extends to this case *mutatis mutandis*.

[50] The condition is called Compliers-Defiers conditions. The subgroup of compliers whose treatment effects are identified are named convivors.

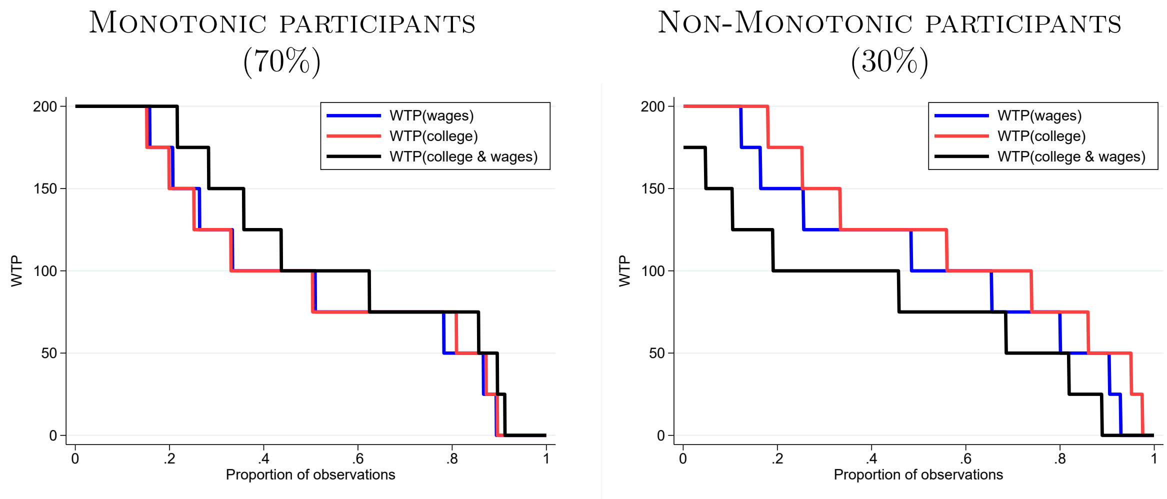


Figure A.2: The demand for information

Notes: A participant is classified as having monotonic preferences if her valuation of all pieces of information is the (weakly) largest. WTP is defined as the largest amount of money a participant would pay for information. The WTP of participants who have switch-backs is calculated as the WTP of the closest pattern, in monetary terms, of behavior without switch-backs. There were 2,389 usable observations of WTP for both pieces of information, 2,409 for salary information, and 2,435 for admissions information.

willing to pay $4 for information on home prices. This is roughly equivalent to 0.2% of the monthly salary of $2160 ($12 per hour).

Table A.1 presents the relationship between WTP and individual covariates from the rich set of baseline variables we collected. We include questions from Castillo et al. (2019), who study dropout rates in a US sample, as well as proxy variables used in that study. We find that, with the exception of risk attitudes, there is little, or no consistent, correlation between WTP and observable covariates.[51] This is consistent with the WTP providing new information about students' decision-making. While these results do not exclude the possibility of omitted variables bias, it helps allay the concern that WTP is correlated with unobservable participants' characteristics rather than their demand for information itself.

We conclude this section by discussing the implication of observing a consistent correlation between risk tolerance and WTP for information. The regression implies that a participant investing all her endowment in the paid lotteries is willing to pay about 30% above the average WTP. As discussed in the theoretical framework, if the average participant is risk averse, this implies that WTP will be less responsive to expected changes in behavior. In the extreme, expected changes in behavior might have little impact on WTP. Importantly, the correlation between risk tolerance and WTP cautions against interpreting

[51]We find little correlation between WTP and other baseline data as well. Results are avaiable from the authors upon request.

lack of a proportional relationship between changes in behavior and WTP as a failure of rational behavior. This confirms that qualitative nature of our test.

A.2.1 The perceived returns of education

Information RCTs providing earnings estimates assume that this information is relevant to students. However, recent research shows that career choices are only partially driven by concerns about earnings (e.g. Arcidiacono et al., 2020; Delavande and Zafar, 2019; Wiswall and Zafar, 2018). It is therefore important to assess whether students' beliefs about their career choices correspond to their expected earnings.

Since the students were asked what they thought their earnings would be for alternative career choices, we can estimate what students think the returns of education are. We follow Arcidiacono et al. (2020)'s methodology for measuring perceived treatment effects. These effects can be calculated because students reported what their earnings would be for every possible alternative. This means that we can measure the perceived treatment effect of any possible combination of individual choices. The treatment effects of interest are the treatment effect on the treated, i.e., the treatment effect on those choosing a particular alternative, and the treatment effect on the untreated, i.e., the treatment effect for each possible alternative that is not chosen. In particular, a measure of the perceived treatment effect on the treated for a particular career choice is the weighted mean of earnings at, say, 30 relative to junior high school. The weight is the probability that a student would choose such a career path. Analogously, the perceived treatment effect on the untreated is the weighted mean of earning at, say, 30 relative to junior high school, where the weight is the probability of not choosing such a career path.[52]

Table A.2 shows these estimates using beliefs on earnings at 30 for different career paths (senior high school, arts in college, commerce in college, and science in college) compared to ending education with junior high school.[53] The table presents separate estimates for the first set of elicited beliefs and the second set of elicited beliefs. We only use data from the information treatment group. First, we observe a clear ordering on the perceived returns of career choices. Senior high school is ranked lowest and science is ranked highest. Second, we observe that the perceived returns of education decreased significantly after the provision of information. For instance, the perceived returns of a career in a science field drop from N121,804 to N65,946. This is almost half of the initially perceived returns for this choice relative to junior high school. The same pattern repeats for other educational choices. Third, we observe that in most cases, estimates of the treatment effect on the treated are larger than

[52] We use probabilities of choosing or not choosing an alternative as a proxy for actual choices.

[53] We bound probabilities to be between 0.01 and 0.99 since some students' beliefs were exactly 0 or 1.

	Non-monotonic	WTP by type of information		
		Admissions	Wages	Both
Female	-0.028	-1.455	3.758	-1.073
	(0.023)	(4.085)	(3.788)	(3.318)
Age	-0.008	0.948	1.208	1.956
	(0.011)	(1.156)	(1.296)	(1.226)
First born	-0.023	-5.344	-3.588	-3.903
	(0.022)	(4.124)	(3.969)	(3.680)
Two-parent household	0.011	2.012	2.503	2.498
	(0.020)	(2.509)	(3.317)	(3.030)
Mother's education level	0.009	1.023	1.418	0.447
	(0.016)	(2.181)	(3.224)	(2.444)
% invested in lotteries	0.001	0.343**	0.294*	0.288*
	(0.001)	(0.147)	(0.147)	(0.151)
Discount rate	-0.029	11.079	13.362	9.575
	(0.055)	(9.495)	(8.913)	(9.656)
Nobody at home helps with homework	0.021	-4.902	-2.055	-6.429
	(0.024)	(4.536)	(5.333)	(4.774)
Repeated a grade	0.010	-0.229	3.995	2.564
	(0.033)	(4.115)	(3.808)	(3.969)
Average grades	-0.030	5.085	24.099*	11.356
	(0.105)	(12.277)	(13.771)	(11.605)
Received a suspension	-0.013	-18.144*	-13.143	-14.493
	(0.078)	(8.856)	(7.813)	(10.255)
Has extra lessons	-0.018	-2.502	-2.210	-2.291
	(0.020)	(2.448)	(2.939)	(2.996)
Currently attends an apprenticeship	0.030	-0.685	-2.180	-4.433
	(0.036)	(4.566)	(3.603)	(3.993)
Best school subject is math	0.015	-2.218	1.726	-0.088
	(0.029)	(4.009)	(3.453)	(4.002)
I pay attention in class	-0.006	4.135	0.359	2.549
	(0.019)	(2.486)	(2.424)	(2.230)
I like being at school	0.027*	-1.871	1.000	-1.909
	(0.016)	(1.825)	(2.397)	(1.910)
I get in trouble at school	0.026**	2.024	1.617	1.187
	(0.011)	(1.206)	(1.324)	(1.388)
Observations	2100	2165	2142	2126
Adj R2	0.000	0.022	0.016	0.016

* p<0.10, ** p<0.05, *** p<0.010. Errors clustered at the school level. Estimates are based on cases with complete data. A subject is classified as having monotonic preferences if her valuation of all pieces of information is weakly largest. WTP is defined as the largest amount of money a subject would pay for information. WTP of subjects who have switch-backs is calculated as the WTP of the closest, in monetary terms, pattern of behavior without switch-backs.

Table A.1: Correlates of the WTP for information estimates of the treatment effect on the untreated. This is consistent with selection based on earnings. Finally, we observe that students perceive that earnings in some careers that they have not chosen are large. Indeed, the difference between the maximum earnings a student could obtain given her beliefs and the expected earnings according to her expected choices is N75,000 (median N30,000). This is consistent with significant perceived nonpecuniary benefits of the chosen careers or significant barriers to education.

Estimates prior to receiving information

	Treatment on the treated		Treatment on the untreated	
	mean	s.e.	mean	s.d
Senior HS	10,953	941	11,883	1,020
Arts	68,723	2,365	58,552	2,071
Commerce	86,969	2,719	79,803	2,804
Science	121,804	3,478	98,283	3,383

Estimates after receiving information

	Treatment on the treated		Treatment on the untreated	
	mean	s.e.	mean	s.d
Senior HS	23,629	1,889	25,640	1,903
Arts	31,025	2,040	21,620	1,963
Commerce	50,508	2,573	43,354	2,821
Science	65,946	3,254	56,537	3,160

Table A.2: Perceived treatment effects of educational choices

Notes: Treatment effect of the treated = weighted mean of earnings at 30 relative to junior high school (weight = probability of choosing the career path). Treatment effect on the untreated = weighted mean of earnings at 30 relative to junior high school (weight = probability of not choosing the career path). Calculations are based on data from the information treatment group ($N = 1925$). The top panel uses data from the first round of belief elicitation and the bottom panel uses data from the second round of belief elicitation. Earnings expectations are winsorized at 1%. Probability beliefs are bounded between 0.01 and 0.99.

A.3 Characterizing compliers

This section provides additional information on treatment effects and their relationship with WTP for information. Table 5 tests whether the size of the information treatment effect varies with the WTP for information. The regression uses the WTP for both pieces of information as a moderator. Figure A.3 graphically shows the implied information treatment effects of the model (using estimates from Table 5, column 2). As theoretically predicted,

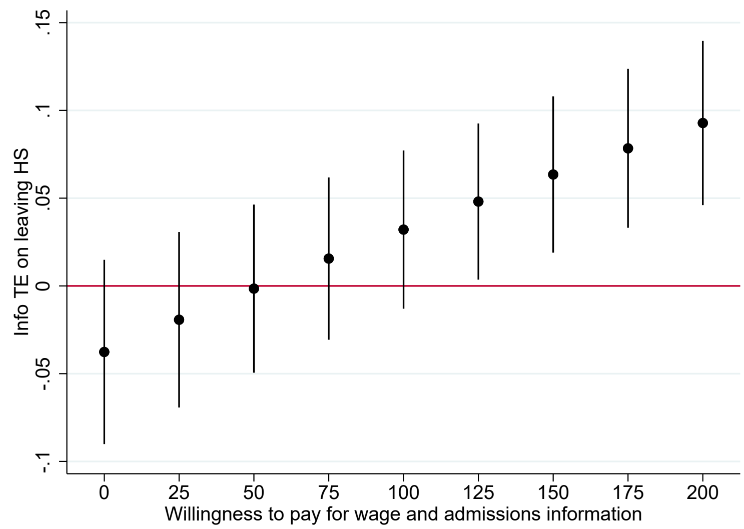


Figure A.3: TREATMENT EFFECTS BY WTP

Notes: The treatment effects for each value of WTP are calculated using the estimates in column (2) of Table 5. The dependent variable is equal to 1 if a student does not register for senior high school, and 0 otherwise.

those who were more willing to pay for information reacted to the information treatment more strongly. While the effect on the treated is 3.7 percentage points, the effect on those willing to pay N200 for both pieces of information is almost 10 percentage points. Column (3) in the Table shows that the result is robust to the inclusion of additional moderating variables. Table 5 allows us to calculate how much WTP is likely to increase due to an increase in expected behavior. We have that a 100% increase in WTP is associated with a 70% increase in the probability of discontinuing education. Inversely, a doubling of the probability of discontinuing education increases WTP for information by 143%.[54]

We can confirm this pattern by analyzing the characteristics of those responding to the information intervention (compliers) under monotonicity. Figure A.4 shows the WTP for each type of information for the different response types. We can see that compliers have a higher WTP for each type of information. Table A.3 shows the characteristics of the different response types for a larger set of variables. Since the complier group is relatively small ($\sim$ 4%), a potential concern is that the results are due to chance. To test whether the differences are significant, we calculate the mean characteristics of compliers and non-compliers using 5,000 bootstrap samples. The penultimate column in the table shows the percentage of times the mean of a variable for the complier population was larger than the mean of the variable for the non-complier population. A two-sided test of significance at the

[54]We reject the hypothesis that the coefficient on the interaction term of information treatment and WTP is equal to 0.1 (p-value = <0.001).

	All	Always-Taker	Never-Taker	Compliers	Pr(Complier ≥ Non-Complier)	$Pr(\frac{WTP_{Complier}}{WTP_{Non-Complier}} \leq \frac{Pr(dropout\|T=1)}{Pr(dropout\|T=0)}$
Percent of population	1.00	0.09	0.87	0.04		
	Mean of each characteristic					
Age	14.03	14.27	13.99	14.51	0.82	
Female	0.47	0.52	0.45	0.83	0.89	
Mother ed. (years)	12.54	13.04	12.51	11.81	0.38	
Christian	0.38	0.35	0.36	0.87	0.98	
Siblings (No.)	3.92	4.03	3.89	4.41	0.58	
Two-parent HH	0.78	0.78	0.79	0.49	0.09	
Grades	0.62	0.62	0.62	0.56	0.16	
Repeated a grade	0.10	0.11	0.09	0.41	0.99	
Discount factor	0.50	0.48	0.51	0.34	0.10	
Lottery investment	40.41	38.07	41.16	27.16	0.11	
Monotonic	0.69	0.60	0.69	1.07	0.95	
WTP admissions	108.31	105.36	105.73	167.91	0.99	0.22
WTP salaries	103.83	106.47	102.16	137.77	0.93	0.47
WTP both	105.60	101.34	102.44	192.75	0.99	0.09

Notes: This analysis is done at the student level. Row 1 shows the share of each compliance group in the sample. The shares of each compliance group are slightly different for the data on WTP since this was collected only for the non pure control conditions. The remaining rows show the means of each student or school characteristic across the different subgroups. The last column shows the probability that the mean characteristic of the complier group is larger than the mean characteristic of non-compliers. P-values are calculated using 5000 bootstrap samples. The mean characteristic uses the approach suggested by Marbach and Hangartner (2020).

Table A.3: Characteristics of response types

10% level corresponds to these proportions being less than 5% or above 95%. We find that the mean of the WTP is significantly larger for compliers for all measures except the WTP for salary information, which is only one-sided significant. Compliers are also more likely to be monotone (no switch-backs in the elicitation task). This adds to the evidence in favor of the instrumental value of information. We also find that compliers are more likely to have repeated a grade and are more likely to declare themselves Christian. Overall, we find that WTP is a good predictor of the effect of information on behavior. The last column of Table A.3 test the hypothesis that the increase in WTP of compliers with respect to non-compliers is proportional to the increase in dropout rates. We cannot reject the hypothesis that this relation is proportional for information on salaries and college admission, but we can reject it for the WTP for both pieces of information (p-value = 0.09). We note that the WTP

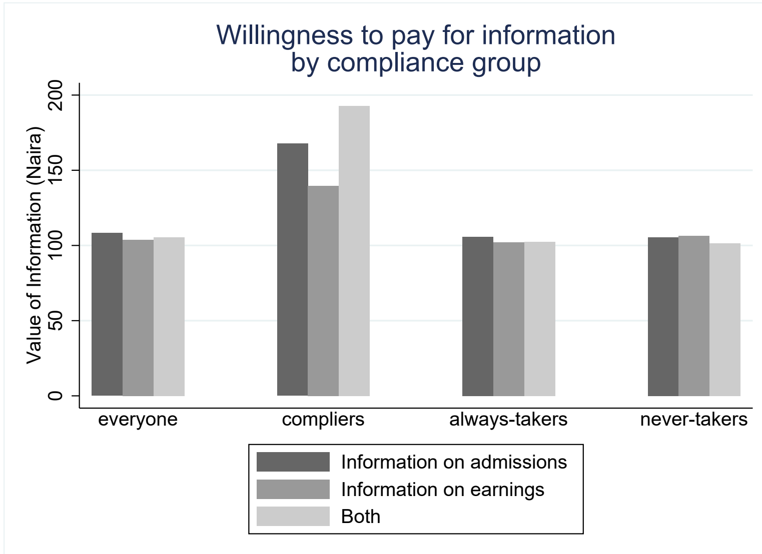


Figure A.4: WTP by response type

Notes: Estimates of characteristics for each information treatment response type are calculated using Marbach and Hangartner (2020)'s approach. The graph includes data from the information treatment and impure control groups.

of compliers is likely to over-estimate the relationship between WTP and expected changes in behavior since it concentrates on those who ex-post change their behavior not those who would change behavior ex-ante.

A.4 The instrumental value of information

Section 2 shows that if students acquire information rationally, we should observe a relationship between the WTP for information and the size of the information treatment effect. Similarly, we should observe that those who react to the intervention (compliers) value the information more. We should also expect that willingness to pay for information is increasing in self-reported expected changes in behavior.

A.4.1 Beliefs

Equation (3) in Section 2 provides a variational representation of the value of information. As discussed there, this equation cannot be evaluated directly because we never observe counterfactual behavior with and without information.[55] Our survey, however, does provide those counterfactuals since they are elicited before and after the information is released. It stands to reason that WTP for information should be larger for those whose beliefs about

[55]Equation (3) in Section 2 generalizes to the case of more than two choices.

career choices change the most.

Table A.4 estimates the relationship between several measures of belief change of career choice on the willingness to pay for information. We measure changes in beliefs using the L1 norm, the L2 norm, and the Kullback-Leibler divergence. We find that both the L2 norm and the Kullback-Leibler divergence are positively related to the willingness to pay for information. This is consistent with the theoretical prediction that those who update their choices the most are willing to pay more for information.

Since the relationship between the value of information and changes in behavior depend on priors, individual preferences and risk attitudes, we take this result as encouraging.[56] Indeed, this makes clear that access to data on beliefs and expected choices as well as willingness to pay for information can be used to improve structural estimations of human capital accumulation models via equation (3) in Section 2. This is a topic of current research.[57]

[56] Frankel and Kamenica (2019) show that no metric can be a valid measure of information. While the Kullback-Leibler is a valid measure of information, it is not clear it is appropriate in this context. As shown in Section 2 the value of information varies with expected changes in behavior under risk neutrality. We take these results as qualitative evidence that willingness to pay for information is related to expected changes in behavior. Their paper focus on ex-post measures of information.

[57] As shown in Section 2, WTP has a closed form solution if one is willing to make assumptions about the uncertainty faced by participants. For instance, if one assume extreme value errors, the ex-ante gross payoff function has a known expression (McFadden, 1978). Importantly, since information structures are equivalent to mean preserving spreads of prior beliefs, this implies that those willing to pay more for information should also be more responsive to information. This implies that the elasticities estimated in Table 4 should be increasing in WTP. The belief data is consistent with this prediction. The elasticity of those with WTP above 100 is 0.36 (s.e. 0.16), and the elasticity of those WTP below 100 is 0.26 (s.e. 0.13).

	WTP	WTP	WTP
	(1)	(2)	(3)
L1	3.836		
	(3.489)		
L2		5.928*	
		(2.880)	
KL			2.576*
			(1.228)
Observations	1209	1209	1209
schools	18	18	18
R2	0.042	0.044	0.044

* p<0.10, ** p<0.05, *** p<0.010

+ clustered at the school level

Table A.4: Belief changes and the value of information

Notes: The dependent variable is equal to WTP for all pieces of information. Each column estimates the effect of a measure in changes in prior and posterior beliefs about career choices. The definition of career choices are the same as in Table 4. The first column uses the L1 norm to measure changes in prior and posterior beliefs of career choices. The second column uses the L2 norm and the third column uses the Kullback-Leibler divergence. All regressions include a dummy for females, age, a dummy for a two-parent household, mother's education, average investment in the lotteries and discount rate. The sample is restricted to those receiving information and those whose beliefs answers add up to 100±5.

B Willingness to pay for information - Instructions

Let's play the following game! From time to time, we will stop during our explanation and allow you to ask questions. The purpose of this game is to help us understand how much value you place on certain information. Remember, if any part of the game makes you feel uncomfortable, you can talk to your school counselor or principal about it. They will be able to help you.

Consider that I want to sell information to you. The information can help you to make better decisions about choosing between science, arts or commercial class in SS1. This type of information is important because it can affect what you become in future. The information tells you the chances of getting admitted into higher institutions based on the type of class you choose or information on average salary for different professions and chances of working full time. For example, you will know the percentage of boys and girls that apply to study Arts, commercial and science courses and what percentage of them get admitted. Imagine you have N200, I would like to know if you will be willing to exchange the money for information. I will offer an amount of money as shown in the table below.

You will play three versions of the game, but only one will be used to pay you. Once you have made the decision for each of the three rounds, we will choose a number from one to three by randomly choosing from numbered balls in a bag. One of your classmates will be the one to pick the ball. The number on the ball chosen will determine which of one of the three versions of the game will be chosen to pay you. Next, I will present the class with 10 cards in a bag which represents prices drawn from N0 to more than N200, and someone in your class will be asked to pick one (the person will not know which card represents what price and I also do not know). The price on the card chosen will be used to determine if you get the information or not. If the price that is drawn from the bag is less than what you select as your value for the information, you will pay the drawn price and receive the information. If, however, the drawn price is strictly greater than what you choose as your valuation, then you do not get the information but keep your money. Think carefully about each decision.

Please listen carefully to the following example of this game: Dele is willing to buy the information at N125 and no more. So, he chooses "yes" for prices N0 - N125 (rows A-F) and chooses "no" for prices 150 and above (options G through I). We present him with a bag that has cards drawn from N0 to more than N200, he puts his hand in the bag and chooses price N0. Since the price N0, is always less than any amount he could have chosen, he will receive the information and keep his N200. And if N250 is chosen, he will not get the information since it is larger than any amount he could have paid.

Do you accept the price?	Yes	No
A: Price: N0 (means you pay: N0) You receive the information AND keep a payment of N200	X	
B: Price: N25 (means you pay: N25) You receive the information AND keep a payment of N175	X	
C: Price: N50 (means you pay: N50) You receive the information AND keep a payment of N150	X	
D: Price: N75 (means you pay: N75) You receive the information AND keep a payment of N125	X	
E: Price: N100 (means you pay: N100) You receive the information AND keep a payment of N100	X	
F: Price: N125 (means you pay: N125) You receive the information AND keep a payment of N75	X	
G: Price: N150 (means you pay: N150) You receive the information AND keep a payment of N50		X
H: Price: N175 (means you pay: N175) You receive the information AND keep a payment of N25		X
I: Price: N200 (means you pay: N200) You receive the information AND keep a payment of N0		X

Do you have any questions?

Consider that I want to sell information to you.

Information to sell: The chances of people getting admitted into university based on the type of class they choose in SS1. For example, you will know the percentage of boys and girls that apply to study Arts, commercial and science courses in university and what percentage of them get admitted. This information is from JAMB.

Imagine you have N200, I would like to know how much of the N200 you would like to use to buy the information. Think carefully on how much you value this information and respond by marking X in the relevant column.

Do you accept the price?	Yes	No
A: Price: N0 (means you pay: N0) You receive the information AND keep a payment of N200		
B: Price: N25 (means you pay: N25) You receive the information AND keep a payment of N175		
C: Price: N50 (means you pay: N50) You receive the information AND keep a payment of N150		
D: Price: N75 (means you pay: N75) You receive the information AND keep a payment of N125		
E: Price: N100 (means you pay: N100) You receive the information AND keep a payment of N100		
F: Price: N125 (means you pay: N125) You receive the information AND keep a payment of N75		
G: Price: N150 (means you pay: N150) You receive the information AND keep a payment of N50		
H: Price: N175 (means you pay: N175) You receive the information AND keep a payment of N25		
I: Price: N200 (means you pay: N200) You receive the information AND keep a payment of N0		

Again, consider that I want to sell information to you.

Information to sell: The average salary for people that go to different classes in SS1 and the chances that they are working full time. For example, you will know the average salary for boys and girls who studied different courses in university and are now working and the chances that they are working full time.

Imagine you have N200, I would like to know how much of the N200 you would like to use to buy the information. Think carefully on how much you value this this information and respond by marking X in the relevant column.

Do you accept the price?	Yes	No
A: Price: N0 (means you pay: N0) You receive the information AND keep a payment of N200		
B: Price: N25 (means you pay: N25) You receive the information AND keep a payment of N175		
C: Price: N50 (means you pay: N50) You receive the information AND keep a payment of N150		
D: Price: N75 (means you pay: N75) You receive the information AND keep a payment of N125		
E: Price: N100 (means you pay: N100) You receive the information AND keep a payment of N100		
F: Price: N125 (means you pay: N125) You receive the information AND keep a payment of N75		
G: Price: N150 (means you pay: N150) You receive the information AND keep a payment of N50		
H: Price: N175 (means you pay: N175) You receive the information AND keep a payment of N25		
I: Price: N200 (means you pay: N200) You receive the information AND keep a payment of N0		

Finally, consider that I want to sell information to you.

Information to sell: The chances of people getting admitted into university based on the type of class they choose in SS1 and the average salary for people that go to different classes in SS1 and the chances that they are working full time. For example, you will know the percentage of boys and girls that apply to study Arts, commercial and science courses in university and what percentage of them get admitted. This information is from JAMB. Also, you will know the average salary for boys and girls who studied different courses in university and are now working and the chances that they are working full time. It combines the two types of information I previously sold to you.

Imagine you have N200, I would like to know how much of the N200 you would like to use to buy the information. Think carefully on how much you value this this information and respond by marking X in the relevant column.

Do you accept the price?	Yes	No
A: Price: N0 (means you pay: N0) You receive the information AND keep a payment of N200		
B: Price: N25 (means you pay: N25) You receive the information AND keep a payment of N175		
C: Price: N50 (means you pay: N50) You receive the information AND keep a payment of N150		
D: Price: N75 (means you pay: N75) You receive the information AND keep a payment of N125		
E: Price: N100 (means you pay: N100) You receive the information AND keep a payment of N100		
F: Price: N125 (means you pay: N125) You receive the information AND keep a payment of N75		
G: Price: N150 (means you pay: N150) You receive the information AND keep a payment of N50		
H: Price: N175 (means you pay: N175) You receive the information AND keep a payment of N25		
I: Price: N200 (means you pay: N200) You receive the information AND keep a payment of N0		