

Pre-Analysis Plan

Perceptions of Consultants about Management Consultancy Across Sectors

Alessandra Fenizia, Renata Lemos

April 14, 2026

1 Introduction

This study aims to explore management consultants' perceptions of the effectiveness of management consultancy across different types of organizations. While there is a substantial body of research on the impact of consultancy interventions—particularly in the private sector—there is limited evidence on what consultants themselves view as important and impactful. To address this gap, we plan to conduct a survey of management consultants in Greece. Specifically, we seek to understand which organizational settings consultants consider most conducive to improvements in organizational outcomes, and whether there are perceived barriers to implementing such interventions in public sector institutions. Gaining insight into which settings are viewed as more appropriate for management consultancy services can also help contextualize and rationalize findings from our earlier study, shedding light on why this approach appears more effective in some environments than in others.

2 Research Design

We have partnered with the Hellenic Association of Management Consulting Firms (SESMA) to conduct a survey experiment with approximately 200 management consultants in Greece. The survey will be anonymous, self-administered online via Qualtrics, and participants will be recruited by SESMA. It will be conducted in English and is expected to take approximately 30 minutes to complete. At the end of the survey, respondents will have the option to participate

in a lottery for a gift card for an online retailer. The probability of winning increases with their accuracy score on the belief-elicitation tasks, ensuring that more accurate respondents are more likely to be rewarded. To participate, respondents will be asked to provide their email address.

The survey will begin with a set of questions collecting demographic information and details about the respondents' professional experience as management consultants. This will be followed by a discrete choice experiment (DCE) designed to elicit consultants' judgments about which organizational characteristics they associate with greater improvement potential from management consultancy. The survey will also capture information about barriers to management consultancy in the public sector. Lastly, the survey will include a belief-updating exercise.

2.1 Demographics and Professional Experience

The survey will begin with demographic and industry-specific questions, including tenure, expertise in management consultancy, and past experience with management consultancy in the public sector.

2.2 Discrete Choice Experiment

Respondents will complete a discrete choice experiment (DCE) designed to elicit their beliefs about which management practices at baseline are most conducive to improvement through management consultancy. In this DCE, respondents are presented with 8 pairs of hypothetical organizations. The attributes describing these organizations include: frequency of performance reviews, time horizon of targets, how the organization responds to problems, type of performance bonus for managers, primary method of promotion for non-managers, time required to reassign or dismiss underperforming non-managers, and organization budget per employee. The attribute values are based on, and align with, frameworks from the World Bank's Enterprise Survey (WBES) and the Management and Organizational Practices Survey (MOPS) conducted by the US Census Bureau. For each respondent, the order of attributes will be randomized.

Task Description. For each of the 8 choice tasks, respondents are shown the following prompt: "We will now show you 8 pairs of organizations along with their current management practices. For each pair, please select the organization you believe would experience the greater

improvement in performance if both were to implement the recommendations offered by management consultants. The profiles are hypothetical, but based on real-life organizations and practices. All other characteristics not shown are the same.”

This framing makes clear that the task is a counterfactual judgment, that is, respondents assess the relative impact of an identical consultancy intervention across two organizations that differ only in their baseline management practices.

Omission of Sector Labels. The hypothetical organizations presented in the DCE are deliberately not identified as public or private sector organizations. This serves two related purposes. First, it avoids priming respondents: a sector label would risk activating stereotypes unrelated to the specific management practices shown, contaminating the preferences we seek to recover. Second, it maximizes generalizability by ensuring that estimated attribute weights reflect consultants’ judgments about management practices as such, rather than about a particular organizational characteristic.

Choice Set Construction. To select the choice sets, we employ a D-efficient design. The choice sets were generated in R using the `cbcTools` package, applying the coordinate exchange algorithm (`method = "cea"`) with 4 blocks, 2 alternatives per question, 8 questions per block, and a maximum of 1,000 iterations.¹ The D-efficient design produced 32 choice sets in total, divided into 4 blocks of 8 choice sets each. Each respondent is randomly assigned to one block and completes all 8 choice tasks in that block. Random assignment to blocks ensures that the full set of 32 choice sets is covered across the sample and that no systematic differences exist between respondents who see different blocks.

Task Format and the Absence of an Opt-Out Option. Unlike DCEs designed to elicit consumer demand, the task here is inherently evaluative. As reflected in the task wording above, respondents are asked to select one of the two organizations. They are not deciding whether to accept a consultancy mandate, but forming a professional judgment about relative improvement potential. An opt-out option would therefore be conceptually inconsistent with the research design. This forced-choice format mirrors the comparative assessments consultants routinely make in practice and is consistent with established use of DCEs in evaluative and expert-judgment contexts.

¹To ensure reproducibility, the full code used to generate the design is available from the authors upon request. The D-efficient approach minimizes the determinant of the variance-covariance matrix of the parameter estimates, resulting in more statistically efficient data relative to a purely random or full factorial design. The algorithm ensures that attribute levels vary independently across choice sets, appear a balanced number of times, and avoid dominant or near-identical options within a pair — conditions that would provide little identifying information. The design also maintains sufficient variation in implied utilities across alternatives, which is important for estimating preference parameters precisely.

Hypothesis. For the six management practice attributes, we do not impose directional priors. Two competing mechanisms are plausible: consultants may favor organizations with stronger baseline practices, believing that better managerial practices are necessary to implement recommendations effectively; or they may favor organizations with weaker practices, believing that greater scope for improvement drives larger gains. The DCE is designed to empirically identify which mechanism dominates, and whether it does so uniformly across attributes or differently depending on the specific practice in question.

Table 1: DCE Attributes and Levels

Attribute	Levels
How often performance is tracked	Once a year Every month
How the target horizon is set	Short-term Short- and long-term
How problems are handled	Fix the problem only Fix the problem and prevent recurrence
Whether manager performance bonuses are given	No bonuses Some bonuses
How staff are promoted	Merit-based Tenure/connections
When poor performers are reassigned	Within 3-6 months After 2 years
Manager's budget per employee	30% below average 20% below average 10% below average average 10% above average 20% above average 30% above average

2.3 Barriers To Management Consultancy in The Public Sector

Respondents will be asked a few questions eliciting perceived barriers to the use of management consultancy in public sector organizations.

2.4 Belief-Updating

The belief-updating exercise will begin by eliciting respondents' beliefs about the effectiveness of management consultancy interventions in public sector organizations, specifically within Vocational Training Institutions (VTIs) in Greece. Respondents will be presented with a brief but detailed case study describing a management consultancy intervention implemented in VTIs, including an overview of its objectives and methodology. Importantly, at this stage, respondents will not be informed that the case study is based on a real intervention conducted by Fenizia et al. (2026). They will then be asked to estimate the intervention's take-up rate and its intention-to-treat (ITT) effects on two key performance indicators: dropout rates and apprenticeship assignment. The belief elicitation will follow the methodology of Iacovone et al. (2025), which allows researchers to recover the full prior distribution of beliefs.

For half of the sample, this elicitation exercise will be followed by disclosure of the actual take-up and ITT effects of the intervention (henceforth referred to as the "treatment"). Only at this point will these respondents be informed that the case study was based on a real study. Because participants are recruited by SESMA on a rolling basis, assignment to treatment will be conducted upon entry using Qualtrics' randomization feature, with a 50% probability of being assigned to the treatment group.

Immediately following the disclosure of the VTI intervention results, respondents in the treatment group will be asked to recall a specific result from the study: the ITT effect on apprenticeship assignment. Responses will be entered as numeric values. We will use the response to this question to evaluate respondents' attention.

Respondents assigned to the control group will proceed directly to the second case study without receiving any information about the actual results of the VTI intervention.

All respondents will then be presented with a second brief but detailed case study, describing a management consultancy intervention implemented in emergency rooms of Greek hospitals. This case study will closely mirror the structure and content of the first one, enabling a meaningful comparison across settings. Respondents will be asked to predict the take-up and treatment effects of the intervention on patient wait times in this new but analogous context. This second elicitation will allow us to assess whether, and how, consultants update their beliefs about the effectiveness of management consultancy interventions in other public sector organizations. Elicitation will follow the same approach as in the previous exercise, allowing us to retrieve the full distribution of posterior beliefs.

Following the second belief elicitation, all respondents will be asked a paired question about

their overall perception of the effectiveness of management consultancy in the public and private sectors. Specifically, respondents will be asked: "What percentage of management consultancy engagements do you believe result in meaningful and sustained performance improvements, in (a) private sector organizations and (b) public sector organizations?" Both items will be presented simultaneously on the same screen to avoid sequential anchoring and to make the public-private comparison explicit. Respondents will report a single point estimate for each. This paired question is designed to test whether exposure to the VTI evidence generalizes beyond the specific management intervention structure and hospital context, and, in particular, whether it narrows consultants' perceived gap between public- and private-sector consultancy effectiveness.

3 Main Outcomes

The primary outcomes of interest are:

1. The organizational characteristics that management consultants associate with greater improvement potential from management consultancy, as elicited through the discrete choice experiment.
2. The prior beliefs about the take-up and effectiveness of management consultancy interventions in public VTIs, as elicited before treatment assignment. These are descriptive outcomes characterizing the baseline distribution of consultant beliefs about public sector consultancy effectiveness, including the mean, dispersion, and shape of beliefs across the sample, independently of any treatment effect analysis.
3. The posterior beliefs about the take-up and effectiveness of management consultancy interventions in public sector organizations, specifically in the context of emergency rooms. The effect of treatment on the mean and median posterior is the primary estimand of interest, estimated via equation (2). The effect of treatment on the posterior variance is a secondary estimand, also estimated via equation (2) with the posterior variance as the dependent variable. The shape of the posterior distribution, including potential changes in bimodality, will be examined descriptively using Hartigan's dip test applied to the pooled distribution of posterior beliefs in each group separately.
4. The learning rates defined as the difference between the respondent's posterior and prior (Haaland et al., 2023).

Secondary outcomes are:

1. The perceived barriers to management consultants in the public sector
2. The perceived public-private effectiveness gap, defined as the difference between the respondent’s general belief about public sector consultancy effectiveness and their general belief about private sector consultancy effectiveness (as elicited by the paired post-exercise question). The effect of the treatment on the perceived public-private effectiveness gap. This is a secondary outcome measuring whether evidence disclosure specifically narrows consultants’ perception that the public sector is categorically less amenable to consultancy interventions, as distinct from any general shift in optimism about consultancy effectiveness overall. The public sector response and private sector response will also be reported separately.

4 Estimation Strategy

4.1 Discrete Choice Experiment

For the DCE, we estimate two models. As a baseline, we first estimate a simple conditional logit, which assumes homogeneous preferences across respondents and provides a tractable benchmark for comparison. Our preferred specification is a mixed logit regression, which relaxes the assumption of homogeneous preferences and allows coefficients to vary across respondents:

$$U_{ij} = \alpha + \beta_i W_{ij} + \varepsilon_{ij} \tag{1}$$

where U_{ij} is the latent utility respondent i derives from choosing organization j (rather than the alternative), and W_{ij} is a vector of attributes of organization j defined in the DCE (frequency of performance reviews, time horizon of targets, how the organization reacts to problems, type of performance bonus for managers, primary way non-managers are promoted, time to reassign or dismiss underperforming non-managers, and organization budget per employee). The observed binary choice outcome y_{ij} — taking value 1 if respondent i chooses organization j and 0 otherwise — is determined by whether U_{ij} exceeds the utility of the alternative. We assume ε_{ij} follows an i.i.d. type-I extreme-value distribution. Standard errors will be clustered at the respondent level.

The coefficients (β_i) represent the marginal utility associated with each organizational attribute and are assumed to follow a multivariate normal distribution $\beta_i \sim N(b, \Sigma)$ across

respondents. The mixed logit is estimated by maximum simulated likelihood. A larger coefficient indicates that the attribute makes an organization more attractive relative to the alternative, thereby increasing the likelihood that it will be chosen. The exact change in choice probability depends on the combination of attributes in the choice set, so the coefficients reflect the relative strength of preferences rather than a fixed change in probability.

The model requires two normalizations, which do not affect the relative preferences. First, we normalize the scale of the coefficients by setting the coefficient on the budget to one, making other coefficients interpretable in terms of budget trade-offs. Second, we normalize the location by setting the mean utility of the baseline option to zero to provide a reference point.

To quantify the relative importance of management practice attributes, we compute marginal rates of substitution between each attribute and the budget per employee. Since the budget coefficient is normalized to one, the estimated coefficient on each management practice attribute can be directly interpreted as the budget equivalent of that attribute — that is, the change in budget per employee that would make a consultant indifferent between an organization with a weaker practice and one with a stronger practice on that dimension.

As a complementary framing, these marginal rates of substitution can also be interpreted as willingness to accept: the reduction in budget per employee that a consultant would tolerate in exchange for an organization having a more favorable management practice. A larger coefficient implies that consultants place greater value on that attribute and would require less budget compensation to overlook a weakness in it. These trade-off estimates will be reported for all six management practice attributes, providing a common metric for comparing the relative weight consultants assign to each organizational characteristic.

Provided the sample size allows, we will split the sample by the following characteristics:

- Tenure in management consultancy, measured as years of professional experience.
- Expertise in management consultancy, measured as self-reported area of specialization. We will distinguish between generalist consultants and those with a specialization.
- Past experience with management consultancy in the public sector. We will distinguish between respondents who have any prior public sector consultancy experience and those who do not.
- Gender

and carry out heterogeneity analyses based on them.

4.2 Belief-Updating

For the belief-updating experiment, we will calculate the effect of revealing the results from Fenizia et al. (2026) on the average and median posterior beliefs, as measured in the elicitation exercise of the effect of management consultancy interventions on emergency rooms. The following model will be estimated:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i \quad (2)$$

where Y_i is the mean or median belief about the effectiveness of management consultancy interventions in public emergency rooms (the “posterior”), T_i is a binary indicator for whether respondent i was assigned to the treatment group (i.e., was revealed the results from Fenizia et al. (2026)), and X_i is the mean or median belief about the effectiveness of management consultancy interventions in public VTIs (the “prior”). We will include covariates selected by LASSO as a robustness check.

This model will be estimated using an OLS regression. Standard errors will be clustered at the respondent level.

In addition to the mean and median posterior, we will estimate equation (2) using the posterior variance as the outcome variable, to test whether the treatment reduces consultants’ uncertainty about the effectiveness of public sector management consultancy — independently of any directional shift in beliefs. A reduction in posterior variance among the treatment group would indicate that the disclosure of the VTI results increases consultant confidence, even among those whose point estimate of effectiveness does not change substantially. We will also examine whether the treatment changes the shape of the belief distribution, testing specifically for a reduction in bimodality using Hartigan’s dip test applied to the distribution of posterior beliefs in treatment and control groups separately.² Visualization of the full empirical prior and posterior distributions for both groups will complement these tests.

As a test of generalization, we will estimate equation (2) using two additional outcomes derived from the post-exercise paired question. The primary generalization outcome is the within-respondent public-private effectiveness gap, defined as the public sector response minus the private sector response. This structure within the respondent controls for general optimism or pessimism about consultancy, isolating the treatment effect on the perceived public-private distinction specifically. We will also estimate equation (2) using the public

²With approximately 100 respondents per group, the Hartigan’s dip test may have limited power to reliably detect bimodality or its reduction. This test should be interpreted as exploratory.

sector response alone and the private sector response alone as separate specifications. We pre-specify the following expectations: (i) the treatment will reduce the public-private gap (i.e., the gap will be smaller in absolute terms for the treatment group), driven primarily by an upward shift in the public sector response rather than a downward shift in the private sector response; (ii) the treatment coefficient on the public sector general belief will be smaller in magnitude than on the hospital-specific posterior, reflecting partial but not full generalisation of the evidence to broader public sector contexts; and (iii) the treatment coefficient on the private sector general belief will be small and statistically indistinguishable from zero, serving as a placebo check (since the VTI evidence provides no direct information about private sector consultancy effectiveness, any large treatment effect on this outcome would suggest demand effects or diffuse persuasion rather than genuine belief updating about the public sector).

We will also split the sample by the following characteristics:

1. The degree to which respondents underestimated the effectiveness of management consultancy interventions in public VTIs, measured as a continuous prior-signal perceived gap as in Haaland et al. (2023). We will also estimate the impact of treatment and its heterogeneous treatment effects along the perceived gap on learning rate as shown in Haaland et al. (2023).
2. Tenure in management consultancy, measured as years of professional experience.
3. Expertise in management consultancy, measured as self-reported area of specialization. We will distinguish between generalist consultants and those with a specialization.
4. Past experience with management consultancy in the public sector. We will distinguish between respondents who have any prior public sector consultancy experience and those who do not.
5. Gender.

and carry out heterogeneity analyses on them.

We will probe the validity of our design in two ways. First, we will test for similarity between priors' empirical CDFs in control and treatment groups using a Kolmogorov-Smirnov test.

Second, we will evaluate whether participants assigned to the treatment and control groups look similar along demographic characteristics and experience as consultants.

The main potential form of non-compliance in our setting would arise from spillover effects. Specifically, a respondent assigned to the control group could be indirectly exposed to the

treatment if they hear about the findings of Fenizia et al. (2026) from a colleague who has already completed the survey and was assigned to the treatment group.

We believe this risk is minimal. Only a small number of employees within each organization are invited to participate, and individuals are not necessarily aware that their colleagues are also taking the survey. Moreover, the probability that respondents have encountered the findings of Fenizia et al. (2026) outside of our experiment is extremely low, as the study has only been presented at a few conferences and has not yet been circulated widely or posted online.

To address potential non-compliance, we will estimate regressions using treatment assignment rather than treatment status. With perfect compliance, the coefficient of interest recovers the Average Treatment Effect (ATE). With imperfect compliance, it estimates the ITT effect. In either case, if spillover effects were present, our estimates would represent a lower bound on the true effect.

References

- Fenizia, A., Lemos, R., and Pantelaiou, I. (2026). Management consultancy in the public sector: Evidence from greece. Manuscript in preparation.
- Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40.
- Iacovone, L., McKenzie, D., and Meager, R. (2025). Bayesian impact evaluation with informative priors: An application to a colombian management and export improvement program. *Econometrica*, 93(5):1915–1935.