Received: 12 October 2021 | Revised: 23 April 2022 | Accepted: 26 April 2022

DOI: 10.1111/padm.12860

REPLICATION

public administration WILEY

A replication of "an experimental test of the expectancy-disconfirmation theory of citizen satisfaction"

Wenna Chen[1] | Binzizi Dong[2] | Chih-Wei Hsieh[2] | M. Jin Lee[3] | Ning Liu[2] | Richard M. Walker[2,4] | Yao Wang[2] | Bo Wen[2] | Wen Wen[2,5] | Peiyi Wu[6] | Xia Wu[2,7] | Jiasheng Zhang[2]

[1]Center for Chinese Public Administration Research and School of Government, Sun Yat-sen University, Guangzhou, China

[2]Laboratory for Public Management and Policy, Department of Public Policy, City University of Hong Kong, Hong Kong, China

[3]Department of Public Administration, Faculty of Business and Economics, University of Malaya, Kuala Lumpur, Malaysia

[4]Shenzhen Research Institute, City University of Hong Kong, Shenzhen, China

[5]School of Economics and Management, China University of Petroleum, Qingdao, China

[6]School of Public Administration, Beihang University, Beijing, China

[7]Department of Public Administration, Renmin University of China, Beijing, China

Correspondence
Peiyi Wu, School of Public Administration, Beihang University, Beijing 100191, China.
Email: wupeiyi@buaa.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 71974164; the Central Policy Unit (now the Policy Innovation and Coordination Office) of the Government of the Hong Kong Special Administrative Region, Grant/Award Number: 2015.A1.031.16A; the Hong Kong Scholars Program, Grant/Award Number: 2020-215

Abstract

To better understand citizen satisfaction with public services, public administration research has adopted the expectancy-disconfirmation model in recent years. This model proposes that satisfaction is a function of perceived performance and expectations. Recent quantitative and experimental studies of the expectancy-disconfirmation model have supported the framework. However, few replications have been conducted and none outside western contexts. We conducted two narrow, robust experimental replications of Van Ryzin (2013, *Journal of Policy Analysis and Management*, 32(3), pp. 597–614) in the Chinese cities of Hong Kong (in 2017) and Shenzhen (in 2021). We found support for the findings reported in Van Ryzin (2013) and

concluded that the expectancy-disconfirmation model holds promise in a variety of settings as a framework for measuring citizen satisfaction with public services.

1 | INTRODUCTION

The expectancy-disconfirmation model has become the dominant model in studies of citizen satisfaction in the field of public administration (Van Ryzin, 2004, 2006, 2013). The expectancy-disconfirmation model holds promise for public administration research because it delves into citizens' thoughts, offering psychological insights into their behavior (Zhang et al., 2021). The central tenet of the expectancy-disconfirmation model is that citizens' expectations of public services are as important as their perceived performance of these services. Together, expectations and perceived performance result in citizen (dis)satisfaction. Building on research in the field of marketing (Oliver, 1980), Van Ryzin introduced this model to public administration, initially empirically testing the framework using survey data (Van Ryzin, 2004, 2006) and then experimentally manipulating expectations and performance (Van Ryzin, 2013).

A meta-analysis of the expectancy-disconfirmation model (Zhang et al., 2021) found broad support for the validity of the framework. Of the 17 studies reviewed by Zhang et al. (2021), six used experimental methods to manipulate expectations and satisfaction (Andersen & Hjortskov, 2016; Filtenborg et al., 2017; Grimmelikhuijsen & Porumbescu, 2017; Noda, 2019; Thomassen et al., 2017; Van Ryzin, 2013), and two explicitly sought to replicate and extend Van Ryzin's work. All of these studies, including the replications, were conducted in Western settings (Denmark, The Netherlands, and the USA). No replication study has been conducted in non-western contexts.[1] This article contributes to knowledge on the expectancy-disconfirmation model by conducting a narrow replication of Van Ryzin's (2013) original study and extending it to two Chinese contexts: Hong Kong and Shenzhen. Such replication efforts help to test the boundary conditions of existing knowledge, enhance external validity, and establish the generalizability of existing theories (Walker et al., 2019). In this case, the existing theory is the expectancy-disconfirmation model of citizen satisfaction. We identify China as an ideal critical case owing to her distinctive tradition of collectivism (Hofstede, 2001), a cultural orientation that may mitigate citizen reactions to public services which aim to fulfill the public interest more than individual needs (Frank et al., 2015).

This article is structured as follows. First, we provide a synopsis of the original study. Second, following the best practice recommendations of Walker et al. (2019) for replicating experimental research in public administration, we discuss the key replication processes of our two replication studies. We then report on the participants, experimental procedures, variable measures, and analysis. This is followed by a summary of the research findings. We then discuss the findings of the original study and our replications, and conclude that the expectancy-disconfirmation model holds promise in a variety of settings as a framework to measure citizen satisfaction with public services.

2 | SYNOPSIS OF THE ORIGINAL STUDY

Before Van Ryzin (2004) introduced the expectancy-disconfirmation model to public administration, research on citizen satisfaction focused on the individual, jurisdictional, and city-specific determinants of citizen ratings of service quality (DeHoog et al., 1990; Hero & Durand, 1985) and how the actual quality and performance of government services and institutions affected citizen satisfaction (Charbonneau & Van Ryzin, 2012; Favero & Meier, 2013; Van Ryzin & Immerwahr, 2007). Drawing on research in the field of marketing (e.g., Oliver, 1980), Van Ryzin (2004, 2006) introduced the expectancy-disconfirmation model to public administration research and empirically validated the model using survey data.

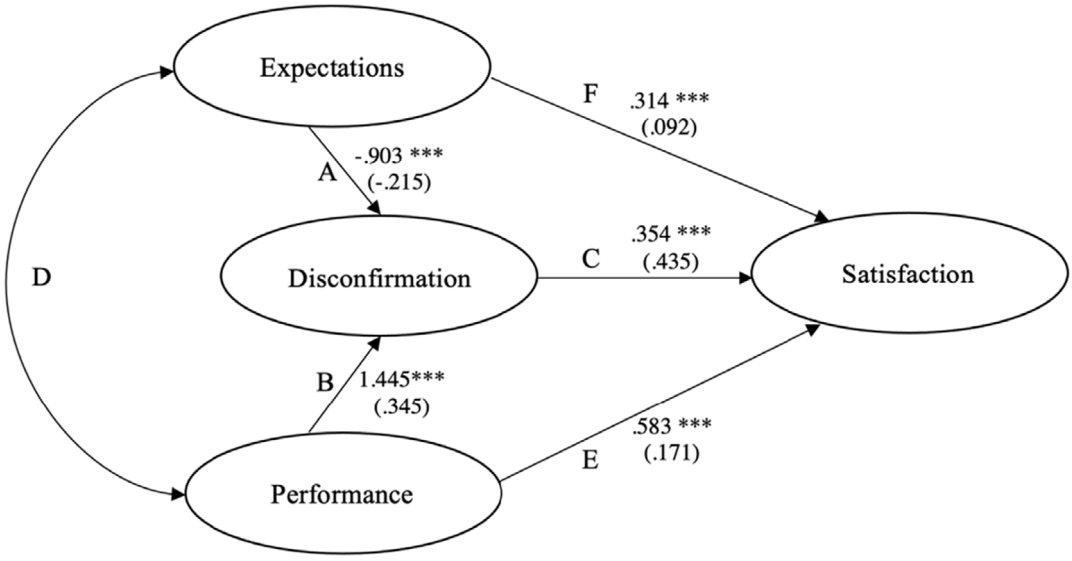

FIGURE 1 Expectancy disconfirmation with performance model. Adapted from Van Ryzin (2004). Coefficients are from Van Ryzin (2013): unstandardized regression coefficients shown, with standardized coefficients in parentheses. Model $R^2 = 0.26$; sample size $n = 905$; $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 1 depicts the expectancy-disconfirmation model and shows the key hypothesized relationships. The core relationships are Links A, B, and C. Expectations and perceived performance (Links A and B, respectively) jointly affect disconfirmation, which leads to satisfaction (Link C). Satisfaction is achieved if perceived performance meets or exceeds expectations, whilst dissatisfaction occurs if perceived performance falls short of expectations. Expectations and perceived performance are correlated (Link D), although the direction of the relationship is not specified. There are direct relationships between perceived performance and satisfaction (Link E) and between expectations and satisfaction (Link F). Expectations and performance are conceptualized as exogenous variables in the model because they are assumed to exist prior to disconfirmation (Van Ryzin, 2004, 2006, 2013).

In his 2013 study, Van Ryzin reviewed the previous studies on expectancy-disconfirmation theory in the public management literature and pointed out that these studies produce conflicting findings with respect to expectation, and most were cross-sectional raising causality concerns. Against this background, Van Ryzin suggested that it is necessary to use experimental research designs to independently manipulate expectations and performance to establish the casual structure of these complex relationships in the expectancy-disconfirmation model. Therefore, in the study, he used experiments to estimate the direction and magnitude of the presumed causal relationships in the model. Specifically, the participants first received either a high- or low-expectations statement from a hypothetical public official in a fictitious American municipality. Then, participants were exposed to either a high- or low-performance picture of street cleanliness. After being presented with the photograph, the participants were asked to rate the cleanliness of the streets and their satisfaction with the government's performance. The participants' demographic information, including age, gender, political ideology, geographical location, and home city size, was also collected. The experiment was conducted in 2010 with a nationally representative sample of 964 US citizens. Van Ryzin's experimental interventions successfully manipulated expectations and performance, and his findings are included in Figure 1. His study experimentally demonstrated that the central thesis behind the expectation-disconfirmation model could be manipulated, offering additional evidence of the model's applicability.

The meta-analysis of citizen satisfaction with public services conducted by Zhang et al. (2021) identified 17 studies in the public administration literature that used key concepts from the expectancy-disconfirmation model. Fifteen of these studies were conducted in Western settings such as Europe and the US, with one study conducted in

South Korea and one in Japan. Noda (2019) studied citizen satisfaction at different levels of government and Grimmelikhuijsen and Porumbescu (2017) focused on trust in government incorporating the expectancy-disconfirmation model. These studies assumed that the theory would have validity in the Japanese and South Korean contexts and did not identify Japan or South Korea as replication sites. Of the 17 studies, six were published using experimental methods to manipulate expectations and satisfaction (Andersen & Hjortskov, 2016; Filtenborg et al., 2017; Grimmelikhuijsen & Porumbescu, 2017; Noda, 2019; Thomassen et al., 2017), of which two explicitly sought to replicate and extend Van Ryzin's work. For example, Filtenborg et al. (2017) conducted a close replication in Denmark verifying the findings of the expectancy-disconfirmation model in a different context. More recently, Favero and Kim (2021) extended research on the expectancy-disconfirmation model by experimentally manipulating normative and predictive expectations finding that normative expectations were negatively associated with satisfaction while predictive expectations had a very weak statistical relationship with satisfaction.

3 | REPLICATION METHOD

The replication method drew on the framework proposed by Walker et al. (2019). Below, we discuss the feasibility of replicating the original study and the boundary conditions arising from selecting China as a critical test case. The narrow replications we conduct are empirical generalizations; measurement and analysis from the original study is kept as far as is practicable and tested on new populations—Hong Kong and Shenzhen (Tsang & Kwan, 1999).

The original study could be successfully replicated because full materials were available and internal threats to validity were judged to be low. The original study was guided by a well-specified theoretical perspective, with measures and treatments well fitted to the theoretical concepts. Operationally, the participants were drawn from a nationally representative sample of US citizens, and vignettes were randomly assigned across all participants. However, the original study did not report statistical power; hence, power analysis was conducted to determine the number of participants (see below).

Hong Kong and Shenzhen are ideal critical cases to further test the boundaries of the expectation-disconfirmation model because replication has not been attempted in China, a region associated with collectivist values including consensus, harmony, unity, and community. Even though studies have documented that the dimension of individualism–collectivism somewhat influences service satisfaction (e.g., Lee et al., 2009; Sharma et al., 2012; Tsoukatos & Rand, 2007), the relative consistency in the findings of empirical tests of the expectancy-disconfirmation model suggested that the findings would be replicated in our study (see below for further discussion).

The Hong Kong and Shenzhen contexts are very different from the original study. The Hong Kong study was conducted in 2017, which was before the social disruption of 2019 and the greater integration with mainland China after the enactment of the National Security Law in the summer of 2020. In 2017, Hong Kong was governed by the "one country, two systems" principle, under which flexibly in policy development and implementation was permitted, and the government reported directly to Beijing. The Hong Kong government was responsible, as a unitary authority, for the provision of all public services, many of which were contracted to the nonprofit and private sectors (Scott, 2010). At the time of the study, the Hong Kong government was relatively stable and effective in a managerial sense, with a small government attitude and a technical approach to governance and administration. Shenzhen is located in Guangdong province and abuts Hong Kong. While the two cities are co-located, Shenzhen is in an area operating under an authoritarian form of governance (Zhang, 2012). As the first "special economic zone" in China, Shenzhen enjoys flexible economic policies and government measures, particularly in comparison with some other Chinese cities. Therefore, it is at the forefront of service innovation, information disclosure, and performance evaluation. As two major cities in South China, Hong Kong and Shenzhen share similar Asian characteristics in many ways but have very different institutional arrangements: Hong Kong is typified as a city where East meets West, and Shenzhen, an area operating under an authoritarian form of governance, is at the forefront of China's socio-economic transformation. Importantly, both cities are very different compared with the fictitious American city in the original study.

Although the boundary conditions in the two cities of our replications differ, and were different than those in the original study, we expected the findings to hold because previous studies that used the expectancy-disconfirmation model in international settings (including in business settings in mainland China and Hong Kong) mostly confirmed the original findings (Zhang et al., 2021). This suggests that differences in the context-delineating variables were unlikely to result in uncertain boundary conditions (Busse et al., 2017). We propose this because the underpinnings of the expectancy-disconfirmation model focus on citizens' psychological processes, and the replications, like the original study, focused on street cleanliness, a frequently experienced and technical public service. Therefore, theoretical expectations were likely to hold in the Hong Kong and Shenzhen contexts. We thus anticipated that the accuracy of the theoretical predictions of the expectancy-disconfirmation framework would remain high. In short, the context-delineating variables were not likely to substantially influence the operations of the model, which led us to expect the results to validate our theoretical prediction. Because uncertainty about boundary conditions was low and the theoretical accuracy of our prediction was high, we adopted following Busse et al. (2017) an "inside-out approach" to the question of boundary conditions.

The expectancy-disconfirmation model has been widely used in public administration research and has a long history in the management and marketing literature, including translation into languages other than English. Therefore, it has strong face validity. In keeping with a narrow replication of the original study, we also examined street cleanliness in the Hong Kong and Shenzhen studies.

4 | EXPERIMENTAL DESIGN

4.1 | Participants and sample

The Hong Kong participants were members of My Citizen Panel, which was drawn from university alumni and responses by citizens to Google ads to participate in a research study. The survey was delivered to the participants using Qualtrics, and the participants could respond in Cantonese or English. In Shenzhen, the participants were found using the sample database provided by the online survey platform WJX.CN. WJX.CN has more than 2.6 million members, varying in age, income, employment, and city of residence. Nearly 15% (14.81%) of the respondents are located in Guangdong Province, and 75.41% of the respondents' age below 30 years old. The survey instrument was developed in English. The translation of the survey into Cantonese (in Hong Kong) and Mandarin (in Shenzhen) included back-translation to English with checks for consistency. Ethical approvals were received from the City University of Hong Kong's Human Subjects Ethics Sub-committee (H000904 and H001986).

Power analysis for the Hong Kong study suggested a sample size of 174 with an alpha error of 0.05 with four treatment groups, resulting in a proposed sample of 696 (Walker et al., 2017). This experiment was part of a larger longitudinal study, and the sample size was therefore increased to 1000 due to concerns about attrition throughout the rounds of the survey. For the Shenzhen study, a more stringent a priori power analysis was conducted to determine the sample size required for our project to detect an effect. Van Ryzin's (2013) study reported path model results based on 905 observations. We conducted a power analysis for the path model (structural equation modeling) based on the specification of an alpha error (0.05), the desired power (0.8), root mean square error of approximation (lower than 0.05), goodness of fit index (GFI, higher than 0.95), and adjusted goodness of fit index (AGFI, higher than 0.9). The result showed that 1238 participants were needed to achieve 80% power.

To ensure their familiarity with the public services in each city, the participants were screened to include only those who had lived or worked in the city for more than 1 year. In Hong Kong, data were collected between February and March 2017, and 1006 participants completed the survey. Data in Shenzhen were collected in June 2021, and 1302 participants completed the survey. After cleaning the data and eliminating incomplete responses, the final sample for analysis contained 992 valid responses from Hong Kong and 1228 from Shenzhen.

The samples were not representative of the populations of Hong Kong and Shenzhen, and there were sampling differences between the two cities (see Table S1). The participants from Hong Kong and Shenzhen were typically younger and more educated than the city population, and in Hong Kong more likely to be working in the public and NGO/non-profit sectors. Between the two cities, there were variations in the sample of subjects that were statistically different in relation to age (Hong Kong participants were younger than those in Shenzhen), length of time living in the city (Hong Kong participants had lived in the city longer), employment sector (more Hong Kong participants worked in government and the NGO/non-profit sectors in comparison to Shenzhen), and education levels (Hong Kong participants were more likely to hold undergraduate and postgraduate degrees) (see Table S1 for Pearson's χ^2). These differences are a potential limitation to the study because participants were not representative of their city and the two samples are not directly comparable.

4.2 | Experimental procedure

The Hong Kong and Shenzhen replications closely resembled the original experiment design, procedure, and measurement. Each replication was designed as a 2 × 2 factorial experiment with two expectation manipulations (high and low) and two performance manipulations (high and low). Figure 2 shows the experiment flow and the random allocation of the participants to the arms of the study. The Hong Kong replication was not preregistered, whilst the Shenzhen study was preregistered at AsPredicted.org prior to data collection.

The context of Van Ryzin's original study was a mid-sized American city, "Hometown." The introductory text of the vignette and the expectation message were delivered by the city's administrator, and referred to district TC in Hong Kong and A in Shenzhen. To ensure that the vignettes had face validity in this study's two cities, the message was delivered by a district councilor in Hong Kong and by a director of the finance bureau in Shenzhen.

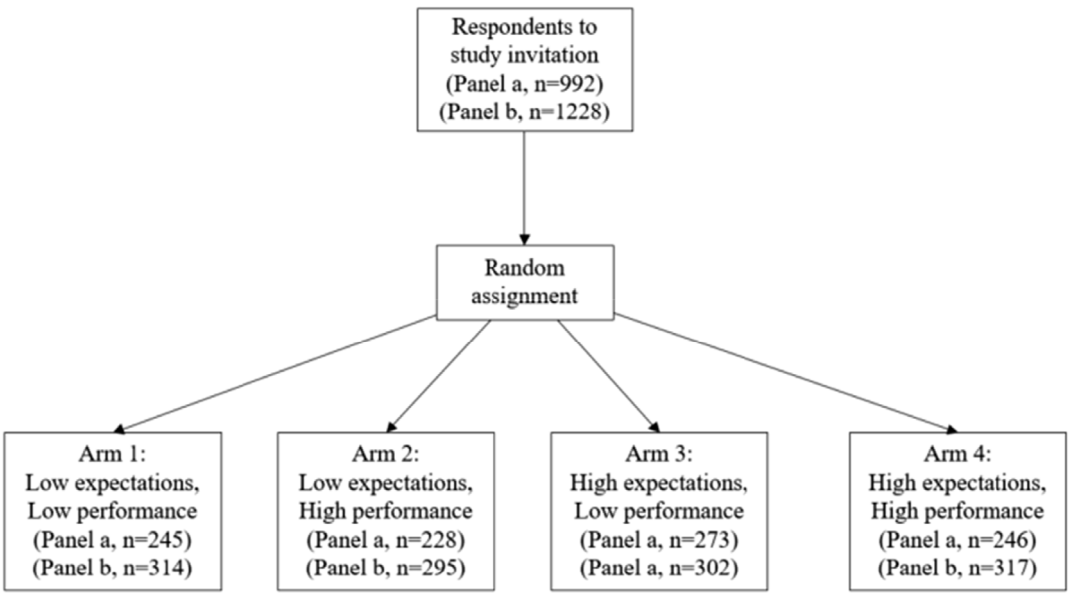

FIGURE 2 Experiment design. Panel a = Hong Kong, Panel b = Shenzhen. Note that based on what you have read about TC/A, how would you rate your expectations for the district council/government's performance? (1 = very low expectations to 7 = very high expectations). Test of difference in means for Hong Kong: $t = -10.59$, $df = 990$, $p < 0.001$; test of difference in means for Shenzhen: $t = -4.54$, $df = 1226$, $p < 0.001$

The expectation text in Hong Kong (study conducted in 2017) used the example of the city's "Small Government" rhetoric to contextualize the message, and the context of COVID-19 was used to contextualize the message in Shenzhen (study conducted in 2021).[2] The rest of the text in the vignette remained as close as possible to the original text. After the introduction of the survey, the participants were exposed to the first randomized factor, which involved either a high- or low-expectations statement from a hypothetical public official (see Table S2). In both the Hong Kong and Shenzhen replications, the low-expectation statements informed the participants that the city was experiencing an economic recession and budget cuts and that they should expect a decline in the quality of public services. The high-expectation statements informed the participants of a similar economic situation, but they were assured that all city services would be maintained to their usual high standards. After the presentation of the expectation statement, the participants were asked to rate their expectations regarding the city government's performance as follows: "Based on what you have read about TC/A, how would you rate your expectations for the district council/government's performance? (1 = very low expectations to 7 = very high expectations)." (following Van Ryzin, 2013, p. 604).

Next, the participants were exposed to the second randomized factor: either a high- or low-performance picture of street cleanliness (see Table S3). The pictures used to represent high performance in street cleanliness were taken from normal, clean city streets in the two cities. The images were edited to include litter for the low-performance vignette. After being presented with the photographs, the participants were asked to rate the cleanliness of the street as follows: "How would you rate the cleanliness of city streets in TC/A? (1 = poor to 7 = excellent)." (Van Ryzin, 2013, p. 604).

The participants were then asked the following question about their satisfaction with the government: "Based on the cleanliness of TC/A, how satisfied would you be with the district council/government's performance? (1 = very dissatisfied to 7 = very satisfied)" (following Van Ryzin, 2013, p. 605). This question served as the dependent variable in the test of the full expectancy-disconfirmation model. The participants' demographic information was also collected. We included an attention check question in the replications that was not included in the original study.

4.3 | Measurement

We estimated the full expectancy-disconfirmation model using the same variables as in the original study by Van Ryzin (2013). These variables included the following.

- Expectations (1 = high-expectation statement, 0 = low-expectation statement).
- Performance (1 = high-performance photograph, 0 = low-performance photograph).
- Disconfirmation (we subtracted the self-reported expectation score from the self-reported performance score to calculate the level of disconfirmation on a scale from −6 to 6).
- Satisfaction (1 = very dissatisfied to 7 = very satisfied).

4.4 | Analysis

Adopting a narrow empirical generalization replication of the Van Ryzin (2013) study, we used the same analytical techniques as the original study to allow for an easy comparison of the findings. Specifically, we used regression-based path analysis to test the model.

5 | RESULTS

We first present the descriptive results of the manipulations, then the results of the full model. The mean scores for the participants randomly assigned to the high-expectation statement were higher than the mean for those randomly

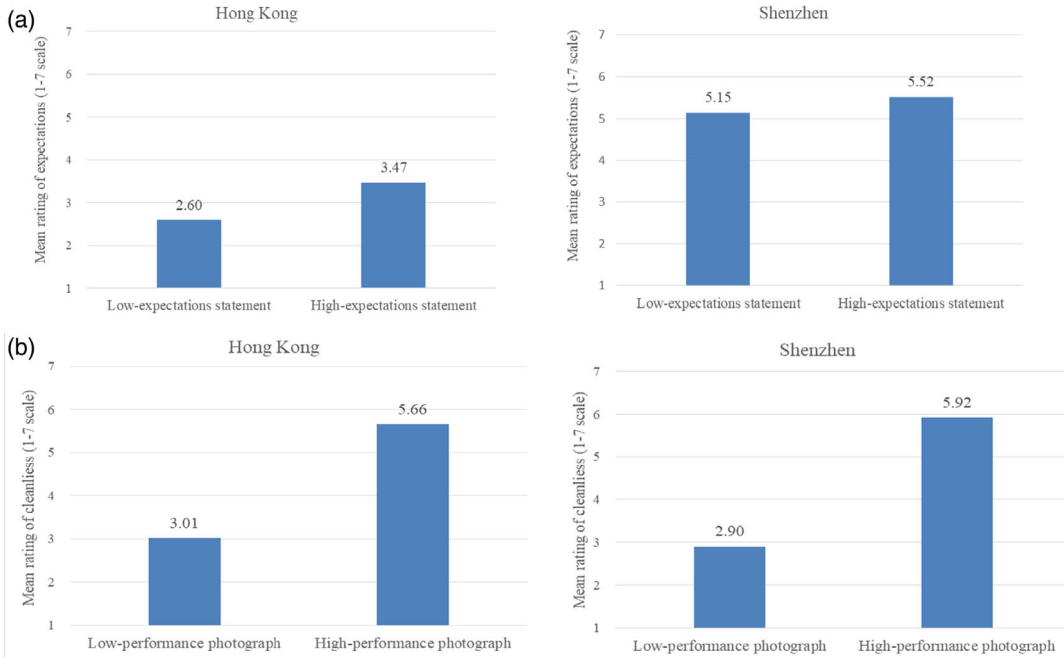

FIGURE 3 (a) Mean expectations ratings by group. How would rate the cleanliness of city streets in TC/A? (1 = poor to 7 = excellent). Test of difference in means for Hong Kong: $t = -36.03$, $df = 990$, $p < 0.001$; Test of difference in means for Shenzhen: $t = -32.37$, $df = 1226$, $p < 0.001$. (b) Mean performance ratings by group. Unstandardized regression coefficients are shown, with standardized coefficients in parentheses. Hong Kong Model $R^2 = 0.45$; sample size $n = 992$. Shenzhen Model $R^2 = 0.51$; sample size $n = 1228$. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

assigned to the low-expectation statement, showing that the manipulation of expectations worked (Figure 3a). It is interesting to note that the mean scores in Shenzhen were substantially higher than in Hong Kong and also similar for the low (5.15) and high-expectation (5.52) statements. We return to these differences between Hong Kong and Shenzhen in the discussion section. Similarly, the performance manipulation also had its expected effect: the participants assigned to the low-performance photograph rated the streets as much less clean than did those assigned to the high-performance photograph (Figure 3b).

Following Van Ryzin (2013), we conducted regression-based path analysis to outline the treatment effects on disconfirmation and satisfaction (Figure 4). The empirical model explained 45% and 51% of the variance in satisfaction for the Hong Kong and Shenzhen experiments, respectively. Across the two cities, all of the core relationships were statistically significant and in the same direction as predicted in the expectation-disconfirmation model and in Van Ryzin's (2013) original study (the relationship between expectations and satisfaction for the Shenzhen participants was significant at the lower threshold of $p < 0.1$). High expectations led to nearly a 1 point (−0.968) decline in disconfirmation for the Hong Kong participants and a 0.371 point decline in Shenzhen. The effect of performance on disconfirmation was positive and relatively large, representing 2.507 points for Hong Kong and 3.051 for Shenzhen. For the relationship between disconfirmation and satisfaction, the results showed that each point increase in disconfirmation led to a 0.274 point increase in satisfaction for the Hong Kong participants and a 0.390 point increase in satisfaction for the Shenzhen participants.

There were two variations in the details of the findings in Hong Kong and Shenzhen compared to the original study. The direct effects of expectations on satisfaction were both positive and significant for the two cities, confirming Van Ryzin's (2013) finding, though the coefficient for Shenzhen (0.154, $p < 0.1$) was approximately half of that for Hong Kong (0.33). In addition, in Shenzhen, the positive direct effect of expectation on satisfaction (0.154) was nearly reduced by the negative indirect effect of expectations on satisfaction through disconfirmation

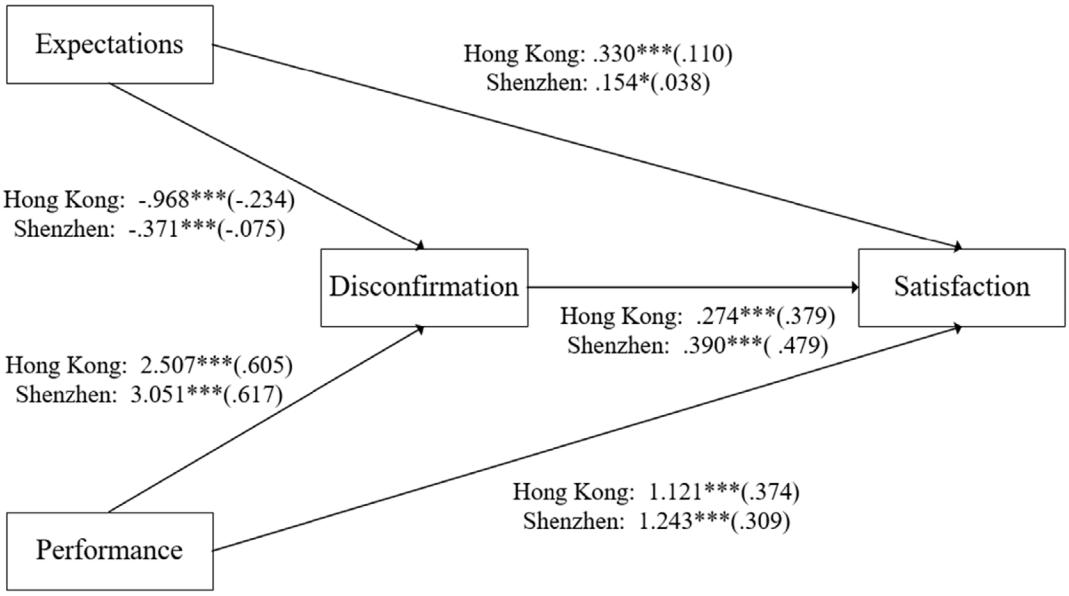

FIGURE 4 Regression-based path analysis. Based on the cleanliness of TC/A, how satisfied would you be with the district council/government's performance? (1 = very dissatisfied to 7 = very satisfied). For both Hong Kong and Shenzhen, the performance main effect is significant ($p < 0.01$) but the expectations main effect and the interaction effect are not (see Table 1 for the statistics)

($-0.371 \times 0.390 = -0.14$). Similarly, the standardized total effect of expectations on satisfaction for Shenzhen sample was also canceled out. This finding is very close to Van Ryzin's original finding. However, in Hong Kong, the positive direct effect of expectations on satisfaction (0.33) could not be completely offset by the negative indirect effect of expectations ($-0.968 \times 0.274 = -0.265$), and the standardized effect in Hong Kong was also positive and different from zero.

Consistent with Van Ryzin's (2013) finding, the direct effect of performance on satisfaction was positive and significant for both experiments. There was over a 1 point increase (Hong Kong: 1.121, Shenzhen: 1.243) in satisfaction when the participants were exposed to the high-performance photograph. When combining the direct effect with the indirect effect of performance on satisfaction through disconfirmation, the total effect was 1.808 for Hong Kong and 2.433 for Shenzhen.

We performed regression analysis to replicate the final stage of Van Ryzin's (2013) analysis.[3] Similar to the path analysis, we also found a comparable result in the regression model. Table 1 shows that performance had a positive and significant effect on satisfaction, and that expectations only had a positive but non-significant main effect on satisfaction. Although not statistically significant, the interaction effect was negative in Hong Kong, whereas such effect was positive in Shenzhen, which is consistent with Van Ryzin's (2013) finding. These findings mean that high expectations reduced the effect of performance on satisfaction for Hong Kong but amplified the effect of performance on satisfaction for Shenzhen. Figure 5 graphically presents these results showing that high performance and low and high expectations resulted in higher satisfaction ratings compared to the vignettes that presented participants with low performance.

6 | DISCUSSION AND CONCLUSION

Since Van Ryzin (2004), public administration scholars have used the expectancy-disconfirmation model as the predominant framework to explain citizen satisfaction (see Zhang et al., 2021 for a review), but few have tested its

TABLE 1 Regression analysis

	(1) Hong Kong	(2) Shenzhen
Factor 1 (high expectations 1; low expectations −1)	0.031 (0.038)	0.005 (0.046)
Factor 2 (high performance 1; low performance −1)	0.906** (0.038)	1.216** (0.046)
F1 × F2 interaction	−0.024 (0.038)	0.017 (0.046)
Constant	4.202** (0.038)	4.602** (0.046)
N	992	1228
R^2	0.364	0.365
Adjusted R^2	0.362	0.363
F	188.823**	234.350**

Note: unstandardized coefficients shown; standard errors in parentheses. *p < 0.05, **p < 0.01.

generalizability, and none in China. In this article, we undertook a narrow replication of Van Ryzin's (2013) study and conducted an empirical generalization by extending the original study to Hong Kong and Shenzhen to see if the findings hold and are valid in this context. Our results are consistent with those of Van Ryzin and others who have implemented experimental research designs in different contexts (Andersen & Hjortskov, 2016; Filtenborg et al., 2017; Grimmelikhuijsen & Porumbescu, 2017; Noda, 2019; Thomassen et al., 2017; Van Ryzin, 2013). Specifically, expectations and performance both had a direct positive effect on citizen satisfaction, and they also indirectly influenced satisfaction through disconfirmation. While the effect of expectations on disconfirmation was negative, the effect of performance was positive. We can therefore conclude that we replicated the study by Van Ryzin (2013) and that the consistency of the findings suggests that the expectancy-disconfirmation model holds promise in a variety of settings as a framework for measuring citizen satisfaction with public services.

However, our results revealed the subtleties of the expectancy-disconfirmation model in the Chinese context. First, our results showed that the net total effect of expectations on satisfaction was positive in Hong Kong, which differs from Van Ryzin (2013), who found that this net total effect was largely zero. Our regression analysis in Table 1 also confirmed that the main effect of expectations on satisfaction was positive, although not statistically significant. It is likely that in a Chinese context, citizens are more psychologically dependent on the government and are thus more sensitive to expectation cues, leading to a more salient effect of expectations on satisfaction than in other contexts. This is particularly likely to be the case in Shenzhen, as can be seen in the high rating for the low expectations statement (Figure 3). Having said this, for the sake of face validity, we modified the contextual cues in the vignette to economic recession and small government rhetoric in Hong Kong and the Covid-19 pandemic in Shenzhen. These causes of budget cuts are a potential limitation and may lead participants to have different service expectations, notably given the severe economic impact of Covid-19. In the original study in the US, however, citizens have a more negative or critical view of government (Jost et al., 2009) that cannot be easily manipulated, and the net effect of expectations is null. Our findings have potential practical implications for government officials to manage citizens' expectations through different messaging strategies to maximize their satisfaction.

Second, the regression results in Table 1 showed that the interaction effect of expectations and performance on satisfaction was positive in Shenzhen but negative in Hong Kong. That is, high expectations seemed to amplify the effect of performance on satisfaction judgments in Shenzhen but weaken the effect of performance on satisfaction in Hong Kong. Although the finding in Shenzhen is consistent with that of Van Ryzin (2013), the finding in Hong Kong does not confirm previous findings. One possible explanation concerns the difference in citizens' responses to expectation cues. In the original study in the US, the participants rated their expectations of public services at above-average levels when asked to read two statements related to expectations (low versus high). Similar situations occurred in Shenzhen. However, the ratings of expectations in the Hong Kong participants were below average, suggesting that they generally had low expectations of government services. Although Shenzhen is under an

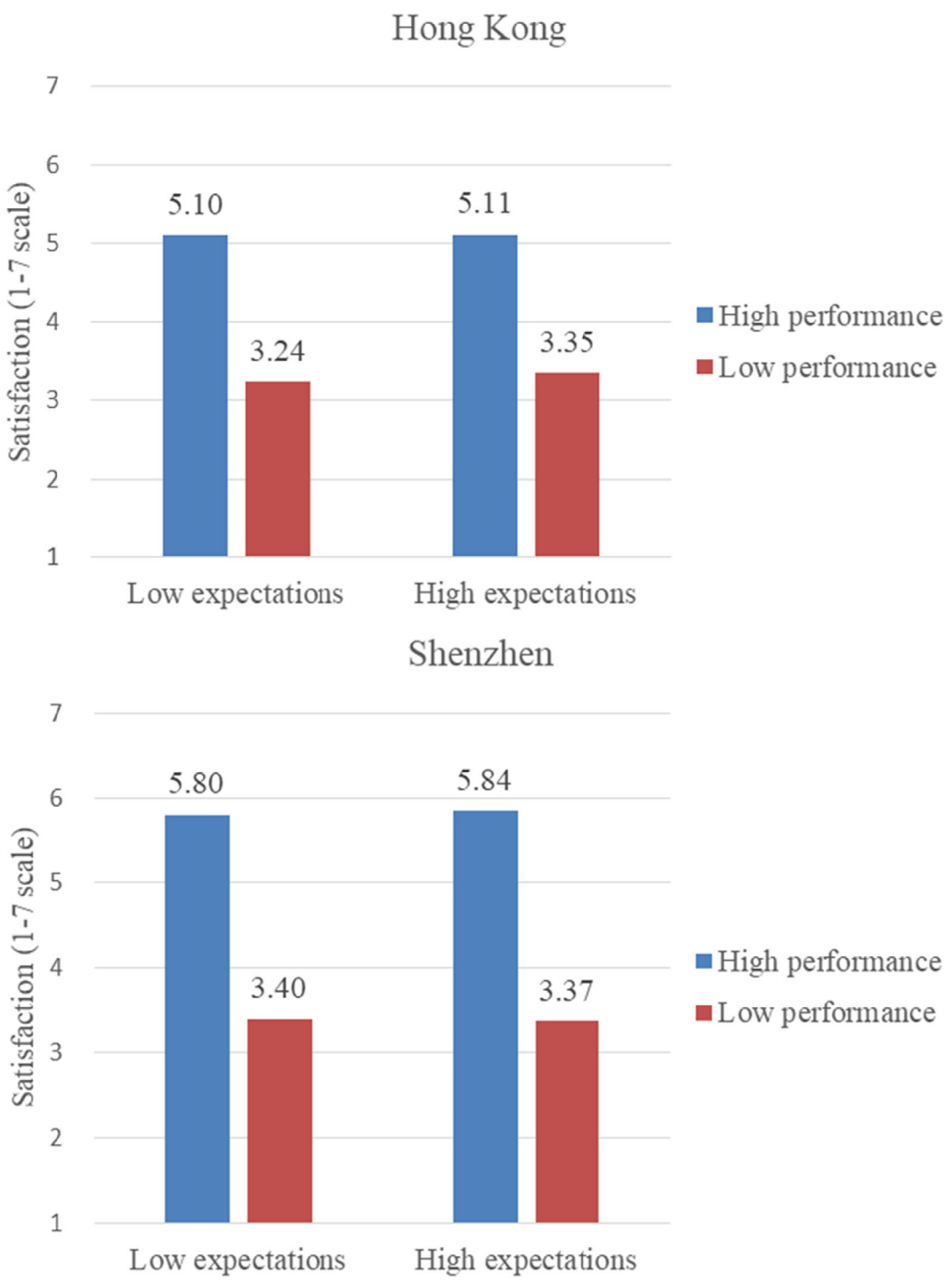

FIGURE 5 Graph of satisfaction ratings by experimental arm

authoritarian form of governance, it enjoys flexible economic policies and government measures and is at the forefront of service innovation, information disclosure, and performance evaluation. As Shenzhen is catching up to Hong Kong quickly in terms of economic indicators and public service quality, it is likely that citizens in Shenzhen have

higher expectations in the government than citizens in Hong Kong. That is, citizens' expectation levels are sensitive to what the government provides. From this perspective, our study responds to Van Ryzin's (2013) call to explore the interaction effect of expectations and performance in different contexts and points to a potential direction for future research. Future studies should consider the absolute level of citizen expectations when examining the interaction effect of expectations and performance on satisfaction.

Overall, our replication findings offer robust evidence supporting the external validity of previous experimental findings, greatly increasing the generalizability of the expectancy-disconfirmation model across contexts. Future research should delve deeper into the concept of expectations. For example, there are two types of expectations: empirical (predictions of what is likely to happen based on prior experiences) and normative (standards of what should happen) (Favero & Kim, 2021; James, 2011). Following Van Ryzin, our study focused on empirical expectations only, and future studies could compare how the two types of expectations differ in influencing citizen satisfaction. Second, citizens form their expectations of public services over the long term in a cumulative process by interacting with the government, watching social media, and communicating with neighbors and friends. The manipulation of expectations in experimental studies may not reflect how expectations operate in the real world (Hjortskov, 2020) or may shift citizens' empirical expectations but not their normative expectations. Future studies could explore better ways to manipulate expectations, especially normative expectations.

In addition, studies have focused predominantly on Western settings. Our study extends this line of research to two Chinese cities. However, the two cities may not represent the dynamics of citizen–state interactions in all of China due to cultural and socioeconomic variations across the country. Future studies could examine the applicability of the model elsewhere in China and in other Asian contexts (Kim, 2010). As Van Ryzin (2013) noted, street cleanliness is an easily perceived, commonplace public service, meaning that citizens can observe differences in performance and make evaluative judgments of government accordingly. However, different services vary substantively in terms of their policy salience, technical complexity, and relevance to different groups of citizens (Hasenfeld, 2010). For example, citizens may assess the performance of health care services by their accessibility, efficiency, and equity, characteristics that may not be discernible through street cleanliness. Further research could also use more representative and comparable sample of citizens. While this is a limitation of the study, the replication of policy area of street cleanliness should reduce this influence because it is a service all citizens are likely to experience. Future studies could further examine the model in different policy areas and use different performance dimensions to capture the quality of public services.

In conclusion, this study contributes to theoretical, methodological, and policy debates on the expectancy-disconfirmation model, which describes a causal process through which citizens form their judgments about the quality of public services and in turn government performance. Our empirical results largely confirmed Van Ryzin's (2013) findings but revealed the subtleties of the expectancy-disconfirmation model in the Chinese context. The combined insights help to theorize the process of citizen satisfaction across settings. In summary, public administration research should put more emphasis on applying replication frameworks to test the boundary conditions of current theory and practice (James et al., 2017; Sievert, 2021).

AUTHOR CONTRIBUTIONS

Richard M. Walker conceived of the study. M. Jin Lee and Richard M. Walker designed the Hong Kong study. M. Jin Lee collected the data and conducted the analysis for the Hong Kong study. Wenna Chen, Binzizi Dong, Chih-Wei Hsieh, Ning Liu, Yao Wang, Bo Wen, Peiyi Wu, Xia Wu, and Jiasheng Zhang designed the Shenzhen study. Peiyi Wu implemented the Shenzhen study and collected and analyzed the data. Peiyi Wu, Richard M. Walker, and Jiasheng Zhang drafted the article. Wenna Chen, Binzizi Dong, Chih-Wei Hsieh, M. Jin Lee, Ning Liu, Yao Wang, Bo Wen, Wen Wen, Peiyi Wu, Xia Wu, and Jiasheng Zhang commented on the article.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data and doi file are openly available in Harvard Dataverse at: https://doi.org/10.7910/DVN/HGYXKL.

ORCID

Wenna Chen https://orcid.org/0000-0003-2543-8271
Binzizi Dong https://orcid.org/0000-0002-1471-2661
Chih-Wei Hsieh https://orcid.org/0000-0003-1775-4709
M. Jin Lee https://orcid.org/0000-0001-5191-1697
Ning Liu https://orcid.org/0000-0001-9418-7694
Richard M. Walker https://orcid.org/0000-0003-1503-1814
Yao Wang https://orcid.org/0000-0001-6129-8291
Bo Wen https://orcid.org/0000-0003-2287-473X
Wen Wen https://orcid.org/0000-0002-3694-2441
Peiyi Wu https://orcid.org/0000-0001-7867-736X
Xia Wu https://orcid.org/0000-0003-0650-2899
Jiasheng Zhang https://orcid.org/0000-0002-2994-6217

ENDNOTES

[1] Noda (2019) conducted an original study of citizen satisfaction to examine vertical relationships between subnational governments. Noda (2019) replicated the research design of Van Ryzin (2013) and Filtenborg et al. (2017) but did not seek to conduct a replication by verifying the findings of the expectancy-disconfirmation model in a different context.

[2] Achieving face validity is essential for a successful replication. However, achieving face validity involves changes in measurement from the original study, in this, case the contextual information in the expectations cue. Furthermore, our two replications were undertaken in two different jurisdictions in China, and at different times. It is possible that the contextual information may result in a different interpretation of expectations. We return to this issue in our Discussion section.

[3] We were unable to conduct some of the sub-analysis performed by Van Ryzin (2013) because our participants were from two cities rather than from across a country, and the political identification variables in the original study could not be replicated in the Hong Kong and Shenzhen contexts.

REFERENCES

Andersen, S.C. & Hjortskov, M. (2016) Cognitive biases in performance evaluations. *Journal of Public Administration Research and Theory*, *26*(4), 647–662.

Busse, C., Kach, A. & Wagner, S.M. (2017) Boundary conditions: why we need them, and when to consider them. *Organizational Research Methods*, *20*(4), 574–609.

Charbonneau, É. & Van Ryzin, G.G. (2012) Performance measures and parental satisfaction with new York City schools. *The American Review of Public Administration*, *42*(1), 54–65.

DeHoog, R.H., Lowery, D. & Lyons, W.E. (1990) Citizen satisfaction with local governance: a test of individual, jurisdictional, and city-specific explanations. *The Journal of Politics*, *52*(3), 807–837.

Favero, N. & Kim, M. (2021) Everything is relative: how citizens form and use expectations in evaluating services. *Journal of Public Administration Research and Theory*, *31*(3), 561–577.

Favero, N. & Meier, K.J. (2013) Evaluating urban public schools: parents, teachers, and state assessments. *Public Administration Review*, *73*(3), 401–412.

Filtenborg, A.F., Gaardboe, F. & Sigsgaard-Rasmussen, J. (2017) Experimental replication: an experimental test of the expectancy disconfirmation theory of citizen satisfaction. *Public Management Review*, *19*(9), 1235–1250.

Frank, B., Enkawa, T. & Schvaneveldt, S.J. (2015) The role of individualism vs. collectivism in the formation of repurchase intent: a cross-industry comparison of the effects of cultural and personal values. *Journal of Economic Psychology*, *51*, 261–278.

Grimmelikhuijsen, S. & Porumbescu, G.A. (2017) Reconsidering the expectancy disconfirmation model. Three experimental replications. *Public Management Review*, *19*(9), 1272–1292.

Hasenfeld, Y. (Ed.). (2010) *Human services as complex organizations*, 2nd edition. Thousand Oaks: Sage.

Hero, R.E. & Durand, R. (1985) Explaining citizen evaluations of urban services: a comparison of some alternative models. *Urban Affairs Quarterly*, *20*(3), 344–354.

Hjortskov, M. (2020) Interpreting expectations: normative and predictive expectations from the citizens' viewpoint. *Journal of Behavioral Public Administration*, *3*(1), 1–11.
Hofstede, G. (2001) *Culture's consequences: comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks: Sage.
James, O. (2011) Managing citizens' expectations of public service performance: evidence from observational and experimentation in local government. *Public Administration*, *89*(4), 1419–1435.
James, O., Jilke, S.R. & Van Ryzin, G.G. (2017) Behavioural and experimental public administration: emerging contributions and new directions. *Public Administration*, *95*(4), 865–873.
Jost, J.T., Federico, C.M. & Napier, J.L. (2009) Political ideology: its structure, functions, and elective affinities. *Annual Review of Psychology*, *60*, 307–337.
Kim, S. (2010) Public trust in government in Japan and South Korea: does the rise of critical citizens matter? *Public Administration Review*, *70*(5), 801–810.
Lee, K., Joshi, K. & Bae, M. (2009) A cross-national comparison of the determinants of customer satisfaction with online stores. *Journal of Global Information Technology Management*, *12*(4), 25–51.
Noda, Y. (2019) Citizen expectations and satisfaction of service performance: lessons from subnational governments in Japan. *Asia Pacific Journal of Public Administration*, *41*(3), 142–156.
Oliver, R.L. (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469.
Scott, I. (2010) *The public sector in Hong Kong: government, policy, people*. Hong Kong: University of Hong Kong Press.
Sharma, P., Chen, I.S. & Luk, S.T. (2012) Exploring the role of IND-COL as a moderator in the comprehensive service evaluation model. *Journal of International Consumer Marketing*, *24*(1–2), 129–142.
Sievert, M. (2021) A replication of "representative bureaucracy and the willingness to coproduce". *Public Administration*, *99*(3), 616–632.
Thomassen, J.P., Leliveld, M.C., Van de Walle, S. & Ahaus, K. (2017) Compensating citizens for poor service delivery: experimental research in public and private settings. *Public Administration*, *95*(4), 895–911.
Tsang, E.W.K. & Kwan, K.-M. (1999) Replication and theory development in organizational science: a critical realist perspective. *Academy of Management Review*, *24*(4), 759–780.
Tsoukatos, E. & Rand, G.K. (2007) Cultural influences on service quality and customer satisfaction: evidence from Greek insurance. *Managing Service Quality: An International Journal*, *17*(4), 467–485.
Van Ryzin, G.G. (2004) Expectations, performance, and citizen satisfaction with urban services. *Journal of Policy Analysis and Management*, *23*(3), 433–448.
Van Ryzin, G.G. (2006) Testing the expectancy disconfirmation model of citizen satisfaction with local government. *Journal of Public Administration Research and Theory*, *16*(4), 599–611.
Van Ryzin, G.G. (2013) An experimental test of the expectancy-disconfirmation theory of citizen satisfaction. *Journal of Policy Analysis and Management*, *32*(3), 597–614.
Van Ryzin, G.G. & Immerwahr, S. (2007) Importance-performance analysis of citizen satisfaction surveys. *Public Administration*, *85*(1), 215–226.
Walker, R.M., Brewer, G.A., Lee, M.J., Petrovsky, N. & van Witteloostuijn, A. (2019) Best practice recommendations for replicating experiments in public administration. *Journal of Public Administration Research and Theory*, *29*(4), 609–626.
Walker, R. M., Lee M. J., Hills, P., Kwok R. C.-W., Ho S.M.Y., Mok J.K.H., et al. (2017) The "Citizen Satisfaction Assessment Tool": applying expectancy disconfirmation theory to public services in Hong Kong. Central Policy Unit, Hong Kong Special Administrative Region Government, Hong Kong.
Zhang, J. (2012) From Hong Kong's capitalist fundamentals to Singapore's authoritarian governance: the policy mobility of neo-liberalising Shenzhen, China. *Urban Studies*, *49*(13), 2853–2871.
Zhang, J., Chen, W., Petrovsky, N. & Walker, R.M. (2021) The expectation-disconfirmation model and citizen satisfaction with public services: a meta-analysis and an agenda for best practice. *Public Administration Review*, *82*(1), 147–159.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Chen, W., Dong, B., Hsieh, C.-W., Lee, M. J., Liu, N., Walker, R. M., Wang, Y., Wen, B., Wen, W., Wu, P., Wu, X., & Zhang, J. (2022). A replication of "an experimental test of the expectancy-disconfirmation theory of citizen satisfaction". *Public Administration*, 1–14. https://doi.org/10.1111/padm.12860