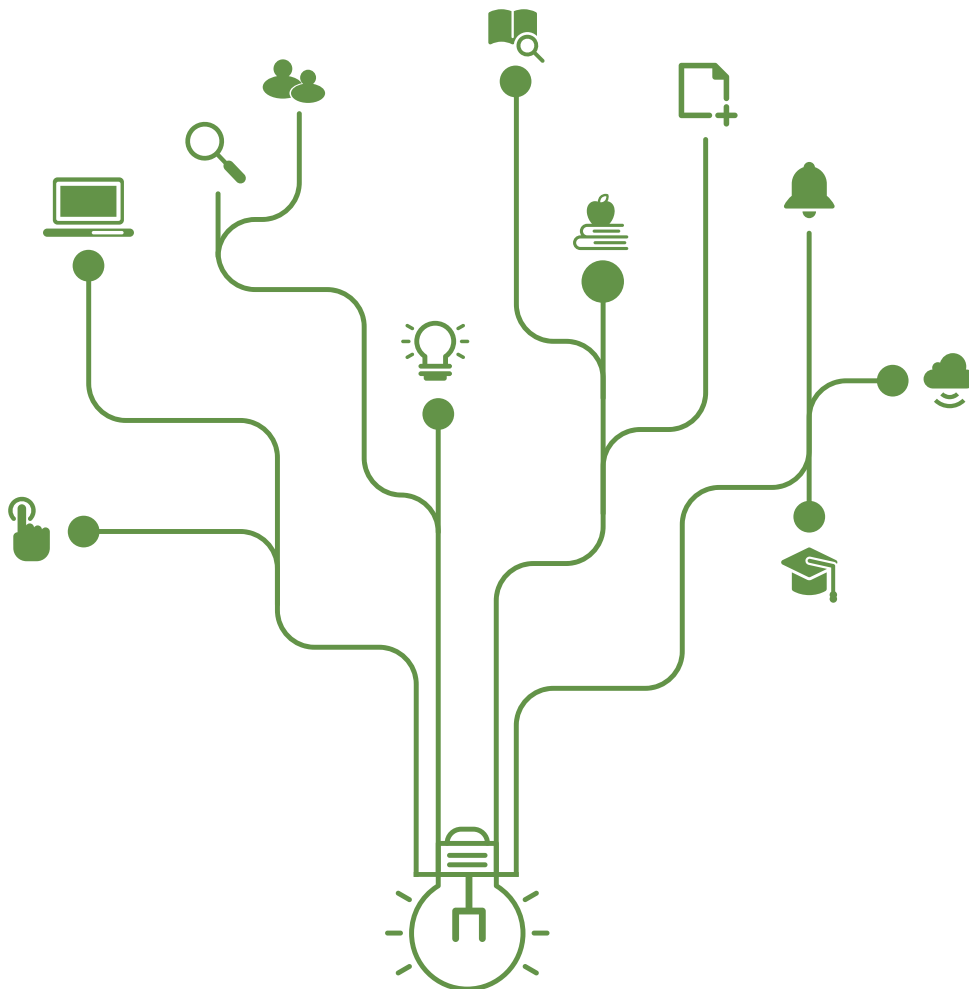


# A Systematic Review of Experimental Approaches on Public Service Motivation

*Kee Hoon Chung (KDI School of Public Policy and Management)*

*Inbok Rhee (KDI School of Public Policy and Management)*

*Cheol Liu (KDI School of Public Policy and Management)*



# **A Systematic Review of Experimental Approaches on Public Service**

## **Motivation**

### **Abstract**

Recent studies on public service motivation (PSM) have increasingly adopted experimental approaches. This article provides a systematic review of the existing experimental studies on PSM and suggests guidance for future research. We assess the existing literature based on two criteria that differentiate experimental methods from observational ones: whether PSM is measured through an experiment and whether the research design operates treatment intervention for PSM. After reviewing a sample of 26 published studies, we find that only one study met both criteria. Most studies measure PSM through a self-reported survey and lack treatment effects, making it difficult to differentiate them from observational studies. Furthermore, external validity remains a concern, as most studies use students as proxies for civil servants or focus only on Western states. We conclude that experimental studies on PSM remain at a nascent stage with much room for improvement, especially the experimental design.

### **Evidence for Practice**

- Experimental studies on PSM have high value for scholars and practitioners, as they can provide better measurements and causal inference.
- Experimental techniques can be used to study both the effects and determinants of PSM;

so far, the latter is underresearched.

- When designing and presenting an experimental study of PSM, experimental treatment should be clearly specified in relation to the outcome variables of interest.
- Using experimental techniques for sensitive survey questions can be an easy fix to mitigate social desirability bias.
- Expanding the scope of the experimental PSM research to go beyond the small number of civil servants or student samples and/or the Western country contexts will help increase the external validity and replicability of PSM research.

Since Perry and Wise (1990) first published an essay on public service motivation (PSM)— “individual’s predisposition to respond to motives grounded primarily or uniquely in public institutions and organizations (pp. 368)” –it has been one of the most popular research topics in the studies of public administration (Perry, Hondeghem, and Wise 2010; Ritz, Brewer, and Neumann 2016). Subsequent academic research on PSM has mainly revolved around three key issues. The first concerns the impact of PSM—whether and how greater PSM affects job satisfaction, choice of occupation, job performance, and commitment, to name a few (Ritz, Brewer, and Neumann 2016). The second concerns determinants of PSM – which investigates factors positively or negatively affecting PSM. The third concerns how to measure PSM—namely, capturing PSM’s theoretical construct.

Over the last decade, studies in PSM have increasingly sought to address these issues through experimental approaches (Jilke, Van de Walle, and Kim 2016). Since Perry (2012) called for more experimental studies in the field of public administration, there has been a greater recognition that we need to pay more careful attention to causal identification. These trends are displayed in figure 1. In total, we identified 26 studies related to PSM and experiments.

[Insert Figure 1 about here]

Despite the growing popularity of experimental approaches to PSM, relatively little research has systematically reviewed these studies. On the one hand, much effort has been made to conduct a systematic review of non-experimental approaches to PSM, with the most recent from Ritz, Brewer, and Neumann (2016). On the other hand, some efforts have been made in recent

years to review experimental approaches in the field of public administration (Bouwman and Grimmelikhuijsen 2016; Jilke, Van de Walle, and Kim 2016). However, broader research to integrate the two sides in order to conduct a systematic review of experimental studies on PSM has not yet been forthcoming

This article reviews existing studies that leverage experimental approaches in the study of PSM. Specifically, we assess the current state of experimental research on PSM, discuss their strengths and weaknesses, and offer recommendations for improving experimental approaches on PSM for the future.

Some of the key findings this article offers are as follows. First, the causal effects of PSM remain significantly underrepresented through experimental approaches. Most experimental studies related to PSM utilize experiments to measure the outcome of interest, not PSM. Moreover, most use an observational approach—a self-reported survey—to measure PSM; only two studies measure PSM through experiments. Furthermore, many studies do not capture PSM’s causal effects through a treatment intervention, but rather focus on measuring a variable of interest through an experiment. Surprisingly, only one study measures PSM via experiment and captures the treatment effect of PSM.

Second, most experimental studies on PSM utilize samples from specific regions, primarily Western countries, and rely on students as proxies for public servants.

Third, only two experimental studies examine the determinants of PSM. Both experiments do not yield positive and significant results, suggesting that more experiments are needed to verify the findings of the observational studies.

Our findings suggest that experimental research on PSM is at a nascent stage with room

for improvement. In the discussion and conclusion, we further discuss the limitations of these studies in detail and elaborate on recommendations, recommending improving the research design of future experimental studies.

This article begins with an overview of PSM's experimental studies and reviews them by three clusters: 1) Studies examining effects of PSM on various outcomes of interest 2) Studies examining determinants of PSM 3) Studies on measuring PSM. We then discuss their strengths and weaknesses in the context of their internal and external validity.

## **STATE OF EXPERIMENTAL RESEARCH ON PSM**

Prior works on PSM have primarily centered around observational studies (Ritz, Brewer, and Neumann 2016). Although observational studies have their strengths, their primary weakness is the threat of not controlling for confounding biases. For example, an observational study such as a self-reported survey may find that an individual with high PSM is more patient and perseveres during difficult times. However, it is also possible that some other determinants may confound this effect. In this example, the individual with high PSM may be more religious than those with low PSM, thus more tolerant and patient. Even if a respondent lacks such religiosity, he may still be prone to social desirability bias. The bias occurs because the respondent reacts to the survey by conforming to their societal norms rather than responding honestly. If the respondent assumes that PSM is a socially desirable value, the observational measure will capture the socially desirable bias with PSM.

Such an observational measure may suffer from measurement error since it is not capturing the intended concept. As a result, studies using such a measure to examine a causal effect would largely be unreliable since the variable of interest correlates with the error term. That is, the internal

validity of these measures would seriously be threatened.

Experimental approaches can complement observational studies by resolving these problems. Experimental research introduces an intervention or a treatment effect in a randomized controlled trial to induce a response from the treatment group but not from a controlled group. This approach is considered the gold standard for strengthening internal validity by generating exogenous measures not correlated with the error term (Jilke, Van de Walle, and Kim 2016). In other words, this approach generates reliable evidence compared to observational studies.

Although experimental approaches may have multiple criteria, we offer two conditions that meet the gold standard of minimizing bias and confounding factors. First and most obviously, an experiment must have a treatment that affects the explanatory variable of interest but not the outcome variable. It is also essential that the experiment takes place as a randomized trial. Both control and treatment groups are randomly assigned to assume that their characteristics are equal (Rajasekar and Kumar 2019). Second, an experiment must measure the variable of interest through an experimental approach rather than an observational approach. Should an experimental study not meet both criteria, it would be difficult to argue that such a study has established a reliable causal effect and thus may not be differentiated from an observational study.

Based on these criteria, this article reviews existing articles using experiments on PSM. To systematically review these studies, we first searched the Web of Science database for articles published at journals of the Social Science Citation Index (SSCI), which have the phrase “public service motivation” and “experiment” in their main text. The search, conducted on June 30, 2020, yielded 48 publications indexed as SSCI.<sup>1</sup> We manually inspected these publications with the following two criteria: (1) whether the main methodological approach of a given article is experimental research (i.e. involving a random assignment of treatment); and (2) whether the study

considers PSM as the main topic of the research.. This process yielded 26 articles, published since 2012, that conduct experimental research related to PSM, and which are displayed chronologically by type of experiments in table 1. We provide some statistics of the articles before we go further into their substantive details.

[Insert Table 1 about here]

These articles have mostly been published in journals with high impact factors in the field of public administration. The *Public Administration Review* (PAR) tops the list with six articles, followed by the *International Public Management Journal* (IPMJ) with five articles, and the *Journal of Public Administration Research and Theory* (JPART) with four. Only one article was published in a non-public administration journal—the *Social Behavior and Personality* (SBP).

Table 1 also shows information for the location of the targeted samples. The United States tops the list with seven experimental studies, followed by four studies from the Netherlands and Italy, three studies from Germany, and two studies from Denmark.

Table 1 classifies these articles by the three types of experiments: survey, lab, and field experiments. Of the three types, survey experiments have been the most popular method with 13 articles, accounting for more than half of the total. Lab experiments rank second with eight articles, followed by field experiments with five articles. Arguably, field experiments require an immense amount of resources. We focus our review based on the three types of experiments.

Further, table 1 categorizes the articles based on the role of the PSM variable within the articles. Most existing PSM studies examine PSM's effects on various outcomes of interest, as 12 studies treat PSM as an independent variable. The number increases to 18 when we include



moderating variables and increases to 20 if we include PSM as a control variable.

However, only two studies treat PSM as an outcome variable (Jensen, Andersen, and Jacobsen 2019; Kroll and Porumbescu 2019), and four articles emphasize measuring PSM (Kim and Kim 2016a, 2016b, 2017; Marvel and Resh 2019). We categorize 20 studies treating PSM as an independent, control, and moderating variables into a single category—effect. This leaves us three clusters of PSM: effect, outcome, and measurement.

## **REVIEW OF EXPERIMENTAL APPROACHES TO PSM**

This section reviews the three clusters by the type of experiments. Table 2 summarizes the clusters. The first cluster, effects of PSM, has twenty articles, eight articles each for survey and lab experiments and four articles for field experiments. The second cluster, PSM as an outcome, consists of two articles, one using a survey experiment and the other adopting a field experiment. The third cluster, PSM as measurement, has four studies dedicated to measuring PSM through survey experiments. We select a few studies from each cluster and summarize them, beginning with PSM's effects.

[Insert Table 2 about here]

### **PSM as Effect**

Table 3 categorizes 20 studies on the effects of PSM by the outcome. Behavioral (Esteve et al. 2016; Tepe 2016; Tepe and Vanhuysse 2017; Meyer-Sahling, Mikkelsen, and Schuster 2019)

and performance-related outcomes (Bellé 2013, 2014; Caillier 2020; Sun, Wu, and Chen 2019) have been the most researched with four studies each; work effort (Bellé and Cantarelli 2015; Pedersen 2015; Resh, Brewer, and Neumann 2018) and collaboration (Bouwman et al. 2018, 2019; Esteve, van Witteloostuijn, and Boyne 2015) have three studies each; and fairness (Pedersen, Stritch, and Taggart 2017; Tepe and Vanhuysse 2017) and preferences (Bellé and Cantarelli 2018; Tepe and Prokop 2018) have two studies each.

[Insert Table 3 about here]

The following section will review these studies based on the three types of experiments: (1) survey experiment; (2) lab experiment; and (3) field experiment. Of the 20 studies examining the effects of PSM, both survey and lab experiments have eight articles each, and the field experiment has four studies. We select and summarize a few articles for each type.

***Survey Experiments.*** Most of the survey experiments examining PSM's effects use vignette survey experiments—one of the more commonly used experimental methods, whereby a subject responds to questionnaires from hypothetical scenarios provided in a written text.

Table 4 summarizes the details of eight studies utilizing the vignette survey. Moynihan (2013) examines the causal effects of PSM on budget maximization. To make budget allocation decisions, a respondent receives a short hypothetical scenario with information on the budget amount for various programs from previous years, including programs relevant to PSM, such as welfare support and social services. The findings show that high PSM does not lead to a higher budget. Pedersen, Stritch, and Taggart (2017) examine how PSM moderates the hiring process and

procedural fairness perception. The treatment group received a scenario wherein the applicant's friend, who was connected to the hiring company manager, put in a complimentary word for the applicant. The study finds that the treatment group scored lower fairness perception than the control group, and the treatment imposed a greater impact for the respondents with a higher level of PSM.

Surprisingly, none of the eight studies utilizing survey experiments measure PSM through an experiment. Instead, they rely on an observational method—a self-reported survey—to measure PSM.

Two studies, although using a self-reported survey, try to minimize potential biases associated with a self-reported survey by designing a randomized survey experiment (Meyer-Sahling, Mikkelsen, and Schuster 2019; Pedersen 2015). Pedersen (2015) shows that a simple external intervention such as additional texts can activate an individual's public service motivation and increase their willingness to spend time completing a task without extrinsic rewards. Two treatment groups received additional texts about PSM in their survey, whereas the control group did not. The texts read: "Your participation will help ensure that citizens in need are aided in the best possible way" and "Your participation will help ensure the development of society and thus serves the public interest" (Pedersen 2015: 738). The results show that the PSM treatment recipients display a greater willingness to spend more time completing a task than the control group.

Meyer-Sahling, Mikkelsen, and Schuster (2019) try to minimize social desirability bias and other endogeneity problems in a self-reported survey. Examining PSM's effects on ethical behaviors, they introduce a novel survey design that randomizes the order of survey questions for the control and treatment groups. The control group received survey questions on the outcome of

interest first, whereas the treatment group received survey questions on PSM first and then received questions about the outcome. Asking about PSM of a survey respondent itself may affect PSM endogenously. By randomizing the order of the questions for PSM, they tried to capture exogenous variations of PSM.

[Insert Table 4 about here]

**Lab experiments.** Table 5 displays that most lab experiments examining PSM's effects rely on various lab games (Bouwman et al. 2018; Esteve, van Witteloostuijn, and Boyne 2015; Esteve et al. 2016; Tepe 2016; Tepe and Vanhuyse 2017).

In a prisoner's dilemma game and a public goods game, individual players choose to cheat or collaborate, where cheating may maximize self-interest at the expense of collective interest (Krueger and Acevedo 2007). These games are used to examine cooperative behavior (Esteve, van Witteloostuijn, and Boyne 2015; Esteve et al. 2016). Designing their own prisoner's dilemma game, which gauges subject's willingness to cooperate when not cooperating maximizes their self-interest. Esteve, van Witteloostuijn, and Boyne (2015) examine how PSM affects collaboration and find that high PSM participants are more likely to collaborate. Esteve et al. (2016) adopt a public goods game, which gauges subject's cooperative behavior within a group where incentives to freeride prevails, and examine PSM's effect on prosocial behavior. They designed two versions of the game. In the first, a player makes a decision, without having been provided any information, regarding the other player's decision on prosocial behavior. In the second, a player selected at random has access to information about the other player's decision. The result shows that an

individual with high PSM, on average, is more likely to behave prosocially, regardless of whether they have information about the other player's decision on prosocial behavior.

A trust game randomly divides game participants into two groups and allows one group to give a resource—as much of it as they deem fit—to the other group. After receiving the resource, the second group can return a certain amount of it. The first group demonstrates a degree of trust; the second group, in deciding how much to send back to the giver, demonstrates trustworthiness (Tepe 2016). Using a laboratory experiment with a monetary reward, Tepe (2016) examines the effect of students' self-reported PSM on the selection of their university major and their trust and trustworthy behaviors. The result shows that self-reported PSM is positively associated with trust behavior without a significant difference across the majors. Tepe concludes that self-reported PSM is more likely to affect a low-cost decision of small monetary incentives than a high-cost decision, such as selecting a university major.

A dictator game captures the degree of altruism between players. Each player serves as a dictator and a receiver, where a dictator offers a resource that a receiver cannot veto. Each player adopts both roles (Tepe and Vanhuyse 2017). An ultimatum game captures fairness (Güth, Schmittberger, and Schwarze 1982), wherein a player endowed with a resource makes an offer to the other player, as in a dictator game. Unlike a dictator game, however, both players do not receive any resources if the receiver rejects the offer. Both players get the resource according to the initial offer only when the receiver accepts the offer. Using these games, Tepe and Vanhuyse (2017) examine if self-reported PSM is associated with job traits, study choice, and prosocial behavior for different subject groups studying law, business, and public administration. A higher level of PSM is not associated with study preferences for public administration. A higher PSM is positively associated with altruism but negatively with strategic fairness.

Like the aforementioned studies using a survey experiment, all lab experiments examining PSM effects rely on a self-reported survey to measure PSM. Thus, they use an experimental approach to measure an outcome of interest, not PSM. We consider the implications of this in the discussions section of this article.

[Insert Table 5 about here]

**Field Experiments.** Four studies use a field experiment to examine the effects of PSM, as displayed in table 6. Bellé (2013) adopts a randomized control group experiment and examines PSM's effect on job performance with 90 nurses from an Italian hospital. The control group received instructions on assembling surgical kits. Two treatment groups received additional treatments, so-called beneficiary contact or self-persuasion intervention. The first treatment group directly contacted health professionals who received the surgical kits the nurses assembled. The professionals told the treatment group how the kits were helpful and vital for the professionals to complete their work. The second treatment group received information through a presentation that the assembled surgical kits would be delivered to all hospitals in the region to improve patients' health. The result shows that PSM associates positively with the job performance of nurses in the second treatment group.

Linos (2018) examines which type of recruitment message via postcard is more effective in attracting a diverse and higher number of police officer applicants. A message emphasizing career benefits and challenges is more effective than one centering around PSM and underlining the importance of sacrifice and contribution to the community. Although not measuring PSM itself

through a field experiment, Linos captures PSM messaging effects through a field experiment.

Bellé and Cantarelli (2018) provide an exemplary work of measuring PSM through a field experiment and adopt a treatment intervention for PSM. Adopting two discreet choice experiments, they examine how different leadership styles activate motivational forces, including PSM, and affect job preference for 102 Italian public sector employees. The experiment participants are asked to choose a job from a pair of job offers. The treatment intervention related to PSM manipulates the participants' opportunity to serve a few citizens versus many citizens through their jobs. The result implies that PSM and other motivations significantly impact public employees' preference for job positions.

Like most of the experimental studies mentioned above, two out of four field experiment studies also rely on a self-reported survey to measure PSM. The following section reviews experimental studies dedicated to measuring PSM to minimize potential biases from a self-reported survey.

[Insert Table 6 about here]

### **PSM as Outcome**

Only two experimental studies treat PSM as an outcome variable (Jensen, Andersen, and Jacobsen. 2019; Kroll and Porumbescu 2019), which are summarized in table 7.

Using a field experiment with 364 managers and a survey with 3,470 employees in Denmark, Jensen, Andersen, and Jacobsen. (2019) examine if transformational and transactional

leadership affects PSM. The experiment takes place in a leadership training program for public and private managers with three treatment groups and a control group. The control group consists of managers without any leadership training, whereas each treatment group consists of managers receiving a one-year training program for transformational leadership, transactional leadership, or both. The research finds a higher level of PSM in the treatment groups than the control group with no leadership program.

Adopting a vignette survey experiment, Kroll and Porumbescu (2019) examine if extrinsic rewards affect public workers' PSM levels. The control group receives information with an emphasis on low extrinsic rewards for joining the public sector. The treatment group receives the opposite information that the public sector offers higher extrinsic rewards. The result shows that extrinsic rewards did not significantly affect PSM in both groups.

These two studies that treat PSM as an outcome variable capture PSM's extent through a self-reported survey. The next section will discuss its implication.

[Insert Table 7 about here]

## **PSM as Measurement**

Most studies examining the effects of PSM have measured PSM through a self-reported survey. To be exact, 18 out of 20 studies rely on a self-reported survey. This increases to twenty if we include those treating PSM as an outcome variable. The two exceptions are Linos (2018) and Bellé and Cantarelli (2018). Arguably, a self-reported survey may have various biases that



contaminate the measurement. In particular, social desirability bias (SDB) refers to a tendency that an individual may respond to a survey in a way that conforms to society's norms, rather than expressing themselves frankly. Thus, a study dedicated to measure a construct itself and minimize the bias is significant.

Of the 26 experimental studies on PSM, four experimental studies focus on measuring PSM. Three works by Kim and Kim (2016a, 2016b, 2017) introduce a measurement strategy—a list experiment (item count technique) that can capture the magnitude of SDB in a self-reported measure of PSM. A list experiment is a type of survey experiment and can capture SDB as follows. First, it randomly splits a survey sample into a control group and a treatment group. The control receives a list of non-sensitive items and is asked to write how many items they support. The treatment group receives the same items that the control group receives and an additionally sensitive topic. A mean difference for the items agreed by both the control and experimental groups is assumed to equal zero for both groups being randomly selected. Thus, the difference between the two groups represents the magnitude of SDB. This technique tests whether SDB exists in measuring racial attitudes (Kuklinski, Cobb, and Gilens 1997), religion (Kane, Craig, and Wald 2004), and voting behavior (Gonzalez-Ocantos et al. 2012). Utilizing the technique, Kim and Kim (2016a, 2016b, 2017) find that a considerable SDB exists in the measurement of PSM in South Korea, Japan, the United States, the Netherlands, and across different ethnicities.

Similarly, Marvel and Resh (2019) capture PSM through an implicit association test (IAT). The measurement technique aims to measure a concept through one's unconscious state. Arguably, it can reduce bias from normative aspects such as a bias caused by a desire to adhere to societal norms. With 16 samples of graduate students and adults, Marvel and Resh (2019) measure PSM through IAT and conduct an experimental survey to obtain an explicit measure of PSM as altruism.

They find that the correlation between implicit and explicit measures of PSM is weak, suggesting that the IAT captures something different from a standard survey.

## **DISCUSSION**

This section will discuss the strengths and weaknesses of the existing experimental studies on PSM regarding internal validity and external validity. We will also provide recommendations for enhancing future experimental research on PSM.

### **Internal Validity**

To assess the strength of the existing experimental studies' internal validity, we check if their experimental approaches satisfy an experiment's requirements. There are two primary requirements for an experiment. First, the variable of interest should be measured. Second, its research design should include a particular treatment intervention. Without these two requirements, it would not be easy to distinguish an experimental study from an observational one.

Based on this argument, we construct two checklists—manipulation checks 1 and 2—and ask the following questions:

1. Does the study measure PSM through an experiment?
2. Does the study have any treatment intervention related to PSM?

We also check when or where a study relying on a self-reported survey tries to measure PSM. A researcher should be careful in locating the PSM measurement in the survey questionnaire. If a self-reporting PSM level is measured after treatment is intervened, the PSM measure shall not

be free from a reverse causality problem. Thus, we categorize the experiments based on whether they measure PSM before or after treatment.

We apply three conditions to the studies treating PSM as an independent variable, excluding studies treating PSM as a mediating, moderating, or control variable. In general, a variable in these categories shall not be a targeted variable of an experimental treatment. Finally, there remain 13 studies for our analysis. Table 8 classifies them into three categories.

[Insert Table 8 about here]

Of the 13 studies, two studies satisfy the manipulation check 1. The rest of the studies use an experiment to measure the outcome of interest but rely on a self-reported survey to measure PSM. Five studies satisfy the manipulation check 2. Lab experiments lack treatment intervention because they focus on measuring the outcome of interest through various lab games. Only one study satisfies both manipulation checks—Bellé and Cantarelli (2018).

For the third condition, most studies measure PSM either before the treatment or before and after the treatment, without explicitly addressing a potential reverse causality issue. At least four studies measure PSM via self-reporting after the treatment effect (Moynihan 2013; Tepe 2016; Tepe, and Vanhuysse 2017; Tepe and Prokop 2018). We cannot locate when the measurement took place in a study (Caillier 2020). Note again that measuring PSM after a treatment seems problematic because it can further contaminate PSM measure through reverse causality. In other words, the treatment of the dependent variable can affect the measure of PSM.

For example, Moynihan (2013) conducts a survey to measure PSM after completing a

vignette experiment, rather than before the experiment (pre-treatment). We predict that a respondent's self-report on his PSM might be affected by the experiment beforehand. Meyer-Sahling, Mikkelsen, and Schuster (2019) confirm such an impact. To address the issue, they randomize the order of survey questions for the control and treatment groups. They show a substantial difference between the two groups, which implies the order of survey questions impacts a respondent's response to a survey.

### **External Validity**

External validity is another critical issue for experimental research. Chen et al. (2019) argue that the effects of PSM differ across countries and regions. However, most targeted samples of the 26 experimental PSM studies we reviewed are from so-called advanced countries in Western Europe and the United States, with only a few studies dealing with Latin American and Asian countries (Kim and Kim, 2016a, 2016b, 2017; Sun, Wu, and Chen 2019). Table 9 displays the relevant statistics.

From the Table 9, we suggest two main points of interest. First, four studies use samples from Asia (South Korea, Japan, and China), and only one study targets a sample from Chile. Thus, the existing experimental studies on PSM would benefit from research using samples from diverse regions and countries. Second, most existing experimental studies on PSM utilize a student group as a proxy for public officials. Of the 26 studies we reviewed, only six studies used public employees as subjects of their experiments. The majority of studies used students as subjects and several studies targeted students majoring in public administration as a proxy for public employees. However, this assumption requires more research to corroborate, as there is not much empirical evidence to suggest that students of public administration and public officials behave similarly.

Likewise, a student does not have to major in public administration to become a public official, requiring further research to validate the assumption.

[Insert Table 9 about here]

## **RECOMMENDATIONS FOR EXPERIMENTAL IMPROVEMENTS**

This section provides several recommendations for an experimental study on PSM. Most importantly, we stress that an experimental study's research design must be improved.

An experimental study's most significant feature distinct from an observational one is that a researcher can operate an experiment's design to capture some causality between variables. First, if a researcher measures a construct through an experiment, the measurement shall be more exogenous than those measured by an observational method. Second, if a researcher devises treatment through external intervention, they can capture a causal effect of the treatment on the outcome of interest. Ideally, experimental research should have both. To our understanding, however, only one study satisfies both requirements among the existing experimental studies on PSM (Bellé and Cantarelli 2018).

Given the situation, our recommendations are as follows. First, an effort to measuring PSM by an experiment is required. Marvel and Resh (2019) show that substantial bias may exist in measuring PSM in an explicit manner. Although four existing studies try to measure PSM through experiments and minimize social desirability bias, none examine PSM's causal effect. A future study should measure PSM by an experiment and examine PSM's causal effects together.

If a researcher resorts to a self-reported survey to measure PSM, they should design a survey in a way that avoids contaminating the PSM measure due to reverse causality (Meyer-Sahling, Mikkelsen, and Schuster 2019).

A researcher may measure PSM before an experiment, or both before and after the experiment, to address such a problem. It is also possible to benchmark Meyer-Sahling, Mikkelsen, and Schuster (2019), who randomize the order of survey questions for the treatment and control groups and reduce a bias generated from the questions on PSM asked by the survey. In their experiment, the control group receives the survey questions related to the outcome of interest first, whereas the treatment group receives the survey questions capturing PSM first. A future experimental study on PSM may consider randomizing the order of the survey questions.

Second, a future experimental study on PSM should explicitly clarify how its treatment affects the outcome of interest. Many existing studies fail to design the treatment effect. Even for those who have a treatment effect in their research design, it is difficult for a reader to understand whether the treatment affects the outcome of interest directly or through PSM. They should therefore clarify what the treatment intervention is and how such an intervention affects PSM.

Third, to parse out the treatment's intended effect and increase the experimental results' general validity, we suggest that more experiments differentiate their treatments into the low- and high-cost environment. Two existing studies utilize such a scheme effectively. Bellé and Cantarelli (2018), as treatment, provide texts on PSM, where the difference between low and high intensity of treatment is serving more people. Conversely, Tepe (2016) differentiates a low-intensity decision from a high-intensity decision. A low-intensity decision is to accept or refuse cooperation with the other player within a game for a monetary reward. In contrast, a high-intensity decision incurs a lifetime consequence, such as an early career decision. In this regard, a future experimental

study on PSM may adopt a similar research design and examine PSM effects in the context of a real-life decision.

Fourth, it is necessary to differentiate a sample of students from public employees, both in terms of conducting experiments and advancing research on PSM. While previous experiments on PSM have proxied students as civil servants, future studies should take advantage of using students in experiments. Chen et al. (2019) show that the civil service examination in an Asian country may dampen the PSM of the candidates preparing for the exam because it is tremendously competitive. Their findings suggest that future experimental study on PSM should not only compartmentalize samples of potential civil servants from current civil servants, but also develop theories for potential civil servants. In sum, this article suggests that future experiments and research on PSM should adopt a multi-layered approach, differentiating research agenda for existing civil servants versus potential civil servants.

Fifth, it is also desirable for a future experimental study on PSM to consider the different context across countries. The existing literature is limited in that most of them primarily concern Western countries.

Finally, more experiments should focus on the determinants of PSM. Only two studies treat PSM as an outcome of interest. Jensen, Andersen, and Jacobsen (2019) find that leadership affects PSM negatively. Kroll and Porumbescu (2019) show that extrinsic rewards do not significantly affect PSM. The findings suggest that the identified 26 experimental studies on PSM fail to find a determinant positively associated with PSM.

## **CONCLUSION**

Our review of the existing experimental studies on PSM concludes that only one study satisfies the two primary requirements for an experiment. In other words, only one experimental study measures PSM through an experiment and operates a treatment effect on PSM. The literature is there undoubtably still at its nascent stage. It is necessary to conduct experiments examining PSM's causal effects and encourage researchers to operate field experiments. Of the 26 identified experimental studies on PSM, most of them adopt lab and survey experiments. There are five field experiments in total, and only three of them examine the causal effects of PSM.

Overall, we recommend that the research design of the experimental studies on PSM needs to change. It is necessary to measure PSM through an experimental method rather than an observational method. A researcher should order their survey questions randomly or effectively to minimize contamination from a bias, such as reverse causality bias. The mechanism of treatment effect should be articulated to convince readers. It is also desirable to seek the determinants of PSM, separate students from civil servants, and accommodate diverse countries. Through these findings, we hope that future experimental approaches to PSM would benefit from improving causal identification.

## **Notes**

1. Articles published after this date are therefore unlikely included in this study.
2. We dropped an article by Neumann (2016) because although it discusses PSM, it did not measure PSM as a variable.

## **References**



- Bellé, Nicola. 2013. Experimental Evidence on the Relationship Between Public Service Motivation and Job Performance. *Public Administration Review* 71(3): 143–53.
- . 2014. Leading to Make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation. *Journal of Public Administration Research and Theory* 24(1): 109–36.
- Bellé, Nicola, and Paola Cantarelli. 2015. Monetary Incentives, Motivation, and Job Effort in the Public Sector: An Experimental Study with Italian Government Executives. *Review of Public Personnel Administration* 35(2): 99–123.
- . 2018. The Role of Motivation and Leadership in Public Employees' Job Preferences: Evidence from Two Discrete Choice Experiments. *International Public Management Journal* 21(2): 191–212.
- . 2019. Do Ethical Leadership, Visibility, External Regulation, and Prosocial Impact Affect Unethical Behavior? Evidence from a Laboratory and a Field Experiment. *Review of Public Personnel Administration* 39(3): 349–71.
- Bouwman, Robin, and Stephan Grimmelikhuisen. 2016. Experimental Public Administration from 1992 to 2014. *International Journal of Public Sector Management* 29(2): 110–31.
- Bouwman, Robin, Sandra van Thiel, Ad van Deemen, and Etiënne Rouwette. 2018. Accountability and Coalitions: Evidence from a Negotiation Experiment. *Public Administration Review* 78(1): 37–47.
- . 2019. Cooperation and Competition in Public-Sector Negotiations: A Laboratory Experiment. *Public Performance and Management Review* 42(5): 1164–85.
- Bozeman, Barry, and Xuhong Su. 2015. Public Service Motivation Concepts and Theory: A Critique. *Public Administration Review* 75(5): 700–10.
- Caillier, James Gerard. 2020. Testing the Influence of Autocratic Leadership, Democratic

- Leadership, and Public Service Motivation on Citizen Ratings of An Agency Head's Performance. *Public Performance and Management Review* 43(4): 918–41.
- Chen, Chung-An, Don-Yun Chen, Zhou-Peng Liao, and Ming-Feng Kuo. 2019. Wining Out High-PSM Candidates: The Adverse Selection Effect of Competitive Public Service Exams. *International Public Management Journal* 23(4): 443–64.
- Esteve, Marc, Diemo Urbig, Arjen van Witteloostuijn, and George Boyne. 2016. Prosocial Behavior and Public Service Motivation. *Public Administration Review* 76(1): 177–87.
- Esteve, Marc, Arjen van Witteloostuijn, and George Boyne. 2015. The Effects of Public Service Motivation on Collaborative Behavior: Evidence from Three Experimental Games. *International Public Management Journal* 18(2): 171–89.
- Gonzalez-Ocantos, Ezequial, Chad Kiewiet De Jonge, Carlos Meléndez, Javier Osorio, and David W. Nickerson. 2012. Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua. *American Journal of Political Science* 56(1): 202–17.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization* 3(4): 367–88.
- Jensen, Ulrich Thy, Lotte Bøgh Andersen, and Christian Bøthcer Jacobsen. 2019. Only When We Agree! How Value Congruence Moderates the Impact of Goal-Oriented Leadership on Public Service Motivation. *Public Administration Review* 79(1): 12–24.
- Jilke, Sebastian, Steven Van de Walle, and Soonhee Kim. 2016. Generating Usable Knowledge Through an Experimental Approach to Public Administration. *Public Administration Review* 76(1): 69–72.
- Kane, James G., Stephen C. Craig, and Kenneth D. Wald. 2004. Religion and Presidential Politics in Florida: A List Experiment. *Social Science Quarterly* 85(2): 281–93.
- Kim, Seung Hyun, and Sangmook Kim. 2016a. National Culture and Social Desirability Bias in

- Measuring Public Service Motivation. *Administration & Society* 48(4): 444–76.
- . 2016b. Social Desirability Bias in Measuring Public Service Motivation. *International Public Management Journal* 19(3): 293–319.
- . 2017. Ethnic Differences in Social Desirability Bias: Effects on the Analysis of Public Service Motivation. *Review of Public Personnel Administration* 37(4): 472–91.
- Kroll, Alexander, and Gregory A. Porumbescu. 2019. When Extrinsic Rewards Become “Sour Grapes”: An Experimental Study of Adjustments in Intrinsic and Prosocial Motivation. *Review of Public Personnel Administration* 39(4): 467–86.
- Krueger, Joachim I., and Melissa Acevedo. 2007. Perceptions of Self and Other in the Prisoner’s Dilemma: Outcome Bias and Evidential Reasoning. *The American Journal of Psychology* 120(4): 593–618.
- Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. Racial Attitudes and the “New South”. *The Journal of Politics* 59(2): 323–49.
- Linos, Elizabeth. 2018. More than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police. *Journal of Public Administration Research and Theory* 28(1): 67–85.
- Marvel, John D., and William D. Resh. 2019. An Unconscious Drive to Help Others? Using the Implicit Association Test to Measure Prosocial Motivation. *International Public Management Journal* 22(1): 29–70.
- Meyer-Sahling, Jan-Hinrik, Kim Sass Mikkelsen, and Christian Schuster. 2019. The Causal Effect of Public Service Motivation on Ethical Behavior in the Public Sector: Evidence from a Large-Scale Survey Experiment. *Journal of Public Administration Research and Theory* 29(3): 445–59.
- Moynihan, Donald P. 2013. Does Public Service Motivation Lead to Budget Maximization?

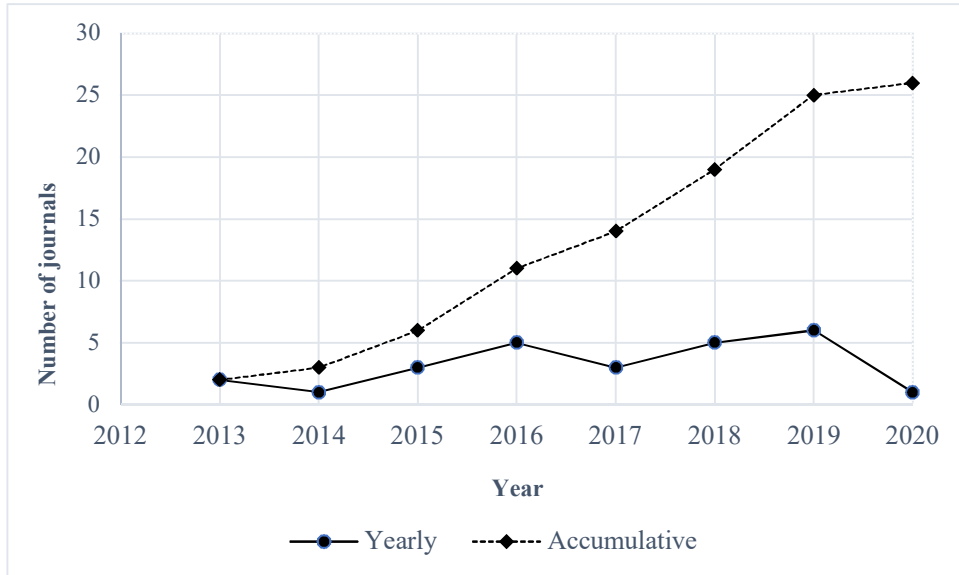
- Evidence from an Experiment. *International Public Management Journal* 16(2): 179–96.
- Neumann, Oliver. 2016. Does Misfit Loom Larger than Fit? Experimental Evidence on Motivational Person–Job Fit, Public Service Motivation, and Prospect Theory. *International Journal of Manpower* 37(5): 822–39.
- Pedersen, Mogens Jin. 2015. Activating the Forces of Public Service Motivation: Evidence from a Low-Intensity Randomized Survey Experiment. *Public Administration Review* 75(5): 734–46.
- Pedersen, Mogens Jin, Justin M. Stritch, and Gabel Taggart. 2017. Citizen Perceptions of Procedural Fairness and the Moderating Roles of “Belief in a Just World” and “Public Service Motivation” in Public Hiring. *Public Administration* 95(4): 874–94.
- Perry, James L. 2012. How Can We Improve our Science to Generate More Usable Knowledge for Public Professionals? *Public Administration Review* 72(4): 479–82.
- Perry, James L., Annie Hondeghem, and Lois Recascino Wise,. 2010. Revisiting the Motivational Bases of Public Service: Twenty Years of Research and an Agenda for the Future. *Public Administration Review* 70(5): 681–90.
- Perry, James L., and Lois Recascino Wise. 1990. The Motivational Bases of Public Service. *Public Administration Review* 50(3): 367–73.
- Rajasekar, Arvina, and Vinoth Kumar. 2019. Randomized Controlled Trials: Gold Standard of Evidence. *Drug Invention Today* 12(6): 1215–17.
- Resh, William G., John D. Marvel, and Bo Wen. 2018. The Persistence of Prosocial Work Effort as a Function of Mission Match. *Public Administration Review* 78(1): 116–25.
- Ritz, Adrian, Gene A. Brewer, and Oliver Neumann. 2016. Public Service Motivation: A Systematic Literature Review and Outlook. *Public Administration Review* 76(3): 414–

26.

- Sun, Xiaoling, Mingzheng Wu, and Qian Chen. 2019. Ego Depletion and Self-Control: The Moderating Role of Public Service Motivation. *Social Behavior and Personality: An International Journal* 47(11): 1–14.
- Tepe, Markus. 2016. In Public Servants We Trust?: A Behavioural Experiment on Public Service Motivation and Trust among Students of Public Administration, Business Sciences and Law. *Public Management Review* 18(4): 508–38.
- Tepe, Markus., and Christine Prokop. 2018. Are Future Bureaucrats More Risk Averse? The Effect of Studying Public Administration and PSM on Risk Preferences. *Journal of Public Administration Research and Theory* 28(2): 182–96.
- Tepe, Markus, and Pieter Vanhuysse. 2017. Are Future Bureaucrats More Prosocial? *Public Administration* 95(4): 957–75.

## Figures and Tables

**Figure 1** Published Articles Related to PSM and Experiments



**Table 1** Classification by the Types of Experiments

Types of Experiments	No. of studies	Author(s)	Journal	Location	Role of PSM
Survey	13	Moynihan (2013)	IPMJ	US	Independent
		Pedersen (2015)	PAR	Denmark	Independent
		Bellé and Cantarelli (2015)	ROPPA	Italy	Moderator
		Kim and Kim (2016a, 2016b, 2017)	AS, IPMJ, RPPA	Japan, Korea, the Netherlands, US	Measurement
		Pedersen, Stritch, and Taggart (2017)	PA	US	Moderator
		Resh, Brewer, and Neumann (2018)	PAR	US	Moderator
		Kroll and Porumbescu (2019)	ROPPA	US	Outcome
		Meyer-Sahling, Mikkelsen, and Schuster (2019)	JPART	Chile	Independent
		Marvel and Resh (2019)	IPMJ	US	Measurement
		Campbell (2019)	PPMR	Korea	Control
		Caillier (2020)	PPMR	Not Specified	Independent
Lab	8	Esteve et al. (2015)	IPMJ	The Netherlands	Independent
		Esteve et al. (2016)	PAR	The Netherlands	Independent
		Tepe (2016)	PMR	Germany	Independent
		Tepe and Vanhuyse (2017)	PA	Germany	Independent
		Tepe and Prokop (2018)	JPART	Germany	Moderator
		Bouwman et al. (2018)	PAR	The Netherlands	Control
		Sun, Wu, and Chen (2019)	SBP	China	Moderator
		Bouwman et al. (2019)	PPMR	The Netherlands	Independent
Field	5	Bellé (2013)	PAR	Italy	Independent
		Bellé (2014)	JPART	Italy	Moderator
		Bellé and Cantarelli (2018)	IPMJ	Italy	Independent
		Linos (2018)	JPART	US	Independent
		Jensen, Andersen, and Jacobsen y(2019)	PAR	Denmark	Outcome

\*\* AS (Administration & Society); IPMJ (International Public Management Journal); JPART (Journal of Public Administration Research & Theory); PA (Public Administration); PAR (Public Administration Review); PMR (Public Management Review); PPMR (Public Performance Management Review); ROPPA (Review of Public Personal Administration); SBP (Social Behavior & Personality)

**Table 2** Role of PSM by Types of Experiments

Role of PSM	No. of Studies	Survey	Lab	Field
Effect	20	8	8	4
Outcome	2	1		1
Measurement	4	4		
Total	26	13	8	5

**Table 3** Effects of PSM by Outcome Types

Outcome Types	Author(s)
Budget maximization	Moynihan (2013)
Job and self-control performance, performance rating (4)	Bellé (2013, 2014); Sun, Wu, and Chen (2019); Caillier (2020)
Negotiation, cooperation, and collaboration (3)	Esteve, van Witteloostuijn, and Boyne (2015); Bouwman et al. (2018, 2019)
Work effort (3)	Pedersen (2015); Bellé and Cantarelli (2015); Resh, Brewer, and Neumann (2018)
Prosocial, trustworthy, and ethical behavior (4)	Esteve et al. (2016); Tepe (2016); Tepe and Vanhuysse (2017); Meyer-Sahling, Mikkelsen, and Schuster (2019)
Fairness	Pedersen, Stritch, and Taggart (2017)
Risk and job preferences (2)	Tepe and Prokop (2018); Bellé and Cantarelli (2018)
Applying rate for police officer job	Linos (2018)
Perception of red tape	Campbell (2019)



**Table 4** Effects of PSM by Survey Types

Author(s)	Survey Type	Outcome Variable	Direction of Causality (+/-)	Statistical Significance	Measurement of PSM
Moynihan (2013)	Vignette	Budget maximization	NA	No	Self-reported survey (post-treatment)
Pedersen (2015)	Randomized survey	Amount of time willing to spend completing a task	+	Yes	Self-reported survey (pre-treatment)
Bellé and Cantarelli (2015)	Vignette	Work effort	NA	No	Self-reported survey (pre-treatment)
Pedersen, Stritch, and Taggart. (2017)	Vignette	Citizen perceptions of procedural fairness	-	Yes	Self-reported survey (not clear)*
Resh, Brewer, and Neumann. (2018)	Vignette	Prosocial work effort	NA	No	Self-reported survey (pre-treatment)
Meyer-Sahling, Mikkelsen, and Schuster (2019)	Randomized survey	Ethical behavior	+	Yes	Self-reported survey (mixed for treat and control groups)
Campbell (2019)	Vignette	Perception of red tape	NA	No	Self-reported survey (pre-treatment)*
Caillier (2020)	Vignette	Performance ratings	+	Yes	Self-reported survey (not clear)*

**Table 5** Effects of PSM by Lab

Author(s)	Lab Type	Outcome Variable	Direction of Causality (+/-)	Statistical Significance	Measurement of PSM
Esteve, van Witteloostuijn, and Boyne (2015)	Prisoner dilemma game	Collaboration	+	Yes	Self-reported survey (pre-treatment)
Esteve et al. (2016)	Public goods game	Prosocial behavior	+	Yes	Self-reported survey (pre-treatment)
Tepe (2016)	Trust game	Trust and trustworthy behavior	+	Yes No	Self-reported survey (post-treatment)
Tepe and Vanhuysse (2017)	Dictator, public good, ultimatum	Major (PA) prosocial behavior strategic fairness	NA + -	No Yes Yes	Self-reported survey (post-treatment)
Bouwman et al. (2018)	Coalition game	Negotiation outcomes	NA	No	Self-reported survey (post-treatment)
Tepe and Prokop (2018)	Lottery choice experiment	Risk preferences (risk-averse behavior)	+	Yes	Self-reported survey (post-treatment)
Sun, Wu, and Chen (2019)	Taste-perception task	Self-control performance	+	Yes	Self-reported survey (both)
Bouwman et al. (2019)	Negotiation game	Level of cooperation	+	Yes	Self-reported survey (post-treatment)

**Table 6** Effects of PSM by Field Experiments

Author(s)	Outcome Variable	Direction of Causality (+/-)	Statistical Significance	Measurement of PSM
Bellé (2013)	Job performance	+	Yes	Self-reported survey (both)
Bellé (2014)	Job performance	+	Yes	Self-reported survey (both)
Linos (2018)	Apply for police officer job Drop out rate	-	Yes	Written post card (not self-reported survey)
Bellé and Cantarelli (2018)	Job preferences	+	Yes	Written text (experiment) (not a self-reported survey)

**Table 7** Studies on Determinants of PSM

Author(s)	Experiment Type	Independent Variable	Direction of	Statistical Significance	Measurement of PSM
-----------	-----------------	----------------------	--------------	--------------------------	--------------------

			Causality (+/-)		
Jensen, Andersen, and Jacobsen (2019)	Field	Leadership	-	Yes	Self-reported survey
Kroll and Porumbescu (2019)	Survey (vignette)	Extrinsic rewards	NA	No	Self-reported survey

**Table 8** Internal Validity Issues for PSM as Independent Variable

Author(s)	Manipulation Check 1: Measuring PSM?	Manipulation Check 2: Treatment for PSM?	PSM Measured Pre or Post Treatment?
Moynihan (2013)	Self-reported survey	Yes	Post
Bellé (2013)	Self-reported survey	Yes	Both
Esteve, van Witteloostuijn, and Boyne (2015)	Self-reported survey	No	Pre
Pedersen (2015)	Self-reported survey	Yes	Pre
Esteve et al. (2016)	Self-reported survey	No	Pre
Tepe (2016)	Self-reported survey	No	Post
Tepe and Vanhuyse (2017)	Self-reported survey	No	Post
Linos (2018)	Field experiment	No	Field experiment
Bellé and Cantarelli (2018)	Field experiment	Yes	Field experiment
Tepe and Prokop (2018)	Self-reported survey	Yes	Post
Bouwman et al. (2019)	Self-reported survey	No	Both
Meyer-Sahling, Mikkelsen, and Schuster (2019)	Self-reported survey	No	Random
Caillier (2020)	Self-reported survey	No	Not clear

**Table 9** Sample Types, Size, and Location

Types of Samples (No. of Studies)	Author(s)	Sample Size	Location
Public employees (6)	Bellé (2013)	90	Italy
	Bellé (2014)	138	Italy
	Bellé and Cantarelli (2015)	296	Italy
	Bellé and Cantarelli (2018)	102	Italy
	Sun, Wu, and Chen (2019)	95	China
	Meyer-Sahling, Mikkelsen, and Schuster (2019)	4,763	Chile
Students (11)	Moynihan (2013)	140	US
	Esteve, van Witteloostuijn, and Boyne (2015)	320	The Netherlands
	Pedersen (2015)	528	Denmark
	Esteve et al. (2016)	263	The Netherlands
	Tepe (2016)	208	Germany
	Tepe and Vanhuysse (2017)	252	Germany
	Bouwman et al. (2018)	87	The Netherlands
	Tepe and Prokop (2018)	252	Germany
	Campbell (2019)	114	South Korea
	Bouwman et al. (2019)	104	The Netherlands
	Kroll and Porumbescu (2019)	129	Florida
Mixed (students + employees) (public + private)	Jensen, Andersen, and Jacobsen (2019)	365 (managers) 3,470 (employees)	Denmark
	Marvel and Resh (2019)	16	US
	Resh, Brewer, and Neumann (2018)	600	US
Applicants for police officer job	Linos (2018)	21,878	US
General samples (not targeted)	Kim and Kim (2016a)	2400 (per country)	Japan, Korea, the Netherlands, US
	Kim and Kim (2016b)	2400	Korea
	Pedersen, Stritch, and Taggart (2017)	910	US
	Kim and Kim (2017)	4,000	US

	Caillier (2020)	599	N/A
--	-----------------	-----	-----