




Digital transformation of citizens' evaluations of public service delivery: evidence from China

Chengwei Wang^{1,2} · Liang Ma² 

Received: 11 October 2022 / Accepted: 21 October 2022 / Published online: 4 November 2022
© The Author(s), under exclusive licence to Institute for Global Public Policy, Fudan University 2022

Abstract

Thanks to the development of digital technologies and their applications in public service delivery, ratings by citizens of service quality have been transforming from paper/phone surveys to digital interfaces, similar to what has happened in e-commerce. How does digitalization drive the change in ratings by citizens? Are citizens more satisfied when they rate public services online? We provide empirical answers to these questions based on data from the government service evaluation system in a Chinese city. We found that digital interfaces facilitate the behavior of citizens making evaluations and boost citizen satisfaction compared with offline channels. Specifically, interfaces displayed on mobile applications significantly facilitate evaluations by citizens and improve citizen satisfaction. The results provide theoretical and practical implications for understanding the digital transformation of public service ratings.

Keywords Evaluations by citizens · E-government · Digital transformation · Administrative burden · Public service performance

1 Introduction

Since the beginning of the twentieth century, government performance measurement has dramatically experienced two transformations. The first is the conversion of performance information sources. Performance information is not only derived from objective measurement data; citizen satisfaction surveys have become crucial evaluation tools for government performance evaluation (Gao, 2012), as they help

✉ Liang Ma
liangma@ruc.edu.cn

Chengwei Wang
chengweiw96@qq.com

¹ School of Public Administration and Policy, Renmin University of China, 59 Zhongguancun Avenue, Haidian District, Beijing 100872, China

² Department of Public Policy, City University of Hong Kong, Hong Kong, China

formalize governments' responsiveness to society through public engagement (So, 2014). The second is the digitalization of evaluation methods, similar to what has happened in e-commerce, such as Amazon's user review. With government use of information and communication technologies (ICTs), performance measurement is heading towards digitalization. Evaluations by citizens are no longer limited to paper/phone questionnaires but consists of various digital interfaces, such as QR codes, government apps, and government websites.

To follow the transformation trend, the State Council (the cabinet) in China initiated the Government Service Evaluation System (GSES) to transform local public service evaluation at the end of 2019 (Fan et al., 2022). It was implemented nationwide in 2021. The GSES is a top-down public service evaluation reform initiated by the Chinese central government. The GSES enables citizens to rate online services more efficiently, and ratings made by citizens can be traced and provide feedback to help improve public service performance. Local governments in China, as policy responders, have experienced an unbalanced transformation process. During this process, public service evaluation interfaces have been gradually digitalized from offline to online across the country, which provides an excellent context in which to observe the digital transformation of evaluations by citizens of public service delivery.

As a large stream of literature in Public Administration, ratings and evaluations by citizens have been discussed and studied for a long period, thereby providing numerical evidence for this research topic (Ma, 2017). Most studies have concentrated on: (1) how to use data on evaluations by citizens for performance measurement (Swindell & Kelly, 2000); (2) how the government responds to evaluations by citizens (Wandaogo, 2022); and (3) what drives citizens to make ratings and evaluations (Ma & Wu, 2020). However, with the wide application of ICTs and digitalization of ratings by citizens, little is known about how digital transformation of evaluations by citizens is taking place and what it will bring about.

Following the practical exploration of digital government and theoretical research of public administration, this study expands the e-government research by borrowing the concept of administrative burden in government-citizen interaction studies to explain how ICTs and digitalization reduce the administrative burden of citizens and result in a boost of evaluation behavior and citizen satisfaction. In this study, we use the case of the GSES to explore how rating by citizens changes and drives such a change.

We specifically aim to answer three research questions:

- (1) Does the number of evaluations by citizens increase with digital transformation? How do the citizen-rating patterns change with digital transformation?
- (2) Do digital interfaces facilitate evaluations by citizens? Are citizens more willing to make evaluations through digital interfaces?
- (3) Do digital interfaces improve citizen satisfaction? Are citizens more satisfied with digital citizen–state interactions?

We used the data from the GSES in a Chinese city and found that digital interfaces facilitate citizens to make evaluations and boost citizen satisfaction compared

with offline channels. Specifically, interfaces displayed on mobile applications significantly facilitate evaluations by citizens and improve citizen satisfaction. The results provide theoretical and practical implications for understanding the digital transformation of public service ratings.

The rest of this paper is structured as follows. “Literature review and theoretical hypotheses” Section discusses the influence of digitalization on citizen–government interaction, followed by the relationship between e-government and administrative burden. “Empirical Context: GSES in China” Section introduces the GSES in China, which was the context of this research. In “Data and Methods” Section, we propose three hypotheses based on theory. “Result” Section provides detailed information on our data and method. We finally present the results, discussions and conclusions.

2 Literature review and theoretical hypotheses

2.1 Digitalization and citizen–government interaction

How does digitalization influence evaluations by citizens? This theoretical question can be traced back to the interaction between government and citizens since citizens' behavior of making evaluations is a significant feedback link in the loop of citizen–government interaction. The relationship between officials and the public became a focal point in the late 1970s. Starting from the 1980s, many competing theories focused on the interaction between citizens and government, including the concept of street-level bureaucracy raised by Lipsky (1983, 2010), administrative burden theorized by Moynihan and colleagues (Moynihan et al., 2015), and the public encounter notion introduced by Goodsell (1981). The first two theories have been well developed since 1980, while the public encounter theory is scarcely explored. Scholars are curious about “what” is the in-between of citizens and public professionals, while “how” it happens is the core concern of public encounter theory (Bartels, 2013).

Goodsell (1981) first introduced the concept of “public encounter” in his book *The Public Encounter: Where State and Citizen Meet*, investigating the form-related issues inside government–citizen interaction. Goodsell defined public encounter as “the interaction of citizens and officials as they communicate to conduct business”. Since then, the public encounter has been scatteringly studied in different subjects but has never formed a systematic research stream. In 2012, Bartels (2013) proposed a framework of public encounters to examine the everyday interaction between public professionals and citizens.

Based on the previous studies, we define public encounter as form-related issues within government–citizen interactions, such as who initiates the encounter, the purpose of the encounter, and the encounter's timing and scope.

The public encounter is summarized into four dimensions (Lindgren et al., 2019), including (1) the nature and purpose of the encounter; (2) the communication forms and setting in which the encounter occurs; (3) the central actors involved; and (4) the encounters' initiation, timing, and scope. Based on these four dimensions, we

Table 1 A framework of traditional and digital channels of evaluations by citizens

Dimension	Traditional encounter	Digital encounter
Initiator	Public official; citizens	Public official; citizens; digital system
Provider	Public official; self-service machines	Public official; self-service machines; AI
Interaction form	Telephone; face-to-face; SMS; letter and visit, etc	Website; app; social media, etc.
Interaction setting	Government office; institution building; shop; home	Computer; smartphone
Timing	Office hours	Office hours; any time

developed a framework for analyzing the digital transformation of public encounters as shown in Table 1.

Practically speaking, with the government application of ICTs, many form-related issues have subtly changed in public encounters. Take the consumer coupons released by governments in the COVID-19 pandemic as an example. In the past, consumption coupons could only be obtained by queuing up offline, but now citizens are able to collect and consume them on multiple digital payment platforms under a hybrid mode. During this offline-to-hybrid transition, the administrative burden faced by citizens was significantly reduced, which in turn changed citizens' consumption behavior. The core impact of administrative burden calls for a need to borrow the notion of administrative burden to fill in the public encounter framework.

2.2 Public encounter framework with the notion of administrative burden

Administrative burden has been explored in different streams of research, including business environments (Liao, 2020), street-level bureaucracy (Bell et al., 2021), red tape (Brown et al., 2021), and government program management (Keiser & Miller, 2020). Moynihan and colleagues (Moynihan et al., 2015) first conceptualized administrative burden as three categories of costs that individuals experience in citizen–government interactions, including learning, compliance, and psychological costs.

In public encounters, the learning, psychological, and compliance costs that citizens feel and experience constitute administrative burdens, which directly pose challenges to policy implementation. For example, Australia's 2018 National Disability Services survey shows that in relation to service providers, administrative burden is the challenge most commented on (Carey et al., 2020). Administrative burdens pose other challenges to citizen–government interactions, such as enhancing racial discrimination and reducing government transparency, fairness, and effectiveness (Heinrich, 2018).

Empirical evidence shows the negative influence of administrative burden. However, limited attention has been paid to how to reduce it effectively. Scattered evidence shows that e-government will significantly reduce the administrative burden of businesses (Arendsen et al., 2014). What remains to be further explored is whether and how digitalization will change individuals' administrative burdens

during citizen–government interaction. This research gap motivated us to link the digitalization of public encounters with administrative burden theory (see Table 2).

Based on previous research and vivid experience within government practices, we expect that the application of ICTs will transform traditional public encounters into digital forms, decreasing administrative burdens in citizen–government interactions, facilitating their participation behavior, and boosting their satisfaction. The four core elements of public encounters concluded in Table 1 provide a framework to illustrate deeply how the chain process operates.

2.2.1 Initiators

The first aspect is the *initiators* of public encounters. Depending on the purpose and nature of the public encounter, citizens and public officials can initiate interactions. For example, citizens may start an encounter by applying for social assistance. The government may also create an encounter for the replacement of ID cards.

However, compared with traditional encounters, the initiator of digital encounters could be a machine; some are created automatically by the system. For example, the Wisconsin government started an auto-enrollment when implementing Badger Care Plus, a Medicaid program in the USA. The auto-enrollment technique uses ICTs and government data to automatically identify citizens who meet the eligibility criteria for a government program, which increased the take-up of health care services in the state of Wisconsin (Herd et al., 2013).

This initiation system is introduced as “robotic bureaucracy” in many countries (Bozeman & Youtie, 2020), which initiates public encounters automatically via an online system or machine. This process eliminates learning costs for individuals (Herd et al., 2013) since citizens do not need to check their eligibility before applying for a service.

In the case of the GSES, the initiator of an evaluation by a citizen is also transformed from street-level bureaucracy to “robotic bureaucracy”. Before this reform, most public encounters were offline, and the corresponding initiators of ratings by citizens were civil servants who just performed services (see Fig. 1). They were willing to selectively initiate evaluations by citizens driven by a motivation to achieve higher ratings. Without their initiative, recommendation, and help, it costly for citizens to make a rating behavior. However, after the implementation of the GSES, “robotic bureaucracy” automatically issues a rating request for citizens via their smartphones or self-service machines (see Fig. 2), which boosts their rating behavior.

We expect the number of ratings by citizens to increase with the development of the GSES, and propose the first hypothesis as follows:

H1: The number of ratings by citizens will increase with the development of the GSES.

2.2.2 Providers

As for the *providers*, in traditional public encounters, public services are provided by frontline public officials and self-service machines in government buildings.

Table 2 Administrative burdens in citizen–state interactions

Type of cost	Definition	Example
Learning costs	Cost of collecting information about public services and assessing whether they are relevant	Tom spends money checking whether he is eligible under the criterion of medical welfare before applying for it
Psychological costs	Cost of negative perceptions and feelings during application and participation	Mike feels ashamed when he lines up for food aid for the poor
Compliance costs	Cost of following rules and requirements	Vivian completes application forms and prints them out before applying for a driving license



Fig. 1 Traditional initiators of evaluations by citizens

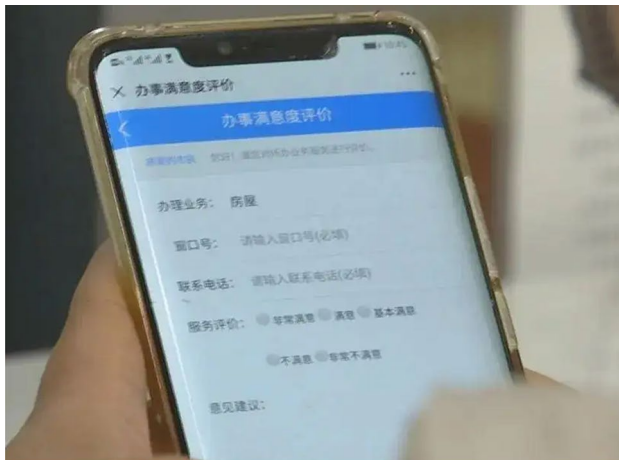


Fig. 2 “Robotic bureaucracy” initiators of evaluations by citizens

However, as for digital encounters, artificial intelligence (AI) could provide services. For example, many government hotline operators in China are robots.

The use of AI by public service providers reduces the psychological costs of individuals. Previous evidence points to unfair and inefficient services that street-level bureaucracy intentionally provides (Barnes & Henly, 2018; Peeters et al., 2018). When confronted with such unfair treatment, citizens may have negative perceptions. However, when services are administered by AI robots, their

non-discriminatory attitude will reduce the psychological burden that individuals face, which will also facilitate citizens' satisfaction with government.

2.2.3 Interactions

Concerning the *interaction between frontline officials and citizens*, traditional public encounters occur through limited media and settings, including face-to-face, telephone, mobile SMS, and even letters. As governments all around the world begin to apply modern technologies in the provision of public services, in addition to the traditional ways, citizens can interact with the government through official websites, apps, and even social media (e.g., WeChat, Alipay, and Twitter).

The phenomenon is deeply rooted in the change in *interaction settings*. Traditional public encounters typically take place in government offices or agency buildings. Besides, in some regions, local people can pay for public services (such as paying electricity and water bills) at convenience stores because the government has excellent collaborations with enterprises. In addition, some public encounters take place in citizens' homes. As for digital encounters, citizens can easily interact with government officials through computers and mobile phones.

With the change of interaction settings from physical to virtual, learning costs, psychological costs, and compliance costs faced by individuals will all be reduced. Citizens can browse related information online whenever they want, which hugely lessens the learning costs (Herd et al., 2013). In addition, the self-service process on social media and online platforms lowers the chances of experiencing negative feelings due to burdensome face-to-face contact with street-level bureaucracy. Furthermore, digital platforms allow the submission of digital instead of paper materials (Brown et al., 2021), thereby reducing compliance costs.

2.2.4 Duration

Regarding timing, traditional public encounters could only occur during office hours. However, with the help of ICTs, citizens can interact with the government at any time through the internet. Changing the timing from office hours to any time eliminates compliance costs. For example, individuals do not have to go to government buildings weekly just to submit an application form. More seriously, in real-life situations, when citizens apply for a public service offline, if any provided application materials are missing, it is necessary to redo the paper materials. Sometimes it takes four to five trips to a government agency during office hours just to successfully submit a public service application, which badly squeezes working time. The added flexibility hugely decreases the compliance costs faced by office workers.

In summary, with the digital transformation of public encounters, the initiator, provider, interaction, and timing of citizen–government interaction have undergone massive shifts towards a low-burden approach, as shown in Fig. 3. Citizens can save money and time owing to public services being provided by digital interfaces. The administrative burden faced by individuals is primarily reduced by the digital transformation of citizen–government interaction.

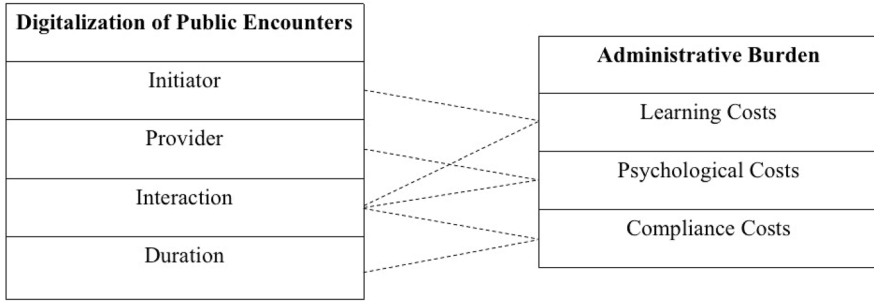


Fig. 3 Linkage of public encounters and administrative burdens

Will the reduction of administrative burden by digital interfaces facilitate citizens' behavior of making evaluations? Practically speaking, eliminating administrative burden could directly enhance the willingness and possibility of citizens to participate in public encounters and improve citizen satisfaction. For example, previous research shows that reduced administrative burden increases health care service take-up (Herd et al., 2013). Evidence has also been found in the area of animal welfare that burdensome paperwork results in farmers failing to comply with statutory requirements (Escobar & Demeritt, 2016).

Consistent with past studies, as a type of citizen–state interaction, we expect digital transformation to facilitate citizens' behavior of making ratings in the following ways.

First, as for *interaction settings*, instant evaluation after real experience makes evaluation an easy task without extra effort to recall or check for related information. There is often a considerable time lag between traditional paper or SMS evaluations and public service delivery, resulting in increased learning costs for evaluations by citizens. When evaluations are transformed from conventional to digital channels, after receiving public services, the system will pop up an evaluation window on websites, apps, and social media, making the evaluations a no-brainer process and lowering the learning cost.

Figure 4 shows the digital interface of a government app. Citizens only need to click the digital button to choose one option from the five-point Likert scale, which reduces their compliance burden to the least.

The second concern is the interaction settings. Compared with face-to-face evaluations, online ratings hugely reduce the psychological costs faced by individuals. When confronted with terrible services, some citizens are afraid to give bad face-to-face evaluations for fear of reprisals. For example, news in 2021 showed that a man ordered a takeaway and gave a bad delivery rating. After the evaluation, the deliveryman came to his apartment to bargain with him. This real-life story reveals the psychological burdens of a face-to-face assessment. However, anonymous online evaluation protects the personal information of individuals, which reduces psychological burdens.

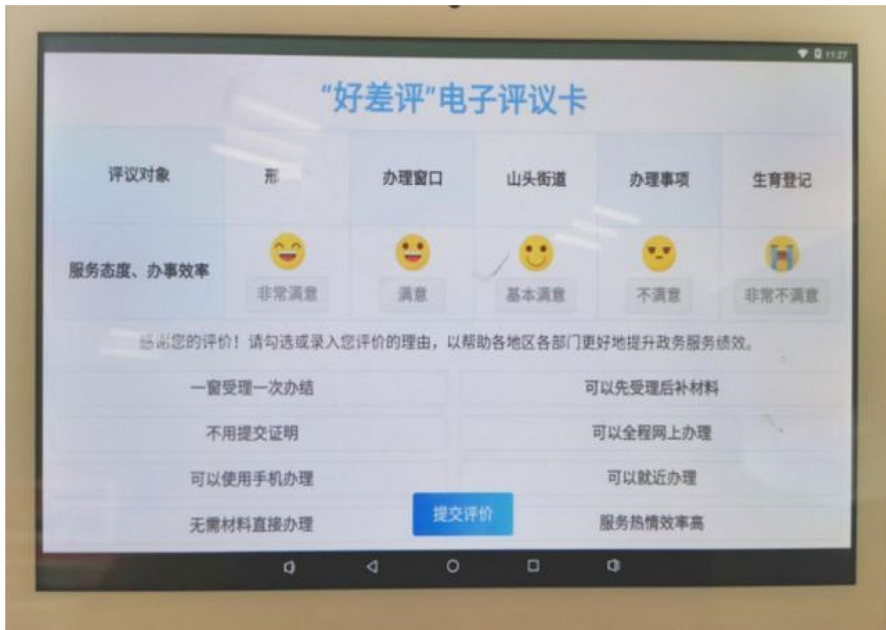


Fig. 4 Digital interface of evaluations by citizens

To summarize, compared with the traditional face-to-face evaluation method, the learning, psychological, and compliance costs of evaluation via digital interfaces will become lower. We thus propose the second hypothesis as follows:

H2: Compared with face-to-face encounters, digital interfaces facilitate ratings by citizens.

We expect that digital interfaces will improve citizen satisfaction. This is an intriguing practical question that lacks empirical evidence. Scholars believe that digital interfaces will improve citizen satisfaction, but a recent between-subject lab experiment found no significant causal relationship between these two factors (Prokop & Tepe, 2021). The lab experiment is often limited in external validity, and we want to retest it using second-hand data.

We expect that increased citizen satisfaction is driven by the digitalized evaluation process rather than the evaluated services per se. Given the same evaluated services, what we would like to explore is how the digital evaluation process leads to a higher level of satisfaction among citizens. Digital interfaces will decrease the administrative burdens of citizen–state interactions and improve citizens’ satisfaction with the government.

First, as for providers, the use of AI by public service providers reduces the possibility of intentionally unfair and inefficient services being provided by street-level bureaucracy, which will increase citizen satisfaction. The introduction of digital evaluation also enables citizens to more strongly hold government accountable for better public services.

Second, digital interfaces lessen administrative burdens in three aspects, as illustrated in the framework, which will increase citizen satisfaction (Tummers et al., 2016). Given the reduced administrative burdens, we expect that citizens would be more likely to have higher levels of satisfaction with public services.

Based on the above illustration, in our research context, we view the evaluation process as a kind of citizen–government interaction and argue that the digital evaluation process will lead to better satisfaction compared with the traditional evaluation process. To summarize, we propose the third hypothesis as follows:

H3: Compared with face-to-face encounters, digital interfaces will produce a higher level of citizen satisfaction.

Despite plenty of empirical evidence for the relationship between digitalization and administrative burden (Arendsen et al., 2014; Wandaogo, 2022), as well as that for administrative burden, citizen participation (Escobar & Demeritt, 2016; Herd et al., 2013), and citizen satisfaction (Tummers et al., 2016), there is a lack of empirical evidence regarding the theoretical question of whether digital transformation will affect citizens' participation in and satisfaction with citizen–state interactions. Based on the above arguments, we will provide empirical evidence of the GSES in China to fill this research gap.

3 Empirical context: GSES in China

In January 2019, the State Council of China declared the implementation of the GSES across the country by the end of 2020.¹ This proposal aimed to develop a comprehensive online and offline evaluation system in China, through which citizens could make ratings of all kinds of government services through complete coverage of service channels. The GSES system transforms evaluation by citizens from traditional offline evaluation to a combination of offline and digital evaluation, providing an ideal context for exploring the changing dynamics of ratings made by citizens.

Specifically, as a national reform of evaluation by citizens, the GSES advocates online evaluation and reduces the administrative burden of offline evaluation. Citizens are encouraged to make evaluations after consuming public services. The GSES has expanded the traditional evaluation objects to full coverage of multi-level hierarchical governments. Individuals can submit ratings to central agencies and all four local governments (e.g., provincial, municipal, district/county, and township/subdistrict governments). Local governments are encouraged to include all kinds of public services in the evaluation system. Evaluations made by citizens go deep into all aspects of public services, making evaluation by citizens something that is at your fingertips.

How can the central government implement this reform across the whole country? The GSES is mainly implemented through Chinese hierarchical bureaucracy. Specifically, the provincial governments took responsibility for the construction

¹ See the website of the Chinese central government, http://www.gov.cn/zhengce/content/2019-12/17/content_5461829.htm.

work, and the municipal governments gave guidance. By the end of 2020, the GSES had been fully established nationwide.

The developing process of the GSES shows the digital transformation of evaluation by citizens in China. What is unique in this case is that the development of the reform is diverse across regions. Some regions attach great importance to this work and started the transformation process very early; however, in some places, they started late and finished the reform on the deadline day. Therefore, from the beginning of 2020 to the end of 2021, the transformation varied across provinces, cities, and even counties, which provides an excellent window to observe the digital transformation of evaluation by citizens regarding public service delivery.

Besides, the implementation of the GSES not only facilitates evaluations by citizens but also provides feedback for better government decision-making. First, the digital interface can help collect more realistic evaluation data. When making face-to-face evaluations, citizens are more inclined to praise the government in order to avoid retaliations. Consequently, the data drawn from the digital interface is more authentic, which will help promote government performance improvement.

Second, data drawn from the GSES can be used for agency performance measurement, similar to the case of CitiStat in the US (Behn, 2008). According to the interview that we conducted in Xi'an, evaluations made by citizens are used in various ways. A) Local governments conduct one-to-one return visits to citizens who have made negative evaluations and help improve the related services. B) Big data analytics are used to screen out problems with an abnormal number of complaints, and the focal issues will be checked and corrected. C) Upper-level governments will reward departments which receive positive evaluations. Consequently, a digital interface helps to obtain authentic data to facilitate future performance improvement.

Among all cities in China, the city of Xi'an also carried out this work. Xi'an is the capital of Shaanxi province, a sub-provincial city in western China. Xi'an completed the GSES by the end of 2020, and the reform has been selected as a best practice in business environment improvement by the National Development and Reform Commission. We chose Xi'an as our research context mainly for two reasons: (1) The good practices of the GSES in Xi'an have accumulated a large amount of data. The high-quality and authentic evaluation data from citizens can provide solid support for deep research; (2) we have formal cooperation with the Xi'an government and have access to all the relevant evaluation data. Based on the case of Xi'an, we aimed to explore the implications of the GSES reform for ratings made by citizens and performance improvement.

4 Data and methods

4.1 Data structure and sources

We created a set of variables from three data sources, including official documents of the Xi'an government, the 7th National Census Data, and Xi'an statistical yearbook in 2021. We collected these second-hand data and manually coded all the variables, resulting in a panel dataset with 441 observations of 21 regions in 21 months.

It should be noted that region-month was the unit of analysis. Twenty-one research objects comprised city-level government, 11 districts, 2 counties, and 7 national and provincial key development zones. All the observations were drawn from February 2020 to October 2021, as the reform started in January 2020, and the GSES was implemented by the end of 2021.

It should be noted that the COVID-19 pandemic since early 2020 did not pose a significant threat to the model specifications despite it having the potential to be a major factor involved in digital and AI transformation. On the one hand, the data were collected after the epidemic outbreak, so the impact of the epidemic on the dependent variable was overall rather than sudden. On the other hand, we used the Hause–Taylor model to control for the time effect, regional effect, and individual effect, which can effectively manage the impact of the epidemic without biasing the estimation.

4.2 Dependent variables

The dependent variables were drawn from monthly official documents issued by the Xi'an Bureau of Administration Examination and Approval Service from February 2020 to October 2021. There were four dependent variables, including the number of evaluations (*Number*), the monthly growth rate of evaluations (*Growth*), the share of positive evaluations (*Positive*), and the share of negative evaluations (*Negative*).

The growth rate of ratings was calculated from raw evaluation data by the following formula. $Evaluation_{i,t}$ represents the number of evaluations made by citizens in region i in month t :

$$Growth_{i,t} = (Evaluation_{i,t} - Evaluation_{i,t-1}) / Evaluation_{i,t-1}$$

Among 21 regions in 21 months, 1 region had missing values for population data, and 2 areas had missing values for GDP data.

4.3 Independent variables

As for the independent variables, we used the proportion of five evaluation channels to measure the use of online and offline evaluations. There are five main channels for ratings by citizens in the GSES, including government websites, government apps, offline public service windows, QR codes posted in public affairs centers, and mobile phone SMS. Government websites, QR codes, and government apps represent online evaluations among the five channels, while offline public service windows and mobile phone SMS are offline evaluations. Comparing the proportion of online and offline assessments could provide us with a measurement of digitalization.

4.4 Control variables

We included six control variables in our models. The data about type, population, education, and gender were drawn from the 7th National Census. Data on GDP were

Table 3 Descriptive statistics of key variables

Category	Variable	<i>N</i>	Mean	SD	Min	Max
Dependent variables	Number	441	14,579.1	21,666.59	0	188,413
	Growth	396	19.83292	184.871	- 1	3186
	Positive	441	0.94	0.23	0	1
	Negative	441	0	0.02	0	0.25
Independent variables	Website	441	7.00%	19.00%	0.00%	100.00%
	App	441	2.00%	8.00%	0.00%	100.00%
	QR	441	15.00%	28.00%	0.00%	100.00%
	Offline	441	46.00%	35.00%	0.00%	100.00%
	SMS	441	5.00%	16.00%	0.00%	100.00%
Control variables	Type 1	441	0.05	0.21	0	1
	Type 2	441	0.52	0.5	0	1
	Type 3	441	0.10	0.29	0	1
	GDP	399	595.75	565.42	45	2410
	GDPs	252	2.67%	4.96%	- 5.90%	15.30%
	Population	420	4.50%	2.39%	0.17%	9.28%
	Education	441	11.81	1.27	9.44	14.15
Gender	441	108.91	14.93	94.79	168.33	

drawn from Xi'an statistical yearbook in 2021. Three dummy variables were used to control for different types of government, including “*Type 1*” (dummy = 1 if evaluation comes from city level), “*Type 2*” (dummy = 1 if evaluation comes from one of the 11 districts), “*Type 3*” (dummy = 1 if evaluation comes from one of the two counties). The population represents the proportion of the local population among all citizens, ranging from 0.17 to 9.28%. *Education* represents the average years of education, ranging from 9.44 to 14.15. Gender represents the ratio of females to males (with females as 100), ranging from 94.79 to 168.33. The unit of GDP is RMB100,000,000. We also included the GDP growth rate as a control variable, ranging from - 5.90 to 15.30%. Table 3 lists the summary statistics of all variables.

4.5 Model specifications

We applied the Hausman–Taylor model to link the treatments with dependent variables, including the growth rate of evaluations and the probability of positive and negative evaluations (Baltagi et al., 2003). We chose the Hausman–Taylor model instead of fixed-effect or random-effect models due to our panel data structure.

The independent and dependent variables were monthly data. However, the control variables were drawn from statistical books or censuses, which are annual data. If we had used a fixed-effect model, we could not have predicted the influence of control variables. If we had used a random-effect model, the estimators would be largely biased. Consequently, we adopted the Hausman–Taylor model, one of the

Table 4 Description of the evaluation data

Year/month	Number	Growth rate	Website	APP	QR	Offline	SMS
202,002	148	–	26.53%	0.37%	0.00%	0.00%	6.43%
202,003	957	546.62%	32.67%	6.61%	0.79%	13.21%	18.11%
202,004	5924	519.02%	22.49%	6.51%	5.99%	17.50%	33.20%
202,005	38,836	555.57%	10.39%	9.92%	20.76%	30.43%	23.73%
202,006	64,406	65.84%	8.53%	1.46%	32.74%	50.38%	6.88%
202,007	129,080	100.42%	2.15%	1.04%	41.78%	51.98%	3.00%
202,008	117,780	– 8.75%	3.43%	0.66%	32.29%	61.64%	1.96%
202,009	146,760	24.61%	1.61%	1.12%	31.86%	64.19%	1.13%
202,010	128,944	– 12.14%	6.23%	0.75%	30.05%	62.74%	0.01%
202,011	198,870	54.23%	2.91%	0.71%	27.69%	68.48%	0.01%
202,012	718,899	261.49%	1.43%	0.30%	5.71%	43.56%	0.00%
202,101	695,668	– 3.23%	1.82%	0.32%	3.75%	48.84%	0.00%
202,102	427,535	– 38.54%	0.51%	0.21%	4.42%	48.53%	0.00%
202,103	670,178	56.75%	1.60%	0.32%	6.58%	51.78%	0.00%
202,104	385,073	– 42.54%	1.72%	0.26%	10.84%	55.06%	0.00%
202,105	462,714	20.16%	2.00%	0.26%	6.83%	54.16%	0.00%
202,106	337,997	– 26.95%	2.39%	0.22%	10.16%	57.58%	0.00%
202,107	551,539	63.18%	3.03%	0.28%	6.72%	50.94%	0.00%
202,108	383,604	– 30.45%	5.66%	0.35%	12.06%	50.85%	0.22%
202,109	403,598	5.21%	4.44%	0.49%	11.69%	49.79%	0.14%
202,110	560,874	38.97%	14.89%	7.25%	9.95%	36.62%	0.10%

mostly used models, to address the problems caused by inconsistent panel datasets and to estimate differences across units.

5 Results

5.1 Descriptive statistics

We first describe our data. As the descriptive statistics show in Table 3, during the GSES reform, evaluations by citizens soared, with a maximum increase of 3,186 times more than the previous month. Citizens generally hold positive attitudes towards public services in Xi'an, with an average proportion of 0.94. The evaluation number of five channels varies substantially, suggesting that it is relevant to explore their variations.

Among the five channels, the proportion of offline evaluations ranked first, with an average rate of 46%. The highest proportion shows the dominant role of offline evaluation even in the digitalization age, suggesting that offline channels are still the main route of public service provision. The government website

Table 5 Regression model estimates

Model variable	Model 1 Number	Model 2 Growth	Model 3 Positive	Model 4 Negative
Website	- 1.79* (1.08)	- 21.99 (109.80)	0.38*** (0.13)	0.02 (0.02)
App	26.73*** (5.95)	6,899.53*** (511.51)	0.97* (0.56)	- 0.20 (0.16)
QR	0.72 (0.75)	49.08* (28.01)	0.14** (0.07)	- 0.00 (0.00)
Offline	1.60*** (0.59)	32.48 (20.49)	0.09 (0.07)	- 0.00 (0.00)
SMS	- 0.67 (1.29)	291.89** (118.12)	0.08 (0.15)	- 0.00 (0.01)
Type 1	-	-	-	-
Type 2	-	-	-	-
Type 3	0.61 (0.63)	37.12* (22.21)	- 0.11** (0.04)	- 0.00 (0.00)
GDP	- 0.00 (0.00)	0.05*** (0.01)	0.00** (0.00)	0.00 (0.00)
GDP growth	- 0.01 (0.05)	0.12 (1.37)	- 0.01** (0.00)	0.00 (0.00)
Population	0.32*** (0.10)	- 4.99 (3.13)	- 0.00 (0.01)	- 0.00 (0.00)
Education	1.36*** (0.25)	6.54 (6.52)	- 0.02 (0.02)	- 0.00 (0.00)
Gender	0.24*** (0.05)	- 0.61 (1.39)	- 0.01* (0.00)	- 0.00 (0.00)
<i>N</i>	231	223	240	240
χ^2	1324.42	35959.03	130.13	128.05
Month fixed effects	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes

Note: The dependent variables in Model 1–Model 4 are the number of evaluations, the growth of evaluations, the proportion of positive evaluations, and the proportion of negative evaluations, respectively. All the independent variables are lagged by 1 month. “Number” is logged. The Hausman–Taylor models are estimated. The time and district dummies are included but not shown to save space. Robust standard errors clustered by region are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed)

accounts for the dominant part of the three online evaluation channels, with an average rate of 7%. GDP growth varies in different areas, with a standard variation of 4.96%. The average number of years of education of citizens in Xi’an is 11.81 years. The average female ratio is 108.91, indicating a higher female proportion in Xi’an.

Table 4 reports the description analysis of evaluation data from February 2020 to November 2021. The number of evaluations by citizens experienced dramatic boosts from February 2020 to December 2020, at the end of the GSES reform. The results show that the number of ratings by citizens increased with the development of the GSES, indicating that H1 is supported.

To conduct further specific analyses of the five channels separately, we found that the development of online and offline evaluations varies. Even though the Xi’an government initiated GSES reforms, most public service delivery still needs to be provided offline. After the boost of online ratings brought about by the reform, the proportion of online ratings decreased, which indicates that citizens are still more inclined to make ratings according to the form of public service provision.

5.2 Regression analyses

Table 5 reports the regression model estimates. The empirical results consistently support all the hypotheses. With the development of the GSES, the number of evaluations by citizens increased. In addition, digital interfaces facilitate evaluations by citizens and improve citizen satisfaction. All the models included jurisdictional and monthly fixed effects, and robust standard errors clustered by regions are reported. All the independent variables lagged by 1 month.

To test the relationship between digital interfaces and evaluations by citizens, we ran Model 1 and Model 2. Model 1 illustrates that the proportion of app evaluations is significantly positively associated with the number of evaluations. Since we took the logarithm of *number*, the result indicates that a 1% increase in app evaluations generates a 26.73% increase in evaluations in the next month. We used the growth rate of evaluations as an alternative dependent variable in Model 2 for robustness checking, and the result was consistent. Thus, H1 is supported.

Specifically, a region with a 1% increase in app evaluations generates a 6899.53 times enhancement in the growth of evaluations in the next month. In addition, the relationship between offline evaluation and the growth of evaluations was not significant in Model 2, indicating its unstable influence. Compared with other evaluation channels, a region with higher app ratings will receive more evaluations by citizens in the future. Thus, the results support H2.

To test the relationship between digital interfaces and citizen satisfaction, we ran Model 3 and Model 4. Model 3 illustrates that the proportion of website, app, and QR evaluations is positively associated with positive evaluations. The result indicates that a 1% increase in website evaluations generates a 0.38 increase in the proportion of positive evaluations; at the same time, a 1% enhancement in app evaluations causes a 0.97 increase in the proportion of positive evaluations in the next month. In addition, a 1% increase in QR code evaluations generates a 0.14 increase in the proportion of positive evaluations. Besides, there was no significant relationship in Model 4. To summarize, compared with other evaluation channels, a region with higher online ratings will receive more positive evaluations in the future, suggesting that H3 is supported.

6 Discussions and conclusion

6.1 Theoretical contributions

Digitalization of evaluations by citizens has rarely been studied, even though it is becoming a worldwide fashionable practice. We need more empirical evidence about whether digital evaluation interfaces will facilitate evaluations by citizens and improve citizen satisfaction. Based on empirical data from the GSES in China, we show that the digitalization reform has enhanced evaluations by citizens to a large extent. In addition, digital interfaces facilitate evaluations by citizens and satisfaction compared to traditional offline channels.

The theoretical implications of our findings are fourfold. First, this research provides empirical evidence to address the research gap between the digitalization of public services and citizens' participation behavior. Existing studies have empirically identified the negative relationship between digitalization and administrative burden (Arendsen et al., 2014), as well as the negative relationship between administrative burden and citizen–state interaction (Bozeman & Youtie, 2020; Herd et al., 2013), leaving a research gap regarding whether digitalization will facilitate citizens' willingness to interact with government. Although sporadic experiments suggest that the relationship is insignificant (Prokop & Tepe, 2021), objective data drawn from vivid policy practice are urgently needed. Empirical results from the GSES in China show a positive relationship between the two core variables.

Second, we combined the public encounter theory with the notion of administrative burden to formalize a theoretical framework to explain the influence of form-related factors on public service provision. The framework could help future studies to explore the relevance of administrative burden and red tape to public encounters.

Third, the results provide empirical support for reducing administrative burden, which helps facilitate citizen–government interaction and policy implementation. The negative effect of administrative burden is a well-developed research agenda, but few studies focus on how to reduce it. This study provides a lens of digitalization of public service provision, which could open up a new starting point for subsequent research.

Fourth, to the best of our knowledge, this research provides the first empirical evidence based on objective data, revealing a significant positive relationship between digitalization and citizens' behavior of making evaluations. The insignificant effect revealed by experimental research could be attributed to a lack of manipulation check. It is hard to manipulate the channels used to make evaluations in experiments. The participants without real experience were asked to imagine using an app, website, or offline channel to make ratings, which is far from reality. In comparison, the data from government archives are more convincing.

6.2 Practical implications

Our research provides insights for practitioners in government to facilitate public service digitalization, reduce administrative burden, and improve citizen satisfaction. First, the results suggest that digital interfaces will facilitate ratings by citizens. Compared with offline evaluation channels, digital interfaces could decrease the administrative burden that citizens face, motivating them to participate in public services. The government should implement reforms of evaluations by citizens and transform from traditional evaluation settings to hybrid forms with various online platforms.

Second, interfaces displayed on apps will facilitate citizens' behavior of making evaluations and boost citizen satisfaction, which shows the priority of the app channel among all online platforms. Empirical findings indicate that compared with other online interfaces, such as QR codes, government websites, and SMS, evaluations

from apps on mobile phones are more convenient and burden-free, making it the most welcoming channel. In addition, government apps are also the most desirable among the five channels. The result suggests that governments could pay more attention to apps when implementing digital transformation of government evaluation systems. Due to limited funding, governments must choose the top reform priorities, and the app channel has the highest potential to boost evaluations by citizens and improve citizens' satisfaction with citizen–state interactions.

Third, the GSES also has implications for developing countries that want to implement administrative reforms. Instead of directly calling for reducing administrative burden or red tape, the digital overhaul and updating of evaluations by citizens could smoothly facilitate government performance improvement. On the one hand, it helps reduce citizens' administrative burden and cultivates the government's mindset of providing public services through digital interfaces. On the other hand, citizens' full use of evaluation data could help governments accurately diagnose problems and difficulties in public service provision. Timely feedback through short-term evaluations can also help governments immediately return to the right track.

6.3 Limitations and future research directions

As an exploratory study, this study is limited in three aspects, and we call for future studies to dig further into the digital transformation of public service ratings. First, the data were drawn on a regional rather than a micro-level, which cannot directly provide nuanced insights into citizens' willingness to make evaluations. We hope that future studies could explore the specific evaluation data of each citizen to examine other interesting questions. Second, we only had access to data from one city, and nationwide data can be used in future research to expand the external validity of our conclusions.

We look forward to future studies that directly and accurately measure changes in administrative burdens brought about by digitalization to clarify the causal relationship between electronic evaluation and administrative burdens. As a multiple-dimension concept, it is difficult to quantify and measure administrative burden, which made it impossible to directly include it in our research. Third, the relationships among variables were not causally inferred, and we hope that future research could use experimental designs to explore the causality.

Acknowledgements The authors would like to thank the National Natural Sciences Foundation of China (Grant/Award Number: 72274203), the Outstanding Innovative Talents Cultivation Funded Programs 2021 of Renmin University of China, and Beijing Social Science Fund (No. 21DTR058) for financial support. The correspondence for this article should go to Liang Ma.

Data Availability Data are not available due to legal restrictions.

Declaration

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

References

- Arendsen, R., Peters, O., ter Hedde, M., & van Dijk, J. (2014). Does e-government reduce the administrative burden of businesses? An assessment of business-to-government systems usage in the Netherlands. *Government Information Quarterly*, *31*(1), 160–169. <https://doi.org/10.1016/j.giq.2013.09.002>
- Baltagi, B. H., Bresson, G., & Pirotte, A. (2003). Fixed effects, random effects or Hausman–Taylor?: a pretest estimator. *Economics Letters*, *79*(3), 361–369. [https://doi.org/10.1016/S0165-1765\(03\)00007-7](https://doi.org/10.1016/S0165-1765(03)00007-7)
- Barnes, C. Y., & Henly, J. R. (2018). They are underpaid and understaffed: how clients interpret encounters with street-level bureaucrats. *Journal of Public Administration Research and Theory*, *28*(2), 165–181. <https://doi.org/10.1093/jopart/muy008>
- Bartels, K. P. R. (2013). Public encounters: the history and future of face-to-face contact between public professionals and citizens. *Public Administration*, *91*(2), 469–483. <https://doi.org/10.1111/j.1467-9299.2012.02101.x>
- Behn, R. D. (2008). Designing performancestat: or what are the key strategic choices that a jurisdiction or agency must make when adapting the compstat/citistat class of performance strategies? *Public Performance & Management Review*, *32*(2), 206–235. <https://doi.org/10.2753/PMR1530-9576320202>
- Bell, E., Ter-Mkrtyan, A., Wehde, W., & Smith, K. (2021). Just or unjust? how ideological beliefs shape street-level bureaucrats' perceptions of administrative burden. *Public Administration Review*, *81*(4), 610–624. <https://doi.org/10.1111/puar.13311>
- Bozeman, B., & Youtie, J. (2020). Robotic bureaucracy: administrative burden and red tape in university research. *Public Administration Review*, *80*(1), 157–162. <https://doi.org/10.1111/puar.13105>
- Brown, J. T., Carey, G., & Malbon, E. (2021). What is in a form? Examining the complexity of application forms and administrative burden. *Australian Journal of Public Administration*, *80*(4), 933–964. <https://doi.org/10.1111/1467-8500.12531>
- Carey, G., Dickinson, H., Malbon, E., Weier, M., & Duff, G. (2020). Burdensome administration and its risks: competing logics in policy implementation. *Administration & Society*, *52*(9), 1362–1381. <https://doi.org/10.1177/0095399720908666>
- Escobar, M. P., & Demeritt, D. (2016). Paperwork and the decoupling of audit and animal welfare: The challenges of materiality for better regulation. *Environment and Planning c: Politics and Space*, *35*(1), 169–190. <https://doi.org/10.1177/0263774X16646771>
- Fan, Z., Christensen, T., & Ma, L. (2022). Policy attention and the adoption of public sector innovation. *Public Management Review*. <https://doi.org/10.1080/14719037.2022.2050283>
- Liao, F. (2020). Singular or plural? Administrative burden and doing business in China. *Journal of Chinese Governance*. <https://doi.org/10.1080/23812346.2020.1859793>
- Gao, J. (2012). How does chinese local government respond to citizen satisfaction surveys? a case study of Foshan City. *Australian Journal of Public Administration*, *71*(2), 136–147.
- Goodsell, C. T. (1981). *The public encounter and its study: Where state and citizen meet*. USA: Indiana University Press.
- Heinrich, C. J. (2018). Presidential address: “a thousand petty fortresses”: administrative burden in U.S. immigration policies and its consequences. *Journal of Policy Analysis and Management*, *37*(2), 211–239. <https://doi.org/10.1002/pam.22046>
- Herd, P., DeLeire, T., Harvey, H., & Moynihan, D. P. (2013). Shifting administrative burden to the state: the case of medicaid take-up. *Public Administration Review*, *73*(s1), S69–S81. <https://doi.org/10.1111/puar.12114>
- Keiser, L. R., & Miller, S. M. (2020). Does Administrative burden influence public support for government programs? evidence from a survey experiment. *Public Administration Review*, *80*(1), 137–150. <https://doi.org/10.1111/puar.13133>
- Lindgren, I., Madsen, C. Ø., Hofmann, S., & Melin, U. (2019). Close encounters of the digital kind: a research agenda for the digitalization of public services. *Government Information Quarterly*, *36*(3), 427–436. <https://doi.org/10.1016/j.giq.2019.03.002>
- Lipsky, M. (1983). *Street-Level Bureaucracy: The Dilemmas of the Individual in Public Service*. NY: Russell Sage Foundation.
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service*. NY: Russell Sage Foundation.

- Ma, L. (2017). Performance management and citizen satisfaction with the government: evidence from Chinese municipalities. *Public Administration*, 95(1), 39–59. <https://doi.org/10.1111/padm.12275>
- Ma, L., & Wu, X. (2020). Citizen engagement and co-production of e-government services in China. *Journal of Chinese Governance*, 5(1), 68–89. <https://doi.org/10.1080/23812346.2019.1705052>
- Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*, 25(1), 43–69. <https://doi.org/10.1093/jopart/muu009>
- Peeters, R., Trujillo Jiménez, H., O'Connor, E., Ogarrio Rojas, P., González Galindo, M., & Morales Tenorio, D. (2018). Low-trust bureaucracy: understanding the Mexican bureaucratic experience. *Public Administration and Development*, 38(2), 65–74. <https://doi.org/10.1002/pad.1824>
- Prokop, C., & Tepe, M. (2021). Talk or type? The effect of digital interfaces on citizens' satisfaction with standardized public services. *Public Administration*, *n/a*(*n/a*). doi:<https://doi.org/10.1111/padm.12739>
- So, B. W. Y. (2014). Civic engagement in the performance evaluation of the public sector in China: Building horizontal accountability to enhance vertical accountability. *Public Management Review*, 16(3), 341–357. <https://doi.org/10.1080/14719037.2013.770055>
- Swindell, D., & Kelly, J. M. (2000). Linking Citizen Satisfaction Data to Performance Measures: A Preliminary Evaluation. *Public Performance & Management Review*, 24(1), 30–52. Retrieved from <http://www.jstor.org/stable/3381075>
- Tummers, L., Weske, U., Bouwman, R., & Grimmelikhuijsen, S. (2016). The impact of red tape on citizen satisfaction: an experimental study. *International Public Management Journal*, 19(3), 320–341. <https://doi.org/10.1080/10967494.2015.1027800>
- Wandaogo, A. A. (2022). Does digitalization improve government effectiveness? Evidence from developing and developed countries. *Applied Economics*, 54(33), 3840–3860. <https://doi.org/10.1080/00036846.2021.2016590>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.