New registration

Review

Only admins can submit a registration.

Metadata

Title

Training Authenticity at Work: A Field Experimental Evaluation in the UK Civil Service

Description

While the concepts of authenticity at work and authentic leadership -emphasizing self-awareness, an internalized moral perspective, balanced processing of information, and relational transparency- have gained prominence in organizational research, causal evidence on their causes and consequences remains limited. This study contributes to the literature by providing casual, field experimental evidence on the effects of training employees in authenticity and authentic leadership. We exploit a staggered roll-out of the 'Achieving Your Potential' training in the UK civil service, with participants randomly assigned to earlier and later start dates of the training, to assess the effects of the training on participants' authenticity at work and authentic leadership.

Contributors

Remove me

Eleanor Florence Woodhouse, Christian Schuster, Ine Steenmans, and marc esteve

Category

Uncategorized

Affiliated institutions

No affiliated institutions

License

CC-By Attribution 4.0 International

Subjects

Psychology Social and Behavioral Sciences Leadership Studies

Public Affairs, Public Policy and Public Administration

Industrial and Organizational Psychology Public Administration

Tags

No tags

Study Information

Hypotheses

Our hypotheses are developed drawing on studies of authentic leadership and authenticity (see, e.g., Gardner, Avolio, Luthans, May, and Walumbwa, 2005). Authenticity is thereby understood as "owning one's personal experiences, be they thoughts, emotions, needs, preferences, or beliefs, processes captured by the injunction to know oneself" and acting in accordance with one's true self (Harter, 2002). We follow Walumbwa, Aviolo, Gardner, Wernsing and Peterson (2008) in understanding authentic leadership as a pattern of leader behavior with four key dimensions: "greater self-awareness, an internalized moral perspective, balanced processing of information, and relational transparency on the part of leaders working with followers, fostering positive self-development."

Self-awareness refers to an understanding of how one makes meaning of the world, of one's strengths and weaknesses, including by gaining insight into oneself through exposure to others, and of one's impact on other people (Kernis, 2003). Relational transparency refers to presenting one's true self to others, openly sharing information, building networks of collaborative relationships and expressing one's true thoughts and feelings all while minimizing displaying inappropriate emotions (Kernis, 2003). Balanced processing refers to objectively analysing all relevant data before coming to a decision, encouraging diverse viewpoints and soliciting views that challenge one's positions (Gardner, Avolio, Luthans, et al., 2005). Internalized moral perspective refers to internalized self-regulation, guided by internal moral standards and values and leading to behavior consistent with one's internalized values (Ryan & Deci, 2003; Gardner, Avolio, Luthans, et al., 2005).

The Achieving Your Potential training content focuses in part on enhancing authenticity of employees in non-managerial roles at work, as well as enhancing their authentic leadership and self-development. Our hypotheses reflect these objectives, and the dimensions of authentic leadership detailed above.

Our core hypothesis is:
H1: Training employees in authenticity and authentic leadership enhances their authenticity and authentic leadership at work

Our secondary hypotheses are:
H2: Training employees in authenticity and authentic leadership increases their self-awareness
H3: Training employees in authenticity and authentic leadership increases their relational transparency
H4: Training employees in authenticity and authentic leadership strengthens their internalized moral perspective
H5: Training employees in authenticity and authentic leadership improves their balanced processing of information
H6: Training employees in authenticity and authentic leadership enhances employee self-development efforts

Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

No blinding is involved in this study.

Is there any additional blinding in this study?

N/A

Study design

We employ a wait-listed field experimental design, in which individuals opting to participate in the 'Achieving Your Potential' training are, through simple random assignment, assigned into a treatment and a wait-listed control group.

Our sample are UK civil servants voluntarily opting to participate in the 'Achieving Your Potential' (AYP) training. AYP is a 4-month programme for aspiring line managers consisting of four modules delivered over 16 weeks. The training is offered free-of-charge by the UK Cabinet Office to UK civil servants, with a focus on employees in non-managerial ranks. Interested UK civil servants could register for the programme online. As a mandatory pre-requisite for enrolment, civil servants had to complete a baseline survey.

This survey included a block of demographic and professional background questions, a block of questions assessing the respondents' understanding of topics covered in the AYP training, a block of questions measuring employee attitudes towards authenticity and authentic leadership, and finally a bock of questions about the frequency with which respondents engage in authenticity-related practices and authentic leadership at work. At the end of the survey, the respondents were asked to insert their email address.

Responders were then, through simple random assignment, assigned into treatment and control groups, with a 50:50 weighting. The treatment group started the AYP training in June 2025 and the control group will start training in November 2025. An endline survey will be fielded to all respondents, before the control group have access to the AYP training.

Please see attachments below for the full baseline survey and endline survey questionnaires. The questionnaire contains additional measures of interest to the UK Civil Service but not included in our pre-registration.

- AYP_baseline.doc
- AYP_endline_treatment.doc
- AYP_endline_control.doc

Randomization

As noted above, the respondents to the baseline survey were randomised with a 50:50 weighting into the treatment (June start date) and control (November start date) groups by the researchers.

Sampling Plan

Existing Data

Registration prior to analysis of the data

Explanation of existing data

NB: Baseline survey completed, and treatment commenced. Endline survey not yet in the field.

Data collection procedures

Our sample are UK civil servants voluntarily opting to participate in the 'Achieving Your Potential' (AYP) training. The training is offered free-of-charge by the UK Cabinet Office to UK civil servants, with a focus on employees in non-managerial ranks. Civil servants were advised in centrally distributed newsletters and blogs as well as through cross departmental talent leads about the training. Interested UK civil servants could register for the programme online. As a mandatory pre-requisite for enrolment, civil servants had to complete a baseline survey in June 2025. We exclude any participants who did not complete the full survey or did not give their consent for us to store their responses or have their responses retained in pseudo-anonymised form for research purposes. 1,134 civil servants completed the survey, gave us consent and provided a valid email address at the end of the survey.

Responders were then, through simple random assignment, assigned into treatment and control groups, with a 50:50 weighting. The treatment group started the AYP training in June 2025 and the control group will start training in November 2025. An endline survey will be fielded to all respondents, before the control group have access to the AYP training.

No files selected

Sample size

1,143 civil servants completed the baseline and were assigned to treatment or control group. Our final sample size is the subset of baseline survey participants who also complete the endline survey and indicate the same email address in the endline survey, thus allowing us to link baseline and endline survey results.

Sample size rationale

The maximum sample size consists of the maximum number of training places available in the 'Achieving Your Potential' training. While there is little evidence on the expected effect sizes of authenticity and authentic leadership training, meta-analyses on leadership training interventions in general (e.g., Lacerenza et al., 2017, Journal of Applied Psychology) report effect sizes with Cohen's d = 0.20 to 0.35 on behavioral and perceptual outcomes. Using conservative assumptions - Cohen's d of 0.2, equal group sizes, power of 0.9, alpha of 0.05 and without baseline covariates - requires a total N of 526. Our analysis is thus well-powered to detect even small training effects.

Stopping rule

Endline survey data collection will stop prior to the first training session of the control group, to ensure the control group remains untreated. The control group will not be able to access the training until November 2025, and the endline survey will be fielded prior to that.

Variables

Manipulated variables

The treatment in our case is assignment to the AYP training programme at the time of endline survey. This will be defined as a binary variable:

Treatment = 1: Individual was assigned to complete the AYP training before the endline survey
Treatment = 0: Individual was assigned to complete the AYP training after the endline survey

As explained previously, the random allocation to treatment groups following completion of the endline survey is what decides allocation to the treatment or control group.

In addition to estimating the average treatment effect (ATE) of assignment to the AYP training, we will also estimate the complier average causal effect (CACE), using random assignment as an instrument for actual training completion. This will allow us to identify the causal effect of the training among those who complied with their assignment.

No files selected

Measured variables

H1: Authenticity at work and authentic leadership
Factor analytic approach to latent measures of self-awareness (H2), relational transparency (H3), internalized moral perspective (H4) and balanced processing of information (H5)

H2: Self-awareness.
Items (measured on an 1-5 agreement scale, from strongly disagree to strongly agree)
- I can list my three greatest strengths
- I am sometimes confused about the way I really feel about things at work [reversed]

Items measured on a frequency scale (In the last month, how frequently have you done each of the following at work?, from “Never” to “Daily”)
- Set time aside to reflect on your strengths and areas for development
- Sought feedback as a way of understanding who you really are as a person

H3: Relational transparency
Items (measured on an 1-5 agreement scale, from strongly disagree to strongly agree)
- I share my honest thoughts and feelings when I talk to others at work

Items measured on a frequency scale (In the last month, how frequently have you done each of the following at work?, from “Never” to “Daily”)
- Admitted your mistakes to others
- Set time aside to build authentic connections with colleagues

H4: Internalized moral perspective
Items (measured on an 1-5 agreement scale, from strongly disagree to strongly agree)
- I can clearly articulate my core personal values
- My actions at work reflect my core values

Items measured on a frequency scale (In the last month, how frequently have you done each of the following at work?, from “Never” to “Daily”)
- Reflected on your personal values and how they relate to Civil Service values

H5: Balanced processing of information
Items (measured on an 1-5 agreement scale, from strongly disagree to strongly agree)
- I listen closely to the ideas of those who disagree with me
- During conversations, I always try to see things from the other person's point of view
- For many decisions at work, I rely heavily upon my instincts [reversed]
- When I am convinced about something, it takes a lot of evidence to the contrary to make me

change my mind [reversed]

H6: Employee self-development efforts
Items (measured on an 1-5 agreement scale, from strongly disagree to strongly agree)
- I take active steps to become the best person I can be
- As time goes on, I feel that I am becoming more my true self at work
- I always encourage myself to try my best at work
- I believe that with effort, I can improve at almost any skill I need for my job

Items measured on a frequency scale (In the last month, how frequently have you done each of the following at work?, from “Never” to “Daily”)
- Taken deliberate action to improve yourself
- Set time aside for your development

Manipulation checks:

Knowledge gain from training, with knowledge scores measured as simple additive index of correct responses to knowledge questions:
- Which of the following are part of self-awareness?
- Which of the following are core Civil Service values?
- “Asking for general feedback can often help you improve more broadly than asking for feedback on very specific areas.” True or false?
- Which of the following are key components of global listening? True or false?

How many of the four webinars did you attend? (Either live or by watching a recording)
0-4 or Don’t know / prefer not to respond

How many problem-solving sessions did you attend?
0-4 or Don’t know / prefer not to respond

Did you read the content on FutureLearn for the module before each webinar?
- Yes, for every module
- Yes, but only for some of the modules
- No, I did not read the FutureLearn content before any of the webinars
- Don’t know / prefer not to respond

Did you complete the workbook tasks for each module?
- Yes, for every module
- Yes, but only for some of the modules
- No, I did not complete any workbook tasks
- Don’t know / prefer not to respond

Did you complete the Learning Journal for each module?
- Yes, for every module
- Yes, but only for some of the modules
- No, I did not complete any Learning Journals
- Don’t know / prefer not to respond / did not attend any sessions

How would you rate the following on a scale from 1–10?
- The overall quality of the training

No files selected

Indices

We outline the details of our approach to creating indices in the "Construct Definition and Item Selection" section of the Analysis Plan.

No files selected

Analysis Plan

Statistical models

HYPOTHESIS TESTING

To estimate the causal effect of the AYP training programme on participants' attitudes towards authenticity at work, and practices related to authenticity and authentic leadership, we will use an ANCOVA model.

The dependent variables will consistent of latent or single-item measures capturing knowledge gains about authenticity at work and authentic leadership (manipulation check), different dimensions of authenticity at work and of authentic leadership, and an overall authenticity at work and authentic leadership outcome measure.

The key independent variable is a binary indicator of whether the participant was randomly assigned to receive the training prior to completing the endline survey.

For each hypothesis, we will estimate the following model:
$Y_{i1} = \alpha + \beta * Treatment_i + \gamma * Y_{i0} + \varepsilon_i$

Where:
Y_{i1} is the endline score on the authenticity measure,
$Treatment_i$ is a binary variable indicating treatment assignment,
Y_{i0} is the corresponding baseline score,
ε_i is the error term.

This specification allows us to estimate the treatment effect while adjusting for baseline levels of the outcome. Including baseline scores improves statistical precision and helps account for individual-level variation in pre-treatment attitudes.

We will estimate this model separately for each pre-registered outcome.

Depending on final item selection, we may use a factor-analytic approach to construct a latent variable or compute an additive index (e.g., based on agreement with statements such as "I share my honest thoughts and feelings when I talk to others at work" or "My actions at work reflect my core values").

In addition to estimating the average treatment effect (ATE) of assignment to the AYP training, we will also estimate the complier average causal effect (CACE), using random assignment as an instrument for actual training completion. This will allow us to identify the causal effect of the training among those who complied with their assignment.

Although participants were randomly assigned to training timing, imbalances may arise by chance, especially in smaller samples. If baseline balance tests reveal substantial differences between groups

on authenticity and authentic leadership or key covariates at baseline, we will complement the ANCOVA with a Difference-in-Differences (DiD) analysis as a robustness check.

We will estimate the following model:

$$Authenticity_it = \alpha + \beta1 * Treatment_i + \beta2 * Post_t + \beta3 * (Treatment_i * Post_t) + \varepsilon_it$$

Where:

Post_t is an indicator for post-training measurement (endline),
Treatment_i * Post_t is the interaction capturing the change in authenticity among treated participants relative to controls,
β3 is the treatment effect of interest.

This model compares within-individual changes over time between the treatment and control groups and will help assess the robustness of the primary results.

CONSTRUCT DEFINITION AND ITEM SELECTION

Our core hypothesis is that the training intervention enhances participants' authenticity and authentic leadership at work. To assess this, we developed questions in which employees self-assess their attitudes and practices in regards to authenticity at work and authentic leadership, adapted and extended from Northouse (2025, chapter 9).

We define the authentic leadership construct as encompassing the following subdimensions: self-awareness (H2), relational transparency (H3), internalized moral perspective (H4) and balanced processing of information (H5).

Each subdimension will be measured using multiple items (typically ~3) and treated as indicators of a higher-order latent construct of authenticity at work and authentic leadership.

All items to be included in the measurement of this construct will be documented in the appendix of the PAP.

Measurement Approach

We distinguish between two measurement strategies:

Additive Indices:
For measures where items are factual or reflect unidimensional knowledge (e.g., post-training knowledge checks), we will construct additive indices by computing the standardised mean across items.

Latent Constructs (Factor Scores):
For constructs that reflect unobservable attributes (including authentic leadership and its dimensions) we will estimate factor scores using confirmatory factor analysis (CFA). CFA will be based on the pre-specified measurement model reflecting the theoretical subdimensions listed above.

Decision Rule for Item Retention and Factor Structure

To ensure consistency and transparency, we will apply the following decision rule for evaluating item performance:

For constructs measured using items on different response scales (e.g., agreement vs. frequency formats), we will use a loading threshold of ≥0.30 for item retention. While this is slightly lower than the conventional threshold, it is justified given that scaling differences tend to reduce inter-item correlations. As argued by Bagozzi & Yi (1988), factor loadings as low as 0.30 can be considered acceptable when scale heterogeneity is expected. We draw on prior measurement approaches, such as Butterworth et al. (2024), who compare Neider and Schriesheim's 2011 study which uses an agreement scale with Walumbwa et al.'s 2008 study which uses a frequency scale. This precedent supports our approach of integrating both response formats in our own latent scale development.

For constructs where all items are measured using the same response scale (e.g., Hypothesis H5 "Balanced processing of information" where all items are measures on a 1-5 agreement scale), we will apply a stricter loading threshold of ≥0.40, in line with standard practice.
In all cases, we will exclude items that exhibit cross-loadings of ≥0.30 on unintended factors.
A factor or subdimension will be retained if at least three items load cleanly (meeting the relevant loading threshold) on it.

If fewer than three items meet this criterion within a subdimension, we will:

(a) collapse across subdimensions and estimate a unidimensional factor model if justified by overall model fit; or
(b) construct a standardized additive index of the original items and interpret results with appropriate caution.

All deviations from the planned measurement structure (e.g., item exclusion, model modifications) will be reported transparently in an appendix to the main analysis.

No files selected

Transformations

To prepare the data for analysis, we will undertake the following data transformations and coding procedures:

Scale Harmonization for Latent Constructs:
Several latent constructs include items measured on different response scales (i.e., 1–5 agreement scale and 1–5 frequency scale). To facilitate latent variable modelling, all items will first be standardized (z-scores) within the sample before estimation (as described in more detail in the "Statistical models" section). This standardization ensures that differences in scale format do not unduly influence factor loadings or bias composite indices.

Reversed Items:
Items worded in a negative or reverse-coded direction will be reverse-scored prior to analysis so that higher values consistently reflect greater presence of the latent construct.

Additive Indices:
Where indicated (e.g., knowledge checks), responses will be scored as binary correct/incorrect and summed to form a total score. These additive indices will not be standardised unless required for modelling purposes.

Categorical Engagement Variables:

Responses capturing participant engagement with programme components (e.g., number of webinars attended, whether workbooks were completed) will be treated as ordinal or categorical variables as appropriate to their response type (i.e. agreement scale = ordinal). These may be dummy-coded or collapsed for subgroup analysis depending on distribution and analytic needs (e.g., 0 = no exposure; 1 = partial exposure; 2 = full exposure).

Centering:
Where relevant (e.g., in models including interaction terms), continuous variables will be mean-centred to aid interpretability and reduce multicollinearity.

Handling of “Don’t Know / Prefer Not to Respond”:
Responses marked as “Don’t know / prefer not to respond” will be treated as missing for the purposes of analysis.

All transformations and coding decisions will be implemented consistently across treatment and control groups and documented in a reproducible script to ensure transparency.

Inference criteria

We will use a conventional null hypothesis significance testing framework to draw inferences about the effect of the training intervention on participants’ sense of authenticity at work. All primary analyses will use two-tailed tests with a standard $\alpha = 0.05$ threshold for statistical significance. For each outcome, we will report p-values, 95% confidence intervals, and standardised effect sizes (e.g., Cohen’s d) to aid interpretation.

If we test multiple dependent variables related to authenticity (e.g., several items or subscales), we will apply a Holm-Bonferroni correction to account for the risk of Type I error due to multiple comparisons, unless we construct a single index or latent factor in advance to serve as the primary outcome.

In addition to p-values, we will assess model fit and robustness by checking for balance on covariates, estimating alternative specifications (e.g., Difference-in-Differences model in case of treatment/control imbalance), and conducting sensitivity analyses. All model assumptions (e.g., homoskedasticity, normality of residuals) will be examined to ensure the validity of inference.

Data exclusion

We will exclude observations with missing data on key variables required for analysis (e.g., treatment assignment, outcome measures, or baseline covariates). Participants who did not complete both the baseline and endline surveys will be excluded from ANCOVA analyses. Outliers on outcome variables will be inspected, and we may conduct robustness checks using winsorized variables if extreme values appear to influence results.

We will include an additional form of manipulation check in the endline survey. Specifically, we will collect self-reported data on participants’ engagement with the training (e.g., number of sessions attended, proportion of online content completed). This will allow us to construct measures of treatment intensity, such as a binary compliance indicator or a continuous index. These will be used in exploratory analyses to explore heterogeneous impacts by level of engagement.

Missing data

We will begin by assessing the extent and patterns of missing data for all key variables, including treatment status, authenticity outcomes, and relevant covariates. If item nonresponse is low and

appears to be missing completely at random (MCAR), we will proceed with complete case analysis. If there is evidence of systematic missingness, we will consider using multiple imputation to recover information, particularly for baseline covariates and outcome measures. For key outcome variables, we will report the rate of missingness and assess robustness by comparing results across complete case and imputed samples, where applicable.

Exploratory analysis

In addition to the pre-registered confirmatory analyses, we may conduct exploratory analyses to better understand heterogeneity in treatment effects. This could include examining whether treatment effects on employee authenticity vary by baseline characteristics such as managerial experience (whether participants are line managers), tenure, or authenticity and authentic leadership at baseline. Specifically, vis-à-vis managerial experience, we are interested in exploring whether treatment effects differ between line managers and non-line managers. Although this subgroup analysis is of substantive interest, we note that the number of line managers in our sample is relatively small (N = 70, compared to 1,058 non-line managers at baseline). Given the limited statistical power for detecting effects in this subgroup, we will interpret any observed differences cautiously, as exploratory analyses. We may also explore relationships between outcomes and engagement with different components of the training programme (e.g. attendance, online content completion), using our training intensity measures. Any such analyses will be clearly marked as exploratory in any reporting and interpreted with appropriate caution. We will also complement the survey data analysis with a set of interviews and participant observations of the training.

Other

Other

ADDITIONAL INFORMATION

This study forms part of a broader evaluation of line management training programmes in the UK Civil Service. Additional information about the evaluation methodology and programme context is available at: https://www.gov.uk/government/publications/evaluating-line-management-capability-training.

REFERENCES:

Avolio, B.J., & Gardner, W.L. 2005. Authentic leadership development: Getting to the root of positive forms of leadership. Leadership Quarterly, 16: 315-338

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. Journal of the Academy of Marketing Science, 16, 74-94.

Butterworth, M., Black, J., & Terry, R. (2024). Authentic leadership measures: an authentic measure for authentic leadership?. The Journal of Values-Based Leadership, 17(1), 13.

Gardner, W.L., Avolio, B.J., Luthans, F., May, D.R., & Walumbwa, F.O. 2005. “Can you see the real me?” A self-based model of authentic leader and follower development . Leadership Quarterly, 16: 343-372.

Harter, S. 2002. Authenticity. In C. R. Snyder & S. J. Lopez (Eds.), Handbook of positive psychology: 382-394. London: Oxford University Press.

Kernis, M.H. 2003. Toward a conceptualization of optimal self-esteem. Psychological Inquiry, 14: 1-26.

Neider, L. L., & Schriesheim, C. A. (2011). The authentic leadership inventory (ALI): development and empirical tests. The Leadership Quarterly, 22, 1146-1164.

Northouse, P. G. (2025). Leadership: Theory and practice. SAGE publications.

Ryan, R.M., & Deci, E.L. 2003. On assimilating identities to the self: A self-determination theory perspective on internalization and integrity within cultures. In M. R. Leary & J. P. Tangney (Eds.), Handbook of self and identity: 253-272. New York: Guilford.

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic Leadership: Development and Validation of a Theory-Based Measure†. Journal of Management, 34(1), 89-126.