The Expectancy-Disconfirmation Model and Citizen Satisfaction with Public Services
A Meta-analysis and an Agenda for Best Practice

Zhang, Jiasheng; Chen, Wenna; Petrovsky, Nicolai; Walker, Richard M.

Jiasheng Zhang
Wenna Chen
Nicolai Petrovsky
Richard M. Walker
City University of Hong Kong

The Expectancy-Disconfirmation Model and Citizen Satisfaction with Public Services: A Meta-analysis and an Agenda for Best Practice

Research Article

Jiasheng Zhang is a doctoral candidate in the Askew School of Public Administration and Policy, Florida State University. He works in the Laboratory for Public Management and Policy, the City University of Hong Kong. His research interests include applying quantitative methods, such as experimental design and computational social science methods, to examine issues related to citizen–state interactions, performance information, and collaborative governance.
Email: jzhan39@cityu.edu.hk

Wenna Chen is a doctoral candidate in the Department of Public Policy at the City University of Hong Kong, where she is currently a Research Assistant in the Laboratory for Public Management and Policy. Her research of interest includes public sector innovation, leadership turnover and mobility, policy diffusion, network governance, and performance information.
Email: wennachen2-c@my.cityu.edu.hk

Nicolai Petrovsky is an Associate Professor of public management in the Department of Public Policy at the City University of Hong Kong and Deputy Director of the Laboratory for Public Management and Policy. His research focus is on public service performance (including the role senior officials play in it), citizen–state interactions, and transparent and replicable research.
Email: npetrovs@cityu.edu.hk

Richard M. Walker is the Chan Hon Pun Professor in Behavioral and Policy Sciences, Chair Professor of Public Management in the Department of Public Policy, and Dean of the College of Liberal Arts and Social Sciences at the City University of Hong Kong, where he is the Director of the Laboratory for Public Management and Policy and Co-Director of the Center for Public Affairs and Law. Prof. Walker is a Fellow of the Academy of Social Sciences.
Email: rmwalker@cityu.edu.hk

Abstract: *The expectancy-disconfirmation model has become the predominant approach in explaining citizen satisfaction with public services. It posits that citizens compare the performance of a service against their expectations of that service. Satisfaction occurs if the perceived performance meets or exceeds the expectations. We provide the first meta-analysis of the empirical evidence on this relationship, and find that the model is supported across studies. However, our meta-analysis also indicates that research design choices affect the results and that the scope of public services examined is not comprehensive. We make best practice recommendations for future research to improve the measurement of citizen satisfaction.*

Evidence for Practice

- Across studies and settings, citizens' satisfaction with public services is not only a function of how well they perceive these services to work but also of their expectations of these services. It is therefore important to understand what citizens expect from public services when assessing their satisfaction.
- The expectancy-disconfirmation model (EDM) is a robust tool that governments can implement when assessing citizen satisfaction. We recommend the following best practices:
  - EDM has produced valuable insights and should continue to be applied to examine citizen satisfaction.
  - Citizens' expectations should be captured in satisfaction surveys, with an emphasis on measuring qualities public services *should* have (referred to as normative expectations).
  - It is important to track citizen satisfaction over time (as resources allow) so that changes in perceptions of performance, expectations, and satisfaction can be linked to managerial and environmental changes.
  - Future studies should examine citizens' experiences and satisfaction with individual services and should include human services.

A large literature in public administration has been addressing determinants of performance of public services. There is a rich and still evolving understanding of external and internal factors that distinguish better performing services and organizations from those performing less well, and of what can be done to improve lagging performance. Many services directly affect citizens. Therefore, it is vital to also understand how citizens perceive and process the performance of these services. The literature on citizen satisfaction has shown that there is no one-to-one translation of the performance of a public service into how satisfied citizens are with it (Stipak 1979; Mok, James, and Van Ryzin 2017). In many contexts, decisionmakers about public services are ultimately accountable to citizens, who will consider their satisfaction when evaluating these decisionmakers. Citizen satisfaction can also be linked to engagement with, co-production of, and legitimacy attributed to public services, which is relevant even

in contexts without elections. Citizen satisfaction is therefore of great practical relevance. It is also theoretically important, as the field of public administration is building micro-foundations for key theoretical frameworks. In the area of citizen-state interactions, this includes the mechanisms behind citizen satisfaction.

There is an extensive body of public administration literature examining the external determinants of citizen satisfaction (Brown and Coulter 1983; Jilke 2018; Jilke and Baekgaard 2020; Kelly and Swindell 2002; Vigoda 2000, 2002; Vigoda-Gadot 2006; Vigoda-Gadot and Mizrahi 2006). Research has also explored the challenges associated with measuring citizen satisfaction (Andrews, Boyne, and Walker 2006; Cowell et al. 2012; Parks 1984; Stipak 1979) and made recommendations for the collection of citizens' perceptions of satisfaction (Dalehite 2008; Swindell and Kelly 2000; Van Ryzin and Immerwahr 2007). An important development

has been the application of the expectancy-disconfirmation model (EDM) to understand citizen satisfaction. The EDM goes beyond earlier approaches to explaining satisfaction by focusing on citizens' minds, rather than understanding satisfaction only in relation to external characteristics such as service characteristics and demographic variables. The EDM is also interesting as a topic of study because it is a key example of a psychological theory being introduced to and adapted by the field of public administration. This happened a decade before the notion of "behavioral public administration" (Grimmelikhuijsen et al. 2017) was coined. It can therefore be argued that research on citizen satisfaction using the EDM forms part of the intellectual roots of the current wave of interest in psychologically informed micro-level studies in public administration.

The EDM proposes that citizen satisfaction arises from a process in which citizens compare their perceptions of the performance of a public service against their prior expectations. The EDM originated in the marketing literature (Oliver 2010). Customer satisfaction is a central concept in this literature, and it can refer to several aspects associated with a consumption experience: the process of consumption and what occurs during it (e.g., for a hotel stay, the check-in experience and the comfort of the room), the outcomes associated with consumption (e.g., for a hotel stay, increased relaxation afterwards), and a reflective assessment of whether the level of satisfaction gained did or did not match what the customer expected (Oliver 2010, Ch. 1). The EDM developed the latter aspect and became widely used in the marketing literature, with substantial evidence supporting it as well as indicating needs for further refinements (Oliver 2010, Ch. 4; Szymanski and Henard 2001).

Since the study of Van Ryzin (2004), who first tested the EDM's applicability to public services, the EDM has become the predominant framework for explaining citizen satisfaction. The core insight for scholars and practitioners from this line of research has been that it is necessary to understand the standard (typically, the expectation) against which citizens assess the performance of a public service. Without this knowledge, investments in improving a public service's performance may not result in increased citizen satisfaction. While this has been a useful advance in knowledge, recent work has identified some limitations of the framework, and has especially shown that it is necessary to better theorize the determinants of expectations. There is some evidence that expectations, while quite stable overall, are somewhat affected by satisfaction in the past (Hjortskov 2019). In addition, our review of the literature points to multiple ways in which the EDM framework has been implemented. Thus, the time is ripe to take stock of the literature testing the EDM and to identify fruitful directions for citizen satisfaction research going forward.

We proceed in two stages. First, we ask two questions: How well supported are the theoretical linkages proposed by the EDM in the empirical literature? Do the findings differ based on different operationalization and study design? To answer them, we conduct a meta-analysis of all published studies testing the EDM. We use meta-regression to determine whether different study design and operationalization affect the conclusions about this approach in explaining citizen satisfaction. Second, drawing on our review of research and the findings of our analysis, we identify the best

practices for testing the EDM, including a greater focus on complex human services.

Our meta-analysis makes three contributions. First, we show how existing published studies all broadly support the core hypothesis of the EDM. Second, we determine exactly how study design and operationalization lead to different findings. Finally, we break down the findings and examine areas where the evidence base is strong and those for which it is weak or nonexistent, and we use this to make a number of best practice recommendations.

We set out by briefly introducing the EDM. We then provide an overview of how it has been tested, and identify different operationalization and research design choices across the existing studies. This is followed by a description of our meta-analytic procedures for the 17 studies, 24 samples, and 163 correlations identified in our search. We then present and interpret our findings. This is followed by a discussion of recommendations for best practices including a suggestion for a greater focus on human services, which would have important payoffs both for the generalizability of knowledge about citizen satisfaction and the management of human-centered public services.

The Expectancy-Disconfirmation Model of Citizen Satisfaction

Compared to prior work, the crucial innovation of the EDM is to draw attention to the determinants of satisfaction that are internal to individuals. The main idea of the model, originally developed in consumer behavior research, is that satisfaction or dissatisfaction is a function of both a referent (a standard against which comparison is made) and perceived performance (Oliver 1977, 1980). Typically, expectations provide the referent against which people assess performance (Oliver 2010, pp. 63–64).[1] Positive disconfirmation (better performance than expected) results in satisfaction, and negative disconfirmation (worse performance than expected) results in dissatisfaction (Oliver 1980; Spreng, MacKenzie, and Olshavsky 1996). The greater the value of disconfirmation, the greater is the difference between performance and expectation.

The complete EDM is depicted in Figure 1 (following Van Ryzin 2004, 435), including the hypothesized relationships.

Expectations, which are correlated with perceived performance (link D), affect disconfirmation (link A), together with perceived performance (link B).[2] If perceived performance meets or exceeds expectations (known as "positive disconfirmation"), satisfaction ensues. If perceived performance falls short of expectations (known as "negative disconfirmation"), dissatisfaction ensues. Link C depicts the disconfirmation–satisfaction relationship, which is the core of the EDM. In addition, perceived performance affects satisfaction directly (link E), as do expectations (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).

Variation between EDM studies
Beginning with Van Ryzin's (2004) study, which was the first to test the EDM in public administration, the literature initially followed variations of the same observational research strategy:
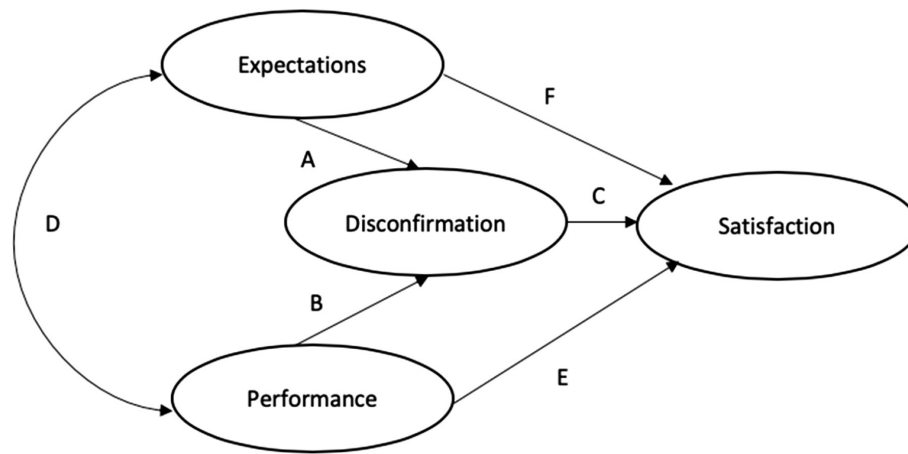
Figure 1   Expectancy-Disconfirmation Model. Adapted from Van Ryzin (2004)

operationalize most or all of the concepts in the EDM as survey items, collect them from a sample of citizens in a single, one-time survey, and then use a multivariate method of analysis to test the linkages hypothesized by the EDM. Methods of analysis included linear regression and binary choice models. Some authors estimated most of the relationships in a single model (e.g., James 2009; Petrovsky, Mok, and León-Cázares 2017), whereas others estimated a series of linear regressions (e.g., Poister and Thomas 2011) and then entered the path coefficients in a single results graph. Since Van Ryzin (2013), several studies have used randomized experimental manipulations in survey experiments. Looking at the variety of research designs, one could say in this article that we are comparing apples, oranges, and tomatoes. Beyond the differences, it is valuable to recognize that they are all still fruits and can be compared according to different aspects of their nutritional value. To make the comparison appropriate and to identify salient differences between study findings, we consider several sources of heterogeneity.

Type of Expectation. Expectations have become the subject of research in their own right. James (2011) showed that information about previous performance shapes citizens' expectations about future performance: high performance in the past leads to high expectations, and low performance leads to low expectations. These expectations then shape citizen satisfaction, along with the level of perceived performance at the time of experiencing the service in question. More recently, Jacobsen, Snyder, and Saultz (2015) took a broader view of expectations, distinguishing expectations about performance (the focus of the studies cited thus far) from expectations about goals, that is, what services should be provided, and how. Indeed, there are multiple ways for citizens to set a standard. Oliver (2010), the original creator of the EDM, noted that both external and internal factors can be the origins of expectations in consumers' minds. These factors include promotional claims, word of mouth, third-party information, and product cues on the external side and ease and vividness of recall on the internal side (Oliver 2010, pp. 74–77). To date, the public administration literature has focused on the distinction between normative and positive/empirical expectations (Hjortskov 2019; James 2011). Normative expectations refer to the qualities public services ought to have, whereas positive/empirical expectations refer to the qualities public services are likely to have. The former are

fairly stable but the latter are potentially amenable to adjustments based on new experiences with services. First, James (2011) found that positive, but not normative expectations were influenced by information about past performance in an experiment (and by performance as measured using archival indicators in an observational study). Then, Hjortskov (2019) corroborated this pattern in a different setting with a different research design. There is some consistent evidence on the greater stability of normative expectations then.

More specifically, based on his findings from panel data, Hjortskov (2019) suggests that citizens' expectations consist of a stable and an adaptive component. The former maps onto normative expectations and the latter onto empirical ones. In his study, changes in empirical expectations are not only explained by changes in performance, which raises concerns for managers about how to respond to those. The stable normative expectations, by contrast, offer more of a reliable North Star to aim for developing a strategy for a service. For a quantitative summary of the pattern of findings in the literature, it is then helpful to distinguish between empirical expectations, normative expectations, and expectations that encompass both these aspects. We thus include expectations as a moderator in our meta-analysis, in particular to test whether empirical expectations more strongly correlate with perceptions of performance and more strongly affect disconfirmation and satisfaction.

Type of Disconfirmation. When conducting a citizen satisfaction survey that takes into account the key lesson of the EDM (expectations matter), scholars and agencies have the choice of either determining disconfirmation indirectly by measuring the difference between the respondent's stated performance perception and their stated expectation, or by including an item that directly measures disconfirmation (along the lines of "thinking back to your expectations for the service, did the performance of the service meet, exceed, or fall short of your expectations?"). The former subtractive approach was first adopted in the literature. Later, Van Ryzin (2006) introduced the direct measure of disconfirmation to the public administration literature. As disconfirmation is the central concept in the EDM, it is important to know whether the way it is operationalized affects the findings. We thus include it as a moderator in our meta-analysis.

Type of Service. In a number of areas, we know that citizens perceive and evaluate government and public services differently depending on whether they are thinking about them in general terms or focusing on a specific agency or service (Cowell et al. 2012). We aim to determine whether this distinction between general and specific services moderates the findings on how expectations and perceived performance affect citizen satisfaction. We therefore code the focus of each study as either general (if the survey asked respondents about services in general) or specific (if the survey asked respondents about specific services such as garbage collection or road maintenance) and include this dummy variable as a moderator in our meta-analysis.

Country (United States or Elsewhere). The initial studies testing the EDM were all conducted in the United States, which is also where the theory originated in the marketing literature. As the United States is a particularly customer- and market-centered society, it is not clear ex ante whether a theoretical framework developed with these foundations works as well in other settings. Because the model has now been tested in multiple studies outside the United States as well, we include a dummy for that country as a moderator in our meta-analysis to test whether the findings are contingent on country.

Study Type (Experimental or Observational). Experimental tests of (aspects of) the EDM are different in that they do not all rely on the recall of expectations or performance by the survey respondents. Rather, expectations and/or levels of performance become treatments manipulated by the researcher. In other words, rather than assuming that expectations are exogenous as in the observational studies, they are made exogenous by design. We seek to test whether this fundamental difference in study design materially affects the findings. We therefore include a dummy for study type as a moderator in our meta-analysis.

Data and Methods
Data
We incorporated several data sources to conduct a comprehensive article search. First, we searched articles on EDM in public administration journals using Web of Science. We used the terms "expecta*" and "disconfirmation" in the Web of Science "topic" search engine. A total of 38 studies under the discipline of public administration were found. By carefully reviewing these studies, 15 articles were identified that directly tested the EDM. Second, we conducted a supplementary search in Google Scholar on the words "citizens" and "expectancy disconfirmation" in public sources; 71 items were identified but no new studies were found. Third, we conducted a supplementary search on studies that cited Van Ryzin's 2004, 2006, and 2013 articles in Google Scholar; one dissertation and one study from the discipline of urban studies were identified and added to the data. This search procedure produced 17 studies in total, which included 24 samples and 163 correlations across the six linkages in the EDM. Our search process is depicted in Figure 2.

Study Characteristics
Table 1 provides an overview of our dataset of the published EDM studies in public administration, including which paths in the EDM were tested by the study and the values of the moderators. The

average sample size was 3,250, ranging from a minimum of 157 to a maximum of 18,611. Five studies examined all EDM paths, four studies examined five paths, one study examined three paths, five studies examined two paths, and the remaining two studies examined a single path. The expectations were operationalized empirically nine times, and normative and generic expectations were both used on six occasions. A direct measure of disconfirmation was used six times and a subtractive measure nine times. Six studies did not include a measure of disconfirmation.

There was variation in the types of service examined. Four studies examined services in general and asked questions such as the following: "How satisfied are you with the performance of your local authority's overall services?" (James 2009, 121). Of these studies, three examined local governments and one the federal government in the United States. Table 1 indicates that 14 studies examined individual services, and the majority focused on technical services such as street cleanliness, roads, refuse, and traffic flow.[3] More human-oriented services included education, with three of the four education studies exclusively examining this service, and policing, with one of the two studies examining police–community relationships. The majority of the 17 studies were conducted in a single country. Eight studies were undertaken in the United States, three each from Denmark and the United Kingdom, and one each from Japan, Mexico, and South Korea. One study took samples from the Netherlands and the United States. The bulk of the studies used an observational research design (11); experimental designs were used seven times.[4] For informational purposes, the final column in Table 1 presents the statistical techniques used in the studies reviewed.

Meta-analytical Procedure
We conducted a meta-analysis, or an "analysis of analyses," to aggregate the quantitative results from different studies into a single, integrated, and quantitative literature review. A meta-analysis is defined by Glass (1976, 3) as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings." We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews and meta-analysis (Liberati et al. 2009). Our meta-analysis relied on the guidelines proposed by Ringquist (2013), and we conducted it using the statistical software R, with the "*meta*" and "*metafor*" packages. We conducted the analysis in four stages: (i) calculation of effect sizes; (ii) overall meta-analysis of the effect sizes in the EDM; (iii) meta-regression to test whether the type of expectation, the type of disconfirmation, the country setting, and the research design moderate the effect sizes; and (iv) examination of potential publication bias.

We iteratively developed a detailed code sheet containing all details of how to code (cf. Ringquist 2013, 86). When identifying a relationship between, for instance, perceived performance and satisfaction, we used the regression coefficients (or other measures of association) reported in the paper. Commonly, in an observational study, a test of the effect of perceived performance on satisfaction contains other variables: (i) other explanatory variables of theoretical interest and (ii) control variables. Suppose we reviewed an article where an analysis explaining satisfaction is done using linear regression with the following explanatory variables:
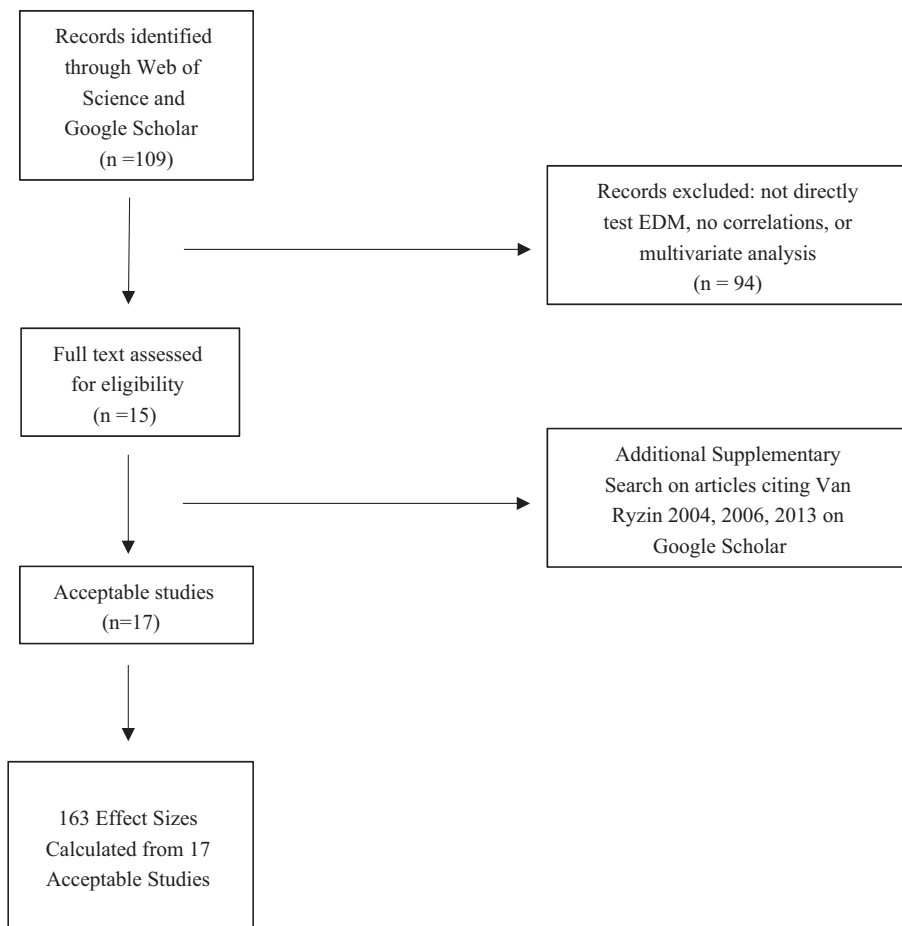
Figure 2 Literature Search Protocol and Results

perceived performance, disconfirmation (the other explanatory variable of interest), age, and gender (control variables). Regression coefficients and standard errors are reported for each of these four explanatory variables. We take the estimated coefficient on perceived performance and the associated standard errors and use it to calculate the size of the effect of perceived performance on satisfaction in this analysis. Our coding of studies includes a note on whether control variables are included in an analysis. (Code sheets and data will be uploaded to an open data repository on publication of the manuscript.)

To code studies, we followed the canonical version of the EDM, introduced above (cf. Figure 1). We reviewed each of the analyses in the articles/papers in our database to see whether it contained one or more of the six effects/correlations of the canonical EDM:

A: Expectations affect disconfirmation.
B: (Perceived) performance affects disconfirmation.
C: Disconfirmation affects satisfaction.
D: Expectations are correlated with performance.
E: (Perceived) performance affects satisfaction.
F: Expectations affect satisfaction.

Then we coded coefficients or correlations pertaining to these six types in our data set. The majority of the articles in our data included estimation of more than one model. For instance, Poister

and Thomas (2011) had three different areas of user satisfaction with Georgia's highways estimated in three different models: (i) condition and ride quality, (ii) traffic flow and lack of congestion, and (iii) safety. As suggested by Ringquist (2013, pp. 73–74), we coded each of the effect sizes from these models separately, but we noted that they stemmed from the same survey of the same sample of citizens. We accounted for the non-independence of observations from the same sample in our analysis by clustering the standard errors by sample.

To begin, we calculated population effect sizes across all correlations and studies. We used a random-effects (RE) meta-analysis with Fisher's r-to-Z transformation. The RE model is preferable to a fixed-effects model because we seek to generalize the study findings to a broader population (i.e., not only the studies incorporated into the meta-analysis). We used Fisher's r-to-Z transformation because it is a convention of meta-analysis (Ringquist 2013). In addition to calculating the population effect size, its significance, and its 95% confidence interval, we identified the heterogeneity of effect sizes using a χ^2 test.

Next, we conducted moderator analysis using random-effects meta-regression with clustered standard errors at the sample level (Ringquist 2013). In these models, the dependent variables consisted of the Z-transformed effect sizes, and the hypothesized moderators were added as explanatory variables. We reported

Table 1 Studies on Expectancy-Disconfirmation Model

Article	Data Size	Number of Effects for Each Link						Expectation	Disconfirmation	Type of Service	Country	Study Type	Method
		A	B	C	D	E	F						
Van Ryzin (2004)	1,504	1	1	1	1	1	1	Generic	Subtractive	9 city services[a]	USA	Observational	Standardized path analysis
Roch and Poister (2006)	1,321		1	3		3	3	Generic	Direct	Trash, police, school	USA	Observational	Ordered probit
Van Ryzin (2006)	615	4	4	3	3	3	3	Generic	Subtractive, direct	11 city services[b]	USA	Observational	Standardized path analysis
James (2009)	3,613			2			2	Normative	Subtractive	Refuse, general satisfaction	UK	Observational	Binary probit
James (2011)	2,495, 355				13			Empirical, normative	N/A	Local services in general	UK	Observational, experimental	OLS, ordered probit, binary probit, t-test
Poister and Thomas (2011)	1,006	3	3	3		3	3	Normative	Direct	Road conditions, traffic flow, highway safety	USA	Observational	Standardized path analysis
Morgeson (2012)	1,480	1	1	1	1	1	1	Generic	Direct	Federal service in general	USA	Observational	Standardized path analysis
Van Ryzin (2013)	964	1	1	1	1	1	1	Empirical	Subtractive	Street cleanliness	USA	Experimental	Regression path model, standardized path analysis, bivariate correlation
Berryman (2015)	1,207	4	4	4	1	4	4	Normative	Direct	Education	USA	Observational	
Andersen and Hjortskov (2016)	1,272, 944[c]			1		1	1	Empirical	Subtractive	Education	Denmark	Experimental	Binary logit
Petrovsky, Mok, and León-Cázares (2017)	1,302			3			3	Empirical, normative	Subtractive	Three aspects of service during in-person visit to municipal gov. office	Mexico	Observational	Binary probit
Filtenborg, Gaardboe, and Sigsgaard-Rasmussen (2017)	1,425	3	3	3		3	3	Empirical	Subtractive	Street cleanliness, school, library	Denmark	Experimental	Standardized path analysis
Grimmelikhuijsen and Porumbescu (2017)	309, 173, 165[d]	3	3	3		3	3	Empirical	N/A	Street cleanliness	USA	Experimental	Anova
Porumbescu (2017)	907					2	2	Generic	N/A	Local services in general	South Korea	Observational	OLS
Thomassen et al. (2017)	157, 937[e]			2				Generic	N/A	Driving license, package, visa	Netherlands, USA	Experimental	Anova
Hjortskov (2019)	17,411, 17,833, 18,611[f]	3	3	3	4	5	5	Empirical	N/A, Subtractive	Education	Denmark	Observational	OLS panel path, bivariate correlation
Noda (2019)	2,000	2	2	2	2	2	2	Empirical	Subtractive	Street cleanliness	Japan	Experimental	Standardized path analysis

[a]Street cleanliness, subway, buses, roads, library, parks, fire, schools, police (aggregated into one variable for analysis).
[b]Street cleanliness, roads, library, parks, fire, schools, ease of car travel, police, garbage, police–community relationship, public transportation (aggregated into one variable for analysis).
[c]Two experimental studies.
[d]Three experimental studies.
[e]Sample of 157 for Netherland, 937 for USA.
[f]Survey data was collected in 2009, 2011, and 2013.

the meta-regression regression coefficients for each hypothesized moderator and its significance and standard error, allowing us to identify the significant sources of variation in effect sizes overall.

To detect potential issues of publication bias, we conducted both visual and statistical tests. Publication bias implies that null findings are less likely to be published (Rosenthal 1979). A meta-analysis like ours, largely based on articles published in academic journals, might therefore overestimate population effects. We followed the recommendations of Ringquist (2013) and presented symmetric funnel plots as a visual assessment of potential publication bias in addition to the Egger test and Stanley and Doucouliagos's (2012) extension of this test to identify the significance of publication bias in our meta-analysis.

Findings
Meta-analysis
Table 2 presents the results of the RE meta-analysis using Fisher's r-to-Z transformation.[5] For each of the six EDM linkages, we found a statistically significant population effect size. Links B (performance affecting disconfirmation) and C (disconfirmation affecting

satisfaction) exhibit the largest effect sizes. Using Cohen's (1992) guidelines, we can describe the effect size for link E (performance affecting satisfaction) as moderate, and those for links A (expectations affecting disconfirmation), D (performance correlating with satisfaction), and F (expectations affecting satisfaction) as small to moderate. These results indicate that the EDM does fit data from multiple settings and services. However, there is substantial variation between studies: Q, the test statistic of the χ^2 test of homogeneity (Hedges and Olkin 1983), is statistically significant for all six linkages in the EDM, meaning there is substantial variation in effect sizes. We therefore next examined how the moderators introduced above explain this variation by using RE meta-regressions.

Meta-regression Analysis
Table 3 presents the results of the RE meta-regression analyses with clustered standard errors at the sample level. We allowed for correlations between effect sizes from the same set of respondents because some of the EDM studies analyzed the satisfaction of respondents with multiple services, and they did so in separate models. Each of the six linkages in the EDM is examined in a

Table 2 Results of Meta-analysis

Correlations	Number of Correlations	Population Effect Size	95% CI	Q	I^2	τ^2	t	
A (Expectation-disconfirmation)	22 (9 studies)	−0.248	[−0.434, −0.061]	4,793.92 ***	99.60%	0.176	−2.76	*
B (Performance-disconfirmation)	22 (9 studies)	0.467	[0.346, 0.589]	3,465.54 ***	99.40%	0.074	8.00	***
C (Disconfirmation-satisfaction)	32 (14 studies)	0.397	[0.298, 0.496]	2,855.39 ***	98.90%	0.074	8.18	***
D (Performance-expectation)	23 (7 studies)	0.174	[0.059, 0.290]	3,436.33 ***	99.40%	0.070	3.13	**
E (Performance-satisfaction)	30 (12 studies)	0.318	[0.232, 0.404]	6,317.22 ***	99.50%	0.051	7.58	***
F (Expectation-satisfaction)	34 (14 studies)	0.157	[0.067, 0.247]	3,272.27 ***	99.00%	0.064	3.56	***

*$p<.05$; **$p<.01$; ***$p<.001$.

Table 3 Results of the Meta-regressions[a]

Moderators	Model A: Expectation-disconfirmation	Model B: Performance-disconfirmation	Model C: Disconfirmation-satisfaction	Model D: Performance-expectation	Model E: Performance-satisfaction	Model F: Expectation-satisfaction
Constant	0.703***	0.476*	−0.732	0.402*	0.443***	−0.864*
	(0.146)	(0.215)	(0.460)	(0.088)	(0.107)	(0.335)
Disconfirmation (reference = Direct)						
Subtractive	−0.948***	−0.128	0.542*	−0.085	−0.167	0.663**
	(0.141)	(0.209)	(0.245)	(0.088)	(0.100)	(0.171)
Expectation (reference = Empirical)						
Generic	−0.498**	−0.043	−0.010	−0.265	−0.105	0.432
	(0.120)	(0.180)	(0.352)	(0.088)	(0.258)	(0.238)
Normative	−0.573**	−0.457	0.118		−0.158	0.201
	(0.147)	(0.232)	(0.351)		(0.284)	(0.242)
Empirical and Normative[b]			0.385			0.431
			(0.351)			(0.242)
Country (reference = Non-US)						
US	0.027	0.012	0.665*	0.307*	−0.103*	0.309
	(0.035)	(0.047)	(0.306)	(0.088)	(0.039)	(0.241)
Study type (reference = Experiment)						
Observational	−0.240***	0.307***	0.118		0.104	0.046
	(0.035)	(0.047)	(0.199)		(0.257)	(0.052)
Service type (reference = General)						
Specific			0.320			0.249
			(0.303)			(0.241)
Number of observations	22	22	30	8	25	27
Number of clusters	19	19	28	7	23	25
R^2	0.91	0.37	0.44	0.58	0.00	0.78
Q	160.97***	14.99*	27.59***	10.48**	1.96	83.86***

Notes:
[a]A negative sign for a coefficient indicates that compared to reference group, the moderator weakens the relationship.
[b]This moderator exists when regression model uses binary variable to capture empirical expectation (= 1) and normative expectation (= 0).

separate column of Table 3. Variation in the effect sizes is better explained by the moderators for some linkages in the EDM than for others. We now review each moderator in turn.

Type of Expectation. In the studies we reviewed, the expectations (or more broadly, the referent) against which citizens compare the performance of a service were either empirical (predictions of what is likely to happen), normative (standards of what should happen), or generic (referents that implicitly combine empirical and normative expectations). The only linkage in the EDM for which this moderator makes a statistically discernible difference is the effect of expectations on disconfirmation (linkage A), with generic and especially normative expectations leading to lower effect size. In other words, if an expectation includes a normative element, it is less likely to be met in the eyes of respondents in the studies we reviewed. This finding has an intuitive appeal. For example, Poister and Thomas (2011) and James (2009) both used normative expectations in terms of how well services should be performed, and found that normative expectations were inversely related to disconfirmation and satisfaction, meaning that citizens with higher expectations were more likely to be disappointed.

Type of Disconfirmation. When conducting a citizen satisfaction survey that considers the key lesson of the EDM (expectations matter), scholars and agencies can determine disconfirmation indirectly by taking the difference between each respondent's stated performance perception and his or her stated expectation. Alternatively, they can include an item that directly measures disconfirmation (along the lines of "thinking back to your expectations for the service, did the performance of the service meet, exceed, or fall short of your expectations?"). Some of the studies we reviewed use the former subtractive approach, whereas others used the latter direct approach. In our meta-regression analyses, we found that the choice of how to operationalize disconfirmation led to significantly different effect sizes for linkages A (expectations affect disconfirmation), C (disconfirmation affects satisfaction), and F (expectations affect satisfaction). In particular, expectations were strongly negatively related to disconfirmation when the subtractive operationalization was chosen. At the same time, subtractive operationalization led to much more positive effects of disconfirmation and expectations on satisfaction.

For example, Van Ryzin (2006) found that subtractive disconfirmation as compared to direct disconfirmation exerts much more influence on satisfaction. In addition, Filtenborg, Gaardboe, and Sigsgaard-Rasmussen (2017) and Roch and Poister (2006) examined the relationship between disconfirmation and satisfaction by using subtractive and direct measures, respectively, and the effect of subtractive disconfirmation was found to be more salient. Van Ryzin (2006) and Morgeson (2012) both found a positive relationship between expectations and disconfirmation when direct disconfirmation measures were used. As they explained, a subtractive measure of disconfirmation, although may make sense conceptually, introduces a statistical bias in the modeling results, whereas direct measure is more accurate.

Taken together, it is possible that these differences in effect sizes are due to the construction of the disconfirmation model, which under a subtractive operationalization was not measured separately but

calculated as the difference of performance and expectation items.

Type of Service. Four of the 17 studies in our meta-analysis analyzed citizen satisfaction with government services in general without specifying which service (e.g., schools, roads) respondents should think of. We could test this variable as a moderator only for two linkages: C (disconfirmation affecting satisfaction) and F (expectations affecting satisfaction), as we did not have variation on it for the other linkage. For neither linkage C nor F was there statistically discernible evidence that referring to a specific service type affected the direction or strength of the relationship.

Porumbescu (2017) examined general local services and found that there was no significant relationship between expectations of public sector performance and citizens satisfaction. This finding is consistent with Van Ryzin (2013), who also found little or no net effect of expectations on satisfaction in street cleanliness services.

Country (United States or Elsewhere). Studies conducted in the United States showed significantly different effect sizes for linkages C (disconfirmation affects satisfaction), D (performance correlating with expectation), and E (performance affecting satisfaction). For the first two linkages, the effect size was much larger in the US-based studies. At the same time, the effect size of performance affecting satisfaction was a tenth of a standard deviation smaller in the US-based studies.

Study Type (Experimental or Observational). As noted above, experimental studies of linkages in the EDM tended to differ substantially from observational studies. We find some reflection of this in the meta-analysis. Effect sizes for linkages A (expectations affecting disconfirmation) and B (performance affecting disconfirmation) were significantly different for studies with experimental designs. We note that the effect of expectations on satisfaction was stronger in the experimental studies, where researchers used treatments to set expectations.

Assessment of Study Heterogeneity and Possible Publication Bias. Figure 3 presents funnel plots, which are visual tests for potential publication bias (or, more broadly, heterogeneity between studies). The plots show that there is substantial heterogeneity between studies, but they do not indicate the pattern most expected under publication bias, which is a kind of hole (absence of studies) on the left side (small effect sizes) and especially in the upper left quadrant (small effect sizes estimated with high precision). A large number of correlations fell outside the funnel plot confidence limits. This may be an indication of publication bias. To statistically test this bias, we conducted Egger's test.

We first used the test proposed by Egger et al. (1997) (see Equation (1)) to examine the funnel plot asymmetry, which tests for $\gamma_0 = 0$ from a linear regression of normalized effect estimate (estimate divided by its standard error) against precision (reciprocal of the standard error of the estimate):

$$effect_i / Se_i = \gamma_0 + \gamma_1 \left(1 / Se_i \right) + e_i \tag{1}$$

If the intercept is significantly different from 0, there is evidence of funnel plot asymmetry. The results in Table 4 (upper part) show that only the intercept for Model E (Performance-satisfaction) is
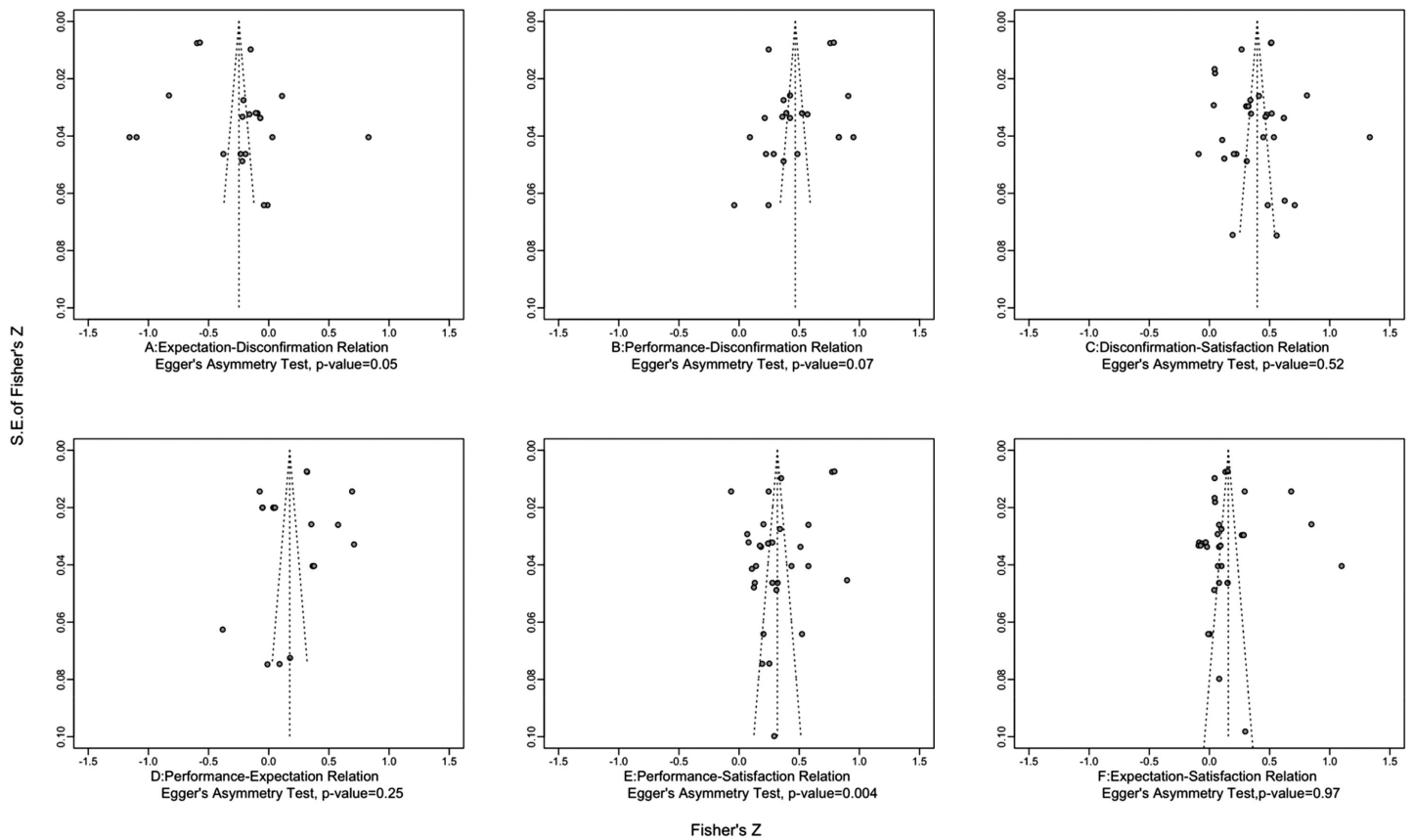
Figure 3 Funnel Plot of Effect Sizes

Table 4 Funnel Asymmetry Test: Inverse Standard Errors and Square Root of Sample Size

Explanatory Variables	Model A: Expectation-disconfirmation	Model B: Performance-disconfirmation	Model C: Disconfirmation-satisfaction	Model D: Performance-expectation	Model E: Performance-satisfaction	Model F: Expectation-satisfaction
Inverse standard errors	−0.562***	0.722***	0.439***	0.315**	0.691***	0.166**
	(0.088)	(0.076)	(0.057)	(0.083)	(0.075)	(0.058)
constant	9.501	−7.515	−1.745	−5.560	−11.366**	0.096
	(4.629)	(3.991)	(2.744)	(4.708)	(3.680)	(2.856)
R^2	0.67	0.82	0.66	0.41	0.75	0.21
F	40.79***	90.72***	58.63***	14.46***	86.04***	8.29**
N	22	22	32	23	30	34
SQR of sample size	−0.561***	0.723***	0.434***	0.305**	0.664***	0.164**
	(0.090)	(0.076)	(0.058)	(0.080)	(0.079)	(0.057)
constant	10.333*	−7.569	−1.469	−4.920	−10.574*	0.196
	(4.847)	(3.996)	(2.749)	(4.562)	(3.939)	(2.829)
R^2	0.66	0.82	0.65	0.41	0.72	0.21
F	38.52***	90.74***	56.99***	14.44***	70.93***	8.29***
N	22	22	32	23	30	34

Notes: Standard errors (in parentheses).
*p < .05; **p < .01; ***p < .001.

statistically significant (p < .01), suggesting that the funnel plot is asymmetric and that publication bias exists. However, Egger's test has been criticized for its low power in small-sample studies. Stanley and Doucouliagos (2012) proposed a more rigorous test that takes sample size into account by using the square root of the sample size as a precision measure (see Equation (2)):

$$effect_i / Se_i = \gamma_0 + \gamma_1 \left(SQR_{SampleSize} i \right) + e_i \qquad (2)$$

The results in Table 4 (bottom part) show that in addition to Model E, the intercept for Model A becomes statistically significant (p < .05),

suggesting that the funnel plot is asymmetric and publication bias exists.[6] Overall, publication bias may be an issue for Model A and E, and some caution is needed when interpreting the results.

Discussion and Outlook: Refocusing Empirical Tests to Enhance the Relevance of Expectancy-Disconfirmation Research

Satisfaction with public services is of critical importance to citizens and managers alike, and advancements in understanding its determinants are highly relevant to research and practice. To improve knowledge of the determinants of citizen satisfaction

with public services, we conducted a meta-analysis to provide a quantitative summary, comparison, and integration of 15 years of empirical research published in the field of public administration. Our meta-analysis found that the EDM is generally supported in studies of public services, with variation in the strength of relationships explained by research setting and research design choices.[7]

These findings have implications and insights for the development of knowledge of citizen satisfaction. We discuss six implications covering behavioral public administration, dynamically measuring citizen satisfaction with services, normative expectations, operationalization of disconfirmation through direct questions, satisfaction with specific services, and the extent to which the model can be extended across the full range of human services. All of these items have implications for research and practice, and we accordingly make eight recommendations for best practice.

A major advance resulting from the adoption of the EDM to study citizen satisfaction was the inclusion of determinants of satisfaction that are internal to citizens. In particular, satisfaction or dissatisfaction is a function of both a referent and perceived performance. The findings of this meta-analysis provide support for the model and for the importance of a standard against which to compare perceived performance, and they are consistent with a large body of findings in the marketing literature (Szymanski and Henard 2001). With its focus on considering the psychology of citizens, the EDM parallels the development of behavioral public administration, which combines insights from public administration and psychology to provide enhancements to practices and outcomes (Grimmelikhuijsen et al. 2017). While all frameworks have limitations (see below), the inclusion of this essential behavioral facet in the EDM marks an important advance on prior approaches to the collection of citizen satisfaction data. In short, the EDM is a productive approach that has produced valuable insights, and our first recommendation is that it should continue to be applied to examine citizen satisfaction.

Second, while the EDM is now prevalent in academic research on citizen satisfaction, governmental satisfaction surveys still frequently do not capture expectations or other referents.[8] However, the inclusion of a referent advances our understanding of the origins of citizens' satisfaction or dissatisfaction, which is important in choosing appropriate courses of action to address dissatisfaction with a service. The EDM draws attention to the relationship between expectations and perceptions of performance, indicating a certain degree of feedback between the two. Research evidence from Hjortskov (2019, 365) has suggested "a certain amount of temporal stability in expectations formation," and that "prior satisfaction and expectations can influence future expectations." These findings point to the importance of measuring satisfaction as a dynamic process, helping managers to assess the extent to which their efforts are recognized by citizens. Our second recommendation is therefore that citizens' expectations be captured by satisfaction surveys.

A third recommendation is to track citizen satisfaction dynamically over time (as resources allow), such that it can be better understood how citizens attribute changes in perceptions of performance, expectations, and satisfaction to managerial and environmental changes.

Hjortskov's (2019) panel data study shows a way forward to ensure that the temporal sequence implied in the EDM can be operationalized properly and expectations are measured at the relevant point in time, rather than as a recall. This is certainly more costly and raises concerns about differential attrition but, on balance, it will solve many more problems than it creates. Not the least, doing so will allow for an assessment of the validity of the direct disconfirmation question that we generally recommend. A fourth recommendation is that to ensure that the EDM is properly operationalized, expectations should be measured at the relevant point.

Expectations (or the referent against which citizens compare the performance of a service) are a core element of the EDM. In the research we covered in this meta-analysis, they are either based on (i) predictions of what is likely to happen (empirical), perhaps based on prior experiences, or (ii) standards of what should happen (normative). Both approaches have merit and reflect the ways in which citizens experience public services. One study reviewed in our meta-analysis undertook a more detailed examination of expectations. Hjortskov's (2019) research indicated both the high importance of relatively stable normative expectations and an affective (not strictly performance-based) set of more marginal adjustments to expectations as a result of prior satisfaction or dissatisfaction. Put simply, when a citizen repeatedly walks down a dirty street, she is likely to still expect it to be clean. As Hjortskov's (2019) research indicates, normative expectations affect satisfaction more strongly than empirical expectations. He conjectures that this is especially important when citizens assess "services that concern their children" (p. 365). We therefore recommend that both future academic research and public managers measure both types of expectations. The latter in particular will likely want to pay attention to normative expectations, the more so, the more salient their service is. Normative standards are often directly upheld by citizens, particularly for valence issues, when citizens across the political spectrum agree on desired outcomes (Green 2007; Stokes 1963). Hjortskov (2020) provides evidence that normative expectations are the most fundamental and important types of expectations among citizens. Most recently, Favero and Kim (2020) have also provided evidence that normative expectations have a stronger relationship with satisfaction. As Hjortskov (2016) argues, stable normative expectations can be seen as the standard or yardstick for the empirical expectations and this intersection between the two should be considered when conceptualizing expectations. At the managerial level, these standards are often further operationalized or extended through vision and mission statements that are made publicly available, and sometimes even specific information on performance.

Additional research is clearly necessary to unpack the antecedents of expectations. Our meta-analysis results pointed to lower disconfirmation when the referent point is normative, suggesting that it is a more demanding standard. At the same time, normative expectations are important to citizens, to candidates for elected office running on valence issues, and to government agencies with published standards of how they should perform. We therefore find normative expectations to be the most salient referent for public organizations, and the fifth recommendation is for research to explicitly measure them.

There are two ways to measure disconfirmation: the subtractive approach and a direct subjective question. As the subtractive approach requires the subtraction of two numbers from each other that may differ in mean and variance, the direct approach is less likely to introduce concerns about artifacts. In addition, a direct question is more directly anchored in citizens' experiences. Given this, as a sixth recommendation, future EDM studies should include a direct question on expectations.

In some studies, researchers have focused on public services in general, that is, satisfaction with local government (James 2009, 2011; Morgeson 2012), with a number of services (Roch and Poister 2006), and aggregates of services (Van Ryzin 2006), while others have examined specific services (Filtenborg, Gaardboe, and Sigsgaard-Rasmussen 2017; Grimmelikhuijsen and Porumbescu 2017; Noda 2019). The choice of taking the government as a whole or an individual service as the unit of analysis is likely to influence citizens' assessment of satisfaction. Research in England shows why this matters: Cowell et al. (2012) reported differing patterns of satisfaction for individual services and local governments as a whole, with more positive assessments of individual services.[9] This is likely to reflect citizens' experiences of using individual services (Van Ryzin and Charbonneau 2010) rather than their experiences with a more amorphous entity such as the local government as a whole. Therefore, the seventh recommendation is that future studies examine citizens' experiences and satisfaction with individual services.

Finally, citizen experiences of public services vary widely depending on the nature of public services. They range from life-and-death situations (treatment of a serious condition in a hospital) to the most mundane (user friendliness of a local government website). They also range from the highly complex, involving multiple interactions with providers who rely on expert knowledge, to the very straightforward, such as waiting for a traffic light to turn green. Besides, public services vary in the degree of choice citizens have (Brown 2007). For instance, in some settings, citizens can choose between different schools or transit providers, whereas in others there is no choice. As Hjortskov (2016) notes, the degree of choice may affect expectations, which has yet to be fully explored by empirical research. Indeed, looking back at Table 1, one of the main conclusions of our meta-analysis is the near-absence of EDM-informed studies of satisfaction with human services, certainly those directly experienced by citizens. Inclusion of human services in studies of citizen satisfaction provides additional evidence about citizen satisfaction and increases the generalizability of knowledge about citizen satisfaction with public services and the management of human-centered public services. However, interactions between citizens and such services are more complex in three ways. First, the general issue of multiple dimensions of performance is particularly acute here (Andrews, Boyne, and Walker 2006). Second, citizens typically experience these different dimensions of performance in temporal sequence, and expectations are likely to have a cumulative influence across these dimensions. That is, negative experiences at the outset (e.g., a long wait time to obtain an appointment with a medical specialist or to check into a hospital) are likely to lead citizens to update their expectations of other aspects of the service (e.g., supportiveness of nurses, quality of care) (Hasenfeld 2010). Third,

the affective dimension to the experience of human services is central (Hasenfeld 2010). For other services, its salience varies based on how important these services are to citizens' directly experienced quality of life, and this may in turn depend on thresholds. For instance, street sweeping, a technical service, will not normally be activating a salient dimension unless its performance is poor. That is, citizens will not likely react emotionally to differences between very clean and somewhat clean streets, but a dirty street will evoke a strong response.

Measuring citizen satisfaction with human services will require assessment of expectations and satisfaction at key stages in the process of their consumption, which makes it more important to record satisfaction dynamically and reference it to normative expectations of public service providers. Our final recommendation is that future EDM studies of citizen satisfaction include human services.

Conclusion

Citizen–state relationships are of central interest to students of public administration. To better understand citizen satisfaction with public services, scholars have turned to the EDM, which offers advantages over traditional approaches to measuring citizen satisfaction by taking account of citizens' behaviors, primarily by examining their expectations. In our meta-analysis of 17 studies, 24 samples, and 163 correlations, we found broad support for the model, and we recommend it for future use in research and practice. However, we noted variation in these findings based on the operationalization of the model, research design, and context. Eight best practices were highlighted to improve the validity and reliability of the model in future citizen satisfaction studies. We encourage others to implement these best practice recommendations to improve the knowledge and practice of the measurement of citizen satisfaction.

Along with a growing number of systematic English language public management studies conducted in settings other than the United States or Northern and Western Europe, there has been a recent focus on developing a theory on how context affects theorized relationships and can be systematically incorporated into research programs (e.g., O'Toole Jr. and Meier 2017; Walker et al. 2019). It is worth considering several of the broad contextual categories that O'Toole Jr. and Meier (2017, 20) discuss in light of their relevance to citizen satisfaction research. We focus on contextual categories that most directly affect performance and expectations, the two key components of the EDM. We conjecture the following: A more munificent environment, more social capital, and clear and consistent goals facilitate performance. Contextual variation in expectations is more difficult to theorize. We do suggest, however, that variation in expectations between citizens is larger where social capital is less. Based on these considerations, social capital would be the foremost contextual factor to vary when choosing where to conduct future tests of the EDM with an eye on identifying the extent of its generalizability and determining its boundary conditions. Among the set of studies we review, it appears that most are conducted in settings with relatively high levels of social capital.[10] Much insight is therefore to be gained from conducting tests of the EDM in settings with a high and low levels of social capital, ideally in a similar service area.

Notes

1. As Oliver (2010) points out, hopes, wishes, and anticipations can serve as the referent, not only predictions of future performance (p. 62).

2. In observational studies, data on perceived performance is obtained by directly asking respondents to rate the performance of services. In experimental studies, subjects' level of perceived performance is presumed to be evoked by vignettes with high or low performance information (Andersen and Hjortskov 2016), or pictures showing instances of high or low performance (Van Ryzin 2013). In experimental studies, the performance treatment has been shown to be significantly and positively related to perceived performance, by asking subjects a closed-end survey question. It indicates the success of the manipulation process by randomly assigning subjects into either high or low performance conditions.

3. James (2009) examined local government in general and one service.

4. James (2011) included observational and experimental research designs. The experimental studies we review manipulate both expectations and performance except one paper, which measured expectations as the dependent variable (James 2011). Therefore, we are not able to construct a moderator to differentiate the manipulation of performance and expectation.

5. We used the R packages *metafor* (Viechtbauer 2010), version 2.1–0, and *meta* (Schwarzer 2007), version 4.13–0, for our analysis. Upon publication, we will make the data and annotated code available on Dataverse.

6. In Table 4, two tests regress the standardized effect sizes on their precisions. If there is asymmetry, with smaller studies showing effects that differ systematically from larger studies, the regression line will not run through the origin. The intercept of the regression provides a measure of asymmetry—the larger its deviation from zero, the more pronounced the asymmetry. Thus, an intercept significantly different from zero (Stanley, 2008) provides evidence of publication bias.

7. The possibility of a file drawer problem exists in principle, that is, there are unpublished studies altering our conclusions that have been buried in researchers' "file drawers" in our searches and conversations with contributors to the expectancy-disconfirmation literature, we did not come across a single unpublished paper that would make a major difference to the insights we report.

8. For instance, the World Bank (2018) notes the relevance of expectations and includes related items in its bank of sample questions, but notes that only some models include expectations (p. 5).

9. There is a parallel to Fenno's (1978) paradox, that is, Americans have much more favorable views of their own representatives than of Congress as a whole.

10. According to different social capital indices (e.g., World Economic Forum Global Competitiveness Index), Denmark, Netherlands, the United States and United Kingdom are all listed as high social capital countries (top 10). Fourteen out of 17 studies in our analysis are conducted in these high social capital countries.

References

Andersen, Simon Calmar, and Morten Hjortskov. 2016. Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory* 26(4): 647–62 *.

Andrews, Rhys, George A. Boyne, and Richard M. Walker. 2006. Subjective and Objective Measures of Organizational Performance: An Empirical Exploration. In *Public Services Performance: Perspectives on Measurement and Management*, edited by George. A. Boyne, Kenneth.J. Meier, Laurence J. O'Toole, Jr., and Richard M. Walker, 14–34. Cambridge: Cambridge University Press.

Berryman, Anita. 2015. Modeling Public Satisfaction with School Quality: A Test of the American Customer Satisfaction Index Model. Doctoral Dissertation, Georgia State University.*

Brown, Trevor. 2007. Coercion Versus Choice: Citizen Evaluations of Public Service Quality across Methods of Consumption. *Public Administration Review* 67(3): 559–72.

Brown, Karin, and Philip B. Coulter. 1983. Subjective and Objective Measures of Police Service Delivery. *Public Administration Review* 43(1): 50–8.

Cohen, Jacob. 1992. A Power Primer. *Psychological Bulletin* 112(1): 155–9.

Cowell, Richard, James Downe, Steve Martin, and Alex Chen. 2012. Public Confidence and Public Services: It Matters What you Measure. *Policy & Politics* 40(1): 123–43.

Dalehite, Esteban G. 2008. Determinants of Performance Measurement: An Investigation into the Decisions to Conduct Citizen Surveys. *Public Administration Review* 68(5): 891–907.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. Bias in Meta-Analysis Detected by a Simple, Graphical Test. *British Medical Journal* 315: 629–34.

Favero, Nathan, and Minjung Kim. 2020. Everything Is Relative: How Citizens Form and Use Expectations in Evaluating Services. *Journal of Public Administration Research and Theory.* https://doi.org/10.1093/jopart/muaa048.

Fenno, Richard F. 1978. *Home Style: House Members in their Districts*. Boston: Little, Brown.

Filtenborg, Anders Foged, Frederik Gaardboe, and Jesper Sigsgaard-Rasmussen. 2017. Experimental Replication: An Experimental Test of the Expectancy Disconfirmation Theory of Citizen Satisfaction. *Public Management Review* 19(9): 1235–50 *.

Glass, Gene V. 1976. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher* 5(10): 3–8.

Green, Jane. 2007. When Voters and Parties Agree: Valence Issues and Party Competition. *Political Studies* 55(3): 629–55.

Grimmelikhuijsen, Stephan, and Gregory A. Porumbescu. 2017. Reconsidering the Expectancy Disconfirmation Model. Three Experimental Replications. *Public Management Review* 19(9): 1272–92 *.

Grimmelikhuijsen, Stephan, Sebastian Jilke, Asmus L. Olsen, and Lars G. Tummers. 2017. Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review* 77(1): 45–56.

Hasenfeld, Yeheskel, ed. 2010. *Human Services as Complex Organizations*, 2nd ed. Thousand Oaks: Sage.

Hedges, L.V., and I. Olkin. 1983. Clustering Estimates of Effect Magnitude from Independent Studies. *Psychological Bulletin* 93(3): 563–73.

Hjortskov, Morten. 2016. Citizen Satisfaction: Political Voice and Cognitive Biases. Doctoral Dissertation, Forlaget Politica. https://www.researchgate.net/profile/Morten_Hjortskov/publication/317259071_Citizen_Satisfaction_Political_Voice_and_Cognitive_Biases/links/592e9d34aca272fc55bdf9d4/Citizen-Satisfaction-PoliticalVoice-and-Cognitive-Biases.pdf (Retrieved February 3, 2021, from).

———. 2019. Citizen Expectations and Satisfaction over Time: Findings from a Large Sample Panel Survey of Public School Parents in Denmark. *The American Review of Public Administration* 49(3): 353–71 *.

———. 2020. Interpreting Expectations: Normative and Predictive Expectations as Seen by Citizens. *Journal of Behavioral Public Administration* 3(1): 1–11.

Jacobsen, Rebecca, Jeffery W. Snyder, and Andrew Saultz. 2015. Understanding Satisfaction with Schools: The Role of Expectations. *Journal of Public Administration Research and Theory* 25(4): 831–48.

James, Oliver. 2009. Evaluating the Expectations Disconfirmation and Expectations Anchoring Approaches to Citizen Satisfaction with Local Public Services. *Journal*

of *Public Administration Research and Theory* 19(1): 107–23 *.

———. 2011. Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government. *Public Administration* 89(4): 1419–35 *.

Jilke, Sebastian. 2018. Citizen Satisfaction Under Changing Political Leadership: The Role of Partisan Motivated Reasoning. *Governance* 31(3): 515–33.

Jilke, Sebastian, and Martin Baekgaard. 2020. The Political Psychology of Citizen Satisfaction: Does Functional Responsibility Matter? *Journal of Public Administration Research and Theory* 30(1): 130–43.

Kelly, Janet M., and David Swindell. 2002. Service Quality Variations Across Urban Space: Frist Steps Towards a Model of Citizen Satisfaction. *Journal of Urban Affairs* 24(3): 271–88.

Liberati, A., D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gøtzsche, J.P.A. Ioannidis, M. Clarke, P.J. Devereaux, J. Kleijnen, and D. Moher. 2009. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions: Explanation and Elaboration. *Journal of Clinical Epidemiology* 62(1): e1–e34.

Mok, Jue Young, Oliver James, and Gregg G. Van Ryzin. 2017. Expectations of and Satisfaction with Public Services. In *Experiments in Public Management Research: Challenges and Contributions*, edited by Oliver James, Sebastian R. Jilke, and Gregg G. Van Ryzin, 345–60. Cambridge, UK: Cambridge University Press.

Morgeson, Forrest V. 2012. Expectations, Disconfirmation, and Citizen Satisfaction with the US Federal Government: Testing and Expanding the Model. *Journal of Public Administration Research and Theory* 23(2): 289–305 *.

Noda, Yu. 2019. Citizen Expectations and Satisfaction of Service Performance: Lessons from Subnational Governments in Japan. *Asia Pacific Journal of Public Administration* 41(3): 142–56 *.

Oliver, Richard L. 1977. Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation. *Journal of Applied Psychology* 62(4): 480–6.

Oliver, Richard L. 1980. A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research* 17(4): 460–9.

Oliver, Richard L. 2010. *Satisfaction: A Behavioral Perspective on the Consumer*, 2nd ed. Armonk: M. E. Sharpe.

O'Toole, Laurence J., Jr., and Kenneth J. Meier. 2017. Comparative Public Management: A Framework for Analysis. In *Comparative Public Management: Why National, Environmental, and Organizational Context Matters*, edited by Kenneth J. Meier, Amanda Rutherford, and Claudia N. Avellaneda, 1–26. Washington, DC: Georgetown University Press.

Parks, Roger B. 1984. Objective and Subjective Measures of Performance. *Public Administration Review* 44(2): 118–27.

Petrovsky, Nicolai, Jue Young Mok, and Filadelfo León-Cázares. 2017. Citizen Expectations and Satisfaction in a Young Democracy: A Test of the Expectancy-Disconfirmation Model. *Public Administration Review* 77(3): 395–407 *.

Poister, T.H., and J.C. Thomas. 2011. The Effect of Expectations and Expectancy Confirmation/Disconfirmation on Motorists' Satisfaction with State Highways. *Journal of Public Administration Research and Theory* 21(4): 601–17 *.

Porumbescu, Gregory. 2017. Not all Bad News after all? Exploring the Relationship between Citizens' Use of Online Mass Media for Government Information and Trust in Government. *International Public Management Journal* 20(3): 409–41 *.

Ringquist, Evan J. 2013. *Meta-Analysis for Public Management and Policy*. San Francisco: Jossey-Bass.

Roch, Christine H., and Theodore H. Poister. 2006. Citizens, Accountability, and Service Satisfaction: The Influence of Expectations. *Urban Affairs Review* 41(3): 292–308*.

Rosenthal, Robert. (1979). The File Drawer Problem and Tolerance for Null Results. *Psychological Bulletin* 86(3): 638–641*.

Schwarzer, G. 2007. Meta: An R Package for Meta-Analysis. *R News* 7(3): 40–5.

Spreng, Richard A., Scott B. MacKenzie, and Richard W. Olshavsky. 1996. A Reexamination of the Determinants of Consumer Satisfaction. *Journal of Marketing* 60(3): 15–32.

Stanley, T.D., and Hristos Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business*. New York: Routledge.

Stipak, Brian. 1979. Citizen Satisfaction with Urban Services: Potential Misuse as a Performance Indicator. *Public Administration Review* 39(1): 46–52.

Stokes, Donald E. 1963. Spatial Models of Party Competition. *American Political Science Review* 57(2): 368–77.

Swindell, David, and Janet M. Kelly. 2000. Linking Citizen Satisfaction Data to Performance Measures: A Preliminary Evaluation. *Public Performance & Management Review* 24: 30–52.

Szymanski, David M., and David H. Henard. 2001. Customer Satisfaction: A Meta-Analysis of the Empirical Evidence. *Journal of the Academy of Marketing Science* 29(1): 16–35.

Thomassen, Jean Pierre, Marijke C. Leliveld, Steven Van de Walle, and Kees Ahaus. 2017. Compensating Citizens for Poor Service Delivery: Experimental Research in Public and Private Settings. *Public Administration* 95(4): 895–911 *.

Van Ryzin, Gregg G. 2004. Expectations, Performance, and Citizen Satisfaction with Urban Services. *Journal of Policy Analysis and Management* 23(3): 433–48 *.

———. 2006. Testing the Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government. *Journal of Public Administration Research and Theory* 16(4): 599–611 *.

———. 2013. An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction. *Journal of Policy Analysis and Management* 32(3): 597–614*.

Van Ryzin, Gregg G., and Etienne Charbonneau. 2010. Public Service Use and Perceived Performance: An Empirical Note on the Nature of the Relationship. *Public Administration* 88(2): 551–63.

Van Ryzin, Gregg G., and Stephen Immerwahr. 2007. Importance-Performance Analysis of Citizen Satisfaction Surveys. *Public Administration* 85(1): 215–26.

Viechtbauer, W. 2010. Conducting Meta-Analyses in R with the Metafor Package. *Journal of Statistical Software* 36(3): 1–48.

Vigoda, Eran. 2000. Are you Being Served? The Responsiveness of Public Administration to Citizens' Demands: An Empirical Examination in Israel. *Public Administration* 78(1): 165–91.

———. 2002. From Responsiveness to Collaboration: Governance, Citizens, and the Next Generation of Public Administration. *Public Administration Review* 62(5): 527–40.

Vigoda-Gadot, Eran. 2006. Citizens' Perceptions of Politics and Ethics in Public Administration: A Five-Year National Study of their Relationship to Satisfaction with Services, Trust in Governance, and Voice Orientations. *Journal of Public Administration Research and Theory* 17(2): 285–305.

Vigoda-Gadot, Eran, and Shlomo Mizrahi. 2006. Public Sector Management and the Democratic Ethos: A 5-Year Study of Key Relationships in Israel. *Journal of Public Administration Research and Theory* 18(1): 79–107.

Walker, Richard M., Gene A. Brewer, M. Jin Lee, Nicolai Petrovsky, and Arjen van Witteloostuijn. 2019. Best Practice Recommendations for Replicating Experiments in Public Administration. *Journal of Public Administration Research and Theory* 29(4): 609–26.

World Bank. 2018. Europe and Central Asia Development of EU Governance Indicators: Indicators of Citizen-Centric Public Service Delivery – Final Report. Report no: AUS0000222. Washington, DC: World Bank.

* indicates studies included in the meta-analysis