KU LEUVEN

DEPARTMENT OF ECONOMICS

The Potential of Tutoring in Higher Education: Students' Preferences, Consumption and the Role of Information.

Willem De Cort and Kristof De Witte

FACULTY OF ECONOMICS AND BUSINESS



DISCUSSION PAPER SERIES DPS 24.13

October 2024

The Potential of Tutoring in Higher Education: Students' Preferences, Consumption and the Role of Information

Willem De Cort and Kristof De Witte*

Abstract

Tutoring programs are highly effective in improving students' outcomes in higher education. However, little is known about students' demand for tutoring or the optimal design of tutoring programs. Additionally, while privately provided tutoring can threaten social mobility, little is known about students' consumption of private tutoring. Using a discrete choice experiment, this paper estimates the preferences, perceptions and consumption of tutoring for 1,200 Flemish first-year higher education students. We find that students' willingness-to-pay for tutoring is high, as they perceive it to be highly effective. However, they perceive the added value of having an experienced lecturer and smaller class sizes in these tutoring programs as low. Our simulations suggest that tutoring could significantly improve private and social welfare, but that the optimal size of tutoring groups is larger than commonly considered. Despite its potential, only 8% of students purchase privately provided tutoring. Using an information experiment, we show that biased perceptions of the returns to education cannot explain why students underinvest in tutoring. We argue that supply-side restrictions and behavioral biases such as status-quo bias and social image concerns can explain this low uptake of private tutoring. Finally, the demand for tutoring is highest among students with low socioeconomic status, but they are less able to purchase tutoring on the private market due to its high prices. These findings suggest that public investments in tutoring could increase social welfare and social mobility cost-effectively, while lowering barriers to students seeking academic support.

Keywords: Tutoring; Higher Education; Choice Experiment; Information Experiment
JEL Codes: H52; I21; I23; I28

1 Introduction

Most higher education students in OECD countries do not graduate within the prescribed duration of their bachelor's program, while more than a fifth of students fail to graduate at all (OECD, 2022). This results in substantial costs to students and society, including lower lifetime earnings and an increased public cost of higher education (Blundell et al., 2022). Recent

*De Cort: KU Leuven, Leuven, Belgium (corresponding author: willem.decort@kuleuven.be)(orcid.org/0000-0001-6106-4890); De Witte: KU Leuven, Leuven, Belgium & UNU-Merit, Maastricht University, Maastricht, The Netherlands (orcid.org/0000-0003-0505-8642).

This work was supported by the Flemish Research Foundation FWO (grant number 11D0723N), and the Horizon Europe project EFFEct (grant 101129146). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the granting authority. Neither the European Union nor the granting authority can be held responsible for them. Approval was granted by the Social and Societal Ethics Committee (project number G-2022-5566). We are grateful to Matthew Kraft and numerous conference and seminar participants.

evidence shows that tutoring - defined as supplemental course-specific instruction either individual or in smaller groups - is a potentially highly effective intervention for improving students' higher education outcomes (Gordanier et al., 2019; Hardt et al., 2023). However, little evidence exists on the demand for tutoring programs, or on their optimal design and cost-effectiveness (Kraft and Falken, 2021). Additionally, while a growing private market for tutoring threatens the social mobility function of K-12 education (Byun et al., 2018), little research has examined higher education students' consumption of privately provided tutoring.

This paper investigates the potential of tutoring in higher education to improve educational outcomes and social mobility. Rather than evaluating one specific tutoring program, we estimate students' preferences for and perceptions of a multitude of tutoring programs. We then combine these estimates with information on tutoring programs' costs and returns to determine their cost-effectiveness. Additionally, we examine students' consumption of privately provided tutoring, and examine the role of information in this decision-making process.

First, we employ a discrete choice experiment to estimate students' preferences for in-person tutoring and their perceptions of its effectiveness. Students' preferences shape the effectiveness of tutoring programs because they determine whether students attend these programs. Additionally, students' preferences determine the supply of tutoring by both private providers and higher education institutions. Nevertheless, no prior research has estimated students' preferences for tutoring. We use the results to simulate whether investing in tutoring could improve private and social welfare, and how this tutoring should be structured to maximize welfare gains. We focus on the optimal class size and the tutor's teaching experience. These factors are instrumental in determining the effectiveness of tutoring, managing its costs and ensuring its feasibility when implemented at scale (Kraft and Falken, 2021; Nickow et al., 2020). Additionally, we measure students' consumption of privately provided tutoring and assess how this private tutoring might affect social mobility. Finally, we examine whether students might underinvest in private tutoring due to informational biases on the returns to education, which are known to cause sub-optimal decisions related to, e.g., financial aid, enrolment, and major choices (Damgaard and Nielsen, 2018). To test this, we design an information experiment that randomly assigns students information about the returns to educational attainment and estimates the effect of this information on students' willingness-to-pay for tutoring.

We distribute these experiments to over 1,200 first-year students from three higher education institutions in Flanders, a region which has one of the highest expenditure rates per student among OECD education systems (OECD, 2022). A large demand for additional tutoring in Flanders therefore suggests a similar potential in countries that spend less on higher education.

The remainder of this paper is structured as follows. Section 2 describes the design of the experiments. Section 3 describes the econometric framework and sample. Section 4 describes the results of the analysis, while Section 5 further discusses the results and concludes.

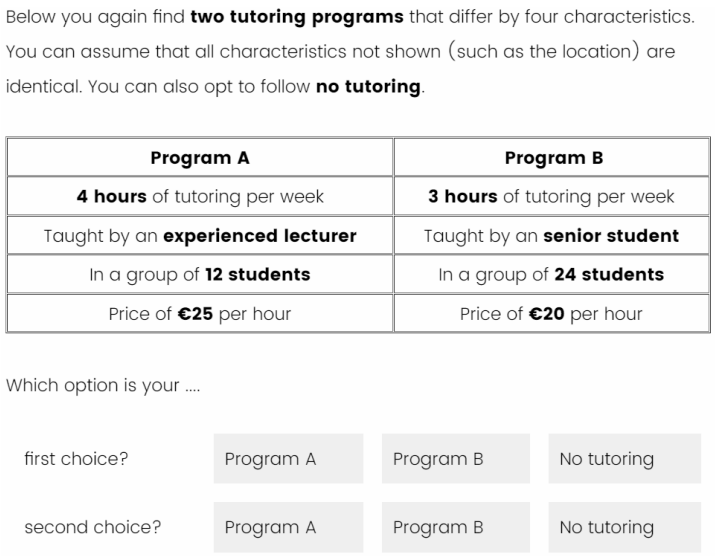
2 Experimental design

To examine our research questions, we distributed a survey containing a discrete choice experiment, questions about the perceived effectiveness of tutoring, and an information experiment.

2.1 The discrete choice experiment

In our discrete choice experiment, respondents repeatedly choose between two different tutoring packages as well as the option to follow no tutoring. Four attributes describe the tutoring packages. The first three attributes are the most critical determinants of tutoring's effectiveness: the amount, the tutor and the class size (Nickow et al., 2020). The amount of tutoring ranges

Figure 1: Example of a choice set from the discrete choice experiment



from 1 hour per 2 weeks to 4 hours per week. The tutor could either be an experienced lecturer or a senior student. The class sizes range from one-on-one lessons to groups of 64 students. This wide range of class size values reflects the diversity of class sizes in higher education (Ake-Little et al., 2020). The final attribute is the price per hour of tutoring. This allows us to estimate students' willingness-to-pay for different tutoring packages. The values of the price attribute range from €0 to €50 per hour, with €50 being a commonly observed market price in Flanders for one-on-one tutoring by an experienced lecturer. Analyzing these choices allows for a precise identification of the respondents' preferences for tutoring programs. The estimated preferences correspond well with actual behavior (Johnston, 2020).

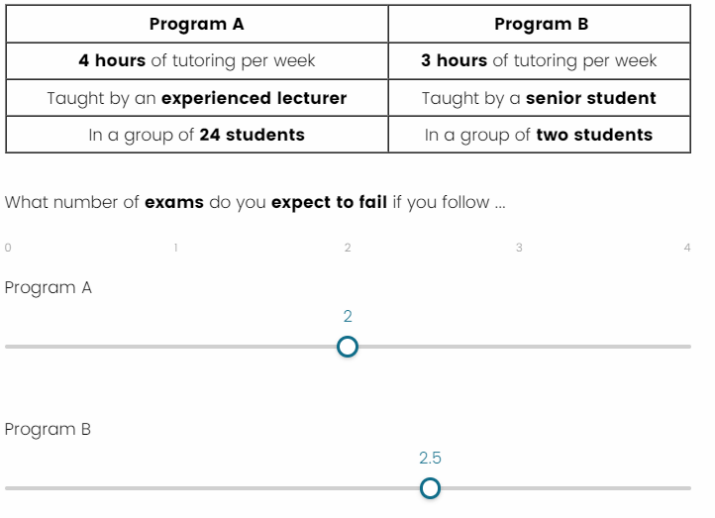
An example of a choice set is included in Figure 1. Further details on the experimental design such as the attention check as well as the full list of attribute values can be found in the Supplementary Materials Section 2.1.

2.2 Perceived effectiveness of tutoring

Students reported their expected GPA or the number of courses they expect to fail conditional on following different tutoring programs from the choice experiment, as shown in Figure 2, and conditional on following no tutoring. This allows us to estimate how students expect that different tutoring programs will improve their educational outcomes.

Additionally, we measure students' perception of the relationship between the number of failed courses this term and their probability of graduating (in time). However, a direct question might be biased as failed exams are a signal of students' ability. We control for this bias by asking students to estimate their perceived probabilities of graduating (in time) both when they pass and when they fail a certain number of courses by the smallest possible margin, which is considered a matter of chance. Further details on this question as well as the empirical framework are provided in the Supplementary Materials Section 2.2.

Figure 2: Example of a question about the effectiveness of different tutoring programs



2.3 The information experiment

Privately provided tutoring could be a good investment if it increases students' probability of graduating (in time) by a large enough margin to justify its costs. However, students would underinvest in tutoring if they underestimated the returns associated with graduating (in time), as for example in Jensen (2010). This information experiment evaluates how informing students about the benefits of tutoring affects their preferences for tutoring.

Students are randomly assigned to a treatment or control group. The treatment informs students using a short quiz at the beginning of the survey that failing to graduate (in time) is common and has important economic consequences. Students received the correct information after providing their response. This 'gamified' format pushes students to actively process the information, while also measuring students' initial beliefs. The control group receives no additional information.

The first question demonstrates the percentage of higher education students that fails to graduate. The second question informs students about the labor income premium associated with their degree. The final question highlights the foregone labor income due to not graduating in time. For all questions, the correct answer was the one that implied the largest benefits of tutoring (see Supplementary Table 1). This approach ensured that students were less likely to overcorrect their beliefs when viewing the potential answers, which could lead to the treatment backfiring (Damgaard and Nielsen, 2018).

3 Methodology & Sample

3.1 Empirical framework

Preferences for tutoring

We analyse students' responses to the choice experiment using a mixed logit model with correlated coefficients. The model assumes that preference coefficients differ between individuals but are drawn from a common distribution, allowing us to describe the amount of heterogeneity in

the coefficients. It is the most commonly used and flexible model (Vass et al., 2022).

Assume that the utility U_{ni} of individual n is an additive function of the attributes of a tutoring program X_i, being its price per hour, class size, tutor type, amount of tutoring and a tutoring constant[1], with individual-specific preference coefficients β_n and an error term ϵ_{ni}. In a second specification, we also include an interaction effect between tutor-type and class size and a second-order polynomial to describe preferences for class size.

$$U_{ni} = \beta_n' X_i + \epsilon_{ni}$$

Assuming that ϵ_{ni} is i.i.d. extreme value distributed and that β_n is not known to the researcher but is drawn from a distribution $f(\beta)$ leads to the following unconditional probability that a program i is chosen out of the set of alternatives J (Train, 2009):

$$P_{ni} = Prob\Big(U_{ni} > U_{nj} \quad \forall j \in J\Big)$$

$$P_{ni} = \int \left(\frac{e^{\beta' X_i}}{\sum_{j=1}^{J} e^{\beta' X_j}} \right) f(\beta) d\beta$$

In specifying $f(\beta)$, we follow earlier literature. All coefficients are drawn from a random distribution, except for interaction variables (Yoo and Ready, 2014). For variables of which we can assume that their effect is non-negative, we decided between a normal or log-normal distribution using the log-likelihood (Revelt and Train, 1998; Train, 2009). Other random variables are specified to be normally distributed. For each attribute, the mixed logit model estimates a mean effect, a standard deviation, and the correlation with other coefficients. The standard deviation represents the amount of unobserved heterogeneity in the coefficient's size among students.

We estimate the parameters that define these distributions of preference coefficients by numerically approximating the maximum simulated likelihood function (Hole, 2007). We then estimate the average willingness-to-pay per hour for a tutoring program WTP_i using the estimated mean preferences for the attributes $\overline{\beta}$ as follows (Train, 2009):

$$WTP_i = \frac{\overline{\beta_1} ClassSize_i + \overline{\beta_2} TutorType_i + \overline{\beta_3} Amount_i + \overline{\beta_4} TutoringConstant_i}{\overline{\beta_5} Price_i}$$

We additionally estimate how students' preferences for tutoring differ by their gender, grades, academic year, father's education level and consumption of private tutoring. For each background characteristic, we estimate a model including interaction effects γ_n of that background characteristic z_n with each attribute from the tutoring program X_i:

$$U_{ni} = \beta_n' X_i + \gamma_n' X_i z_n + \epsilon_{ni}$$

Stated preference methodologies like discrete choice experiments can suffer from hypothetical bias, i.e. students not reporting their true preferences for example due to social desirability bias. As a robustness check, we compare students' WTP with the expected returns to tutoring, conditional on students' own perception of the effectiveness of tutoring, and to total private spending on tertiary education. Supplementary Materials Section 2.3 describes in detail the possible sources of hypothetical bias as well as the methodology used to assess its role.

[1] This 'tutoring constant' is the preference for a baseline program of 1 hour per 2 weeks of individual tutoring by a senior student relative to following no tutoring.

Perceived effectiveness and the role of information

Additionally, we estimate how students expect that a tutoring program would improve their educational performance Y_{ni}, being either the expected number of exams failed or their expected GPA, relative to following no tutoring Y_{n0}, as measured by the type of question shown in Figure 2. This improvement, $Y_{ni} - Y_{n0}$, is estimated as a function of the program's class size, tutor type, and amount, as well as a constant. Tutor type and class size are interacted with the amount of tutoring, as the effectiveness of, e.g., an experienced lecturer at improving GPA depends on the amount of tutoring a student receives from that experienced lecturer.[2]

$$Y_{ni} - Y_{n0} = \beta_0 + \beta_1(ClassSize_i \times Amount_i) + \beta_2(TutorType_i \times Amount_i) + \beta_3 Amount_i + \epsilon_{ni}$$

The information treatment would only change students' preference for an attribute to the extent that they perceive this attribute to be effective at improving their educational performance. Hence, we directly assess the impact of the treatment on students' preferences for tutoring programs they anticipate will enhance their success rate in passing exams. More specifically, we estimate students' choices in the choice experiment as a function of the price, the amount of tutoring, and the student's perceived effectiveness of that tutoring package at reducing their number of retakes as measured by the type of question shown in Figure 2. We then interact these variables with a dummy variable indicating whether they received the treatment.

$$\begin{aligned} U_{ni} = \beta_{n1} ExamsPassed_{ni} + \beta_{n2} Price_i + \beta_{n3} Amount_i + \\ Treatment_n \times (\beta_{n4} ExamsPassed_{ni} + \beta_{n5} Price_i + \beta_{n6} Amount_i) + \epsilon_{ni} \end{aligned}$$

3.2 Setting and sample

The final survey contained the information experiment, the discrete choice experiment, questions about the perceived effectiveness of tutoring, questions on students' background characteristics and questions on students' consumption of private tutoring (see Supplementary Table 2 for an overview). Flanders has both research-oriented universities (RU) and universities of applied sciences (UAS). This second type of institution provides professional and vocational education at the bachelor level, as for example in Germany, Switzerland, and Finland. We distributed this survey in three Flemish higher education institutions, two of which are UAS. One of these UAS distributed the survey to students during class time in a wide selection of study programs and study years to achieve a representative sample and avoid selection bias. The other UAS advertised the survey to its students providing financial lottery-type incentives for participation. Both data collection methods were used for the third institution, a more traditional RU. Data collection was restricted to courses taught in the first year of bachelor's programs in economics, business, and engineering.

Our total sample consists of 1,976 students after dropping those that failed the attention check. Their characteristics are described in Appendix Table A1. Given the critical differences between UAS and RU students, they are analyzed separately. Furthermore, we limit the main analyses to first-year students, as they are most likely to struggle academically (Bovill et al., 2011).

[2]Interaction effects with the amount of tutoring are not necessary in estimating students' preferences as we express these preferences in willingness-to-pay *per hour*, while we express the perceived effectiveness in educational improvement *per semester*.

4 Results

4.1 Preferences for and perceived effectiveness of tutoring

Preferences for tutoring

Understanding students' preferences is crucial to optimally designing and scaling higher education tutoring interventions. Figure 3 plots the mean WTP estimates for six different tutoring packages for UAS and RU students from the main mixed logit analysis of the choice experiment.

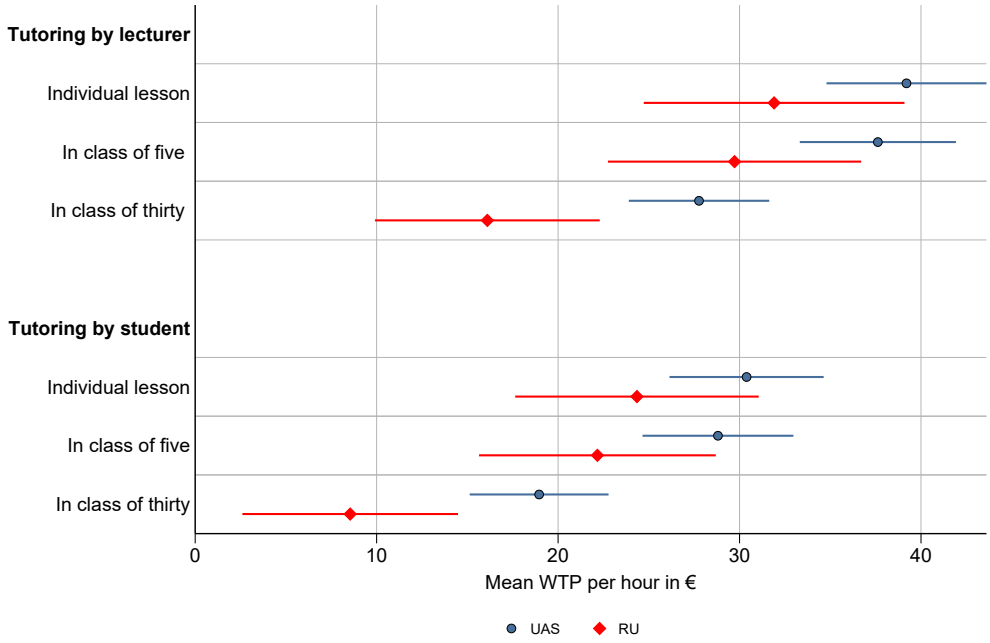
Figure 3 shows how most students strongly prefer having supplementary tutoring, as demonstrated by their sizeable WTP for the different tutoring packages. We observe that the average RU student has a WTP of €29.7 per hour for tutoring by an experienced lecturer in a small class setting of five students. Additionally, while the average RU student has no preference for the amount of tutoring per week, UAS students have on average a small preference for more tutoring. The substantial standard deviations observed for the tutoring constant point towards substantial heterogeneity in the preferences for tutoring. In particular, the mixed logit model predicts that 27% of students would prefer no tutoring over a tutoring program taught by an experienced lecturer in a class of five even at no cost. These students might believe that individual study time is more effective.

Figure 3 further indicates that students' WTP for higher-quality tutoring is modest relative to the increase in costs. RU and UAS students are willing to pay €9.1 and €14.5 per hour extra respectively for a tutoring program with an experienced lecture as a tutor rather than a senior student. Decreasing class sizes from five students to individual tutoring only increases students' WTP per hour by about €2 for both RU and UAS students, while costs would increase by about a factor of 5. These results change little when we include interaction effects between class size and tutor type and the second-order polynomial for class size preferences as shown in Supplementary Table 5. Both extra coefficients are either statistically insignificant or small. For reasons of parsimony and consistency, the main effects model will stay as our baseline specification, though the results of later sections are robust to different specifications.

The results of the heterogeneity analyses are presented in Supplementary Table 6. These analyses include second and third year students to increase power and to assess differences with respect to academic year. Students who currently pay for private tutoring have a significantly stronger preference for tutoring, affirming the validity of the choice experiment. Additionally, female students, first-year students, students without a higher-educated father, and students with a lower GPA have a stronger preference for tutoring. The fact that mostly weaker students and lower SES students would utilize tutoring aligns with the literature on secondary education (Byun et al., 2018; Dang and Rogers, 2008). Tutoring could therefore reduce differences in performance between students, as in Hardt et al. (2023).

As a robustness check to the discrete choice experiment, we assess whether students overstated their WTP for tutoring due to hypothetical bias. To do so, we compare students' WTP with the expected returns to tutoring, conditional on students' own perception of the effectiveness of tutoring, as further elaborated in Supplementary Materials Section 2.3. We find that we would expect students' WTP for tutoring to be significantly *larger* than estimated, given their perception of its effectiveness. Additionally, we find that implementing the estimated WTP implies a private share of spending on higher education that is still relatively small at 15.2% compared to the OECD average of 30.4%. While these findings do not entirely rule out the possibility of hypothetical bias, they instill confidence that the obtained estimates of students' WTP are not unrealistically high. The results in the following sections concerning welfare gains and behavioral biases would be more robust if the private welfare gains to tutoring are indeed larger than students' estimated WTP.

Figure 3: Average WTP for different tutoring packages



Notes: WTP estimates are derived from the mean coefficients of the mixed logit analysis of the discrete choice experiment using main effects, as reported in Supplementary Table 3. 95% confidence intervals are estimated using the delta method. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students.

Perceived effectiveness of tutoring

Students will only want to attend tutoring sessions if they perceive it to increase their probability of passing the exams, and, therefore, increase their probability of graduating (in time). This subsection estimates students' perception of the effectiveness of tutoring in increasing their educational performance as a function of its amount, class size, and tutor type.

The results in Table 1 demonstrate that students perceive tutoring to be highly effective in improving their educational performance. For example, one-to-one tutoring provided by a senior student for 1 hour per two weeks is perceived by UAS students to yield a substantial decrease in the number of exams to retake, as indicated by the significant coefficient for the tutoring constant. The magnitude of this effect, approximately 0.63 standard deviations (SD) or 1.32 less exams expected to retake, is larger than the effect size of the online tutoring program in Hardt et al. (2023) with an effect size of 0.3 SD.

Table 1: Perceived effectiveness of tutoring at reducing retakes and improving GPA

	Retakes (Std.)		GPA (Std.)	
	UAS	RU	UAS	RU
Amount (in hours per week)	-0.045***	-0.082***	0.200***	0.185***
	(0.011)	(0.019)	(0.029)	(0.034)
Tutoring constant	-0.633***	-0.582***	0.468***	0.396***
	(0.021)	(0.041)	(0.067)	(0.086)
Experienced lecturer x Amount	-0.041***	-0.027*	0.039	0.019
	(0.009)	(0.016)	(0.025)	(0.027)
Class size x Amount	0.003***	0.002***	-0.005***	-0.005***
	(0.000)	(0.001)	(0.001)	(0.001)
Number of packages rated	5442	2104	3108	816
Number of respondents	717	256	345	90

Notes: The outcome variables, standardized, reflect students' expected change in the number of retaken courses and overall GPA for the semester when following the tutoring program. On average, UAS students expect 2.10 retaken exams (SD = 1.41) and a GPA of 12.06 out of 20 (SD = 2.00) without tutoring, while RU students expect 2.25 retaken exams and a 11.14 GPA. Independent variables represent the attributes of the alternatives in the discrete choice experiment for which students reported the perceived effectiveness (see Figure 2). UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. * p<0.10, ** p<0.05, *** p<0.01

Students perceive little added value of additional hours of tutoring beyond the baseline of 1 hour per two weeks, with each additional hour resulting in a marginal reduction of only 0.05 SD in the number of retakes. This perception seems contradictory to the general recommendation in the tutoring effectiveness literature in secondary education of having high-dosage tutoring interventions (Kraft and Falken, 2021). However, this can be partially explained by using a different outcome variable, namely whether one passes or fails an exam. Indeed, this non-linearity is less pronounced when considering tutoring's impact on students' GPA.

On the merits of having an experienced lecturer and smaller classes, students' perception of the effectiveness aligns with their preferences. Two hours of one-to-one tutoring provided by an experienced lecturer results in an 11% larger reduction in the number of expected retakes than when provided by a senior student. This modest added value of more experienced tutors aligns with experimental evidence (Feld et al., 2020; Guryan et al., 2023; Nickow et al., 2020).

Increasing class sizes is also perceived to have a small effect on the effectiveness of tutoring. For example, increasing the class size of a tutoring intervention from 1 to 30 students only leads students to expect an 11% smaller decrease in the expected number of retakes.

Finally, on the relationship between the number of exams passed and the probability of graduating (in time), students on average perceive one less failed exam to lead to a 2.6 percentage point rise in the probability of obtaining a degree, all other factors held constant (see Supplementary Table 7). Furthermore, it corresponds to a decrease of 5.5 percentage points in the probability of needing an extra year to complete their degree.

4.2 Welfare analysis and the optimal design of a tutoring intervention

The previous results show that students have a sizeable WTP for tutoring, as they perceive it to be highly effective. This section estimates how tutoring could improve private and social welfare, and how it should be structured to maximize these welfare gains.

Setup of the problem

The optimal tutoring intervention can be defined as the amount of tutoring A, tutor-type T, and class size C that maximizes private welfare, W_p, or social welfare, W_s. We measure private welfare gains by the average WTP of a student for the tutoring intervention as estimated by the choice experiment, $WTP(A, T, C)$. We provide a baseline estimate of the intervention's marginal cost, $MC(T)$, using the universities' annual accounts and market prices for tutoring, described in Supplementary Materials Section 2.4. To express these costs at the student level, they are divided by the class size C. The domain of the attributes of the tutoring intervention is restricted to the values used in the discrete choice experiment.

$$\max_{A,T,C} W_p = WTP(A, T, C) - \frac{MC(T) * A}{C} \tag{1}$$

The social welfare maximization problem additionally considers two types of social returns: (i) reduced expenditure on higher education due to more students graduating on time, R_{GT}, and (ii) higher tax revenue due to more students graduating, R_G. Details on the calculations of these returns using EU-SILC data and the estimation procedure can be found in Supplementary Materials Section 2.4. We describe both the probability of students graduating and graduating in time, P_G and P_{GT}, as a function of the absolute number of exams passed, $E(\tilde{A}(A, T, C), T, C)$, which is a function of the tutoring program. Section 4.1.2. provides a baseline estimate for both these relationships using students' perceptions. Note that the amount of exams passed is a function of the average amount of tutoring that maximizes students' utility, $\tilde{A}(A, T, C)$, not A, thus taking into account non-attendance.

$$\max_{A,T,C} W_s = W_p(A, T, C) + P_{GT}\Big(E\big(\tilde{A}(A, T, C), T, C\big)\Big) * R_{GT} + P_G\Big(E\big(\tilde{A}(A, T, C), T, C\big)\Big) * R_G$$

One can think of many reasons why students overestimate the effectiveness of tutoring. Additionally, estimating the social returns and marginal costs of tutoring requires assumptions that can be questioned. Therefore, we extensively consider the sensitivity of our results to the marginal costs and to the social welfare gains of tutoring.

Optimal tutoring intervention and welfare gains

Table 2 describes the results of the maximization problem for UAS students using the baseline estimates of tutoring's effectiveness and marginal costs. The results for RU students are similar across the board (see Supplementary Table 8).

We estimate that the optimal tutoring intervention would lead to large private and social welfare gains. The socially optimal intervention with an experienced lecturer, consisting of 4 hours of tutoring in a class of 12 students, would lead to a yearly net private welfare increase of €3,540 per student. The yearly net social returns are similar at €3,945 yearly, consisting for 54.2% of increased government revenue due to higher graduations rates and for 45.8% of decreased spending due to faster graduation rates. These social returns are equivalent to 24% of government spending on higher education per student. The cost-effectiveness ratio is similarly remarkable with for example a 11.4 benefit-cost ratio for the socially optimal intervention using an experienced lecturer. Together, these results suggest that tutoring in higher education could be a highly cost-effective investment regardless of how its costs are distributed between students and public resources.

We also find that the optimal class size ranges from 7 to 20 depending on the type of welfare one maximizes and the tutor-type. In contrast, the literature on tutoring effectiveness almost exclusively examined interventions with class sizes smaller than six students (Dietrichson et al., 2017; Gordanier et al., 2019; Hardt et al., 2023; Nickow et al., 2020). These class size estimates are in line with what many higher education institutions are typically already providing in the form of TA-sessions.

Third, we find that interventions employing either experienced lecturers or senior students as tutors can result in nearly identical welfare gains. The higher marginal costs and higher effectiveness of having an experienced lecturer are compensated by smaller class sizes when using senior students as tutors. However, the intervention with student tutors is significantly more cost-effective. Combined with the seemingly easier scalability of finding a sufficient number of tutors, this makes investments in tutoring by senior students perhaps the more attractive alternative.

Finally, the optimal amount of tutoring consistently remains 4 hours per week, representing the upper limit within its domain. This is a consequence of the average student's WTP not diminishing with the amount of tutoring. This result takes into account that a substantial proportion (23%) of UAS students would opt out of any tutoring, even when provided with the socially optimal tutoring program.

Sensitivity analysis and robustness checks

This subsection shows how the main findings hold under significantly more conservative assumptions on tutoring's costs and social returns. It also shows how the findings do not depend on the chosen specification. We focus on using senior students as tutors as this was the most promising intervention according to the earlier results, though the results hold for tutoring with an experienced lecturer.

Figure 4 shows how the cost-effectiveness of public investments in the optimal tutoring intervention by a senior student changes with the factor by which we underestimate marginal costs and overestimate the social returns to tutoring.[3] Even when the social returns to tutoring are five times lower and marginal costs are three times larger than in our baseline estimate, the cost-benefit ratio for public investments remains positive at around 1.5. Note that this calculation only includes the purely social returns, not the private welfare gains. Larger class sizes compensate for these increases in marginal costs (see Supplementary Figure 1).

Finally, the optimal tutoring intervention and welfare gains change very little when we use the mixed logit model with a second-order polynomial for class size and an interaction effect between tutor type and class size (see Supplementary Table 9).

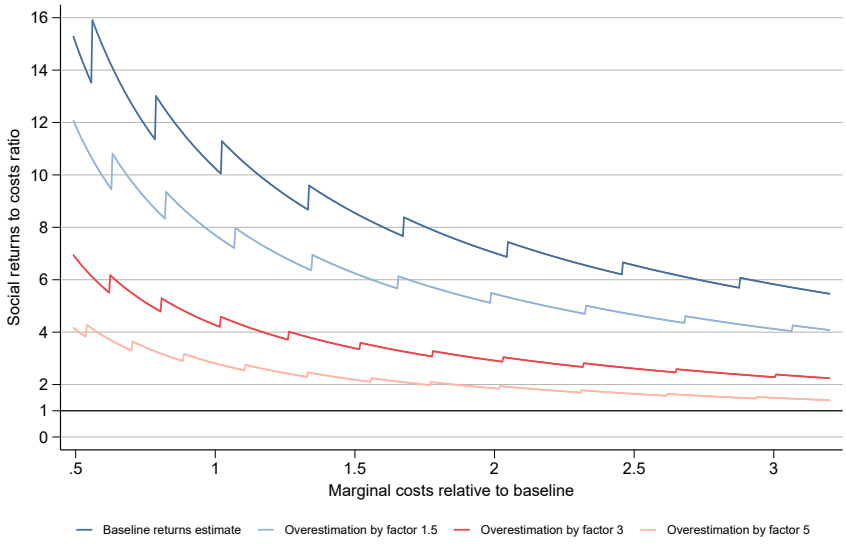
[3]Note that the cost-benefit ratio occasionally increases with higher marginal costs. This occurs because the underlying maximization problem aims to maximize welfare gains, not cost-effectiveness.

Table 2: Optimal tutoring intervention design and welfare gain for university of applied science students

	Without social returns		With social returns	
	Student-tutor	Lecturer-tutor	Student-tutor	Lecturer-tutor
Optimal amount of hours per week	4 [4,4]	4 [4,4]	4 [4,4]	4 [4,4]
Optimal class size	11 [10,12]	20 [19,22]	7 [6,8]	12 [10,13]
Net welfare gain per student per year	€3,327 [3127,3526]	€3,374 [3165,3586]	€7,038 [6191,7862]	€7,485 [6598,8431]
Benefit-cost ratio	15.4 [13.9;17.1]	9.2 [8.5;10.2]	19.0 [16.8;21.4]	11.4 [10.2;12.9]

Notes: Simulations using individual-specific coefficients from main mixed logit and mean coefficients of effectiveness analysis of tutoring (see Supplementary Table 3 and Table 1). Net welfare gains with social returns include students' willingness-to-pay, and the expected increase in government revenue and decrease in expenditure due to students' improved probability of graduating (in time) minus the program's costs. 95% bootstrapped confidence intervals between brackets using 500 iterations.

Figure 4: The cost-effectiveness of public investment in tutoring with a student tutor as a function of marginal costs and the social returns to tutoring



Notes: Y-axis represents the ratio of social returns (excluding private welfare) to total costs. Sensitivity to overestimation of returns to tutoring and underestimation of marginal costs considered relative to baseline estimates. Optimal intervention consists of an experienced lecturer for 4 hours per week, with the class size depending on the marginal costs and returns (see Supplementary Figure 1). The baseline estimate of its marginal cost per hour was €25.

4.3 Consumption of private tutoring and the role of information

Previous results showed that students have a sizeable WTP for tutoring, and that tutoring could lead to large welfare gains. This final section describes students' consumption of tutoring on the private market. Additionally, we assess whether information about the returns to educational attainment influences students' preferences for tutoring.

Students' consumption of private tutoring

When higher education institutions do not meet students' demand for tutoring, students can resort to private providers. Within our sample of UAS and RU students, 4.3% and 11.2%, respectively, reported engaging in or intending to consume private tutoring during the evaluated term for an average price of €28.5 per hour. This equates to a weighted average of 7.6% of first-year students in Flanders.

This uptake of private tutoring is lower than expected based on students' sizeable WTP for tutoring as outlined in Section 4.1, for which we consider four potential explanations. First, 83% of these students pay for one-on-one private tutoring. Private tutoring in group is thus relatively rare, perhaps due to the lack of scale of private providers. This causes the average hourly price of private tutoring to be significantly larger than our baseline cost estimates of publicly provided tutoring. A second explanation for the low consumption of private tutoring is that 70% of students stated they preferred tutoring from their own university over a private provider all else being equal (though only 23% had a strong preference). Third, two prominent behavioral biases could explain why students fail to translate their preferences for tutoring into purchasing privately provided tutoring. First, status-quo bias can lead to students failing to take action despite potential welfare gains (Damgaard and Nielsen, 2018). Second, consuming private tutoring implies admitting to need help and might be considered as an unfair advantage. Both of these aspects of tutoring might hurt students' self-image or social image, which are prominent behavioral biases (Damgaard and Nielsen, 2018). All these biases are particularly prominent among university students and would lead to a lower uptake of private tutoring than their preferences suggest (Damgaard and Nielsen, 2018). Additionally, these three potential explanations for students' low take-up would be partially or fully alleviated under university-provided tutoring. Finally, some hypothetical bias in our estimates of students' WTP might explain part of this discrepancy, despite the earlier findings that students' WTP are not unrealistically high. However, we show in Supplementary Materials Section 2.4 how the large potential of public tutoring to improve private and social welfare holds even under a significant degree of hypothetical bias.

Appendix Table A2 presents the results of a linear regression analysis examining the relationship between various background variables and students' decisions to pursue private tutoring, as well as the corresponding hourly prices they pay. The results align well with students' preferences as estimated by the choice experiment and with earlier literature on the prevalence of private tutoring in K-12 education (Byun et al., 2018; Dang and Rogers, 2008). Students with a lower expected GPA are more likely to pay for private tutoring, suggesting that tutoring is mainly used for remedial purposes. Students without a higher-educated father are less likely to pay for private tutoring. Their higher price sensitivity seems to outweigh their stronger preference for tutoring as measured by the choice experiment. Additionally, students receiving a need-based scholarship and students with a migration background follow private tutoring programs of a lower price and thus, perhaps, a lower quality. These results highlight how the unequal access to private tutoring by socioeconomic status.

The effect of information about the returns to educational attainment

One potential concern is that students underinvest in tutoring because they underestimate the returns to educational attainment. The information experiment lends some evidence to this. Not a single question in the quiz was correctly answered by most students (see Supplementary Table 1), though part of this could be explained by central-tendency bias. For example, more than half of the students in the information treatment underestimated the yearly labor market premium by at least a factor of three.

Table 3 reports whether the treatment that informed students about the returns to a degree, the average dropout rate, and the foregone labor income of an extra year of study increased students' preferences for tutoring.[4] We find that the treatment increased students' WTP for passing an extra course this term by about €1.4 per week for the pooled sample through a lower price sensitivity in the choice experiment. The effect is not statistically significant for UAS students, but it is for RU students and the pooled sample. Results are very similar when using a multinomial logit model (see Supplementary Table 10, or when restricting the treatment group to those who answered at the most one question of the treatment quiz correct.

We conclude that students' underestimate the returns to education, but that correcting this bias has only a small effect on students' WTP for having fewer retakes. Therefore, informational biases seem a relatively unimportant reason why students underinvest in tutoring.

5 Discussion and conclusion

Little is known about students' preferences regarding tutoring, nor how to structure tutoring programs in a cost-effective way. Additionally, while private tutoring has the potential to threaten social mobility, little is known about students' consumption of private tutoring. This paper estimates higher education students' preferences for tutoring and their perception of its effectiveness. We use these estimates as inputs to simulate the cost-effectiveness of different tutoring programs. Additionally, we measure students' consumption of private tutoring and evaluate the role of informational biases in this decision using an information experiment.

Most students have a sizeable willingness-to-pay for tutoring, as they perceive it to be highly effective at increasing their educational attainment. However, their willingness-to-pay for having an experienced lecturer and smaller class sizes is limited. Using these results as baseline inputs combined with an extensive sensitivity analysis, we find that tutoring could be a highly cost-effective intervention to improve private and social welfare. This result takes into account that about a quarter of students would not attend the optimal tutoring interventions even at no marginal cost. However, the optimal class size is always larger than what is considered in most tutoring RCTs, and more in line with what higher education institutions are already providing, for example in the form of TA-sessions. This is not surprising given that higher education institutions compete to optimally meet students' demand in a cost-effective way (Epple et al., 2006) and the literature on the effect of class size reductions on learning outcomes (Ake-Little et al., 2020; Leuven and Oosterbeek, 2018). These results suggest that policies expanding TA sessions or tutoring programs by senior peers have a large potential to increase students' educational outcomes cost-effectively.

We also find that academically struggling students are more likely to use tutoring, as they both display a stronger preference for tutoring in the discrete choice experiment and are more

[4]Preferences for retakes was specified to be normal rather than log-normal as in the previous estimation. In case of a log-normal distribution, the interaction effect allows to circumvent the log-normal restriction on the coefficient's sign. Indeed, if preferences for retakes is specified to be log-normal, using a dummy variable indicating receiving the treatment or a dummy variable indicating not receiving the treatment results in almost the same (negative) coefficient.

Table 3: Effect of information treatment on WTP for retakes using mixed logit

	(1) UAS	(2) RU	(3) Pooled
Mean effects			
Treat x Amount of retakes	-0.080	0.075	-0.067
	(0.092)	(0.224)	(0.084)
Treat x Cost per week	0.002	0.008**	0.003**
	(0.002)	(0.003)	(0.001)
Treat x Hours of tutoring per week	-0.004	-0.155	-0.025
	(0.059)	(0.136)	(0.054)
Hours of tutoring per week	0.472***	0.568***	0.488***
	(0.045)	(0.110)	(0.042)
Amount of retakes	-0.209***	-0.433**	-0.235***
	(0.065)	(0.172)	(0.060)
Cost per week	-0.023***	-0.027***	-0.024***
	(0.001)	(0.004)	(0.001)
Standard deviation			
Hours of tutoring per week	0.705***	0.736***	0.711***
	(0.040)	(0.108)	(0.038)
Amount of retakes	1.057***	1.074***	1.054***
	(0.064)	(0.162)	(0.059)
Cost per week	0.024***	0.011	0.023***
	(0.003)	(0.010)	(0.003)
Number of choices	19104	4608	23712
Number of respondents	1592	384	1976

Notes: The dependent variable is students' preferences in the choice experiment. Treatment effects are assumed to be non-random. Preferences for the amount of retakes was specified to be normal, while (the inverse of) preferences for retakes and costs per week were specified to be log-normal. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

likely to purchase tutoring on the private market. Publicly providing tutoring could also increase social mobility. Students without a higher-educated father have a stronger preference for tutoring but are unable to purchase tutoring on the private market because they are more sensitive to the high prices.

Despite the considerable potential of tutoring to improve private welfare, we find that only 7.6% of students currently purchase tutoring on the private market. We also find that we would expect students' willingness-to-pay for tutoring to be even larger than estimated, suggesting that hypothetical bias does not cause this discrepancy between stated preference and behavior. One explanation is that the current supply of private tutoring cannot meet this demand efficiently. Indeed, current private tutoring consists almost exclusively of expensive one-on-one tutoring. A lack of scale advantages among private providers might explain this inefficiency. Behavioral biases are a second likely explanation for students underinvesting in private tutoring. Purchasing tutoring outside higher education institutions requires students to overcome their status-quo bias and might negatively influence their self and social image (Damgaard and Nielsen, 2018). Policies providing public tutoring programs could not only be a cost-effective investment to improve social welfare and social mobility, but also help lower these behavioral barriers to students seeking academic support.

Finally, we find that a biased perception of the returns to education does not cause students to underinvest in tutoring. While students underestimate the returns to educational attainment, correcting this bias has only a small effect on students' willingness-to-pay. Like other educational decisions, behavioral nudges might have more potential to increase students' consumption of tutoring.

These results raise the question why higher education institutions are not already providing the optimal amount of tutoring. One can interpret these unrealized potential welfare gains as a welfare loss caused by a supply shortage of educational support (Dang and Rogers, 2008). Price controls on tuition fees, which exist in Flanders as in many other education systems (OECD, 2020), are a likely explanation for this market failure. Policies relaxing these tuition fee caps would allow higher education institutions to realize these potential welfare gains, while increasing social mobility by limiting the role of private providers of tutoring.

While these results point to a large potential for tutoring to improve both private and social welfare, they have their limitations. First, the results are based on a sample of only one higher education system. Most importantly, the calculation of social returns relies on students' perception of the effectiveness of tutoring, which could significantly differ from its actual effectiveness. Students might imagine some sort of 'ideal' tutoring settings which could significantly deviate from what universities are able to offer. Additionally, both marginal costs and the returns to education might be lower due to general equilibrium effects. In all these aspects, the robustness of the results to underestimation of marginal costs and overestimation of the effectiveness of tutoring is reassuring. Additionally, even if students' perception of the limited added value of smaller classes is wrong, their perception still matters. Competing higher education institutions are not only incentivized to maximize learning outcomes but also to meet students' demand (Epple et al., 2006).

These results and their limitations point to various directions for future research. More research is needed on the importance of class size in the effectiveness of tutoring. Most recent research, including in K-12 education, has focused on the potential of tutoring with class sizes smaller than six students (Dietrichson et al., 2017; Nickow et al., 2020). The literature on the effectiveness of tutoring in higher education is smaller but has similarly focused on interventions in small groups (Gordanier et al., 2019; Hardt et al., 2023). To the best of our knowledge, no previous study has assessed the differential effect of tutoring programs of different class sizes in one experimental setting. These results can help inform the design of such experiments.

References

Ake-Little, E., von der Embse, N., and Dawson, D. (2020). Does class size matter in the university setting? *Educational Researcher*, 49(8):595–605.

Blundell, R., Green, D. A., and Jin, W. (2022). The UK as a technological follower: Higher education expansion and the college wage premium. *The Review of Economic Studies*, 89(1):142–180.

Bovill, C., Bulley, C. J., and Morss, K. (2011). Engaging and empowering first-year students through curriculum design: perspectives from the literature. *Teaching in Higher Education*, 16(2):197–209.

Byun, S.-y., Chung, H. J., and Baker, D. P. (2018). Global patterns of the use of shadow education: Student, family, and national influences. *Research in the Sociology of Education*, 20:71–105.

Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19:7–42.

Chavez, D. E., Palma, M. A., Nayga Jr, R. M., and Mjelde, J. W. (2020). Product availability in discrete choice experiments with private goods. *Journal of choice modelling*, 36:100225.

Court of Audit (2022). Report of by the Court of Auditors about study progress in higher education [Verslag van het Rekenhof over studievoortgang in het hoger onderwijs]. Technical report, Court of Audit (Rekenhof).

Damgaard, M. T. and Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*, 64:313–342.

Dang, H.-A. and Rogers, F. H. (2008). The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? *The World Bank Research Observer*, 23(2):161–200.

Dietrichson, J., Bøg, M., Filges, T., and Klint Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, 87(2):243–282.

Epple, D., Romano, R., and Sieg, H. (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928.

Expertise Center on Budgets and Financial Well-being (2023). Studying in tertiary education: how much does that cost? [Studeren in het hoger onderwijs: hoeveel kost dat?]. Technical report, Thomas More.

Feld, J., Salamanca, N., and Zölitz, U. (2020). Are professors worth it? The value-added and costs of tutorial instructors. *Journal of Human Resources*, 55(3):836–863.

Gordanier, J., Hauk, W., and Sankaran, C. (2019). Early intervention in college classes and improved student outcomes. *Economics of Education Review*, 72:23–29.

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., Farkas, G., Fryer Jr, R. G., Mayer, S., Pollack, H., et al. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review*, 113(3):738–765.

Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H., and Lancsar, E. (2021). Hypothetical bias in stated choice experiments: Part ii. conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. *Journal of choice modelling*, 41:100322.

Hardt, D., Nagler, M., and Rincke, J. (2023). Tutoring in (online) higher education: Experimental evidence. *Economics of Education Review*, 92:102350.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata journal*, 7(3):388–401.

Horiuchi, Y., Markovich, Z., and Yamamoto, T. (2022). Does conjoint analysis mitigate social desirability bias? *Political Analysis*, 30(4):535–549.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Jobat (2022). Salary compass 2023 [Salariskompas 2022]. https://www.jobat.be/nl/art/hoeveel-levert-een-extra-diploma-op. Accessed: 2022-04-23.

Johnston, A. C. (2020). Teacher preferences, working conditions, and compensation structure. IZA Discussion Paper No. 13121, Forschungsinstitut zur Zukunft der Arbeit.

Kemptner, D. and Tolan, S. (2018). The role of time preferences in educational decision making. *Economics of Education Review*, 67:25–39.

Kraft, M. A. and Falken, G. T. (2021). A blueprint for scaling tutoring and mentoring across public schools. *AERA Open*, 7.

Leuven, E. and Oosterbeek, H. (2018). Class size and student outcomes in europe. Technical report, EENEE Analytical Report No. 33.

Moore, M. A., Boardman, A. E., Vining, A. R., Weimer, D. L., and Greenberg, D. H. (2004). "Just give me a number!" Practical values for the social discount rate. *Journal of Policy Analysis and Management*, 23(4):789–812.

Nickow, A., Oreopoulos, P., and Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. *National Bureau of Economic Research*. Working paper 27476. http://www.nber.org/papers/w27476.

OECD (2020). Resourcing higher education: Challenges, choices, and consequences. Technical report, OECD Publishing, Paris.

OECD (2022). Education at a glance 2022. Technical report, OECD Publishing, Paris.

Phelps, E. S. and Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2):185–199.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, 80(4):647–657.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Trapero-Bertran, M., Rodríguez-Martín, B., and López-Bastida, J. (2019). What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. *PloS one*, 14(7):e0219905.

Vass, C., Boeri, M., Karim, S., Marshall, D., Craig, B., Ho, K.-A., Mott, D., Ngorsuraches, S., Badawy, S. M., Mühlbacher, A., et al. (2022). Accounting for preference heterogeneity in discrete-choice experiments: an ispor special interest group report. *Value in Health*, 25(5):685–694.

Yoo, J. and Ready, R. (2014). Preference heterogeneity for renewable energy technology. *Energy Economics*, 42:101–114.

A Appendix

Table A1: Descriptive statistics of sample

	UAS (N=1592)		RU (N=384)	
	Share (in %)	Missing observations	Share (in %)	Missing observations
Background characteristics				
Male	36.4	81	57.1	27
Migration background	41.6	64	27.1	23
Higher educated father	40.4	64	69.3	23
Higher educated mother	49.8	64	82.5	23
Academic year				
In first year	51.3	84	74.8	23
In second year	31.6	84	11.6	23
In third year	17.0	84	13.6	23
Academic Domain				
Economics & Business	33.0	87	64.3	56
STEM	16.7	87	35.7	56
Healthcare	37.7	87	0.0	56
Others	12.7	87	0.0	56

Notes: UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. 'Migration background' means that at least one of their grandparents was born outside of Belgium. 'Higher educated' means at least ISCED-5 according to the ISCED-2011 classification. Amount of observations differs slightly due to data availability. Students were allowed to not report their gender. On their academic domain, they were allowed to opt for 'Other, being ...', which some students selected without filling in the subsequent open question. The share of men in the population of Flemish UAS students is 41.6%. The share of men in the population of Flemish RU students in Economics, Business and STEM is 61.2%. Missing observations are due to students dropping out after filling the the discrete choice experiment, but before having filled out all the question on the background characteristics.

Table A2: Regression analysis of who follows private tutoring and its hourly price as a function of personal background characteristics

	(1) Following private tutoring	(2) Hourly price of private tutoring (in €)
Male	-0.024**	-6.607*
	(0.012)	(3.333)
Highly educated father	0.025**	-1.698
	(0.011)	(3.198)
Needs-based scholarship	0.009	-7.562*
	(0.013)	(3.944)
Migration background	0.009	-8.299**
	(0.013)	(3.470)
In first year	0.012	5.362
	(0.010)	(3.797)
Economics program	0.039***	-0.356
	(0.014)	(4.050)
STEM program	0.001	-3.241
	(0.016)	(3.985)
Research university	0.037**	10.619**
	(0.018)	(4.968)
Expected amount of retakes (Std.)	0.008	-4.052*
	(0.006)	(2.058)
Expected GPA (Std.)	-0.033***	-2.723
	(0.008)	(2.174)
Constant	0.016*	30.744***
	(0.010)	(4.235)
Number of respondents	1813	89

Notes: The dependent variable in the first model is a dummy indicating whether a student stated paying or expecting to pay for privately provided tutoring this semester. The dependent variable in the second model is the reported hourly price. 'Migration background' means that at least one of their grandparents was born outside of Belgium. 'Higher educated' means at least ISCED-5 according to the ISCED-2011 classification. Amount of observations differs slightly due to data availability. Students were allowed to not report their gender. On their academic domain, they were allowed to opt for 'Other, being ...', which some students selected without filling in the subsequent open question. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

B Supplementary Materials and Information

B.1 Supplementary Tables & Figures

Supplementary Table 1: The information treatment for university students

Question	Possible answers	Frequency of response (in %)
'How many students that start in higher education fail to graduate?'	1 in 10	8.9
	1 in 8	25.7
	1 in 7	35.9
	1 in 5	29.6
'A higher education degree leads to a higher labour income. With an academic master, your yearly net wage is ... higher.	€1,000	13.8
	€2,500	42.3
	€5,000	32.3
	€7,500	11.6
'If you fail too many courses, you might need an extra year to finish your studies, which means an extra year in which you don't earn a full-time labour income. What would be the yearly net salary that you would miss out on on average?'	€5,500	13.1
	€11,000	17.8
	€22,000	30.8
	€33,000	38.4

Notes: The treatment for university of applied science students used the returns to education and foregone labour income of their degree. The answer in bold is correct, as found by Court of Audit (2022) using Flemish administrative data for the first question, and by Jobat (2022) based on a large survey of Belgian employees for the second and third question.

Supplementary Table 2: Overview of survey structure

Information experiment	Students were randomly assigned to either the treatment or control group. Treatment students received information about the returns to educational attainment, as described in Supplementary Table 1
Discrete choice experiment	Students ranked their preferences between hypothetical tutoring programs and the option not to follow any tutoring. See Figure 1 in the main manuscript for an example and Supplementary Table 11 for all possible attribute values.
Perceived effectiveness of tutoring	Students stated their expected GPA and amount of exams failed this semester without tutoring, as well as conditional on following some tutoring programs. See Figure 2 in the main manuscript for an example. Students also described how they believe barely passing or failing one or more additional exams would influence their probability of graduating (in time).
Consumption of private tutoring	Students stated whether they were following any private tutoring this semester or were planning on doing so. Students that stated doing so were then asked what price they paid and what group size they were taught in.
Background characteristics	Students reported their gender, parental education level, migration background, academic year and study domain. See Table A1 in the Appendix of the main manuscript for the results.

Supplementary Table 3: Main mixed logit analysis of discrete choice experiment: mean and standard deviation

	(1) UAS		(2) RU	
	Mean	SD	Mean	SD
Amount (in hours per week)	0.0461**	0.3252***	0.0467	0.4114***
	(0.0227)	(0.0345)	(0.0425)	0.0328***
Class size	-0.0272***	0.0271***	-0.0371***	(0.0045)
	(0.0019)	(0.0027)	(0.0034)	0.6801***
Experienced lecturer	0.6075***	0.9204***	0.5143***	0.6801***
	(0.0594)	(0.0818)	(0.0894)	(0.1487)
Price per hour	-0.0689***	0.0463***	-0.0681***	0.0364***
	(0.0030)	(0.0036)	(0.0046)	(0.0056)
Tutoring constant	2.0949***	3.5975***	1.6570***	3.3924***
	(0.1560)	(0.1730)	(0.2511)	(0.2790)
Number of choices	9.288		3.240	
Number of respondents	774		270	

Notes: In each analysis, the outcome variable is students' preferences for tutoring programs in the discrete choice experiment. All random coefficients are specified to be normally distributed. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Supplementary Table 4: Main mixed logit analysis of discrete choice experiment: estimated correlations between coefficients

	(1) UAS	(2) RU
Amount & tutoring constant	0.1278*	0.1509
	(0.0695)	(0.0946)
Amount & class size	0.1117	0.0210
	(0.1338)	(0.1697)
Amount & experienced lecturer	0.0620	-0.0707
	(0.1257)	(0.2048)
Amount & Price	0.0760	0.1194
	(0.1151)	(0.1943)
Tutoring constant & class size	-0.3449***	-0.5214***
	(0.0741)	(0.0961)
Tutoring constant & experienced lecturer	-0.0900	0.1196
	(0.0765)	(0.1662)
Tutoring constant & price	-0.1248*	-0.1384
	(0.0729)	(0.1477)
Class size & experienced lecturer	-0.1622	-0.0618
	(0.1107)	(0.2295)
Class size & price	-0.0711	-0.3735*
	(0.1108)	(0.1939)
Experienced lecturer & price	-0.0944	0.0883
	(0.1056)	(0.2292)
Number of choices	9288	3240
Number of respondents	774	270

Notes: All random coefficients are specified to be normally distributed. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Supplementary Table 5: Robustness checks for analysis of discrete choice experiment: with interaction effects and using multinomial logit

	(1) UAS	(2) UAS	(3) RU	(4) RU
Mean effects				
Amount (in hours per week)	0.0430**	0.0287**	0.0705	0.0133
	(0.0188)	(0.0129)	(0.0436)	(0.0233)
Class size	-0.0203***	-0.0118***	-0.0722***	-0.0180***
	(0.0040)	(0.0010)	(0.0077)	(0.0019)
Experienced lecturer	0.5746***	0.2355***	0.5753***	0.2247***
	(0.0812)	(0.0323)	(0.1372)	(0.0507)
Price per hour	-0.0654***	-0.0322***	-0.0728***	-0.0334***
	(0.0029)	(0.0013)	(0.0049)	(0.0023)
Tutoring constant	1.6798***	1.0773***	4.2558***	0.9531***
	(0.3317)	(0.0694)	(0.6182)	(0.1142)
Exp. lecturer x class size	-0.0008		-0.0020	
	(0.0036)		(0.0063)	
Class size squared	-0.0001		0.0006***	
	(0.0001)		(0.0001)	
Standard deviations				
Amount (in hours per week)	0.1071***		0.4293***	
	(0.0233)		(0.0571)	
Class size	0.0243***		0.0351***	
	(0.0025)		(0.0044)	
Experienced lecturer	0.8760***		0.6219***	
	(0.0739)		(0.1505)	
Price per hour	0.0461***		0.0386***	
	(0.0031)		(0.0052)	
Tutoring constant	3.4790***		3.5243***	
	(0.1689)		(0.2943)	
Number of choices	9288	9288	3240	3240
Number of respondents	774	774	270	270

Notes: The dependent variable in all four models is students' preferences in the choice experiment. All random coefficients are specified to be normally distributed. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. * p<0.10, ** p<0.05, *** p<0.01

Supplementary Table 6: Heterogeneity analysis of discrete choice experiment using interaction effects with background characteristic for university of applied sciences students including second and third year

Interacting background : characteristic	(1) ... Follows private tutoring	(2) ... Male	(3) ... Father tertiary educated	(4) ... First year	(5) ... GPA (Std.)
Mean effects					
Hours of tutoring	0.008	0.039*	0.057***	-0.034*	0.012
	(0.015)	(0.021)	(0.017)	(0.020)	(0.013)
Class size	-0.023***	-0.029***	-0.021***	-0.023***	-0.024***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Experienced lecturer	0.529***	0.573***	0.590***	0.533***	0.538***
	(0.036)	(0.051)	(0.050)	(0.056)	(0.038)
Price per hour	-0.062***	-0.068***	-0.067***	-0.063***	-0.064***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Tutoring constant	1.435***	2.135***	1.480***	1.155***	1.691***
	(0.102)	(0.138)	(0.142)	(0.155)	(0.107)
Interaction effects					
Hours of tutoring x ...	0.108	-0.090***	-0.122***	0.081***	-0.027**
	(0.073)	(0.035)	(0.027)	(0.027)	(0.013)
Class size x ...	-0.019***	0.008***	-0.006**	-0.004	-0.001
	(0.006)	(0.003)	(0.003)	(0.003)	(0.001)
Experienced lecturer x ...	0.002	0.032	-0.095	0.042	0.051
	(0.175)	(0.082)	(0.076)	(0.076)	(0.037)
Price per hour x ...	0.011	0.002	0.006*	-0.004	0.002
	(0.009)	(0.004)	(0.004)	(0.004)	(0.002)
Tutoring constant x ...	3.024***	-0.416*	-0.052	0.715***	-0.360***
	(0.437)	(0.215)	(0.212)	(0.207)	(0.113)
Standard deviation					
Hours of tutoring	0.253***	0.357***	0.090***	0.089***	0.095***
	(0.025)	(0.025)	(0.018)	(0.017)	(0.017)
Class size	0.027***	0.030***	0.029***	0.029***	0.029***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Experienced lecturer	0.709***	0.842***	0.781***	0.795***	0.776***
	(0.056)	(0.059)	(0.056)	(0.057)	(0.056)
Price per hour	0.041***	0.046***	0.044***	0.045***	0.044***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Tutoring constant	3.492***	3.680***	3.685***	3.578***	3.598***
	(0.108)	(0.125)	(0.127)	(0.124)	(0.123)
Number of choices	19104	19104	19104	19104	19104
Number of respondents	1592	1592	1592	1592	1592

Notes: The dependent variable in each model is students' preferences in the choice experiment. Each model includes the average preference for all attributes, which are specified to be normally distributed, as well as an interaction effect between one background characteristic and all the attributes. The row at the top of the table indicates with which background characteristic the attributes were interacted. For example, the coefficients 'Hours of tutoring x ...' in the column '... Follows private tutoring' indicates that the coefficient describing the preference for hours of tutoring is 0.090 lower for a student that follows private tutoring than a student that doesn't follow private tutoring. The first four background characteristics are dummy variables. GPA variable is standardized. * p<0.10, ** p<0.05, *** p<0.01

Supplementary Table 7: Regression analysis of students' expected change in probability to graduate (in time) as a function of the number of exams barely passed relative to having barely failed these exams

	(1) Change in the probability of graduating	(2) Change in the probability of graduating in time
Number of exams barely passed	2.578*** (0.365)	5.477*** (0.412)
Number of respondents	537	611

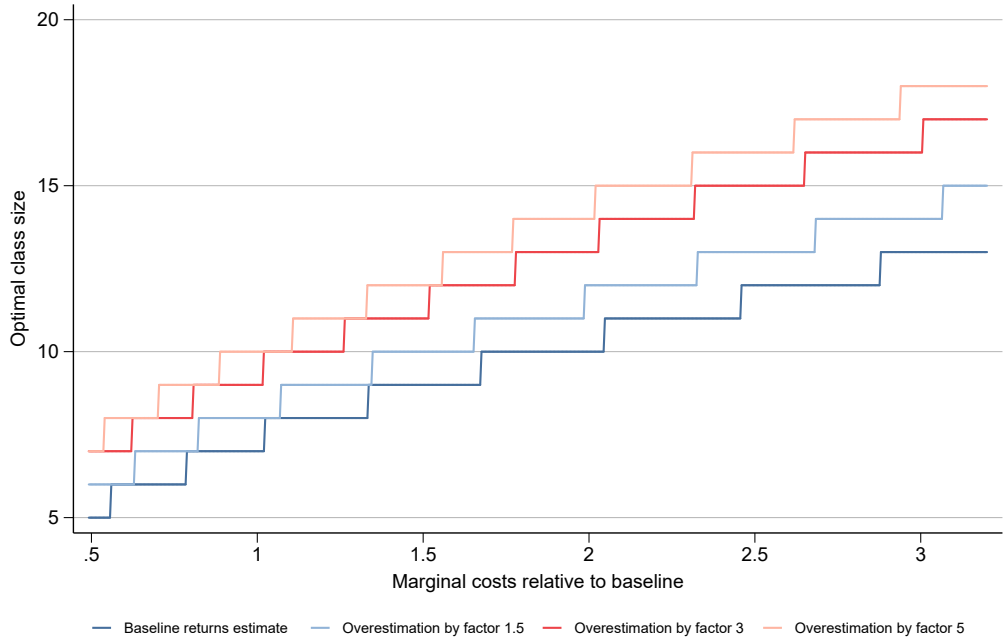
Notes: The set of questions used in this estimation is described in detail in Supplementary Section B.2.2. * p<0.10, ** p<0.05, *** p<0.01

Supplementary Table 8: Optimal tutoring intervention design and welfare gain for research university students

	Without social returns		With social returns	
	Student-tutor	Lecturer-tutor	Student-tutor	Lecturer-tutor
Optimal amount of hours per week	4 [4,4]	4 [4,4]	4 [4,4]	4 [4,4]
Optimal class size	9 [8,10]	16 [15,18]	6 [5,7]	11 [9,13]
Net welfare gain per student per year	€2,908 [2545,3243]	€2,855 [2493,3171]	€7,565 [6557,8771]	€7,802 [6708,9083]
Benefit-cost ratio	11.1 [9.6;12.9]	6.6 [5.7;7.5]	18.6 [15.4;22.3]	11.0 [9.1;13.4]

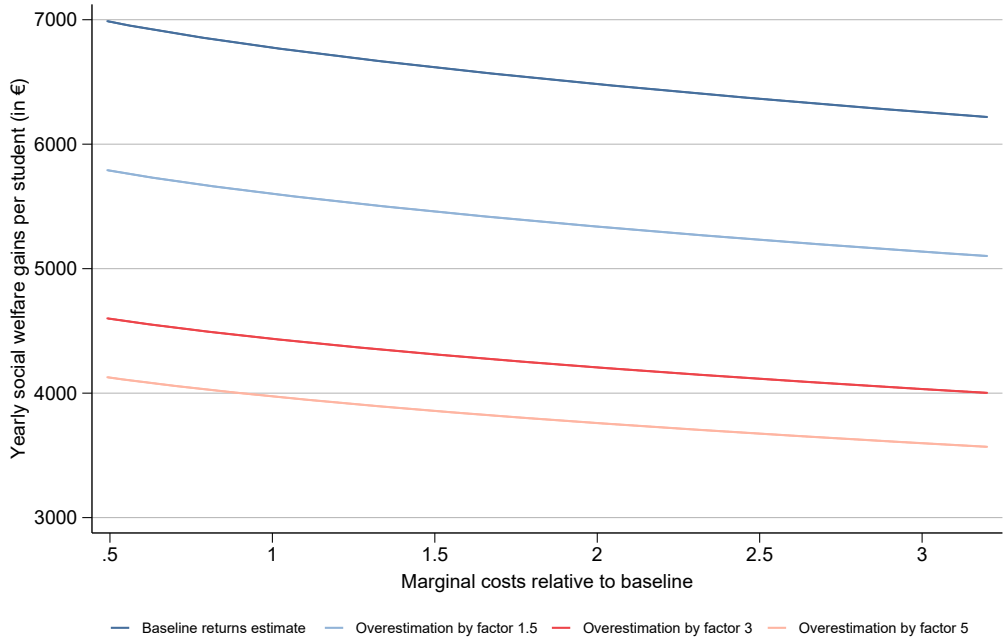
Notes: Simulations using coefficients from main mixed logit and efficacy analysis (see Supplementary Table 3 and see Table 1 in the main manuscript). Net welfare gains with social returns include students' willingness-to-pay, and the expected increase in government revenue and decrease in expenditure due to students' improved probability of graduating (in time) minus the program's costs. 95% boostrapped confidence intervals reported between brackets using 500 iterations.

Supplementary Figure 1: Class size maximizing social welfare with a student-tutor as a function of marginal costs and the social returns to tutoring



Notes: This graphs shows how the optimal class size from a social welfare perspective of a tutoring program by a senior student for four hours per week varies by the returns to tutoring and underestimated of marginal costs relative to the baseline estimates. The baseline estimate of this program's marginal cost was €25 per hour.

Supplementary Figure 2: Social welfare gains of optimal tutoring intervention by a senior student as a function of marginal costs and social returns



Notes: This graphs shows how the social welfare gains of a tutoring program by a senior student for four hours per week varies by the returns to tutoring and underestimated of marginal costs relative to the baseline estimates. The baseline estimate of this program's marginal cost was €25 per hour.

Supplementary Table 9: Optimal tutoring intervention design and welfare gain using interaction effects

	Without social returns		With social returns	
	Student-tutor	Lecturer-tutor	Student-tutor	Lecturer-tutor
UAS				
Optimal amount	4	4	4	4
of hours per week	[4,4]	[4,4]	[4,4]	[4,4]
Optimal	12	21	6	10
class size	[11,14]	[19,23]	[5,7]	[9,12]
Net welfare gain per	€3,030	€3,102	€6,773	€7,024
student per year	[2637,3450]	[2733,3450]	[5802,7647]	[6074,8022]
Benefit-cost ratio	15.2	8.8	15.4	9.6
	[13.7;17.0]	[7.9;9.7]	[13.4;17.7]	[8.5;10.9]
RU				
Optimal amount	4	4	4	4
of hours per week	[4,4]	[4,4]	[4,4]	[4,4]
Optimal	8	16	4	8
class size	[7,9]	[15,18]	[4,6]	[7,10]
Net welfare gain per	€4,433	€4,268	€9,395	€9,309
student per year	[3673,4849]	[3500,4723]	[8047,10747]	[7945,10807]
Benefit-cost ratio	15.0	9.1	16.5	10.0
	[12.9;17.3]	[7.8;10,6]	[13.6;21.0]	[8.3;12.3]

Notes: Simulations using coefficients from mixed logit with interaction effects(see Supplementary Table 5) and the efficacy analysis with interaction effects (available on request). 95% boostrapped confidence intervals between brackets using 500 iterations.

Supplementary Table 10: Effect of information treatment: robustness check using multionomial logit

	(1) UAS	(2) RU	(3) Pooled
Treat x Amount of retakes	-0.026	-0.014	-0.028
	(0.051)	(0.116)	(0.047)
Treat x Cost per week	0.000	0.004*	0.001
	(0.001)	(0.002)	(0.001)
Treat x Hours of tutoring per week	-0.004	-0.086	-0.016
	(0.039)	(0.089)	(0.035)
Hours of tutoring per week	0.203***	0.250***	0.210***
	(0.027)	(0.067)	(0.025)
Amount of retakes	-0.129***	-0.242***	-0.142***
	(0.037)	(0.088)	(0.034)
Cost per week	-0.009***	-0.013***	-0.010***
	(0.001)	(0.002)	(0.001)
Number of choices	19104	4608	23712
Number of respondents	1592	384	1976

Notes: The dependent variable in each model is students' preferences in the choice experiment. Preferences for the amount of retakes was specified to be normal, while (the inverse of) preferences for retakes and costs per week were specified to be log-normal. Interaction effects with treatment dummy are assumed to be non-random. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. * p<0.10, ** p<0.05, *** p<0.01

B.2 Supplementary information

B.2.1 Further notes on the design of the discrete choice experiment

Supplementary Table 11 provides the full list of attribute values used in the discrete choice experiment.

Supplementary Table 11: Attributes & attribute values in DCE

Attribute	Possible values
Amount of tutoring	1 hour every 2 weeks 1 hour every week 2 hour every week 3 hour every week 4 hour every week
Tutor	A senior student An experienced lecturer
Class size	Individual In a group of {2; 4; 8; 12; 16; 24; 32; 48; 64} students
Price	Between €0 and €50 per hour in increments of €5

In the description of the discrete choice experiments, students that were already paying for private tutoring were asked to consider a situation in which they were not yet doing so. Thus, we measure the demand for tutoring supplemental to the one provided by the public institution (and possibly free help from friends and family) but in the absence of privately provided tutoring. Students were also instructed to consider that all attributes not explicitly listed were identical between the two tutoring packages. This further reduces the potential for omitted variable bias (Johnston, 2020). We also asked students what they believe the first and second choice of their parents would be, as they might play an important role in students' willingness-to-pay for tutoring if they share costs. Results for parents' preferences were very similar and are therefore not further discussed in this paper. The full text can be found below:

> In this part of the survey, you are able to choose between two different tutoring programs and the option not to follow any tutoring in addition to the current offer of educational support from your university. The tutoring programs are equal on all aspects not mentioned, such as location. You are to describe which option would be your first choice and which would be your second.
>
> The table below shows two tutoring programs with four characteristics. The left column describes tutoring program A, the right column tutoring program B. The characteristics are the amount of tutoring per week, who would be the tutor, the size of the group in which you would follow the tutoring and how much the tutoring would cost per hour.

We also deliberately did not specify whether these tutoring programs were provided by their university or privately provided. Our rationale is that students generally have a preference for tutoring itself irrespective of the provider. Because students might have varying perceptions regarding the desirability of university-provided versus privately provided tutoring, we address this topic using in a separate question in Section 4.3. 'Students' consumption of private tutoring'. Not specifying this has the disadvantage that different students might have imagined

different settings (university or privately provided) of the different tutoring programs in the choice experiment, which then might have influenced their preferences. However, given that our primary objective is to highlight the potential of university-provided tutoring, and that students, on average, exhibit only a slight preference for university-provided tutoring, this approach would result in a slight underestimation of students' preferences and perceived effectiveness of tutoring. Both students' preferences are perceived effectiveness of tutoring were already considerable, and any underestimation would further reinforce the subsequent conclusions of the welfare analysis.

Using the four attributes and their corresponding attribute values of the choice experiment, we constructed three blocks of six hypothetical choice sets. Students were randomly assigned to one of the three blocks. Rather than randomly choosing attributes values to construct choice sets, we utilize a D-efficient design procedure. Such a procedure maximizes the statistical efficiency of the constructed choice sets based on some assumptions of the coefficients such as their sign (Johnston, 2020). We assume for example that students prefer a more experienced lecturer, smaller class sizes and lower prices *all other things equal*. To understand why statistical efficiency is relevant, consider the extreme example of a choice between two alternatives where one alternative clearly dominates the other (i.e. both have the same amount of tutoring, but one has a more experienced teacher, a smaller class size and is cheaper). The dominating alternative will always be chosen by attentive respondents and the choices would there provide no statistical information on how respondents trade off the different attributes.

Additionally, the experiment included an attention check. The attention check was an additional choice set with two identical tutoring packages except that one was more expensive. Students who preferred the more expensive tutoring package were considered inattentive and consequently excluded from the analysis of the discrete choice experiment, such as in Chavez et al. (2020). We also conducted a large scale pilot study with 279 respondents to ensure the survey and its experiments were clear and easy to follow. Given the limited changes, these students were also included in the full analysis to maximize power. Additional analyses show that the results without these pilot students are almost identical.

Finally, we purposefully did not include the course subject in the discrete choice experiment, as the available courses differ widely between study programs. It seems likely that students will prefer tutoring for the courses in which they perceive the lowest probability of passing, or for the courses in which they believe the marginal effect of tutoring relative to individual study time is highest.

B.2.2 Perceived effectiveness of tutoring and the relationship between exams failed and the probability of graduating (in time)

Before measuring students' perception of the effectiveness of tutoring, students first stated whether they would use tutoring to (i) decrease the amount of courses they expect to fail, (ii) to increase their GPA for courses for which they are already sure to pass, or (iii) for both outcomes. Students then stated the expected value of the outcome variable they just chose (amount of courses passed, their GPA or both) for 5 sets of tutoring packages. If students responded wanting both to increase the amount of courses for which they pass and improve their GPA, they were asked to rate their expected grades and their expected amount of retakes for 2 (possibly different) choice sets to reduce cognitive load.

A direct question on students' perception of the relationship between exams failed and the probability of graduating (in time) directly might be biased. Students' estimated probability of graduating (in time) will depend on the amount of courses they have to retake and their perception of their ability on which they are imperfectly informed. If students were told to imagine failing a certain number of courses, they might consider this a negative signal of their ability. A direct question on their perceived probability of graduating conditional on failing

a number of courses could not distinguish between a change in probability due to having to retake exams or due toa change in students' beliefs of their ability. We control for this bias by comparing students' perceived probabilities of graduating (in time) in two situations: one in which students pass a certain amount of courses by the smallest possible margin and one in which students fail that same amount of courses by the smallest possible margin. Barely passing or barely failing a course is generally considered to be more a matter of luck than skill, making this a hypothetical discontinuity design.

More concretely, we presented students with two scenarios that differed in the amount of courses they barely passed or failed. The baseline number of courses was determined by their initial estimate of their amount of failed courses to assure realism, while the number of courses barely passed or failed was randomly selected to be one, two or three. For example, imagine a student that reported expecting to fail about two courses this semester. This student was then asked to estimate her perceived probability of graduating if she failed two courses, out of which either one or two (randomly chosen) were failed by the smallest possible margin, which is a 9 out of 20. Then, this same student was asked to estimate her perceived probability of graduating if she passed this same number of courses, either one or two, by the smallest possible margin, which is a 10 out of 20. We then assume that any difference between the reported probabilities are due to passing or failing an extra exam, as such a small difference in grade is generally considered a matter of luck, and thus should not be interpreted as a signal of ability. More formally, assuming a student's probability of graduating $Pr_G(A, E)$ is a function of their ability A and the amount of exams they barely pass E, and assuming that this probability is a linear function of small changes in the amount of exams barely passed $E - E'$ leads us to estimate the following equation:

$$Pr_G(A, E) - Pr_G(A, E') = \beta(E - E') + \epsilon$$

This equation is then estimated using ordinary least squares regression with robust standard errors. The equation for the probability of graduating in time is identical. The results of these estimations can be found in Supplementary Table 7 and are described in Section 4.1.2 in the main manuscript. Note that we pool UAS and RU students here to increase the sample size, as the number of observations per students was significantly lower than in the estimation of the perceived effectiveness of tutoring. The coefficients did not significantly differ between UAS and RU students when analyzed separately.

B.2.3 Hypothetical bias in estimated WTP

Students might misrepresent their true preferences in the discrete choice experiment, a phenomenon known as hypothetical bias. Two important sources of hypothetical bias could cause concern in our setting. First, the experiment lacks incentive compatibility, which can lead to respondents disregarding their budget constraint (Haghani et al., 2021). This would only be a problem in our setting if the questions require significant cognitive effort, as truth-telling is the weakly dominant strategy in a low-stakes setting (Camerer and Hogarth, 1999). We believe the required cognitive effort is limited because (i) we limited the number of attributes on which the alternatives vary to 4, which is well below the average choice experiment (Trapero-Bertran et al., 2019), and (ii) the attention check limited potential inattentiveness among respondents. Nevertheless, this source of hypothetical bias can never be fully ruled out. Second, there might be some social desirability bias towards purchasing tutoring, as this might be a way for students to signal that they care about their educational outcomes. Recent evidence finds that discrete choice experiments are much less susceptible to social desirability bias than direct questions (Horiuchi et al., 2022).

This section first demonstrates the realism of our WTP estimates by comparing them to the returns to tutoring and to other private spending on tertiary education. Then, it shows how the main results of the welfare analysis holds even under a significant degree of hypothetical bias.

Comparing students' preferences to the returns of tutoring and to total private spending on tertiary education

To test whether students might have overstated their preferences for tutoring in the discrete choice experiment due to hypothetical bias, we first compare students' preferences to what we would expect the preferences of a risk-neutral rational agent to be, conditional on students' own perception of the effectiveness of tutoring.

One of the most important benefits of tutoring is that it can decrease the amount of courses a student has to retake. This leads to a higher probability of graduating, a higher probability of graduating in time and less time spent on studying for retake exams. We therefore compare students' WTP for having fewer retakes with the expected economic returns of these benefits due to having fewer retakes.[5]

To estimate students' valuation of decreasing the number of retakes, we estimate students' preferences for the tutoring packages in the discrete choice experiment as a function of their individual-specific perception of the effectiveness of those tutoring packages at reducing their expected number of retakes. By adding the tutoring packages' cost per week and amount of tutoring, we control for the main costs of tutoring, namely its direct cost and its opportunity cost. We thus can estimate students' direct WTP for reducing the number of courses they have to retake. Supplementary Table 12 describes the results of this analysis using a mixed logit model. We find that both UAS and RU students are on average willing to pay about €40 per week or €522 per semester to have one exam less to retake, controlling for the amount of time spent in tutoring. We can then compare students' valuation of passing an extra exam with its associated economic value using students' own perception of the relationship between passing an extra exam and a selection of beneficial outcomes: a higher probability to graduate (in time) and less time spent studying for retakes in the summer.

To estimate the private economic returns to graduating (in time), we use the lifetime labor income returns to a bachelor's degree in Flanders and the forgone labor income due to spending an additional year in education again from EU-SILC's individual-level data combined with conservative assumptions on changes in costs and changes in utility. The foregone labor income due to taking an extra year to graduate for an individual with a bachelor's degree in Belgium is on average €23,800. The net benefit for a student to graduate as soon as possible might be lessened by (i) their parents no longer covering their living expenses (though the average age in Belgium to graduate from tertiary education is 23 while the average age to move out of the parental home is 26.2) and (ii) incurring a disutility of no longer living in a student dorm (which about 40% of students do). Assuming that parents charge the full cost of living expenses to the students, which amounts to about €768 per month, as measured by the Flemish government's reference budgets, the net benefit of graduating a year earlier is reduced to €14,584 per year. For those students living in a dorm to continue to live together with friends would reduce their net benefit per year to €10,816 per year, again based on Flemish government's reference budgets. The average opportunity cost, weighted by the number of students that live in a dorm, amounts to €13,077 foregone labor income, conservatively controlling for possible changes in costs and

[5]Because we consider only a subset of the benefits, we underestimate the WTP for less retakes through tutoring by the rational agent, which strengthens our final conclusion. Though, this is somewhat lessened due to the fact that the possible other benefits, such as less time spent studying, higher grades and higher human capital, are correlated with the decrease in retakes.

utility.

Additionally, having one less exam to retake implies that students do not have to study for this retake exam in the summer. We measure in an additional question the opportunity cost of a retake in terms of extra days spent in summer studying for the retake exam, which we then conservatively value at the Belgian minimum hourly wage of a 19-year-old student in Belgium at the time of the survey.

Supplementary Table 12: Measuring aversion towards retakes

	(1) UAS	(2) RU
Mean effect		
Hours of tutoring per week	0.446***	0.451***
	(0.031)	(0.070)
Amount of retakes	-0.923***	-0.950***
	(0.175)	(0.360)
Cost per week	-0.023***	-0.023***
	(0.001)	(0.003)
Standard deviation		
Hours of tutoring per week	0.657***	0.650***
	(0.041)	(0.090)
Amount of retakes	2.264***	2.032***
	(0.147)	(0.326)
Cost per week	0.028***	0.013***
	(0.003)	(0.004)
Number of choices	5442	2104
Number of respondents	717	256

Notes: The dependent variable in each model is students' preferences in the choice experiment. The amount of retakes is derived from questions in which students rate the expected number of retakes conditional on following different tutoring programs from the choice experiment, as shown in Figure 2 in the main manuscript. All random coefficients are specified to be log-normally distributed. UAS refers to the subsample of university of applied science students, while RU refers to the subsample of research university students. * p<0.10, ** p<0.05, *** p<0.01

Supplementary Table 13 combines the calculations from above to compare students' valuation of decreasing their number of retakes through tutoring with its actual returns. We first use discounting factors from Kemptner and Tolan (2018), who estimate German students' time preferences using a structural model and find that a quasi-hyperbolic discounting model of Phelps and Pollak (1968) with an exponential discounting factor of 0.97 and a large present bias of 0.57 best explains educational choices. Using these discounting factors, we find that the weighted sum of the three benefits is three times larger than students' WTP. Note that this underestimation is even somewhat conservative as we control for the main costs of tutoring but consider only a subset of the returns to tutoring. Next, we endogenize students' present bias factor to rationalize preferences, while still using the same exponential discounting factor. An extreme present bias of 0.17 is needed to equalize these three returns to students' valuation of one less retake. Students' WTP for tutoring thus seems to be significantly *lower* than rational, implying that the WTP estimates are not unrealistically high.

Supplementary Table 13: Returns to having one less retake using students' perceived relationship between retakes and graduating (in time)

Benefits of passing an exam	Increase in probability per retake	Total value	Present discounted value	Rationalizing preferences with present bias
Higher probability of lifetime wage premium	2.6 pp	€120,675	€35,729	€10,845
Higher probability of graduating in time	5.5 pp	€13,077	€7,071	€2,146
Less studying time for retake in summer	100 pp	€705	€391	€122
Sum of benefits weighted by probabilities		€4,561	€1,709	€522

Notes: The increase in probability per retake is based on students' own perception on the relationship between passing an extra exam and the probability of graduating (in time). Total value column estimated using EU-SILC data for the higher probability of graduating (in time) and the minimum hourly wage for a student job (€10.45) as a conservative estimate of the opportunity cost of expected increased studying time for having to retake the exam in the summer for an average of 9 days as reported by students. The present discounted value uses the exponential discounting function from Kemptner and Tolan (2018) with an exponential discounting factor of 0.97 and a large present bias of 0.57. In the final column, the present bias needed to rationalize students choice while using Kemptner and Tolan (2018)'s exponential discounting factor is 0.17.

A second way of assessing the realism of students' WTP for tutoring is to compare it to the total private spending on higher education. An average household in Flanders spends about €6,977 per student per year on tertiary education (Expertise Center on Budgets and Financial Well-being, 2023). This figure includes costs related to tuition fees and student dormitories, excludes spending on food and clothing, and weighs these costs by the share of students with a scholarship and the share of students that commute. A research university student is estimated to have a WTP to pay of €23.2 per hour for the socially optimal intervention with a student tutor. Considering the average amount of hours per week from the choice experiment's programs (2.1) and 13 weeks per academic semester, this amounts to a total WTP of €1,267 per academic year. Assuming students would have to pay their willingness-to-pay, this tutoring program would represent a 18.1% increase in private spending on tertiary education. This is sizeable, though students perceive the effectiveness of tutoring to be large. Additionally, Flemish baseline private spending on tertiary education is comparatively low at 12.9% of total costs compared to the OECD average of 30.4%, suggesting substantial room for increased private spending.

In sum, these results instill confidence that any hypothetical bias did not lead to unrealistically high WTP estimates for tutoring. Students' WTP is actually significantly lower than what we would expect the WTP of a rational agent to be. While these calculations are based on the actual economic returns to graduating (in time), the information experiment shows that even students fully informed about these returns have only a slightly higher WTP. Additionally, these WTP estimates would represent a sizeable but realistic increase in private spending on tertiary education if realized.

The robustness of the welfare analysis to hypothetical bias

The subsection above showed that our estimates of students' WTP are not unrealistically high. Nevertheless, it cannot rule out the existence of some hypothetical bias. This subsection shows how the main conclusions of the welfare analysis are robust to a significant degree of upwardly bias in our estimates of students' WTP for tutoring. More concretely, an upward bias in our estimates of students' WTP could affect two results of the welfare analysis: the estimated private welfare gains and students' attendance of the tutoring programs.

On the first result, note that the estimated private welfare gains are large enough to be robust to a significant degree of bias. Using students' WTP, the estimated private welfare gains are 15.4 and 9.2 times larger than the estimated costs of the optimal tutoring program without social returns with a student-tutor and lecturer-tutor respectively. In other words, even if students' WTP for tutoring would be up to three times smaller than estimated, the private welfare gains would still be at least 5.13 and 3 times larger than the programs' costs. If the program's costs were shared equally between private and public spending, the private welfare gains of the privately optimal tutoring programs would exceed private costs even if students' WTP was 30 or 18 times smaller with a student or lecturer as tutor respectively. Perhaps more importantly, the previous subsection showed that the discounted private returns to tutoring are at least three times larger than the estimated students' WTP for tutoring, without considering for example positive effects on employment probability or marriage market advantages. Therefore, students' WTP is a rather conservative estimate of the total private welfare gains of a tutoring programs.

In the estimation of the socially optimal tutoring intervention, we use the distribution of students' WTP for the tutoring program to estimate the share of students that would attend these programs when provided by the university as part of their educational services. In this case, the marginal costs for attending a tutoring session would be zero since the price has already been paid at enrolment, either or both through tuition fees or public subsidies. Students thus attend the program if their WTP is larger than zero. In this case, any finite degree of proportional upwardly bias (i.e. by a multiplicative factor of x) would not influence the estimated attendance, as the sign of the WTP would not change. In other words, the estimated attendance of the tutoring programs would not be biased if the source of bias is that students in the choice experiment disregard their budget constraint, as this would only influence the size of the WTP but not its sign. Therefore, the main source of hypothetical bias in students' WTP is not relevant for estimating students' attendance. Attendance would only be overestimated if some students that state wanting to pay a positive amount for tutoring in reality would prefer not to attend tutoring at no direct financial cost, for example due to social desirability bias. In this case, the main results, i.e. that investments in tutoring provide large private and social welfare gains, would still hold but for a smaller subset of the population, and universities should then organize their offer of tutoring accordingly. Given the small estimated optimal social class sizes of between 7 and 12 and the relative large size of most academic programs, attendance in most programs is likely to be large enough to organize at least one such group. Only the amount of groups would differ with attendance.

In sum, the main results of the welfare analysis, i.e. the large private and social welfare gains to investments in tutoring, would hold even under a large degree of hypothetical bias. The ratio of private welfare gains to costs is large enough to allow for a significant degree of bias while remaining cost-effective, while we argue that the true private returns to tutoring are likely to be even higher. Additionally, students disregarding their budget constraint would not influence our estimate of attendance. Finally, the estimated large social welfare and high cost-benefit ratio could still be obtained with lower attendance for a smaller subset op the population.

B.2.4 Estimating the optimal tutoring intervention

This subsection describes three aspects of estimating the optimal tutoring intervention of Section 4.2. in more detail: (i) the back-of-the-envelope calculations of the marginal costs, (ii) the social returns to tutoring (iii) and how we calculate the individual-specific amount of tutoring followed in calculating the social returns.

For the estimate of the marginal costs of tutoring by an experienced lecturer, we use the official gross hourly wage of a lecturer with ten years of experience, which is €37.8. Providing an hour of education also requires preparation and planning, but as of yet no systematic analysis has been conducted on time spent on different teaching tasks in higher education. In secondary education, teachers in OECD countries spend on average 0.325 hours on lesson preparation and planning per hour of teaching. For teachers in higher education levels this is higher (OECD, 2022), so we assume this to be 1 hour. Additionally, providing tutoring might cost the higher education institutions more furniture, heating, administration, maintenance, ... and require an expansion in its material assets in the form of buildings. We assume that the increase in these costs, as reported in the annual accounts of the UAS that provides the majority of our sample, is proportional to the increase in the amount of education they provide and is written off over 25 years. The final marginal cost, €79.56, which we round up to €80, is close to but larger than what we usually observe as the price for a tutor with teaching experience in the private market (being around €50), which instills confidence in this estimate. We have no official hourly wage on which to base the estimate of the marginal costs of tutoring by a senior student. We, therefore, use the average hourly price of a senior student for tutoring on the private market, which is about €25 per hour.

We utilize OECD data on the Flemish government's annual spending per student in higher education (€16,158) to estimate the returns linked to the increased probability of graduating in time. For the returns to the increased probability of graduating, we rely on EU-SILC survey data to estimate the higher tax revenues resulting from the lifetime labor income returns to a bachelor's degree (€95,266). To estimate the lifetime labor income returns, we estimate yearly income as a function of education, work experience, work experience squared, and an interaction effect between work experience and education. We discount both forms of government returns using a social discount factor of 0.035 as recommended by Moore et al. (2004). Note that we are also likely to underestimate the returns to tutoring, as increasing students' probability to graduate is associated with other returns such as more innovation and a lower probability of unemployment.

Next, in maximizing social welfare, we need to take into account that students' attendance to these tutoring classes is based on their private welfare maximization problem (assuming attendance is not mandatory). In calculating the decrease in retakes, we therefore use the number of hours that a student would attend (i.e. for which their WTP is positive) using individual-level coefficients, as estimated by the mean of the coefficient distribution conditional on the individual's choice pattern (Train, 2009). We recode individual negative WTP-values to zero, as students can choose to not attend these supplementary tutoring classes and therefore don't suffer a welfare loss. Additionally, some WTP estimates on the right-hand side of the distribution seem unrealistically high (the 99th percentile of the WTP for an hour of individual tutoring by an experienced lecturer is €555 per hour) and, more importantly, are far beyond the bounds of the price attribute values in our DCE (being €50 per hour). Indeed, mixed logit models are known to produce unrealistic tail estimates for log-normally specified coefficients (Revelt and Train, 1998). We, therefore, limit WTP values to being 50% above our bound being €75 per hour. Together with the recoding of negative values to zero, these changes ensures that the average WTP of these estimates is equal to the average of the raw WTP estimates. Thus, the overall welfare loss stays the same.

Finally, we solve the social welfare maximization problem as follows. Equation B.2.4 describes the problem as described in Section 4.2..

$$\max_{A,T,C} W_s = W_p(A,T,C) + P_{GT}\Big(E\big(\tilde{A}(A,T,C),T,C\big)\Big) * R_{GT} + P_G\Big(E\big(\tilde{A}(A,T,C),T,C\big)\Big) * R_G$$

Denote the probability of graduating P_G as a function of the probability of graduating without tutoring P_G^0, and linearly depending on the amount of exams passed this term. Consider a similar functional form for P_{GT}. While this functional form is likely to be non-linear, linearity is a reasonable approximation given that we consider only small changes in the argument of the function. Assume a similar functional form for E, with E^0 the amount of exams passed without tutoring, and E^T the number of extra exams passed as a function of the tutoring program.

$$\begin{aligned}\max_{A,T,C} W_s = W_p(A,T,C) + \Big(P_{GT}^0 + \gamma_{GT}\big(E^0 + E^T(\tilde{A}(A,T,C),T,C)\big)\Big) * R_{GT} \\ + \Big(P_G^0 + \gamma_G\big(E^0 + E^T(\tilde{A}(A,T,C),T,C)\big)\Big) * R_G\end{aligned}$$

The private welfare can be straightforwardly derived from students' willingness-to-pay from the discrete choice experiment (see Supplementary Table 3) and the baseline estimates of the marginal costs. P_{GT}^0, P_G^0, and E^0 are not relevant for the maximization problem as they are constants that do not depend on the arguments of the problem. One can interpret this as maximizing the welfare gains from the tutoring program rather than the total welfare, two equivalent problems in this case. γ_G and γ_{GT} are derived from Supplementary Table 7. $E^T(\tilde{A}(A,T,C),T,C)$ is derived from Table 1 in the main manuscript. Finally, $\tilde{A}(A,T,C)$ is constructed as follows, with A^* being the maximal amount of tutoring offered:

$$\tilde{A}(A^*,T,C) = \arg\max_A W_p(A,T,C) \text{ s.t. } A \leq A^*$$

We solve the problem numerically. A brute-force search was conducted over the set of discrete values which is bounded by the attribute values of the choice experiment, i.e. class sizes between 1 and 64 and amount of tutoring between 0.5 and 4 hours per week, to identify the combination that maximizes the objective function. Bootstrapping this procedure including the parameter estimates for 500 iterations provides the confidence intervals.

Copyright © 2024 @ the author(s). Discussion papers are in draft form. This discussion paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.