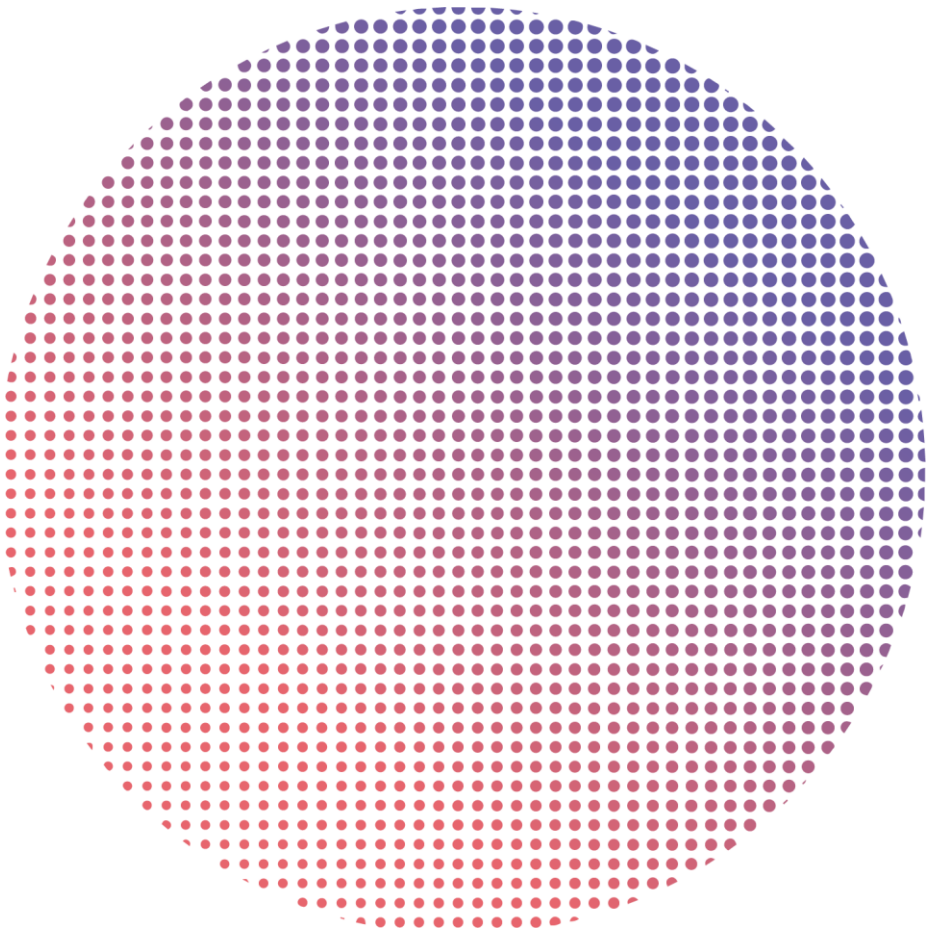
Effective Professional Development Design in a Civil Service Context – A systematic review

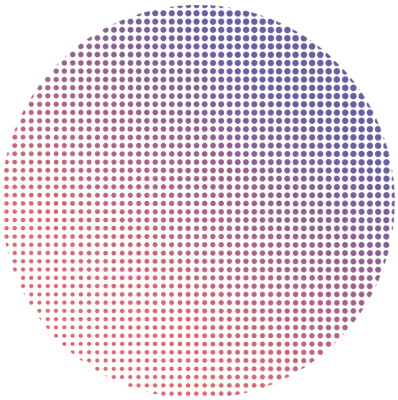
Prepared for Government Skills

Government Skills

May 2025



alma economics



About the authors

alma economics

Alma Economics combines unparalleled analytical expertise with the ability to communicate complex ideas clearly.

www.almaeconomics.com

About the commissioning organisation

Government Skills

Government Skills is part of the Cabinet Office and the Government People Group. Government Skills supports civil servants to develop the skills, knowledge and networks they need to deliver their best for citizens and ministers, now and in the future.

www.governmentskills.com

Table of Contents

Table of Contents 3

Abstract 5

Introduction 7
- Review rationale 7
- Objectives 9

Methodology 11
- Overall approach 11
- Eligibility criteria 11
- Information sources 14
- Search strategy 15
- Selection process 15
- Data extraction process 16
- Data items 17
- Development of taxonomy 22
- Study risk of bias assessment 23
- Effect measures 24
- Evidence mapping and synthesis methods 25
- Assessment of reporting biases 28
- Confidence in cumulative evidence 28

Results 29
- Study selection 29
- Study characteristics 30
- Taxonomy 32
- Risk of bias in studies 33
- Results of individual studies 34
- Results of synthesis 35
- Reporting biases 64
- Certainty of evidence 64

Discussion 66
Discussion of findings 66
Limitations of evidence 67
Limitations of review processes 70
Implications for practice and policy 71
Implications for further research 72
Conclusion 76

Administrative Information 77
Registration and protocol 77
Deviations from protocol 77
Support 78
Competing interests 78
Availability of data, code, and other materials 78

References 79

Appendix A: Search strategy 84

Appendix B: Evidence map 111

Appendix C: Taxonomy 120

Appendix D: Risk of bias in studies 133

Appendix E: Results of individual studies 135

Appendix F: Albatross plot 140

Appendix G: Best practice example of abstract 141

Abstract

Objectives: We conducted a systematic review of the literature to identify and synthesise high-quality literature on professional development in both civil service and adjacent contexts. Our first objective was to identify the characteristics of effective professional development. After identifying relevant reports, we: (i) explored the effectiveness of these interventions in driving improvements in knowledge, skills, networks, work performance, and productivity, (ii) examined the mechanisms, design features, and forms of professional development that drive the greatest impact, and (iii) identified factors that support the successful implementation of professional development programmes in the civil service and adjacent contexts.

Methods: We used the PICO framework to develop a set of criteria utilised to determine the inclusion of studies in our synthesis. We included English-language RCTs and quasi-experimental studies from OECD countries that examined professional development interventions for the most common UK civil service professions, both within the UK civil service sector and the broader public and private sectors. Eligible studies compared interventions to a comparator group and assessed their impact on knowledge, skills, networks, work performance, and productivity. We searched a range of academic bibliographic databases (e.g., Web of Science, Scopus) and grey literature repositories and also conducted a call for evidence and citation searches. The risk of bias within the included studies was assessed using the revised JBI critical appraisal tools for both RCTs and quasi-experimental studies. Due to the limited number of studies identified, we provide a narrative synthesis in addition to vote counting. The search protocol was also registered with the UK government evaluation registry and is available on the UK government website. This work was commissioned by Government Skills, which is part of the UK Cabinet Office and funded by HMT Labour Markets Evaluation and Pilots Fund.

Results: Our search yielded a total of 27 studies in 26 reports, of which 20 studies were independent. Participants in the included studies were most commonly from operational delivery professions (e.g., probation officers, call centre workers) based in North America. RCTs formed the majority of the included papers. The majority of papers found a positive effect of professional development, with skill acquisition being the most commonly assessed outcome, as well as the outcome with the most common significant results. While heterogeneity of study characteristics limited a meta-analysis of effect sizes, vote counting and albatross plots confirmed a strong positive association. Drawing conclusions on the most effective design features was challenging, due to limitations in reporting and the small number of identified studies providing causal evidence. Nonetheless, there was an indication that design features combining different training approaches appeared to yield better outcomes for participants, as all studies that showed statistically significant results used multiple types of professional development methods. The most commonly identified mechanisms for change included behavioural practice, instruction, and feedback on behaviour.

Discussion: Despite consistent evidence of positive effects from professional development interventions relevant to the civil service context, this review highlights significant limitations in the underlying evidence base. The small number of eligible studies, inconsistent outcome measures, and poor reporting practices limit confidence in the findings and constrain conclusions about what makes interventions effective. For practice and policy, findings suggest a need to link skills development more clearly to performance outcomes, ensure organisational support for implementation, and consider multi-component training approaches. Future research should prioritise improved reporting standards, better measurement of productivity and performance outcomes, and further refinement of the taxonomy of underlying mechanisms to support more effective design and evaluation.

Introduction

Review rationale

Globally, the public sector is a major employer, accounting for 17% of total employment and 38% of formal employment, according to the Worldwide Bureaucracy Indicators (World Bank, 2023). As of March 2024, the UK civil service employs around 540,000 people in various departments, agencies, and professions, playing a key role in shaping employment practices, policies, and standards throughout the broader public sector workforce (Cabinet Office, 2024). As part of this role, the civil service makes an ongoing investment in learning and development (L&D), covering a wide range of activities from induction and early career training to the development of senior leaders. This provision spans both core and specialist skills, tailored to departmental and professional needs.

The public sector is also an important employer of high-skilled labour (World Bank, 2023), and its ability to attract, retain, and develop talent is central to delivering high-quality public services. Against the backdrop of a changing labour market – with more people opting out of the workforce or being unable to work, as well as a growing mismatch between current skills and future needs – it is understood that training and development programmes must be effective and well-targeted (Frontier Economics, 2022). This is particularly important given the documented productivity challenges faced by the UK economy and public sector, often linked to chronic underinvestment, including in training and development of skills (e.g., Office for National Statistics, 2023; Van Reenen and Yang, 2024; The Productivity Institute, 2022; The Productivity Institute, 2024).

Strengthening the evidence base on the effectiveness of professional development interventions can support better decisions about where and how to invest. In turn, enhanced productivity can lead to improved service delivery and greater social value for a given level of expenditure – in other words, more efficient and impactful public spending (Romani, 2021). Ensuring that leaders, line managers, and L&D professionals have access to the best available evidence on what works in professional development is central to making this investment as effective as possible. Within the civil service, Government Skills in the Cabinet Office acts as the strategic centre for L&D, while individual departments and professions are responsible for delivering provision aligned to their workforce needs.

In light of the skills and productivity landscape within the UK's public sector, it is important to work towards an understanding of how and under what conditions investment in skills development can be a driver of public sector organisational performance, delivery, efficiency, and productivity. Given the influence of the civil service on the wider public sector, this investment is also likely to have an influence on the wider economy and society.

Currently, a considerable amount is spent by the UK civil service on professional development and learning activities. Previous systematic reviews in the professional

development area (e.g., Filges et al., 2019; Sims et al., 2021) define professional development as facilitated learning opportunities for qualified professionals that aim to enhance professionals' knowledge and skills, in ways that are relevant for application in practice – that is, to serve the ultimate beneficiaries.

In order for the UK civil service to ensure its investments in professional development are as effective as possible, there needs to be extensive and high-quality evidence synthesis work to support future decision-making. To understand the quality of the current evidence base, we undertook an initial overview of the literature as well as scoping interviews with relevant professionals during the initial scoping phase of this project to assist methodological planning. The evidence base consists mostly of work carried out on the design features and mechanisms of effective professional development in other public sectors, as can be seen within comprehensive guides for the teaching and medical professions (e.g., Education Endowment Foundation, 2021; AMEE, 2024).

Filges et al. (2019) conducted a systematic review on the effectiveness of continuing professional development for welfare professionals (i.e., those working within education, social welfare, and crime and justice). No eligible studies were identified for social welfare or crime and justice, with all 51 eligible studies relating to education. However, even within the education-related literature, there are limitations. Sims and Fletcher-Wood (2020) carried out a critical review of reviews identifying characteristics of effective teacher professional development and found the evidence base to employ, at times, unsuitable inclusion criteria and to depend on inference methods that lack robustness. Research into effective professional development in the education sector has also identified a number of concerns, including high workloads and competing priorities of learners, a lack of funding for programmes, and a lack of belief in the efficacy of programmes (Adams et al., 2023).

Similar work to the evidence reviews described above has not been conducted in the context of the civil service. The current evidence base related to the civil service is also of variable quality, fragmented, and from various international contexts (Elliott et al., 2023). This has been a long-standing issue within this field of research, with Chapman (1987) identifying a lack of a unified framework of concepts within public administration teaching, as well as a wide gap between academic research and practitioners.

Against this background, Government Skills has commissioned three projects within an integrated programme of evidence synthesis and evaluation to develop high-quality and actionable evidence on the links between effective learning and development interventions, workforce skills, and productivity. As part of this, the current systematic review synthesises the evidence on the design features (such as duration or mode of delivery) and mechanisms of professional development associated with the effective development of productivity-enhancing skills within a civil service context.

This review employs a systematic search of the literature exploring effective professional development design. Information regarding the 'mechanisms' of interventions was also collated to inform the development of a taxonomy of effective mechanisms for professional development design. Mechanisms are 'active

ingredients' that would make an intervention less effective if they were to be removed (Sims et al., 2021). A synthesis of the findings of the identified literature is then carried out.

This systematic evidence review will inform the creation of evidence-based tools and recommendations for future investment into professional development within the UK civil service. In practice, these tools and recommendations can be utilised to improve professional development interventions, both within and outside of the civil service. The insights generated can be used by a range of audiences, including practitioners, researchers, and learners. These improvements can also help address key concerns identified in research on professional development, such as high workloads and competing priorities of learners, limited funding for programmes, and scepticism about their effectiveness (e.g., Adams et al., 2023). By tackling these challenges, the review can contribute to the prioritisation of professional development across various professions.

Objectives

This review aims to synthesise existing high-quality literature to identify the characteristics of effective professional development in both civil service and adjacent contexts. It seeks to: (i) explore the overall effectiveness of these interventions in driving improvements in knowledge, skills, networks, work performance, and productivity, (ii) examine the mechanisms, design features, and forms (clusters of mechanisms) of professional development associated with the greatest impact on driving these improvements, and (iii) identify factors that support the successful implementation of professional development programmes in the civil service and adjacent contexts.

To achieve this, four research questions are addressed:

1) What are the characteristics of the studies and interventions in the experimental impact evaluation literature on professional development design in the civil service and adjacent contexts?
2) Overall, how effective are professional development interventions in the civil service context at improving knowledge, skills, networks, work performance, and productivity?
    a) Does this vary based on study characteristics (features of the evaluation not specific to the intervention itself)?
    b) Does the effectiveness of the interventions vary based on the types of workers and the target outcomes?
3) Which design features, mechanisms and forms of professional development are associated with the greatest impact?
    a) Which design features (e.g., online versus face-to-face; longer duration versus shorter duration) are associated in the literature with the greatest impact on skills, knowledge, networks, work performance, and productivity?

b) Which mechanisms and forms (clusters of mechanisms) do we observe in the literature?

c) Which mechanisms or forms (combinations of mechanisms) are associated in the literature with the greatest impact on skills, knowledge, networks, work performance, and productivity?

4) What supports the successful implementation of professional development interventions targeted at driving improvements in knowledge, skills, networks, work performance, and productivity in the civil service context and adjacent contexts?

Methodology

Overall approach

We conducted a systematic review (SR) of evidence on professional development in the civil service and adjacent contexts. An SR is ideally suited to providing an unbiased and reliable assessment of the available evidence by using transparent, well-defined, and replicable procedures. By synthesising insights from a fully comprehensive set of literature that satisfies specified criteria, it also delivers a clear view of where gaps are in the evidence base, helping to provide foundations for future research.

Our protocol resolved to use meta-analysis to synthesise the magnitudes of quantitative estimates of effect sizes of professional development, where the comparability of studies and the measures used rendered this feasible. Where these conditions were not satisfied, our protocol outlined that we would use narrative synthesis, according to SWiM guidelines (Campbell et al., 2020). As we explain more fully later, narrative synthesis is the route we ultimately took upon examination of the highly heterogeneous literature. We supplemented this with vote counting – a quantitative synthesis technique that is an alternative to meta-analysis, focused on the strength of evidence around the existence and direction of an association, but not its magnitude.

Eligibility criteria

The eligibility criteria used in this review are built around the PICO framework, which focuses on population, intervention, comparator, and outcome. The framework helps define the scope with precision, making it one of the most appropriate models for structuring intervention-based research questions for systematic reviews.

- **Population**: At the centre of our analysis were studies involving workers employed by the civil service. However, our scoping searches revealed a lack of high-quality evidence when including solely the international civil service workforce. Therefore, we broadened the scope of the search while ensuring the evidence remains relevant to the civil service. Namely, we included studies that target workers operating in UK civil service professions across the civil service, the wider public sector, and the private sector. This approach is based on the assumption that interventions targeting these professions in other settings (e.g., the wider public sector and the private sector) still focus on outcomes relevant to those same professions in the civil service context. We focused on the five civil service professions that employ the most workers in the civil service and cover more than 75% of the civil service workforce in the UK namely, 1) Operational Delivery, 2) Policy, 3) Digital, Data & Technology, 4) Project Delivery, and 5) Tax (Cabinet Office, 2024). Interventions targeting people in the wider public and private sectors who are not part of these five professions were excluded. Employees from the wider public and private sector in roles that

are very general and appear across all professions are also excluded, despite appearing in the professions mentioned above, such as general roles like managers or leaders. As these roles are common, there is a large literature base regarding their professional development, which has been analysed in existing systematic reviews (e.g., Busso et al., 2023; Lacerenza et al., 2017; Lundqvist et al., 2023).

- **Intervention**: Building on previous systematic reviews of professional development design (e.g., Filges et al., 2019; Sims et al., 2021) and for the purposes of this review, professional development is defined as the deliberate process of acquiring and driving knowledge, skills, networks, work performance, and productivity for professionals. Hence, this review focused on interventions aimed at enhancing professionals' knowledge, skills, and networks or interventions targeting improvements in work performance and productivity. Studies with interventions deliberately aiming to enhance professionals' knowledge, skills, networks, work performance, or productivity were included. Studies that did not involve any interventions related to professional development or that focused on interventions unrelated to knowledge, skills, networks, or work performance improvements were excluded.
- **Comparator**: Only studies that explored the effect of an intervention compared to a comparator group (e.g., alternative intervention, no intervention, or pre-intervention period) were included. Comparator groups included alternative forms of a professional development intervention (e.g., a limited form of the intervention), no professional development intervention or a pre-intervention period (e.g., Interrupted Time Series (ITS) studies). Studies that did not have a comparator were excluded.
- **Outcomes**: Only studies that contained a quantitative outcome measure for (or proxies for) changes in skills, knowledge, networks, work performance, or productivity were included. Studies focusing on other outcomes (e.g., job satisfaction, wellbeing, organisational culture, job retention) or without a quantitative outcome measure were excluded. The following definitions of outcome variables were used:
    - **Knowledge**: Knowledge is a familiarity, awareness, or understanding of someone or something, such as facts, information, or descriptions, which is acquired through experience or education, by perceiving, discovering, or learning. Knowledge can refer to a theoretical or practical understanding of a subject (adapted to context from Librarianship Studies & Information Technology, 2017).
    - **Skills**: The ability to apply knowledge and use know-how to complete tasks and solve problems (CEDEFOP, 2014).
    - **Networks**: Interpersonal relationships that connect individuals or organisations, facilitating the exchange of ideas and resources to achieve specific goals. They encompass both informal sources, such as

personal contacts, and formal structures, like organisational connections, contributing to collaboration and information sharing within professional development contexts (adapted to context from Saltiel, 2006).

- **Work performance**: Work performance encompasses not only task proficiency but also adaptive and proactive behaviours that enable individuals to respond to dynamic changes and take initiative in improving processes or outcomes (adapted from López-Cabarcos, Vázquez-Rodríguez, and Quiñoá-Piñeiro, 2022).
- **Productivity**: Productivity refers to the efficiency and effectiveness with which inputs (skills, knowledge, resources) are converted into improved outputs, such as service delivery and social outcomes. This encompasses both technical efficiency (performing current tasks better) and allocative efficiency (ensuring the right activities are pursued to maximise impact) (adapted from Aldridge, Hawkins, and Xuereb, 2016).

- **Geographical focus**: To ensure the relevance of findings to the UK civil service, this review included studies only from OECD member countries. OECD member countries have comparable economic and administrative structures to the UK, making their findings more applicable and relevant to the UK context than non-OECD countries. Studies from non-OECD countries were excluded to maintain high contextual relevance and comparability.
- **Methodology**: Only studies using randomised controlled trials (RCTs) or quasi-experimental research designs were included. To ensure we capture a sufficient amount of evidence while obtaining valuable insights into the effectiveness of interventions, we included all types of quasi-experimental designs, including single-group pre-test/post-test studies. All other research designs were excluded.
- **Date of publication**: To ensure the relevance of our findings to contemporary civil service practices, we only included studies published between 1st January 2004 and 1st September 2024. This 20-year period captures significant public sector reforms and the rapid digital transformation that have reshaped professional development within the civil service. The early 2000s marked the start of major initiatives like the Modernising Government agenda and the Gershon Review, which emphasised efficiency, accountability, and the adoption of digital technologies (UK Government, 1999; see also Dunleavy et al., 2006). These changes necessitated new skills and continuous learning for civil servants, making this period crucial for understanding current professional development trends.
- **Language of publication**: We only included studies written in English. By concentrating on English-language publications, we aimed to maintain a manageable scope for the review while ensuring that the evidence is accessible and understandable to the intended audience. Studies published in other languages were excluded.

- **Types of publication**: Journal articles, working papers, research reports, theses, and dissertations were included. Book chapters were excluded, as they do not follow standardised reporting guidelines and can vary significantly in methodological transparency, making it harder to assess their quality and reliability within the systematic review framework.

The documentation and implementation of the screening process (including the order of screening) are described in the Study Selection section. To maximise the insights we can gain with a limited evidence base, studies in which a subsample of the population and geographical focus fulfils the inclusion criteria have been included. In applying this approach, we assume that a professional development program aimed at civil service leaders (who are eligible) and leaders in the broader public sector (who are not eligible) remains relevant to the review. Further information on how data was extracted from the included studies is presented in the Data items section.

For the narrative synthesis, studies will be grouped by intervention. We anticipated considerable heterogeneity in population characteristics, outcome types, and outcome measurement approaches across the included studies. Grouping the narrative synthesis by intervention allowed us to draw the most overarching conclusions by focusing on the effectiveness of different professional development approaches, despite this variation.

Information sources

We retrieved evidence from academic databases and grey literature repositories, as well as through a call for evidence and through citation searches of papers included via the other sources. We recorded the information source each study was retrieved from, and we applied the inclusion/exclusion criteria to all of them. Table 4 in Appendix A shows the results, dates of search, filters applied, and search strings for each database and repository searched. The PRISMA flow diagram presented in the Study selection section also shows the results for the additional ways of adding papers. We used the following data sources to search for evidence:

- **Academic bibliographic databases:** Education Resources Information Center (ERIC), IDEAS/RePEc, Web of Science, Scopus.
- **Grey literature repositories:** gov.uk (Type filters: ‘Research and statistics’ and ‘Policy Papers and Consultations’ only), OECD iLibrary (Content type filter: Journals, Articles, and Papers), ProQuest (Source type filter: ‘Government & Official Publications’, ‘Reports’, ‘Scholarly Journals’, ‘Dissertations & Theses’ and ‘Working Papers’ only), World Bank Open Knowledge Repository, Campbell Collaboration, Cedefop (Content type filter: Publications only), Google Scholar (limited to first 50 results only).[1]

[1] The Cochrane Handbook highlights that limiting search results from search engines that produce more entries than can be feasibly screened is a sensible approach, particularly given the often-low precision of these searches. Furthermore, as replicability tends to be more challenging with search engines compared to bibliographic databases, we will export and document the first 50 entries and use

- **Additional ways of adding papers.** In addition to the public databases, we collected research and studies from the following sources:
    - **Call for evidence and stakeholder engagement**: We issued a public call for evidence to address potential publication bias. Alongside this, we conducted interviews with key stakeholders, requesting any relevant materials they may have.
    - **Snowballing**: We used snowballing techniques to expand our evidence base, including both backward and forward citation tracking. This involved reviewing the reference lists of studies included in the systematic review via other sources to identify additional relevant papers, as well as examining studies that have cited the included papers. We implemented this process using the online tool SpiderCite, with documentation of the source.

Search strategy

Table 4 in Appendix A presents a list of keywords that were used to identify sources of evidence relevant to our research questions. This list was finalised during the scoping phase, where various combinations of keywords were tested. Based on these documented searches, some keywords were dropped where they did not add more relevant results (e.g., education).

These keywords were combined into search strings using Boolean operators (AND/OR) and other database-specific search operators. This long list of materials was then screened to test if the inclusion criteria outlined in the previous section were met. The three keyword groups were combined with AND operators and were searched for in title, abstract, and indexing terms/keywords (where available).

As each database has different search settings, we recorded all search strings and filters used across the different databases. In line with PRISMA guidelines and to ensure a transparent search process, these details are presented under Appendix A. The outcomes of these searches (i.e., the dates and numbers of retrieved items) are also presented in Appendix A, as well as a research activity sheet that documents the search and screening phase and is available on request.

Selection process

The process followed to select the studies relevant to this review is outlined below.

- **Identification**:
    - We stored all records identified from our database searches and other search methods, described in the Information Sources section above, in Zotero, clearly indicating from which sources the records were identified.

privacy mode in our browser to ensure we minimise the influence of personal recommendations or algorithms on the search results (Lefebvre et al., 2023).

  - After storing the records, we began the selection process by removing duplicate records. The number of duplicates removed is recorded in the PRISMA flow chart below (Figure 1).
- **Screening**:
  - **Tool**: For the screening process, we used Rayyan software. Rayyan is a free-to-use software to support systematic reviews that allows the process of screening to be sped up through the use of a visual, colour-coded interface that highlights keywords associated with the eligibility criteria. It also documents the screening decisions of multiple reviewers for reconciliation and assessing the inter-rater reliability score. While Rayyan includes AI features, such as assigning a ranking of the likelihood of relevance to unreviewed articles based on already screened articles, we screened all articles manually and did not remove any articles based on any automated tools.
  - **Title/Abstract:** Our interdisciplinary team began the screening process by scanning the titles and abstracts of records and applying the eligibility criteria to decide which studies should be retrieved for full-text screening in Rayyan. During this step, we implemented partial double screening, with two reviewers reviewing 10% of the records independently. Single screening commenced only after an agreement between the reviewers was reached at 90%. Throughout the process, reviewers held regular discussions to ensure consistency in study inclusion and exclusion. As 'Date of Publication,' 'Type of Publication,' and 'Language of Publication' are typically accessible through database filters or during the title and abstract screening stage, records lacking this information at that stage were excluded. The inclusion/exclusion decision of any records during the screening of titles and abstracts was recorded in the Research Extraction Sheet (RES) and the PRISMA flowchart (Figure 1).
  - **Full Text:** At this stage, the full text of any record that was not excluded was retrieved. Since not all eligibility criteria can be fully assessed based solely on the title and abstract, a full-text review was conducted in Rayyan for any records that were not excluded in the previous stage. The full-text review was conducted by two reviewers who worked independently. The eligibility criteria were assessed with the following hierarchy: 1) Geographical Focus, 2) Population, 3) Intervention, 4) Outcomes, 5) Methodology, and 6) Comparator. The eligibility decision, along with the reason for exclusion, where applicable, was documented in the RES. Additionally, a summary of the records included at this stage is presented in the PRISMA flowchart (Figure 1).

Completing these steps led to the identification of the studies included in this review.

Data extraction process

Following the selection of the included studies, we extracted the data used for the evidence mapping, data synthesis, and assessment of the evidence. This information

was documented in the ‘Data and Synthesis’ sheet of the RES.

Reviewers then discussed each element of the data extraction table in detail to ensure a shared understanding of each data element. Data extraction was then piloted in order to check whether additional guidance or definitions were required. An independent double extraction was performed on a randomly selected 10% sample of the studies. This was performed by the two reviewers who conducted the full-text review.

Discrepancy resolution procedure

The following procedure was used to address discrepancies in data collection of the sample:

- Any discrepancies were to be discussed in agreement meetings.
- If reviewers did not reach a consensus, a third one would be consulted for a final decision.
- Full agreement on all assigned codes in this subset must be achieved before proceeding with further single extraction.

If concerns about the interpretation were to persist after this phase, the team would assess whether additional double extraction or further refinements to the tool are necessary to resolve these issues.

Following a triangulation of findings from this sample, two reviewers separately extracted study characteristics and numerical outcome data from the included studies.

Data items

Table 1 presents the data items that were extracted from all included studies.

Table 1. Data items

Evidence Map / Additional Coding	Category	Detail
Evidence Map	**Bibliographic information**	• Title • Authors • Type of publication • Publication date • Source • Conflict of interest disclosed: Yes | No • Abstract text • Appendix: Yes (same document) | Yes

		(separate document) | No

Evidence Map	**Study characteristics**	• General:
		  ○ Country of focus
		  ○ Time period covered
		• Methodology:
		  ○ Study Design: RCT | Quasi-experimental
		  ○ Comparator: No intervention | Lighter intervention | Pre-intervention period
		  ○ Number of treatment groups
		  ○ Process evaluation elements reported: Yes | No
		  ○ Process evaluation wording (if process evaluation elements are reported)
		• Population:
		  ○ Professions: Operational delivery | Policy | Digital, Data & Technology | Project Delivery | Tax | Other civil service | Mixed (report all that apply)
		  ○ Sector: Civil Service | Wider public sector (incl. charities/non-profits) | Private sector | Mixed (report all that apply) | Unclear/Unreported
		  ○ In vivo description of job role
		  ○ Managerial role: Yes | No | Mixed (report all that apply) | Unclear/Unreported
		  ○ Tenure: Average tenure (in years) | Unclear/Unreported
		  ○ Gender: % female or non-binary | Unclear/Unreported
		  ○ Ethnicity: % non-white | Unclear/Unreported
		  ○ Age: Average age | Unclear/Unreported
		• Interventions:
		  ○ Level: Individual | Team | Organisation | Unclear/Unreported

  - Group size intervention: Individual | 2-5 | 6-10 | 11-20 | 21-50 | 50+ | Mixed (report all that apply) | Unclear/Unreported
  - Duration of the intervention in total: (Numeric in hours) | Unclear/Unreported (e.g., if an intervention was comprised of ten 2-hour sessions over 1 year, the duration is 1 year)
  - Dosage: Total time participants are actively engaged (e.g., if an intervention was comprised of ten 2-hour sessions over 1 year, the dosage is 10 x 2 hours = 20 hours)
  - Number of sessions: Total count of distinct training meetings (e.g., if an intervention was comprised of ten 2-hour sessions over 1 year, the number of sessions is 10)
  - Type of Engagement: Online – Synchronous | Online – Asynchronous | Face to face – Residential | Face to face – Workplace | Mixed (report all that apply) | Unclear/Unreported
  - Design: Internal | External | Unclear/Unreported
  - Delivery: Internal | External | Self-Administered | Mixed (report all that apply) | Unclear/Unreported
  - Attendance: Voluntary | Compulsory | Mixed | Unclear/Unreported
  - Type: Lecture/Seminar | Coaching | High-tech simulation-based learning | Low-tech simulation-based learning | Coaching | Action learning sets | Other (In vivo) | Mixed (report all that apply) | Unclear/Unreported
  - In vivo category of the type of training
- Outcomes:
  - Category of targeted outcome: Skills | Knowledge | Network | Work performance | Productivity | Mixed

		(report all that apply) ○ Outcome as described in study ○ Outcome measure as described in study
Additional coding for synthesis	**Quantitative data for the Meta-analysis**	• Quantitative data: ○ Sample Size (N Overall, N Intervention, N Control) ○ Type of outcome variable: Binary | Continuous ○ Mean / Outcome (Intervention / Control) ○ Standard error (Intervention / Control) ○ Effect size type (e.g., regression coefficient, Cohen's d) ○ P-value ○ Timing of outcome measurement after intervention: Numeric | Multiple (Report all) | Unclear/Unreported ○ Wording used by the study to report findings related to the outcome
Additional coding for synthesis	**Mechanisms**	• We coded which mechanisms of the developed taxonomy appear in each intervention. The methodology is explained in the Development of taxonomy section, and the findings in the Taxonomy section.
	Risk of bias assessment	• Following tools specified in Study risk of bias section.

Missing data

Authors of five papers were contacted in cases where quantitative information important for our quantitative synthesis was unavailable or where the independence of papers was not clear. Three out of the five did respond to us, but only two were able to provide the information we asked for. We did not find any systematic bias in the non-reporting of these variables.

Dependent effect sizes

Dependent effect sizes were addressed in cases where multiple effect sizes were reported from the same study, or the same participants contributed to multiple outcomes. Failing to address this dependency can lead to underestimated standard errors and inflated Type I error rates. We outlined this issue and our approach for handling it in the protocol in the case of meta-analysis. While we are now replacing the meta-analysis with a vote-counting exercise due to the heterogeneity of the

evidence, the same logic persists. Bushman and Wang (2009) emphasise that the sign test of the vote counting methodology does require independent estimates for valid inference. We followed a reductionist approach to do so (López-López et al., 2018) and used the following rules to determine the main effect size per study:

1) When effect sizes were available for multiple types of relevant outcomes, measures of work performance or productivity take precedence over those related to skill and knowledge acquisition. As outlined in the Review rationale section, this review is focused on skill development within the wider context of efforts to improve productivity. Hence, although learning and development can have other objectives (e.g., worker wellbeing, job satisfaction, and retention), it makes sense in our context to treat skills as an intermediate outcome that, it is hoped, can improve work performance or productivity. Where measures of productivity or performance are directly available, it is therefore these that we focus on, since whether a skill improvement would translate into a productivity improvement is a separate empirical question and not one that can be taken for granted.
2) Preference was given to outcomes that are widely applicable, commonly used in the literature, or easier to compare across studies. This enhanced the consistency of the analysis and allowed for more meaningful cross-study comparisons. This is a practical approach used in quantitative synthesis, where outcomes that are most frequently reported across studies, most familiar to readers, or most frequently measured in practice are selected, as this improves comparability and reduces heterogeneity (López-López et al., 2018).
3) When effect sizes were reported at different levels, individual-level measures were prioritised over organisational-level ones, as the focus of this review lies on the change in outcome for the individuals participating in the intervention.
4) When multiple treatment groups were available, those receiving the more intensive intervention were prioritised (although treatment groups consist of different participants, effect sizes are estimated by comparing their outcomes with the same control group, so the estimated effect sizes are not independent). Stronger interventions are more likely to produce detectable effects, providing clearer insights into the relationship between the intervention and the outcome. This approach also helps ensure that the selected effect sizes reflect the full potential impact of the intervention rather than partial or weaker effects. This decision is supported by evidence from Cordingley et al. (2015), who found that more effective professional development for teachers is typically prolonged, includes iterative follow-up activities, and creates a sustained rhythm of support to enable deeper changes in practice.
5) For studies in which only a subsample of the population and geographical focus fulfil the inclusion criteria (see Eligibility criteria), preference was given to effect sizes focused on the eligible subsample only. If no outcome sub-analysis was done, the findings for the total sample were extracted.

Where multiple rules could be applied, they were applied sequentially. This approach

was systematically applied in all quantitative syntheses or analyses, including vote counting, the albatross plot, and the effect direction plot, to ensure that dependency among effect sizes was appropriately managed. By eliminating dependency, we ensured the validity of the statistical test in our vote counting procedure. Descriptive visual representations of effect sizes (the albatross plot and effect direction plots) also drew only on the set of independent effect estimates, for consistency.

In any other areas of the systematic review, the full set of relevant papers was used. The evidence map (research question 1) and the data extraction on study characteristics were done for all papers included in the evidence review. For the narrative synthesis, we discussed all results that were of interest to the specific research question. Here, we were able to provide the necessary context to acknowledge instances where multiple effect sizes originated from the same study or where two papers were based on the same underlying study. This allowed for a nuanced interpretation of the findings, ensuring transparency regarding the contribution of each study to the synthesis.

To ensure that the rules of the reductionist approach have not accidentally introduced a systematic bias, we have also checked the effect direction reported for the excluded outcomes. The results of this sensitivity check are presented with the results of the vote counting.

Development of taxonomy

During our systematic review of the evidence, we developed a coding framework designed to capture the specific mechanisms underpinning each intervention (research question 3b). The coding framework was applied during the synthesis stage to systematically analyse and compare the effectiveness of different learning and development interventions.

Factoring in steers from Government Skills, sector experts, the advisory board, as well as findings from our own scoping, we chose to use the Behavioural Change Technique (BCT) taxonomy as a base for developing a taxonomy for this systematic review, given the existing precedent of having been successfully applied by Sims et al. (2021) (“BCT Taxonomy (v1): 93 hierarchically-clustered techniques,” n.d.). The BCT taxonomy organises 93 mechanisms into 16 groups. The following three-step process was performed to ensure the adapted taxonomy is fit for purpose and a relevant, useful, and intuitive framework to categorise public sector professional development mechanisms for policymakers and other practitioners:

- **Testing and refining BCT taxonomy on an initial sample of studies:** As adopted by Sims et al. (2021), once the study selection process has been completed, we tested the mechanisms in the BCT taxonomy. Due to the small number of papers found, this was completed for all studies eligible for inclusion in the review. Throughout this process, frequent discussions took place to clarify any uncertainties about the mechanisms that should be applied and identified. The mechanisms in each of these studies were iteratively coded until they were judged to be suitably relevant and applicable. Following this step, to

arrive at a draft taxonomy, we (i) removed irrelevant mechanisms or those that are not applied in a public sector professional development context (66 removed in total), (ii) refined definitions to make them more relevant to the public sector (one mechanism refined in total), and (iii) added mechanisms that were missing from the BCT taxonomy but are present in the public sector professional development literature (two mechanisms added in total). This sample coding also allowed us to iteratively refine and operationalise the definitions of the mechanisms in a consistent and standardised way.

- **Refining the draft taxonomy after initial feedback:** Once the draft taxonomy was developed, we consulted with Government Skills and the project advisory board to validate our approach as part of the quality assurance process. No refinements were made at this stage.
- **Finalising of taxonomy and approach to coding:** During the synthesis, each included study was coded by our team of researchers using the taxonomy. After coding approximately 50% of the studies, we held a review point discussing among the team whether there was a significant indication that the taxonomy required adaptation. We also considered whether grouping or aggregating certain mechanisms would be beneficial for some steps of the synthesis or for the practical use of the taxonomy tool; however, no changes to the original groupings of the BCT taxonomy were made.

Both the original BCT taxonomy and our revised version of the taxonomy are presented in Appendix C.

Study risk of bias assessment

The risk of bias was also assessed for all the included studies. This formed an important component of our review, making us aware of any reported estimates that may be biased in size. The following tools were used to complete this assessment:

- **Randomised controlled trials**: We used the revised JBI critical appraisal tool for the assessment of risk of bias for randomised controlled trials, a comprehensive framework developed for assessing bias in randomised studies (Barker et al., 2023).
- **Quasi-experimental studies**: We used the revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies, which is designed to evaluate studies where participants are not randomly assigned to intervention groups (Barker et al., 2024).

Using both tools allowed for a robust evaluation of study quality across different research designs.

Using these tools, we calculated a risk of bias assessment score. Each tool consists of a series of questions or categories designed to assess factors that may influence a study's bias (e.g., whether treatment groups were randomly assigned or if the study clearly establishes cause and effect). A "yes" response to these questions indicates a lower risk of bias. An independent double assessment was performed on a randomly selected 10% sample of the studies. Discrepancies were discussed in agreement meetings, with a third reviewer consulted if needed, ensuring full consensus on assigned answers before proceeding with further single assessment.

To determine the risk of bias score, we counted the number of affirmative responses and expressed them as a proportion of the total applicable questions in the tool. If a question was "Not applicable" to a particular study, it was excluded from the total count. As a result, the final score represents the proportion of applicable questions that received affirmative responses, providing a measure of the study's overall risk of bias.

Effect measures

We calculated Hedges' g for standardised mean differences (SMD) where the necessary information was available – that is, where the information on the mean and standard deviation of the outcome (after treatment) and the sample size is available or can be derived for the treatment and control group. To do so, we used the effect size calculator of the Campbell Collaboration. We were not able to do this for every included paper due to missing information. Appendix E states whether effect sizes were calculated or taken as reported by the authors.

For p-values, there is a large number of studies for which these are either not reported or reported only as significance levels (which effectively bounds the p-value but does not give its exact value). In these cases, to facilitate comparison through an albatross plot, we assumed that the t-statistic follows a normal distribution and calculated the p-value using the standardised effect size and the standard error, where that information was available.

Evidence mapping and synthesis methods

Eligibility for synthesis

For each of the research questions, we identified the studies that contributed to the results. Table 9 in Appendix E shows, for each paper included in the review, which research question it was deemed relevant for and whether it was included in the vote-counting exercise. Any papers that provided results on the overall link between a professional development intervention and one of the targeted outcomes were potentially eligible for addressing research question 2, but not all of these papers could be used for the quantitative synthesis, as there were cases where more than one paper contained results based on the same underlying experiment. The results of those papers were not used in the quantitative synthesis, since having the same participant appear more than once would invalidate the assumptions underlying the statistical test embodied by vote counting. Our approach to dependent effect sizes is discussed in detail in the Dependent effect sizes section. Further information on the vote-counting exercise is presented in the Synthesis Methods section. However, the results are described in the narrative synthesis, with a clear statement on which papers were not independent from each other. Any papers that presented results on how treatment effect vary based on specific factors were matched with the relevant sub-questions; (i) 2a for study characteristics, (ii) 2b for type of worker or target outcome, (iii) 3a for design features, and (iv) 3c for mechanisms and forms (combinations of mechanisms).

Preparing for synthesis

The methods used to prepare the data for presentation and synthesis are described in detail in the Data items and Effect measures sections. The former describes how we handle missing data and the process for dependent effect sizes, and the latter includes an overview of how we standardise effect sizes and p-values before presentation.

Tabulation and graphical methods

We use the following tabulation and graphical methods in our presentation and synthesis of the evidence:

i) A table summarising the information collected for the evidence map from each individual paper in Appendix B.

ii) A table summarising the information collected for the quantitative synthesis from each individual paper in Appendix E.

iii) A table presenting the results of the vote-counting exercise, which was the quantitative synthesis method used.

iv) An albatross plot that presents the p-values of the included results against the sample sizes of studies (Harrison et al., 2017). This is used to present the overall findings for the main results relevant to research question 2. The plot was made in Stata using a package designed by the inventor of the plot.

v) Effect direction plots to show the number of results favouring the intervention versus the control and whether the results were statistically significant, grouped by specific variables to inform the subgroup analysis. These were made in R using the ggplot package.

vi) Bar plots to inform the descriptives around the mechanisms and forms (clusters of mechanisms) appearing in the literature.

Evidence mapping

Research question 1: What are the characteristics of the studies and interventions in the experimental impact evaluation literature on professional development design in the civil service and adjacent contexts?

The findings of our first research question are presented by employing a systematic evidence map. An evidence map involves conducting a comprehensive search across a broad field, followed by coding the identified studies based on specific characteristics. All studies included in the review were coded according to the criteria of the extraction tool presented in the Data extraction process and Data items sections. The findings of the systematic evidence mapping are presented in a user-friendly table format in Appendix B and summarised in the main text.

Synthesis methods

Research question 2: Overall, how effective are professional development interventions in the civil service context at improving knowledge, skills, networks, work performance, and productivity?

While we intended to conduct a meta-analysis, if possible, this was contingent upon the quantity and the characteristics of included studies. In light of the small number of identified independent studies and the very wide range of outcome types and measurement approaches, we concluded that meta-analysis as a quantitative synthesis method would not be appropriate.

Instead, we opted for vote counting based on the direction of effect as an alternative quantitative synthesis method, as recommended by the Cochrane Handbook for Systematic Reviews of Interventions (Chapter 12) (McKenzie and Brennan, 2024). This approach involves assessing whether individual studies report effects that favour the intervention or favour the comparator, without relying on standardising the magnitudes of effects across a very widely heterogeneous set of measures. In this exercise, the statistical significance of any individual estimated effect is ignored, so as not to disproportionately favour well-powered studies. Instead, the data on estimated effect directions are aggregated across studies in order to statistically test a different null hypothesis: namely, that the true effects across the studies are evenly balanced around zero (equivalently, that 50% are positive). According to Cochrane guidelines, vote counting by effect direction is particularly useful when meta-analysis is not feasible, as it does not rely on being able to meaningfully compare the magnitudes of effects across studies but still allows for a structured synthesis of findings with a rigorous statistical test.

In addition to the quantitative synthesis, we described and synthesised study findings narratively following the SWiM reporting guideline (Campbell et al., 2020). SWiM enhances transparency in systematic reviews by providing clear guidelines for reporting narrative synthesis methods, which improves the credibility and reproducibility of reviews where meta-analysis is not feasible.

Regarding the two sub-questions of research question 2, on how the overall effect varies based on study characteristics, type of worker, and target outcome, we were only able to address the research questions in a descriptive way. This was done by producing effect direction plots and identifying patterns for the following groups:

- Overall sample size - RQ 2a
- Risk of bias assessment scores - RQ 2a
- Target outcome - RQ 2b
- Type of worker - RQ 2b

In addition, we reported the findings of any papers that specifically analysed how the effect of training differs depending on any of the aforementioned variables.

Research question 3: Which design features and forms of professional development are associated with the greatest impact?

While our original plan was to employ elements of qualitative comparative analysis (QCA) to examine how different combinations of mechanisms contribute to the effectiveness of professional development interventions, we had to revise this approach due to limitations in the literature. Specifically, the number of included studies was too small relative to the number of mechanisms considered. Although we had planned for the aggregation of mechanisms in order to address this challenge, it could not address it sufficiently, given the small number of included studies relative to mechanisms.

To address the research question, we used the same approach as outlined for research question 2 and created the following effect direction plots:

- Dosage of training – RQ 3a
- Type of engagement (online, face-to-face, mixed) – RQ 3a
- Type of training (e.g., lecture-based training, coaching) – RQ 3a
- Number of mechanisms

In addition, we described the findings of any paper that specifically analyses how the effects of professional development interventions differ based on design features, mechanisms, and forms used in the intervention.

To describe which mechanisms and forms (clusters of mechanisms) are observed in the literature (RQ 3b), we produced bar plots to display the number of mechanisms per included paper, the most common mechanisms and forms used, and whether they differ depending on the type of worker the interventions target.

Research question 4: What supports successful implementation of professional development interventions targeted at driving improvements in knowledge, skills, networks, work performance, and productivity in the civil service and adjacent contexts?

To synthesise the evidence for the last research question, we focused on studies that included an implementation and process evaluation (IPE) element. The IPE can either take the form of a separate evaluation or be a specific component within the study. During the coding, we recorded which studies included an IPE element. Given the largely qualitative nature of the data collected during an IPE, as well as the heterogeneity in implementing the IPEs, we synthesised the evidence in narrative form using thematic analysis. In the protocol, we outlined that we will follow Sims et al. (2021) and structure the thematic analysis around the following three questions to identify support and barriers to the successful implementation of professional development interventions in the public sector:

- To what extent were interventions implemented as planned?
- What factors supported or obstructed effective implementation?
- What was the nature and effect of programme adaptations?

Assessment of reporting biases

This assessment concerns the risk of bias in the results of a synthesis, such as a meta-analysis, which may arise from missing studies or incomplete results within studies. This type of bias can occur if the decision to publish a study's or report's specific results is influenced by the observed p-value or the magnitude and direction of the effect. For instance, publication bias can occur if studies with statistically non-significant results are not submitted for publication, while selective non-reporting bias can arise if certain statistically non-significant results are omitted from published reports.

To partially mitigate the risk of publication bias, we also reviewed grey literature and sent a call for evidence to identify relevant unpublished papers.

A funnel plot was not deemed possible given the small number of papers included and the heterogeneity in the studies (which did not allow for the calculation of a standardised metric across all papers). However, we created an albatross plot, which shows whether p-values are clustered just below standard statistical significance thresholds (which is evidence of publication or reporting bias). Unlike funnel plots, the albatross plot does not provide a direct estimate of publication bias, nor does it allow formal statistical tests such as Egger's test.

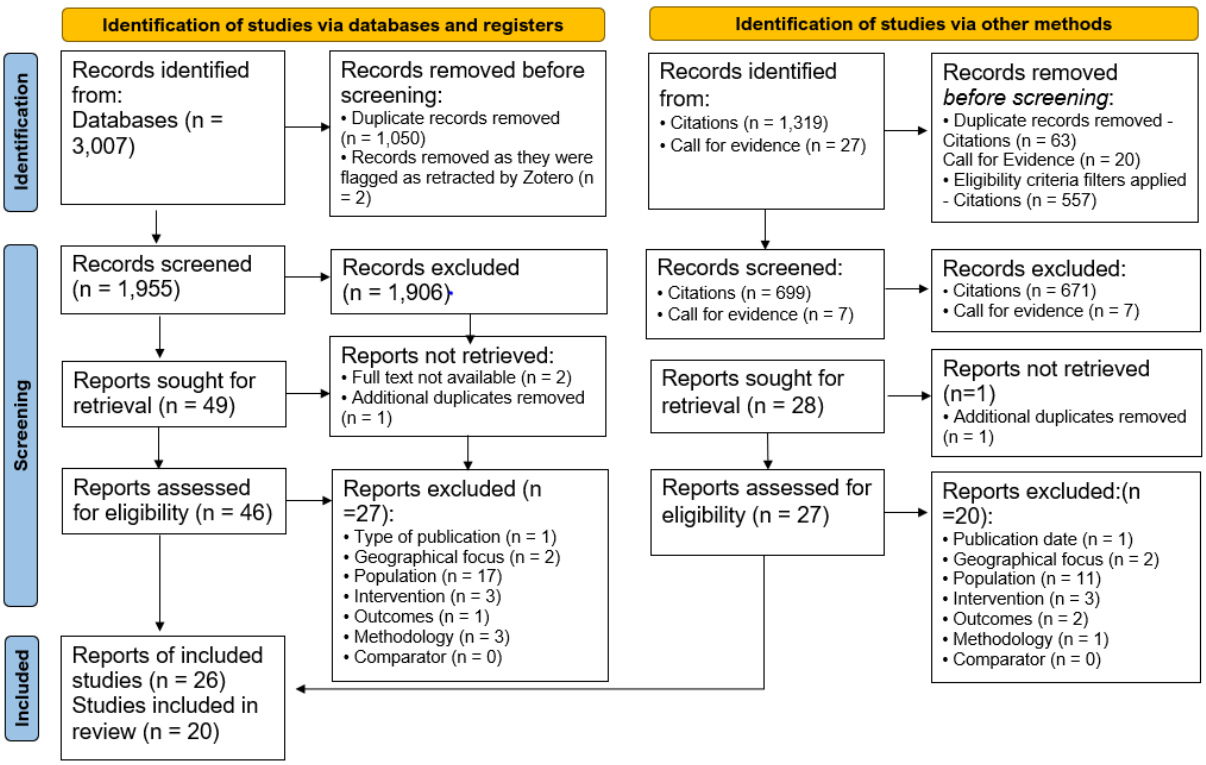
Confidence in cumulative evidence

The small number of papers and their heterogeneous nature, as well as not being able to conduct a meta-analysis, limit the scope of tools typically used for certainty assessment (such as GRADE). When assessing the certainty of the evidence in this systematic review, multiple factors were considered, including risk of bias, consistency, directness, precision, and potential publication bias.

Results

Study selection

Figure 1. PRISMA 2020 flow diagram



As shown in Figure 1, the initial searches focused on databases and registers yielded 3,007 results. After duplicates were removed, 1,955 records were included in the screening process. Based on abstract and title review, 1,906 records were excluded, and 49 full-text reports were sought to assess for eligibility. After three were not retrievable, the other 46 were assessed for eligibility, resulting in 27 exclusions. The most common reasons for exclusion at this stage were inappropriate population (28 exclusions), as well as intervention, geographical focus, and methodology. The result was 19 reports being included in the final review. We used two additional ways of adding papers: (i) a call for evidence and (ii) backward and forward citation of the 26 included reports via the databases and registers. This yielded an additional 1,319 and 27 from citations and the call for evidence, respectively. After screening, we were able to add seven more reports to the list of included reports from these two channels. Overall, the complete search yielded 26 reports, which were based on 20 independent studies. One of the reports consisted of two independent studies (we will treat them as separate reports from here onwards and speak of 27 papers), while seven reports were based on the same study as others included in the review.

Study characteristics

(Research question 1)

What are the characteristics of the studies and interventions in the experimental impact evaluation literature on professional development design in the civil service and adjacent contexts?

This section presents the study characteristics based on the data extraction conducted as part of this systematic review and answers research question 1. It summarises key aspects of the included papers, structured around setting and design, participants, interventions, and outcomes. Information on the study level can be found in the evidence map in Appendix B. A key finding is that the evidence base is highly heterogeneous, with substantial variation especially around participants, intervention types, targeted outcomes, and their measurement. This heterogeneity has important implications for the type of synthesis that can be conducted and the conclusions that can be drawn. Additionally, reporting quality was inconsistent across papers, particularly regarding details on the interventions and participant characteristics, which poses challenges for interpretation and comparison. These issues will be further explored in the discussion on limitations of the evidence and recommendations for the future.

Setting and design

Overall, 19 of the 27 papers utilised Randomised Controlled Trials (RCTs). These RCTs utilised a wide variety of structures, with many creating comparator groups subject to no intervention and many giving comparator groups lighter interventions. There is an approximately even split of papers using one, two, or three different treatment groups. The other eight papers use quasi-experimental designs (QEDs). Six of those eight papers use QED methods like difference-in-differences, which do make use of a comparator group, but not one that was created through randomisation. Two papers are single-group pre- and post-test comparisons.

The majority of the papers included North America. Eight of the papers were set in the United States, three of which were RCTs and five of which utilised quasi-experimental designs. There were five different RCTs run in Canada, along with one QED. Five studied RCTs in Denmark, although four of those concern the same RCT and are written by a combination of the same group of authors. There is only one paper studying workers in the United Kingdom. Others featured the Netherlands (two papers), Australia (one paper), Finland (one paper), Slovenia (one paper), Israel (one paper), and groups across different Latin American countries (one paper). One paper was identified that included two separate studies, each with different settings (the United States and Israel).

Participants

Two specific professions are substantially over-represented in our sample: probation officers and call centre workers. They account for 12 of the 27 papers. Another three

papers study similar operational delivery roles. This is most likely due to the availability of readily quantifiable metrics – such as call handling time and reoffending rates – that allow for clear measures of performance or productivity. Another commonly occurring sector under observation is tax professionals. The four papers from the same RCT in Denmark study leaders in tax offices (as well as other professions), as do two other papers. Other papers in our list study a range of roles, including policymakers, programme administrators, and managers in different civil service branches. Papers report a wide range of demographic distributions within their samples, in terms of gender, age, and tenure. Women are overrepresented in our included papers, averaging 66% of each sample where gender split is reported. The average age of participants is usually around 40, and there is variation by tenure, with some studies' samples having an average tenure of only two to three years and others around 20 years. This variation is often related to the studied sample; for example, call centre workers are likely to have been in the job for far less time than civil service managers.

Interventions

The systematic review captures a highly heterogeneous range of professional development interventions, spanning multiple sectors, professions, and training approaches. Some interventions focus on enhancing job-specific skills, such as probation officers receiving cognitive-behavioural training (e.g., STICS and STARR) to improve client interactions and decision-making, while others target broader competencies, like leadership development programs. The review also includes diverse training formats, such as online behavioural economics courses designed to improve policymakers' problem-solving abilities and computerised cognitive training aimed at helping older workers maintain productivity in high-demand roles. Additionally, structured in-person interventions are common, such as call centre employees receiving coaching in listening techniques to improve customer interactions or career management workshops designed to boost employees' confidence and long-term motivation. The breadth of interventions highlights the wide-ranging nature of professional development efforts, from brief, targeted workshops to extended multi-month programs with structured feedback.

Outcomes

The targeted outcomes for this systematic review were knowledge, skills, networks, work performance, and productivity. No papers were identified that targeted networks as outcomes. Many papers analysed multiple outcomes, either different outcome types or they used different types of measurements for the same outcome. Eighteen of the 27 included papers reported skill outcomes, while only seven reported on work performance outcomes, four on outcomes related to productivity, and two papers included knowledge outcomes. Most papers that included outcomes on productivity and work performance had a target population from the operational delivery profession.

The majority of studies focus on skills acquisition and application, including interventions that taught motivational interviewing, cognitive techniques, leadership

behaviours, and correctional skills. These are primarily measured through behavioural coding systems, structured audiotape ratings, self-reported surveys, and Likert-scale assessments. Work performance outcomes are typically linked to reoffending rates, failure rates, leadership effectiveness, and organisational performance, relying on official records (e.g., police databases), organisational performance metrics, and structured evaluation rubrics. Three studies focus on productivity, assessing work output, efficiency, and handling time, often measured through company data, standardised performance indices, and composite scoring systems. The remaining outcomes focus on knowledge, such as understanding of policies and professional roles, which are evaluated through content assessments, structured surveys, and self-reported measures.

Taxonomy

In this section, we provide further information on the taxonomy developed to capture the specific mechanisms underpinning each intervention. Further information on how the taxonomy was developed is provided in the Development of taxonomy section. The full taxonomy is provided in Appendix C.

As mentioned, the BCT taxonomy was used as the foundation of the taxonomy developed for this synthesis. While some additions and amendments were named, we used the same categorisation of mechanisms (i.e., groups). The groups included in this taxonomy are: 1. Goals and planning, 2. Feedback and Monitoring, 3. Social support, 4. Shaping Knowledge, 6. Comparison of Behaviour, 7. Associations, 8. Repetition and Substitution, 10. Reward and threat, 13. Identity, 14. Scheduled Consequences, and 16. Covert Learning. We use the original grouping numbers from the BCT taxonomy for ease of understanding and comparison. In total, our taxonomy includes 29 mechanisms.

Two mechanisms were added to enhance our categorisation of behavioural change techniques identified in the included studies. “Feedback (unspecified)” was introduced to capture instances where it was unclear whether feedback was focused on behaviour or outcomes. This addition ensures that all forms of feedback are accounted for, since feedback forms an essential component of the included interventions. Second, “Social support (collaboration)” was added to recognise collaborative forms of social support, which were not fully addressed by the existing categories. The BCT taxonomy included mechanisms for social support, which either focused on practical support (e.g., tangible assistance) or unspecified support (e.g., general social support without a defined focus). The new category allows for the identification of mutual engagement and shared responsibility, including joint problem-solving and co-development of strategies for behaviour change.

We also made two amendments to the existing BCT mechanisms to ensure they are better aligned for the purposes of this systematic review. Mechanism 1.7 - “Review outcome goal(s)” was updated to include instances where someone is prompted to review their outcome goals. This amendment enables us to capture cases where goal review is encouraged, a feature commonly observed in training programmes of the

included studies. Additionally, Mechanism 8.3 - “Habit formation” was revised to explicitly mention booster sessions, which were common among the included studies' training programmes. Overall, these updates improve the clarity and applicability of the BCT mechanisms, tailoring the taxonomy for the purposes of this review.

Risk of bias in studies

Tables 7 and 8 in Appendix D display the risk of bias assessment of the 19 randomised controlled trials and eight quasi-experimental designs. As mentioned in the Methodology section, we used the JBI critical appraisal tools for the assessment of RCTs and quasi-experimental studies. The tools facilitate the appraisal by listing a series of questions that assess various components of a study’s methodology. The assessor then marks these questions as Yes/No/Unclear/Not applicable. A greater share of affirmative responses indicates a more robust methodology. This assessment helps us understand the reliability of the evidence base and identify potential weaknesses that may influence how we interpret the findings.

On average, the risk of bias assessment score for RCTs was 43%, while for quasi-experimental studies, this was 70%. Although this score is higher for quasi-experiments than RCTs, this difference is attributed to the quasi-experimental tool assessing a narrower range of methodological components (e.g., nine for quasi-experiments versus 13 for RCTs). This means that risk of bias scores should not be compared directly between study designs, as they are measured using different criteria.

In analysing the key limitations identified among studies, it should be noted that the lowest-scoring areas of the risk of bias appraisal arose mainly due to unclear reporting. This lack of clarity influenced the number of affirmative responses in the assessment tool, consequently affecting the overall quality score. The key limitations identified for RCTs were in relation to internal validity and the validity of statistical conclusions, namely:

- incomplete description and analysis of differences between groups in terms of their follow-up (i.e., post-assignment attrition), which can threaten internal validity if dropouts differ systematically from those who complete the study,
- failure to analyse participants in their originally assigned groups (i.e., intention to treat analysis), potentially compromising the validity of statistical conclusions,
- lack of blinding among those delivering the treatment, which may influence its implementation,
- lack of blinding treatment conditions among participants, increasing the risk of response bias.

Specifically, (1) and (2) arose as limitations as they received the highest number of “No” responses, whereas limitations (3) and (4) arose due to the significant number of papers where reporting on these categories was unclear. Other areas commonly unreported among the studies included whether outcome assessors were blind to treatment assignment, whether true randomisation was used, and whether allocation

to treatment was concealed from participants. These issues are important because they can impact how confident we can be in the findings and whether the studies provide reliable evidence.

The key limitations identified for quasi-experiments, also related to internal validity, were:

1. uncertainty regarding the reliability of outcome measurements,
2. incomplete follow-up (i.e., post-assignment attrition), and
3. potential differences in treatment or care between comparison groups, beyond the exposure or intervention of interest.

Whereas (1) arose as a key limitation due to the number of "No" responses coded for the corresponding quality appraisal question, (2) and (3) emerged as limitations due to the significant number of papers where reporting on these categories was unclear. Both a lack of reporting and a failure to adequately address these issues can weaken our confidence in the results, as they leave uncertainty about the validity and reliability of the study findings.

No study was excluded based on the risk of bias assessment because systematic reviews aim to synthesise all available evidence, and excluding studies solely due to high risk of bias can introduce its own bias and limit the comprehensiveness of the review. Instead, we descriptively looked for patterns between the risk of bias and the effect direction and significance of the results when exploring research question 2a.

Results of individual studies

The quantitative results of each included paper are presented in Appendix E. This includes the following information for each paper where available: (i) sample size, (ii) the main outcome (if multiple, see our discussion in the Dependent effect sizes section), (iii) the type of outcome measured (binary/continuous), (iv) the treatment group (if multiple), (v) the effect size type (Hedges' g if mean differences, standard errors, and sample sizes are not missing), (vi) the effect direction, (vii) the effect size, (viii) the standard error of the effect size, (ix) the p-value, (x) the relevant research question, and (xi) whether the paper was included in the vote-counting exercise.

Any statistics calculated based on information available in the paper rather than directly extracted from the paper are formatted in bold.

Findings of selection process of main outcome for the quantitative synthesis and analysis

Out of the 27 papers identified during the review, 20 reported findings contributing to the overall effect of professional development on the targeted outcomes (research question 2). Out of these, 17 were based on independent studies. To prepare the data for the quantitative synthesis and analysis, we had to determine the main outcome per study, as described in the Dependent effect sizes section. We made the following reductions based on the rules outlined in that section:

- For four studies, no reductions had to be made as they only reported one outcome, and there was only one treatment group.
- Fifteen outcomes across four studies were not selected based on prioritising work performance and productivity outcomes over skills and knowledge outcomes.
- Fourteen outcomes across eight studies were not selected because more widely applicable, commonly used, or easier-to-compare outcomes were available in the study.
- One outcome across one study was not selected due to prioritising individual-level measures over organisational-level ones.
- Four of the 17 independent studies presented multiple treatment groups. Based on the rule of selecting the treatment group with the strongest treatment, six effect sizes were not selected for the quantitative analysis and synthesis.

Results of synthesis

Research question 2

Overall, how effective are professional development interventions in the civil service context at improving knowledge, skills, networks, work performance, and productivity?

Out of the 17 papers that were independent studies and focused their analysis on the effectiveness of professional development, 16 papers reported on the effect direction. The other one mentioned that there was no significant effect, but did not clarify whether the effect direction was positive or negative. The paper's authors were contacted; however, the missing information could not be retrieved, so the Borness et al. (2013) paper had to be removed from the vote-counting exercise.

The 16 studies included in the quantitative synthesis were conducted across ten countries, with the majority based in North America (six in the United States and two in Canada), followed by studies from Denmark (2), Netherlands (1), UK (1), Finland (1), Slovenia (1), Israel (1), and Latin American regions (1). For 11 out of the 16 studies, the target population was employees in operational delivery roles.

The majority of the 16 studies contributing to the quantitative synthesis had an RCT design (11/16). The average risk of bias for the RCT studies was 41%, while it was 70% for the five studies with a quasi-experimental design.

As discussed in the Data items section, ensuring that each participant is included only once in the synthesis is crucial to avoid biased results. Table 2 presents the selected outcome and treatment group and the associated effect direction for each of the included papers, with 15 out of 16 papers favouring the intervention. Hence, there is very strong evidence that the range of professional development approaches studied in our set of included papers does tend to have a positive effect on outcomes related to knowledge, skills, networks, work performance, and productivity (94% (95% CI 72% to 99%), p = 0.001).

Table 2. Results of vote counting based on the direction of effect

Study	Main outcome	Treatment group	Effect direction
Asteris, 2013	Motivational interviewing skills (MITI 3.1.1)	Combination group (MI learning + coaching/feedback)	Favours intervention
Bonta et al., 2011	Recidivism rate	1 treatment group only	Favours intervention
Bonta et al., 2019	Recidivism rate	1 treatment group only	Does not favour intervention
Cotabish and Robinson, 2012	Content assessment	1 treatment group only	Favours intervention
De Grip and Sauermann, 2012	Performance	1 treatment group only	Favours intervention
Dunlop et al., 2015	Understanding of Primary Authority initiative – Vignette 1	1 treatment group only	Favours intervention
Haunstrup and Jensen, 2024	Transformational leadership behaviour	Nudges with static tool	Favours intervention
Itzchakov, 2020 (Israel study)	Perspective-taking skills	1 treatment group only	Favours intervention
Itzchakov, 2020 (US Study)	Perspective-taking skills	1 treatment group only	Favours intervention
Jacobsen et al., 2022	Transformational leadership behaviour	Combination group (transformational + transactional leadership training)	Favours intervention
Labrecque and Viglione, 2021	New arrests	STARR coaches	Favours intervention
Milič Kavčič et al., 2024	Work productivity	1 treatment group only	Favours intervention
Robinson et al., 2012	Failure rate (probation)	1 treatment group only	Favours intervention

Study	Main outcome	Treatment group	Effect direction
Rojas Méndez and Scartascini, 2024	Policy decision-making skills	1 treatment group only	Favours intervention
Salmela-Aro et al., 2012	Career preparedness	1 treatment group only	Favours intervention
Seidle et al., 2016	Leadership performance	1 treatment group only	Favours intervention

Due to the low number of included papers and the heterogeneity of those identified, vote counting is the only quantitative synthesis technique we use in this systematic review. From here onwards, we use a narrative synthesis approach, supported by descriptive quantitative evidence.[2] As the next step in the narrative synthesis, we summarise the focus and findings of each of the 20 studies contributing to research question 2. This includes three additional papers that, while not independent studies, provide useful insights relevant to the research questions. To ensure clarity, it will be explicitly noted where these studies are not independent. For this part of the narrative synthesis, we grouped studies by intervention types, see the Eligibility criteria section for more information around the grouping.

Strategic training for probation officers improves their use of targeted skills, but its impact on recidivism rates remains uncertain

Bonta et al. (2019) was the only study to find results not favouring the intervention. As part of a broader investigation into the impact of probation officer training on client recidivism, this study evaluated whether the Strategic Training Initiative in Community Supervision (STICS) in Alberta, Canada, could reduce reoffending compared to standard probation practices. STICS emphasises cognitive-behavioural techniques, including cognitive restructuring (replacing pro-criminal thoughts with pro-social ones), rapport-building, and targeting criminogenic needs, supported by ongoing development measures such as monthly meetings, structured exercises, refresher courses, and individualised feedback on audio recordings. The study measured recidivism as a new conviction within two years, using official criminal records from the Royal Canadian Mounted Police, and analysed audio recordings of supervision sessions to assess how officers applied the training. In this randomised controlled trial, probation officers were assigned to either STICS training or probation-as-usual. Results showed no significant reduction in overall two-year recidivism rates between clients of trained officers and the control group. However, officers who actively applied cognitive techniques – a core STICS component – had clients who took longer to reoffend. While STICS training alone did not lower recidivism, the findings suggest

[2] Note that the rest of the report is based on a narrative synthesis. During that synthesis, all papers included in the review are discussed. Hence, their risk of bias and study characteristics are summarised in the corresponding overarching sections.

that cognitive techniques may have potential in delaying reoffending when consistently applied. Bonta et al. (2019) was a replication of an earlier RCT by Bonta et al. (2011), which used the same protocol and training manuals but was conducted in the Canadian provinces of British Columbia, Saskatchewan, and Prince Edward Island. The authors found evidence that the STICS training had a significant impact on the behaviour of the probation officers, such as better adherence to the principles underlying STICS and more frequent use of cognitive-behavioural techniques to address the pro-criminal attitudes of their clients. In terms of recidivism rates, they find a positive but not significant link between STICS training and the two-year recidivism rates of clients. Using data from the same study and hence, not being independent from Bonta et al. (2011), Bourgon and Gutierrez (2012) did a similar analysis linking the STICS training to a significant increase in the discussion of pro-criminal attitudes and the use of cognitive techniques for probation officers. While they also assessed the relationship between the usage of these techniques and recidivism, they did not analyse whether there was a significant effect of the STICS training on recidivism rates.

Robinson et al. (2012) analysed the effects of STARR training, the U.S. equivalent of STICS, on probation officers' skill usage and client recidivism rates across ten U.S. districts. Using an RCT design, they found that trained officers demonstrated greater use of targeted skills, and their clients had a lower one-year probation failure rate (26%) compared to the control group (34%). However, while they reported this as a statistically significant effect on the basis that it was significant at the 10% level, it did not meet more conventional significance thresholds ($p = 0.097$). Lowenkamp et al. (2014) extended the analysis to examine recidivism over a two-year period, aligning with the Canadian studies above. Due to data limitations on arrest records, their sample included only offenders on post-conviction supervision, whereas Robinson et al. (2012) had analysed a mixed sample of both post-conviction offenders and pretrial defendants. Using a quasi-experimental design, Lowenkamp et al. (2014) found that the positive effect of STARR training on probation failure rates persisted at two years, but the difference remained statistically insignificant.

Labrecque and Viglione (2021) explored the relationship between STARR training of probation officers and their clients' outcomes in the Middle District of Florida. The authors adopted a quasi-experimental study design, using propensity score matching to match clients of trained and untrained probation officers, as allocation to training was not random for this study. The main aim of their study was to assess whether a difference in training dosage affects the outcomes of the probation officers' clients. This specific element is discussed when answering research question 3a. However, they also presented findings on the general effect of STARR training versus no training, contributing to the overall question of the effectiveness of professional development interventions. Their outcome measures for probation officers' clients spanned a wider range than what is presented in the other papers. Next to new arrests and probation revocations, they also analysed whether STARR training of officers had any impact on their clients' likelihood of technical violations and positive drug tests. They found that clients of non-trained probation officers were less likely to be newly arrested or have their probation revoked compared to clients of STARR

coach-trained probation officers, but only the latter's impact is significant. Interestingly, clients of non-trained probation officers have a lower (not significant) likelihood of getting a technical violation or positive drug test. The authors argue that one reason could be that these two outcomes rely more exclusively on officer discretion, and working more intensely with clients would lead to a higher detection of these violations.

Asteris (2013) is the last paper that focuses on the training of probation officers. This RCT was conducted in Texas, and the intervention differed from the others, by focusing on the acquisition and retention of motivational interview (MI) skills of officers. The study had one control group and three treatment groups. Every participant, including the control group, attended a three-day basic MI course, while the three treatment groups received different levels of training beyond that. The varying effects of the different treatment groups will be discussed in research question 3, while here, we focus on the effect of the most intense treatment group (combination of MI learning and coaching/feedback) and the control group. Motivational interview skills were measured with two tools: the Motivational Interviewing Treatment Integrity 3.1.1, which uses trained raters to evaluate a practitioner's motivational interviewing skills through global ratings and behaviour counts in actual sessions, while the Video Assessment of Simulated Encounters – Revised assesses MI competence by scoring written responses to simulated video client encounters. A positive treatment effect was detected for both measures. However, the author did not test the significance of the difference between post-treatment means of control and combined treatment groups.

Computerised cognitive training helps older workers sustain productivity, but missing data prevents conclusions for tax employees

Milič Kavčič et al. (2024) conducted a randomised controlled trial evaluating computerised cognitive training (CCT) among case managers aged 50 and above at the Pension and Disability Insurance Institute of Slovenia. Given the accelerated aging of the workforce and concerns over cognitive decline impacting productivity, this study aimed to assess whether CCT could help mitigate such effects and preserve work performance among older employees. Participants were randomly assigned either to an intervention group, which completed 24 sessions of a virtual maze navigation CCT over 12 weeks, or to a passive waitlist control group. Outcomes were assessed on cognitive performance, wellbeing, and work productivity – the latter measured using the company's "ponder" metric, a composite score reflecting the volume of normed procedures completed over a set period. The authors find a positive but insignificant effect of the CCT on productivity. Interestingly, the training affected productivity not by increasing it but by maintaining it, as the productivity of participants in the control group dropped significantly. The authors suggest that this decline in the control group may have been driven by increased workplace stress and heavier workloads due to pension and disability legislative changes at the end of the analysed year. However, the intervention group was able to sustain their performance despite dedicating time to training.

Borness et al. (2013) also explored the effect of cognitive training on work productivity in the civil service context. They conducted a randomised controlled trial to assess the

real-world efficacy of CCT in the workplace, addressing its debated benefits for working-age adults. The study involved 135 white-collar employees from an Australian public sector organisation focused on tax, who were randomly assigned to either a 16-week online CT program (20 minutes, three times per week) or an active control (AC) group that watched nature documentaries of equal duration. Outcomes were measured across three time points – baseline, immediately after training, and six months post-training – focusing on cognition, wellbeing, and workplace productivity. Productivity was assessed using the organisation's routinely collected performance data, including task completion times, case processing speed, and quality ratings. The study found no significant short-term or long-term effects of CT on workplace productivity, with overall productivity outcomes being inconclusive due to high levels of missing data (61% at six-month follow-up). Surprisingly, participants in the AC group experienced significant improvements in self-reported wellbeing, including higher quality of life ($p = .003$) and lower stress levels ($p = .03$), suggesting that simple, low-effort respite activities might be more beneficial for workplace wellbeing than cognitive training.

Training in conversation and listening techniques enhances call centre productivity, knowledge spillovers, and employee competence

De Grip and Sauermann (2012) also examined the effect of a professional development intervention for operational delivery personnel on productivity. They conducted an RCT to examine the impact of work-related training on employee productivity, focusing on call centre agents in the private sector in the Netherlands. Participants were randomly assigned to either a training group, which undertook a one-week course on conversation techniques designed to reduce call handling times while maintaining service quality, or a control group that received no training during the study period. Productivity was measured using the call centre's key performance indicator: average handling time (AHT), representing the total time needed to complete a customer interaction, including logging the call in the system. The study found a significant positive difference in productivity between the treatment and control groups after the training. The authors suggested this effect is likely underestimated, as the researchers identified positive externalities from training, as a 10-percentage-point increase in the share of trained co-workers led to a 0.51% productivity boost among untrained peers, suggesting knowledge spillovers and peer learning effects. Importantly, the productivity improvements were not achieved at the expense of service quality, highlighting the effectiveness of targeted skill training in improving both individual and team performance.

Itzchakov (2020) carried out two quasi-experimental studies, one in Israel and one in the United States, to examine whether listening training can empower service employees by reducing anxiety, enhancing perspective-taking, and increasing their sense of competence in difficult customer interactions. In the United States, the training was conducted among customer service employees at a Fortune 500 company, where participants received an 18-hour listening training over two days. Results showed a significant increase in listening perception, a reduction in anxiety, and improvements in perspective-taking and sense of competence. Mediation analysis indicated that the reduction in anxiety led to improved perspective-taking, which in turn

enhanced employees' sense of competence. In Israel, a similar study with a control group was conducted among customer service employees in a nursing service company, with training delivered in two sessions using the Listening Circle Paradigm. Findings mirrored those in the U.S., showing improvements in listening perception, reduced anxiety, and increased perspective-taking, with mediation analysis confirming that anxiety reduction preceded improvements in perspective-taking. Together, these studies suggest that listening training can empower service employees by reducing anxiety and fostering perspective-taking, ultimately strengthening their competence in handling challenging conversations.

Leadership development programs play a positive role in improving employee-perceived transformational leadership and leader performance

Seidle et al. (2016) conducted a panel study to evaluate the impact of leadership training and development on leader and organisational performance in the United States public sector. The study focused on branch, division, and project managers within a large federal agency, comparing those who participated in a leadership training programme with a control group that did not. The intervention combined coaching, classroom instruction, multisource feedback, and experiential training, with participants undergoing a structured six-month development programme. Leader performance was measured using annual performance pay increases, while organisational effectiveness was assessed through quality and timeliness metrics, reflecting output standards and service delivery speed. The findings indicate that leadership training had a significant positive effect on both leader performance and organisational effectiveness. However, the study did not establish a direct causal link between leader performance and organisational effectiveness, highlighting the complexity of measuring leadership impact. The results suggest that well-designed leadership training programmes can yield meaningful benefits for public sector organisations, supporting their continued investment in leadership development initiatives.

Also analysing the effect of leadership training, Jacobsen et al. (2022) conducted a randomised field experiment in Denmark to assess the effects of leadership training on leadership behaviour. The study involved leaders from 463 public and private organisations, including schools, daycare centres, tax offices, and bank branches, who were randomly assigned to one of three training groups – transformational, transactional, or combined training – or a control group. The intervention consisted of a structured nine-month leadership training programme, and leadership behaviour was measured using employee surveys before and after training. For the overall effect, the study compared the control group with the most intensive treatment group (combined training) and found significant improvements in employee-perceived transformational leadership behaviour ($p < 0.01$), along with increased use of contingent pecuniary rewards. The study also examined whether effects varied across treatment groups, a discussion explored further in research question 3a. In addition, Jacobsen et al. (2022) explored the effects on organisational performance, but this analysis was limited to school principals, as reliable performance metrics (exam results and standardised test scores) were only available for schools. An additional and non-independent study by An et al. (2022) used the same data but focused on a

different measurement – the gap between leaders' self-perceptions and employees' perceptions of leadership behaviour. The study's point estimates suggested that combined training reduced this gap, but the effect was not statistically significant. However, there is some evidence that leadership training is more successful in reducing the gap in public organisations versus private sector settings.

The last paper discussing the effects of a professional development intervention on transformational leadership behaviour is Haunstrup and Jensen (2024). The authors conducted a randomised field experiment in Denmark to assess the impact of leadership training combined with just-in-time implementation nudges on transformational leadership behaviour. The study involved 226 public managers from Danish job centres, who were randomly assigned to one of two treatment groups or a control group. The intervention consisted of transformational leadership training combined with either a static or a dynamic software tool designed to help managers implement training content into daily leadership practices. Employee-perceived transformational leadership behaviour was measured using pre- and post-training surveys from the leaders' employees. Comparing the static tool group to the control group, the study finds a significant improvement in transformational leadership behaviour. The study also examines whether effects differ between the static and dynamic tools, a discussion explored further in research question 3a.

Career management training enhances employee confidence and long-term intrinsic motivation

Salmela-Aro et al. (2012) conducted an RCT in Finland to examine whether career management training enhances career preparedness and intrinsic work-goal motivation. The study included 718 employees from 17 organisations, spanning municipalities, universities, federal government agencies, service sector companies, and technology firms. Participants were randomly assigned to an intervention group (career management workshops) or a control group (printed materials only). The study measured career preparedness (including career self-efficacy and preparedness for setbacks) and intrinsic work-goal motivation using self-reported questionnaires before the intervention, immediately after, and again at a seven-month follow-up. Results showed that the intervention significantly increased career preparedness, and this improvement mediated a long-term increase in intrinsic work-goal motivation. The findings suggest that career management interventions can enhance employees' confidence in managing their careers, leading to sustained improvements in their intrinsic motivation at work.

Online training in behavioural economics improves policymakers' problem-solving skills

Rojas Méndez and Scartascini (2024) delivered a randomised experiment to evaluate the impact of behavioural economics training on policymakers' decision-making and problem-solving skills. The study involved approximately 25,000 participants enrolled in an online course provided by the Inter-American Development Bank (IDB), targeting policymakers across Latin America and the Caribbean. Participants were randomly assigned to either a treatment group, which completed the training before taking a decision-making and problem-solving test, or a control group, which took the test

before undergoing the course. The study assessed decision-making performance through a survey test that included problem-solving tasks and public policy decision-making scenarios. Results showed that the treatment group performed significantly better than the control group. These effects were consistent across multiple problem-solving tasks, suggesting that behavioural economics training can enhance policymakers' ability to make more rational and effective policy decisions. The findings highlight the potential of online training programmes to mitigate cognitive biases and improve decision-making in public policy contexts.

Implementing standards through targeted training: Peer coaching improves evaluation practices in education, while regulatory training reinforces awareness of roles and responsibilities

Cotabish and Robinson (2012) carried out a randomised field study in the United States to examine the impact of peer coaching on gifted program administrators' evaluation knowledge. The study involved 200 administrators from a southwestern U.S. state, randomly assigned to a peer coaching intervention or a control group. The training focused on implementing the National Association for Gifted Children (NAGC) Program Standards, particularly for evaluating and expanding access to culturally diverse and low-income gifted learners. Outcomes were measured through content assessments, program evaluation rubrics, and concerns questionnaires, alongside student participation data. Results showed that peer coaching significantly improved evaluation knowledge and skills, with administrators demonstrating stronger planning, goal setting, and alignment with NAGC standards. They also reported fewer concerns about implementing evaluations. While the intervention did not increase student placements, it positively influenced referral rates. The findings suggest peer coaching is an effective tool for strengthening evaluation practices in gifted education.

Dunlop et al. (2015) undertook a quasi-experimental study in the United Kingdom to assess the impact of training on local government inspectors' understanding of regulatory reform. The study focused on Primary Authority (PA) partnerships, introduced by the UK’s Better Regulation Delivery Office (BRDO) in 2009 to streamline business regulation. Local authority inspectors either self-selected into training or remained untrained, allowing for a comparison between the two groups. The intervention consisted of BRDO-led training workshops aimed at improving inspectors' knowledge of PA partnerships and their role in regulatory enforcement. The study measured inspectors' understanding of PA and their professional identity using survey questionnaires and scenario-based vignettes. Findings showed that trained and untrained participants did not differ in their understanding and knowledge of the PA initiative, suggesting that the policy was already well-integrated. However, training had a small but significant effect on reinforcing awareness of regulatory roles and risk-sharing responsibilities, particularly in contentious areas of reform.

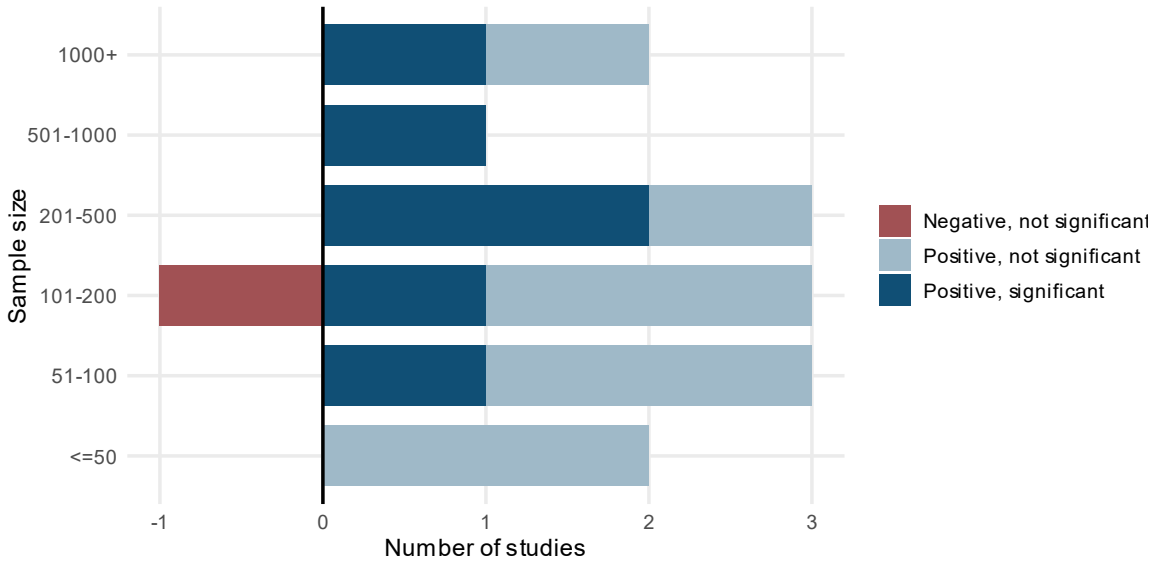
Research question 2a

Does the effectiveness of professional development interventions in improving knowledge, skills, networks, work performance, and productivity vary based on study characteristics (features of the evaluation not specific to the intervention itself)?

To address this research question, we present effect direction plots grouped by sample size and risk of bias assessment scores. We do this for each of the independent studies contributing to the question of the effectiveness of the overall effect of professional development, and where the necessary information was reported.

Figure 2 shows the effect direction plot and also indicates whether each of the results was statistically significant ($p \leq 0.05$) or not. The plot shows the wide heterogeneity in the overall sample sizes of the studies, varying from 33 to 5,655. Only two of the studies had sample sizes higher than 1,000, which is unsurprising, given the costs and logistical challenges involved in implementing professional development interventions of that size. Neither study with a sample size below 50 finds statistically significant effects. It should be noted that with such small sample sizes, standard errors tend to be large, and hence only large estimated effect sizes could be statistically significant. These are the Itzchakov (2020) listening training for customer service employees in a nursing service company in Israel, and the Milič Kavčič et al. (2024) paper that analyses computerised cognitive training among older case managers at the Pension and Disability Insurance Institute of Slovenia. Beyond this, there is no clear systematic pattern (based on this admittedly small sample of papers) linking sample size to statistical significance. Labrecque and Viglione (2021), one of the two studies with the sample size exceeding 1,000, which analysed the impact of STARR training of probation officers on 1,444 of their clients' outcomes in the middle district of Florida, did not find a statistically significant effect on the reduction of new arrests.

Figure 2. Effect direction plot grouped by sample size



Figures 3 and 4 present the effect direction plots grouped by the risk of bias assessment scores. As discussed in the Risk of bias section, we used different assessment tools for the quasi-experimental and RCT study designs. They are not directly comparable, and hence we present the effect direction plots separately.

Grouping the results by risk of bias assessment scores helps us understand whether studies with a higher risk of bias are skewing the overall results and if the findings hold true for the more reliable studies. This analysis ensures we're not drawing conclusions based on less trustworthy research, giving us more confidence in what we are reporting. While the sample is small, it is still important to check whether any patterns emerge.

Figure 3 shows the results for studies with RCT designs. The risk of bias assessment scores for the RCT studies contributing to the findings of the overall effect of professional development on the targeted outcomes varied from 23% to 54%. There is no obvious pattern visible between the risk of bias assessment score and the direction and significance of the effects. Each bucket of scores contained at least one significantly positive result.

Figure 3. Effect direction plot by risk of bias assessment scores for RCT designs

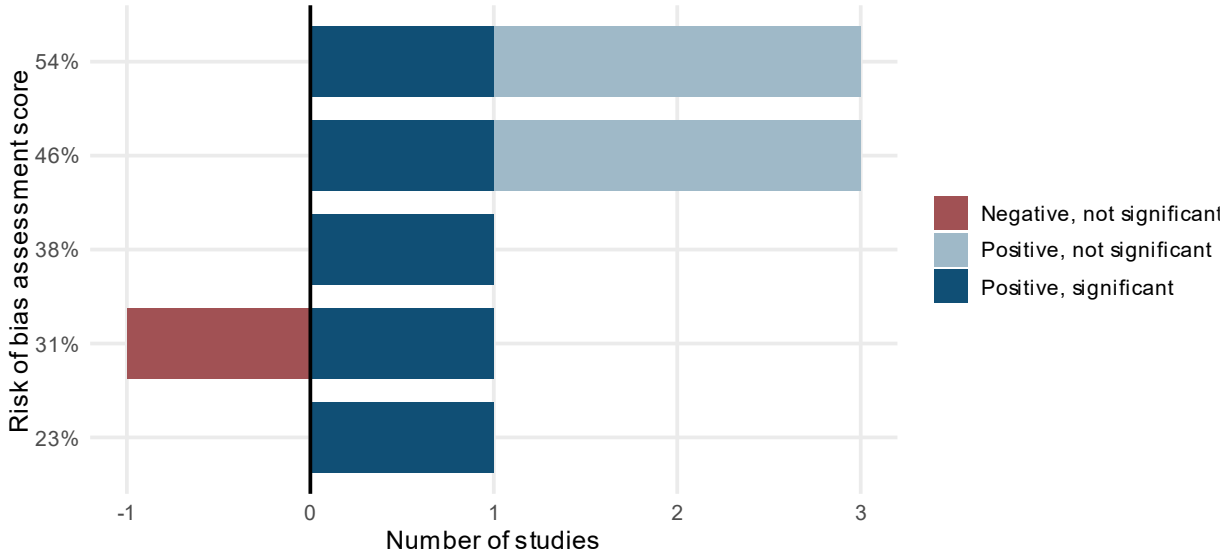
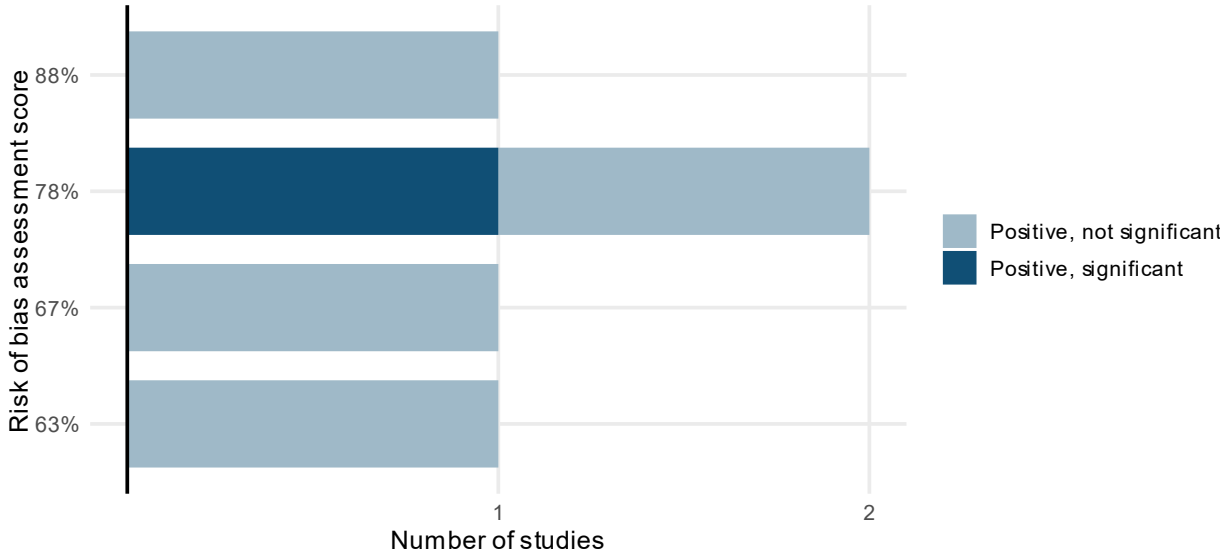


Figure 4 shows the same plot for the quasi-experimental independent studies contributing to the findings of the overall effect. Again, there is no obvious link between the effect direction and the significance of the results and the risk of bias assessment scores, which range between 63% and 88%.

Figure 4. Effect direction plot by risk of bias assessment scores for quasi-experimental designs



Research question 2b

Does the effectiveness of the interventions vary based on the types of workers and the target outcomes?

To address this research question, we present effect direction plots grouped by target outcome and types of workers. We do this for each of the independent studies contributing to the question of the effectiveness of the overall effect of professional development, and where the necessary information was reported. In addition, a few of the papers identified specifically explore how training outcomes vary based on, e.g., the type of worker. We will describe the findings of those papers in this section.

Figure 5 displays the effect direction plot by the main target outcome for the 15 studies that contribute to the overall effect research question, and where the p-value is not missing. Most of these studies (7) focused on outcomes related to skills, four on work performance, and two each on productivity and knowledge. No studies that had networks as a target outcome were found in the systematic review. Nearly all studies that identified statistically significant effects were focused on skills. This is reminiscent of the pattern described for the studies focused on probation officer training, for which the immediate effect on use of targeted skills had a strong evidence base, but it was more difficult to demonstrate that this then led to changes in recidivism rates. We can conclude that the evidence base for positive impacts of professional learning programmes on performance and productivity is weaker than for skills.

Figure 5. Effect direction plot by target outcome

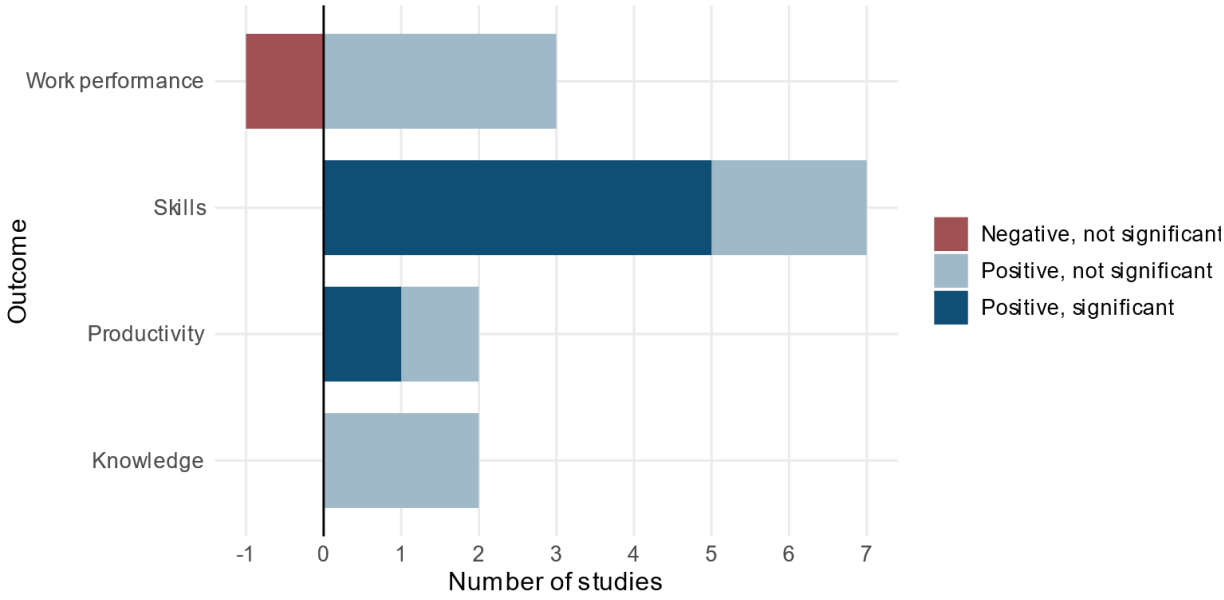
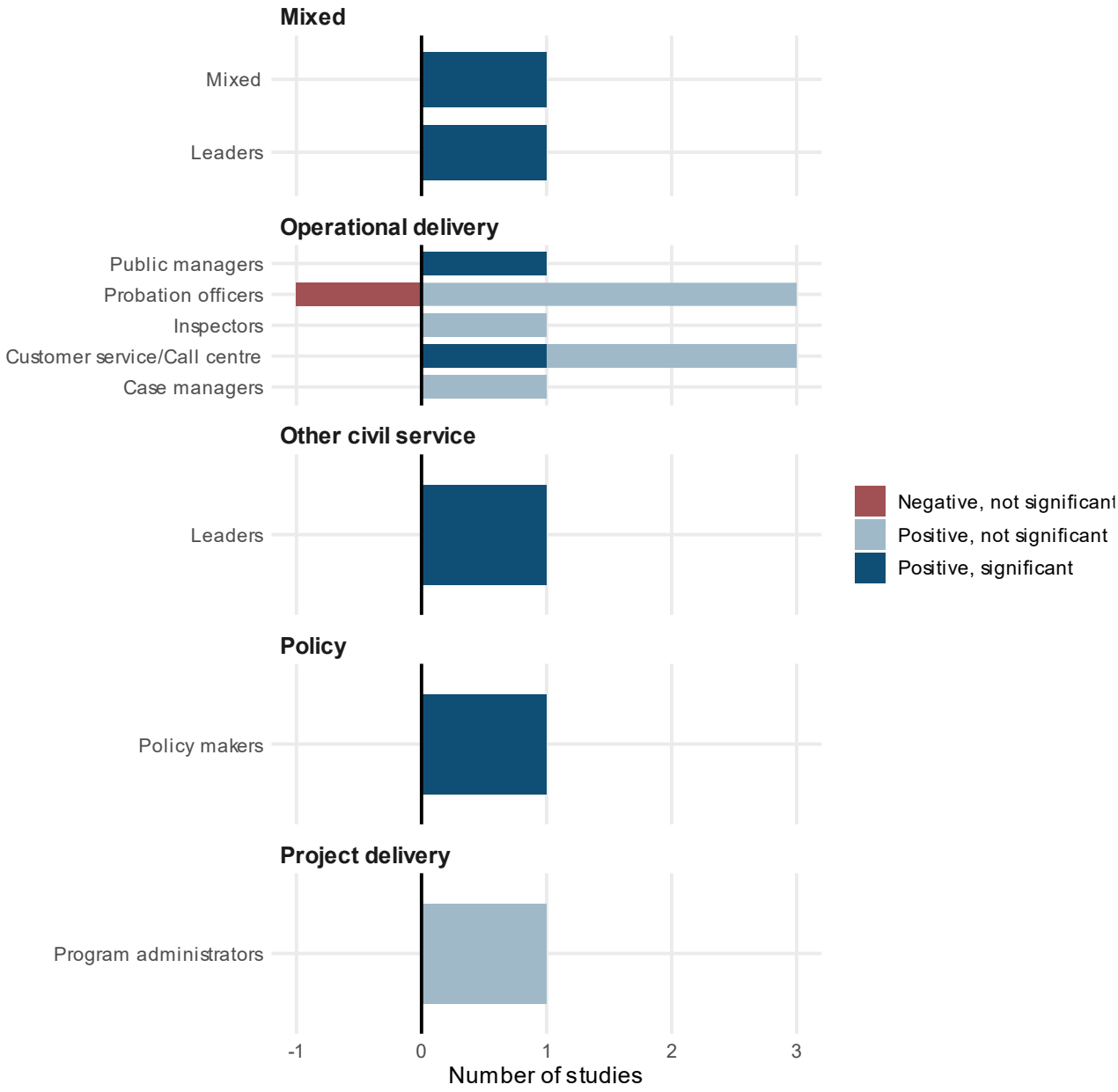


Figure 6 shows the effect direction plot for different types of workers. The studies are grouped based on the civil service professions they belong to and the specific roles the workers experiencing professional development do. Most studies' target populations belong to the operational delivery profession, such as probation officers and customer service agents. Due to the high degree of heterogeneity among the studies, drawing strong conclusions is challenging. However, the variation in effect direction and significance within the operational delivery profession likely reflects the previously discussed distinction between skills and knowledge versus productivity and work performance. The emphasis on productivity and performance in studies on operational delivery roles stems from the fact that these aspects are more measurable in this profession, unlike other civil service roles, where evaluating work performance and especially productivity is more complex.

Figure 6. Effect direction plot grouped by type of worker



We also identified three papers that specifically analysed how the effect of training differs depending on the type of worker.

Transformational leadership training is more effective for women after accounting for prior leadership levels, but shows no significant difference between public and private sector managers

Two of the papers use data from the same study on Danish leaders as Jacobsen et al. (2022) and An et al. (2022) and use the same outcomes measured as in Jacobsen et al. (2022): transformational leadership behaviour, contingent verbal rewards, and contingent pecuniary rewards based on survey data from the leaders' employees before and after training. An and Meier (2021) investigate whether the effect of

leadership training is significantly different based on the gender of the participants. The study finds that women exhibit stronger transformational leadership behaviours than men, even before training. While both transformational and combined leadership training improve leadership behaviours, initial models show no significant gender difference. However, when controlling for pre-existing leadership and absences, women benefit significantly more. For transactional leadership, women initially show greater gains, but after adjusting for prior leadership levels, gender differences become insignificant, suggesting that only transformational leadership training is distinctly more effective for women. An et al. (2019) explored whether the effectiveness of training in transformational and transactional leadership differed between workers coming from the public and the private sectors. The population of the study were Danish leaders from a mixed background, including managers in taxation, education (both schools and daycare centres), and banking. They found a positive but not statistically significant correlation between combined leadership training and perceived transformational leadership behaviour in both the private and public sectors, with the public sector showing a slightly larger, albeit non-significant effect.

Employer-sponsored training has a greater impact on call centre employees who feel reciprocal loyalty

Lastly, Sauermann (2021) used an RCT design based on the same data as De Grip and Sauermann (2012) to examine whether the impact of employee training depends on reciprocal attitudes – specifically, the affinity of certain employees to reciprocate their employer's investment. The study focused on call centre employees in the private sector in the Netherlands, selecting a sample to undergo employee-sponsored training designed to enhance efficiency by reducing average call handling time. Employee performance served as the key measure of training effectiveness. The study found a greater and statistically significant treatment effect among employees with reciprocal attitudes, suggesting that individuals with reciprocal attitudes respond to the training by increasing their efforts, potentially to reciprocate their employer's investment.

Overall, the findings indicate that the effectiveness of professional development interventions varies by worker type and target outcome. While some patterns emerge, such as stronger evidence for skill development than productivity or performance gains, the small number of studies and high heterogeneity make firm conclusions difficult.

Research question 3a

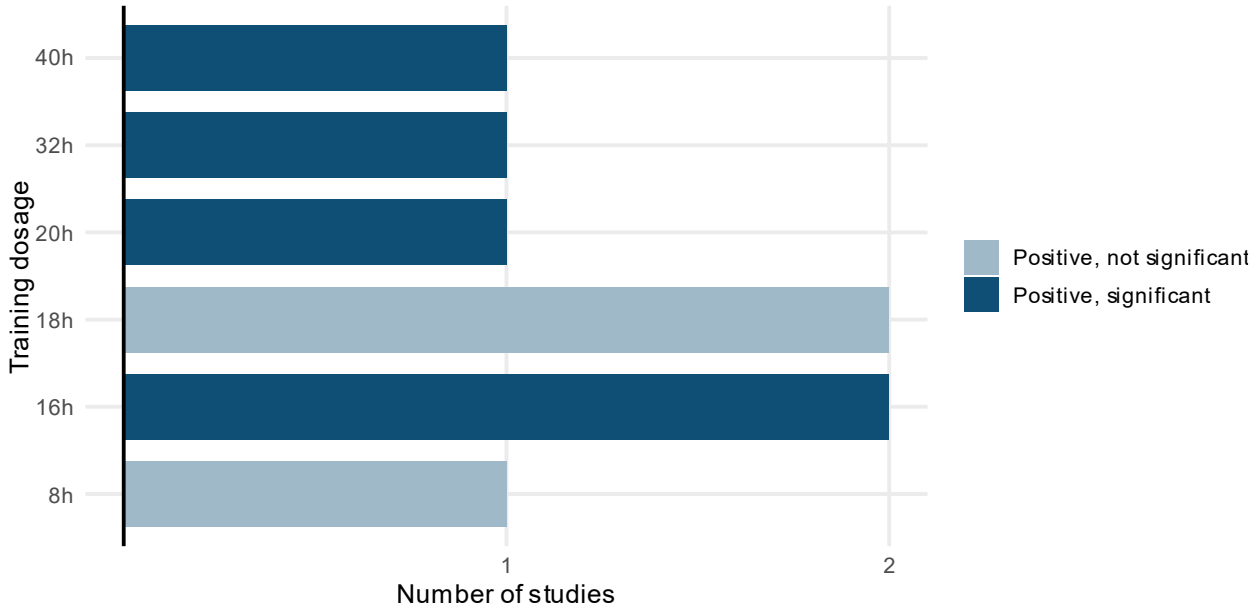
Which design features (e.g., online versus face-to-face; longer duration versus shorter duration) are associated in the literature with the greatest impact on skills, knowledge, networks, work performance, and productivity?

To address this research question, we present effect direction plots grouped by dosage of training, type of engagement, and type of professional development. We do this for each of the independent studies contributing to the question of the effectiveness of the overall effect of professional development, and where the necessary information was reported. In addition, some of the papers identified

specifically explore how the effectiveness of professional development interventions varies based on design features. We will describe the findings of those papers in this section.

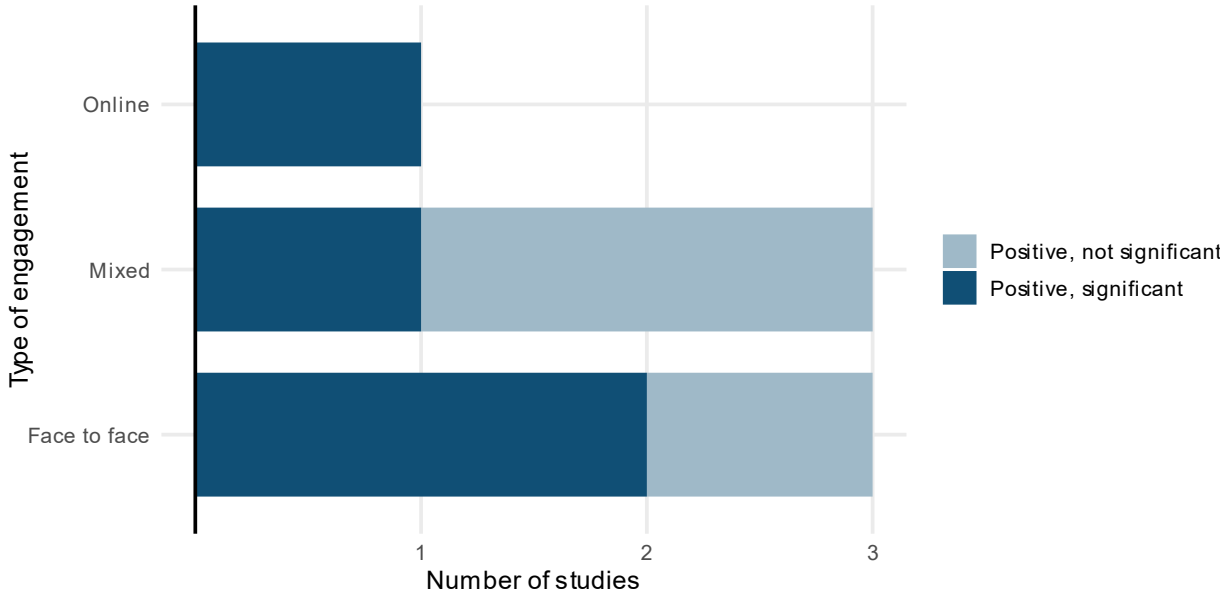
Figure 7 displays the effect direction plot by dosage of training received. The dosage captures the total time that participants were actively engaged in a professional development intervention. For this purpose, we assumed that one day equals eight hours and one week 40 hours. Only six of the studies contributing independently to the overall research question reported the dosage of their intervention. This makes it difficult to draw firm conclusions, although it is notable that the three interventions with the longest dosages all produced statistically significant results.

Figure 7. Effect direction plot by training dosage



The effect direction plot for the type of engagement is presented in Figure 8. Seven of the relevant studies reported on whether their professional development intervention was conducted online, face-to-face, or a mix of the two types of engagement. There is no obvious pattern visible between the type of engagement and the effect direction or the statistical significance of the results.

Figure 8. Effect direction plot by type of engagement

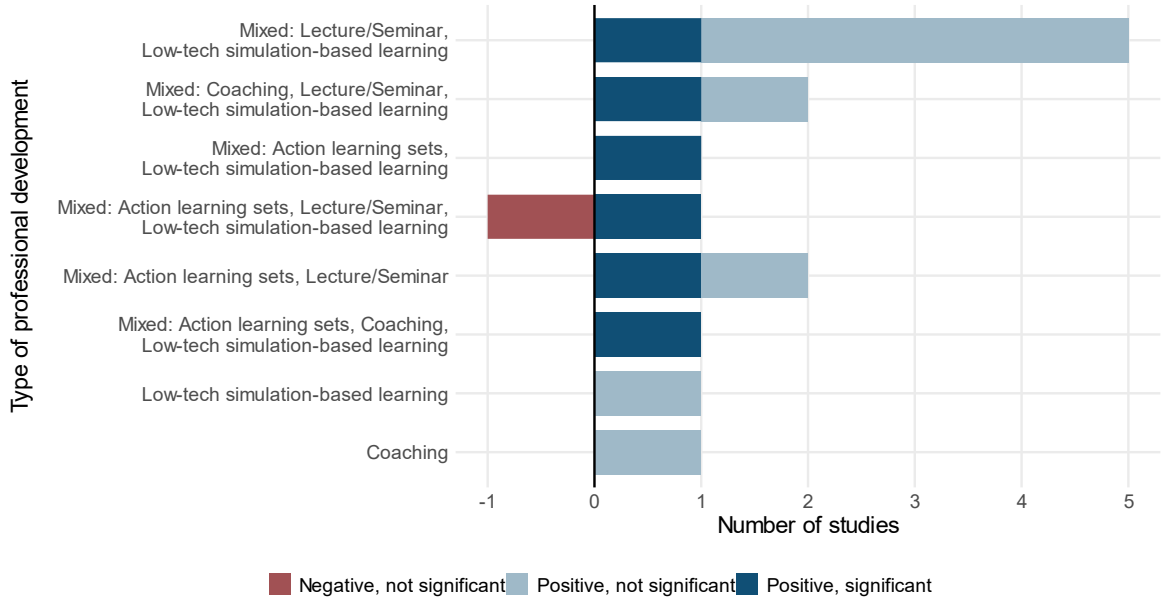


Next, we look at the effect direction and statistical significance by type of professional development. Professional development encompasses various learning methods designed to enhance skills and knowledge in the workplace: (i) Lecture/Seminar sessions are traditional instructor-led formats where information is presented to a group, often followed by discussions or Q&A to reinforce understanding, (ii) coaching, on the other hand, offers a personalised approach, providing one-on-one or small-group guidance to help individuals refine specific skills and achieve their goals with tailored feedback, (iii) high-tech simulation-based learning leverages advanced technology, such as virtual reality and computer simulations, to create immersive, hands-on training environments, (iv) while low-tech simulation-based learning uses simpler methods, like role-playing or physical props, to provide realistic practice experiences. Another approach is (v) action learning sets, where small groups collaborate to address real-world challenges, reflecting on their actions and learning through shared experiences.[3]

Figure 9 shows that a large majority of the relevant studies applied more than one type of professional development in their intervention, with only two studies focusing on only one type: coaching and low-tech simulation-based learning. The combination that was used most frequently was a mix of Lecture/Seminar and low-tech simulation-based learning. All studies that produce statistically significant results use more than one type of professional development.

[3] This typology was defined in advance, as outlined in the Data items section, to categorise common professional development methods. An "Other (In vivo)" category was included to capture any interventions that did not fit the predefined classifications, ensuring flexibility in the analysis.

Figure 9. Effect direction plot by type of professional development



We identified eight papers that address this research question by analysing how training differs depending on design features of the professional development intervention.

One-on-one peer coaching strengthens programme evaluation skills beyond standard training

While we already discussed the overall effect of the program evaluation training for program administrators, analysed in Cotabish and Robinson (2012) for research question 2, their findings are also relevant for this research question, as the control group also received some training. All participants, both in the experimental and control groups, initially attended two one-day professional development institutes focused on programme evaluation. These institutes covered key concepts, such as defining evaluation, setting goals, identifying areas for improvement, and aligning evaluations with the National Association for Gifted Children (NAGC) Program Standards. However, only the experimental group received additional, intensive support through one-on-one peer coaching, while the control group received no further training. Peer coaching provided tailored guidance in applying these principles, including setting evaluation goals, refining assessment tools, and addressing programme weaknesses. In comparing minimal professional development (the institutes alone) with a more intensive approach (institutes plus peer coaching), the study found that administrators who received peer coaching demonstrated significantly greater knowledge of programme evaluation, particularly in goal setting, planning, and linking evaluation questions to programme standards.

Workshops are more effective than passive information sharing for career preparedness

Similarly, the findings for the career preparedness intervention in Salmela-Aro et al. (2012) also have some relevance for understanding the role of intervention design,

since even the control group received some treatment. All participants were randomly assigned to either an intervention or a control group. Those in the intervention group attended a series of five half-day workshops, designed to enhance career preparedness and intrinsic work-goal motivation through interactive learning, goal setting exercises, and career management training. In contrast, the control group received only printed materials covering similar content. While both groups were exposed to career management information, the study found that interactive workshops led to significantly greater improvements in career preparedness, career goal setting confidence, and intrinsic work-goal motivation compared to passive information sharing alone.

Combining training approaches yields stronger effects across several studies

As discussed for research question 2, Jacobsen et al. (2022) found that combined leadership training, incorporating both transformational and transactional elements, significantly improved transformational leadership behaviour compared to no training. Here we will discuss whether the authors find that different training types – transformational, transactional, or combined – had varying effects on leadership behaviour. The study assigned leaders to three groups. Transformational training focused on developing and communicating a vision, while transactional training emphasised contingent rewards to align employee behaviour with organisational goals. Combined training integrated both approaches to maximise their effects. The findings show that combined training had the strongest and most consistent positive impact, with effects significant at the five per cent level. Transformational training alone also showed positive effects, but was statistically significant only in a balanced panel analysis at the ten per cent level, indicating a weaker impact. Transactional training improved the use of contingent rewards but did not significantly enhance transformational leadership behaviour. Overall, the results indicate that combining transformational and transactional leadership training is more effective than either approach alone, reinforcing the idea that leadership behaviours are complementary and best developed together.

In addition to assessing the impact of STARR training on probation officers on their clients' re-arrest rates after 12 months and 24 months, the study by Lowenkamp et al. (2014) also provides insight into how different treatment groups responded to the intervention. Probation officers either received no training, STARR training only, or STARR training combined with motivational interviewing (MI). The STARR-only group received training in core correctional practices aimed at improving officer-client interactions, while the STARR + MI group received additional training in motivational interviewing techniques designed to enhance client engagement and behavioural change. The findings indicated that STARR training alone had a positive impact on moderate-risk offenders, reducing their 24-month re-arrest rate by 13% compared to the control group. However, for high-risk offenders, STARR training alone did not produce a significant reduction in re-arrest rates. In contrast, high-risk offenders supervised by officers trained in both STARR and MI showed significantly lower re-arrest rates at 12 months and a notable, though smaller, reduction at 24 months. This suggests that while STARR training can improve outcomes for moderate-risk

offenders, high-risk individuals may require additional interventions, such as MI, to see meaningful reductions in recidivism.

While the overall effect of the training on probation officers' acquisition and retention of motivational interviewing (MI) skills has been discussed in research question 2, the study by Asteris (2013) is also relevant here due to its comparison of different treatment groups. All participants, including the control group, attended a three-day basic MI course. Beyond this, the treatment groups received varying levels of additional support: one group participated in MI learning groups, which involved regular meetings to discuss MI principles, review recorded sessions, and engage in structured practice; another received one-on-one coaching and feedback, where trained coaches provided individualised guidance based on recorded officer-client interactions; and a third group received both interventions, combining peer learning with personalised coaching. The findings showed that while the learning groups and coaching interventions alone led to some skill improvement, neither was sufficient for significant gains. Only the combination of both produced meaningful improvements in MI competence, suggesting that interactive, sustained training methods are necessary for skill development, while standalone workshops or feedback may not be enough to ensure long-term proficiency.

Labrecque and Viglione (2021) explored the impact of different intensities of the Staff Training Aimed at Reducing Rearrest (STARR) training on probation officers and their clients' offence rates. The general effect of the training on officers' skills and clients' outcomes is discussed in research question 2, but the authors also explored whether effects varied based on training intensity. Their study categorised probation officers into three groups: untrained officers (control), STARR users (first treatment group), and STARR coaches (second treatment group). STARR users received standard training in the program, while STARR coaches underwent additional advanced training to mentor and support the STARR users. The results indicated that clients of both STARR-trained users and coaches were significantly less likely to have their probation revoked compared to those supervised by untrained officers. However, differences emerged in other outcomes. Clients of STARR users were more likely to have a positive drug test and commit a technical violation than those supervised by untrained officers, with these increases reaching statistical significance. In contrast, while clients of STARR coaches also showed slight increases in these outcomes, the differences were not significant. The authors suggest that one explanation for this pattern is that STARR coaches, having undergone additional training, may be better at handling minor infractions informally, reducing the need for formal documentation of violations. In contrast, STARR users, with less advanced training, may be more likely to formally record these incidents.

Mandating skill use increases STARR application but favours simpler techniques

In a related study, Viglione and Labrecque (2021) investigated the impact of a policy mandating probation officers to use at least eight STARR skills per month to address low skill application following training. The mandate required officers to log their skill use, reinforcing the district's commitment to evidence-based supervision practices.

Using a pre-post study design, they found a significant increase in skill use, with the average number rising from 4.6 to 13.3 skills per officer per month. The proportion of officers who met the requirement of using at least eight skills per month increased significantly, from only 14.6% before the mandate to 89.6% after. However, simpler techniques like role clarification and reinforcement were used more frequently than more advanced cognitive-behavioural strategies. The study concluded that while policy mandates can successfully boost skill use, they do not guarantee quality or consistency.

There was no significant difference in using static or dynamic implementation tools to increase transformational leadership behaviour

Haunstrup and Jensen (2024) investigated the effect of transformational leadership training coupled with different types of implementation nudges on transformational leadership behaviour among public managers in job centres in Denmark. We already discussed the general effect of the intervention under research question 2. They investigated the impact of transformational leadership training combined with two different types of implementation nudges – static and dynamic software tools – on transformational leadership behaviour. As part of that, they also assessed whether the effectiveness of training depended on the format of these post-training nudges. The static software tool relied on existing systems, such as PDFs, email, and calendar reminders, requiring managers to manually engage with and apply the materials. The dynamic software tool was a more interactive system, providing automated prompts, reminders, and structured guidance to integrate leadership behaviours into daily routines. However, the authors find no statistically significant differences in the effect of transformational leadership behaviour depending on the type of implementation nudge used. The paper unfortunately did not analyse whether the effect of leadership training without an implementation nudge differed from leadership with the implementation nudge, as the control group received no training.

For research question 3a, the findings suggest that design features such as training dosage, engagement type, and professional development format influence effectiveness, with longer interventions and combined training approaches generally yielding stronger results. However, patterns are not always consistent, and some studies found no significant differences based on specific design variations, such as online versus face-to-face delivery or different post-training support tools.

Research question 3b

As mentioned in the Development of Taxonomy section, we adapted the BCT taxonomy to establish a coding framework that captures specific mechanisms or ‘active ingredients’ that would make an intervention less effective if they were to be removed (Sims et al., 2021). These mechanisms are presented in our taxonomy (available in Appendix C).

Which mechanisms and forms (clusters of mechanisms) do we observe in the literature?

Figure 10 depicts the distribution of the number of mechanisms identified for each included intervention. We observe that the majority of interventions incorporated more than four mechanisms as part of their training programmes. The highest number of mechanisms recorded per paper was in the three Danish studies from An et al. (2019, 2021, 2022), which included analysis of the same underlying intervention.

Figure 10. Number of mechanisms identified for each included paper

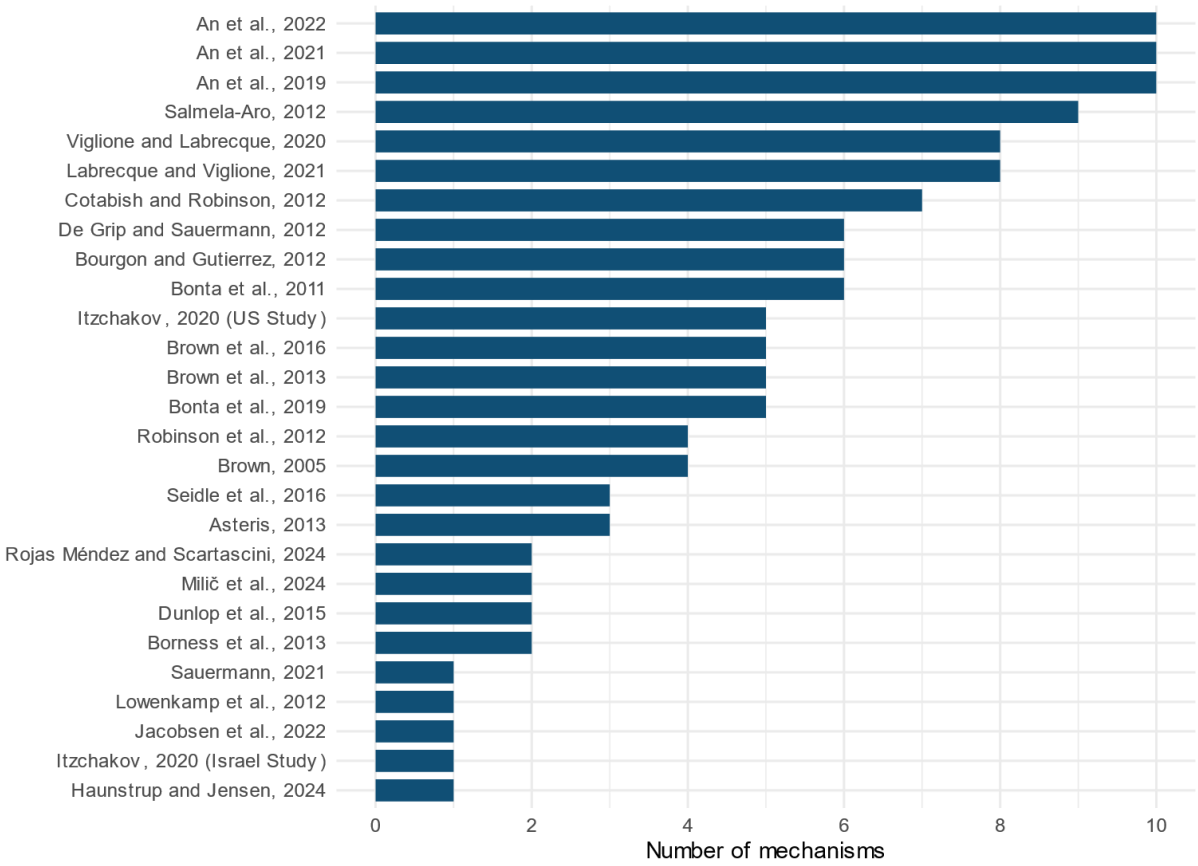


Figure 11 presents the ten most common mechanisms identified in the included papers. The most common mechanism used by the included papers is behavioural practice/rehearsal (identified in 18 studies). This mechanism encompasses the repeated practice or rehearsal of the behaviour in situations where it is not immediately required, to strengthen habit formation and improve skill development. For example, as part of a listening training programme, attendees may actively engage in practising listening techniques to improve their understanding and attention to customers. By challenging common misconceptions about listening (e.g., being automatic or effortless), participants are encouraged to repeatedly apply new strategies during the training session, reinforcing skill development and habit formation.

The second most common mechanism identified was instruction on how to perform a given behaviour. The prevalence of this mechanism is anticipated, given the focus of this review on training programmes. In practice, this mechanism covers behavioural

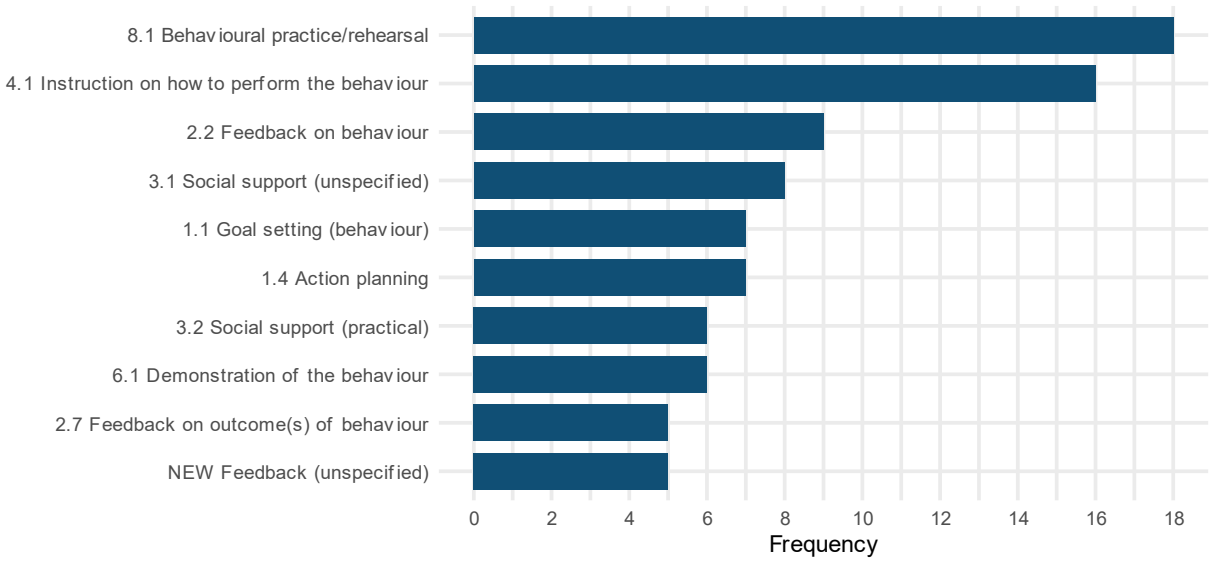
change, which is facilitated through the provision of guidance on how to perform the desired behaviour. For example, as part of a professional development workshop, a facilitator may provide step-by-step guidance on how to conduct effective performance appraisals, including techniques for delivering constructive feedback. For skills-related outcomes, instruction often involved step-by-step guidance on performing specific behaviours required to acquire these skills. For example, leadership development programmes guided participants on how to deliver contingent verbal rewards to employees.

The third most prevalent mechanism for behaviour change identified was feedback on behaviour. This includes the tracking and provision of informative or evaluative feedback on the execution of the behaviour (e.g., form, frequency, duration, intensity). For example, facilitators may observe participants in mock leadership meetings and provide feedback on their communication clarity, engagement frequency, and areas for improvement.

The fourth most commonly identified mechanism was general social support, where it was unclear whether the support was practical in nature or derived from mutual engagement and shared responsibility. In the included studies, for instance, this involved participants sharing their goals with peers who were also setting goals, with the peers then providing feedback on those goals.

The fifth most common mechanism identified was goal setting, where the goal was clearly defined in terms of the specific behaviour to be achieved. For example, participants may be prompted to develop a set of goals on the specific initiatives to implement in their team or unit, particularly regarding how they perform goal-oriented development dialogues with their employees.

Figure 11. Most common mechanisms identified in included papers



It is also interesting to know whether the mechanisms used vary depending on the type of workers. As 15 out of the 27 included studies analyse interventions targeting workers from the operational delivery profession, and the rest are dispersed across

the other professions, we focus the analysis on whether the mechanisms frequently used for interventions for the operational delivery populations vary from those used in the other studies. Figure 12 presents the results of that analysis. Behavioural practice/rehearsal and Instruction on how to perform the behaviour are still the most common mechanisms across both groups. However, after that, the most frequently used mechanisms vary. Within the interventions targeting the operational delivery professions, there is a much stronger focus on social support, with practical collaboration and unspecified social support all being among the most common mechanisms. In addition, habit formation is commonly used in interventions targeting operational delivery workers. For the other professions, three mechanisms around feedback are among the most common: feedback on behavior, unspecified feedback, and feedback on outcomes of behaviour. Action planning is also one of the three most common mechanisms outside of operational delivery, despite not being among the most used mechanisms within operational delivery.

Figure 12. Most common mechanisms in papers targeting operational delivery profession vs others

Given the small number of papers and large number of mechanisms, the identification of any forms (clusters/combinations of mechanisms) is difficult. To identify any combinations that appear multiple times, we first aggregate the mechanisms to the BCT groups, as described in the Taxonomy section. Table 3 shows which clusters of mechanism groups appear multiple times across all 27 studies. There are six combinations that appear multiple times, but three of them are based on a single mechanism group only. The three papers by An et al. are not independent and are

based on the same professional development intervention. They all use a combination of mechanisms across (i) Goals and Planning, (ii) Feedback and Monitoring, (iii) Shaping Knowledge, (iv) Associations, (v) Repetition and Substitution, and (vi) Identity. The Israel study of Itzchakov (2020) that analyses listening training of customer service agents, Milič Kavčič et al. (2024) who explore an intervention for Slovenian case managers, and Borness et al. (2013) who estimate the impact of cognitive training on Australian tax personnel all only use one mechanism group, Repetition and Substitution. Bonta et al. (2011) and Bourgon and Gutierrez (2012) are based on the same study on probation officer training, so it is unsurprising that the intervention described uses the same mechanism groups: (i) Feedback and Monitoring, (ii) Social support, (iii) Shaping knowledge, (iv) Comparison of Behaviour, and (v) Repetition and Substitution. Both Rojas Méndez and Scartascini (2024) and Dunlop et al. (2015) analyse an intervention that uses a mix of mechanisms from the (i) Shaping Knowledge and (ii) Repetition and substitution groups. Lastly, there are two sets of papers that only use mechanisms from the Goals and Planning group and two that only use mechanisms from the Social Support group.

Table 3. Overview of forms (combinations of mechanism groups) that appear multiple times

Count of forms	References	Mechanism groups
3	An et al., 2019; An and Meier, 2021; An et al., 2022	1 Goals and Planning, 2 Feedback and Monitoring, 4 Shaping Knowledge, 7 Associations, 8 Repetition and Substitution, 13 Identity
3	Itzchakov, 2020 (Israel study); Milič Kavčič et al., 2024; Borness et al., 2013	8 Repetition and Substitution
2	Bonta et al., 2011; Bourgon and Gutierrez, 2012	2 Feedback and Monitoring, 3 Social Support, 4 Shaping knowledge, 6 Comparison of Behaviour, 8 Repetition and Substitution
2	Rojas Méndez and Scartascini, 2024; Dunlop et al., 2015	4 Shaping knowledge, 8 Repetition and substitution
2	Brown et al., 2016; Brown et al., 2013	1 Goals and Planning
2	Lowenkamp et al., 2014; Sauermann, 2021	3 Social Support

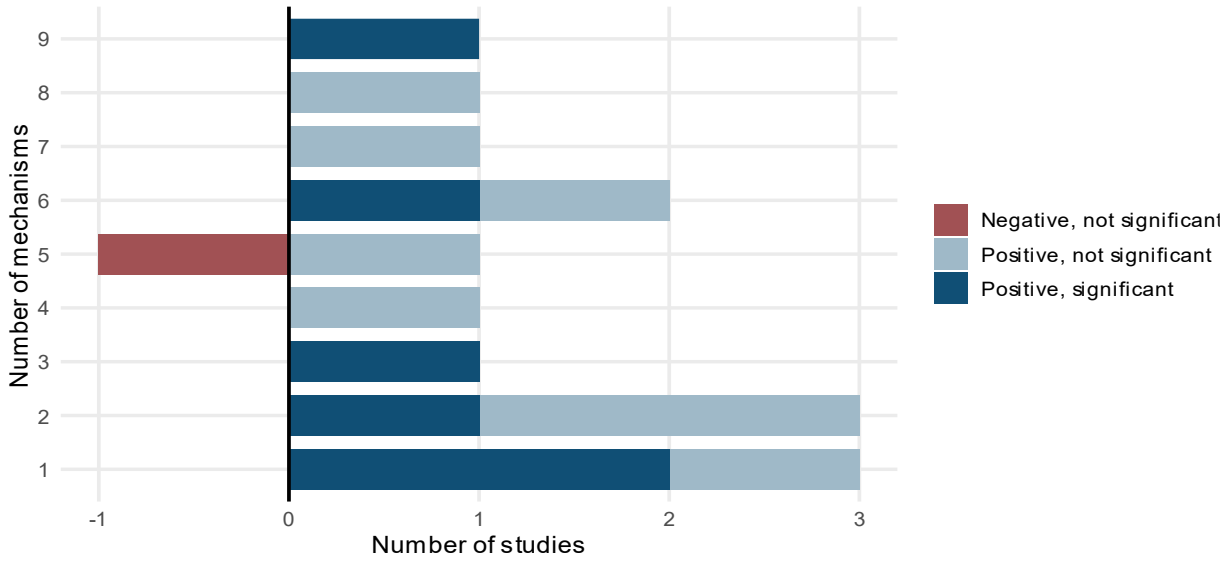
Research question 3c

Which mechanisms or forms (combinations of mechanisms) are associated in the literature with the greatest impact on skills, knowledge, networks, work performance, and productivity?

Given the small number of papers included in the review and the many combinations of mechanisms, we are severely limited in how we are able to address this research question. We do, however, present the effect direction plots grouped by the *number* of mechanisms. In addition, a few of the papers identified specifically explore how the effectiveness of professional development interventions varies based on the mechanisms used. We will describe the findings of those papers in this section.

Figure 13 shows the effect direction plot grouped by the number of mechanisms used in the studies. There is no obvious link visible between the number of mechanisms and the effect direction and statistical significance of the results. Both studies with only one to three mechanisms used in the interventions and those with a maximum of nine found positive and significant results. The one study that found a non-significant negative effect had five mechanisms.

Figure 13. Effect direction plot grouped by number of mechanisms



Three similar but independent studies by Brown et al. present findings related to how the effectiveness of a professional development intervention varies depending on the mechanisms used.

Structured goal setting after training does not necessarily lead to better outcomes than a do-your-best approach

Brown (2005) investigated the impact of goal setting after training on various outcomes, including self-efficacy, in a quasi-experimental study conducted in Canada with federal and provincial government employees. All participants attended a one-day self-awareness training programme before being assigned to one of three goal setting conditions. The "do your best" (DYB) group was encouraged to apply the training content as best they could without setting specific goals. The distal goal group was

asked to set a long-term, specific, and challenging goal regarding how often they would use the training skills over the following six weeks. The proximal + distal goal group set both short-term (two-week and four-week) benchmark goals alongside their long-term goal, providing more immediate feedback and motivation. Outcomes were measured through self-reported surveys assessing self-efficacy and training transfer. After six weeks, the researchers found no significant differences in self-efficacy among the three groups. However, for training transfer, both the DYB group and the proximal + distal goal group showed greater application of skills in the workplace compared to the distal goal group. This suggests that distal goals alone may not be effective in encouraging skill transfer, whereas breaking down long-term goals into smaller, actionable steps may enhance workplace application. The findings highlight the potential limitations of distal goal setting as a transfer-of-training strategy and suggest that more structured, incremental goal setting approaches could be more effective.

Similarly, Brown et al. (2016) also investigate the impact of goal setting on training transfer, but with some key differences in design and findings. Like the 2005 study, it used a quasi-experimental design and was conducted in Canada with public-sector employees, but it focused specifically on management development training rather than self-awareness training. Instead of assessing the effect of distal and proximal goals, this study compared three types of behavioural goals – behavioural outcome goals, behavioural-specific goals, and rank-ordered behavioural goals – against a do-your-best (DYB) control group. Findings showed no clear advantage of setting behavioural goals over the do-your-best condition. While behavioural outcome goals led to higher self-reported transfer, behavioural-specific goals actually reduced transfer, as rated by workplace observers, compared to the other groups. The rank-ordered goals showed no significant difference from DYB. Overall, the results suggest that not all forms of goal setting enhance training transfer, and in some cases, encouraging employees to simply "do their best" may be just as effective compared to structured goal setting interventions.

Lastly, Brown et al. (2013) also explored the effect of different goal setting conditions after training on self-efficacy. The baseline training for this study was a two-day performance coaching programme for Canadian public-sector employees. Participants were then grouped into three study conditions: the first was instructed to “do their best”, the second was instructed to set behavioural outcome goals, and the third was prompted to set learning goals. As with Brown et al. (2005) and Brown et al. (2016), there was no statistically significant relationship identified between the goal setting intervention and self-efficacy.

Research question 4

What supports successful implementation of professional development interventions targeted at driving improvements in knowledge, skills, networks, work performance, and productivity in the civil service and adjacent contexts?

Our systematic review included 26 studies in total. Out of these, ten explicitly reported some elements related to Implementation and Process Evaluation (IPE). However, even within these studies, the reporting was not extensive, with most providing only

limited insights into how interventions were implemented in practice. No study conducted a standalone IPE, and in many cases, discussions around implementation were brief, anecdotal, or embedded within broader analyses of intervention effectiveness rather than forming a systematic evaluation.

Extent of implementation as planned and the nature and effect of programme adaptations

Across the 13 studies that reported some level of process evaluation, there was limited discussion about whether interventions were implemented fully as planned. Where implementation fidelity was addressed, the focus was generally on changes made during the implementation process, rather than structured assessments of adherence to the original intervention model. In some cases, interventions were actively adapted in response to emerging barriers, but these adaptations were rarely evaluated for their impact.

Viglione and Labrecque (2021) provided one of the few studies that systematically assessed an implementation adaptation. Their study was part of a broader evaluation of the STARR training programme for probation officers. During the implementation period, agency leaders identified that trained officers were not applying the STARR techniques as frequently as intended. To address this, the agency introduced a policy requiring all trained officers to use at least eight STARR skills per month. The effectiveness of this policy was evaluated in the study, and while they found that the policy successfully increased the frequency of skill use, officers still applied STARR techniques in less than half of their supervision interactions, suggesting that while mandates can drive compliance, they may not ensure full integration of skills into daily practice.

In contrast, Bonta et al. (2019) found that the STICS supervision model, which included monthly meetings, refresher training, and feedback, was not implemented as planned. Participation in these activities was low, and monthly meetings became sporadic and eventually stopped. Unlike in the STARR study, no structured adjustments were made to address these challenges, leading to a gradual breakdown in intervention fidelity.

Overall, the evidence suggests that while most professional development interventions are introduced with structured frameworks, their real-world implementation is often shaped by external constraints, leading to deviations from the original design. However, few studies systematically evaluated these adaptations, making it difficult to determine whether they improved or hindered the effectiveness of the interventions. This highlights a significant gap in the literature: future research should place greater emphasis on understanding how professional development programmes evolve in practice, assessing which adaptations are beneficial and which may undermine intervention goals.

Factors supporting or obstructing implementation

Several studies provided insights into the factors that influenced successful implementation, whether positively or negatively. While some factors were supported

by empirical evidence, others were identified as potential challenges based on the authors' reflections and theoretical considerations.

Organisational enablers, such as financial stability, smaller size, and a supportive work environment, may play a role in determining the effectiveness of training implementation. Seidle et al. (2016) noted that organisations with stronger financial resources were better able to sustain leadership development programmes, while those with weaker financial positions faced competing demands that diverted attention away from implementation. The study also found that larger organisations experienced greater challenges in ensuring training translated into measurable leadership improvements, as structural complexity and slower decision-making processes hindered the uptake of new approaches. A supportive work environment was also highlighted as a potential facilitator of implementation. Brown (2005) noted that organisations involved in in-depth needs assessments and actively engaged in training design were more supportive of professional development efforts.

Conversely, time constraints and competing work demands were frequently mentioned as key obstacles to effective implementation. Brown et al. (2016) found that many participants struggled to apply new skills due to high workloads and lack of time for reflection and practice, with one respondent stating that “time is the biggest obstacle”. This finding aligns with Bonta et al. (2019a), where probation officers reported that excessive administrative burdens made it difficult to fully engage with the STICS training model, which led to a drop in participation.

Another identified barrier was implementation fatigue. De Grip and Sauermann (2012) found that participants experienced a temporary decline in performance immediately after training, which they attributed to cognitive overload and fatigue from intensive learning sessions. This suggests that even well-designed interventions can have unintended short-term effects if implementation does not take workplace realities into account.

Participant attrition also posed a challenge to implementation. Bourgon and Gutierrez (2012) observed that, despite participation being voluntary, attrition rates were high, with one-third of officers dropping out before completion. This raised concerns about whether compulsory participation might yield better engagement.

Lastly, Borness et al. (2013) discussed dilution of training as a potential barrier. They described an informal adaptation, where a 16-hour training programme was stretched over 16 weeks due to workplace constraints. While necessary for feasibility, this dilution may have reduced the intervention’s effectiveness, and it was argued that more concentrated and extended doses could have been more successful, though no formal evaluation was conducted.

Overall, these findings suggest that successful implementation is influenced not only by intervention design but also by broader structural and organisational factors. Ensuring that training is well-integrated into workplace schedules, securing leadership support, and addressing workload constraints are all important considerations for improving intervention uptake.

Reporting biases

Given the small number and variability of studies, a funnel plot was not possible, so an albatross plot was used to check for p-value clustering as an indication of potential reporting bias.

The figure in Appendix F presents the albatross plot with the p-values of the included studies plotted on the x-axis. If publication bias were present, we would typically expect p-values to cluster just below the conventional significance threshold of 0.05, as studies with statistically significant results are more likely to be published, while non-significant findings may go unpublished or selectively reported. In the figure, studies with p-values below and above 0.05 are colour-coded for clarity. While no strong indications of bias are apparent, the limited number of studies prevents definitive conclusions.

Certainty of evidence

To assess how reliable the findings in this review are, we considered several factors: risk of bias, consistency across studies, directness (how closely the studies relate to the research question), precision of results, and potential publication bias.

The vote counting approach provided a quantitative summary along with a formal statistical test, helping to assess the level of uncertainty in the findings. Of the 16 relevant and included studies, 15 reported positive effects of professional development interventions on outcomes, such as knowledge, skills, work performance, and productivity. This corresponds to 94% of studies showing a positive effect, with a 95% confidence interval of 72% to 99%. This gives us strong statistical confidence that the majority of professional development interventions studied are associated with positive outcomes.

However, because a meta-analysis was not possible, and much of the synthesis relied on narrative methods, there are limitations. Specifically: (i) we are not able to determine how strong these positive effects are, and (ii) the findings associated with what makes a professional development intervention effective are not accompanied by formal statistical tests or precise estimates of uncertainty.

The quality of studies (measured by risk of bias assessment) varied substantially across studies, with an average score of 43% for RCTs and 70% for quasi-experimental designs. For RCT designs, common issues included poor reporting of how participants were randomly assigned, high dropout rates, lack of blinding for those assessing the outcomes, and failure to analyse all participants as originally assigned. In quasi-experimental studies, additional concerns arose around how participants were selected, incomplete follow-up, and other factors that could skew comparisons between groups. Missing or unclear data in several studies further reduced confidence in the results.

Publication bias remains a concern, as studies with null or negative findings may be underrepresented, potentially overestimating the effectiveness of interventions. We used an albatross plot, presented in Appendix F, to check for signs of reporting bias.

This type of analysis looks at patterns in study results to see if certain findings are more likely to be published than others. In this case, we did not see strong signs of bias, but because there were so few studies, we cannot be completely certain.

How much confidence should we have in the findings of this review?

Overall, there is consistent evidence suggesting that professional development interventions in civil service settings tend to have positive effects. The findings are encouraging, but the nature of the existing evidence, including variation in study design and outcomes measured, did not allow for a meta-analysis, which means it is not possible to draw reliable overall conclusions about the size of these effects or how the degree of effectiveness depends on other factors. In addition, the certainty of the evidence around what makes professional development effective in the civil service is moderate to low, due to the small number of identified studies and methodological limitations. A substantially greater body of high-quality research is needed to provide stronger, more conclusive evidence on the size of the effects and how to design interventions in the most effective way.

Discussion

Discussion of findings

This systematic review set out to assess the effectiveness of professional development interventions within the civil service and related sectors, focusing on outcomes such as skills, knowledge, work performance, productivity, and networks. The review aimed not only to synthesise evidence around the overall effect of professional development interventions but also to draw conclusions about how to design and implement learnings effectively. Overall, the findings point to consistent evidence that such interventions generally yield positive outcomes, particularly in relation to skills development. However, the review also highlights significant limitations in the nature, scope, and quality of the evidence base – limitations that affect the degree of confidence with which these findings can inform policy and practice.

Professional development interventions appear promising, but effect sizes and drivers of effectiveness remain unclear

The weight of evidence shows that the majority of included studies reported results favouring the intervention. This consistency is encouraging and aligns with a broader body of evidence from the fields of education and leadership training (Busso et al., 2023; Lacerenza et al., 2017, Sims et al., 2021) that professional development interventions have positive outcomes. However, the substantial heterogeneity in study designs, outcomes, and measurement tools, as well as the small number of studies overall, limited the richness of the conclusions we could draw. By preventing a meta-analysis, these factors limited our ability to synthesise the strength of effects or to understand how they vary across different contexts or intervention designs – a key gap in the existing literature.

Strongest evidence found for skill development; productivity outcomes remain unclear

A pattern that emerged is that there is strong evidence for positive effects on skills, contrasted with a very thin evidence base regarding improvements in work performance and productivity. A similar pattern is evident in wider organisational contexts. Arthur et al. (2003) conducted a meta-analysis of training effectiveness across sectors, finding similar effect sizes for learning, behaviour, and results-level outcomes. However, results-level outcomes – such as productivity or organisational performance – were assessed far less frequently, accounting for only 7% of the studies. This was attributed to the distal nature of these outcomes and the practical difficulties of isolating causal effects amid external variables.

The civil service context brings additional complexities. Productivity in knowledge-intensive roles is especially challenging to define and measure, and this was evident in the limited treatment of productivity outcomes across the studies reviewed. Where outcomes related to work performance or productivity were measured, they were

largely drawn from operational delivery roles, such as probation services or call centres, where outputs are relatively easy to define and then quantify. In this respect, the review highlights a significant gap: while many interventions are designed with productivity improvements in mind, there is limited empirical evidence demonstrating a clear causal link between professional development and productivity gains in civil service contexts.

Limited evidence on effective intervention design

Although the review provides some early indications about important features and mechanisms of professional development, the evidence is not strong enough to draw reliable conclusions about what makes an intervention effective. Design features were not consistently described, and few studies directly compared different formats. While some multi-modal, sustained, and contextually embedded interactive approaches appeared more promising, these patterns were neither robust nor systematically tested. As such, the review cannot yet support strong claims about optimal design. However, there is also some indication of this in related literature. For example, Sims et al. (2021) found that professional development interventions for teachers that included more mechanisms tended to have a higher impact on pupil test scores.

Structural and organisational barriers to implementation

Reporting on implementation barriers and enablers was limited, with few studies offering systematic assessments of contextual factors. However, where these issues were described, several common themes emerged. Time constraints and competing operational demands limited participants' ability to engage with training or apply new skills consistently. In some cases, implementation weakened over time due to a lack of follow-up or reduced support. Conversely, organisational features such as stable funding, supportive leadership, and opportunities to shape or adapt training to local needs were described as helpful in enabling delivery.

Although these themes were not studied in detail or by large numbers of our included studies, they do chime with wider evidence. Adams et al. (2023) found similar barriers and enablers in a national study of teachers: 66% reported lack of time due to workload as a key barrier to professional development, followed by cost (42%), and lack of cover (41%). Conversely, leadership support and schools prioritising staff development were identified as important enablers. International survey data analysed by the OECD show comparable patterns across sectors, with common barriers including lack of time due to work (29%), family responsibilities (16%), financial cost (16%), and lack of employer support (7%) (OECD, 2019).

Limitations of evidence

Despite a rigorous search and selection process, several limitations were identified in the available evidence, which have implications for the strength and generalisability of our findings.

Lack of studies meeting eligibility criteria

One of the primary challenges encountered was the limited number of studies that fulfilled the eligibility criteria. The number of relevant studies identified was significantly lower than initially expected, restricting the breadth of the evidence base. This shortage can be attributed, in part, to the inherent difficulties of conducting experimental studies within the civil service context. The high costs associated with implementing randomised controlled trials or other robust quasi-experimental designs in this setting make such studies rare.

Additionally, measuring outcomes related to productivity is notoriously difficult in the public sector. One key reason for this is that market prices are not available with which to value the output that the workforce produces. In addition, because much of the civil service is a part of the "knowledge economy", even the output itself can be difficult to identify precisely. Key counterexamples illustrate the point: where we do have productivity or performance measures, they are overwhelmingly in professions like call centre workers or probation officers, where certain outputs or outcomes, like call volumes or recidivism rates, can be straightforwardly quantified. These are the exceptions rather than the rule within the civil service.

Heterogeneity in outcome measures

A significant limitation across the included studies was the high degree of heterogeneity in outcome measures. The studies examined used a wide range of indicators to assess the impact of interventions, making comparisons challenging (and, relatedly, ruling out a meaningful meta-analysis). This variation is partly due to the diverse nature of the civil service workforce, which encompasses a wide array of roles, responsibilities, and organisational structures. This interacts with the low volume of literature overall; if there were more, then the challenge of heterogeneity could be partially addressed by conducting syntheses separately for different parts of the civil service. That said, even among studies examining similar worker populations or training interventions, inconsistencies in measurement approaches persisted. The lack of standardised outcome measures reduces the ability to synthesise results meaningfully and draw definitive conclusions.

Reporting issues

A key limitation identified throughout this review was the quality of reporting practices across the included studies. Specifically, the following areas could be improved:

- **PICO elements in abstracts:** Many abstracts fail to adequately report the essential PICO (Population, Intervention, Comparator, Outcome) elements. In this review, our search strategy relied on title and abstract keyword searches, meaning that studies with unclear reporting may have been overlooked. While full-text searching and screening help ensure comprehensiveness, the need to go beyond the abstract to extract this information makes the process significantly more resource-intensive. In health sciences, the inclusion of PICO elements in abstracts is more standardised, enabling clearer identification of relevant studies. Improved reporting of these elements would enhance the ease and accuracy of literature searches in future systematic reviews.

- **Details around how RCTs are conducted:** Many studies fail to report key aspects of how their RCTs were conducted, which complicates the assessment of risk of bias. Commonly used tools like ROBINS and the JBI assess factors such as randomisation procedures, blinding of assessors and participants, and handling of missing data. If these details are not clearly reported, the study receives a lower risk of bias score, even if these procedures were followed in practice. This lack of transparency can lead to inaccurate assessments of the study's quality and reliability. Properly reporting these details is essential for accurate risk of bias evaluations and for understanding the true validity of the study.
- **Participant and intervention details:** Many studies lack sufficient detail regarding participant characteristics, such as age, gender, and other demographic information. These details are essential for understanding the generalisability of findings and conducting sub-analyses to assess which interventions work for specific groups. Additionally, information about the intervention itself – such as its duration, location, and design – is often insufficient. Comprehensive descriptions of the training interventions would allow for better comparisons and enable insights into which components contribute to success. A lack of information about the details of the intervention also makes it difficult to assess with rigour the likely cost implications and scalability of the interventions. Variables such as dosage, duration, and delivery mode could be key determinants of those costs. Although our research questions did not specify that the "effectiveness" of professional learning design would include an explicit assessment of cost-effectiveness, this would clearly be an important part of the broader calculus for policy, and our review highlights that it would be extremely difficult to gauge cost-effectiveness from the information provided in many of the papers. Lastly, this also has implications for the taxonomy developed in this review. While we coded mechanisms used in the interventions, a lack of detailed reporting around the interventions could mean that we were not able to systematically capture the mechanisms employed during professional development.
- **Implementation and process evaluation elements:** Process and implementation evaluation elements, which assess how strategies are applied in practice, are often either underreported or not examined in sufficient depth. These evaluations are essential for understanding the factors that support or hinder successful implementation. Additionally, there is often a lack of clear protocols outlining the planned approach, which would allow for the identification of adaptations or implementation issues in a systematic way. Pre-registering studies and conducting separate process and implementation evaluations would offer valuable insights into the real-world application of professional development programmes, helping to identify best practices, areas for improvement, and key factors that contribute to their success or failure.

Our recommendations for future research due to this limitation and best practices are discussed in the Implications for further research section.

Limitations of review processes

Every systematic review involves a series of methodological decisions that inherently shape the scope of the evidence identified. While these decisions are necessary to ensure a focused and manageable review, they also introduce limitations that should be acknowledged.

One key limitation is the choice of keywords, which, despite substantial time invested during the scoping phase, may not have captured all relevant studies due to variations in terminology. Additionally, we limited the search to titles and abstracts, which in this field are often less structured and less informative than in the health sciences. Furthermore, the review was conducted using a more restricted set of databases than would typically be available for the most comprehensive systematic reviews, due to access constraints. As a result, some relevant studies may have been missed, though citation searches and the call for evidence were designed to mitigate this.

Eligibility criteria also influenced the scope. This review did not prioritise softer outcomes such as job satisfaction, motivation, turnover, mental health, or attitudes, meaning some relevant insights may be underrepresented. Similarly, this review focused on OECD countries to ensure the findings were relevant to the UK civil service. OECD member countries have comparable economic and administrative structures, making their research more applicable to the UK context than studies from non-OECD countries. While this geographical focus may be seen as a limitation, we sought to mitigate this by assessing whether expanding the eligibility criteria to include OECD key partner countries would add meaningful evidence. However, a targeted scoping exercise found no additional high-quality studies, suggesting that broadening the scope would not have significantly impacted the findings.

In addition, the focus was on RCTs and quasi-experimental studies, as these methodologies allow for stronger causal evidence of intervention effects. While this approach ensures rigour and comparability, it also introduces limitations, particularly in the civil service context. Many civil service interventions operate in complex environments where experimental designs are difficult to implement, meaning that valuable evidence from implementation studies, qualitative research, or mixed-methods approaches may have been excluded.

The synthesis approaches present other limitations. Given the heterogeneity of studies, we focused mostly on narrative synthesis, alongside vote counting as a method of formal quantitative synthesis. Both deliver key insights, but neither provides rigorous information about, or comparisons between, the magnitudes of effects, as would have been possible if the literature had allowed for meta-analysis.

Finally, we used Behaviour Change Techniques (BCT) as a starting point to develop our own taxonomy. However, BCT is behaviourally informed, and some professional development training may not take the same perspective or use comparable terminology. This could mean that certain interventions were not naturally classified within the BCT framework. To address this, we expanded the taxonomy where necessary to capture relevant approaches not fully covered by BCT.

These limitations reflect necessary trade-offs in balancing breadth, depth, and rigour. While they shape the scope of findings, they help cut through the noise to provide a focused and meaningful synthesis.

Implications for practice and policy

Notwithstanding the limitations of the evidence base that we have discussed, there are a number of insights that should help shape considerations around the design and implementation of professional development strategies. Below, we outline recommendations for practitioners and policymakers to keep in mind when designing and shaping professional development in the civil service context.

Professional development interventions tend to have positive effects

The systematic review demonstrates strong evidence that professional development interventions have a positive impact on skills and, to a lesser extent, in terms of evidence strength, work performance and productivity. This suggests that investing in professional development is a worthwhile endeavour for civil service organisations seeking to enhance employee capabilities.

- **Recommendation:** Civil service agencies should continue to prioritise professional development as a key workforce strategy.

Bridging the gap between skills and knowledge development and work performance and productivity

Findings suggest that the metrics by which training interventions are assessed are typically skills metrics. However, this review is concerned with skills development in the broader context of improving productivity and work performance. The focus on skills in practice is likely related to the fact that they are easier to measure – an issue we return to in implications for future research. For practitioners, it is therefore important not to lose sight of the fact that skill improvements are typically an intermediate outcome.

- **Recommendation:** Professional development programmes should explicitly define both the immediate skills they aim to improve and the justification for this in terms of the expected performance or productivity outcomes that this would lead to. Practitioners should design interventions that link skill-building with performance improvements and, where possible, this should be integrated with the data and measurements used as part of evaluations of the programme.

Overcoming structural and organisational barriers to implementation success

Time constraints and competing demands on employees present a significant barrier to training implementation. At the same time, organisational factors such as leadership support, financial resources, and a positive learning culture can act as enablers for successful implementation.

- **Recommendation:** The design of training interventions should include careful consideration not only of the content and mode of training, but also of how it

integrates with people's ongoing professional commitments. The joining up of learning design and workforce planning offers the possibility of better minimising workload conflicts and making training and work genuinely complementary.

Considering multi-component approaches when designing interventions

There is some evidence suggesting that training interventions that combine multiple methods or provide sustained support may be more effective than single-method approaches. Workshops with interactive elements appear to yield greater improvements than passive information sharing, and blended leadership training that integrates different strategies may enhance behavioural outcomes. Additionally, training that includes peer coaching, follow-up reinforcement, or higher intensity may be more impactful, particularly for skill retention and application. However, the overall evidence base remains limited, and further research is needed to confirm these patterns.

- **Recommendation:** While more evidence is required, practitioners should consider incorporating a mix of training methods, interactive elements, and sustained support when designing interventions, as these features may enhance effectiveness. This would also facilitate an expansion of the evidence base around these key dimensions of learning design.

Implications for further research

The need for more high-quality research on professional development in the civil service to inform effective design

The limited evidence base, particularly outside of operational delivery roles, makes it difficult to draw granular conclusions about what makes for effective design of professional development across the civil service. To build a stronger foundation, future research should prioritise rigorous studies across a broader range of roles and functions.

Developing a framework for translating private sector evidence to the civil service to bridge the evidence gap

There is relatively little high-quality research on professional development within the civil service itself. However, given the wide range of work conducted in both the civil service and wider economy, there is likely to be much that each can learn from the other. With this in mind, in this review, we considered evidence from comparable professions in both the public and private sectors, determining relevance based on specific roles.

One helpful avenue for further research would be to systematically examine what would be the most robust and practical for identifying insights that can be transferred from outside of the civil service to the civil service – whether based on professions (as we did in this review), common core skills, organisational features, or professional learning environments. There are many ways in which this could be done, and understanding the utility of each is a research agenda in itself. Making progress on

this question would allow for better-calibrated use of evidence from the private sector to fill evidence gaps in the civil service. This is particularly worthy of consideration given that some of the barriers to the evaluation of professional learning are likely to be persistently higher in the civil service relative to the private sector (e.g., the greater challenges in measuring productivity, as discussed below).

Promoting standardised outcome measures in professional development research within the civil service context

A key finding of this review is the diversity in outcome types and measurement approaches across studies, even for similar interventions. This variability makes it difficult to compare findings and prevents a quantitative synthesis of the magnitudes of effects (through meta-analysis), especially given the wide range of roles, responsibilities, and implementation settings within the civil service. To generate more robust and comparable evidence, future research should prioritise the development and adoption of more consistent, standardised outcome measures tailored to the civil service. This may involve creating generalised, widely accepted metrics, but given that the diversity of roles, interventions, and contexts makes full standardisation challenging, this is not the only solution. Rather, efforts could focus on providing clearer guidance on measuring specific outcomes in a way that accounts for sectoral diversity. Aligning outcome measurement approaches would enable more meaningful comparisons and strengthen the overall evidence base.

One notable gap is the measurement of productivity and work performance outcomes in the civil service context, other than for some operational delivery roles. While professional development interventions are consistently found to improve skills – an outcome more easily assessed through structured evaluations – evidence linking these interventions to productivity remains much more limited. Given that skill development ultimately aims to enhance job performance and organisational efficiency, it is crucial to expand the evidence base in this area.

To address this, future studies should attempt to develop productivity and performance measures suited to the civil service. We recognise that there are reasons why this has not already happened: objective measures of output in the knowledge economy are not always readily available, and the value of that output (which in principle offers a convenient metric by which to standardise, even across highly diverse lines of work) is hard to estimate, given the lack of market prices for the output of the public sector. Making progress on these challenges will likely require a concerted research effort involving academic researchers, statisticians, and government bodies, particularly the Office for National Statistics (ONS), which is already highly engaged in efforts to measure public service productivity (e.g., ONS Public Service Productivity).

Improving reporting standards in professional development research: Key issues and best practices

A key implication for future research is the need for better reporting practices, a significant issue identified throughout this review. To enhance the quality and utility of

professional development research, we recommend the following best practices, based on the issues and implications discussed in the Limitations of evidence section:

1. **Strengthening reporting guidelines and standards**

   Comprehensive reporting is essential for quality, transparency, and impact. Standardised frameworks such as CONSORT for RCTs and TIDieR for more specific intervention descriptions provide clear structures for reporting key study components (Hoffmann et al., 2014; Schulz et al., 2010). Strengthening adherence to these guidelines would improve the consistency and usability of research findings. Our findings align with concerns raised in prior research, such as Ryan et al. (2023), who highlight that inadequate reporting remains widespread despite the availability of established reporting checklists. Their study found that reporting frameworks, such as TIDieR, are inconsistently applied, with many journals failing to require comprehensive intervention descriptions. They further emphasise the role of journal editors in enforcing reporting standards by updating submission guidelines to mandate the use of reporting checklists. While not specifically mentioned in the study, given the civil service context, this is also applicable to those reviewing and signing off on grey literature. Based on this review, the areas that require particular improvement in reporting standards are:

   a. **PICO elements in abstracts:**

      Abstracts should systematically include key PICO (Population, Intervention, Comparator, Outcome) elements to facilitate efficient literature searches. CONSORT also provides guidelines for abstracts, outlining the key elements that should be included to improve the clarity and completeness of trial reporting (Hopewell et al., 2008). One of our included papers, Borness et al. (2013), uses CONSORT checklists, and we believe this provides the best example of an abstract from our review. We present this abstract in Appendix G for illustrative purposes.

   b. **Detailed reporting of intervention and population components:** Studies should provide comprehensive details on participant characteristics (e.g., age, gender, demographics) and intervention components (e.g., dosage, duration, delivery mode, and mechanisms).

   c. **Clear documentation of experimental design:** Transparent reporting of key methodological aspects, including randomisation procedures, blinding, and handling of missing data, is critical for accurate risk of bias assessments.

   d. **Incorporating implementation and process evaluation (IPE) elements:** Understanding how interventions function in practice requires greater emphasis on process and implementation evaluations. Researchers should systematically document how interventions are delivered, whether adaptations occur, and what factors contribute to success or failure. Such reporting would provide valuable insights into the real-world effectiveness of professional development initiatives.

2. **Registration of study protocols:** Pre-registering study protocols on platforms such as the Open Science Framework (OSF) or ClinicalTrials.gov enhances transparency and minimises bias. Registering protocols ensures that planned methodologies, interventions, and analyses are publicly documented before data collection begins, preventing selective reporting and increasing the credibility of findings. This practice also facilitates replication efforts and strengthens trust in research outcomes.

By implementing these best practices, researchers can significantly improve the quality, transparency, and impact of professional development research, ultimately contributing to more effective and evidence-based practices in the field.

Refining the taxonomy of mechanisms: Future directions and considerations

The BCT taxonomy provided a strong and evidence-based foundation for developing our taxonomy of mechanisms, allowing us to systematically categorise the mechanisms underpinning professional development interventions in the civil service context. Our review offers an overview of the types of mechanisms commonly employed in this context. However, given the limited number of studies available, there remains insufficient evidence to determine which mechanisms – or combinations thereof – are most effective in enhancing skills, knowledge, networks, work performance, and productivity. A key limitation observed was the inconsistent or incomplete reporting of mechanisms within intervention descriptions, which restricts the ability to draw firm conclusions about their impact. A structured approach to documenting mechanisms – such as integrating the taxonomy into intervention design and evaluation frameworks – would not only improve transparency but also enhance the ability to assess and compare effectiveness across studies. As more studies emerge, the taxonomy can be refined and expanded, either by incorporating new mechanisms or by further tailoring existing definitions to better reflect their application in the public sector. Additionally, as the evidence base grows, more can be said about the success of specific mechanisms and combinations of mechanisms in driving key outcomes, enabling a more nuanced understanding of what works best in professional development interventions.

Given these considerations, we recommend using our developed taxonomy as the primary framework for future work. The sheer number of mechanisms within the full BCT taxonomy can be overwhelming, whereas our version has been refined with civil service-relevant examples, making it more accessible and easier to interpret in this context. However, the BCT taxonomy remains a valuable resource, and when mechanisms emerge that do not clearly align with those in our taxonomy, the BCT should be consulted as a next step to determine whether additional mechanisms should be incorporated. To ensure ongoing relevance, the taxonomy should be continuously updated as new evidence emerges, supporting an iterative approach to refining civil service professional development strategies.

Conclusion

This systematic review finds that professional development interventions tend to have a positive impact, particularly on skills, across civil service and adjacent contexts. However, the evidence base is limited in both size and consistency, with large heterogeneity and gaps in reporting and design, making it difficult to draw firm conclusions about what works best. Strengthening the quality and quantity of future research will be essential to inform an effective professional development design.

Administrative Information

Registration and protocol

The protocol was registered on the UK government evaluation registry and is available on the UK government website. Any deviations from the protocol have been reported in the next section.

Deviations from protocol

We note the following deviations from the protocol:

1. As part of our risk of bias assessment, we deviated from the original protocol, which proposed using RoB2 for RCTs and ROBINS-I for quasi-experimental studies and instead opted for the JBI tool. This decision was made due to concerns about the suitability of RoB2 and ROBINS-I for our context. These tools are known for their complexity and require specialist knowledge to apply competently. They are also less commonly used in practice, even in health sciences, and social science studies often lack the level of detail necessary to complete them effectively. Given these challenges, we concluded that using these tools would require significant effort while still leaving large gaps in assessment due to missing information. After reviewing alternative tools, we selected JBI, as it provides structured checklists for both quasi-experimental and RCT study designs, offering a balance between comprehensiveness and feasibility.
2. To maximise the insights gained from a limited evidence base, we deviated from the original eligibility criteria by including studies where a subsample of the population and geographical focus met the inclusion criteria. This approach allowed us to capture relevant findings from studies that included both eligible and ineligible groups. For example, we assumed that a professional development programme targeting both civil service leaders (eligible) and leaders in the broader public sector (ineligible) would still provide valuable insights for the review. This adjustment enabled a broader yet still relevant evidence base while maintaining the focus on civil service contexts.
3. For research question 3, we made slight adjustments to the wording to better reflect our focus on both individual mechanisms and broader forms (clusters of mechanisms). Specifically, we added explicit references to mechanisms throughout, as we aim to examine both the role of specific mechanisms and how they interact within different forms of professional development. Additionally, we merged the previous sub-questions 3c and 3d, as the wording of 3d inherently addressed the findings of 3c – both examining whether certain mechanisms or combinations are associated with the greatest impact on skills, knowledge, networks, work performance, and productivity. These refinements ensure greater clarity and alignment with the scope of our analysis.

4. For research question 3, we originally planned to use elements of qualitative comparative analysis (QCA) to explore how different combinations of mechanisms contribute to the effectiveness of professional development interventions. However, due to limitations in the literature, we revised our approach. The small number of included studies relative to the number of mechanisms made QCA infeasible, even with planned aggregation strategies.
5. In the absence of a feasible meta-analysis, we decided to employ vote counting as recommended by the Cochrane handbook. Vote counting allows for a synthesis of a quantitative nature, as we have discussed in the report.

Support

This is an independent report conducted by Alma Economics and commissioned by Government Skills, which is part of the UK Cabinet Office and funded by HMT Labour Markets Evaluation and Pilots Fund. The main contact of Government Skills is Siobhan Dickens (gscu.comms@cabinetoffice.gov.uk).

An Expert Advisory Group consisting of members of the civil service and external experts has also been established for monitoring and oversight over the systematic review.

Both Government Skills and the Expert Advisory Group supported the development of the scope, analytic framework, and key questions for this review. However, they had no role in the selection of studies, quality assessment, or synthesis of evidence other than giving expert advice.

We also received expert support from Dr Mark Newman and Dr Janice Tripney from UCL, who advised on best practices for searching, screening, data extraction, and assessing the feasibility of meta-analysis. Julie Glanville, an experienced academic librarian, provided quality assurance throughout the review.

Competing interests

No competing interests to declare.

Availability of data, code, and other materials

Materials associated with this project, such as information extracted for the evidence map and any analytic code used, will be made available for public use through the Government Skills website. Additional information can be made available on request.

References

Adams, L., CoburnCrane, S., Sanders-Earley, A., Keeble, R., Harris, H., Taylor, J., Taylor, B., 2023. Working lives of teachers and leaders – Year 1: Research report. Department for Education.

Aldridge, S., Hawkins, A., Xuereb, C., 2016. Improving public sector efficiency to deliver a smarter state. Civ. Serv. Q. URL https://quarterly.blog.gov.uk/2016/01/25/improving-public-sector-efficiency-to-deliver-a-smarter-state/

AMEE, 2024. AMEE guides [WWW Document]. URL https://amee.org/amee-guides/

An, S.-H., Jensen, U.T., Bro, L.L., Andersen, L.B., Ladenburg, J., Meier, K.J., Salomonsen, H.H., 2022. Seeing eye to eye: can leadership training align perceptions of leadership? Int. Public Manag. J. 25, 2–23. https://doi.org/10.1080/10967494.2020.1763533

An, S.-H., Meier, K.J., 2021. Gender and the Effectiveness of Leadership Training: Results From a Field Experiment. Rev. Public Pers. Adm. 41, 747–770. https://doi.org/10.1177/0734371X20932989

An, S.-H., Meier, K.J., Bøllingtoft, A., Andersen, L.B., 2019. Employee Perceived Effect of Leadership Training: Comparing Public and Private Organizations. Int. Public Manag. J. 22, 2–28. https://doi.org/10.1080/10967494.2018.1497739

Arthur, W., Bennett, W., Edens, P.S., Bell, S.T., 2003. Effectiveness of training in organizations: A meta-analysis of design and evaluation features. J. Appl. Psychol. 88, 234–245. https://doi.org/10.1037/0021-9010.88.2.234

Asteris, M.M., Jr, 2013. An experimental study of interventions for the acquisition and retention of motivational interviewing skills among probation officers. Diss. Abstr. Int. Humanit. Soc. Sci. 73.

Barker, T.H., Habibi, N., Aromataris, E., Stone, J.C., Leonardi-Bee, J., Sears, K., Hasanoff, S., Klugar, M., Tufanaru, C., Moola, S., Munn, Z., 2024. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. JBI Evid. Synth. 22, 378–388. https://doi.org/10.11124/JBIES-23-00268

Barker, T.H., Stone, J.C., Sears, K., Klugar, M., Tufanaru, C., Leonardi-Bee, J., Aromataris, E., Munn, Z., 2023. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evid. Synth. https://doi.org/10.11124/JBIES-22-00430

BCT Taxonomy (v1): 93 hierarchically-clustered techniques, n.d.

Bonta, J., Bourgon, G., Rugge, T., Scott, T.-L., Yessine, A.K., Gutierrez, L., Li, J., 2011. An Experimental Demonstration of Training Probation Officers in Evidence-Based Community Supervision. Crim. Justice Behav. 38, 1127–1148. https://doi.org/10.1177/0093854811420678

Bonta, J., Rugge, T., Bourgon, G., Wanamaker, K.A., 2019. A conceptual replication of the Strategic Training Initiative in Community Supervision (STICS). J. Exp. Criminol. 15, 397–419. https://doi.org/10.1007/s11292-019-09371-4

Borness, C., Proudfoot, J., Crawford, J., Valenzuela, M., 2013. Putting Brain Training to the Test in the Workplace: A Randomized, Blinded, Multisite, Active-Controlled Trial. PLoS ONE 8, e59982. https://doi.org/10.1371/journal.pone.0059982

Bourgon, G., Gutierrez, L., 2012. The general responsivity principle in community supervision: the importance of probation officers using cognitive intervention techniques and its influence on recidivism. J. Crime Justice 35, 149–166. https://doi.org/10.1080/0735648X.2012.674816

Brown, T.C., 2005. Effectiveness of distal and proximal goals as transfer-of-training interventions: A field experiment. Hum. Resour. Dev. Q. 16, 369–387. https://doi.org/10.1002/hrdq.1144

Brown, T.C., McCracken, M., Hillier, T.-L., 2013. Using evidence-based practices to enhance transfer of training: assessing the effectiveness of goal setting and behavioural observation scales. Hum. Resour. Dev. Int. 16, 374–389. https://doi.org/10.1080/13678868.2013.812291

Brown, T.C., Warren, A.M., Khattar, V., 2016. The Effects of Different Behavioral Goals on Transfer from a Management Development Program. Hum. Resour. Dev. Q. 27, 349–372. https://doi.org/10.1002/hrdq.21257

Bushman, B.J., Wang, M.C., 2009. Vote-counting procedures in meta-analysis, in: The Handbook of Research Synthesis and Meta-Analysis, 2nd Ed. Russell Sage Foundation, New York, NY, US, pp. 207–220.

Busso, M., Park, K., Irazoque, N., 2023. The Effectiveness of Management Training Programs: A Meta-Analytic Review.

Cabinet Office, 2024. Civil service statistics 2024: Statistical bulletin [WWW Document]. https://www.gov.uk/government/statistics/civil-service-statistics-2024/statistical-bulletin-civil-service-statistics-2024

Campbell, M., McKenzie, J.E., Sowden, A., Katikireddi, S.V., Brennan, S.E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., Thomson, H., 2020. Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. BMJ 368, 6890. https://doi.org/10.1136/bmj.l6890

CEDEFOP, 2014. Terminology of European education and training policy: A selection of 100 key terms. European Centre for the Development of Vocational Training.

Chapman, R.J.K., 1987. Teaching public administration: Education or training? Aust. J. Public Adm. 46, 363–375. https://doi.org/10.1111/j.1467-8500.1987.tb02579.x

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., Coe, R., 2015. Developing Great Teaching: Lessons from the international reviews into effective professional development.

Cotabish, A., Robinson, A., 2012. The Effects of Peer Coaching on the Evaluation Knowledge, Skills, and Concerns of Gifted Program Administrators. Gift. Child Q. 56, 160–170. https://doi.org/10.1177/0016986212446861

De Grip, A., Sauermann, J., 2012. The Effects of Training on Own and Co-worker Productivity: Evidence from a Field Experiment. Econ. J. 122, 376–399. https://doi.org/10.1111/j.1468-0297.2012.02500.x

Dunleavy, P., Margetts, H., Bastow, S., Tinkler, J., 2006. New public management is dead—Long live digital-era governance. J. Public Adm. Res. Theory 16, 467–494. https://doi.org/10.1093/jopart/mui057

Dunlop, C.A., Kamkhaji, J.C., Radaelli, C.M., 2015. Regulators and Reform: A Quasi-experimental Assessment of the Effects of Training Inspectors. Int. Public Manag. J. 18, 304–322. https://doi.org/10.1080/10967494.2015.1023912

Education Endowment Foundation, 2021. Effective professional development.

Elliott, I.C., Bottom, K.A., O'Connor, K., 2023. The status of public administration teaching in the UK. J. Public Aff. Educ. 29, 262–274. https://doi.org/10.1080/15236803.2023.2202609

Filges, T., Torgerson, C., Gascoine, L., Dietrichson, J., Nielsen, C., Viinholt, B.A., 2019. Effectiveness of continuing professional development training of welfare professionals on outcomes for children and young people: A systematic review. Campbell Syst. Rev. 15, 1060. https://doi.org/10.1002/cl2.1060

Frontier Economics, 2022. Review of skills taxonomies: Report prepared for the SPB. GOV.UK.

Gershon, P., 2004. Releasing resources to the front line: independent review of public sector efficiency. HMSO, London.

Harrison, S., Jones, H.E., Martin, R.M., Lewis, S.J., Higgins, J.P.T., 2017. The albatross plot: A novel graphical tool for presenting results of diversely reported studies in a systematic review. Res. Synth. Methods 8, 281–289. https://doi.org/10.1002/jrsm.1239

Haunstrup, J.S., Jensen, U.T., 2024. Leadership training and just-in-time nudges: A field experiment on the transfer of learning into action. Int. Public Manag. J. 27, 684–708. https://doi.org/10.1080/10967494.2024.2314055

Hoffmann, T.C., Glasziou, P.P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S.E., Dixon-Woods, M., McCulloch, P., Wyatt, J.C., Chan, A.-W., Michie, S., 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 348, g1687–g1687. https://doi.org/10.1136/bmj.g1687

Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D.G., Schulz, K.F., and the CONSORT Group, 2008. CONSORT for Reporting Randomized Controlled Trials in Journal and Conference Abstracts: Explanation and Elaboration. PLoS Med. 5, e20. https://doi.org/10.1371/journal.pmed.0050020

Itzchakov, G., 2020. Can listening training empower service employees? The mediating roles of anxiety and perspective-taking. Eur. J. Work Organ. Psychol. 29, 938–952. https://doi.org/10.1080/1359432X.2020.1776701

Jacobsen, C.B., Andersen, L.B., Bøllingtoft, A., Eriksen, T.L.M., 2022. Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership. Public Adm. Rev. 82, 117–131. https://doi.org/10.1111/puar.13356

Labrecque, R.M., Viglione, J., 2021. The Impact of a Community Supervision Officer Training Program on Client Outcomes: A Propensity Score Modeling Analysis by Officer Training Dosage. Crim. Justice Behav. 48, 315–331. https://doi.org/10.1177/0093854820980788

Lacerenza, C.N., Reyes, D.L., Marlow, S.L., Joseph, D.L., Salas, E., 2017. Leadership training design, delivery, and implementation: A meta-analysis. J. Appl. Psychol. 102, 1686–1718. https://doi.org/10.1037/apl0000241

Librarianship Studies & Information Technology, 2017. Knowledge [WWW Document]. Librariansh. Stud. Inf. Technol. URL https://www.librarianshipstudies.com/2017/11/knowledge.html

López-Cabarcos, M.Á., Vázquez-Rodríguez, P., Quiñoá-Piñeiro, L.M., 2022. An approach to employees' job performance through work environmental variables and leadership behaviours. J. Bus. Res. 140, 361–369. https://doi.org/10.1016/j.jbusres.2021.11.006

López-López, J.A., Page, M.J., Lipsey, M.W., Higgins, J.P.T., 2018. Dealing with effect size multiplicity in systematic reviews and meta-analyses. Res. Synth. Methods 9, 336–351. https://doi.org/10.1002/jrsm.1310

Lowenkamp, C.T., Holsinger, A., Robinson, C.R., Alexander, M., 2014. Diminishing or durable treatment effects of STARR? A research note on 24-month re-arrest rates. J. Crime Justice 37, 275–283. https://doi.org/10.1080/0735648X.2012.753849

Lundqvist, D., Wallo, A., Coetzer, A., Kock, H., 2023. Leadership and Learning at Work: A Systematic Literature Review of Learning-oriented Leadership. J. Leadersh. Organ. Stud. 30, 205–238. https://doi.org/10.1177/15480518221133970

McKenzie, J.E., Brennan, S.E., 2024. Chapter 12: Synthesizing and presenting findings using other methods, in: Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions (Version 6.5).

Milič Kavčič, Z., Kavcic, V., Giordani, B., Marusic, U., 2024. Computerized Cognitive Training in the Older Workforce: Effects on Cognition, Life Satisfaction, and Productivity. Appl. Sci. Switz. 14. https://doi.org/10.3390/app14156470

OECD, 2019. Getting Skills Right: Future-Ready Adult Learning Systems.

Office for National Statistics, 2023. Public service productivity, UK: 1997 to 2022. ONS.

Robinson, C.R., Lowenkamp, C.T., Holsinger, A.M., VanBenschoten, S., Alexander, M., Oleson, J.C., 2012. A random study of Staff Training Aimed at Reducing Re-arrest (STARR): using core correctional practices in probation interactions. J. Crime Justice 35, 167–188. https://doi.org/10.1080/0735648X.2012.674823

Rojas Méndez, A.M., Scartascini, C., 2024. Debiasing Policymakers: The Role of Behavioral Economics Training (IDB Publications (Working Papers) No. 13476). Inter-American Development Bank. http://dx.doi.org/10.1823

Romani, M., 2021. Public sector productivity: Part one – Why is it important and how can we measure it? World Bank.

Ryan, M., Hoffmann, T., Hofmann, R., Van Sluijs, E., 2023. Incomplete reporting of complex interventions: a call to action for journal editors to review their submission guidelines. Trials 24, 176. https://doi.org/10.1186/s13063-023-07215-1

Salmela-Aro, K., Mutanen, P., Vuori, J., 2012. Promoting career preparedness and intrinsic work-goal motivation: RCT intervention. J. Vocat. Behav. 80, 67–75. https://doi.org/10.1016/j.jvb.2011.07.001

Saltiel, I.M., 2006. Networking and network theory, in: English, F.W. (Ed.), Encyclopedia of Educational Leadership and Administration. Sage Publications. https://search.credoreference.com/articles/Qm9va0FydGljbGU6MTE2ODQ4Mg==?aid=96753

Sauermann, J., 2021. Worker reciprocity and the returns to training: Evidence from a field experiment. J. Econ. Manag. Strategy 32, 543–557. https://doi.org/10.1111/jems.12419

Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group, 2010. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 340, c332–c332. https://doi.org/10.1136/bmj.c332

Seidle, B., Fernandez, S., Perry, J.L., 2016. Do Leadership Training and Development Make a Difference in the Public Sector? A Panel Study. Public Adm. Rev. 76, 603–613. https://doi.org/10.1111/puar.12531

Sims, S., Fletcher-Wood, H., 2020. Identifying the characteristics of effective teacher professional development: A critical review. J. Educ. Stud. 47, 47–63. https://doi.org/10.1080/09243453.2020.1772841

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Herwegen, J., Anders, J., 2021. What are the characteristics of teacher professional development that increase pupil achievement? A systematic review and meta-analysis. Educ. Endow. Found.

The Productivity Institute, 2024. What explains the UK's productivity problem? The Productivity Institute.

The Productivity Institute, 2022. Making public sector productivity practical: Capita report. The Productivity Institute.

UK Government, 1999. Modernising Government.

Van Reenen, J., Yang, X., 2024. Cracking the productivity code: An international comparison of UK productivity (POID Special Report. Centre for Economic Performance, LSE.

Viglione, J., Labrecque, R.M., 2021. Core Correctional Practices in Community Supervision: An Evaluation of a Policy Mandate to Increase Probation Officer Use of Skills. Int. J. Offender Ther. Comp. Criminol. 65, 858–881. https://doi.org/10.1177/0306624X20981045

World Bank, 2023. Worldwide bureaucracy indicators dashboard [WWW Document]. World Bank Group. URL https://www.worldbank.org/en/data/interactive/2019/05/21/worldwide-bureaucracy-indicators-dashboard

Appendix A: Search strategy

Table 4 presents the number of results, the search date, the applied filters, and the search strings used per database. Due to character limits, the full search string was not used across all databases. Moreover, in some cases, titles, abstracts, and keywords were searched separately due to the functionality of the database's built-in search field.

Table 4. Results of search strategy

Database	Search strings
Web of Science Results: 974 Date of search: 22/10/2024 Filters: Language: English, Dates: 01/01/2004-01/09/2024	(((TS=("civil service" OR "civil servant" OR "civil servants" OR "public sector" OR "public administration" OR "public administrators" OR "public management" OR "government employee" OR "government employees" OR "government worker" OR "government workers" OR "government workforce" OR "government work force" OR "government staff" OR "government personnel" OR "government agency" OR "government agencies" OR "government official" OR "government officials" OR "government department" OR "government departments" OR "public employee" OR "public employees" OR "public worker" OR "public workers" OR "public workforce" OR "public work force" OR "public staff" OR "public personnel" OR "public manager" OR "public managers" OR "public official" OR "public officials" OR "public service employee" OR "public service employees" OR "public service worker" OR "public service workers" OR "public service workforce" OR "public service work force" OR "public service staff" OR "public service personnel" OR "public service manager" OR "public service managers" OR "public service official" OR "public service officials" OR "public servant" OR "public servants" OR "federal employee" OR "federal employees" OR "federal worker" OR "federal workers" OR "federal workforce" OR "federal work force" OR "federal staff" OR "federal personnel" OR "federal agency" OR "federal agencies" OR "federal official" OR "federal officials" OR "state employee" OR "state employees" OR "state worker" OR "state workers" OR "state workforce" OR "state work force" OR "state staff" OR "state personnel" OR "state agency" OR "state agencies" OR "state official" OR "state officials" OR "local government" OR "local governments" OR "local official" OR "local officials" OR "local authority" OR "local authorities" OR "municipal government" OR "municipal governments" OR "municipal employee" OR "municipal employees" OR "municipal worker" OR "municipal workers" OR "municipal workforce" OR "municipal work force" OR "municipal staff" OR "municipal

Database	Search strings
	personnel" OR "municipal official" OR "municipal officials" OR "regional government" OR "regional governments" OR "regional employee" OR "regional employees" OR "regional worker" OR "regional workers" OR "regional workforce" OR "regional work force" OR "regional staff" OR "regional personnel" OR "regional official" OR "regional officials" OR "operational employee" OR "operational employees" OR "operational worker" OR "operational workers" OR "operational workforce" OR "operational work force" OR "operational staff" OR "operational manager" OR "operational managers" OR "government front office" OR "government back office" OR "public service operations" OR "government operations" OR "court employee" OR "court employees" OR "court worker" OR "court workers" OR "court work force" OR "court workforce" OR "court staff" OR "prison employee" OR "prison employees" OR "prison worker" OR "prison workers" OR "prison workforce" OR "prison work force" OR "prison officer" OR "prison officers" OR "prison staff" OR "prison manager" OR "prison managers" OR "tribunal officer" OR "tribunal clerk" OR "court service officer" OR "border force officer" OR "border force officers" OR "immigration officer" OR "immigration officers" OR "customs officer" OR "customs officers" OR "probation officer" OR "probation officers" OR "civil enforcement officer" OR "public service bailiff" OR "enforcement officer" OR "passport control officer" OR "passport control officers" OR "visa processing officer" OR "visa processing officers" OR "consular employee" OR "consular employees" OR "consular worker" OR "consular workers" OR "consular workforce" OR "consular work force" OR "consular officer" OR "consular officers" OR "consular staff" OR "consular manager" OR "consular managers" OR "job centre employee" OR "job centre employees" OR "job centre worker" OR "job centre workers" OR "job centre workforce" OR "job centre work force" OR "job centre staff" OR "job centre manager" OR "job centre managers" OR "job center employee" OR "job center employees" OR "job center worker" OR "job center workers" OR "job center workforce" OR "job center work force" OR "job center staff" OR "job center manager" OR "job center managers" OR "work coach" OR "work coaches" OR "call centre employee" OR "call centre employees" OR "call centre worker" OR "call centre workers" OR "call centre workforce" OR "call centre work force" OR "call centre staff" OR "call centre manager" OR "call centre managers" OR "call center employee" OR "call center employees" OR "call center worker" OR "call center workers" OR "call center workforce" OR "call center work force" OR "call center staff" OR "call center manager" OR "call center managers" OR "employment advisor" OR "employment advisors" OR "customer service employee" OR "customer service

alma economics

Database	Search strings
	employees" OR "customer service worker" OR "customer service workers" OR "customer service workforce" OR "customer service work force" OR "customer service staff" OR "customer service representative" OR "customer service representatives" OR "customer service advisor" OR "customer service advisors" OR "policy advisor" OR "policy advisors" OR "policy staff" OR "policy officer" OR "policy officers" OR "policy adviser" OR "policy advisers" OR "policy design" OR "policy designing" OR "policy delivery" OR "policy advice" OR "policy briefing" OR "policy implementation" OR "policy implementing" OR "policy evaluation" OR "policy evaluations" OR "policy evaluating" OR "strategy design" OR "strategy designing" OR "strategy delivery" OR "strategy advisor" OR "strategy advisors" OR "strategy advice" OR "strategy briefing" OR "strategy implementation" OR "strategy implementing" OR "strategy evaluation" OR "strategy evaluations" OR "strategy evaluating" OR "policy official" OR "policy delivering" OR "policy advising" OR "policy research" OR "policy consultation" OR "policy legislation" OR "strategy designing" OR "strategy advising" OR "business architect" OR "business architects" OR "data architect" OR "data architects" OR "enterprise architect" OR "enterprise architects" OR "network architect" OR "network architects" OR "security architect" OR "security architects" OR "solution architect" OR "solution architects" OR "technical architect" OR "technical architects" OR "analytics engineer" OR "analytics engineers" OR "data analyst" OR "data analysts" OR "data engineer" OR "data engineers" OR "data ethicist" OR "data ethicists" OR "data governance manager" OR "data governance managers" OR "data scientist" OR "data scientists" OR "machine learning engineer" OR "machine learning engineers" OR "performance analyst" OR "performance analysts" OR "application operations engineer" OR "application operations engineers" OR "business relationship manager" OR "business relationship managers" OR "change and release manager" OR "change and release managers" OR "command and control centre manager" OR "command and control centre managers" OR "end user computing engineer" OR "end user computing engineers" OR "it service manager" OR "it service managers" OR "incident manager" OR "incident managers" OR "infrastructure engineer" OR "infrastructure engineers" OR "infrastructure operations engineer" OR "infrastructure operations engineers" OR "problem manager" OR "problem managers" OR "service desk manager" OR "service desk managers" OR "service transition manager" OR "service transition managers" OR "business analyst" OR "business analysts" OR "delivery manager" OR "delivery managers" OR "digital portfolio manager" OR "digital portfolio managers" OR "product manager" OR "product managers" OR "programme

alma economics

Database	Search strings
	delivery manager" OR "programme delivery managers" OR "program delivery manager" OR "program delivery managers" OR "service owner" OR "service owners" OR "quality assurance testing analyst" OR "quality assurance testing analysts" OR "qat analyst" OR "qat analysts" OR "test engineer" OR "test engineers" OR "test manager" OR "test managers" OR "development operations engineer" OR "development operations engineers" OR "devops engineer" OR "devops engineers" OR "frontend developer" OR "frontend developers" OR "software developer" OR "software developers" OR "accessibility specialist" OR "accessibility specialists" OR "content designer" OR "content designers" OR "content strategist" OR "content strategists" OR "graphic designer" OR "graphic designers" OR "interaction designer" OR "interaction designers" OR "service designer" OR "service designers" OR "technical writer" OR "technical writers" OR "user researcher" OR "user researchers" OR "project delivery" OR "project manager" OR "project managers" OR "project management" OR "project lead" OR "project leader" OR "project leaders" OR "project leadership" OR "project admin" OR "project administration" OR "project administrator" OR "project planning" OR "project analyst" OR "project support" OR "project consultant" OR "project consultants" OR "project consultancy" OR "project consultation" OR "project coordination" OR "project coordinator" OR "project coordinators" OR "project director" OR "project directors" OR "programme delivery" OR "programme manager" OR "programme managers" OR "programme management" OR "programme lead" OR "programme leader" OR "programme leaders" OR "programme leadership" OR "programme admin" OR "programme administration" OR "programme administrator" OR "programme planning" OR "programme analyst" OR "programme support" OR "program delivery" OR "program manager" OR "program managers" OR "program management" OR "program lead" OR "program leader" OR "program leaders" OR "program leadership" OR "program admin" OR "program administration" OR "program administrator" OR "program planning" OR "program analyst" OR "program support" OR "resource delivery" OR "resource manager" OR "resource managers" OR "resource management" OR "resource lead" OR "resource leader" OR "resource leaders" OR "resource leadership" OR "resource admin" OR "resource administration" OR "resource administrator" OR "resource planning" OR "resource analyst" OR "resource support" OR "business case delivery" OR "business case manager" OR "business case managers" OR "business case management" OR "business case lead" OR "business case leader" OR "business case leaders" OR "business case leadership" OR "business case admin" OR "business case administration" OR "business case

alma economics

Database	Search strings
	administrator" OR "business case planning" OR "business case analyst" OR "business case support" OR "tax professional" OR "tax professionals" OR "tax specialist" OR "tax specialists" OR "tax lead" OR "taxation professional" OR "taxation professionals" OR "taxation specialist" OR "taxation specialists" OR "taxation lead" OR "tax centre" OR "tax centres" OR "tax center" OR "tax centers" OR "tax agency")) AND TS=("professional development" OR "PD" OR "career development" OR "talent development" OR "leadership development" OR "executive development" OR "human resource development" OR "skill development" OR "skills development" OR "skill acquisition" OR "skills acquisition" OR "capacity development" OR "capacity building" OR "workforce development" OR "workplace development" OR "work place development" OR "workplace learning" OR "work place learning" OR "continuing development" OR "continuing learning" OR "lifelong development" OR "lifelong learning" OR "personal development" OR "practice based development" OR "practice based learning" OR "professional learning" OR "professional education" OR "career education" OR "leadership education" OR "continuing education" OR "lifelong education" OR "executive education" OR "workplace education" OR "work place education" OR "practice based education" OR "training" OR "CPD" OR "CPE" OR "learning and development" OR "L&D" OR "knowledge acquisition" OR "knowledge transfer" OR "knowledge sharing" OR "upskilling" OR "up-skilling" OR "reskilling" OR "re-skilling" OR "accreditation" OR "coaching" OR "human capital development" OR "leadership development" OR "talent development" OR "mentoring" OR "adult learning" OR "learning culture" OR "learning organisation" OR "competency development" OR "competencies development" OR "competency acquisition" OR "competencies acquisition" OR "network development" OR "network acquisition")) AND TS=("field experiment" OR "field experiments" OR "field study" OR "field studies" OR "natural experiment" OR "natural experiments" OR "quasi experiment" OR "quasi-experiment" OR "quasi experiments" OR "quasi-experiments" OR "quasi-experimental" OR "quasi experimental" OR "experimental design" OR "experimental study" OR "experimental evidence" OR "controlled trial" OR "control trial" OR "controlled trials" OR "control trials" OR RCT* OR "random experiment" OR "random experiments" OR "random assignment" OR "random assignments" OR "random allocation" OR "random allocations" OR "random trial" OR "random trials" OR "random treatment" OR "random treatments" OR "random intervention" OR "random interventions" OR "random comparison" OR "random comparisons" OR "randomised assignment" OR "randomised assignments" OR "randomised allocation" OR "randomised allocations" OR "randomised trial" OR

Database	Search strings
	"randomised trials" OR "randomised treatment" OR "randomised treatments" OR "randomised intervention" OR "randomised interventions" OR "randomised comparison" OR "randomised comparisons" OR "randomized assignment" OR "randomized assignments" OR "randomized allocation" OR "randomized allocations" OR "randomized trial" OR "randomized trials" OR "randomized treatment" OR "randomized treatments" OR "randomized intervention" OR "randomized interventions" OR "randomized comparison" OR "randomized comparisons" OR "randomized controlled trial" OR "randomized controlled trials" OR "pretest posttest" OR "pretest-posttest" OR "pre/post" OR "before-after" OR "difference-in-differences" OR "difference-in-difference" OR "diff-in-diff" OR "propensity score" OR "regression discontinuity" OR "RDD" OR "instrumental variable" OR "instrumental variables" OR "cohort study" OR "cohort studies" OR "control group" OR "control groups" OR "treatment group" OR "treatment groups" OR "impact evaluation" OR "causal analysis" OR "causal inference" OR "matching techniques" OR "covariate matching" OR "inverse probability weighting" OR "nearest neighbor matching" OR "exact matching" OR "kernel matching" OR "interrupted time series" OR "synthetic control" OR "synthetic controls" OR "panel study"))
Scopus Results: 1,450 Date of search: 22/10/2024 Filters: Language: English, Year: 2004-2024, Document Type: Article	(((TITLE-ABS-KEY("civil service" OR "civil servant" OR "civil servants" OR "public sector" OR "public administration" OR "public administrators" OR "public management" OR "government employee" OR "government employees" OR "government worker" OR "government workers" OR "government workforce" OR "government work force" OR "government staff" OR "government personnel" OR "government agency" OR "government agencies" OR "government official" OR "government officials" OR "government department" OR "government departments" OR "public employee" OR "public employees" OR "public worker" OR "public workers" OR "public workforce" OR "public work force" OR "public staff" OR "public personnel" OR "public manager" OR "public managers" OR "public official" OR "public officials" OR "public service employee" OR "public service employees" OR "public service worker" OR "public service workers" OR "public service workforce" OR "public service work force" OR "public service staff" OR "public service personnel" OR "public service manager" OR "public service managers" OR "public service official" OR "public service officials" OR "public servant" OR "public servants" OR "federal employee" OR "federal employees" OR "federal worker" OR "federal workers" OR "federal workforce" OR "federal work force" OR "federal staff" OR "federal personnel" OR

alma economics

Database	Search strings
	"federal agency" OR "federal agencies" OR "federal official" OR "federal officials" OR "state employee" OR "state employees" OR "state worker" OR "state workers" OR "state workforce" OR "state work force" OR "state staff" OR "state personnel" OR "state agency" OR "state agencies" OR "state official" OR "state officials" OR "local government" OR "local governments" OR "local official" OR "local officials" OR "local authority" OR "local authorities" OR "municipal government" OR "municipal governments" OR "municipal employee" OR "municipal employees" OR "municipal worker" OR "municipal workers" OR "municipal workforce" OR "municipal work force" OR "municipal staff" OR "municipal personnel" OR "municipal official" OR "municipal officials" OR "regional government" OR "regional governments" OR "regional employee" OR "regional employees" OR "regional worker" OR "regional workers" OR "regional workforce" OR "regional work force" OR "regional staff" OR "regional personnel" OR "regional official" OR "regional officials" OR "operational employee" OR "operational employees" OR "operational worker" OR "operational workers" OR "operational workforce" OR "operational work force" OR "operational staff" OR "operational manager" OR "operational managers" OR "government front office" OR "government back office" OR "public service operations" OR "government operations" OR "court employee" OR "court employees" OR "court worker" OR "court workers" OR "court work force" OR "court workforce" OR "court staff" OR "prison employee" OR "prison employees" OR "prison worker" OR "prison workers" OR "prison workforce" OR "prison work force" OR "prison officer" OR "prison officers" OR "prison staff" OR "prison manager" OR "prison managers" OR "tribunal officer" OR "tribunal clerk" OR "court service officer" OR "border force officer" OR "border force officers" OR "immigration officer" OR "immigration officers" OR "customs officer" OR "customs officers" OR "probation officer" OR "probation officers" OR "civil enforcement officer" OR "public service bailiff" OR "enforcement officer" OR "passport control officer" OR "passport control officers" OR "visa processing officer" OR "visa processing officers" OR "consular employee" OR "consular employees" OR "consular worker" OR "consular workers" OR "consular workforce" OR "consular work force" OR "consular officer" OR "consular officers" OR "consular staff" OR "consular manager" OR "consular managers" OR "job centre employee" OR "job centre employees" OR "job centre worker" OR "job centre workers" OR "job centre workforce" OR "job centre work force" OR "job centre staff" OR "job centre manager" OR "job centre managers" OR "job center employee" OR "job center employees" OR "job center worker" OR "job center workers" OR "job center workforce" OR "job center work force" OR "job center staff" OR

alma economics

Database	Search strings
	"job center manager" OR "job center managers" OR "work coach" OR "work coaches" OR "call centre employee" OR "call centre employees" OR "call centre worker" OR "call centre workers" OR "call centre workforce" OR "call centre work force" OR "call centre staff" OR "call centre manager" OR "call centre managers" OR "call center employee" OR "call center employees" OR "call center worker" OR "call center workers" OR "call center workforce" OR "call center work force" OR "call center staff" OR "call center manager" OR "call center managers" OR "employment advisor" OR "employment advisors" OR "customer service employee" OR "customer service employees" OR "customer service worker" OR "customer service workers" OR "customer service workforce" OR "customer service work force" OR "customer service staff" OR "customer service representative" OR "customer service representatives" OR "customer service advisor" OR "customer service advisors" OR "policy advisor" OR "policy advisors" OR "policy staff" OR "policy officer" OR "policy officers" OR "policy adviser" OR "policy advisers" OR "policy design" OR "policy designing" OR "policy delivery" OR "policy advice" OR "policy briefing" OR "policy implementation" OR "policy implementing" OR "policy evaluation" OR "policy evaluations" OR "policy evaluating" OR "strategy design" OR "strategy designing" OR "strategy delivery" OR "strategy advisor" OR "strategy advisors" OR "strategy advice" OR "strategy briefing" OR "strategy implementation" OR "strategy implementing" OR "strategy evaluation" OR "strategy evaluations" OR "strategy evaluating" OR "policy official" OR "policy delivering" OR "policy advising" OR "policy research" OR "policy consultation" OR "policy legislation" OR "strategy designing" OR "strategy advising" OR "business architect" OR "business architects" OR "data architect" OR "data architects" OR "enterprise architect" OR "enterprise architects" OR "network architect" OR "network architects" OR "security architect" OR "security architects" OR "solution architect" OR "solution architects" OR "technical architect" OR "technical architects" OR "analytics engineer" OR "analytics engineers" OR "data analyst" OR "data analysts" OR "data engineer" OR "data engineers" OR "data ethicist" OR "data ethicists" OR "data governance manager" OR "data governance managers" OR "data scientist" OR "data scientists" OR "machine learning engineer" OR "machine learning engineers" OR "performance analyst" OR "performance analysts" OR "application operations engineer" OR "application operations engineers" OR "business relationship manager" OR "business relationship managers" OR "change and release manager" OR "change and release managers" OR "command and control centre manager" OR "command and control centre managers" OR "end user computing engineer"

alma economics

Database	Search strings
	OR "end user computing engineers" OR "it service manager" OR "it service managers" OR "incident manager" OR "incident managers" OR "infrastructure engineer" OR "infrastructure engineers" OR "infrastructure operations engineer" OR "infrastructure operations engineers" OR "problem manager" OR "problem managers" OR "service desk manager" OR "service desk managers" OR "service transition manager" OR "service transition managers" OR "business analyst" OR "business analysts" OR "delivery manager" OR "delivery managers" OR "digital portfolio manager" OR "digital portfolio managers" OR "product manager" OR "product managers" OR "programme delivery manager" OR "programme delivery managers" OR "program delivery manager" OR "program delivery managers" OR "service owner" OR "service owners" OR "quality assurance testing analyst" OR "quality assurance testing analysts" OR "qat analyst" OR "qat analysts" OR "test engineer" OR "test engineers" OR "test manager" OR "test managers" OR "development operations engineer" OR "development operations engineers" OR "devops engineer" OR "devops engineers" OR "frontend developer" OR "frontend developers" OR "software developer" OR "software developers" OR "accessibility specialist" OR "accessibility specialists" OR "content designer" OR "content designers" OR "content strategist" OR "content strategists" OR "graphic designer" OR "graphic designers" OR "interaction designer" OR "interaction designers" OR "service designer" OR "service designers" OR "technical writer" OR "technical writers" OR "user researcher" OR "user researchers" OR "project delivery" OR "project manager" OR "project managers" OR "project management" OR "project lead" OR "project leader" OR "project leaders" OR "project leadership" OR "project admin" OR "project administration" OR "project administrator" OR "project planning" OR "project analyst" OR "project support" OR "project consultant" OR "project consultants" OR "project consultancy" OR "project consultation" OR "project coordination" OR "project coordinator" OR "project coordinators" OR "project director" OR "project directors" OR "programme delivery" OR "programme manager" OR "programme managers" OR "programme management" OR "programme lead" OR "programme leader" OR "programme leaders" OR "programme leadership" OR "programme admin" OR "programme administration" OR "programme administrator" OR "programme planning" OR "programme analyst" OR "programme support" OR "program delivery" OR "program manager" OR "program managers" OR "program management" OR "program lead" OR "program leader" OR "program leaders" OR "program leadership" OR "program admin" OR "program administration" OR "program administrator" OR "program planning" OR "program analyst" OR "program

alma economics

Database	Search strings
	support" OR "resource delivery" OR "resource manager" OR "resource managers" OR "resource management" OR "resource lead" OR "resource leader" OR "resource leaders" OR "resource leadership" OR "resource admin" OR "resource administration" OR "resource administrator" OR "resource planning" OR "resource analyst" OR "resource support" OR "business case delivery" OR "business case manager" OR "business case managers" OR "business case management" OR "business case lead" OR "business case leader" OR "business case leaders" OR "business case leadership" OR "business case admin" OR "business case administration" OR "business case administrator" OR "business case planning" OR "business case analyst" OR "business case support" OR "tax professional" OR "tax professionals" OR "tax specialist" OR "tax specialists" OR "tax lead" OR "taxation professional" OR "taxation professionals" OR "taxation specialist" OR "taxation specialists" OR "taxation lead" OR "tax centre" OR "tax centres" OR "tax center" OR "tax centers" OR "tax agency")) AND TITLE-ABS-KEY("professional development" OR "PD" OR "career development" OR "talent development" OR "leadership development" OR "executive development" OR "human resource development" OR "skill development" OR "skills development" OR "skill acquisition" OR "skills acquisition" OR "capacity development" OR "capacity building" OR "workforce development" OR "workplace development" OR "work place development" OR "workplace learning" OR "work place learning" OR "continuing development" OR "continuing learning" OR "lifelong development" OR "lifelong learning" OR "personal development" OR "practice based development" OR "practice based learning" OR "professional learning" OR "professional education" OR "career education" OR "leadership education" OR "continuing education" OR "lifelong education" OR "executive education" OR "workplace education" OR "work place education" OR "practice based education" OR "training" OR "CPD" OR "CPE" OR "learning and development" OR "L&D" OR "knowledge acquisition" OR "knowledge transfer" OR "knowledge sharing" OR "upskilling" OR "up-skilling" OR "reskilling" OR "re-skilling" OR "accreditation" OR "coaching" OR "human capital development" OR "leadership development" OR "talent development" OR "mentoring" OR "adult learning" OR "learning culture" OR "learning organisation" OR "competency development" OR "competencies development" OR "competency acquisition" OR "competencies acquisition" OR "network development" OR "network acquisition")) AND TITLE-ABS-KEY("field experiment" OR "field experiments" OR "field study" OR "field studies" OR "natural experiment" OR "natural experiments" OR "quasi experiment" OR "quasi-experiment" OR "quasi experiments" OR "quasi-

Database	Search strings
	experiments" OR "quasi-experimental" OR "quasi experimental" OR "experimental design" OR "experimental study" OR "experimental evidence" OR "controlled trial" OR "control trial" OR "controlled trials" OR "control trials" OR RCT* OR "random experiment" OR "random experiments" OR "random assignment" OR "random assignments" OR "random allocation" OR "random allocations" OR "random trial" OR "random trials" OR "random treatment" OR "random treatments" OR "random intervention" OR "random interventions" OR "random comparison" OR "random comparisons" OR "randomised assignment" OR "randomised assignments" OR "randomised allocation" OR "randomised allocations" OR "randomised trial" OR "randomised trials" OR "randomised treatment" OR "randomised treatments" OR "randomised intervention" OR "randomised interventions" OR "randomised comparison" OR "randomised comparisons" OR "randomized assignment" OR "randomized assignments" OR "randomized allocation" OR "randomized allocations" OR "randomized trial" OR "randomized trials" OR "randomized treatment" OR "randomized treatments" OR "randomized intervention" OR "randomized interventions" OR "randomized comparison" OR "randomized comparisons" OR "randomized controlled trial" OR "randomized controlled trials" OR "pretest posttest" OR "pretest-posttest" OR "pre/post" OR "before-after" OR "difference-in-differences" OR "difference-in-difference" OR "diff-in-diff" OR "propensity score" OR "regression discontinuity" OR "RDD" OR "instrumental variable" OR "instrumental variables" OR "cohort study" OR "cohort studies" OR "control group" OR "control groups" OR "treatment group" OR "treatment groups" OR "impact evaluation" OR "causal analysis" OR "causal inference" OR "matching techniques" OR "covariate matching" OR "inverse probability weighting" OR "nearest neighbor matching" OR "exact matching" OR "kernel matching" OR "interrupted time series" OR "synthetic control" OR "synthetic controls" OR "panel study"))

Database	Search strings
ERIC – Descriptors (character limit) Results: 2 Date of search: 25/10/2024 Filters: Descriptor only, Dates: 2004-2024	descriptor:("Randomized Controlled Trials" OR "Quasiexperimental Design" OR "Causal Models") AND descriptor:("Training" OR "Coaching (Performance)" OR "Capacity Building") AND descriptor:("Public Sector" OR "Local Government" OR "Government Employees") pubyearmin:2004 pubyearmax:2024
ERIC – Title (character limit) Results: 0 Date of search: 25/10/2024 Filters: Title only, Dates: 2004-2024	title:("controlled trial" OR "quasi-experimental" OR "randomized trial" OR "RCT") AND title:("training" OR "coaching" OR "capacity building") AND title:("public sector" OR "local government" OR "civil service" OR “civil servants”) pubyearmin:2004 pubyearmax:2024

alma economics

Database	Search strings
ERIC – Abstract (character limit) Results: 11 Date of search: 25/10/2024 Filters: Abstract only, Dates: 2004-2024	abstract:("controlled trial" OR "quasi-experimental" OR "randomized trial" OR "RCT") AND abstract:("training" OR "coaching" OR "capacity building") AND abstract:("public sector" OR "local government" OR "civil service" OR “civil servants”) pubyearmin:2004 pubyearmax:2024
IDEAS/RePEc (character limit) Results: 61 Date of search: 25/10/2024 Filters: Dates: 2004-2024 Separately applied to: Title, Abstract, Keywords	("civil service"|"civil servant"|"civil servants"|"public sector"|"public administration"|"government employee"|"government worker"|"government workforce"|"government personnel"|"government employees"|"government agency"|"government agencies"|"government official"|"government department"|"public employee"|"public worker"|"public workforce"|"public personnel"|"public manager"|"public official"|"federal employee"|"federal worker"|"federal workforce"|"state employee"|"state worker"|"state workforce"|"local government"|"local governments"|"local authorities"|"local authority"|"local official"|"municipal government"|"regional government"|"court employee"|"prison employee"|"tribunal officer"|"border force officer"|"immigration officer"|"customs officer"|"probation officer"|"enforcement officer"|"visa processing officer"|"consular officer"|"job center employee"|"call center employee"|"customer service employee"|"policy officer"|"policy advisor"|"strategy advisor"|"data scientist"|"digital portfolio manager"|"devops engineer"|"frontend developer"|"software developer"|"project manager"|"tax specialist") + ("professional development"|"career development"|"leadership development"|"skills acquisition"|"upskilling"|"L&D"|"coaching"|"mentoring"|"training"|"capacity building"|"accreditation"|"knowledge acquisition"|"knowledge transfer") + ("controlled trial"| "quasi-experimental"|"propensity score"|"randomized trial"|RCT*|"experimental design"|"experimental study"| "field experiment"|"randomised trial"|"causal inference"|"propensity score")

alma economics

Database	Search strings
gov.uk* Results: 8 Date of search: 30/10/2024	(“controlled trial” OR "quasi-experimental" OR "randomized trial" OR RCT*)(training OR "coaching" OR "capacity building")(“public sector” OR "local government" OR “civil service” OR “civil servants”) site:www.gov.uk
OECD library – Title (character limit) Results: 0 Date of search: 25/10/2024 Filters: Dates: Pre-2009 - 2024, Content type: Journals, Articles, Paper, Language: English	(Title ‘"civil service" OR "civil servant" OR "civil servants" OR "public sector" OR "public administration" OR "government employee" OR "government worker" OR "government workforce" OR "government personnel" OR "government employees" OR "government agency" OR "government agencies" OR "government official" OR "government department" OR "public employee" OR "public worker" OR "public workforce" OR "public personnel" OR "public manager" OR "public official" OR "federal employee" OR "federal worker" OR "federal workforce" OR "state employee" OR "state worker" OR "state workforce" OR "local government" OR "local governments" OR "local authorities" OR "local authority" OR "local official" OR "municipal government" OR "regional government" OR "court employee" OR "prison employee" OR "tribunal officer" OR "border force officer" OR "immigration officer" OR "customs officer" OR "probation officer" OR "enforcement officer" OR "visa processing officer" OR "consular officer" OR "job center employee" OR "call center employee" OR "customer service employee" OR "policy officer" OR "policy advisor" OR "strategy advisor" OR "data scientist" OR "digital portfolio manager" OR "devops engineer" OR "frontend developer" OR "software developer" OR "project manager" OR "tax specialist"’) (Language ‘en’) AND (Title ‘"professional development" OR "career development" OR "leadership development" OR "skills acquisition" OR "upskilling" OR "L&D" OR "coaching" OR "mentoring" OR "training" OR "capacity building" OR "accreditation" OR "knowledge acquisition" OR "knowledge transfer"’) AND (Title ‘"controlled trial" OR "quasi-experimental" OR "propensity score" OR "randomized trial" OR RCT* OR "experimental design" OR "experimental study" OR "field experiment" OR "randomised trial" OR "causal inference" OR "propensity score"’) AND (‘’) with type(s) subtype/journal OR subtype/article OR subtype/workingpaper

Database	Search strings
OECD library – Abstract (character limit) Results: 0 Date of search: 25/10/2024 Filters: Dates: Pre-2009 - 2024, Content type: Journals, Articles, Paper, Language: English	(Abstract ‘"civil service" OR "civil servant" OR "civil servants" OR "public sector" OR "public administration" OR "government employee" OR "government worker" OR "government workforce" OR "government personnel" OR "government employees" OR "government agency" OR "government agencies" OR "government official" OR "government department" OR "public employee" OR "public worker" OR "public workforce" OR "public personnel" OR "public manager" OR "public official" OR "federal employee" OR "federal worker" OR "federal workforce" OR "state employee" OR "state worker" OR "state workforce" OR "local government" OR "local governments" OR "local authorities" OR "local authority" OR "local official" OR "municipal government" OR "regional government" OR "court employee" OR "prison employee" OR "tribunal officer" OR "border force officer" OR "immigration officer" OR "customs officer" OR "probation officer" OR "enforcement officer" OR "visa processing officer" OR "consular officer" OR "job center employee" OR "call center employee" OR "customer service employee" OR "policy officer" OR "policy advisor" OR "strategy advisor" OR "data scientist" OR "digital portfolio manager" OR "devops engineer" OR "frontend developer" OR "software developer" OR "project manager" OR "tax specialist"’) (Language ‘en’) AND (Abstract ‘"professional development" OR "career development" OR "leadership development" OR "skills acquisition" OR "upskilling" OR "L&D" OR "coaching" OR "mentoring" OR "training" OR "capacity building" OR "accreditation" OR "knowledge acquisition" OR "knowledge transfer"’) AND (Abstract ‘"controlled trial" OR "quasi-experimental" OR "propensity score" OR "randomized trial" OR RCT* OR "experimental design" OR "experimental study" OR "field experiment" OR "randomised trial" OR "causal inference" OR "propensity score"’) AND (‘’) with type(s) subtype/journal OR subtype/article OR subtype/workingpaper

Database	Search strings
ProQuest – Title (character limit) Results: 5 Date of search: 30/10/2024 Filters: Document type: Article, Dissertation/Thesis, Government & Official Document, Report, Dates: 01/01/2004-01/09/2024, English selected as a language filter after completing the initial search	title(("civil service" OR "civil servant" OR "civil servants" OR "public sector" OR "public administration" OR "public administrators" OR "public management" OR "government employee" OR "government employees" OR "government worker" OR "government workers" OR "government workforce" OR "government work force" OR "government staff" OR "government personnel" OR "government agency" OR "government agencies" OR "government official" OR "government officials" OR "government department" OR "government departments" OR "public employee" OR "public employees" OR "public worker" OR "public workers" OR "public workforce" OR "public work force" OR "public staff" OR "public personnel" OR "public manager" OR "public managers" OR "public official" OR "public officials" OR "public service employee" OR "public service employees" OR "public service worker" OR "public service workers" OR "public service workforce" OR "public service work force" OR "public service staff" OR "public service personnel" OR "public service manager" OR "public service managers" OR "public service official" OR "public service officials" OR "public servant" OR "public servants" OR "federal employee" OR "federal employees" OR "federal worker" OR "federal workers" OR "federal workforce" OR "federal work force" OR "federal staff" OR "federal personnel" OR "federal agency" OR "federal agencies" OR "federal official" OR "federal officials" OR "state employee" OR "state employees" OR "state worker" OR "state workers" OR "state workforce" OR "state work force" OR "state staff" OR "state personnel" OR "state agency" OR "state agencies" OR "state official" OR "state officials" OR "local government" OR "local governments" OR "local official" OR "local officials" OR "local authority" OR "local authorities" OR "municipal government" OR "municipal governments" OR "municipal employee" OR "municipal employees" OR "municipal worker" OR "municipal workers" OR "municipal workforce" OR "municipal work force" OR "municipal staff" OR "municipal personnel" OR "municipal official" OR "municipal officials" OR "regional government" OR "regional governments" OR "regional employee" OR "regional employees" OR "regional worker" OR "regional workers" OR "regional workforce" OR "regional work force" OR "regional staff" OR "regional personnel" OR "regional official" OR "regional officials" OR "operational employee" OR "operational employees" OR "operational worker" OR "operational workers" OR "operational workforce" OR "operational work force" OR "operational staff" OR "operational manager" OR "operational managers" OR "government front office" OR "government back office" OR "public service operations" OR "government operations" OR "court employee" OR "court employees" OR "court worker" OR "court workers" OR "court work force" OR "court workforce" OR "court staff"

alma economics

Database	Search strings
	OR "prison employee" OR "prison employees" OR "prison worker" OR "prison workers" OR "prison workforce" OR "prison work force" OR "prison officer" OR "prison officers" OR "prison staff" OR "prison manager" OR "prison managers" OR "tribunal officer" OR "tribunal clerk" OR "court service officer" OR "border force officer" OR "border force officers" OR "immigration officer" OR "immigration officers" OR "customs officer" OR "customs officers" OR "probation officer" OR "probation officers" OR "civil enforcement officer" OR "public service bailiff" OR "enforcement officer" OR "passport control officer" OR "passport control officers" OR "visa processing officer" OR "visa processing officers" OR "consular employee" OR "consular employees" OR "consular worker" OR "consular workers" OR "consular workforce" OR "consular work force" OR "consular officer" OR "consular officers" OR "consular staff" OR "consular manager" OR "consular managers" OR "job centre employee" OR "job centre employees" OR "job centre worker" OR "job centre workers" OR "job centre workforce" OR "job centre work force" OR "job centre staff" OR "job centre manager" OR "job centre managers" OR "job center employee" OR "job center employees" OR "job center worker" OR "job center workers" OR "job center workforce" OR "job center work force" OR "job center staff" OR "job center manager" OR "job center managers" OR "work coach" OR "work coaches" OR "call centre employee" OR "call centre employees" OR "call centre worker" OR "call centre workers" OR "call centre workforce" OR "call centre work force" OR "call centre staff" OR "call centre manager" OR "call centre managers" OR "call center employee" OR "call center employees" OR "call center worker" OR "call center workers" OR "call center workforce" OR "call center work force" OR "call center staff" OR "call center manager" OR "call center managers" OR "employment advisor" OR "employment advisors" OR "customer service employee" OR "customer service employees" OR "customer service worker" OR "customer service workers" OR "customer service workforce" OR "customer service work force" OR "customer service staff" OR "customer service representative" OR "customer service representatives" OR "customer service advisor" OR "customer service advisors" OR "policy advisor" OR "policy advisors" OR "policy staff" OR "policy officer" OR "policy officers" OR "policy adviser" OR "policy advisers" OR "policy design" OR "policy designing" OR "policy delivery" OR "policy advice" OR "policy briefing" OR "policy implementation" OR "policy implementing" OR "policy evaluation" OR "policy evaluations" OR "policy evaluating" OR "policy official" OR "policy delivering" OR "policy advising" OR "policy research" OR "policy consultation" OR "policy legislation" OR "strategy design" OR "strategy designing" OR "strategy

alma economics

Database	Search strings
	delivery" OR "strategy advisor" OR "strategy advisors" OR "strategy advice" OR "strategy briefing" OR "strategy implementation" OR "strategy implementing" OR "strategy evaluation" OR "strategy evaluations" OR "strategy evaluating" OR "strategy designing" OR "strategy advising" OR "business architect" OR "business architects" OR "data architect" OR "data architects" OR "enterprise architect" OR "enterprise architects" OR "network architect" OR "network architects" OR "security architect" OR "security architects" OR "solution architect" OR "solution architects" OR "technical architect" OR "technical architects" OR "analytics engineer" OR "analytics engineers" OR "data analyst" OR "data analysts" OR "data engineer" OR "data engineers" OR "data ethicist" OR "data ethicists" OR "data governance manager" OR "data governance managers" OR "data scientist" OR "data scientists" OR "machine learning engineer" OR "machine learning engineers" OR "performance analyst" OR "performance analysts" OR "application operations engineer" OR "application operations engineers" OR "business relationship manager" OR "business relationship managers" OR "change and release manager" OR "change and release managers" OR "command and control centre manager" OR "command and control centre managers" OR "end user computing engineer" OR "end user computing engineers" OR "it service manager" OR "it service managers" OR "incident manager" OR "incident managers" OR "infrastructure engineer" OR "infrastructure engineers" OR "infrastructure operations engineer" OR "infrastructure operations engineers" OR "problem manager" OR "problem managers" OR "service desk manager" OR "service desk managers" OR "service transition manager" OR "service transition managers" OR "business analyst" OR "business analysts" OR "delivery manager" OR "delivery managers" OR "digital portfolio manager" OR "digital portfolio managers" OR "product manager" OR "product managers" OR "programme delivery manager" OR "programme delivery managers" OR "program delivery manager" OR "program delivery managers" OR "service owner" OR "service owners" OR "quality assurance testing analyst" OR "quality assurance testing analysts" OR "qat analyst" OR "qat analysts" OR "test engineer" OR "test engineers" OR "test manager" OR "test managers" OR "development operations engineer" OR "development operations engineers" OR "devops engineer" OR "devops engineers" OR "frontend developer" OR "frontend developers" OR "software developer" OR "software developers" OR "accessibility specialist" OR "accessibility specialists" OR "content designer" OR "content designers" OR "content strategist" OR "content strategists" OR "graphic designer" OR "graphic designers" OR "interaction designer" OR "interaction designers" OR "service designer"

Database	Search strings
	OR "service designers" OR "technical writer" OR "technical writers" OR "user researcher" OR "user researchers" OR "project delivery" OR "project manager" OR "project managers" OR "project management" OR "project lead" OR "project leader" OR "project leaders" OR "project leadership" OR "project admin" OR "project administration" OR "project administrator" OR "project planning" OR "project analyst" OR "project support" OR "project consultant" OR "project consultants" OR "project consultancy" OR "project consultation" OR "project coordination" OR "project coordinator" OR "project coordinators" OR "project director" OR "project directors" OR "programme delivery" OR "programme manager" OR "programme managers" OR "programme management" OR "programme lead" OR "programme leader" OR "programme leaders" OR "programme leadership" OR "programme admin" OR "programme administration" OR "programme administrator" OR "programme planning" OR "programme analyst" OR "programme support" OR "program delivery" OR "program manager" OR "program managers" OR "program management" OR "program lead" OR "program leader" OR "program leaders" OR "program leadership" OR "program admin" OR "program administration" OR "program administrator" OR "program planning" OR "program analyst" OR "program support" OR "resource delivery" OR "resource manager" OR "resource managers" OR "resource management" OR "resource lead" OR "resource leader" OR "resource leaders" OR "resource leadership" OR "resource admin" OR "resource administration" OR "resource administrator" OR "resource planning" OR "resource analyst" OR "resource support" OR "business case delivery" OR "business case manager" OR "business case managers" OR "business case management" OR "business case lead" OR "business case leader" OR "business case leaders" OR "business case leadership" OR "business case admin" OR "business case administration" OR "business case administrator" OR "business case planning" OR "business case analyst" OR "business case support" OR "tax professional" OR "tax professionals" OR "tax specialist" OR "tax specialists" OR "tax lead" OR "taxation professional" OR "taxation professionals" OR "taxation specialist" OR "taxation specialists" OR "taxation lead" OR "tax centre" OR "tax centres" OR "tax center" OR "tax centers" OR "tax agency") AND ("professional development" OR "PD" OR "career development" OR "talent development" OR "leadership development" OR "executive development" OR "human resource development" OR "skill development" OR "skills development" OR "skill acquisition" OR "skills acquisition" OR "capacity development" OR "capacity building" OR "workforce development" OR "workplace development" OR "work place development" OR "workplace learning" OR "work

alma economics

Database	Search strings
	place learning" OR "continuing development" OR "continuing learning" OR "lifelong development" OR "lifelong learning" OR "personal development" OR "practice based development" OR "practice based learning" OR "professional learning" OR "professional education" OR "career education" OR "leadership education" OR "continuing education" OR "lifelong education" OR "executive education" OR "workplace education" OR "work place education" OR "practice based education" OR "training" OR "CPD" OR "CPE" OR "learning and development" OR "L&D" OR "knowledge acquisition" OR "knowledge transfer" OR "knowledge sharing" OR "upskilling" OR "up-skilling" OR "reskilling" OR "re-skilling" OR "accreditation" OR "coaching" OR "human capital development" OR "leadership development" OR "talent development" OR "mentoring" OR "adult learning" OR "learning culture" OR "learning organisation" OR "competency development" OR "competencies development" OR "competency acquisition" OR "competencies acquisition" OR "network development" OR "network acquisition") AND ("field experiment" OR "field experiments" OR "field study" OR "field studies" OR "natural experiment" OR "natural experiments" OR "quasi experiment" OR "quasi-experiment" OR "quasi experiments" OR "quasi-experiments" OR "quasi-experimental" OR "quasi experimental" OR "experimental design" OR "experimental study" OR "experimental evidence" OR "controlled trial" OR "control trial" OR "controlled trials" OR "control trials" OR RCT* OR "random experiment" OR "random experiments" OR "random assignment" OR "random assignments" OR "random allocation" OR "random allocations" OR "random trial" OR "random trials" OR "random treatment" OR "random treatments" OR "random intervention" OR "random interventions" OR "random comparison" OR "random comparisons" OR "randomised assignment" OR "randomised assignments" OR "randomised allocation" OR "randomised allocations" OR "randomised trial" OR "randomised trials" OR "randomised treatment" OR "randomised treatments" OR "randomised intervention" OR "randomised interventions" OR "randomised comparison" OR "randomised comparisons" OR "randomized assignment" OR "randomized assignments" OR "randomized allocation" OR "randomized allocations" OR "randomized trial" OR "randomized trials" OR "randomized treatment" OR "randomized treatments" OR "randomized intervention" OR "randomized interventions" OR "randomized comparison" OR "randomized comparisons" OR "randomized controlled trial" OR "randomized controlled trials" OR "pretest posttest" OR "pretest-posttest" OR "pre/post" OR "before-after" OR "difference-in-differences" OR "difference-in-difference" OR "diff-in-diff" OR "propensity score" OR "regression discontinuity" OR "RDD" OR "instrumental variable" OR

alma economics

Database	Search strings
	"instrumental variables" OR "cohort study" OR "cohort studies" OR "control group" OR "control groups" OR "treatment group" OR "treatment groups" OR "impact evaluation" OR "causal analysis" OR "causal inference" OR "matching techniques" OR "covariate matching" OR "inverse probability weighting" OR "nearest neighbor matching" OR "nearest neighbour matching" OR "exact matching" OR "kernel matching" OR "interrupted time series" OR "synthetic control" OR "synthetic controls" OR "panel study"))
ProQuest – Abstract (character limit) Results: 418 Date of search: 30/10/2024 Filters: Document type: Article, Dissertation/Thesis, Government & Official Document, Report, Dates: 01/01/2004-01/09/2024, English selected as a language filter after completing the initial search	abstract(("civil service" OR "civil servant" OR "civil servants" OR "public sector" OR "public administration" OR "public administrators" OR "public management" OR "government employee" OR "government employees" OR "government worker" OR "government workers" OR "government workforce" OR "government work force" OR "government staff" OR "government personnel" OR "government agency" OR "government agencies" OR "government official" OR "government officials" OR "government department" OR "government departments" OR "public employee" OR "public employees" OR "public worker" OR "public workers" OR "public workforce" OR "public work force" OR "public staff" OR "public personnel" OR "public manager" OR "public managers" OR "public official" OR "public officials" OR "public service employee" OR "public service employees" OR "public service worker" OR "public service workers" OR "public service workforce" OR "public service work force" OR "public service staff" OR "public service personnel" OR "public service manager" OR "public service managers" OR "public service official" OR "public service officials" OR "public servant" OR "public servants" OR "federal employee" OR "federal employees" OR "federal worker" OR "federal workers" OR "federal workforce" OR "federal work force" OR "federal staff" OR "federal personnel" OR "federal agency" OR "federal agencies" OR "federal official" OR "federal officials" OR "state employee" OR "state employees" OR "state worker" OR "state workers" OR "state workforce" OR "state work force" OR "state staff" OR "state personnel" OR "state agency" OR "state agencies" OR "state official" OR "state officials" OR "local government" OR "local governments" OR "local official" OR "local officials" OR "local authority" OR "local authorities" OR "municipal government" OR "municipal governments" OR "municipal employee" OR "municipal employees" OR "municipal worker" OR "municipal workers" OR "municipal workforce" OR "municipal work force" OR "municipal staff" OR "municipal personnel" OR "municipal official" OR "municipal officials" OR "regional government" OR "regional governments" OR "regional employee" OR "regional employees" OR "regional worker" OR "regional workers"

alma economics

Database	Search strings
	OR "regional workforce" OR "regional work force" OR "regional staff" OR "regional personnel" OR "regional official" OR "regional officials" OR "operational employee" OR "operational employees" OR "operational worker" OR "operational workers" OR "operational workforce" OR "operational work force" OR "operational staff" OR "operational manager" OR "operational managers" OR "government front office" OR "government back office" OR "public service operations" OR "government operations" OR "court employee" OR "court employees" OR "court worker" OR "court workers" OR "court work force" OR "court workforce" OR "court staff" OR "prison employee" OR "prison employees" OR "prison worker" OR "prison workers" OR "prison workforce" OR "prison work force" OR "prison officer" OR "prison officers" OR "prison staff" OR "prison manager" OR "prison managers" OR "tribunal officer" OR "tribunal clerk" OR "court service officer" OR "border force officer" OR "border force officers" OR "immigration officer" OR "immigration officers" OR "customs officer" OR "customs officers" OR "probation officer" OR "probation officers" OR "civil enforcement officer" OR "public service bailiff" OR "enforcement officer" OR "passport control officer" OR "passport control officers" OR "visa processing officer" OR "visa processing officers" OR "consular employee" OR "consular employees" OR "consular worker" OR "consular workers" OR "consular workforce" OR "consular work force" OR "consular officer" OR "consular officers" OR "consular staff" OR "consular manager" OR "consular managers" OR "job centre employee" OR "job centre employees" OR "job centre worker" OR "job centre workers" OR "job centre workforce" OR "job centre work force" OR "job centre staff" OR "job centre manager" OR "job centre managers" OR "job center employee" OR "job center employees" OR "job center worker" OR "job center workers" OR "job center workforce" OR "job center work force" OR "job center staff" OR "job center manager" OR "job center managers" OR "work coach" OR "work coaches" OR "call centre employee" OR "call centre employees" OR "call centre worker" OR "call centre workers" OR "call centre workforce" OR "call centre work force" OR "call centre staff" OR "call centre manager" OR "call centre managers" OR "call center employee" OR "call center employees" OR "call center worker" OR "call center workers" OR "call center workforce" OR "call center work force" OR "call center staff" OR "call center manager" OR "call center managers" OR "employment advisor" OR "employment advisors" OR "customer service employee" OR "customer service employees" OR "customer service worker" OR "customer service workers" OR "customer service workforce" OR "customer service work force" OR "customer service staff" OR "customer service representative" OR

alma economics

Database	Search strings
	"customer service representatives" OR "customer service advisor" OR "customer service advisors" OR "policy advisor" OR "policy advisors" OR "policy staff" OR "policy officer" OR "policy officers" OR "policy adviser" OR "policy advisers" OR "policy design" OR "policy designing" OR "policy delivery" OR "policy advice" OR "policy briefing" OR "policy implementation" OR "policy implementing" OR "policy evaluation" OR "policy evaluations" OR "policy evaluating" OR "policy official" OR "policy delivering" OR "policy advising" OR "policy research" OR "policy consultation" OR "policy legislation" OR "strategy design" OR "strategy designing" OR "strategy delivery" OR "strategy advisor" OR "strategy advisors" OR "strategy advice" OR "strategy briefing" OR "strategy implementation" OR "strategy implementing" OR "strategy evaluation" OR "strategy evaluations" OR "strategy evaluating" OR "strategy designing" OR "strategy advising" OR "business architect" OR "business architects" OR "data architect" OR "data architects" OR "enterprise architect" OR "enterprise architects" OR "network architect" OR "network architects" OR "security architect" OR "security architects" OR "solution architect" OR "solution architects" OR "technical architect" OR "technical architects" OR "analytics engineer" OR "analytics engineers" OR "data analyst" OR "data analysts" OR "data engineer" OR "data engineers" OR "data ethicist" OR "data ethicists" OR "data governance manager" OR "data governance managers" OR "data scientist" OR "data scientists" OR "machine learning engineer" OR "machine learning engineers" OR "performance analyst" OR "performance analysts" OR "application operations engineer" OR "application operations engineers" OR "business relationship manager" OR "business relationship managers" OR "change and release manager" OR "change and release managers" OR "command and control centre manager" OR "command and control centre managers" OR "end user computing engineer" OR "end user computing engineers" OR "it service manager" OR "it service managers" OR "incident manager" OR "incident managers" OR "infrastructure engineer" OR "infrastructure engineers" OR "infrastructure operations engineer" OR "infrastructure operations engineers" OR "problem manager" OR "problem managers" OR "service desk manager" OR "service desk managers" OR "service transition manager" OR "service transition managers" OR "business analyst" OR "business analysts" OR "delivery manager" OR "delivery managers" OR "digital portfolio manager" OR "digital portfolio managers" OR "product manager" OR "product managers" OR "programme delivery manager" OR "programme delivery managers" OR "program delivery manager" OR "program delivery managers" OR "service owner" OR "service owners" OR "quality assurance testing analyst" OR "quality

alma economics

Database	Search strings
	assurance testing analysts" OR "qat analyst" OR "qat analysts" OR "test engineer" OR "test engineers" OR "test manager" OR "test managers" OR "development operations engineer" OR "development operations engineers" OR "devops engineer" OR "devops engineers" OR "frontend developer" OR "frontend developers" OR "software developer" OR "software developers" OR "accessibility specialist" OR "accessibility specialists" OR "content designer" OR "content designers" OR "content strategist" OR "content strategists" OR "graphic designer" OR "graphic designers" OR "interaction designer" OR "interaction designers" OR "service designer" OR "service designers" OR "technical writer" OR "technical writers" OR "user researcher" OR "user researchers" OR "project delivery" OR "project manager" OR "project managers" OR "project management" OR "project lead" OR "project leader" OR "project leaders" OR "project leadership" OR "project admin" OR "project administration" OR "project administrator" OR "project planning" OR "project analyst" OR "project support" OR "project consultant" OR "project consultants" OR "project consultancy" OR "project consultation" OR "project coordination" OR "project coordinator" OR "project coordinators" OR "project director" OR "project directors" OR "programme delivery" OR "programme manager" OR "programme managers" OR "programme management" OR "programme lead" OR "programme leader" OR "programme leaders" OR "programme leadership" OR "programme admin" OR "programme administration" OR "programme administrator" OR "programme planning" OR "programme analyst" OR "programme support" OR "program delivery" OR "program manager" OR "program managers" OR "program management" OR "program lead" OR "program leader" OR "program leaders" OR "program leadership" OR "program admin" OR "program administration" OR "program administrator" OR "program planning" OR "program analyst" OR "program support" OR "resource delivery" OR "resource manager" OR "resource managers" OR "resource management" OR "resource lead" OR "resource leader" OR "resource leaders" OR "resource leadership" OR "resource admin" OR "resource administration" OR "resource administrator" OR "resource planning" OR "resource analyst" OR "resource support" OR "business case delivery" OR "business case manager" OR "business case managers" OR "business case management" OR "business case lead" OR "business case leader" OR "business case leaders" OR "business case leadership" OR "business case admin" OR "business case administration" OR "business case administrator" OR "business case planning" OR "business case analyst" OR "business case support" OR "tax professional" OR "tax professionals" OR "tax specialist" OR "tax specialists" OR "tax lead" OR "taxation

alma economics

Database	Search strings
	professional" OR "taxation professionals" OR "taxation specialist" OR "taxation specialists" OR "taxation lead" OR "tax centre" OR "tax centres" OR "tax center" OR "tax centers" OR "tax agency") AND ("professional development" OR "PD" OR "career development" OR "talent development" OR "leadership development" OR "executive development" OR "human resource development" OR "skill development" OR "skills development" OR "skill acquisition" OR "skills acquisition" OR "capacity development" OR "capacity building" OR "workforce development" OR "workplace development" OR "work place development" OR "workplace learning" OR "work place learning" OR "continuing development" OR "continuing learning" OR "lifelong development" OR "lifelong learning" OR "personal development" OR "practice based development" OR "practice based learning" OR "professional learning" OR "professional education" OR "career education" OR "leadership education" OR "continuing education" OR "lifelong education" OR "executive education" OR "workplace education" OR "work place education" OR "practice based education" OR "training" OR "CPD" OR "CPE" OR "learning and development" OR "L&D" OR "knowledge acquisition" OR "knowledge transfer" OR "knowledge sharing" OR "upskilling" OR "up-skilling" OR "reskilling" OR "re-skilling" OR "accreditation" OR "coaching" OR "human capital development" OR "leadership development" OR "talent development" OR "mentoring" OR "adult learning" OR "learning culture" OR "learning organisation" OR "competency development" OR "competencies development" OR "competency acquisition" OR "competencies acquisition" OR "network development" OR "network acquisition") AND ("field experiment" OR "field experiments" OR "field study" OR "field studies" OR "natural experiment" OR "natural experiments" OR "quasi experiment" OR "quasi-experiment" OR "quasi experiments" OR "quasi-experiments" OR "quasi-experimental" OR "quasi experimental" OR "experimental design" OR "experimental study" OR "experimental evidence" OR "controlled trial" OR "control trial" OR "controlled trials" OR "control trials" OR RCT* OR "random experiment" OR "random experiments" OR "random assignment" OR "random assignments" OR "random allocation" OR "random allocations" OR "random trial" OR "random trials" OR "random treatment" OR "random treatments" OR "random intervention" OR "random interventions" OR "random comparison" OR "random comparisons" OR "randomised assignment" OR "randomised assignments" OR "randomised allocation" OR "randomised allocations" OR "randomised trial" OR "randomised trials" OR "randomised treatment" OR "randomised treatments" OR "randomised intervention" OR "randomised interventions" OR "randomised comparison" OR "randomised comparisons" OR "randomized

alma economics

Database	Search strings
	assignment" OR "randomized assignments" OR "randomized allocation" OR "randomized allocations" OR "randomized trial" OR "randomized trials" OR "randomized treatment" OR "randomized treatments" OR "randomized intervention" OR "randomized interventions" OR "randomized comparison" OR "randomized comparisons" OR "randomized controlled trial" OR "randomized controlled trials" OR "pretest posttest" OR "pretest-posttest" OR "pre/post" OR "before-after" OR "difference-in-differences" OR "difference-in-difference" OR "diff-in-diff" OR "propensity score" OR "regression discontinuity" OR "RDD" OR "instrumental variable" OR "instrumental variables" OR "cohort study" OR "cohort studies" OR "control group" OR "control groups" OR "treatment group" OR "treatment groups" OR "impact evaluation" OR "causal analysis" OR "causal inference" OR "matching techniques" OR "covariate matching" OR "inverse probability weighting" OR "nearest neighbor matching" OR "nearest neighbour matching" OR "exact matching" OR "kernel matching" OR "interrupted time series" OR "synthetic control" OR "synthetic controls" OR "panel study"))
World Bank Open Knowledge Repository (character limit) Results: 28 Date of search: 29/10/2024 Filters: After the search is completed: Date: 2004-2024, Supported language: EN	("public sector" OR "local government" OR "local governments" OR "local authorities" OR "local authority" OR "government agencies" OR "public administration" OR "civil service" OR "civil servants" OR "government employees") AND ("controlled trial" OR "quasi-experimental" OR "propensity score" OR "randomized trial" OR RCT* OR "experimental design" OR "experimental study" OR "field experiment" OR "randomised trial") AND ("training" OR "coaching" OR "capacity building" OR "professional development" OR "mentoring" OR "accreditation" OR "knowledge acquisition" OR "knowledge transfer")

alma economics

Database	Search strings
Campbell collaboration* Results: 0 Date of search: 31/10/2024	(“controlled trial” OR "quasi-experimental" OR "randomized trial" OR RCT*)(training OR "coaching" OR "capacity building")(“public sector” OR "local government" OR “civil service” OR “civil servants”) site:www.campbellcollaboration.org
Cedefop* Results: 0 Date of search: 31/10/2024	(“controlled trial” OR "quasi-experimental" OR "randomized trial" OR RCT*)(training OR "coaching" OR "capacity building")(“public sector” OR "local government" OR “civil service” OR “civil servants”) site:www.cedefop.europa.eu
Google Scholar (character limit) Results: 50 (By default, as we committed to looking at first 50 results) Date of search: 25/10/2024 Filters: Date: 2004-2024, Type: Review articles	(“controlled trial” OR "quasi-experimental" OR "randomized trial" OR RCT*) AND (training OR "coaching" OR "capacity building") AND (“public sector” OR "local government" OR “civil service” OR “civil servants”)

Note: For databases marked with an asterisk (*), the search was conducted using Google Search by applying the search string while restricting results to the database’s URL. This approach was applied due to the limited functionality of the database’s built-in search function.

Appendix B: Evidence map

Table 5. Evidence map

Study	Setting and design	Participants	Interventions	Outcomes
An et al., 2019	RCT in Denmark	Mixed group - Leaders from schools, daycare centres, public tax offices, and private bank offices	A structured leadership training program focused on transformational and transactional leadership. It included classroom instruction, group discussions, action learning, experiential exercises, and recorded sessions with follow-up discussions for remote participants.	Skills: Transformational leadership behaviour, use of contingent verbal rewards, use of contingent pecuniary rewards
An et al., 2021	RCT in Denmark	Mixed group - Leaders from schools, daycare centres, public tax offices, and private bank offices	A structured leadership training program focused on transformational and transactional leadership. It included classroom instruction, group discussions, action learning, experiential exercises, and recorded sessions with follow-up discussions for remote participants.	Skills: Transformational leadership behaviour, use of contingent verbal rewards, use of contingent pecuniary rewards
An et al., 2022	RCT in Denmark	Mixed group - Leaders from schools, daycare centres, public tax offices, and private bank offices	A structured leadership training program focused on transformational and transactional leadership. It included classroom instruction, group discussions, action learning, experiential exercises, and recorded sessions with follow-up discussions for remote participants.	Skills: Transformational leadership behaviour, use of contingent verbal rewards, use of contingent pecuniary rewards (based on self-other agreement gap)

alma economics

Study	Setting and design	Participants	Interventions	Outcomes
Asteris, 2013	RCT in the United States	Operational delivery - Probation officers	A specialised training program for probation officers, where all participants received baseline training on motivational interviewing. After that, some participants attended group learning sessions, others received individual coaching with feedback, and a third group experienced both interventions.	Skill: Motivational interviewing skill acquisition
Bonta et al., 2011	RCT in Canada	Operational delivery - Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application.	Skills: Session structuring skills, relationship skills, behavioural techniques, cognitive techniques, effective correctional skills Work performance: Recidivism rates of individuals on probation
Bonta et al., 2019	RCT in Canada	Operational delivery - Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application.	Skills: Session structuring skills, relationship skills, behavioural techniques, cognitive techniques, effective correctional skills Work performance: Recidivism rates of individuals on probation

Study	Setting and design	Participants	Interventions	Outcomes
Borness et al., 2013	RCT in Australia	Tax - Employees of a large Australian public sector tax organisation	A workplace intervention where employees completed either online cognitive training with exercises targeting memory, attention, language, and problem-solving or an active control task, watching nature documentaries with quizzes.	Skills: Cognitive performance Productivity: Workplace productivity based on task completion time, case processing speed, and work quality ratings.
Bourgon and Gutierrez, 2012	RCT in Canada	Operational delivery - Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application.	Skills: Discussion of pro-criminal attitudes, use of cognitive intervention Work performance: Recidivism rate of individuals on probation
Brown et al., 2013	RCT in Canada	Mixed group – Employees from municipal, provincial, and federal governments, healthcare, law enforcement, and education	All participants received a performance coaching training. Following this, participants were divided into groups with different goal setting strategies (behavioural outcome or learning), where they discussed and committed to specific goals for applying the training content back in their workplace.	Skills: Maintenance, generalisation, self-efficacy

alma economics

Study	Setting and design	Participants	Interventions	Outcomes
Brown et al., 2016	Quasi-experimental in Canada	Mixed group – Employees from public administration, liquor control, municipal government, law enforcement	All participants received a one-day management development program focused on self-awareness and interpersonal skills. Following this, participants were divided into groups with different goal setting strategies (behavioural outcome, behavioural-specific, rank-ordered), where they discussed and committed to specific goals for applying the training content back in their workplace.	Skills: Self-efficacy, transfer, goal commitment
Brown, 2005	RCT in Canada	Unclear/Unreported - Managers in the federal and provincial government	All participants received a one-day self-awareness training. Following this, participants were divided into groups with different goal setting strategies (distal or proximal), where they discussed and committed to specific goals for applying the training content back in their workplace.	Skills: Self-efficacy, maintenance, generalisation
Cotabish and Robinson, 2012	RCT in the United States	Project delivery - Program administrators	A workplace intervention where all participants attended two one-day institutes on program evaluation. The treatment group then received additional one-on-one peer coaching to align programs with national standards, set evaluation goals, and implement best practices, with progress monitored throughout the year.	Knowledge: Evaluation knowledge

alma economics

Study	Setting and design	Participants	Interventions	Outcomes
De Grip and Sauermann, 2012	RCT in the Netherlands	Operational delivery - Private sector call centre employees	Call centre agents took part in a training program aimed at improving conversation techniques and reducing call handling time without compromising quality. The training included formal instruction, group discussions with team leaders, and supervised hands-on practice handling customer calls.	Productivity: Average handling time of a customer call
Dunlop et al., 2015	Quasi-experimental in the United Kingdom	Operational delivery - Local government inspectors	A workplace intervention where local government inspectors received training on the Primary Authority (PA) initiative. The program included one-day workshops with expert-led presentations, group work, and case-based exercises to enhance understanding of PA partnerships and risk-based regulation.	Knowledge: Level of understanding of regulatory policy
Haunstrup and Jensen, 2024	RCT in Denmark	Operational delivery - Public managers working at job centres	A transformational leadership training program combining classroom instruction with goal-oriented development dialogues. Just-in-time nudges, delivered via static or dynamic software, supported leaders in applying learnt behaviours in real-time workplace interactions.	Skills: Transformational leadership behaviour
Itzchakov, 2020	This paper concerns two studies: Study 1: Quasi-experimental in	Study 1: Operational delivery - Customer service employees at a	Study 1: An intervention focused on listening training for customer service employees. It included lectures, exercises, discussions, and role-plays to improve active listening, attention focus, and handling difficult conversations.	Study 1 + 2: Skills: Listening perception, perspective-taking Study 1: Skills: Sense of competence

Study	Setting and design	Participants	Interventions	Outcomes
	United States Study 2: Quasi-experimental in Israel	Fortune 500 company Study 2: Operational delivery - Customer service employees at a nursing service	Study 2: A multi-session listening training for employees in a nursing services company. It focused on perspective-taking, managing emotional reactions, and applying the listening circle method through practice and reflection.	
Jacobsen et al., 2022	RCT in Denmark	Mixed group - Leaders from schools, daycare centres, public tax offices, and private bank offices	A structured leadership training program focused on transformational and transactional leadership. It included classroom instruction, group discussions, action learning, experiential exercises, and recorded sessions with follow-up discussions for remote participants.	Skills: Transformational leadership behaviour, use of contingent verbal rewards, use of contingent pecuniary rewards
Labrecque and Viglione, 2021	Quasi-experimental in the United States	Operational delivery - Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application.	Work performance: Positive drug test, technical violation, probation revocation, new arrest of individuals on probation

alma economics

Study	Setting and design	Participants	Interventions	Outcomes
Lowenkamp et al., 2014	Quasi-experimental in the United States	Operational delivery - Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application.	Work performance: Recidivism rates for individuals on probation
Milič Kavčič et al., 2024	RCT in Slovenia	Operational delivery - Case managers of the Pension and Disability Insurance Institute	A self-administered computerised cognitive training program for older office workers, consisting of virtual maze navigation sessions.	Productivity: Work productivity (total normed procedures completed per period)
Robinson et al., 2012	RCT in the United States	Operational delivery -Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application.	Skills: Role clarification use, use of reinforcement, disapproval, and effective authority; discussing cognitions, peers, or coping skills, use of cognitive techniques. Work performance: Failure on supervision (pretrial clients), arrest for new criminal behaviour (post-trial client)

alma economics

Study	Setting and design	Participants	Interventions	Outcomes
Rojas Méndez and Scartascini, 2024	RCT in Latin America (Andean countries, Mexico and Central America, Southern Cone)	Policy – Policymakers	A self-paced online behavioural economics training for policymakers, covering cognitive biases, decision-making, and policy applications through interactive modules, case studies, and assessments.	Skills: Policy decision-making skills
Salmela-Aro et al., 2012	RCT in Finland	Mixed group - Participants belonged to a range of organisations, including government research institutes, city departments, job centres, and others in the private sector.	A group-based career preparedness intervention led by HR and occupational health trainers, focusing on career management, goal setting, stress management, and resilience-building through active learning and social support.	Skills: Career preparedness

alma economics

Study	Setting and design	Participants	Interventions	Outcomes
Sauermann, 2021	RCT in the Netherlands	Operational delivery - Private sector call centre employees	Call centre agents took part in a training program aimed at improving conversation techniques and reducing call handling time without compromising quality. The training included formal instruction, group discussions with team leaders, and supervised hands-on practice handling customer calls.	Productivity: Average handling time of a customer call
Seidle et al., 2016	Quasi-experimental in the United States	Other civil service profession - Leaders (i.e., branch/ division/project managers) in the Department of Defense	A leadership training program to enhance leader performance and organisational effectiveness. It combined coaching, classroom instruction, feedback, and experiential training, with structured exercises and multisource feedback to improve self-awareness and decision-making.	Work performance: Leader performance, organisational effectiveness
Viglione and Labreque, 2021	Quasi-experimental in the United States	Operational delivery - Probation officers	A specialised training program for probation officers focused on cognitive-behavioural techniques, delivered face-to-face through multi-day training, monthly meetings, and refresher courses. The program combined didactic instruction, group exercises, and interactive role-plays to enhance learning and skill application. A policy mandate was later introduced, requiring trained officers to apply these skills at least eight times per month to improve implementation.	Skills: Overall use of STARR skills

Appendix C: Taxonomy

This appendix displays both the original BCT taxonomy, which was the starting point for the development of our taxonomy, and the results of our revised taxonomy.

Figure 14. Original BCT taxonomy

BCT Taxonomy (v1): 93 hierarchically-clustered techniques

Page	Grouping and BCTs	Page	Grouping and BCTs	Page	Grouping and BCTs
1	**1. Goals and planning**	8	**6. Comparison of behaviour**	16	**12. Antecedents**
	1.1. Goal setting (behavior) 1.2. Problem solving 1.3. Goal setting (outcome) 1.4. Action planning 1.5. Review behavior goal(s) 1.6. Discrepancy between current behavior and goal 1.7. Review outcome goal(s) 1.8. Behavioral contract 1.9. Commitment		6.1. Demonstration of the behavior 6.2. Social comparison 6.3. Information about others' approval		12.1. Restructuring the physical environment 12.2. Restructuring the social environment 12.3. Avoidance/reducing exposure to cues for the behavior 12.4. Distraction 12.5. Adding objects to the environment 12.6. Body changes
3	**2. Feedback and monitoring**	9	**7. Associations**	17	**13. Identity**
	2.1. Monitoring of behavior by others without feedback 2.2. Feedback on behaviour 2.3. Self-monitoring of behaviour 2.4. Self-monitoring of outcome(s) of behaviour 2.5. Monitoring of outcome(s) of behavior without feedback 2.6. Biofeedback 2.7. Feedback on outcome(s) of behavior		7.1. Prompts/cues 7.2. Cue signalling reward 7.3. Reduce prompts/cues 7.4. Remove access to the reward 7.5. Remove aversive stimulus 7.6. Satiation 7.7. Exposure 7.8. Associative learning		13.1. Identification of self as role model 13.2. Framing/reframing 13.3. Incompatible beliefs 13.4. Valued self-identify 13.5. Identity associated with changed behavior
5	**3. Social support**	10	**8. Repetition and substitution**	18	**14. Scheduled consequences**
	3.1. Social support (unspecified) 3.2. Social support (practical) 3.3. Social support (emotional)		8.1. Behavioral practice/rehearsal 8.2. Behavior substitution 8.3. Habit formation 8.4. Habit reversal 8.5. Overcorrection 8.6. Generalisation of target behavior 8.7. Graded tasks		14.1. Behavior cost 14.2. Punishment 14.3. Remove reward 14.4. Reward approximation 14.5. Rewarding completion 14.6. Situation-specific reward 14.7. Reward incompatible behavior 14.8. Reward alternative behavior 14.9. Reduce reward frequency 14.10. Remove punishment
6	**4. Shaping knowledge**	11	**9. Comparison of outcomes**	19	**15. Self-belief**
	4.1. Instruction on how to perform the behavior 4.2. Information about Antecedents 4.3. Re-attribution 4.4. Behavioral experiments		9.1. Credible source 9.2. Pros and cons 9.3. Comparative imagining of future outcomes		15.1. Verbal persuasion about capability 15.2. Mental rehearsal of successful performance 15.3. Focus on past success 15.4. Self-talk
7	**5. Natural consequences**	12	**10. Reward and threat**	19	**16. Covert learning**
	5.1. Information about health consequences 5.2. Salience of consequences 5.3. Information about social and environmental consequences 5.4. Monitoring of emotional consequences 5.5. Anticipated regret 5.6. Information about emotional consequences		10.1. Material incentive (behavior) 10.2. Material reward (behavior) 10.3. Non-specific reward 10.4. Social reward 10.5. Social incentive 10.6. Non-specific incentive 10.7. Self-incentive 10.8. Incentive (outcome) 10.9. Self-reward 10.10. Reward (outcome) 10.11. Future punishment		16.1. Imaginary punishment 16.2. Imaginary reward 16.3. Vicarious consequences
		15	**11. Regulation**		
			11.1. Pharmacological support 11.2. Reduce negative emotions 11.3. Conserving mental resources 11.4. Paradoxical instructions		

The table below outlines each mechanism, identified by its number and label within the BCT taxonomy. We followed the same numbering and groupings as the original BCT taxonomy, making it clear which mechanisms were removed. Any newly added mechanisms are labelled as such in the No. column and are highlighted in grey, and any adaptations to the definitions have been formatted in bold.

Table 6. Revised taxonomy

No.	Category	Label	Definition	Example
1.1	Goals and Planning	Goal setting (behaviour)	Set or agree on a goal defined in terms of the behaviour to be achieved. Note: Only code goal setting if there is sufficient evidence that goal set as part of intervention; if goal unspecified or a behavioural outcome, code 1.3, Goal setting (outcome); if the goal defines a specific context, frequency, duration, or intensity for the behaviour, also code 1.4, Action planning	Participants develop a set of goals on the specific initiatives to implement in their team or unit, particularly regarding how they perform goal-oriented development dialogues with their employees.
1.2	Goals and Planning	Problem solving	Analyse, or prompt the individual to analyse, factors influencing the behaviour and generate or select strategies that include overcoming barriers and/or increasing facilitators Note: barrier identification without solutions is insufficient.	Participants identify potential setbacks, such as missing deadlines due to competing priorities, and generate solutions for them.

No.	Category	Label	Definition	Example
1.3	Goals and Planning	Goal setting (outcome)	Set or agree on a goal defined in terms of a positive outcome of wanted behaviour Note: only code guidelines if set as a goal in an intervention context; if goal is a behaviour, code 1.1, Goal setting (behaviour); if goal unspecified code 1.3, Goal setting (outcome)	Participants set a specific, difficult goal for the Behavioural Observation Scale (BOS) score they wished to receive when their workplace observers assessed their usage of skills back at work using BOS forms.
1.4	Goals and Planning	Action planning	Prompt detailed planning of performance of the behaviour (must include at least one of context, frequency, duration, and intensity). Note: evidence of action planning does not necessarily imply goal setting, only code latter if sufficient evidence.	Participants are encouraged to create a detailed plan for conducting weekly team check-ins, specifying the time and day of the week and the length of each meeting.
1.7	Goals and Planning	Review outcome goal(s)	Review outcome goal(s) jointly with the person **or prompt the person to review their outcome goal(s)** and consider modifying goal(s) in light of achievement. This may lead to resetting the same goal, a small change in that goal, or setting a new goal instead of, or in addition to, the first.	Participants are asked to reflect on the experiences, successes, and failures in applying their leadership development plan, developed as part of the training programme, and review their outcome goals based on that.
1.9	Goals and Planning	Commitment	Ask the person to affirm or reaffirm statements indicating commitment to change the behaviour Note: if defined in terms of the behaviour to be achieved, also code 1.1, Goal setting (behaviour)	Participants are asked to publicly state their goals in discussion with their peers to increase commitment.

No.	Category	Label	Definition	Example
2.1	Feedback and Monitoring	Monitoring of behaviour by others without feedback	Observe or record behaviour with the person's knowledge as part of a behaviour change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behaviour, do not code; if feedback given, code only 2.2, Feedback on behaviour, and not 2.1, Monitoring of behaviour by others without feedback; if monitoring outcome(s) code 2.5, Monitoring outcome(s) of behaviour by others without feedback; if self-monitoring behaviour, code 2.3, Self-monitoring of behaviour.	Facilitators observe participants during role-play scenarios, with their knowledge, to track how they handle conflict resolution in workplace situations as part of improving interpersonal effectiveness.
2.2	Feedback and Monitoring	Feedback on behaviour	Monitor and provide informative or evaluative feedback on performance of the behaviour (e.g., form, frequency, duration, intensity). Note: If feedback is on outcome(s) of behaviour, code 2.7, Feedback on outcome(s) of behaviour; if there is no clear evidence that feedback was given, code 2.1, Monitoring of behaviour by others without feedback.	Facilitators observe participants in mock leadership meetings and provide feedback on their communication clarity, engagement frequency, and areas for improvement.

No.	Category	Label	Definition	Example
2.3	Feedback and Monitoring	Self-monitoring of behaviour	Encourage the person to monitor **or** record their behaviour(s) as part of a behaviour change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behaviour, do not code; if monitoring of outcome of behaviour, code 2.4, Self-monitoring of outcome(s) of behaviour; if monitoring is by someone else (without feedback), code 2.1, Monitoring of behaviour by others without feedback.	Participants are asked to keep a daily log of how they manage their time, noting tasks completed and interruptions, to identify patterns and improve time management skills.
2.4	Feedback and Monitoring	Self-monitoring of outcome(s) of behaviour	Establish a method for the person to monitor and record the outcome(s) of their behaviour as part of a behaviour change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behaviour, do not code ; if monitoring behaviour, code 2.3, Self-monitoring of behaviour; if monitoring is by someone else (without feedback), code 2.5, Monitoring outcome(s) of behaviour by others without feedback.	Participants are provided with a template to track the completion rates of their assigned projects over a month, enabling them to evaluate their productivity and identify areas for improvement.

No.	Category	Label	Definition	Example
2.5	Feedback and Monitoring	Monitoring of outcome(s) of behaviour without feedback	Observe or record outcomes of behaviour with the person's knowledge as part of a behaviour change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behaviour, do not code; if feedback given, code only 2.7, Feedback on outcome(s) of behaviour; if monitoring behaviour code 2.1, Monitoring of behaviour by others without feedback; if self-monitoring outcome(s), code 2.4, Self monitoring of outcome(s) of behaviour.	Progress toward the goals identified by participants in the training programme is monitored by programme facilitators.
2.7	Feedback and Monitoring	Feedback on outcome(s) of behaviour	Monitor and provide feedback on the outcome of performance of the behaviour. Note: Feedback on outcome(s) of behaviour; if feedback is on behaviour code 2.2, Feedback on behaviour; if there is no clear evidence that feedback was given code 2.5, Monitoring outcome(s) of behaviour by others without feedback; if feedback on behaviour is evaluative e.g. praise, also code 10.4, Social reward.	Employees managed by participants complete a survey providing feedback on whether their leaders set a clear vision, align them with organisational goals, clarify contributions to achieving those goals, and foster cooperation among employees.

No.	Category	Label	Definition	Example
NEW	Feedback and Monitoring	Feedback (unspecified)	Monitor and provide feedback related to the behaviour (general). Note: used in situations where detailed information on whether the feedback focused on behaviour or outcome of behaviour.	Participants of a training programme are provided with multisource feedback six months after completing the programme.
3.1	Social support	Social support (unspecified)	Advise on, arrange, or provide social support (e.g., from friends, relatives, colleagues,' buddies' or staff) or non-contingent praise or reward for the performance of the behaviour. It includes encouragement and counselling, but only when it is directed at the behaviour. Note: attending a group class and/or mention of 'follow-up' does not necessarily apply this BCT; support must be explicitly mentioned; if practical, code 3.2, Social support (practical).	Participants shared their goals with goal setting peers, who provided feedback on these goals.

No.	Category	Label	Definition	Example
3.2	Social support	Social support (practical)	Advise on, arrange, or provide practical help (e.g., from friends, relatives, colleagues, 'buddies', or staff) for performance of the behaviour. Note: if emotional, code 3.3, Social support (emotional); if general or unspecified, code 3.1, Social support (unspecified) If only restructuring the physical environment or adding objects to the environment, code 12.1, Restructuring the physical environment or 12.5, Adding objects to the environment; attending a group or class and/or mention of 'follow up' does not necessarily apply this BCT, support must be explicitly mentioned.	As part of a professional development training, a mentor arranges regular check-ins with a participant to offer guidance on how to improve project management skills and provide feedback on their progress.
NEW	Social support	Social support (collaboration)	Facilitate or encourage mutual engagement and shared responsibility among individuals (e.g., peers, colleagues, or group members) to support and encourage each other in performing the behaviour. This includes collaborative activities, such as group discussions, knowledge exchanges, joint problem-solving, and co-developing strategies for behaviour change.	During a team-based training session, participants are grouped to collaborate on solving a case study, encouraging them to share insights, support each other's ideas, and co-develop strategies for improving efficiency in their department.

No.	Category	Label	Definition	Example
4.1	Shaping knowledge	Instruction on how to perform the behaviour	Advise or agree on how to perform the behaviour (includes 'Skills training'). Note: when the person attends classes, such as exercise or cookery, code 4.1, Instruction on how to perform the behaviour, 8.1, behavioural practice/rehearsal and 6.1, Demonstration of the behaviour.	As part of a professional development workshop, a facilitator provides step-by-step guidance on how to conduct effective performance appraisals, including techniques for delivering constructive feedback.
6.1	Comparison of Behaviour	Demonstration of the behaviour	Provide an observable sample of the performance of the behaviour, directly in person or indirectly, e.g., via film, pictures, for the person to aspire to or imitate (includes 'Modelling'). Note: if advised to practice, also code, 8.1, Behavioural practice and rehearsal; If provided with instructions on how to perform, also code 4.1, Instruction on how to perform the behaviour.	Training programme facilitators role model important learning aspects, for instance, showing by example how confidentiality is developed and maintained (by sharing information on their own workplace and illustrating how a translation process would look in their own context).
6.3	Comparison of Behaviour	Information about others' approval	Provide information about what other people think about the behaviour. The information clarifies whether others will like, approve of, or disapprove of what the person is doing or will do.	In a development workshop, a participant is informed that their team members have expressed appreciation for their recent efforts in improving meeting efficiency, which encourages them to continue using the new techniques.
7.1	Association	Prompts/cues	Introduce or define	Following the training

No.	Category	Label	Definition	Example
	s		environmental or social stimulus with the purpose of prompting or cueing the behaviour. The prompt or cue would normally occur at the time or place of performance. Note: when a stimulus is linked to a specific action in an if-then plan, including one or more of frequency, duration or intensity, also code 1.4, Action planning.	programme, participants received regular emails from the programme facilitators reminding them of tips and providing advice covered during the training.
8.1	Repetition and substitution	Behavioural practice/rehearsal	Prompt practice or rehearsal of the performance of the behaviour one or more times, in a context or at a time when the performance may not be necessary, in order to increase habit and skill. Note: if aiming to associate performance with the context, also code 8.3, Habit formation.	As part of a training program, participants are asked to role-play conflict resolution scenarios during a practice session to build their skills and increase confidence in handling such situations in the future.
8.2	Repetition and substitution	Behaviour substitution	Prompt substitution of the unwanted behaviour with a wanted or neutral behaviour.	Participants are encouraged to replace procrastination with a practice of setting short, focused work intervals followed by brief breaks to help improve productivity.

No.	Category	Label	Definition	Example
8.3	Repetition and substitution	Habit formation	**Encourage practice and repetition of the behaviour within the same context, including through booster training sessions, to strengthen the association between the context and the behaviour.** Note: also code 8.1, behavioural practice/rehearsal.	Participants are asked to repeatedly practice delivering clear and concise messages during team meetings, so that the act of speaking in meetings becomes a trigger for using these skills effectively.
8.7	Repetition and substitution	Graded tasks	Set easy-to-perform tasks, making them increasingly difficult, but achievable, until behaviour is performed.	Participants are asked to conduct a cognitive training programme, which consists of maze exercises with increasing difficulty.
10.3	Reward and threat	Non-specific reward	Arrange delivery of a reward if and only if there has been effort and/or progress in performing the behaviour (includes 'Positive reinforcement').	Employees are able to participate in further professional development training if they demonstrate the ability to apply the skills covered in the initial training programme.

No.	Category	Label	Definition	Example
10.6	Reward and threat	Non-specific incentive	Inform that a reward will be delivered if and only if there has been effort and/or progress in performing the behaviour (includes 'Positive reinforcement'). Note: if incentive is material, code 10.1, Material incentive (behaviour), if social, code 10.5, Social incentive and not 10.6, Non-specific incentive; if incentive is for outcome, code 10.8, Incentive (outcome). If reward is delivered, also code one of: 10.2, Material reward (behaviour); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).	Before participating in the training programme, employees are informed that if, following completion, they are able to demonstrate the skills covered, they will be eligible to participate in another, more advanced, professional development programme.
13.2	Identity	Framing/reframing	Suggest the deliberate adoption of a perspective or new perspective on behaviour (e.g., its purpose) in order to change cognitions or emotions about performing the behaviour (includes 'Cognitive structuring').	When introducing employees to the leadership development program, it is framed as an essential component of the organisation's broader business strategy, highlighting how their participation and growth directly support the organisation's objectives.

No.	Category	Label	Definition	Example
14.5	Scheduled Consequences	Rewarding completion	Build up behaviour by arranging reward following final component of the behaviour; gradually add the components of the behaviour that occur earlier in the behavioural sequence (includes 'Backward chaining').	In a customer service training, participants are initially rewarded for successfully completing the final step of resolving customer complaints. Gradually, earlier steps – such as acknowledging the complaint and gathering information – are added to the sequence, with rewards given once the entire process is completed effectively.
16.3	Covert Learning	Vicarious consequences	Prompt observation of the outcomes experienced by others when they perform this behaviour (e.g., the positive and negative consequences they encounter as a result of performing the behaviour).	During a leadership training session, participants watch a video of a colleague effectively managing a difficult conversation with a team member, highlighting the positive outcomes, such as improved team morale and respect, to encourage similar behaviour.

Appendix D: Risk of bias in studies

Table 7. Risk of bias assessment scores for RCTs

Study	Risk of bias assessment score
An et al., 2019	23%
An and Meier, 2021	38%
An et al., 2022	23%
Asteris, 2013	31%
Bonta et al., 2011	54%
Bonta et al., 2019	31%
Borness et al., 2013	54%
Bourgon and Gutierrez, 2012	54%
Brown et al., 2013	46%
Brown, 2005	46%
Cotabish and Robinson, 2012	46%
De Grip and Sauermann, 2012	54%
Haunstrup and Jensen, 2024	38%
Jacobsen et al., 2022	23%
Milič Kavčič et al., 2024	54%
Robinson et al., 2012	46%
Rojas Méndez and Scartascini, 2024	31%
Salmela-Aro et al., 2012	46%
Sauermann, 2021	75%

Table 8. Risk of bias assessment scores for quasi-experimental studies

Study	Risk of bias assessment score
Brown et al., 2016	56%
Dunlop et al., 2015	63%
Itzchakov, 2020 (Israel study)	78%
Itzchakov, 2020 (US study)	67%
Labrecque and Viglione, 2021	88%
Lowenkamp et al., 2014	56%
Seidle et al., 2016	78%
Viglione and Labrecque, 2021	75%

Appendix E: Results of individual studies

Table 9 displays the following information for each paper where available: (i) sample size, (ii) the main outcome (if multiple, see our discussion in the Dependent effect sizes section), (iii) the type of outcome measured (binary/continuous), (iv) the treatment group (if multiple), (v) the effect size type (Hedges' g if mean differences, standard errors, and sample sizes are not missing), (vi) the effect direction, (vii) the effect size, (viii) the standard error of the effect size, (ix) the p-value, (x) the relevant research question, and (xi) whether the paper was included in the vote counting exercise. For (xi), papers are either not included in the vote counting exercise because either they are not independent studies or because they do not contribute to the overarching research question 2, where (x) clearly states which research questions they help to answer. Any statistics calculated based on information available in the paper rather than directly extracted from the paper are formatted in bold.

Table 9. Results of individual studies

Study	Sample size	Main outcome	Type of outcome measures	Treatment group	Effect size type	Effect direction	Effect size	Standard error	P-value	Relevant research question	Entered vote counting?
An et al., 2019	N/A	Transformational Leadership Behaviour	Continuous	Combination	Interaction Effect	N/A	2.67	3.30	**0.42**	RQ 2b	No
An et al., 2021	275	Transformational Leadership Behaviour	Continuous	Combination	Interaction Effect	N/A	2.70	2.08	**0.19**	RQ 2b	No
An et al., 2022	N/A	SOA (self-other agreement) gap for transformational leadership	Continuous	Combination	Regression Coefficient	Favours intervention	-0.02	0.09	**0.86**	RQ 2	No, not independent
Asteris, 2013	12	MITI 3.1.1 - Motivational interviewing (MI) skill acquisition	Continuous	Combination	**Mean Difference**	Favours intervention	**4.67**	N/A	N/A	RQ 2 + RQ 3a	Yes

Study	Sample size	Main outcome	Type of outcome measures	Treatment group	Effect size type	Effect direction	Effect size	Standard error	P-value	Relevant research question	Entered vote counting?
Bonta et al., 2011	Officers: 52, Ind on probation (retrospective): 185 (prospective): 112	Recidivism rate	Binary	1 treatment group	**Hedges' g**	Favours intervention	**-0.38**	**0.23**	**0.10**	RQ 2	Yes
Bonta et al., 2019	Officers:27, Ind on probation: 120	Recidivism rate	Binary	1 treatment group	**Hedges' g**	Does not favour intervention	0.07	**0.21**	**0.75**	RQ 2	Yes
Borness et al., 2013	88	Workplace Productivity	Continuous	1 treatment group	N/A	N/A	N/A	N/A	N/A	RQ 2	Yes
Bourgon and Gutierrez, 2012	Officers: 52, Ind on probation: 142	Recidivism rate	Binary	1 treatment group	N/A	N/A	N/A	N/A	N/A	RQ 2	No, not independent
Brown et al., 2013	N/A	Self-efficacy	Continuous	Behavioural outcome goals	**Mean Difference**	Does not favour intervention	**-0.41**	N/A	N/A	RQ 3c	No
-	N/A	Self-efficacy	Continuous	Learning goals	**Mean Difference**	Does not favour intervention	**-2.70**	N/A	N/A	RQ 3c	No
Brown et al., 2016	N/A	Self-efficacy	Continuous	Behavioural outcome goals	**Mean Difference**	Does not favour intervention	**-7.80**	N/A	N/A	RQ 3c	No
-	N/A	Self-efficacy	Continuous	Behavioural specific goals	**Mean Difference**	Does not favour intervention	**-6.01**	N/A	N/A	RQ 3c	No

Study	Sample size	Main outcome	Type of outcome measures	Treatment group	Effect size type	Effect direction	Effect size	Standard error	P-value	Relevant research question	Entered vote counting?
-	N/A	Self-efficacy	Continuous	Rank-ordered behavioural goals	**Mean Difference**	Does not favour intervention	**-4.19**	N/A	N/A	RQ 3c	No
Brown, 2005	33	Self-efficacy	Continuous	Proximal + distal goals	**Hedges' g**	Does not favour intervention	**-0.36**	**0.34**	**0.29**	RQ 3c	No
-	33	Self-efficacy	Continuous	Distal outcome goals	**Hedges' g**	Does not favour intervention	**-0.22**	**0.34**	**0.51**	RQ 3c	No
Cotabish and Robinson, 2012	78	Content Assessment (Evaluation Knowledge and Skills)	Continuous	1 treatment group	**Hedges' g**	Favours intervention	**0.44**	**0.23**	**0.06**	RQ 2 + RQ 3a	Yes
De Grip and Sauermann, 2012	63	Performance	Continuous	1 treatment group	**Hedges' g**	Favours intervention	**0.79**	**0.26**	**0.00**	RQ 2	Yes
Dunlop et al., 2015	172	Understanding of Primary Authority initiative - Vignette 1	Binary	1 treatment group	**Hedges' g**	Favours intervention	**0.07**	**0.18**	**0.72**	RQ 2	Yes
Haunstrup and Jensen, 2024	140	Transformational Leadership Behaviour	Continuous	Nudges (Static tool)	Regression Coefficient	Favours intervention	0.34	0.08	**0.00**	RQ 2 + RQ 3a	Yes
-	133	Transformational Leadership Behaviour	Continuous	Nudges (Dynamic tool)	Regression Coefficient	Favours intervention	0.28	0.09	**0.00**	RQ 2 + RQ 3a	No

Study	Sample size	Main outcome	Type of outcome measures	Treatment group	Effect size type	Effect direction	Effect size	Standard error	P-value	Relevant research question	Entered vote counting?
Itzchakov, 2020 (Israel Study)	33	(Israel) Perspective-taking	Continuous	1 treatment group	**Hedges' g**	Favours intervention	**0.58**	**0.35**	**0.10**	RQ 2	Yes
Itzchakov, 2020 (U.S. Study)	61	(U.S.) Perspective-taking	Continuous	1 treatment group	**Hedges' g**	Favours intervention	**0.42**	**0.26**	**0.10**	RQ 2	Yes
Jacobsen et al., 2022	234	Transformational Leadership Behaviour	Continuous	Combination	Regression coefficient	Favours intervention	3.47	**1.21**	0.00	RQ 2 + RQ 3a	Yes
Labrecque and Viglione, 2021	Officers: 65, Ind on probation: 1444	New arrest	Binary	STARR coaches (more advanced)	**Hedges' g**	Favours intervention	**-0.08**	**0.21**	**0.71**	RQ 2 + RQ 3a	Yes
	Officers: 75, Ind on probation: 1568	New arrest	Binary	STARR users	**Hedges' g**	Favours intervention	**-0.09**	**0.12**	**0.44**	RQ 2 + RQ 3a	No
Lowenkamp et al., 2014	Ind on probation: 220	Failure rates	Binary	STARR training + MI (12 months)	**Hedges' g**	Favours intervention	**-0.37**	**0.18**	**0.03**	RQ 2	No, not independent
	Ind on probation: 186	Failure rates	Binary	STARR training + MI (24 months)	**Hedges' g**	Favours intervention	**-0.29**	**0.16**	**0.07**	RQ 2 + RQ 3a	No
	Ind on probation: 186	Failure rates	Binary	STARR training (24 months)	**Hedges' g**	Favours intervention	**-0.32**	**0.17**	**0.07**	RQ 3a	No
Milič Kavčič et al., 2024	42	Work productivity	Continuous	1 treatment group	**Hedges' g**	Favours intervention	**0.37**	**0.31**	**0.23**	RQ 2	Yes

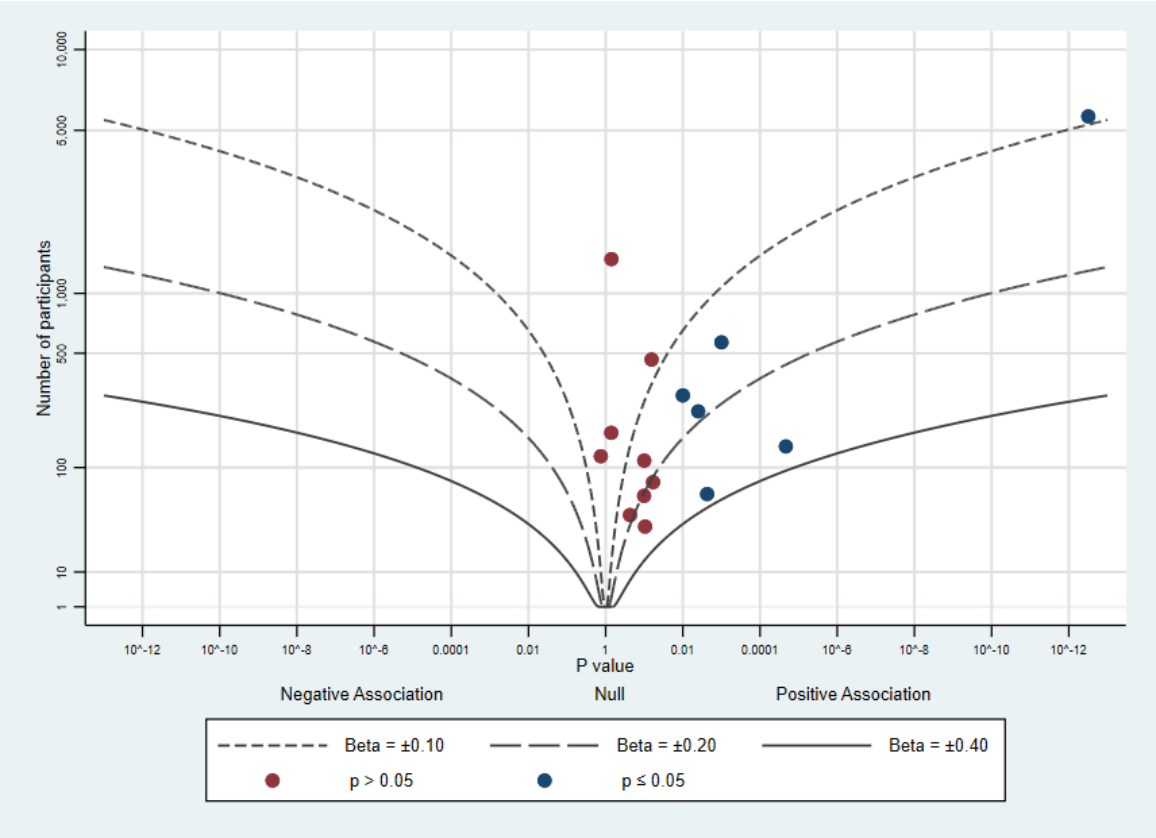
Study	Sample size	Main outcome	Type of outcome measures	Treatment group	Effect size type	Effect direction	Effect size	Standard error	P-value	Relevant research question	Entered vote counting?
Robinson et al., 2012	Officers: 59, Ind on probation: 462	Failure on supervision (pretrial clients), arrest for new criminal behaviour (post-trial client)	Binary	1 treatment group	**Hedges' g**	Favours intervention	**-0.21**	**0.11**	**0.06**	RQ 2	Yes
Rojas Méndez and Scartascini, 2024	5655	Policy decision-making skills	Continuous	1 treatment group	Regression coefficient	Favours intervention	0.60	0.03	**0.00**	RQ 2	Yes
Salmela-Aro, 2012	570	Career preparedness	Continuous	1 treatment group	Regression Coefficient	Favours intervention	0.21	N/A	0.00	RQ 2 + RQ 3a	Yes
Sauermann, 2021	63	Performance	Continuous	1 treatment group	Interaction Effect	N/A	0.02	0.01	**0.01**	RQ 2b	No
Seidle et al., 2016	291	Leadership performance	Continuous	1 treatment group	Regression coefficient	Favours intervention	1.36	N/A	<0.01	RQ 2	Yes
Viglione and Labrecque, 2021	48	All STARR use	Continuous	1 treatment group	**Hedges' g**	Favours intervention	**1.74**	**0.24**	**0.00**	RQ 3a	No

Appendix F: Albatross plot

One tool commonly used in narrative synthesis is an albatross plot. Figure 15 presents the albatross plot for the studies relevant to research question 2. The albatross plot displays p-values and sample sizes, with points further from the centre on the x-axis indicating lower p-values, suggesting a lower likelihood that the observed associations occurred by chance. Positive associations are shown on the right, while negative associations appear on the left. Points in blue represent statistically significant results with a p-value of 0.05 or below, while points in red indicate non-significant results. For context, the contour lines indicate the relationship between p-value and sample size that would hold in a hypothetical setting, whereby the dependent and independent variables are both standardised (i.e., divided by their respective standard deviations), for different effect sizes (indicated by beta).

Of the 17 papers identified, 15 either provided the necessary information to calculate the p-value or reported it directly. The plot again shows that only one study favours the control group, while all others have positive associations with the intervention. Out of the 15 studies in the plot, six report statistically significant effects ($p \leq 0.05$). There is no clear pattern visible between the number of participants and statistical significance, though it would be hard to confidently detect such a pattern with the low number of papers that we have.

Figure 15. Albatross plot showing relationship between p-values of results and number of participants in papers



Appendix G:
Best practice example of abstract

Best practice example of abstract included in this review (Borness et al., 2013)

Abstract

Background: Cognitive training (CT) is effective at improving cognitive outcomes in children with and without clinical impairment as well as older individuals. Yet whether CT is of any preventative health benefit to working age adults is controversial. Our objective was therefore to investigate the real-world efficacy of CT in the workplace, involving employees from across the working-age spectrum and addressing many of the design issues that have limited trials to date.

Methods and Findings: 135 white collar employees of a large Australian public sector organization were randomised to either 16 weeks (20 minutes three times per week) of online CT or an active control (AC) program of equal length and structure. Cognitive, wellbeing and productivity outcome measures were analysed across three timepoints: baseline, immediately after training and 6 months post-training. CT effects on cognitive outcomes were limited, even after planned subgroup analyses of cognitive capacity and age. Unexpectedly, we found that our AC condition, which comprised viewing short documentaries about the natural world, had more impact. Compared to the CT group, 6 months after the end of training, those in the AC group experienced a significant increase in their self-reported Quality of Life (Effect Size g = .34 vs 2.15; TIME6GROUP p = .003), decrease in stress levels (g = .22 vs 2.19; TIME x GROUP p = .03), and overall improvement in Psychological Wellbeing (g = .32 vs 2.06; TIME6GROUP p = .02).

Conclusions: CT does not appear to positively impact cognition or wellbeing amongst white collar office workers; however, short time-out respite activities may have value in the promotion of psychological wellbeing. Given looming challenges to workplace productivity, further work-based interventional research targeting employee mental health is recommended.

Trial Registration: This trial was registered with the Australian New Zealand Clinical Trials Registry: ACTRN12610000604000 (http://www.anzctr.org.au/TrialSearch.aspx).

+44 20 8133 3192 43 Tanner Street, SE1 3PL, London, UK

LinkedIn company/alma-economics almaeconomics

alma economics