From Normative Awareness to Workflow Confidence: Self-Reported Data Science Literacy in China's Local Public Sector

Haisheng Tang[1,2]*

[1]School of Public Administration, Guangzhou Nanfang College, Guangzhou 510970, Guangdong, China

[2]Party School of the CPC Chaozhou Municipal Committee, Chaozhou 521011, Guangdong, China

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100700252

Received: 15 July 2026; Accepted: 20 July 2026; Published: 29 July 2026

ABSTRACT

Digital-government reform depends on public-sector personnel who can move evidence from problem framing to accountable action, yet empirical research often measures statistical skill or broad digital literacy rather than the full workflow. This study examines self-reported confidence across data generation, collection, cleaning, management, analysis, communication, application, and ethics. A cross-sectional online survey recruited personnel through professional networks in Hunan, Guangdong, and Sichuan, China. Of 512 returns, two fell below the prespecified 60-second completion threshold, leaving 510 cases. The analysis reports internal consistency, psychometric diagnostics, descriptive profiles, correlations, and exploratory Welch group comparisons, together with duration- and scoring-rule sensitivity checks. The item-weighted overall mean was near the five-point scale midpoint (M = 3.048, SD = .771). Ethics ranked first (M = 3.565), followed by data generation (M = 3.541); collection ranked last (M = 2.813), with communication (M = 2.867) and analysis (M = 2.929) also below the midpoint. Respondents expressed least confidence in programming, web acquisition, application programming interfaces, metadata, advanced methods, and interactive visualization. The gender mean difference was negligible (0.040 points, 95% CI [-0.104, 0.184], Hedges g = .052). Welch tests showed omnibus variation across age, education, organizational level, and position, but small groups, unusually low within-group variance, and empty cross-classifications preclude independent or causal interpretations. The study contributes a workflow-completeness lens and identifies a normative-operational confidence gap in this sample while keeping reported confidence distinct from demonstrated performance and organizational capability. Under these evidentiary limits, it supports authentic public-data projects, role-differentiated development, and ethical judgment embedded throughout technical work.

Keywords: Public managers, data science literacy, digital government, public administration education, workforce development

INTRODUCTION

A government can buy software, connect databases, and still change little. Digital transformation reaches administrative practice only when routines, services, and professional roles change with the technology [11, 19]. Public-sector personnel must then frame an administrative problem in data terms, locate usable evidence, prepare it, judge an analysis, explain uncertainty, and act within legal and ethical limits. The OECD treats skills, an enabling work environment, and long-term workforce planning as interdependent foundations of digital maturity[13]. When those foundations lag, data systems can accelerate weak routines, deepen dependence on vendors, and grant opaque models unwarranted authority[2, 9, 12].

China offers a useful setting in which to examine this implementation problem. National policy links digital-government construction to process redesign, governance innovation, and administrative performance[18]. The Overall Layout Plan for Digital China also calls for stronger digital thinking and skills among public officials and places interdisciplinary digital-talent development on the higher-education agenda[3]. This combination of ambitious policy, rapid digitalization, and a multilevel administrative system makes the case internationally relevant: it exposes a problem faced well beyond China--how to convert digital strategy into reliable everyday data work.

Public-administration educators have mapped many of the necessary competencies. Frameworks for MPA and MPP programs combine statistics, computing, communication, and ethics[6, 8, 15]. Yet conventional methods teaching often begins with a clean file and concentrates on research design and statistical analysis, leaving less room for acquisition, documentation, reproducibility, and operational communication[7, 16]. Workplace research reaches a related conclusion. Data-informed decisions rely on analytical skill, domain knowledge, teamwork, stakeholder engagement, critical judgment, and political awareness[4]. These capacities are distributed across people and organizations rather than located in a single employee[5, 14, 17].

Two gaps remain. First, curriculum frameworks specify what programs should teach, but empirical studies less often ask serving personnel where they experience difficulty across an end-to-end workflow. Second, self-reported individual literacy is easily conflated with demonstrated performance or organizational capacity. Existing Chinese research has developed a six-domain model from interviews with 37 civil servants[20], but larger workflow-based diagnostics remain limited. Subgroup evidence also needs care because education, position, administrative level, access to data, and task exposure overlap inside public organizations.

This study makes three bounded contributions. Conceptually, it treats workflow completeness--including upstream acquisition and downstream communication--as the object of diagnosis. Empirically, it maps self-reported confidence across eight theory-informed content areas in a targeted sample of 510 public-sector personnel. Practically, it connects the observed profile to a division of responsibility among universities, public-service academies, and workplaces. The eight areas are a content map, not a claim that an eight-factor reflective scale has been validated. Three research questions guide the analysis.

RQ1: What is the overall level and domain profile of self-reported data science literacy among the surveyed public managers?

RQ2: What pattern emerges between normative/problem-framing awareness and operational workflow confidence?

RQ3: How does overall self-reported literacy vary across gender, age, education, organizational level, and position, and what can these patterns defensibly imply for education and workforce development?

The argument is deliberately narrow. Valuing data has little administrative force when personnel cannot obtain, prepare, evaluate, or communicate them. I use the phrase **normative-operational confidence gap** to describe a difference in reported confidence, not a latent trait or observed performance deficit. The subgroup analysis remains exploratory because the survey is cross-sectional, nonprobability based, and structurally sparse. The educational claim follows from those boundaries: universities can build transferable foundations, public-service academies can place them in an administrative setting, and workplaces can provide repetition, mentoring, and complementary team expertise.

Theoretical Framework

Data science literacy as workflow completeness

Data-literacy assessments in the public service range from self-ratings to performance tasks, and some measure individual skill while others examine organizational capacity[1, 14]. The term therefore needs a clear boundary. In this study, **data science literacy in the public sector** means reported confidence in framing a public problem in data terms and then generating, collecting, cleaning, managing, analyzing, communicating, and applying data under ethical constraints. This is a study-specific operational definition. It does not imply that agreement with a survey statement demonstrates competence. The definition begins with the administrative problem because software proficiency has value only when it helps personnel move that problem through an accountable evidence workflow.

This workflow reading builds on curriculum frameworks that bring computing, statistics, communication, and ethics into public administration[8, 15, 16]. It also exposes a familiar classroom shortcut. Analysis is taught with a clean file already on the screen, although public employees usually begin with an ill-defined question, scattered records, uncertain provenance, and missing documentation. The trouble starts upstream. A poorly framed question shapes the wrong data, and weak collection or cleaning compromises everything that follows. Downstream failures matter too. Even a sound calculation remains unused when the result cannot be explained or connected to a decision, and an overall mean can hide any one of these bottlenecks.

Table 1 sets out the eight content areas used in the inventory. The order offers a practical map. Administrative data projects rarely move in a straight line. A manager may return to the question after inspecting the records, revise a collection plan when privacy concerns surface, and then revisit earlier choices.

Table 1. Workflow Areas in the Data Science Literacy Diagnostic Inventory

Content area	Items	Public-management function	Illustrative self-report item
Data generation	4	Translate public problems into measurable data needs	I can identify and articulate my data needs when formulating a question.
Data collection	7	Select sources and obtain data in usable formats	I understand application programming interfaces and can use them to access data.
Data cleaning	12	Identify and address errors, missingness, and outliers	I can identify outliers and select an appropriate treatment.
Data management	8	Store, document, classify, protect, and retrieve data	I understand metadata and can create a data dictionary.
Data analysis	12	Apply appropriate computational and statistical methods	I can use an analytical tool such as R or Python proficiently.
Data communication	8	Explain evidence, uncertainty, and implications	I can create interactive charts or dashboards to present results.
Data application	6	Integrate evidence into decisions and service improvement	I can use analytical results to support decisions or forecasts.
Data ethics	5	Protect rights and ensure responsible data use	I consider privacy when collecting and communicating data.

The normative-operational confidence gap

Public employees can learn that privacy, fairness, and accountability matter long before they gain regular access to APIs, code, metadata, or untidy administrative records. Problem framing also calls for judgment about what deserves measurement and whose interests a measure may miss. Operational confidence grows through a different set of encounters: finding data, correcting it, recording what changed, choosing an analysis, explaining uncertainty, and taking a result into a decision. The two forms of learning can move at different speeds.

I use **normative-operational confidence gap** as a descriptive label for this possible unevenness. Rules and integrity education can strengthen awareness of responsible use while offering little practice in tracing a source, debugging a script, documenting a cleaning choice, or defending a visualization before a skeptical audience. Technical training creates the opposite risk when public values sit in a separate module. Ethics therefore belongs inside collection, cleaning, analysis, communication, and deployment, where choices acquire public consequences. Whether the same gap appears in observed performance remains an empirical question beyond a self-report survey.

From individual literacy to multilevel capability

A self-rating belongs to an individual. Digital-government capability is produced across people, teams, and organizations. Most consequential data work draws on domain knowledge, analytical skill, critical judgment, stakeholder engagement, and political awareness, a combination that rarely resides in one employee[4]. OECD and recent public-administration research make the same distinction between individual skills and the leadership, team composition, work environment, and career systems that turn those skills into public value[13, 17].

This matters when reading differences by age or position. Either variable may stand in for years of task exposure, access to a data team, or distance from hands-on work, but the category itself reveals no mechanism. Senior managers need enough literacy to question a model, demand documentation, and understand uncertainty even if they never write production code. Frontline staff may spend more time collecting and recording data. Specialists carry deeper technical tasks. Everyone still needs a common floor in data quality, ethics, and evidentiary limits, while technical depth can follow the work assigned in practice.

Universities occupy one part of this learning path. Public-administration programs can teach concepts that travel across jobs and give students their first sustained encounter with a public-data project. Skill then changes through continuing education and use at work. One conventional methods course has too much to carry[16]. International curriculum initiatives now place students in iterative, user-centered, ethical, and multidisciplinary work because a software checklist ages quickly[5]. I return to this division of responsibility when interpreting the survey.

METHODOLOGY

Research design and recruitment

I administered a cross-sectional online survey to serving public-sector personnel. Recruitment moved through professional networks in Hunan, Guangdong, and Sichuan, with help from contacts involved in cadre education and training. The questionnaire recorded organizational and position levels but did not separately ask whether each respondent supervised staff. Accordingly, the paper uses *public-sector personnel* for participants and reserves *public managers* for the broader professional and educational implications. The aim was to reach personnel working at village/community, township/subdistrict, county/district, and prefecture-or-higher levels.

Recruitment was targeted and nonprobability based. The number of invitations, number of participating networks, and denominator needed for a response rate were not retained. The sample therefore supports a diagnosis of these respondents and exploratory comparisons within the group; national, provincial, and organizational prevalence estimates lie beyond the design.

The platform recorded 512 submissions between 27 and 29 September 2024. Under the prespecified screening rules, a response was removed when completion took less than 60 seconds or when the same option appeared across all 62 literacy items. Two submissions took 59 seconds and were removed. None of the 512 responses had an invariant item pattern, and the retained records had no missing literacy responses. The analytical sample contained 510 cases. Because 60 seconds is permissive for a 67-question instrument, I also repeated the profile with 90- and 120-second thresholds.

Participants

Table 2 describes the 510 retained respondents. Most worked in county/district organizations, and 76.3% held a bachelor's degree. Only 10 respondents were aged 51 years or above, 12 worked in village/community organizations, and 26 held an associate degree. Estimates for those groups deserve particular restraint.

Table 2. Sample Characteristics (N = 510)

Variable	Category	n	%
Gender	Men	356	69.8

Gender	Women	154	30.2
Age	18-30 years	111	21.8
Age	31-40 years	212	41.6
Age	41-50 years	177	34.7
Age	51 years or above	10	2.0
Education	Associate degree	26	5.1
Education	Bachelor's degree	389	76.3
Education	Postgraduate degree or above	95	18.6
Organizational level	Village/community	12	2.4
Organizational level	Township/subdistrict	40	7.8
Organizational level	County/district	282	55.3
Organizational level	Prefecture level or above	176	34.5
Position	Section member	191	37.5
Position	Deputy at township/section level	153	30.0
Position	Principal at township/section level	135	26.5
Position	Deputy at county/division level	31	6.1

Diagnostic inventory and scoring

Five background questions preceded 62 literacy statements. The Chinese-language item pool was assembled from workflow domains and competencies discussed in the NASPAA framework and public-administration studies of data literacy and data-science education[8, 14-16, 20], then adapted to the public-sector setting. Each statement used a five-point agreement scale running from 1 (strongly disagree) to 5 (strongly agree). The inventory assigned 4 items to data generation, 7 to collection, 12 to cleaning, 8 to management, 12 to analysis, 8 to communication, 6 to application, and 5 to ethics. Supplementary Table S1 provides the fielded Chinese wording and an English translation for transparency.

For each content area, I averaged the assigned items. The primary overall score is the mean across all 62 statements, so areas with more items carry more weight. A sensitivity score gives equal weight to the eight area means. Higher scores reflect greater reported confidence, not observed proficiency. I treat the instrument as a **diagnostic inventory** because the adaptation did not include independent expert validation or cognitive interviews, and the resulting scores have not undergone confirmatory factor analysis, test-retest assessment, measurement-invariance testing, or comparison with observed performance. I therefore report reliability coefficients and exploratory factor-analytic diagnostics as evidence about score behavior in this sample, not as proof that the inventory is a validated eight-factor scale. Public-service measurement work makes the same separation among self-assessment, objective performance, and organizational capacity[1].

Statistical analysis

I ran the analysis in Python 3.8.2 with pandas 2.0.3, NumPy 1.24.4, SciPy 1.10.1, and statsmodels 0.14.1. Cronbach's alpha, means, standard deviations, item ranks, and Pearson correlations describe the inventory.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

Group comparisons used Welch's independent-samples *t* test for gender and Welch's one-way analysis of variance for age, education, organizational level, and position because group sizes and variances differed markedly. For gender, I report the mean difference, a 95% confidence interval, and Hedges *g*. For the multi-category comparisons, I report pooled one-way omega squared as a descriptive share of variance; it is not an exact effect size for the Welch statistic. Conventional ANOVA and Kruskal-Wallis tests served as sensitivity checks. All tests were two-sided at alpha = .05.

These are exploratory omnibus tests. They indicate whether sample means vary somewhere across categories but do not identify which background characteristic produced a difference. The cross-tabulations explain why this boundary matters. All 26 associate-degree respondents came from township/subdistrict organizations, all 12 village/community respondents were section members, and all 10 respondents aged 51 or above were deputies at township/section level. Several combinations do not exist in the data. Platform metadata also contained repeated network identifiers, which may reflect shared organizational networks but make independent-observation assumptions uncertain. Those identifiers were not used as respondent identities or exported to the analytical tables. I fitted a full OLS model only as a diagnostic check; its residuals were heteroskedastic and several coefficients depended on comparisons outside observed combinations. I therefore draw no adjusted or causal claim from that model and do not conduct pairwise post-hoc tests that would invite unsupported category rankings.

RESULTS

Internal consistency and domain profile

Table 3 reports the N = 510 estimates. Alpha ranged from .874 for data collection to .961 for cleaning, with .986 across all 62 items. The items clearly move together. The unusually high overall coefficient, coupled with strong correlations among several content areas, may also reflect conceptual or wording overlap. These coefficients support score consistency in this sample but do not establish structural validity.

Table 3. Internal Consistency and Descriptive Profile

Content area	**Items**	**Cronbach's alpha**	**M**	**SD**	**Rank**
Data ethics	5	.944	3.565	.997	1
Data generation	4	.919	3.541	.926	2
Data management	8	.908	3.146	.821	3
Data cleaning	12	.961	3.006	.886	4
Data application	6	.935	2.992	.873	5
Data analysis	12	.954	2.929	.865	6
Data communication	8	.953	2.867	.893	7
Data collection	7	.874	2.813	.843	8
Overall inventory	62	.986	3.048	.771	-

At 3.048, the item-weighted overall mean lay close to the midpoint of the five-point scale. The equally weighted mean of the eight content areas was 3.107 and correlated .995 with the primary score, so the scoring rule did not change the interpretation. Ethics and data generation occupied the top two places, management sat lower, and cleaning and application clustered around 3.0. Collection was at the bottom, with communication and analysis

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

just above it. Respondents thus felt surer about recognizing responsible use and stating a data need than about carrying out the technical and communicative work that follows.

The item means make the contrast easier to see. Recognizing that data misuse can harm others received the highest rating (M = 3.749), closely followed by the ability to state one's data needs (M = 3.724). At the other end were R or Python (M = 2.510), web crawling (M = 2.512), APIs (M = 2.539), metadata (M = 2.557), and advanced statistical methods (M = 2.567). Interactive visualization also drew a low mean (M = 2.665). These numbers capture perceived confidence. They leave actual performance open, but they show where respondents themselves locate the steepest learning curve.

Correlations among the content areas ranged from .394 to .906, with every p value below .001. Analysis and communication had the closest association (r = .906). Management and analysis (r = .874) and cleaning and analysis (r = .873) followed, showing how confidence in analysis travels with confidence in preparing, managing, and explaining data. Correlation alone says nothing about direction, and it offers no confirmation of eight separate factors.

I also treated the item matrix as a psychometric stress test. The Kaiser-Meyer-Olkin statistic was .976 and Bartlett's test of sphericity was significant ($p < .001$), indicating that the correlation matrix was factorable. The deeper pattern was less supportive of the proposed eight content areas as separate reflective factors. The first component accounted for 54.4% of the item variance, and parallel analysis retained six factors rather than eight. The first eight components together accounted for 78.5% of the variance, but this descriptive result does not establish the intended structure. These diagnostics reinforce the paper's use of the inventory as a workflow-based profile of reported confidence, not as a completed measurement scale.

Group patterns

Table 4 places the group means and standard deviations beside the Welch tests and effect-size summaries. Gender produced almost no difference. Omnibus variation appeared across age, education, organizational level, and position, although severe imbalance, extremely small within-group variance in a few cells, and structural clustering narrow the meaning of those results.

Table 4. Overall Literacy by Demographic and Occupational Group

Predictor	Category	n	M	SD	Welch statistic	p	Effect-size summary
Gender	Men	356	3.060	.781	t(302.15) = .544	.587	g = .052
	Women	154	3.020	.749			
Age	18-30 years	111	2.874	.958	F(3, 255.60) = 10.858	< .001	omega squared = .012
	31-40 years	212	3.049	.803			
	41-50 years	177	3.144	.584			
	51 years or above	10	3.248	.035			
Education	Associate degree	26	2.744	.409	F(2, 70.50) = 6.560	.002	omega squared = .005
	Bachelor's degree	389	3.056	.706			

	Postgraduate degree or above	95	3.098	1.043			
Organizational level	Village/community	12	2.327	.058	F(3, 139.68) = 143.435	< .001	omega squared = .040
	Township/subdistrict	40	3.123	.617			
	County/district	282	2.961	.831			
	Prefecture level or above	176	3.219	.674			
Position	Section member	191	2.746	.912	F(3, 139.35) = 85.034	< .001	omega squared = .216
	Deputy at township/section level	153	3.113	.303			
	Principal at township/section level	135	3.554	.674			
	Deputy at county/division level	31	2.377	.301			

Note. Hedges *g* is reported for gender. Omega squared is a pooled descriptive effect size and is not the exact effect size of the heteroskedastic Welch test.

Gender. The mean for men exceeded the mean for women by .040 points (95% CI [-.104, .184], Hedges g = .052, p = .587). The estimate is negligible in this sample and gives no basis for gender-specific training.

Age. Means rose across the four age bands, from 2.874 among respondents aged 18-30 to 3.248 among those aged 51 or above. Experience may sit behind part of that climb, though the survey contains no measure of task exposure, training, tenure, or frequency of data use. The oldest group is especially fragile. It contains 10 people, all in the same broad position category, and its SD of .035 is unusually small. Future work needs a direct measure of experience before attaching meaning to age itself.

Education. Associate-degree respondents averaged 2.744, compared with 3.056 for bachelor's-degree holders and 3.098 for the postgraduate group. The apparent gap comes with a serious design problem: all 26 associate-degree respondents worked in township/subdistrict organizations. Recruitment, role, and organizational placement are tangled with education in these data. The postgraduate group also had a much larger SD of 1.043. Since the survey recorded neither curriculum exposure nor matched career histories, it gives no estimate of what an additional degree contributed.

Organizational level. Means ran from 2.327 in village/community organizations to 3.219 at prefecture level or above, but the middle categories reversed the expected ordering: township/subdistrict respondents averaged 3.123 and county/district respondents 2.961. The village/community figure comes from 12 section members and has an SD of only .058, so the four means do not support a linear hierarchy.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

Position. Principals at township/section level had the highest overall mean (3.554). Deputies at county/division level had the lowest (2.377). For the latter group, the weakness lay mainly in collection, cleaning, analysis, communication, and application, while ethics remained comparatively high. Distance from hands-on data work is one plausible explanation. Recruitment composition or unmeasured job assignments could produce the same pattern, and the survey recorded no task allocation, so the result is best treated as a hypothesis for workplace research.

Sensitivity checks and analytical boundary

Restoring the two 59-second questionnaires moved the overall mean from 3.048 to 3.049. Raising the threshold to 90 seconds retained 474 cases (M = 3.043), while a 120-second threshold retained 415 (M = 3.032). Under all three thresholds, ethics and generation remained first and second, and collection remained last. The primary N = 510 analysis follows the prespecified rule; the alternative screens show that the domain ordering does not depend on two borderline records.

The omnibus conclusions were directionally similar under Kruskal-Wallis tests, but conventional and Welch ANOVA did not always agree. The education comparison, for example, produced $p = .107$ under conventional ANOVA, $p = .002$ under Welch ANOVA, and $p = .026$ under Kruskal-Wallis. Combined with the very small omega-squared estimate (.005), this disagreement is a reason not to turn the education categories into a substantive ranking.

The exploratory OLS model returned R-squared = .310 and adjusted R-squared = .293, while a Breusch-Pagan test found pronounced heteroskedasticity ($p < .001$). Empty cross-classifications create the harder problem. For example, there is no observed associate-degree respondent outside a township/subdistrict organization, so an adjusted comparison with bachelor's-degree respondents asks the model to fill combinations the survey never captured. I draw no independent-predictor claim from those coefficients.

DISCUSSION

A normative-operational confidence gap

Ethics and data generation both cleared 3.5, while collection, communication, and analysis remained below 3.0. Respondents reported greater confidence in recognizing responsible use and stating what data they needed than in obtaining those data, analyzing them, and explaining what the result meant. Calls for "more digital skills" are too loose to capture that shape. In this sample, the weak points sit in the middle of the evidence process, where a question must become usable data and then a defensible administrative claim. Because no performance task was administered, the result is a confidence profile rather than proof of a capability deficit.

Collection deserves closer attention. It was the lowest domain, and four of the weakest items concerned web acquisition, APIs, metadata, or data dictionaries. By the time a poorly sourced file reaches an analyst, many important choices have already been made. Communication presents a second choke point. Its correlation with analysis reached .906, which makes sense in practice because a result must be explained, questioned, and tied to a public consequence before it can guide action. A methods course that begins with a clean file misses much of this work.

The high ethics mean needs a modest reading. Agreement with statements about privacy or harmful data use records awareness, while ethical conduct appears in concrete choices about minimization, proxies, bias, explanation, and contestability. Those choices were not observed here. Algorithmic accountability research places responsibility in procedures and answerability as well as individual awareness[2]. A final ethics module is too late. Students and officials need to confront ethical choices while they collect, clean, model, visualize, and deploy data.

Individual literacy and distributed capability

Table 4 resists a tidy hierarchy. Older respondents had higher means, but the oldest group contained 10 people in one broad position category. Postgraduate and bachelor's-degree means were close despite very different dispersion. Principals at township/section level sat at the top of the position table, while deputies at county/division level sat at the bottom, a sequence that does not form a simple ladder. A role-based organizational reading fits the evidence better because data work draws on hybrid teams whose members carry different combinations of analytical, domain, relational, and political competence[4]. A common literacy floor remains necessary, and depth can follow responsibility.

The deputy county/division result makes the issue visible. Their relatively low operational scores may reflect distance from hands-on tasks, selection into different roles, or quirks in recruitment. The survey cannot separate those explanations. Still, a gap between decision authority and familiarity with data work would create a governance risk. Senior officials can delegate programming, yet they remain responsible for asking where the data came from, what assumptions entered a model, how uncertainty was handled, and who can challenge the output[2, 10], which makes evaluative literacy necessary at that level.

The age result also redirects the research question. A label such as "digital native" says little about years spent cleaning administrative records or working with an analyst. Future surveys should record task exposure, project participation, training dates, mentoring, frequency of data use, and access to specialist colleagues. Following employees across several years and changes of assignment could show whether repeated use builds capability and whether a move away from operational work gradually erodes it.

From workflow profile to a bounded theoretical contribution

The paper's contribution is not a new validated scale. It is a way of organizing a familiar public-administration problem. A workflow-completeness lens keeps acquisition, cleaning, documentation, communication, and application visible alongside statistical analysis. The normative-operational confidence gap then names an empirically observable ordering without pretending that ethics and technical skill are opposite ends of one latent dimension. Finally, the multilevel reading prevents an individual self-rating from standing in for team or organizational capability. Together, these distinctions connect curriculum design, workforce development, and digital-government implementation while preserving the limits of the evidence.

Contribution to public administration education

The survey contains no evidence about the courses these respondents took, so it cannot assign the gap to universities. It does identify work that public administration programs are well placed to address. Prior studies find that MPA and related programs concentrate data education in conventional methods courses, where acquisition, documentation, tool use, and operational communication receive less time[6, 16]. Those are also the areas in which respondents reported the least confidence.

That correspondence supports a curriculum-design proposition rather than a causal claim about existing programs. Give students a public problem with imperfect data and ask them to produce a traceable data product, an analysis, a communication artifact, a decision recommendation, and an ethics record. The work should loop when new information changes the question. Such projects bring methods, technology, policy analysis, communication, and public values into the same assignment, which is also the direction taken by international digital-era curriculum initiatives[5, 13, 17].

Relevance beyond China

The numerical estimates should not be exported to other countries, but the diagnostic questions travel. Governments with very different administrative traditions still need personnel who can move from problem framing to documented evidence and accountable use. The Chinese case shows how a strong national digital agenda can coexist with uneven reported confidence at the point of implementation. Comparative work can now ask whether similar profiles appear under different recruitment systems, levels of administrative decentralization, and models of public-service education. The transferable contribution is therefore the workflow and multilevel logic, not a claim that 3.048 is a benchmark for public services elsewhere.

Implications For Education And Workforce Development

The survey did not evaluate a training intervention. The recommendations below are therefore design propositions grounded in the observed profile and the cited literature, not claims that a particular program will produce a measured effect.

Build a common floor and role-differentiated depth

A common floor should cover answerable questions, suitable sources, data quality, basic interpretation, uncertainty, rights, and the point at which specialist help is needed. Technical depth can then follow the job. A data specialist needs reproducible programming, deeper analysis, data engineering, and careful documentation. Program and policy managers spend more of their time interpreting results, evaluating claims, procuring systems, and translating evidence for a decision. Senior leaders need enough model-governance knowledge to challenge assumptions and keep responsibility visible.

Uniform technical depth would waste training time, while total delegation would leave leaders unable to question the systems for which they remain accountable. Role-based depth offers a practical middle ground. It also fits the hybrid-competency account of teams that combine analytical, domain, relational, and political capabilities[4]. This proposition follows from task differences and the literature; Table 4 does not establish a causal training need for any rank.

Redesign learning around authentic workflow products

Disconnected software demonstrations rarely resemble public work. A cumulative project can begin with a short data-needs statement, then require a source log, a data dictionary, recorded cleaning decisions, an analysis, an accessible visualization, a decision memo, and an ethics review. Students should choose tools to fit the problem. R or Python may support a reproducible analysis, while a spreadsheet, database, or visualization package may be the better choice elsewhere. Assessment can then reward traceability, judgment, and technical accuracy in the same piece of work, with each choice open to review.

Connect universities, public-service academies, and workplaces

Capability changes across a career. Universities can teach concepts and methods that travel across posts. Public-service academies--including party schools in the Chinese context--can place those foundations inside current administrative responsibilities, and employers can provide supervised projects, mentoring, and colleagues who review the work. A shared competency map and a portable portfolio would make that progression visible. Short courses need a follow-on assignment and feedback. This division of labor speaks directly to China's policy objective of strengthening the digital thinking and skills of public officials[3, 18] and to the OECD's emphasis on sustained workforce systems[13].

Embed ethics in technical decisions

During collection, students should state the lawful purpose, examine consent, and minimize what they take. Cleaning and modeling assignments need an account of exclusions, proxy variables, and possible bias. Uncertainty belongs on the visualization. Each proposed application should identify its human overseer, appeal route, logging requirements, and responsible office. These details give instructors and supervisors something they can inspect.

Table 5 pairs each observed weakness with a proposed training response and a concrete assessment indicator for future implementation, none of which was evaluated in this survey.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

Table 5. Evidence-to-Action Matrix

Diagnostic evidence	**Capability risk**	**Education/workforce response**	**Illustrative assessment indicator**
Collection was the lowest domain. API, web acquisition, metadata, and data dictionaries were weak	Analysts begin with poorly sourced or undocumented data	Require acquisition and documentation in every applied project	Reproducible source log and complete data dictionary
Analysis and communication were weak and strongly correlated	Results cannot be independently checked or used in decisions	Pair analysis with visualization, uncertainty, and decision writing	Executable analysis plus plain-language policy brief
Ethics awareness was high but behavior was not observed	Normative confidence may not translate into accountable practice	Integrate privacy, fairness, security, and contestability into technical tasks	Workflow-stage ethics and risk audit
Position profiles differed in operational and oversight areas	Decision authority may be separated from evaluative data literacy	Create a common literacy floor with role-specific advanced modules	Role-based demonstration and team capability map

Limitations And Future Research

Professional networks produced a targeted sample during a three-day window, from 27 to 29 September 2024. The invitation denominator and network-level recruitment counts were not retained, so a response rate cannot be calculated. That design offers no national, provincial, or organizational prevalence estimate. The networks operated in Hunan, Guangdong, and Sichuan, while the questionnaire itself recorded no workplace province. Repeated technical network identifiers suggest that some responses may also be clustered within shared connections or organizations; the survey did not record a non-identifying organizational cluster variable that would permit a defensible multilevel analysis. A stronger replication should use probability-based sampling where a sampling frame is available, or at least a transparent multistage quota design that covers more provinces, agency types, administrative levels, and functional departments. Wider coverage would allow researchers to distinguish territorial variation, agency mission, organizational resources, and individual role more cleanly than this study can.

The inventory records agreement with statements about one's own ability. Social desirability, confidence, reference-group standards, and familiarity with technical terms can all shape such answers. The ethics result makes this visible. So do the tiny SDs in a few small groups. A stronger assessment would pair self-report with objective competency tests, practical data tasks, scenario-based judgment items, data-quality diagnosis, visualization interpretation, and reproducible analysis exercises scored with transparent rubrics[1]. Such measures would show whether reported confidence corresponds to actual skill and whether respondents can transfer that skill to administrative problems.

The eight content areas came from a theory-informed workflow framework. The adaptation did not include an independent expert panel or cognitive interviews, and the areas have not been validated as eight reflective factors. High alpha coefficients and strong content-area correlations answer neither the question of discriminant validity nor the question of stability over time. Future validation should begin with expert review, cognitive interviews, and pilot testing, then use separate samples for exploratory and confirmatory factor analysis. Reliability reporting should extend beyond alpha to coefficient omega and test-retest evidence, and measurement-invariance tests should examine whether the inventory behaves similarly across regions, agencies, education levels, and positions. Criterion evidence should come from comparison with objective tasks. The English wording in

Supplementary Table S1 is a translation for reporting and requires separate linguistic validation before use as an English-language instrument.

Subgroup comparisons face a different problem. The age-51+ group had 10 people, the village/community group had 12, and the associate-degree group had 26. Several cross-classifications were empty. Welch tests handle unequal variance better than conventional procedures, but they cannot supply cases that were never sampled. A replication should stratify recruitment, or set quotas, so that education, organization, and position can be compared within observed combinations.

Several mechanisms discussed in the paper remain unmeasured. The questionnaire omitted job tenure, task exposure, training history, data-use frequency, access to specialists, infrastructure, learning climate, and leadership support. Those omissions block a mechanism test. Longitudinal research would be especially useful. A training evaluation could measure employees before instruction, after instruction, and after a period of supervised workplace use, combining self-report with practical tasks and portfolio evidence. The strongest designs would also examine whether gains translate into better public-sector decision-making, for example through clearer data requests, better-documented analyses, more transparent uncertainty communication, and more accountable use of evidence in policy or service decisions.

The platform did not ask for names, yet it retained timestamps and IP addresses. I excluded those fields from the literacy analysis. Any shared analytical file must remove them. Most importantly, submission should wait until the applicable ethics approval or exemption, the consent procedure, and the lawful handling of platform metadata have been verified and reported accurately.

CONCLUSION

Among the 510 retained respondents, ethics averaged 3.565 and data generation 3.541. Collection averaged 2.813, communication 2.867, and analysis 2.929. That spread is the paper's clearest result. Public-sector personnel in this sample reported greater confidence in responsible use and problem framing than in moving data through acquisition, analysis, and explanation. The finding is a diagnostic confidence profile, not a performance test, but it draws attention to the full chain on which digital-government capability depends.

The group means add texture, but small groups, low within-group variance, and empty cells prevent an independent-effects reading because education, organizational level, position, and likely task exposure are intertwined. A defensible development strategy begins with a common literacy floor and then varies depth by role and task. Teams can supply complementary expertise, while organizations determine whether personnel have data, specialist help, time to practise, and reasons to keep using what they learn.

For public administration education, the practical implication is concrete: students should finish a data course with a public product whose source, cleaning decisions, analysis, uncertainty, recommendation, and ethical choices can all be inspected. Government training needs the same discipline. A short course should lead into supervised use, mentoring, and a role-appropriate demonstration of skill. Comparative and longitudinal research can test whether this workflow-based diagnosis travels across administrative systems, whether training changes observed performance rather than confidence alone, and whether those gains improve the quality of public-sector decisions.

Declarations

Consent to participate

The questionnaire preface described the study purpose, estimated completion time, research-only use of responses, and researcher contact information, and the form did not request participant names. The platform nevertheless retained timestamps and IP addresses; these fields were excluded from the literacy analysis.

Conflict of interest

The author declares no conflict of interest.

Funding

This work was supported by the Guangdong Higher Education Association project, “Building and Practicing an Industry-University-Research Collaborative Education Model for Public Administration Programs Oriented toward Digital Governance” (Grant No. 23GYB96), and the Guangdong Provincial Department of Education project, “Coupling County Economic Resilience and Rural Revitalization in Eastern, Western, and Northern Guangdong under Industrial Transfer” (Grant No. 2023WTSCX153).

Author contributions

Haisheng Tang was responsible for conceptualization, methodology, investigation, data curation, formal analysis, validation, writing--original draft, writing--review and editing, project administration, and funding acquisition.

ACKNOWLEDGEMENTS

Author Completion Required: Add any acknowledgements with permission, or replace this sentence with "None."

Declaration of generative AI and AI-assisted technologies

During manuscript revision, generative-AI-assisted tools were used to support English translation, language editing, reference checking, code-based reproducibility checks, and document formatting. The author reviewed the source data, statistical outputs, citations, interpretations, and final text and accepts full responsibility for the work. No AI tool recruited participants or generated the survey responses.

REFERENCES

1. Bonikowska, A., Sanmartin, C., & Frenette, M. (2019). Data literacy: What it is and how to measure it in the public service. Statistics Canada, Analytical Studies: Methods and References, No. 22. https://publications.gc.ca/site/eng/9.877889/publication.html
2. Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. Public Administration Review, 81(5), 825-836. https://doi.org/10.1111/puar.13293
3. Central Committee of the Communist Party of China & State Council of the People's Republic of China. (2023). Overall layout plan for the construction of Digital China [In Chinese]. https://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/202302/t20230228_1048220.html
4. Dingelstad, J., Borst, R. T., & Meijer, A. (2022). Hybrid data competencies for municipal civil servants: An empirical analysis of the required competencies for data-driven decision-making. Public Personnel Management, 51(4), 458-490. https://doi.org/10.1177/00910260221111744
5. Eaves, D., Mergel, I., Clarke, A., Fetter, J., & Steinberg, T. (2026). Teaching public service in the digital age: Collectively building a global curriculum to teach digital-era competencies to public managers. Information Polity, 31(1), 43-60. https://doi.org/10.1177/15701255251382598
6. Ho, K. K. W., Li, N., & Sayama, K. C. (2024). Equip public managers with data analytics skills: A proposal for the new generation of MPA/MPP programs with data science track. Library Hi Tech, 42(4), 1286-1304. https://doi.org/10.1108/LHT-07-2022-0320
7. Hsu, Y.-C. (2024). Mapping the landscape of data science education in higher general education in Taiwan: A comprehensive syllabi analysis. Education Sciences, 14(7), Article 763. https://doi.org/10.3390/educsci14070763
8. Kleinschmit, S. (2019). An inclusive framework for data science competencies in NASPAA MPA and MPP programs [White paper]. Network of Schools of Public Policy, Affairs, and Administration.

https://www.naspaa.org/sites/default/files/docs/2019-09/NASPAA%20Whitepaper%20-%20Kleinschmit.pdf

9. Lavertu, S. (2016). We all need help: "Big data" and the mismeasure of public administration. Public Administration Review, 76(6), 864-872. https://doi.org/10.1111/puar.12436
10. Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. Government Information Quarterly, 40(1), Article 101774. https://doi.org/10.1016/j.giq.2022.101774
11. Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. Government Information Quarterly, 36(4), Article 101385. https://doi.org/10.1016/j.giq.2019.06.002
12. Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. Public Administration Review, 76(6), 928-937. https://doi.org/10.1111/puar.12625
13. Organisation for Economic Co-operation and Development. (2021). The OECD framework for digital talent and skills in the public sector (OECD Working Papers on Public Governance No. 45). OECD Publishing. https://doi.org/10.1787/4e7c3f58-en
14. Ongena, G. (2023). Data literacy for improving governmental performance: A competence-based approach and multidimensional operationalization. Digital Business, 3(1), Article 100050. https://doi.org/10.1016/j.digbus.2022.100050
15. Overton, M., & Kleinschmit, S. (2022). Data science literacy: Toward a philosophy of accessible and adaptable data science skill development in public administration programs. Teaching Public Administration, 40(3), 354-365. https://doi.org/10.1177/01447394211004990
16. Overton, M., & Kleinschmit, S. (2023). Transforming research methods education through data science literacy. Teaching Public Administration, 41(2), 149-169. https://doi.org/10.1177/01447394221084488
17. Ruijer, E., & Lember, V. (2026). Public sector capabilities and individual skills needed for the digital transition: A multilevel perspective on teaching digital governance. Information Polity, 31(1), 4-12. https://doi.org/10.1177/15701255261423332
18. State Council of the People's Republic of China. (2022). Guiding opinions on strengthening the construction of digital government [In Chinese] (State Council Document No. 14). https://app.www.gov.cn/govdata/gov/202206/23/486412/article.html
19. United Nations Department of Economic and Social Affairs. (2024). UN e-government survey 2024: Accelerating digital transformation for sustainable development. United Nations. https://desapublications.un.org/publications/un-e-government-survey-2024
20. Zhang, H., & Yang, T. (2023). The concept, structure and measurement of civil servants' data literacy [In Chinese]. Digital Library Forum, (1), 58-65. https://doi.org/10.3772/j.issn.1673-2286.2023.01.008