Investigating the Structure and Meaning of Public Service Motivation across Populations: Developing an International Instrument and Addressing Issues of Measurement Invariance

Sangmook Kim,[*] Wouter Vandenabeele,[†,‡] Bradley E. Wright,[§] Lotte Bøgh Andersen,[**,††] Francesco Paolo Cerase,[‡‡] Robert K. Christensen,[§§] Céline Desmarais,[***] Maria Koumenta,[†††] Peter Leisink,[†] Bangcheng Liu,[‡‡‡] Jolanta Palidauskaite,[§§§] Lene Holm Pedersen,[††] James L. Perry,[****] Adrian Ritz,[††††] Jeannette Taylor,[‡‡‡‡] Paola De Vivo[‡‡]

**Seoul National University of Science & Technology; †Utrecht University; ‡Catholic University of Leuven; §Georgia State University; **Aarhus University; ††Danish Institute of Governmental Research; ‡‡Federico II University of Naples; §§University of Georgia; ***Université de Savoie; †††Cardiff University; ‡‡‡Shanghai Jiao Tong University; §§§Kaunas University of Technology; ****Indiana University; ††††University of Bern; ‡‡‡‡University of Western Australia*

ABSTRACT

The growth in international research on public service motivation (PSM) raises a number of important questions about the degree to which the theory and research developed in one country can contribute to our understanding of PSM in other counties. To help address this issue, this study revisits the conceptual and operational definitions of PSM to address weaknesses previously noted in the literature. Although some important steps have been taken to both improve and internationalize the PSM scale, this work has been done incrementally. In contrast, this study takes a more systematic and comprehensive approach by combining the efforts of international PSM scholars to develop and then test a revised measurement instrument for PSM in 12 countries. Although the resulting four dimensional 16-item measure of PSM reported here provides a better theoretical and empirical foundation for the measurement of PSM, our results suggest that the exact meaning and scaling of PSM dimensions are likely to differ across cultures and languages. These results raise serious concerns regarding the ability to develop a single universal scale of PSM, or making direct comparisons of PSM across countries.

Its earlier versions were delivered at the Annual Conference of the European Group for Public Administration, Toulouse, France, September 8–10, 2010, and at the 11th National Public Management Research Conference at Syracuse University, Syracuse, NY, June 2–4, 2011. Address correspondence to the author at smook@seoultech.ac.kr.

Professor Jolanta Palidauskaite died on September 10, 2011, while this article was under review. She was a valued collaborator on this research. She will be missed by all her co-authors.

doi:10.1093/jopart/mus027
Advance Access publication September 10, 2012

INTRODUCTION

Given the service and community-oriented nature of public sector tasks, "an individual's orientation to delivering services to people with a purpose to do good for others and society" has garnered considerable interest among public management scholars (Perry and Hondgehem 2008, vii). Although the initial research on this individual orientation commonly referred to as public service motivation (PSM) began in the United States (Perry and Wise 1990; Rainey 1982), scholars in many countries have increasingly contributed to the accumulation of knowledge about PSM's antecedents and consequences (Perry and Hondeghem 2008; Perry, Hondeghem, and Wise 2010). This growth in international research on PSM raises a number of important questions about the degree to which the theory and research developed in one country can contribute to our understanding of PSM in other counties. In fact, recent research suggests that PSM may have cultural differences not only in its conceptual and operational definitions (Cerase and Farinella 2009; Giauque et al. 2011; Leisink and Steijn 2009; Liu, Tang, and Zhu 2008; Ritz and Waldner 2011; Vandenabeele 2008b), but also its prevalence, antecedents, and consequences (Houston 2011; Vandenabeele and Van de Walle 2008).

To produce more definitive answers regarding these cultural differences, however, requires that we compare the same construct in each country. In other words, we need to know to what degree PSM has the same theoretical structure and psychological meaning across different languages and cultures. If PSM measures do not have the same psychometric properties across different cultures then we cannot have confidence that the empirical findings in one country can generalize to another or that any cultural differences (or similarities) are not merely artifacts of measurement (Chen 2008). Given that the most commonly used measures of PSM were developed in the United States, it is not surprising that scholars have questioned the appropriateness of these measures for explaining and predicting public service-related motivations and behavior internationally (Cerase and Farinella 2009; Coursey and Pandey 2007; Giauque et al. 2011; Ritz and Waldner 2011; Vandenabeele 2008b; Wise 2000). Still others have noted a general need to develop a more appropriate measure of PSM that can be used consistently and confidently regardless of the context or culture in which it is studied (Castaing 2006; Leisink and Steijn 2009; Liu, Tang, and Zhu 2008; Taylor 2007; Vandenabeele 2008a; Wright 2008; Wright and Pandey 2008).

This study describes the results of the first large-scale effort to develop a PSM measurement instrument for international use. In addition to developing a measurement instrument with stronger and more consistent psychometric properties, this study also addresses broader theoretical issues regarding PSM research being conducted across populations whose language, culture, and values differ. After an international effort to identify measurement dimensions and items that avoid culturally specific values or wording, a survey was conducted in 2010 collecting data (n = 2,868) from 12 countries. The revised dimensions and items were then tested cross-nationally to isolate a set of universal dimensions and determine the extent to which the meaning of these dimensions were shared across the 12 countries studied. Based on these analyses, we propose a four dimensional (4D) 16-item measure of PSM as well as

some important guidelines regarding the generalizability of PSM measures. Although the revised measure provides a better theoretical and empirical foundation for the measurement of PSM, tests of measurement invariance suggest that the exact meaning and scaling of PSM dimensions are likely to differ across cultures and languages. As a result, our study raises serious concerns regarding the ability to develop a single universal scale of PSM, or making direct comparisons of PSM across countries.

MEASURING PSM AND THE GLOBALIZATION OF PSM RESEARCH

Although research has relied on a number of different measures of PSM, there is growing recognition that PSM is composed of multiple dimensions (Wright 2008). First suggested 20 years ago, Perry and Wise (1990) proposed that PSM can have dimensions associated with three types of motives: affective, norm-based, and rational. Building on this multidimensional framework, Perry (1996) later developed a 24-item measure of PSM that identified four dimensions of the PSM construct: attraction to public policy making, commitment to the public interest/civic duty, compassion, and self-sacrifice (SS). Since then, many scholars have used Perry's PSM dimensions and items to measure PSM (Kim 2009b; Wright 2008). Even though this 4D structure and its respective measurement items were originally developed based on samples within the United States, they have been used (often with changes) to measure PSM across other cultures and languages including Australia (Taylor 2007), Belgium (Vandenabeele 2008b), China (Liu, Tang, and Zhu 2008; Liu et al. 2011), Germany (Ritz and Waldner 2011), Italy (Cerase and Farinella 2009), Malta (Camilleri 2006), South Korea (Kim, 2009a, 2009b), the Netherlands (Leisink and Steijn 2009), and Switzerland (Giauque et al. 2011).

The growing use of Perry's (1996) 4D measure of PSM has raised some specific concerns regarding the generalizability of its 4D framework in other countries. Although some international scholars have provided evidence the four original dimensions must be supplemented with culturally specific dimensions (Cerase and Farinella 2009; Giauque et al. 2011; Vandenabeele 2008b), others have suggested that some of the dimensions developed in the United States must be omitted (Leisink and Steijn 2009; Liu, Tang, and Zhu 2008) or even combined (Ritz and Waldner 2011; Vandenabeele 2008a) when conducting research in other countries.

At first glance, these findings suggest some meaningful differences in the theoretical structure and psychological meaning of PSM across different languages and cultures but, unfortunately, these findings may only reflect limitations in the original instrument as studies in the United States have raised similar concerns. Consistent with studies in China (Liu, Tang, and Zhu 2008) and the Netherlands (Leisink and Steijn 2009), for example, studies in the United States have raised concerns about the reliability of the compassion dimension (Christensen and Wright 2009; Coursey and Pandey 2007a; Moynihan and Pandey 2007). Similarly, studies in the United States (Christensen and Wright 2009; Perry 1996) and Belgium (Vandenabeele 2008a) have raised questions regarding the discriminant validity of the public interest and SS dimensions. In fact, regardless of the nationality of the sample studied, scholars have

noted the difficulty in getting the full 24-item scale or its abridged versions to achieve acceptable psychometric properties (Christensen and Wright 2009; Clerkin, Paynter, and Taylor 2009; Coursey and Pandey 2007; Coursey et al. 2008; Giauque et al. 2011; Kim 2009a, 2009b; Leisink and Steijn 2009; Liu, Tang, and Zhu 2008).[1] Measurement, therefore, remains a key obstacle to our ability to interpret the findings of international studies with any confidence.

To overcome this barrier, a more concerted effort must be taken to develop a PSM scale that can be used consistently and confidently across languages and cultures. At the very least, a core set of dimensions and items need to be developed that represent aspects of PSM that are shared across cultures. Only when the common elements of PSM are identified, and a reliable way of measuring them is developed, can scholars begin to understand the differences in PSM across cultures including the possibility of additional culturally specific dimensions. Although some important steps have been taken to internationalize the PSM scale (Kim 2009a; Taylor 2007; Vandenabeele 2008b), to date this work has been done incrementally, one country at a time. This study, however, takes a more systematic and comprehensive approach by combining the efforts of international PSM scholars to develop and then test a revised measurement instrument for PSM in 12 countries.

DEVELOPING THE MEASUREMENT INSTRUMENT OF PSM FOR INTERNATIONAL USE

Refining the Multidimensional PSM Framework

Given its popularity and use in different countries around the world, Perry's (1996) measure of PSM consisting of four dimensions and 24 items provides a good foundation for an international measure of PSM. Reviewing past empirical work using this measure suggests a number of specific areas in which the dimensions and its measurement indicators can be improved. Not surprisingly, scholars in the United States and other countries have identified many of the same conceptual weaknesses in how the dimensions were operationalized. In particular, the items of the dimension of attraction to policy making lack face and content validity (Camilleri 2006; Coursey and Pandey 2007; Coursey et al. 2008; Kim 2009b; Ritz 2011). Others have suggested that commitment to public interest dimension needs to concentrate more on a personal disposition to pursue public values (Castaing 2006; Leisink and Steijn 2009; Taylor 2007), and the items of compassion do not always represent affective motives (DeHart-Davis, Marlowe, and Pandey 2006; Moynihan and Pandey 2007; Wright 2008).

Using lessons from this past research, Kim and Vandenabeele (2010) have recently proposed a number of changes to the current multidimensional measures of PSM that would address its weaknesses and support the development of a more universal measure that can be used globally. Building on Perry's (1996) multidimensional measure of PSM, they argue that PSM should continue to be conceived as a 4D construct, with SS as the foundational concept representing the altruistic or pro-social origins of PSM. In addition to SS, they propose three other dimensions—attraction to public participation

1 In addition to weak over all model fit indices, the unacceptable psychometric properties include low reliability scores for dimensions as well as low item factor loadings.

(APP), commitment to public values (CPV), and compassion—that, respectively, represent instrumental, value-based, and affective motives. Instrumental motives, for example, focus on the extent to which individuals want to participate in the public policy process or other activities that contribute to their community or society. Based on this motive, Kim and Vandenabeele (2010) propose redefining Perry's (1996) attraction to policy making dimension as APP and developing items that better represent such instrumental motives. Value-based motives, on the other hand, reflect the extent to which an individual's interest in public service is driven by their internalization of and interest in pursuing commonly held public values such as equity, concern for future generations, accountability and ethics. Using this motive, Kim and Vandenabeele (2010) propose redefining Perry's (1996) commitment to public interest dimension as CPV and identifying measurement items that specifically represent the degree to which individuals share these common values. The final dimension is based on identification motives and it emphasizes an individual's affective commitment to or concern for the needs of specific individuals and groups. Although Perry's original compassion dimension represents the identification motives, Kim and Vandenabeele (2010) note the need to develop measurement items that better capture the degree to which individuals identify with the needs and suffering of others. Although this new framework provides the conceptual basis for the development of an internationally robust measurement instrument of PSM for cross-national research, the next step is to identify specific measurement items for each dimension and test these items using data from diverse international samples.

The next section of this article will briefly explain the processes for developing the international measurement instrument. The survey results are analyzed and discussed in order to confirm the dimensional structure and validate the items of PSM. An examination of measurement invariance is conducted to determine whether the items and the underlying dimensions mean the same thing across the 12 countries.

Item Generation

Based upon the refinement and clarification of the operational dimensions of PSM, items were generated to measure each of the four dimensions: APP, CPV, compassion (COM), and SS. For each dimension of PSM, reflective indicators were identified or generated that satisfy the following criteria (Jarvis, MacKenzie, and Podsakoff 2003): a relative homogeneity and interchangeability of indicators pertaining to each dimension; a high degree of covariation among indicators of each dimension; and the expectation that the indicators of each dimension are likely to be affected by the same antecedents and have the same consequences.

Using these general criteria, items from previously published PSM measures were selected when the item wording fit the revised framework (Kim and Vandenabeele 2010) and the item loaded highly on PSM dimensions in multiple studies (Giauque et al. 2011; Kim 2009a; Perry 1996; Vandenabeele 2008b). In addition to these previously used items, new items were drafted by the authors. In total, 35 items were initially proposed. For the dimension of APP that focuses on individual disposition to work in the public sector, to participate in the public policy process and in activities for community and social development, seven possible items were suggested. Eleven potential items were developed for the CPV dimension concerning the personal disposition to

pursue public interest and public values. For the dimension of COM, which focuses on affective bonding with the identified objects such as other members of a social category or of a political system, 10 possible items were suggested. For the SS dimension, Perry's original items (1996) were the primary source of items.

Next, a group of researchers interested in developing an international measure of PSM were asked to review the 35 items via e-mail in November 2009.[2] The objective was to develop a measurement instrument for PSM with a robust comparative character that might lead not only to a valid instrument on a national basis, but also to an instrument that could be used internationally. For the first-round review, two criteria were employed: (1) whether each item seemed to reflect the intended dimension and (2) whether each item was likely to be meaningful across various national contexts, and in particular, to the situation of civil servants in the reviewers' country. The expert reviewers were also asked to add new items that might better reflect a particular dimension or to rephrase items that were already on the list.

A summary of comments and suggestions emerging from the first round was distributed to the expert reviewers to stimulate comments in a second review round, in December 2009. During this second round, 11 additional items suggested from the first round of reviews were considered. Because some items required further discussion, a third expert review round was conducted in January 2010. One issue that remained unresolved in the discussions related to the dimension of CPV. Some experts argued that it is reasonable to divide CPV into two sub-dimensions, because committing to public interests is different from pursuing public values and there are many important public values in democratic countries. Others contended that CPV can cover all important public values including public interest and there is no real benefit to add one more dimension. As these differences centered on the number of dimensions rather the inclusion of any particular item, this issue was left to be resolved through empirical research.

Based on the comments received during the expert review process, 33 possible items for an international survey were selected. As part of this newly developed 33-item index, seven items were chosen for the dimension of APP, 13 items for CPV (four for CPV1, and nine for CPV2), six items for COM, and seven for SS.

DATA COLLECTION

To generate the instrument to test this measure in multiple countries, each item was translated into the official (native) language of each of the 12 countries surveyed. As this raised the possibility of inaccurate translations, the back-translation of the items into native languages was validated by asking two researchers to independently translate the English-language items. The researchers were instructed to consult with each other over differences between translated items and to test the survey items with some native respondents.

In the survey questionnaire, a five-point Likert-type scale (1 = strongly disagree, 5 = strongly agree) was used. This scale is among the most widely used response formats

2 In addition to several of the study's authors, Emmanuel Camilleri, David Giauque, Gerhard Hammerschmid, Isabel Egger-Peitler, Renate Meyer, and Cristian Pliscoff also participated in the review processes.

in empirical investigations within the social and behavioral science in which there are no "correct" or "incorrect" responses (Spector 1992). The items of PSM were randomly distributed in the survey format. Sex, age, education, length of service, and employment status were also measured in this survey. The resulting survey questionnaire was distributed to all the collaborators on February 9, 2010, to conduct a survey in each country.

The international surveys, fielded from March to September 2010, were administered to civil servants in local governments in Australia, Belgium, China, Denmark, France, Italy, Korea, Lithuania, the Netherlands, Switzerland, the United Kingdom, and the United States.[3] More specifically, because the structure and functions of local governments can differ widely even within countries, the survey in each country was focused on town hall bureaucrats in city, county, or township government.[4] The respondents were permanent employees from both managerial and non-managerial levels and from both administrative departments (i.e., no direct contact with the public) and service departments (i.e., likely to have direct contact with the public).

We aimed to receive 250 responses for each country, as this would enable us to carry out the analyses we wanted to perform. Although there is no ideal sample size for a confirmatory factor analysis in structural equation modeling, one study recommends a sample of 200 to provide a sound basis for estimation (Hair et al. 2010). As for the method of data collection, collaborators were encouraged to use web surveys, but paper-pencil surveys were deemed acceptable in cases where it was not possible to conduct a web survey.[5] Researchers were encouraged to contact just one municipality, except in cases where it was not possible to gather enough data. In such instances, responses from multiple municipalities were accepted.

By the end of September 2010, a total of 2,868 responses were obtained.[6] The researchers selected municipalities because of their accessibility and proximity, and a convenience sample was used in most countries. Seven countries surveyed one municipality, but seven local councils were surveyed in Australia. Web surveys were used in

3 Because the samples collected in the various countries should be comparable to the largest possible extent, we ruled out national and state government employees, as the competencies of national governments, and also of state governments, if they exist constitutionally, are different among most of the countries included in the research.

4 Despite being employed by local governments in many of the countries, police officers, firefighters, school teachers, public transport workers, artists and musicians, as well as nurses and doctors were excluded from the sample.

5 In some cases, not everybody in the sample had access to or frequently used the internet.

6 In Australia, seven local councils, each with 70–232 employees, were surveyed. Paper-pencil surveys were used in two councils, whereas the rest relied on a web survey. The local governments limited distribution of paper-pencil surveys to only one or two sections of the council, e.g., administration. The response rate was approximately 30%. However, it should be noted that not everyone working for the council has direct access to a computer. In Belgium, three municipalities were surveyed on a census base by means of a web survey as all employees had access to a computer. This rendered a response of 120 (22.6%), 79 (75%), and 16 (53%) responses for the respective municipalities, resulting in an overall response rate of 33.1% with 662 invitations being sent out. In China, two departments; social work and administration, were surveyed in one municipality, using paper-pencil surveys. In the social work department, 65 out of 110 surveys were returned, and 165 out of 211 responded from the administration department, giving an overall response rate of 76.7%. In Denmark, a total of 1,282 respondents in a municipality were sent an initial e-mail invitation to a web survey and up to two e-mail reminders, and the response rate was 47.3%. In France, all the employees in two municipalities were surveyed. In one municipality where both web and paper-pencil surveys were utilized, 128 out of 685 questionnaires were returned, whereas in another city, using only a paper-pencil survey, 138 out of 652 were received. The average

Table 1
Background of Respondents (n = 2,868)

Variables	Characteristics	Respondents	%
Sex	Male	1,209	42.2
	Female	1,622	56.6
	N/A	37	1.3
Length of service (years)	0–10	1,126	39.3
	10–20	723	25.2
	20–30	484	16.9
	30+	491	17.1
	N/A	44	1.5
Organizational status:	No	1,472	51.3
"Do you supervise employees?"	Yes	1,076	37.5
	N/A	320	11.2
Country	Australia	249	8.7
	Belgium	214	7.5
	China	230	8.0
	Denmark	249	8.7
	France	266	9.3
	Italy	162	5.6
	Korea	253	8.8
	Lithuania	236	8.2
	The Netherlands	249	8.7
	Switzerland	250	8.7
	United Kingdom	260	9.1
	United States	250	8.7

Note: N/A = no answer.

four countries, whereas three countries used both web and paper-pencil surveys. The number of respondents varied among the countries depending on the numbers of questionnaires distributed, the survey method, and the response rate. In most countries, we used approximately 250 responses but only 162 from Italy.[7] The majority of the respondents were women (56.6%), whereas in terms of service length those in the largest group (39.3%) had worked for fewer than 10 years. Table 1 shows the distribution of respondents' sex, length of service, employment status, and country.

response rate was 19.9%. In Italy, in principle the link to the questionnaire was made available to all employees (about 4,000) in one city, but for some of them access may have been difficult, yielding a response rate of 4.5%. In Korea, 500 employees in a city received a printed questionnaire and 253 responses were returned, yielding a response rate of 50.6%. In Lithuania, 236 responses from 54 municipalities were gathered by web survey. In the Netherlands, three municipalities were surveyed on a census base by means of a web survey. In total, 319 responses of 910 invitations were returned (35%)—90 (45%), 115 (38%), and 114 (28%) for the respective municipalities. In Switzerland, 516 out of 1,326 local employees with e-mail access in a municipality completed the survey, giving a response rate of 38.9%. In the United Kingdom, the survey was posted online and 1,360 employees both in administrative jobs and front-line services of one municipality received an email to complete the survey. The response rate was 19%. In the United States, the total number of employees surveyed was 626 in one city, using both electronic and paper-pencil surveys, resulting in a response rate of 58.8%.

7 In countries where significantly more than 250 surveys were collected, we randomly selected only 250 cases for analysis.

ANALYSES

Using *SPSS 18.0*, descriptive statistics were computed for individual items, as shown in Table 2. One indicator (APP3) was deleted because it was omitted in the survey conducted in the United States. In the assessment of normality, no items showed a skew or kurtosis value greater than the cutoffs of |3| or |8| recommended by Kline (2005). Based on inspection of item-total correlations, one item (COM4) was dropped from further analysis because it was weakly correlated with the overall index (.011). The remaining 31 items used in the next stage of analysis have low rates of missing data, ranging from 0.16% to 1.43%. Even with such low levels of missing data, mean substitution was used to calculate replacement values for the missing values in each country's data set to help avoid estimation errors (Hair et al. 2010).

To test the hypothesized five-factor model, a series of confirmatory factor analyses (CFA) were conducted in *LISREL 8.72* (Jöreskog and Sörbom 2005) using the robust maximum likelihood estimation (RMLE) method (Yang-Wallentin, Jöreskog, and Luo 2010).[8] The CFA model in the initial analysis hypothesized, *a priori*, that: (1) responses to the 31-item PSM measurement instrument could be explained by five factors: APP, CPV1, CPV2, COM, and SS; (2) each item would have a non-zero loading on the PSM factor that it was designed to measure and zero loadings on all other factors; (3) the five factors would be correlated; and (4) measurement error terms would be uncorrelated.

The evaluation of model fit is based on an inferential goodness-of-fit index, called the chi-square value, in combination with several descriptive indices. In addition to the Satorra-Bentler-scaled chi-square statistic ($SB\chi^2$) which is likely to reject model fit for the large sample sizes used here, we followed recent recommendations (Williams, Vandenberg, and Edwards 2009) and assumed the model achieves an acceptable fit to the data when comparative fit index (CFI) exceeds 0.95, the root mean square error of approximation (RMSEA) is below 0.08 (the 90% confident interval will be reported within parentheses) and the standardized root mean residual (SRMR) is below 0.10. The resulting CFA with RMLE (reported in Table 3) suggested that the initial five-factor model was a good fit to the data, $SB\chi^2$ [$df = 424$] = 3989.0, $p < .05$; CFI = 0.972; RMSEA = 0.053 [.051, .054]; SRMR = 0.058. However, at this stage, APP was highly correlated with CPV1 ($r = .993$), suggesting the two dimensions were not unique and lacked adequate discriminant validity.

To address concerns regarding discriminant validity, the APP and CPV1 dimensions were combined to form a single dimension: attraction to public service (APS). The items of CPV1 may be regarded as action-oriented by the respondents because dedication to public service, community, and common good is underlined in these

8 Although using ordinal scales violates the maximum likelihood (ML) assumption that the observed variables are continuous and normally distributed, recent evidence suggests that robust ML is relatively robust to violations of the multivariate normality assumption and is generally endorsed for most uses (Byrne and van de Vijver 2010; Iacobucci 2010; Olsson et al. 2000; Yang-Wallentin, Jöreskog, and Luo 2010). Given that there is also an argument that suggests that diagonally weighted least squares (DWLS) estimation may be better than ML (Coursey and Pandey 2007; Kim 2011; Moynihan, Pandey, and Wright 2012), we also conducted all of the analysis using DWLS. The DWLS results, however, did not substantively change any of the empirically findings produced by the RMLE estimation results reported. One disadvantage of using the DWLS results was that the invariance tests relied solely on the ΔCFI test as the Bryant and Satorra (2012) scaling correction for calculating the $\Delta\chi$ cannot be used when using DWLS estimation in LISREL.

Table 2
Possible Indicators of PSM and Descriptive Statistics (n = 2,868)

Dimensions and Items	Mean	Standard Deviation	Item-Total Correlation
APP			
APP1: I am interested in helping to improve public service	4.12	.696	.524
APP2: I am satisfied when I see people benefiting from the public programs I was involved in	4.41	.662	.421
APP3: I like to discuss topics regarding public programs and policies with others	3.55	.917	.482
APP4: I believe that public sector activities contribute to our general welfare	3.93	.753	.412
APP5: I admire people who initiate or are involved in activities to aid my community	4.23	.672	.493
APP6: Contributing to public programs and policies helps me realize myself	3.51	.881	.478
APP7: It is important to contribute to activities that tackle social problems	4.09	.734	.552
CPV			
Sub-dimension for Public Interests (CPV1)			
CPI1: Meaningful public service is very important to me	4.18	.695	.525
CPI2: It is important for me to contribute to the common good	4.09	.662	.596
CPI3: I would prefer seeing public officials do what is best for the whole community, even if it harmed my interests	3.48	.899	.480
CPI4: Serving the public interest is more important than helping a single individual	3.37	.995	.251
Sub-dimension for Public Values (CPV2)			
CPV1: I think equal opportunities for citizens are very important	4.28	.730	.484
CPV2: It is important that citizens can rely on the continuous provision of public services	4.16	.704	.437
CPV3: It is fundamental that public services respond to the needs of the citizens	4.33	.710	.356
CPV4: Decisions regarding public services should be democratic despite the time and effort it takes	3.94	.852	.325
CPV5: Everybody is entitled to a good service, even if it costs a lot of money	3.67	.932	.376
CPV6: It is fundamental that the interests of future generations are taken into account when developing public policies	4.26	.689	.439
CPV7: To act ethically is essential for public servants	4.41	.681	.462
CPV8: I believe that public employees must always be aware of the legitimacy of their activities	4.34	.692	.402
CPV9: I personally identify with the aim of protecting individual liberties and rights	4.14	.763	.392

continued

Table 2 *(continued)*

Dimensions and Items	Mean	Standard Deviation	Item-Total Correlation
Compassion (COM)			
COM1: It is difficult for me to contain my feelings when I see people in distress	3.54	.951	.399
COM2: I feel sympathetic to the plight of the underprivileged	3.87	.825	.511
COM3: I empathize with other people who face difficulties	3.88	.707	.389
COM4: I have little compassion for people in need who are unwilling to take the first step to help themselves	2.99	1.083	.011
COM5: I get very upset when I see other people being treated unfairly	4.12	.755	.487
COM6: Considering the welfare of others is very important	3.91	.765	.502
Self-Sacrifice (SS)			
SS1: Making a difference to society means more to me than personal achievements	3.57	.904	.508
SS2: I am prepared to make sacrifices for the good of society	3.29	.872	.584
SS3: I believe in putting civic duty before self	3.39	.949	.454
SS4: I am willing to risk personal loss to help society	3.02	.982	.531
SS5: People should give back to society more than they get from it	3.45	.888	.412
SS6: Serving other citizens would give me a good feeling even if no one paid me for it	3.63	1.049	.368
SS7: I would agree to a good plan to make a better life for the poor, even if it costs me money	3.50	.895	.546

Table 3
SFLs and Correlations for the 31-Item PSM Index (n = 2,868)

Dimensions and Items		SFL	Alpha	
APP	APP1	.670	.699	
	APP2	.564		
	APP4	.506		
	APP5	.634		
	APP6	.523		
	APP7	.654		
CPV1	CPI1	.672	.542	
	CPI2	.749		
	CPI3	.502		
	CPI4	.303		
CPV2	CPV1	.692	.745	
	CPV2	.628		
	CPV3	.510		
	CPV4	.454		
	CPV5	.459		
	CPV6	.600		
	CPV7	.674		
	CPV8	.575		
	CPV9	.516		
COM	COM1	.537	.687	
	COM2	.611		
	COM3	.566		
	COM5	.665		
	COM6	.676		
SS	SS1	.597	.782	
	SS2	.783		
	SS3	.585		
	SS4	.768		
	SS5	.491		
	SS6	.435		
	SS7	.706		
Measure of fit (RMLE)	MLχ^2 (df = 424) = 6713.0, p < .05; NTWLSχ^2 (df = 424) = 7580.9, p < .05; SBχ^2 (df = 424) = 3789.0, p < .05; CFI = .972; RMSEA = .053, [.051, .054]; SRMR = .058			
Inter-factor correlations				
	1	2	3	4
1. APS				
2. CPV1	.993			
3. CPV2	.905	.809		
4. COM	.790	.759	.792	
5. SS	.702	.772	.520	.722

items. Consequently, it is reasonable to combine APP with CPV1 to form APS, a dimension that focuses more on disposition to serve the public, to work for the common good, and to participate in public policy processes.

After the overall number of dimensions was reduced to four, using standardized factor loadings (SFLs) and modification indices as criteria, the items with the lowest factor loading in each dimension were deleted and the items whose modification index indicated association with multiple dimensions were eliminated. Using these exclusion criteria in an iterative process, successive CFA models were tested until a model with very good fit to the data was achieved.

The outcome of this refinement process was a 16-item index of 4 factors, and overall model fit is quite strong (SBχ^2 [*df* = 98] = 564.1, $p < .05$; CFI = 0.989; RMSEA = 0.041 [.038, .044]; SRMR = 0.042). The estimation results for the 4D model are presented in Table 4. The results provide support for convergent validity. The composite reliability (CR) of the set of reflective indicators for each dimension of PSM ranged from .716 to .824, and indicated adequate internal consistency.[9] The resulting factor structure showed a four-factor structure with all items loading significantly onto their *a priori* dimension ($p < .05$), and the SFLs ranging from .556 to .812. The results also provide support for discriminant validity as the correlation estimates between the four factors ranged from .482 to .851 and the confidence intervals (±2 standard errors) around the correlation estimates between the two factors did not include 1.00 (Anderson and Gerbing 1988). To test whether the 4D model is superior to other designs such as the 2D or 3D models, a series of CFA were conducted, and the results show that the 4D model is a significantly better fit than other models.[10]

We tested the cross-country equivalence of the measuring instrument in a set of hierarchical multi-group CFA where each model imposes additional constraints to determine the extent to which the items and the underlying constructs have a shared meaning across all 12 countries (Cheung and Rensvold 2002). Although different levels or degrees of measurement invariance exists, our study focuses on testing three forms most commonly recommended and used for validating the use of a measure across different groups (Chen 2008; Schmitt and Kuljian 2008).

The first of these levels concerns *configural variance*. Configural variance tests whether the measure has the same hypothesized factorial structure (the same number of factors and the same items loading on each factor) in each country. Although this test cannot rule out the existence of additional culture-specific dimensions, it is important because so many previous studies have failed to confirm the same four PSM dimensions even when using the same measurement items (Christensen and Wright 2009; Clerkin, Paynter, and Taylor 2009; Leisink and Steijn 2009; Liu, Tang, and Zhu 2008; Moynihan and Pandey 2007; Vandenabeele 2008a). Assuming the conditions for configural invariance are met, the next level is *metric invariance*. Metric invariance tests whether the strength of the relationship between each factor and its associated items (reflected in the factor loading) are the same across countries. If the factor loadings are equal, then the unit of measurement is assumed to be the same across groups;

9 CR is interpreted similarly to Cronbach's alpha but better fits the measurement assumptions of confirmatory factor analysis (Raykov 1997).

10 Given the especially strong ($r > .787$) bivariate correlations between three of the dimensions reported in Table 4, we tested a series of alternative factor structures that combine one or more of these dimensions. The hypothesized four-factor model, for example, was a significantly better fit than three-factor models combining the APS and CPV ($\Delta\chi^2$ [*df* =3] = 182.38, $p < .05$), CPV and COM ($\Delta\chi^2$ [*df* = 3] = 310.35, $p < .05$), or APS and COM ($\Delta\chi^2$ [*df* = 3] = 333.19, $p < .05$) dimensions. The four-factor model was also better than a two-factor model combining APS, CPV, and COM dimensions ($\Delta\chi^2$ [*df* = 5] = 542.01, $p < .05$).

Table 4
SFLs and Goodness-of-Fit Indices for the 16-Item PSM Index (n = 2,868)

Dimensions and Items	SFL
APS	
APP5: I admire people who initiate or are involved in activities to aid my community	.683
APP7: It is important to contribute to activities that tackle social problems	.646
CPI1: Meaningful public service is very important to me	.710
CPI2: It is important for me to contribute to the common good	.744
CPV	
CPV1: I think equal opportunities for citizens are very important	.692
CPV2: It is important that citizens can rely on the continuous provision of public services	.634
CPV6: It is fundamental that the interests of future generations are taken into account when developing public policies	.618
CPV7: To act ethically is essential for public servants	.660
COM	
COM2: I feel sympathetic to the plight of the underprivileged	.591
COM3: I empathize with other people who face difficulties	.556
COM5: I get very upset when I see other people being treated unfairly	.651
COM6: Considering the welfare of others is very important	.686
SS	
SS2: I am prepared to make sacrifices for the good of society	.812
SS3: I believe in putting civic duty before self	.594
SS4: I am willing to risk personal loss to help society	.796
SS7: I would agree to a good plan to make a better life for the poor, even if it costs me money	.723
Measure of fit (RMLE)	$ML\chi^2$ (df = 98) = 1266.6, p < .05; $NTWLS\chi^2$ (df = 98) = 1253.0, p < .05; $SB\chi^2$ (df = 98) = 564.1, p < .05; CFI = .989; RMSEA = .041, [.038, .044]; SRMR = .042

	CR	Inter-factor correlations		
1. APS	.790	1. APS	2. CPV	3. COM
2. CPV	.747	.851		
3. COM	.716	.787	.831	
4. SS	.824	.580	.482	.696

a condition necessary before predictive relationships (PSM's antecedents and consequences) can be meaningfully compared across countries (Chen 2008). As such, it is a necessary condition for quantitative comparative correlational research or comparative construct validation.

Assuming metric invariance is met, a third test is conducted to ascertain *scalar invariance*. Scalar invariance requires that the intercepts of the regression equations of the items on their hypothesized factors are equivalent across countries. This third level of invariance is important because both the factor loadings (metric invariance) and intercepts (scalar invariance) must be identical before factor means can be compared across groups (Chen 2008).

Consistent with the steps described above, we tested the degree to which the revised 4D measure of PSM was measurement invariant in a series of multi-group factor analyses where each model imposes additional constrains on the common four-factor model. The first model (M1) tested configural invariance of the four dimensions across countries while allowing factor loadings, intercepts, and error variances to vary across countries. The fit indices for M1 were χ^2 (df = 1176) = 1759.33, $p < .05$; CFI = 0.9884, RMSEA = 0.046 [.041, .050], SRMR = 0.063. These indices indicated that the four-factor model of the PSM instrument exhibits good fit across the samples from the 12 countries. Thus, we can conclude that the conditions for configural variance were met and the countries did not differ in the number or composition of the four factors represented in the measurement instrument.

Given that configural variance was achieved, we tested for metric invariance by running a second model (M2) that constrained all factor loadings to be equal across all 12 countries but still allowed the intercepts and error variances to vary. Although these fit statistics (with the exception of the chi-square) all suggest the model provides a good overall fit to the data across the 12 countries ($SB\chi^2$ [df = 1308] = 2011.38, $p <$ .001; CFI = 0.9860, RMSEA = 0.048 [.043, .052], SRMR = 0.067), the test for metric invariance requires that the fit of the factor loading constrained model (M2) not be significantly worse than the fit of the model only constraining the factor structure (M1). Testing the differences in nested models to determine measurement invariance requires comparing the differences between the model in the chi-square and CFI fit indices. Metric invariance would be accepted if the difference in the chi-square ($\Delta\chi^2$) between M1 and M2 is not statistically significant ($p > .05$) and the difference in the CFI (ΔCFI) across the two models is less than 0.002 (Meade, Johnson, and Braddy 2008).[11] Unfortunately, the measure failed to achieve metric invariance across the 12 countries on either test ($\Delta SB\chi^2_{(132)}$ = 308.85, $p < .05$; ΔCFI = 0.0024).[12] Given that metric invariance is required before stricter invariance tests be conducted, we can assume that the measure is not scalar invariant without additional tests.

To investigate and illustrate the support for the four-factor structure in more detail, the validity of the baseline model was tested separately for each country and reported in Table 5. The goodness-of-fit indices indicated a good model fit for each country. The SFLs of the baseline CFA model for each country are provided in Table 5 and ranged from 0.30 to 0.92. Following the evaluation standards suggested by Comrey and Lee (1992), the vast majority (93.75%) of the factors loadings in each country specific test were good (0.55 or higher), and another 2.6% were considered of fair or acceptable strength (0.45≥ and <0.55). The few factor loadings that did not meet even the lower of the two standards (those less than 0.45) were only found in 2 of the 4 dimensions (CPV and COM) and in only 6 of the 12 countries (Belgium,

11 Although others have suggested that a more lenient threshold of 0.01 could be used for the ΔCFI (Cheung and Rensvold 2002), this recommendation was developed based on tests where the CFI was derived from the minimum fit function chi-square (as is the case when using software programs such as AMOS and Mplus). The 0.002 threshold was recommended when using software (like LISREL used here) that derives the CFI from the normal theory weighted least squares chi-square.

12 As the difference between the Satorra-Bentler scaled chi-square statistic ($SB\chi^2$) for two models is not chi-square distributed, the nested chi-square test was calculated using the Sattora-Bentler scaling corrections outlined by Bryant and Satorra (2012).

Table 5
SFLs and Goodness-of-Fit Indices for the Baseline Model in 12 Countries (RMLE)

Factors and Items	Australia	Belgium	China	Denmark	France	Italy	Korea	Lithuania	Netherlands	Switzerland	United Kingdom	United States
APS												
APP5	.790	.792	.657	.663	.597	.697	.666	.764	.651	.624	.721	.795
APP7	.594	.475	.672	.743	.608	.767	.546	.621	.755	.751	.687	.636
CPI1	.802	.782	.716	.448	.805	.617	.773	.532	.718	.684	.792	.832
CPI2	.920	.696	.692	.736	.732	.783	.724	.743	.667	.677	.787	.773
CPV												
CPV1	.562	.699	.704	.631	.704	.808	.636	.612	.626	.684	.778	.752
CPV2	.541	.744	.619	.624	.749	.712	.655	.689	.633	.425	.615	.635
CPV6	.698	.488	.651	.423	.716	.438	.645	.599	.610	.638	.648	.661
CPV7	.666	.624	.658	.765	.618	.531	.697	.698	.595	.706	.570	.674
COM												
COM2	.765	.809	.660	.764	.651	.553	.623	.304	.770	.689	.689	.753
COM3	.766	.680	.759	.559	.501	.334	.640	.565	.649	.499	.670	.759
COM5	.717	.506	.724	.744	.685	.553	.580	.710	.554	.681	.541	.633
COM6	.705	.405	.692	.684	.846	.746	.765	.510	.368	.770	.851	.732
SS												
SS2	.823	.803	.904	.848	.883	.845	.825	.713	.789	.794	.755	.821
SS3	.779	.603	.742	.749	.740	.586	.704	.488	.733	.595	.840	.598
SS4	.806	.745	.852	.859	.814	.757	.839	.728	.773	.724	.863	.794
SS7	.664	.807	.821	.670	.749	.692	.717	.673	.695	.684	.626	.808
Measures of fit												
MLχ^2	762.5*	447.3*	331.9*	332.9*	361.9*	422.0*	414.5*	382.1*	360.8*	327.5*	429.3*	357.4*
NTWLSχ^2	582.6*	390.4*	328.4*	291.7*	335.5*	393.9*	398.7*	337.0*	340.5*	323.3*	367.1*	333.7*
SBχ^2	149.2*	179.3*	139.7*	134.6*	138.0*	143.7*	151.7*	131.1*	155.7*	160.3*	152.0*	129.4*
df	98	98	98	98	98	98	98	98	98	98	98	98
CFI	.988	.972	.992	990	.992	.983	.988	.989	.984	.985	.989	.995
RMSEA	.046	.062	.043	.039	.039	.054	.047	.038	.049	.051	.046	.036
SRMR	.102	.091	.072	.068	.065	.095	.078	.089	.070	.063	.078	.063

Denmark, Italy, Lithuania, the Netherlands, and Switzerland). Even so, the overall pattern of factor loadings seems to compensate for any item-level weaknesses because the CF was strong (greater than 0.70) in 3 of the dimensions (APS, CPV, and SS) in all 12 of the countries tested. In contrast, the internal reliability of the compassion dimension was only strong in 9 of the 12 countries. In the remaining three countries—Italy, Lithuania, and the Netherlands—the CF of the compassion dimension was above 0.60.[13]

Given the questions regarding the discriminant validity of the four dimensions raised by the previous research noted above, we investigated the strength of the four-factor structure further by examining the correlations between each dimension for each of the 12 counties. In only 4 of the 12 countries did the 95% confidence intervals (±2 standard errors) around the bivariate correlation estimates between dimensions include a perfect (1.0) correlation (Anderson and Gerbing 1988). In the Italian sample, for example, the correlations suggest that three of the dimensions (APS, CPV, and COM) were not empirically distinct and could be combined. Similarly, we could not rule out the possibility that the CPV and COM dimensions were perfectly correlated in Lithuania and Korea or that the APS and CPV dimensions were not empirically distinct in the Netherlands.[14]

Although the results of the test of configural variance, combined with the evidence of convergent and discriminate validity found at the country level, provide strong support for the international use of the measures' dimensions and items in at least 8 of the 12 countries studied, there are important limitations. The measures' lack of metric and scalar invariance suggests that the PSM measure test here does not have exactly the same meaning or scaling across the 12 countries. There are a number of reasons why a measure can fail to achieve metric invariance. In addition to the possibility that the item content can be more appropriate (or interpreted differently) in one culture than in another, it can also be due to errors in the translation of the items across languages or differences in response sets such as the propensity for some cultures to avoid using extreme response categories (Chen 2008).

In an attempt to minimize the potential cultural differences and test for measurement invariance under the most conducive conditions, we limited the countries included in the analysis to just three—Australia, the United Kingdom, and the United States—that share many of the characteristics likely to affect or reflect culture. In addition to having English speaking, Christian, and Caucasian majorities, all three countries are often classified as having similar views on welfare, work, and equality (Leibfried 1992; Siaroff 1994; Wildeboer Schut, Vrooman, and de Beer 2001). Although the first three characteristics are likely to generally influence the values and experiences that shape a culture, a recent study has provided evidence that the welfare regime type can influence the levels of obligation-based intrinsic motives like PSM (Houston 2011).

13 The CF for the compassion index was 0.68 in the Netherlands and 0.64 in both Italy and Lithuania.

14 We do not report the individual correlations between the dimensions for each country because the measure failed to achieve sufficient invariance to allow meaningful comparisons of the relative strength or pattern of correlations across countries. That said, the differences that can occur when using this measure across countries is illustrated by how the correlations between the dimensions varied across the 12 counties (the variation in the correlation found between the APS and SS dimensions was fairly typical of that found between each pairing of dimensions and ranged from a low of 0.41 to a high of 0.80).

When we limited our analysis to these three countries, the conditions for configural variance were met and we found no cross-country differences in the number or composition of the four factors represented in the measurement instrument (SB χ^2 [df = 294] = 432.61, $p < .05$; CFI = 0.9907, RMSEA = 0.043 [.034, .052], SRMR = 0.063). We then tested for metric invariance by running a second model that constrained all factor loadings to be equal across these three countries while still allowing the intercepts and error variances to vary. The results from this test suggested that the measure was metric invariant across Australia, the United Kingdom, and the United States. The overall fit statistics (with the exception of the chi-square) suggested this constrained model provided a good overall fit to the data across the three countries (SBχ^2 [df = 318] = 460.72, $p < .05$; CFI = 0.9905, RMSEA = 0.042 [.033, .050], SRMR = 0.068) and did not provide a significantly worse fit than the model only constraining the factor structure ($\Delta\chi^2$ [df = 24] = 34.27, $p > .05$; ΔCFI = −0.0002).[15] We then tested scalar invariance by constraining the intercepts of the regression equations of the items on their hypothesized factors as equivalent across these three countries. The results of this test failed to support scalar invariance. Although the overall fit statistics (with the exception of the chi-square) were consistent with model fit (SBχ^2 [df = 342] = 561.76, $p < .05$; CFI = 0.9853, RMSEA = 0.051 [.043, .058], SRMR = 0.070), both the ΔCFI and $\Delta\chi^2$ suggested that this factor loading and intercept constrained model provided a significantly worse fit than the model only constraining the factor structure and loadings.

RECOMMENDATIONS AND CONCLUSIONS

Much of the work on PSM is now being done across cultures/countries which raises a number of important issues/questions about the applicability of PSM across cultures. In fact, one recent international study of PSM suggested "that although public service motivation is a more or less universal concept…the focus and empirical nature tends to differ due to a different or partial implementation of similar ideas" (Vandenabeele and Van de Walle 2008, 236). Although this conclusion is consistent with more recent studies that suggest that while PSM seems to exist in many countries (Houston 2011), the number, type, and nature of PSM dimensions are not always consistent with the theoretical expectations originally developed in the United States. Given that many of the differences in the structure and dimensions of PSM identified by studies conducted outside the United States mirror empirical findings in United States samples, some of the concerns raised regarding the cultural differences in PSM may only be an artifact of the limitations of currently available PSM measures. In order to better understand the true nature of these past findings and advance the ability of scholars to conduct international research on PSM, our study represents the first comprehensive international effort to develop and test a revised measure of PSM with universal applicability. Although we were not able to develop a truly universal measure of PSM, our findings do have at least three important implications for future research on PSM.

15 As noted previously, the $\Delta\chi^2$ test was calculated using the scaling corrections outlined by Bryant and Satorra (2012).

First, by addressing many of the conceptual concerns previously raised about dimensions and items in Perry's (1996) original instrument, this study provides a better conceptual framework and measurement instrument for scholars to use as a starting point when studying PSM in single country research. Using the revised construct of PSM and data from an international survey conducted in 12 countries, we found that the APS dimension measures the instrumental motives of PSM, the CPV dimension determines the value-based motives of PSM, whereas the COM and SS dimensions gauge the identification motives and self-sacrifice, respectively. These improvements culminated in a core set of dimensions and measurement items that produced the same four-factor structure (configural invariance) across the 12 countries, with the psychometric properties of this measure consistently strong across all metrics in eight of these 12 countries. Although there is a high degree of correlation between these four dimensions, we encourage scholars to use all four of the dimensions because they represent related but distinct aspects of PSM that past research suggests can have different antecedents and consequences (DeHart-Davis, Marlowe, and Pandey 2006; Moynihan and Pandey 2007; Pandey and Stazyk 2008; Perry 1997).

Second, this study suggests that any PSM measurement instrument developed for international use is likely to have a number of inherent weaknesses. Even with extensive input from international PSM scholars, our results suggest that the exact meaning and scaling of PSM dimensions are likely to differ across cultures and languages. As a result, PSM scholars in one country must take care when building on theory and findings produced in another country. When conducting literature reviews, for example, scholars should clearly specify when the theory and findings they use were developed in cultural contexts that differ from their own study as the lack of measurement invariance suggests that they must confirm (rather than assume) that findings in one country will generalize to another. In addition to recognizing that the PSM dimensions may not have the same strength or direction of relationships with other variables in each country, PSM scholars also cannot rule out the possibility that any differences in these relationships are due to the cultural differences in the way that PSM is conceptualized. Multinational researchers should also be cautious about any direct comparisons of PSM (i.e., means tests) across countries as our findings suggest that the dimensions do not always have the same meaning and scaling across cultures.

Although the four dimensions and 16 items reported here provide a better theoretical and empirical foundation for the measurement of PSM, the lack of metric and scalar invariance across all of the countries in our study suggest that additional work (combining, omitting, and even adding dimensions) is likely needed to adapt and validate the PSM measure on a country-by-country basis. Among the countries studied here, four—Italy, Korea, Lithuania, and the Netherlands—necessitate measurement modification, whereas eight may only require qualification. To the latter, our 16-item index has immediate research potential within Australia, Belgium, China, Denmark, France, Switzerland, the United Kingdom, and the United States as long as researchers avoid pooling data and making cross-country comparisons. In short, while we have made some progress here within some of the countries studied, future research is needed to understand the differences in PSM across cultures/languages, not just in terms of its antecedents and consequences (or their relative importance), but also in terms of its meaning and operationalization.

Although these findings may suggest some conceptual or operational weaknesses in the PSM construct, they also reflect the realities associated with cross-cultural research in organizational behavior. Similar to our results regarding the lack of invariance in PSM, recent research has failed to find consistent evidence that common measures of personal values and goals are invariant across cultures (Grouzet et al. 2005).[16] In fact, important cultural differences in both meaning and relationships have been found in some of the most fundamental organizational behavior constructs. For example, research suggests that the meaning of job satisfaction may only be equivalent in countries with the same language and similar cultural histories (Liu, Borg, and Spector 2004) and intrinsic job characteristics are more strongly associated with job satisfaction in countries with greater wealth and individualistic values (Huang and Van de Vliert 2003). Evidence also suggests that important cultural differences exist in organizational commitment (Hattrup, Mueller, and Aguirre 2008; Meyer et al. forthcoming). Of particular relevance to PSM is that the relationship between postmaterial job quality (i.e., the ability to be useful to society and help other people) and organizational commitment was found to be stronger in countries with collectivistic cultures than countries with individualistic cultures (Andolsek and Stebe 2004). Thus, although our findings not only highlight the need for more systematic research on the cultural differences in PSM, they should also serve as a warning for all public management scholars conducting comparative research or even research on one country that relies heavily on measures and findings developed in other countries (Chen 2007).

Third, although a shared meaning and scaling of PSM may be uncommon, there may be identifiable patterns in PSM across countries that share similar cultural values, language, levels of development, or geographic proximity. One recent study, for example, found that PSM is stronger in countries with Anglo-Saxon (means tested) welfare states (Houston 2011). Consistent with this, our results show that the PSM measure reported here has greater shared meaning and scaling among three countries—Australia, the United Kingdom, and the United States—that share many characteristics likely to influence cultural values and dispositions. Although this does not guarantee that the theory or empirical findings regarding the antecedents and consequences of PSM developed in any one of these countries can directly generalize and contribute to our understanding of PSM in the other two, it does suggest that the unit of measurement in these three countries is identical and predictive relationships (i.e., the strength of the relationships between PSM and other variables) can be meaningfully compared. Even in these three countries, however, the lack of evidence of scalar invariance raises legitimate questions regarding the extent to which direct comparisons can be made regarding the prevalence of PSM (i.e., the average or mean on each dimension) among their populations. Future research should continue this line of work to identify populations for which PSM has more of a shared meaning as well as characteristics and existing theories that may explain important cultural differences in the meaning, prevalence, antecedents, and consequences of PSM.

16 Of the eight measures of personal goals and aspirations tested for metric invariance across 15 cultures, only two (self-acceptance and affiliation) met the thresholds for metric invariance on both the $\Delta\chi^2$ and ΔCFI tests. None of the measures of personal goals and aspirations met the thresholds for scalar invariance on both tests.

As is true of any study, the research reported here has a number of limitations that need to be addressed in future research. Perhaps most notable, in addition to relying on only a single sample from each country, our revised PSM measure was also tested only using local government samples in order to maximize the comparability of the samples across countries. The findings and conclusions are based on PSM among local government employees and might be different when surveying central government employees or other specific groups in the public sector. Thus, the potential impact of the sample bias needs to be considered in understanding this study. It also means that the survey items for measuring PSM should be examined with different samples in various areas and at different government levels.

Notwithstanding these cautions and limitations, this study represents an important contribution to our understanding of PSM and the growing emphasis on cross-national research and constitutes a significant step along the way to developing and refining more robust measures of PSM.

FUNDING

National Natural Sciences Foundation of China (Project 71002035); SMC-Chenxing Young Scholar Program (Shanghai Jiao Tong University) awarded to Bangcheng Liu.

REFERENCES

Anderson, James C., and David D. Gerbing. 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin* 103:411–23.

Andolsek, Dana Mesner, and Janez Stebe. 2004. Multinational perspectives on work values and commitment. *International Journal of Cross Cultural Management* 4 (2):181–209.

Bryant, Fred, and Albert Satorra. 2012. Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling* 19 (3):372–98.

Byrne, Barbara M., and Fons J.R. van de Vijver. 2010. Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing* 10:107–32.

Camilleri, Emanuel. 2006. Towards developing an organizational commitment: Public service motivation model for Maltese public service employees. *Public Policy and Administration* 21:63–83.

Castaing, Sébastien. 2006. The effects of psychological contract fulfillment and public service motivation on organizational commitment in the French civil service. *Public Policy and Administration* 21:84–98.

Cerase, Francesco P., and Domenica Farinella. 2009. Public service motivation: How does it relate to management reforms and changes in the working situation of public organisations: A case study of the Italian Revenue Agency. *Public Policy and Administration* 24 (3):281–308.

Chen, Fang Fang. 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling* 14:464–504.

Chen, Fang Fang. 2008. What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology* 95:1005–18.

Cheung, Gordon W., and Roger B. Rensvold. 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling* 9:233–55.

Christensen, Robert K., and Bradley E. Wright. 2009. Public service motivation: Testing measures, antecedents and consequences. Paper presented at the International Public Service Management Research Conference, July 7–10, Bloomington, IN.

Clerkin, Richard M., Sharon R. Paynter, and Jami Kathleen Taylor. 2009. Public service motivation in undergraduate giving and volunteering decisions. *American Review of Public Administration* 39 (6):675–98.

Comrey, Andrew L., and Howard B. Lee. 1992. *A first course in factor analysis*. Hillsdale, NJ: L. Erlbaum Associates.

Coursey, David H., and Sanjay K. Pandey. 2007. Public service motivation measurement: Testing an abridged version of Perry's proposed scale. *Administration & Society* 39:547–68.

Coursey, David H., James L. Perry, Jeffrey L. Brudney, and Laura Littlepage. 2008. Psychometric verification of Perry's public service motivation instrument: Results for volunteer exemplars. *Review of Public Personnel Administration* 28:79–90.

DeHart-Davis, Leisha, Justin Marlowe, and Sanjay K. Pandey. 2006. Gender dimensions of public service motivation. *Public Administration Review* 66:873–87.

Giauque, David, Adrian Ritz, Frédéric Varone, Simon Anderfuhren-Biget, and Christian Waldner. 2011. Putting public service motivation into context: A balance between universalism and particularism. *International Review of Administrative Sciences* 77 (2):227–53.

Grouzet, Frederick M.E., Tim Kasser, Aaron Ahuvia, Jose M. F. Dois, Youngmee Kim, Sing Lau, Richard M. Ryan, Shaun Saunders, Peter Schmuck, and Kennon M. Sheldon. 2005. The structure of goal contents across 15 cultures. *Journal of Personality and Social Psychology* 89 (5):800–16.

Hair, Joseph F., Jr., William C. Black, Barry J. Babin, and Rolph E. Anderson. 2010. *Multivariate data analysis*, 7th ed. Upper Saddle River, NJ: Pearson Education.

Hattrup, Keith, Karsten Mueller, and Pancho Aguirre. 2008. An evaluation of the cross-national generalizability of organizational commitment. *Journal of Occupational and Organizational Psychology* 81:219–40.

Houston, David J. 2011. Implications of occupational locus and focus for public service motivation: Attitudes toward work motives across nations. *Public Administration Review* 71 (5):761–71.

Huang, Xu, and Evert Van de Vliert. 2003. Where intrinsic job satisfaction fails to work: National moderators of intrinsic motivation. *Journal of Organizational Behavior* 24:159–79.

Iacobucci, Dawn. 2010. Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology* 20:90–8.

Jarvis, Cheryl Burke, Scott B. MacKenzie, and Philip M. Podsakoff. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30:199–218.

Jöreskog, Karl G., and Dag Sörbom. 2005. *LISREL 8.72* (Computer Software). Lincolnwood, IL: Scientific Software International.

Kim, Sangmook. 2009a. Revising Perry's measurement scale of public service motivation. *American Review of Public Administration* 39:149–63.

Kim, Sangmook. 2009b. Testing the structure of public service motivation in Korea: A research note. *Journal of Public Administration Research and Theory* 19:839–51.

Kim, Sangmook. 2011. Testing a revised measurement of public service motivation: Reflective versus formative specification. *Journal of Public Administration Research and Theory* 21(3):521–46.

Kim, Sangmook, and Wouter Vandenabeele. 2010. A strategy for building public service motivation research internationally. *Public Administration Review* 70:701–9.

Kline, Rex B. 2005. *Principles and practice of structural equation modeling*, 2nd ed. New York, NY: Guilford Press.

Leibfried, Stephan. 1992. Towards a European welfare state? Integrating poverty regimes into the European community. In *Social policy in a changing Europe*, ed. Zsuzsa Ferge and Jon Eivind Kolberg, 245–79. Boulder, CO: Westview.

Leisink, Peter, and Bram Steijn. 2009. Public service motivation and job performance of public sector employees in the Netherlands. *International Review of Administrative Sciences*, 75:35–52.

Liu, Bangcheng, Chun Hui, Jin Hu, Wensheng Yang, and Xinli Yu. 2011. How well can public service motivation connect with occupational intention? *International Review of Administrative Sciences* 77:191–211.

Liu, Bangcheng, Ningyu Tang, and Xiaomei Zhu. 2008. Public service motivation and job satisfaction in China: An investigation of generalisability and instrumentality. *International Journal of Manpower* 29:684–99.

Liu, Cong, Ingwer Borg, and Paul E. Spector. 2004. Measurement equivalence of the German job satisfaction survey used in a multinational organization: Implications of Schwartz's culture model. *Journal of Applied Psychology* 89(6):1070–82.

Meade, Adam W., Emily Johnson, and Phillip W. Braddy. 2008. Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology* 93:568–92.

Meyer, John P., David J. Stanley, Timothy A. Jackson, Kate J. McInnis, Elyse R. Maltin, and Leah Sheppard. 2012. Affective, normative and continuance commitment levels across cultures: A meta-analysis. *Journal of Vocational Behavior* 80 (2):225–45.

Moynihan, Donald P., and Sanjay K. Pandey. 2007. The role of organizations in fostering public service motivation. *Public Administration Review* 67:40–53.

Moynihan, Donald P., Sanjay K. Pandey, and Bradley E. Wright. 2012. Setting the table: How transformational leadership fosters performance information use. *Journal of Public Administration Research and Theory* 22 (1):143–64.

Olsson, Ulf H., Tron Foss, Sigurd V. Troye, and Roy D. Howell. 2000. The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling* 7 (4):557–95.

Pandey, Sanjay K., and Edmund C. Stazyk. 2008. Antecedents and correlates of public service motivation. In *Motivation in Public Management*, ed. James L. Perry and Annie Hondeghem, 101–17. Oxford, UK: Oxford University Press.

Perry, James L. 1996. Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory* 6:5–22.

Perry, James L. 1997. Antecedents of public service motivation. *Journal of Public Administration Research and Theory* 7:181–97.

Perry, James L., and Annie Hondeghem. 2008. *Motivation in public management*. Oxford, UK: Oxford University Press.

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* 50:367–73.

Rainey, Hal G. 1982. Reward preference among public and private managers: In search of the service ethic. *American Review of Public Administration* 16:288–302.

Raykov, Tenko. 1997. Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement* 21:173–84.

Ritz, Adrian. 2011. Attraction to public policy making: A qualitative inquiry into improvements in PSM measurement. *Public Administration* 89 (3):1128–47.

Ritz, Adrian, and Christian Waldner. 2011. Competing for future leaders. A study of attractiveness of public sector organizations to potential job applicants. *Review of Public Personnel Administration* 31 (3):291–316.

Schmitt, Neil, and Goran Kuljanin. 2008. Measurement invariance: Review of practice and implications. *Human Resource Management Review* 18:210–22.

Siaroff, Alan. 1994. Work, welfare and gender equality: A new typology. In *Gendering welfare states*, ed. Diane Sainsbury, 82–100. London: Sage Publications.

Spector, Paul E. 1992. *Summated rating scales construction: An introduction*. Newbury Park, CA: Sage.

Taylor, Jeannette. 2007. The impact of public service motives on work outcomes in Australia: A comparative multi-dimensional analysis. *Public Administration* 85:931–59.

Vandenabeele, Wouter. 2008a. Government calling: Public service motivation as an element in selecting government as an employer of choice. *Public Administration* 86:1089–105.

Vandenabeele, Wouter. 2008b. Development of a public service motivation measurement scale: Corroborating and extending Perry's measurement instrument. *International Public Management Journal* 11:143–67.

Vandenabeele, Wouter, and Steven Van de Walle. 2008. International differences in public service motivation: Comparing regions across the world. In *Motivation in public management*, ed. James L. Perry and Annie Hondeghem, 223–44. Oxford, UK: Oxford University Press.

Wildeboer Schut, J. M., J. C. Vrooman, and Paul de Beer. 2001. *On worlds of welfare: Institutions and their effects in eleven welfare states*. The Hague: Social and Cultural Planning Office of the Netherlands.

Williams, Larry J., Robert J. Vandenberg, and Jeffrey R. Edwards. 2009. Structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals* 3:543–605.

Wise, Lois Recascino. 2000. The public service culture. In *Public administration concepts and cases*, 7th ed., ed. Richard J. Stillman II, 342–53. Boston: Houghton Mifflin.

Wright, Bradley E. 2008. Methodological challenges associated with public service motivation research. In *Motivation in public management*, ed. James L. Perry and Annie Hondeghem, 80–98. Oxford, UK: Oxford University Press.

Wright, Bradley E., and Sanjay K. Pandey. 2008. Public service motivation and the assumption of person-organization fit: Testing the mediating effect of value congruence. *Administration & Society* 40:502–21.

Yang-Wallentin, Fang, Karl G. Jöreskog, and Hao Luo. 2010. Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling* 17 (3):392–423.