

ARTICLES

From Concepts to Measures: A Mixed Methods Framework for Measuring Latent Constructs

Srijeeta Mitra[†], Ujjayini Das[†], Tejjwansh S. Anand², Bertrand A. Stoffel²

¹ Joint Program in Survey Methodology, University of Maryland, College Park, ² Robert H. Smith School of Business, University of Maryland, College Park

Keywords: Measurement framework development and validation, Measurement of latent constructs, Exploratory sequential mixed methods design, Nonprobability sample, Mass imputation

<https://doi.org/10.29115/SP-2026-0008>

Survey Practice

Vol. 21 Special Issue, 2026

Mixed methods research offers survey researchers a way to link conceptual development with rigorous measurement and inference, especially when the interest lies in underspecified domains. In this article, we implement an exploratory sequential mixed methods design that moves in a structured way from latent concepts to measured and tested constructs. We begin with a targeted literature review and semi-structured expert interviews to identify and refine key dimensions of a latent complex topic. We use these qualitative insights to develop a provisional multidimensional framework and to create survey questions informed by practitioner-used vocabulary. We then conduct a web survey to validate the developed framework. The web survey was administered on a small non-probability sample due to constrained resources, which is very common in practice. Although it apparently hinders our ability to draw generalizable conclusions for the target population, we implement statistical adjustments to leverage the data we collected. We address the selection bias in our sample by linking it to a large reference probability survey through a mass imputation approach, modeling study outcomes as a function of shared demographic variables and projecting them onto the probability sample. Using the mass imputed outcome in the reference sample, we further conduct confirmatory factor analysis to determine whether there is empirical evidence to support the conceptual framework we developed at the qualitative phase of the research. While our specific interest is focused on measurement of impact of artificial intelligence in federal government, we present this research as a practical example of how survey researchers can integrate qualitative and quantitative evidence to develop and validate new and latent constructs along with handling small, non-probability samples to produce externally valid conclusions.

1. Introduction

Survey researchers are increasingly asked to address questions that are conceptually complex and empirically dynamic. Projects often begin with broad theoretical or policy motivations but quickly require multiple forms of

[†] Joint first authors

evidence to develop and test those ideas. Designing a research process that coherently integrates literature, qualitative insights, and quantitative evidence can therefore be as important as the findings themselves. In this article, we present a methodological roadmap for building that connection, moving systematically from a literature review to qualitative exploration and then to quantitative validation. This roadmap is intended as a transferable approach that other researchers can adapt to their own substantive questions.

The approach developed here consists of two complementary components that together expand the methodological toolbox for survey research. The first focuses on translating themes derived from literature and qualitative interviews into measurable survey constructs. This stage demonstrates how qualitative insights can inform instrument design, operationalize abstract concepts, and strengthen content validity before data collection. The second component addresses a persistent challenge in survey research: drawing credible and externally valid conclusions from small convenience samples. By benchmarking survey results against an existing representative reference sample, we show when and how carefully structured statistical analysis can improve generalizability even when probability-based data collection is not feasible.

Although we illustrate this approach using an applied study of organizational change in the federal government, the contribution of this paper is primarily methodological. Throughout the article, we outline transferable procedural steps, data integration strategies, and analytic checks for research settings where surveys intersect with qualitative inquiry or sampling options are limited.

2. Research Design

We employed a three-phase mixed methods design intended to connect conceptual construct development with empirical validation. The goal was to illustrate how researchers can move from identifying themes in existing knowledge to developing and testing measurable constructs using qualitative and quantitative techniques.

First, we conducted an extensive literature review to conceptualize the focal construct and identify dimensions that potentially define it, using relevant keywords. Because such constructs are often multidimensional and interpreted differently across disciplines, we reviewed both academic and gray literature. This step enabled us to synthesize existing perspectives, identify conceptual gaps, and outline an initial set of subconstructs and relationships that then guided our interview guide and our thinking about a framework to measure the primary latent construct.

Second, we refined and operationalized the framework derived from theory. Through semi-structured interviews with subject matter experts (SMEs), we explored how the construct manifests in real-world settings, how practitioners

understand its dimensions, and what contextual factors shape its interpretation. These qualitative insights allowed us to revise, expand, and structure the framework in a way grounded in both lived experience and theoretical guidance.

Third, we conducted a quantitative validation study to empirically assess the framework. Developing a structured web survey from the refined dimensions and indicators, we collected data from individuals who match our target population. The purpose of this phase was to examine the reliability and validity of the proposed framework and to determine whether the theorized dimensions coherently capture the underlying latent construct.

Together, these three components: a theory-driven literature review, qualitative interview for framework development, and a web survey for quantitative validation, form a coherent mixed-methods strategy that enables the rigorous conceptualization, construction, and empirical assessment of a latent construct that would otherwise be difficult to measure directly.

In this study, we applied this mixed-methods approach to examine how artificial intelligence (AI) is understood, governed, and integrated within U.S. federal agencies. We first synthesized existing academic and gray literature on AI adoption in federal agencies to identify the core dimensions (Fluency, Adoption, Adaptation) relevant to assessing its organizational impact. We then refined these dimensions through qualitative interviews with SMEs involved in federal AI initiatives. Finally, we developed and administered a web-based survey to validate the resulting framework and evaluate perceptions around operational and organizational effects of AI across the federal population.

3. Qualitative Phase: Literature Review and Expert Interviews

In the qualitative phase, we focused on clarifying key concepts and developing a framework that we could later test with survey data. Our goal in this phase was not to produce generalizable estimates, but to move in a structured way from existing work to concrete dimensions and candidate measures that would define our latent construct of interest.

3.1. Literature Review

We began with a targeted review of academic and gray¹ literature around adoption of AI in the public sectors and landscape of AI usage in the government using a standard set of relevant keywords as search terms and iteratively adding more terms as they emerged since we were interested in measuring impact of AI on federal agencies. We used Google Scholar, Scopus, Policy Commons and SSRN along with conventional web-based

¹ Literature published outside peer-reviewed channels mostly by corporate organizations, think tanks and industry researchers.

search engines (Google and Bing) to trace the literature. We recruited graduate assistants to code the key findings from the identified literature into a corpus, that we later synthesized to identify the core themes around our primary topic of interest – impact of AI adoption in public and private sectors. We narrowed down to four core themes that can shape the idea of ‘impact’ in our case. See Appendix for details. The resulting narrative review highlighted the importance of measuring the impact of AI adoption on government agencies more holistically as opposed to focusing on segregated aspects of AI adoption. Thus, we developed a theoretical framework that connects these identified themes overall.

3.2. Expert Interviews

Next, we conducted semi-structured interviews with SMEs to refine and operationalize the themes we saw in the literature. We interviewed 9 Chief Information Officers (CIO) in federal government positions who had direct experience with the development, implementation, or oversight of these tools in federal or related settings. We recruited participants using snowball sampling, starting from contacts in our professional networks and asking interviewees to recommend additional experts. Interviews took place between Fall of 2022 and Summer of 2023, were conducted by two of the authors, lasted about 45 – 60 minutes and were held via videoconferencing. We used an interview guide focusing on specific projects within participants’ agencies that involved AI-based solutions as case studies. The guide elicited detailed information on their roles in planning and implementation, the challenges encountered, and their overall perceptions of the use of AI within their agencies (full guide in Appendix).

We audio-recorded all interviews with participants’ consent and then transcribed them. We analyzed the transcripts using a thematic approach. We first created a set of deductive codes based on the concepts from the literature, and we then added inductive codes that emerged as we read the interviews. Two members of the research team worked together on the codebook, double-coded an initial set of transcripts, and resolved differences through independent inputs from a third member of the research team. We then grouped related codes into broader themes that helped us see how the literature-based concepts played out in practice and where experts’ views suggested adjustments or additions.

We used the combined insights from literature and the interviews to define a small set of core dimensions that describe our latent construct. We then arranged these dimensions into a provisional framework that specifies how they relate to one another. For each dimension, we identified subdimensions

and concrete indicators that could be asked about in a survey², and we drafted items that reflected the language experts used in the interviews. The specific details regarding the definitions of the dimensions, and the conceptualized framework in our example can be found in the Appendix.

Our qualitative sample was small: we interviewed 9 experts due to resource constraints. This limits what we can say about how common any specific view is in the wider expert community. Studies on thematic saturation (Hennink and Kaiser 2022; Hennink et al. 2017) suggest that samples of this size are often sufficient to identify many of the main themes in a relatively focused domain, but they are less suited to fully exploring all nuances or rare perspectives. For this reason, we treated these interviews as an exploratory step that helps us define and organize key dimensions, rather than as a standalone source for generalizable conclusions.

The interviews still played an important role in our design. They allowed us to connect ideas from literature to the language and experiences of practitioners, and they helped us turn broad concepts into concrete dimensions and candidate indicators. We then used the survey to examine how well the framework holds up when empirically measured with survey data.

4. Quantitative Phase: Web Survey

In the quantitative phase, we used a web survey to assess the framework that we developed from literature and interviews. We designed the survey instrument³ to mirror the structure of the framework. For each core dimension, we created multiple items that reflected the subdimensions identified in the qualitative phase and used wording that was close to the language experts used in the interviews. We also included items to capture demographic characteristics of the respondents.

The target population for the survey was federal employees in the US (including consultants and contractors to the government). We recruited the participants through professional events⁴, listservs⁵, and individual networks. The survey was conducted online in November 2024 using Qualtrics and remained open for three months. In total, 85 respondents participated, out of which we received 24 completes. Due to the small number of consultants and contractors, we excluded them from further analysis, restricting the target population to current and former federal employees as of November 2024 ($n = 20$). Our sample S_{NP} had the two following features:

² Refer to the mapping table in the supplementary materials.

³ Full questionnaire is provided in the supplementary materials.

⁴ Morris Hansen Lecture, 2024; Annual Conference of Federal Committee of Statistical Methodology, 2024

⁵ American Statistical Association chapters (SRMS, GSS, SOC, HPSS) and Washington Statistical Society

Table 1. Demographic Distribution in FEVS and S_{NP}

Variables	Categories	Demographic Composition*	
		FEVS	S_{NP}
Age	Under 40	24.4%	50%
	40 or Older	64.7	50
Race	Black or African American	12.9	5
	White or Caucasian	58.2	80
	Asian	5.2	NA**
	Others	7	15***
	Hispanic	11.4	15
Ethnicity	Non-Hispanic	76.3	80
	Male	51.8	55
	Female	36.8	30
Supervisory Status	Non-supervisor/Team Leader	73.7	70
	Supervisor/Manager/Executive	18.7	30
Duration in Federal Govt.	10 years or fewer	39.3	50
	11-20 years	31.2	30
	More than 20 years	21.7	20
Sample Size		1691257	20

*Percentages of missing values in each variable has not been shown in the table; in general, that is equivalent to the difference between 100 and the sum of the reported column percentages for each variable.

**No Asian individual participated in our survey

***This includes individuals who chose 'prefer not to say' or selected more than one race

1. There was no sampling frame, i.e., the selection probability for the individuals in the target population was unknown.
2. The selection probabilities for certain individuals in the population could be zero.

These precluded us from traditional design-based population inference. Hence, we combined our survey with a reference probability survey, Federal Employees Viewpoint Survey (FEVS)⁶ that is large enough ($n = 1,691,257$) and is designed to represent our target population (Valliant 2020). This illustrates one way survey researchers can leverage auxiliary information to strengthen conclusions from small or nonprobability samples (Chen et al. 2020; Kim et al. 2021; Schonlau et al. 2007). [Table 1](#) shows the demographic distribution in FEVS and S_{NP} .

FEVS includes the same core demographic variables as our survey, but it does not include our key outcomes of interest. Following Kim et al. 2021, we fit an ordinal regression model in our survey data that predicted the outcome(s) from the shared demographic variables to account for the ordered response scales. Under standard assumptions of positivity and transportability—that everyone in the population has non-zero probability of being included in

⁶ <https://www.opm.gov/fevs/about/>

FEVS and that the model fitted on our survey also holds in FEVS—we then used this model to predict the outcomes for all FEVS respondents. This “mass imputation” step created a synthetic outcome for each unit in the large representative probability sample for further inference. Refer to Appendix for the mathematical details.

However, our nonprobability survey is very small ($n = 20$) relative to the reference probability survey, which limits how far we can push population level inference based on mass imputation. With so few cases, the outcome model is noisy, and any misspecification can directly bias the projected estimates. Valid inference also depends on strong, hard to verify assumptions about self-selection and model transportability, so we treat our integrated estimates as illustrative outputs with clearly stated assumptions rather than definitive population statistics. Thus, the main contribution is to show, in a concrete example, how mass imputation can be implemented and what kinds of assumptions and limitations survey researchers should consider when combining a small nonprobability sample with a large probability survey.

5. Validating Conceptualized Framework

Based on the FEVS data with mass-imputed outcomes of our interest we conducted a confirmatory factor analysis (CFA) that sought to validate the theoretical framework we developed at the qualitative phase of this study. We treated the main dimensions in the framework as latent constructs, each measured by a set of survey questions. First, we specified separate measurement models for each dimension and estimated factor loadings to see how strongly each item relates to its intended factor. Then we combined these into a structural model that included the directional relationships among the latent constructs, reflecting the structure suggested by the qualitative work (in our case, from fluency to adoption to adaptation).

We compared two competing CFA models. We started with the framework developed in the earlier phases and then made theory-driven modifications, such as adding covariances between item residuals when the earlier qualitative findings suggested that two items are closely related. We used standard fit indices to decide our final model, as shown below:

In our final CFA model, the observed indicators loaded meaningfully onto the three latent dimensions—**Fluency**, **Adoption**, and **Adaptation**—supporting the proposed measurement structure. We can interpret this based on the factor loading estimates and latent covariances obtained from the summary of the CFA model. **Fluency** showed generally strong loadings, indicating a well-defined construct. **Adoption** displayed a more heterogeneous pattern, reflecting variability in how different adoption-related indicators contribute to the construct. **Adaptation** was characterized by strong overall loadings, capturing organizational responses to AI implementation, while also reflecting perceived shifts in strategic focus.

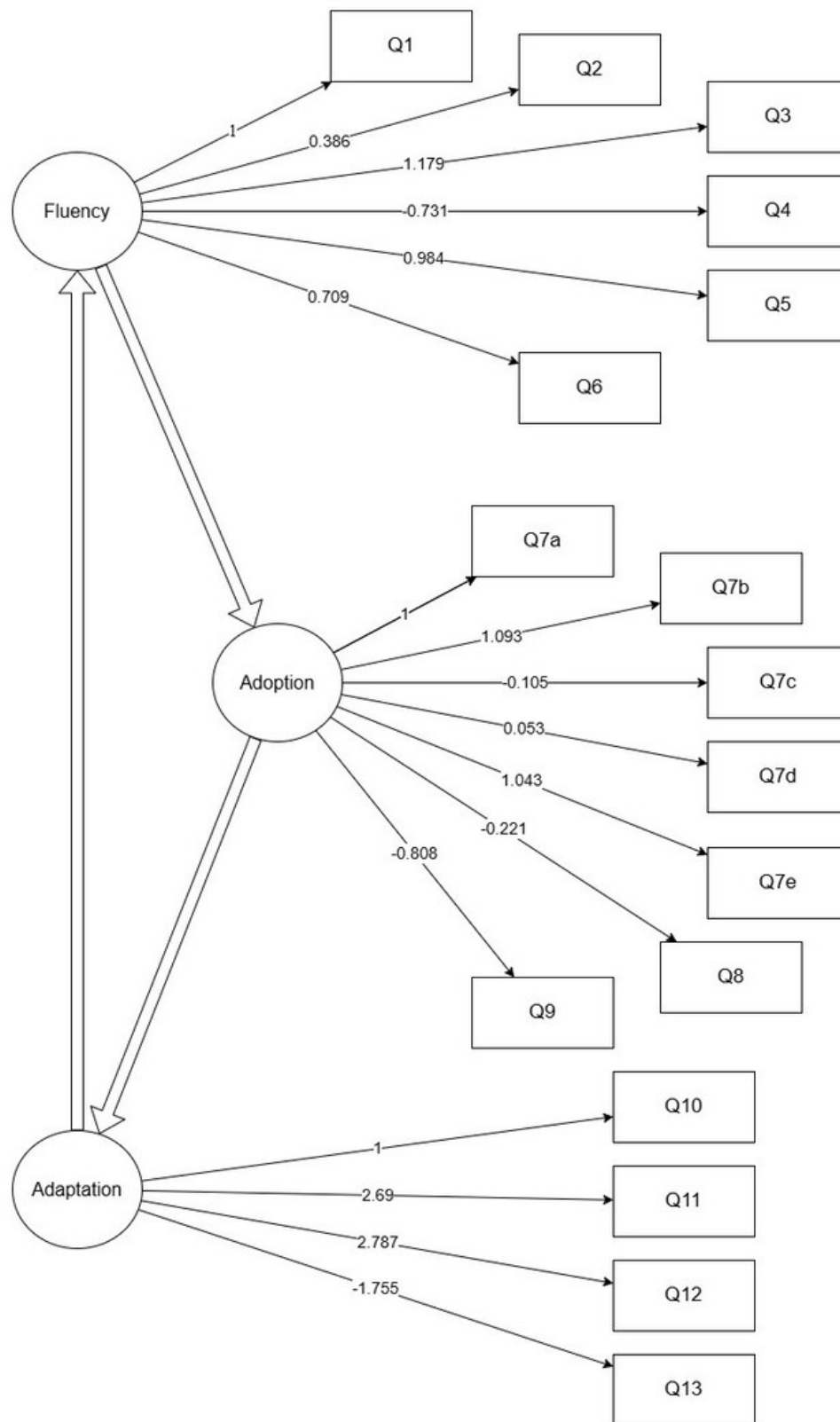


Figure 1. Visualization of the Fitted CFA Model

Finally, Fluency, Adoption, and Adaptation are all strongly associated, indicating that the three latent dimensions are closely related, though the directions of these associations vary.

Table 2. Model Fit Indices

Model	CFI	Chi Square Statistic (in Million)	RMSEA
1	0.776	10.372	0.414
2	0.91	4.171	0.272

6. Implications

This study shows how a mixed methods design can help survey researchers move from broad ideas to tested measures in a structured way. By starting with a focused literature review, using a small set of expert interviews to refine and name key dimensions, and then converting those dimensions into survey items, we demonstrate a practical path from concepts to an empirically tested framework.

Our approach illustrates an exploratory sequential design in which qualitative work does not stand alone but feeds directly into survey development and validation. The interviews help us identify language, subdimensions, and relationships that matter in practice; the survey later lets us test whether these hold up empirically. Using confirmatory factor analysis, we show how one can assess whether items derived from qualitative insights form coherent scales and represent the intended latent constructs.

A second contribution is our use of mass imputation to link a small nonprobability survey with a large reference probability survey, illustrating how this strategy can attenuate selection bias; however, one should be cautious with the sample size of the receiver survey before making valid population-level inferences. Taken together, these steps, i.e., literature to interviews to framework development, framework to survey items, and nonprobability survey data linked to a probability survey for further analyses, offer a set of concrete procedures that other survey researchers can adapt to design and strengthen survey-based research in settings with constrained samples, or emerging constructs that have not yet been well defined or measured.

.....

Author Contribution

Ujjayini Das and Srijeeta Mitra contributed equally to the planning and execution of this research. They share joint first authorship on this article. Tejawansh S. Anand as Principal Investigator and Bertrand A. Stoffel as Co-Principal Investigator supported at different stages of the work.

Conflict of Interest

The authors declare no conflict of interest.

Statement on Funding

This work was supported through the Deloitte Initiative for Artificial Intelligence and Learning (DIAL) at the University of Maryland for the academic year 2023–2024.

Acknowledgments

We extend our gratitude to the research team at Robert H. Smith School of Business for contributing to background research, with a special mention to Ishita Joshi, for her assistance in initial qualitative coding.

Corresponding author contact information

Ujjayini Das

Email: ujstat@umd.edu

Address: 1218 Lefrak Hall, 7251 Preinkart Drive, College Park, MD 20742

Srijeeta Mitra

Email: smitra98@umd.edu

Address: 1218 Lefrak Hall, 7251 Preinkart Drive, College Park, MD 20742

Submitted: November 13, 2025 EDT. Accepted: February 17, 2026 EDT. Published: August 23, 2026 EDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-ND-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc-nd/4.0> and legal code at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> for more information.

REFERENCES

- Chen, Yilin, Pengfei Li, and Changbao Wu. 2020. "Doubly Robust Inference With Nonprobability Survey Samples." *Journal of the American Statistical Association* 115 (532): 2011–21. <https://doi.org/10.1080/01621459.2019.1677241>.
- Deloitte. 2022. "State of AI in the Enterprise 2022 | Deloitte US." <https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/articles/state-of-ai-2022.html>.
- Engstrom, David Freeman, Daniel E. Ho, Catherine M. Sharkey, and Mariano-Florentino Cuéllar. 2020. "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3551505>.
- Frank, Darius-Aurel, Christian T. Elbæk, Caroline Kjær Børsting, Panagiotis Mitkidis, Tobias Otterbring, and Sylvie Borau. 2021. "Drivers and Social Implications of Artificial Intelligence Adoption in Healthcare during the COVID-19 Pandemic." *PLOS ONE* 16 (11): e0259928. <https://doi.org/10.1371/journal.pone.0259928>.
- Hennink, Monique, and Bonnie N. Kaiser. 2022. "Sample Sizes for Saturation in Qualitative Research: A Systematic Review of Empirical Tests." *Social Science & Medicine* 292 (January): 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>.
- Hennink, Monique M., Bonnie N. Kaiser, and Vincent C. Marconi. 2017. "Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough?" *Qualitative Health Research* 27 (4): 591–608. <https://doi.org/10.1177/1049732316665344>.
- Horowitz, Michael C., and Lauren Kahn. 2021. "What Influences Attitudes about Artificial Intelligence Adoption: Evidence from U.S. Local Officials." *PLOS ONE* 16 (10): e0257732. <https://doi.org/10.1371/journal.pone.0257732>.
- Kim, Jae Kwang, Seho Park, Yilin Chen, and Changbao Wu. 2021. "Combining Non-Probability and Probability Survey Samples Through Mass Imputation." *Journal of the Royal Statistical Society Series A: Statistics in Society* 184 (3): 941–63. <https://doi.org/10.1111/rssa.12696>.
- Madan, Rohit, and Mona Ashok. 2022. "A Public Values Perspective on the Application of Artificial Intelligence in Government Practices: A Synthesis of Case Studies." In *Advances in Electronic Government, Digital Divide, and Regional Development*, edited by Jose Ramon Saura and Felipe Debas. IGI Global. <https://doi.org/10.4018/978-1-7998-9609-8.ch010>.
- Mikalef, Patrick, Kristina Lemmer, Cindy Schaefer, et al. 2022. "Enabling AI Capabilities in Government Agencies: A Study of Determinants for European Municipalities." *Government Information Quarterly* 39 (4): 101596. <https://doi.org/10.1016/j.giq.2021.101596>.
- Neumann, Oliver, Katharina Guirguis, and Reto Steiner. 2024. "Exploring Artificial Intelligence Adoption in Public Organizations: A Comparative Case Study." *Public Management Review* 26 (1): 114–41. <https://doi.org/10.1080/14719037.2022.2048685>.
- Pencheva, Irina, Marc Esteve, and Slava Jankin Mikhaylov. 2020. "Big Data and AI – A Transformational Shift for Government: So, What next for Research?" *Public Policy and Administration* 35 (1): 24–44. <https://doi.org/10.1177/0952076718780537>.
- Schonlau, Matthias, Arthur H. O. Van Soest, and Arie Kapteyn. 2007. "Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring?" *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1006108>.
- Sun, Tara Qian, and Rony Medaglia. 2019. "Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare." *Government Information Quarterly* 36 (2): 368–83. <https://doi.org/10.1016/j.giq.2018.09.008>.

- Valliant, Richard. 2020. "Comparing Alternatives for Estimation from Nonprobability Samples." *Journal of Survey Statistics and Methodology* 8 (2): 231–63. <https://doi.org/10.1093/jssam/smz003>.
- Wirtz, Bernd W., Jan C. Weyerer, and Benjamin J. Sturm. 2020. "The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration." *International Journal of Public Administration* 43 (9): 818–29. <https://doi.org/10.1080/01900692.2020.1749851>.

Appendix

1. Interview Guide

First, let us thank you for taking the time to talk with us. As we mentioned during our previous exchange, we are researchers at the Smith School of Business, University of Maryland. Our goal in this research project is to better understand how public agencies approach artificial intelligence and how they implement it. Because this process is largely guided by individuals in those agencies, we are trying to talk to as many people in the mix as possible to get a better sense of how this works. So again, thank you for your time.

Before we proceed, let us review our consent form and see if you have any questions in this regard.

Now that we have your consent, let's start with some general questions:

1. How about we talk about your position in [agency] when AI-based solutions were first considered.
2. Can you describe how AI relates to the mission of the agency?

Because it is nice for us as academics to be as concrete as possible, we want to use a couple of specific projects as case studies. We have read about some of the projects at your agency, and know the contours of the projects, as well as the details that were made public. What I don't know is the information that was not revealed to the public. So, I hope you can help me unpack what types of discussions unfold when it comes to using artificial intelligence, who is involved in planning and executing AI-based solutions, and how this gets implemented in practice.

Let's take an AI project within your agency that you feel most comfortable talking about.

1. Can you describe the project?
2. Why do you think AI was important in this project?
3. Can you tell us more about how this project was selected?
4. What was your role in it?
5. How was it implemented?
6. How would you assess the result of this project?
7. What were the key factors in the success of this project?
8. What were the key factors in the lack of success of this project?

Is there another project that you would like to talk about? What about this project?

1. Can you describe the project?
2. Why do you think AI was important in this project?
3. Can you tell us more about how this project was selected?
4. What was your role in it?
5. How was it implemented?
6. How would you assess the result of this project?
7. What were the key factors in the success of this project?
8. What were the key factors in the lack of success of this project?

Let us finish with some final reflective questions:

1. How do you compare your agency's use of AI with what other agencies are doing?
2. Is there anything you would like to talk about that we did not cover in our discussion?
3. Is there anyone else that we should talk to?

2. Insights from Literature Review: Case Study-specific

For conventional library catalogs, some of the keywords we used are, “AI” and “adoption”, “framework”, “adoption”, “governance”, as well as “driving AI adoption”, “factors impacting AI adoption”, and “slow down AI adoption”. For the broader search of the gray literature, we simply used the keywords “AI” and “report”, followed by a manual selection of relevant reports among the first 100 hits. Our review of studies on AI adoption in public and private agencies revealed four key themes:

- **AI Literacy:** AI literacy reflects an organization's capacity to deploy and utilize AI technologies. Deloitte's State of AI in the Enterprise report categorizes organizations as “transformers”, “pathseekers”, “starters”, and “underachievers” based on AI deployment and outcomes (Deloitte 2022). Other studies define literacy by the sophistication level of use cases, measured by the technical complexity and expertise required for maintenance and

oversight, rather than outcomes. For instance, Engstrom et al. (2020) classifies logistic regression as low in sophistication and deep learning as high.

- **AI Adoption Frameworks:** Research conceptualizing AI adoption in public administration generally encompasses three domains: technology and data, people and organization, and environment and governance (Madan and Ashok 2022; Mikalef et al. 2022). In a study of eight Swiss public organizations, Neumann et al. (2024) used the technology-organization-environment (TOE) framework, showing that these factors vary by an organization's "AI maturity", where simpler applications often precede advanced ones, and early adopters influence diffusion.
- **Drivers of AI Adoption:** Researchers differ on what enables or constrains AI/ML implementation in public services. Some emphasize organizational and environmental factors like financial costs, innovation culture, regulatory support, and governmental pressure, as key enablers (Mikalef et al. 2022; Neumann et al. 2024). Others stress social influences, particularly public trust in AI, which shapes adoption when citizens are directly affected (Frank et al. 2021).
- **AI Governance and Policies:** Finally, another major strand examines ethical and governance challenges in AI-enabled public administration. This literature highlights how moral values, norms, and experiences shape support for AI (Horowitz and Kahn 2021; Pencheva et al. 2020; Sun and Medaglia 2019), emphasizing fairness, accountability, and transparency as essential principles. Scholars argue that governance models must explicitly address these ethical tensions (Engstrom et al. 2020; Madan and Ashok 2022; Wirtz et al. 2020).

3. Insights from Expert Interviews: Case Study-specific

Building on themes from the literature, we refined our understanding of AI adoption and its impact in the public sector through qualitative interviews. These interviews helped define "impact" across three core dimensions: **AI fluency**, **AI adoption**, and **AI adaptation**. Drawing on expert insights and the literature review, we define these as follows:

- **AI Fluency:** The extent to which individuals or organizations understand and apply AI concepts, critically evaluate tools, and responsibly integrate AI into decision-making, policy, and service delivery.

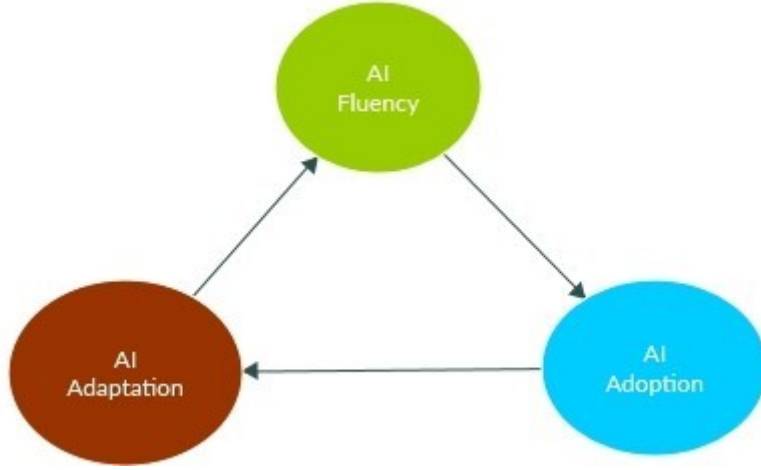


Figure A1. AI Impact Framework

- **AI Adoption:** The degree to which AI is embedded in workflows, operations, and governance, which signifies a transition from experimentation to sustained institutional use, linking fluency to practice.
- **AI Adaptation:** The organizational, social, and policy transformations resulting from the use of AI, including increased efficiency, accountability, transparency, and public value.

Based on the combined qualitative insights, we conceptualize the **AI Impact Framework (AIF)** as a sequential relationship among these dimensions (Fig. A1), where AI fluency facilitates adoption, adoption drives adaptation, and adaptation in turn reinforces fluency.

4. Mass Imputation on FEVS

Let δ be the indicator variable for an individual i being included in the nonprobability sample S_{NP} , i.e., $\delta = 1 \forall i \in S_{NP}$ and $\delta = 0 \forall i \notin S_{NP}$. We assume the following for the inference:

1. **Positivity:** Each individual in the population has a non-zero probability of being included in S_{NP} , i.e., $P(\delta = 1 | \mathbf{X} = \mathbf{x}) > 0$ for all $\mathbf{x}$ in the support of $\mathbf{X}$.
2. **Transportability:** The prediction model $f(y|\mathbf{x})$, which can be estimated using the observed $(Y, \mathbf{X})$ from S_{NP} if $f(y|\mathbf{x}, \delta = 1) = f(y|\mathbf{x})$ is transportable to FEVS.

Under these two assumptions, we consider a mass imputation estimator based on a semi-parametric model-based imputation approach as suggested by Kim et al. 2021. We consider a model for S_{NP} with the first conditional moment specified as

$$E(Y|X = x) = m(x, \beta_0)$$

for some unknown $p \times 1$ vector β_0 and a known function $m(\cdot; \cdot)$.

Among all the demographic variables listed in Table 1 in the article, the most relevant ones to model our outcome of interest are race, ethnicity, age, sex, supervisory status, and duration in federal government. Therefore, considering these four demographic variables as predictors and given that our outcome variables are categorical and ordinal in nature, we employ a proportional odds cumulative logit model as the following:

$$\text{logit}(p_{ik}) = \mu_k + X_i' \beta; i = 1, \dots, n_a, k = 1, \dots, 5 \text{ or } 7 \quad (1)$$

where $p_{ik} = P(Y_i \leq k|X_i)$ denotes the likelihood that the outcome Y_i for individual i falls in category k or below on the response scale, given the predictors $X_i = (X_{i_{age}}, X_{i_{fed}}, X_{i_{sup}}, X_{i_{sex}}, X_{i_{race}}, X_{i_{ethnicity}})'$. The model is fitted under a proportional odds assumption, which ensures that the effect of an explanatory variable is the same across all levels of ordinal outcome. Alternatively, the relationship between every pair of response levels is identical, and the coefficients for each predictor category have parallel slopes.

Under the transportability condition, we estimate the model parameters β using S_{NP} . We assume that $\hat{\beta}$ is the unique solution to the estimating equations

$$\hat{U}(\beta) = \frac{1}{n_{NP}} \sum_{i \in S_{NP}} \{y_i - m(x_i; \beta)\} h(x_i; \beta) = 0$$

for some p -dimensional vector function $h(x; \beta)$. Here n_{NP} denotes the sample size of our survey i.e., $n_{NP} = 20$ in our case.

After fitting model in equation 1, we estimate the model parameters $\hat{\beta}$ using S_{NP} where both Y and X are observed. Using these estimates, we then obtain the predicted values $\hat{y}_i = m(\mathbf{x}_i, \hat{\beta})$ for all $i \in FEVS$. Consequently, the mass imputation estimator for the finite population (of size N) proportion $\theta_N = N^{-1} \sum_{i=1}^N y_i$ is computed as $\hat{\theta}_I = \frac{1}{\hat{N}} \sum_{i \in FEVS} w_i \hat{y}_i$; where w_i are the survey weights for the probability sample FEVS. Assuming that the population size N is unknown, we estimate it as $\hat{N} = \sum_{i \in FEVS} w_i = 1,691,257$.

SUPPLEMENTARY MATERIALS

Supplementary materials 1

Download: <https://www.surveypractice.org/article/158334-from-concepts-to-measures-a-mixed-methods-framework-for-measuring-latent-constructs/attachment/332881.docx>

Supplementary information 2

Download: <https://www.surveypractice.org/article/158334-from-concepts-to-measures-a-mixed-methods-framework-for-measuring-latent-constructs/attachment/332882.zip>

Supplementary information 3

Download: <https://www.surveypractice.org/article/158334-from-concepts-to-measures-a-mixed-methods-framework-for-measuring-latent-constructs/attachment/332883.docx>
