<u>Research Article</u>

Short and Sweet: Measuring Experiences of Administrative Burden

Sebastian Jilke[†], Martin Baekgaard[‡], Pamela Herd[†], Donald Moynihan[†]

Abstract: Emerging research on administrative burden has highlighted the need for survey measures that capture people's experience of government as onerous. Such indicators can connect research and practice, and fulfill government mandates to identify and reduce burdens. This study presents a measure of experienced administrative burden, based on a survey of social welfare users. Using psychometric scale development techniques and a split sample approach, we develop both a single-item and three-item scale that can be applied in user surveys for both research and practitioner purposes. We use predictive validity tests to assess both measures, finding that people are more likely to report burden when they have poorer health, lower education, experience short-term financial scarcity, are younger and have less program-specific experience. We also provide evidence to demonstrate how the measure captures burden distinct from the customer experience indicators used by the US federal government, while also being shorter and thus less burdensome for respondents.

Keywords: Administrative burden, customer experience, scale development, citizen-state interactions

Supplements: Open Materials

Introduction

Over the past decade, public administration and policy scholars have paid increasing attention to the administrative burdens that people encounter in public services (Herd and Moynihan 2022). This body of research has drawn from a variety of literatures, including but not limited to policy take-up, policy feedback, social policy studies, and red tape, to develop the concept, and provide empirical evidence on the creation of such burdens and their effects (Madsen, Mikkelsen and Moynihan 2022). Since the concept focuses on *individual experience,* some quantitative studies have sought to directly capture those experiences, but there is not, as yet, a standard, validated, subjective survey-based measure that captures people's experience of burdens.

Having a standard survey-based measure has a number of advantages. First, such a measure is highly relevant in a growing research area: it could serve both as a dependent variable in its own right, as well as an independent variable that predicts outcomes such as program take-up or policy feedback effects. Second, it would increase comparability across studies and thus help expand our common knowledge base. It would also make it possible to map and compare experiences across target groups, programs, and countries. Third, the development of a scale could help to better connect research and practice. As governments become more attentive to administrative burdens (see, for example, President Biden's Customer Experience agenda[1]), many

---

† McCourt School of Public Policy, Georgetown University
‡ Department of Political Science, Aarhus University
Address correspondence to Sebastian Jilke at sebastian.jilke@georgetown.edu.

---

[1] See Executive Order on Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government and revised guidance for the implementation of the Paperwork Reduction Act (Memorandum-22-10).

will mandate that agencies identify and reduce such burdens. Measures of user experiences with programs provide one simple and standardized way of doing so, providing a basis by which researchers can help government to address real-world problems.

This study seeks to develop such a measure of people's experience of administrative burden which can be used in multiple settings.

Principles for Measuring Burdens

There are multiple ways to make valid inferences about experiences. Behavioral responses to increases and decreases in state actions likely to result in burden is a common approach: for example, are you more or less likely to apply for a service if you are face shorter forms, clearer explanations, or informational nudges (e.g., Arbogast, Chorniy and Currie 2022; Barofsky et al 2022; Linos, Quan and Kirkman 2020; Moynihan, Gianella, Herd and Sutherland 2022)? This approach has the advantage of revealing behavioral responses without having to rely on subjective accounts. However, it does not directly measure the experience of burdens itself – rather the experience of burden is inferred based on a behavioral response to state action.

Researchers can also measure physiological responses to exposure to burdens (e.g., Hattke et al 2020). This has the benefit of relying on objective indicators, but is costly and limits the researcher to the lab in most cases. Another suggested approach is to use objective measures of experienced costs (Madsen, Mikkelsen, and Moynihan 2022). We can track, for example, what people might spend in an administrative encounter, how far they traveled, or how long they have to wait (e.g., Holt and Vonopal 2023). This works less well for harder-to-observe and more subjective costs, i.e., learning and especially psychological costs.

While these approaches are valid and valuable, subjective measures are sometimes required and offer some inherent advantages. For policy feedback processes, people's interpretation of their experiences matter, because that is what shapes their subsequent beliefs and behavior. In other cases, a behavioral response may be hard to observe, or a treatment that might change experienced burdens is inaccessible or unethical. It is also easier to embed items in user surveys of public services, enabling research partnerships with government on a broad scale.

To develop our approach we reviewed other efforts to measure burdens (Baekgaard et al. 2021; Bell et al 2023; de Bruijn 2021; Döring and Madsen 2022; Johnson and Kroll 2021; Madsen, Baekgaard and Kvist 2023; Thomsen, Baekgaard and Jensen 2020). While prior studies have measured the experience of burden in specific settings, wording items are often closely tied to aspects of the program that make it difficult to adapt them to other settings. These measures have also not been rigorously tested. Indeed, studies have repeatedly called for better measures of burden (e.g., Baekgaard and Tankink 2022; Bell et al 2023; Madsen, Mikkelsen and Moynihan 2022). We also engaged with officials interested in burdens in government and non-profit sectors. This helped us develop concept-specific principles for measurement beyond standard measurement concerns.

Measures should reflect experiences: The administrative burden framework focuses on individual experience (Madsen, Mikkelsen and Moynihan 2022). Thus, questions about burdens should be anchored in experience. The further the questions are from direct experience (e.g., questions regarding how burdensome an organization is, rather than a specific interaction, or being asked to judge the experiences of others) the greater the error. For example, people may systematically overestimate their capacity to deal with hassles, and underestimate how much burdens affect them or others. Questions that ask people how they would respond to a hypothetical compliance cost, such as compiling documentation, will likely generate responses that are inconsistent with behavior. This makes survey vignette experiments, where subjects speculate about how they or others might respond to a hypothetical situation, unreliable.

The more specific the experience, the better, generating more grounded recall and minimizing halo effects. Thus, researchers should purposely sample a cohort that experienced a particular program or task, and ask them about their experiences (Baekgaard et al. 2021; Bell et al. 2022). Subjects could also report on the experience of a specific task created by the researcher, such as filling out a form (Hattke et al 2020). For practitioners, the easiest opportunity is to embed burden items in user surveys at the completion of an administrative process.

Capture multiple dimensions of burdens: While there are other valid ways to conceptualize administrative burden (most notably Heinrich 2016), we develop items consistent with the conceptualization of burdens as a series of learning, compliance, and psychological costs (Moynihan, Herd and Harvey 2015). We develop a specific item for each of these categories, as well as a single summary item. Learning costs incorporate both the search for relevant information (information may be widely available, but understanding what is salient for you is another question) and understanding the content of the information. For compliance costs, the values of effort and time are frequently evoked. There may be other, more specific aspects of compliance costs – such as money spent, or distance traveled – but it seems reasonable to assume these would correlate with people's understanding of time and effort. Psychological costs include a sense of shame or stigma, a sense of loss of personal autonomy, stress or frustration with the process and associated outcomes.

In some settings, some aspects of the concept will not be relevant. For example, consider a lab experiment where subjects are asked to complete a form that induces compliance costs and possibly psychological costs. If the researcher gave them the form to complete, questions about searching for forms would be irrelevant. Thus, we warn against reifying any battery of items to the point that items seek to measure aspects of burdens that cannot reasonably be assumed to exist in that particular setting.

Measures should be adaptable to different contexts and experiences: Administrative burdens arise in many places. A global measure of the concept is less valid if the context is unclear to the subject. In other words, if administrative burdens are the experience of policy implementation as onerous, researchers need to specify which aspect of the policy implementation process or specific interaction they are asking respondents about.

Thus, the emphasis on experience and specificity also implies the need for some measure of flexibility in the design of survey items, at least to the point that the wording can make reference to a specific process, program or experience that would otherwise be unclear. Indeed, many of the existing items we reviewed were very much tied to the specific process or program they studied. Thus, our goal is to develop a set of items that could be applied, with minor modification, to very different interactions, such as providing documentation, completing forms, waiting for service, participating in compulsory meetings etc.

The need for flexibility and the recognition that not all sub-dimensions of burdens apply to every situation suggests that we are not so much creating a single gold-standard set of measures, as much as we are developing a tool that may need to be tweaked depending on the context. It may also mean that our scale does not apply well to all dimensions. Because much of the existing work focuses on applications for welfare benefits, they are less easily applicable to situations that look different, such as use of those benefits (for instance, the concept of redemption costs, identified by Barnes 2021).

Measuring burden should not be burdensome: A longer scale may better capture nuance, but it comes with costs. The more items in a scale, the more burdensome it is to participants. The consequences are irritated users, low response rates and less relevance for surveys with pre-defined limitations on the number of questions. We thus sought to develop a short but adaptable scale, while leaving open the possibility to develop a longer scale at a later point in time.

Data

Based on the principles above, we determined to test a short scale of actual users of a major social program. We chose the Supplemental Nutrition Assistance Program (SNAP), a benefit program used by more than 41 million beneficiaries in the U.S. in 2021 (CBPP 2022).

The wording of items sought to be specific enough to anchor the respondents' recall to a particular program (SNAP) and aspect of the administrative process of that program (applying for and renewing benefits), but generic enough that the wording of questions could be adapted for other programs with minimal effort. We specified what we meant by learning and compliance costs in the question framing. Because of the multi-dimensional nature of psychological costs, we tested a variety of items, including feelings of frustration, being in control, stress and a sense of being respected (see table 1).

All items were measured using a 5-points Likert scale. Item order was randomized and all questions had the same introductory text. The response categories for the overall item, learning and compliance costs were: very difficult, somewhat difficult, neither easy nor difficult, somewhat easy, very easy. The response categories for the psychological cost items were extremely, very, moderately, slightly or not at all. The psychological cost items were randomized as either being presented as a 4-items battery or 4 single items that were presented on their own. This allowed us to find out which of the items had the best measurement properties for being included in the final scale.

Table 1: Initial list of survey items

Question wording *We want to hear about your most recent experiences with SNAP, also known as food stamps. This includes applying for AND/OR renewing your benefits. Please think about your most recent experience with the program when you respond to the question.*	Cost category	Included in final instrument
Item #1: How would you describe this experience overall?	*Overall*	Yes (single-item)
Item #2: How difficult was the process of finding information about the program, such as how to apply or what you needed to do to renew your benefit?	*Learning*	Yes (3-items scale)
Item #3: How was the process of filling out the paperwork, providing proof of eligibility (such as pay stubs, proof of residence, birth certificates, etc.), and/or attending interviews?	*Compliance*	Yes (3-items scale)
Item #4: Please describe how you felt during these experiences? *Frustrated*	*Psychological*	Yes (3-items scale)
Item #5: Please describe how you felt during these experiences? *In control*	*Psychological*	No
Item #6: Please describe how you felt during these experiences? *Stressed*	*Psychological*	No
Item #7: Please describe how you felt during these experiences? *Respected*	*Psychological*	No

We conducted two initial pilot tests that resulted in minor changes to wording of the response items before launching the full survey in December 2022. The survey, as well as the two rounds of pilot testing, were conducted with SNAP users, who were screened based on their income and response to questions about benefits. The survey was pre-registered and subject to IRB approval at Georgetown University, with data collected via Cloud Research.

The survey included an attention check question right after the informed consent form. Respondents who failed this check were directed to the end of the survey. We also collected measures of respondents' beliefs, including trust in government, perceptions of corruption, and political ideology. We asked individuals about their experience of short-term financial scarcity, health, duration of use of SNAP benefits, age, gender, education, race, and income. Financial scarcity is measured based on respondents indicating how often (i.e., never, rarely, sometimes, fairly often, very often) they 1) were preoccupied with thoughts about their personal financial situation; 2) thought about future expenses, some of which may be unexpected; 3) were worried about having enough money to make ends meet; and 4) were troubled about coping with ordinary bills

(adapted from Carvalho et al. 2016). The higher the score, the higher their situational financial stress. Self-rated health is a standard global measure indicating whether the respondent considered their health excellent, very good, good, fair or poor. The higher the score the better one's health. The full survey instrument can be found in the online appendix.

Method

We collected 3,819 valid responses of which 3,817 responded to the full set of administrative burden items. The single psychological cost items were presented to between 815 to 715 respondents each, and the 4-item battery was completed by 749 respondents. Our goal was to have more than 600 respondents for each item-specific sample – which allowed us to use a split sample approach with more than 300 respondents per item-specific split sample. This is significantly above the threshold of recommendations for factor analytical modeling (e.g., Kline 2005). On this basis, we set a reproducible seed and randomly split the sample to conduct exploratory factor analysis on one split sample and confirmatory factor analysis on the other. Indeed, the advantage of the split sample approach is that it allows us to conduct exploratory and confirmatory tests using two random sub-samples derived from the same sample – which is considered best practice in scale development (DeVellis 2016). In a first set of models, we ran exploratory factor analyses using oblique rotation, which is the appropriate rotation method for correlated items. As a second step, we constructed confirmatory factor models to validate the exploratory factor structure. In addition to factor analytical approaches, we conducted tests of predictive validity.

Results

Scale development

First, we present descriptive statistics for all administrative burden items. Table 2 shows the overall means and corresponding standard deviations. There is substantial variability in people's experiences of burdens, indicating that the items indeed offer useful variation to explore further.

Table 2: Descriptive statistics

Item	Mean (Standard deviation)	Sample size
#1: Overall	3.553 (1.223)	3,819
#2: Learning costs	3.729 (1.178)	3,819
#3: Compliance costs	3.395 (1.178)	3,819
#4: Psychological costs (frustrated – battery)	3.713 (1.334)	749
#4: Psychological costs (frustrated – stand-alone)	3.740 (1.208)	749
#5: Psychological costs (in control – battery)	2.982 (1.273)	749
#5: Psychological costs (in control – stand-alone)	2.681 (1.196)	715
#6: Psychological costs (stressed – battery)	3.454 (1.337)	749
#6: Psychological costs (stressed – stand-alone)	3.404 (1.326)	815
#7: Psychological costs (respected – battery)	2.680 (1.224)	749
#7: Psychological costs (respected – stand-alone)	2.649 (1.133)	790

After that, we examined correlations between all items. All items are moderately strong correlated with each other, with the correlation coefficients ranging between 0.763 and 0.317. The psychological cost item that correlated most strongly with learning (item #2) (0.476) and compliance (item #3) costs (0.547) was item #4 (frustrated).

As a next step, we took a random split sample and conducted a series of exploratory factor analyses (see table 3). Model #1 includes all cost items, including the 4-item psychological cost battery. It produces three factors with Eigenvalues greater than 1. Factors 1 and 3 are two distinct aspects of psychological costs, while factor 2 is a combination of learning (item #2) and compliance (item #3) costs. Since the goal of our scale development exercise is a single cohesive scale of administrative burdens, we continued testing different

variants of the scale using each of the 4 items separately (models #2-5). Out of these models, model #2 has the best fit in terms of Eigenvalue and Cronbach's alpha values. Neither of its factor loadings are below 0.6.

We also examined the scale's configural equivalence by re-estimating Model #2 for men and women separately. The respective factor structure performs well for both sub-samples, although slightly better for men. This is evidence in favor of the configural equivalence of our 3-item scale (model #2).

Table 3: Results Exploratory Factor Analysis

Model #	Items included	Eigenvalue	Factor loadings	Cronbach's alpha	Sample size
#1	#2: Learning #3: Compliance #4: Psychological-Frustrated #5: Psychological-In-control #6: Psychological-Stressed #7: Psychological-Respected	Factor 1: 2.574 Factor 2: 2.381 Factor 3: 2.232	Factor 1 item #4: 0.769 item #6: 0.825 Factor 2 Item #2: 0.746 Item #3: 0.664 Factor 3 Item #5: 0.665 Item #7: 0.652	0.775	367
#2	#2: Learning #3: Compliance #4: Psychological-Frustrated	Factor 1: 1.486	Item #2: 0.722 Item #3: 0.735 Item #4: 0.652	0.861	360
#3	#2: Learning #3: Compliance #4: Psychological-In-control	Factor 1: 1.363	Item #2: 0.716 Item #3: 0.752 Item #5: -0.534	0.822	380
#4	#2: Learning #3: Compliance #4: Psychological-Stressed	Factor 1: 1.413	Item #2: 0.726 Item #3: 0.753 Item #6: 0.565	0.850	413
#5	#2: Learning #3: Compliance #4: Psychological-Respected	Factor 1: 1.285	Item #2: 0.703 Item #3: 0.737 Item #7: -0.497	0.797	388

We then tested model #2 using a confirmatory factor analytical (CFA) framework. The following standardized factor coefficient were recovered (standard errors in parentheses):

- Learning costs (item #2): 0.773 (0.030);
- Compliance costs (item #3) 0.848 (0.028):
- Psychological costs (item #4; frustrated): 0.719 (0.032).

CFA fit is determined by a series of global fit indices, which all pointed towards an excellent model fit (RMSEA: 0.000; CFI: 1.000; TLI: 1.000; SRMR: 0.000), thus confirming our 3-items factor structure. The likelihood ratio test statistic (i.e., χ^2) versus the baseline model is 425.563 (p<0.000) with 3 degrees of freedom, and the model's modification index did not suggest any changes to the factor structure that would improve its likelihood ratio by more than 3.841 (for 1 degree of freedom) which confirms the existing factor structure of model #2.

We also examined whether model #2 has a better fit with and without the general burden item (item #1), finding that its inclusion leads to a decrease in model fit on all respective fit indices (RMSEA: 0.066; CFI: 0.996; TLI: 0.987; SRMR: 0.014). When directly comparing the log-likelihood of both models using a log likelihood-ratio test, the inclusion of the overall item does not lead to a statistically significant increase in

model fit (χ^2=-906.77 (3 df); p>0.100). Thus, we conclude that model #2 best fits the data, and that item #1 should not be part of the multi-item scale.

Predictive validity and relationship with single-item measure
We next examined the validity of the single-item administrative burden measure and how it compares to the 3-items solution we discussed above. To do so, we construct a sum scale of administrative burden (Min 1, Max 5) and correlate it with the single-item measure, yielding a 0.764 correlation coefficient (p<0.000).

We next completed a set of predictive validity tests. Predictive validity holds if measures of administrative burdens exhibit similar relationships with concepts of importance as found in the literature. Prior work suggests that lower human capital, health or the experience of short-term financial scarcity will increase the experience of burden (Christensen et al. 2020; Bell et al. 2022; Madsen, Baekgaard and Kvist 2023). We therefore examined how well both measures are correlated using a simple linear regression model where experiences of administrative burdens are predicted by health, self-reported financial scarcity, age, current gender, education and years of SNAP use[2].

The standardized regression coefficients for both models are similar, as are their respective p-values (see table 4). As anticipated, we find that poorer health, higher levels of perceived financial scarcity, and lower levels of education are correlated with higher levels of burden. We also find that people who have used SNAP for longer periods report reduced experiences of burden. This may reflect familiarity and learned knowledge about how to navigate program demands, reducing their perception of them as burdensome. Younger people report greater experiences of burden, which may reflect lower stocks of administrative literacy (Döring and Madsen 2022), or generational differences in terms of expectations for what constitutes reasonable burdens. We do not find differences between men and women, but transgender people report lower burden, though this is a response category that is based on only 34 observations. Model 1 explains about 9% of the total variance in people's experiences of administrative burdens while model 2 explains about 12%.

Table 4: Relationships between administrative burdens and health, financial scarcity, education, program duration, gender, and age

Independent Variable	Model #1 Burden overall single-item	Model #2 Burden 3-items scale
Self-reported health	-0.089 (0.026)	-0.090 (0.019)
Perceived financial scarcity	0.112 (0.003)	0.136 (0.000)
Age	-0.114 (0.003)	-0.188 (0.000)
Current gender (reference: woman)		
Man	0.013 (0.719)	0.033 (0.364)
Transgender	-0.057 (0.042)	0.006 (0.878)
Education (reference: less than high school)		
High school	-0.038 (0.640)	0.004 (0.953)
Associate/ junior college	0.035 (0.626)	0.034 (0.568)
Bachelor's	0.147 (0.011)	0.171 (0.000)
Graduate	0.028 (0.552)	0.069 (0.000)

---

[2] Note that 74 respondents indicated that they did not use SNAP; we included them in all analyses but excluding them does not substantially alter any results reported in this study.

SNAP experience (reference: less than 1 year)		
Not at all	0.051 (0.269)	-0.010 (0.792)
1-3 years	-0.086 (0.055)	-0.105 (0.021)
4-6 years	-0.078 (0.082)	-0.082 (0.064)
7-10 years	-0.023 (0.596)	-0.040 (0.327)
10+ years	-0.152 (0.000)	-0.170 (0.000)
Sample size	740	740
R-squared	0.086	0.115

Note: Standardized regression coefficients are reported with respective p-values in parenthesis.

These findings offer broad support for a three-items scale that captures aspects of learning, compliance, and psychological costs, but also shows that the single overall item works almost as well, and may be easier to use in surveys where space is a consideration. Both items are also correlated with variables in ways that are consistent with previous administrative burden theory and empirical findings.

Comparing both measures to a federal customer experience measure
In a final series of tests, we also compared both measures of administrative burden to the federal customer experience scale (or federal CX scale). The Office of Management and Budget recommends that government agencies use this scale, including as part of new reforms in the Biden administration that are framed broadly as customer experience, but explicitly direct attention to identifying and measuring burdens. But the CX scale was adapted from the private sector, and not developed to directly measure burdens[3]. Instead, it includes 7 items including service users' perceived satisfaction, need, ease, timeliness, fairness and quality of the interaction itself. The origins and nature of the CX scale therefore suggests it may not directly capture the experience of burdens, pointing to the value of examining the metrics we develop here. On the other hand, if the two scales are functionally equivalent, there is little need for items we propose.

A simple look at correlations suggests that the CX and burden scales capture different phenomena. As a first comparison, we report correlations between the single item measure of burdens, the multi-item sum scale with a sum scale of all 7 measures – all items are measured using 5-points Likert scales. Our measures of administrative burden are moderately correlated with the federal CX sum scale, with correlations ranging from -0.478 to -0.539 (table 5). When examining the full set of CX items instead of the sum scale, correlation coefficients range from -0.233 to -0.640 (see table 5). Indeed, more than 65% of the correlation coefficients exhibit a low correlation (i.e., below 0.5). Thus, the two scales are correlated, as one would expect, but not highly so, suggesting they capture different aspects of people's interactions. The federal CX measure captures people's broader customer experience including aspects that are correlated with different burden costs (like ease of completing an application), while the burden scale addresses those costs directly. We argue that the CX measures are, in parts, the result of people's experiences with administrative burdens and more alike measures of service quality and satisfaction.

Table 5: Correlations between Administrative Burden and Federal CX measures

	Burden overall	Burden scale	Burden item #2	Burden item #3	Burden item #4
CX scale	-0.576	-0.623	0.478	0.539	0.539

---

[3] See, for example, Section 280 of the governmentwide Circular A-11 guidance, Managing Customer Experience and Service Delivery https://www.performance.gov/cx/assets/files/2022-OMB-Circular-A11-Section-280.pdf. The inclusion of this scale means it can be used without seeking prior permission from OMB, unlike other items, thus facilitating its adoption by agencies.

CX item 1 *I am satisfied with the service I received from SNAP.*	-0.512	-0.560	-0.419	-0.479	-0.500
CX item 2 *This interaction increased my confidence in SNAP.*	-0.503	-0.518	-0.399	-0.454	-0.451
CX item 3 *My need was addressed.*	-0.467	-0.493	-0.386	-0.431	-0.428
CX item 4 *It was easy to complete what I needed to do.*	-0.569	-0.640	-0.500	-0.573	0.525
CX item 5 *It took a reasonable amount of time to do what I needed to do.*	-0.311	-0.271	-0.252	-0.279	-0.233
CX item 6 *Employees I interacted with were helpful.*	-0.453	-0.513	-0.388	-0.428	-0.445
CX item 7 *I was treated fairly.*	-0.484	-0.535	-0.398	-0.443	-0.473

Note: all entries are statistically significant at p<0.000.

To the degree these measures are intended to capture potential heterogeneity in responses in citizen-state interactions, which can serve as sources of inequality, the shorter burden scale appears to perform better. We re-ran the predictive validity tests we did on the administrative burden single- and multi-items scales with the federal CX scale using a similar model as we did in table 4 but with the federal CX scale as dependent variable (see table 6). The final model exhibits a substantially smaller R-squared than the 3-items burden scale, and even the single-item overall burden measure. The three-item burden scale explains twice the variation in administrative burden than the seven-item CX scale, and the single-item scale explains one-and-a-half the variation. While regression coefficients and respective p-values lead to substantially similar conclusions than to those in table 6, they reveal some substantial differences in effect sizes. For instance, the relationship between the federal CX measure and health is more than 40% stronger than that of health and our 3-item burden measure (95% confidence intervals of health from table 5 are -0.053 – -0.128 and table 6 are 0.097 – 0.159). Given the low to moderate, yet statistically significant correlation between CX and our burden measures, this is not surprising.

Table 6: Relationship between the federal CX measure and health, financial scarcity, education, program duration, and gender

Variable	Model #3 Federal CX measure
Self-reported health	0.128 (0.000)
Scarcity	-0.112 (0.000)
Age	0.061 (0.000)
Current gender (reference: woman)	0.019 (0.247)
Man	-0.027 (0.138)
Transgender	
Education (reference: less than high school)	
High school	0.029 (0.412)
Associate/ junior college	0.003 (0.929)
Bachelor's	-0.041 (0.096)
Graduate	0.001 (0.964)
SNAP use (reference: less than 1 year)	
Not at all	-0.087 (0.000)
1-3 years	0.029 (0.156)
4-6 years	0.029 (0.121)
7-10 years	0.039 (0.032)
10+ years	0.069 (0.000)
Sample size	3,760
R-squared	0.055

Note: Standardized regression coefficients are reported with respective p-values in parenthesis.

Conclusion

The accumulation of scientific knowledge depends upon well-defined concepts that capture important phenomena, as well as the ability to measure those concepts in a way that is feasible and allows empirical evidence to be broadly comparable.

The scale developed here represents the first step, rather than the last word, on measurement of people's experiences of administrative burden. It comes with some caveats that should encourage future work. The population of respondents represented an online panel, and thus will have some basic digital literacy. The validity of the scale across different countries, populations, programs, types of interactions, and field settings has to be further explored. The items proposed point the way for such exploration.

The items are structured to be adaptable to a wide (though not exhaustive) array of citizen-state encounters and we believe they are well-suited to be applied to other social programs, including those that require more direct interactions with frontline personnel. Yet, more specific items about direct encounters with frontline workers could be added. However, as we discuss below, this will be at the expense of the scale brevity and add additional response burden. Indeed, the scale's brevity ensures that it is not burdensome for participants, while also making it feasible for digital or phone-based user and research surveys. This includes

surveys that examine which processes people experience as more onerous, and whether experimental testing of alternative processes improves experiences. It also allows for cross-programmatic rankings of administrative processes, to determine which one's users judge to be more burdensome. Our focus on brevity also means that we do not focus on all aspects of every cost. Future work can do this better, using a variety of approaches that seek both to measure different aspects of costs, as well as experiment with conditions that might trigger those costs. For example, we center our measure of learning costs on finding information. Future work might also measure understanding and comprehension as other aspects of learning costs, or exploit variation in complexity of information. Physiological measures could also seek to capture cognitive demands placed on subjects. Similarly, while we measure four different types of psychological costs, future work could explore others, or exploit conditions likely to trigger such costs, such as the experience of uncertainty as a trigger.

Our goal of developing a scale about burdens that is short and not burdensome for respondents also prevents us from capturing a more detailed level of burdens that people experience with a given program. Asking respondents about their 'most recent experience' may include applying for or renewing benefits – which are two qualitatively different interactions that come along with their own set of burdens. Our goal was to capture people's 'overall' experience, which captures both aspects. A more detailed scale would be able to get at this level of nuance, however, very likely at the expense of response rates and implementation feasible by practitioners. Still a burden scale like ours provides usable guidance about the levels of overall burdens of a specific program. Future work could also expand measurement to reflect other aspects, such as comprehension, interpretation, or understanding. This could be pursued via traditional survey work, but also via physiological indicators that are related to cognitive effort.

The administrative burden framework has provided a community of both practitioners and scholars a common language to describe people's onerous experiences interacting with public services. This scale can provide a common measurement tool to help that community to work together in assessing our effectiveness in actually reducing those burdens.

References

Arbogast, I., Chorniy, A., & Currie, J. (2022). *Administrative Burdens and Child Medicaid Enrollments* (No. w30580). National Bureau of Economic Research.

Baekgaard, M., Mikkelsen, K. S., Madsen, J. K., & Christensen, J. (2021). Reducing compliance demands in government benefit programs improves the psychological well-being of target group members. *Journal of Public Administration Research and Theory*, *31*(4), 806-821.

Baekgaard, M., & Tankink, T. (2022). Administrative burden: Untangling a bowl of conceptual spaghetti. *Perspectives on Public Management and Governance*, *5*(1), 16-21.

Barnes, C. Y. (2021). "It takes a while to get used to": The costs of redeeming public benefits. *Journal of Public Administration Research and Theory*, *31*(2), 295-310.

Barofsky, J., Vargas, A., Rodriguez, D., Matos, E., & Barrows, A. (2021). Putting out the 'unwelcome mat:' The Announced Public Charge Rule reduced safety net enrollment among exempt noncitizens. *Journal of Behavioral Public Administration*, *4*(2).

Bell, E., Christensen, J., Herd, P., and Moynihan, D. (2023). Health in citizen-State Interactions: How Physical and Mental Health Problems Shape Experiences of Administrative Burden and Reduce take-Up. *Public Administration Review 83*(2): 385–400.

Center on Budget and Policy Priorities. (2002). Policy Basics: The Supplemental Nutrition Assistance Program (SNAP). https://www.cbpp.org/research/food-assistance/the-supplemental-nutrition-assistance-program-snap

Carvalho, L. S., Meier, S., & Wang, S. W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American economic review*, *106*(2), 260-284.

Christensen, J., Aarøe, L., Baekgaard, M., Herd, P., & Moynihan, D. P. (2020). Human capital and administrative burden: The role of cognitive resources in citizen-state interactions. *Public Administration Review*, *80*(1), 127-136.

de Bruijn, E.-J. (2021). The Financial and Administrative Burden of Financially Vulnerable Households. https://doi.org/10.18174/ 549093

DeVellis, R. F. (2018). *Scale Development: Theory and Applications*. (Fourth Edition). Sage Publications

Döring, Matthias and Jonas Krogh Madsen. 2022. Mitigating Psychological Costs–The Role of Citizens'

Administrative Literacy and Social Capital. Public Administration Review. 82(4): 671-681.

Hattke, F., Hensel, D., & Kalucza, J. (2020). Emotional responses to bureaucratic red tape. *Public Administration Review*, *80*(1), 53-63.

Heinrich, Carolyn J. "The bite of administrative burden: A theoretical and empirical investigation." *Journal of Public Administration Research and Theory* 26, no. 3 (2016): 403-420.

Herd, P., & Moynihan, D. (2022). Behavior and Burdens: Introduction to the Symposium on Behavioral Implications of Administrative Burden. *Journal of Behavioral Public Administration*, *5*(1).

Holt, Stephen B., and Katie Vinopal. "Examining inequality in the time cost of waiting." *Nature Human Behaviour* (2023): 1-11.

Johnson, D., & Kroll, A. (2021). What makes us tolerant of administrative burden? Race, representation, and identity. *Journal of Behavioral Public Administration*, *4*(1).

Kline, R. (2005). Principles and Practice of Structural Equation modeling. Third Edition. Guilford Publications

Linos, E., Quan, L. T., & Kirkman, E. (2020). Nudging early reduces administrative burden: Three field experiments to improve code enforcement. *Journal of Policy Analysis and Management*, *39*(1), 243-265.

Madsen, J. K., Baekgaard, M., & Kvist, J. (2023). Scarcity and the Mindsets of Social Welfare Recipients: Evidence from a Field Experiment. *Journal of Public Administration Research and Theory*. 33(4): 675-687.

Madsen, J. K., Mikkelsen, K. S., & Moynihan, D. P. (2022). Burdens, sludge, ordeals, red tape, oh my!: A user's guide to the study of frictions. *Public Administration, 100*(2), 375-393.

Moynihan, D., Giannella, E., Herd, P., & Sutherland, J. (2022). Matching to categories: Learning and compliance costs in administrative processes. *Journal of Public Administration Research and Theory*, *32*(4), 750-764.

Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*, *25*(1), 43-69.

Thomsen, M. K., Baekgaard, M., & Jensen, U. T. (2020). The psychological costs of citizen coproduction. *Journal of Public Administration Research and Theory*, *30*(4), 656-673.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.