Measuring the construct of public sector creativity: Development of a validated scale

Glenn Houtgraaf | Peter Kruyen | Sandra van Thiel

1Erasmus University Rotterdam, Rotterdam, The Netherlands
2Radboud University, Nijmegen, The Netherlands

Correspondence
Glenn Houtgraaf, Erasmus University Rotterdam, Thomas Morelaan 1, 3062 PA Rotterdam, The Netherlands.
Email: Houtgraaf@essb.eur.nl

Abstract
Public sector creativity—public servants coming up with novel and useful ideas—is the origin of solutions and innovations central to public sector organizations’ ability to optimally serve society’s interests. Despite its relevance and argued limitations, an adequate scale and framework to assess public servants’ creativity remained absent. Using three quantitative (n = 2434; n = 1157; n = 621) and two qualitative datasets, this state-of-the-art preregistered study provides a novel, valid, reliable, comprehensive but succinct multidimensional measurement tool allowing detailed assessment of states and changes of public servants’ creativity by academics and practitioners. To aid academics and practitioners’ understanding of the scale and public sector creativity, it also provides a supplementary theoretical framework in line with the data structure of the scale that compresses and structures extant theory on creativity while emphasizing what is indicated as important in the public sector. The findings shed light on the nature of public sector creativity as an aggregate and multidimensional construct.

Evidence for practice
• The scale provided as a result of this study enables multidimensional measurement of public servants’ creativity and the theoretical framework aids in understanding the scale, data, and construct
• Public servants’ creativity can vary on unipolar dimensions of incrementalism and radicalism (magnitude), reactivism and proactivism (trigger), and realism and idealism (perspective)
• Two categories of creativity appear to exist in the public sector: pragmatic and pioneering
• Public servants appear predominantly “pragmatic” creative, generally scoring high on dimensions of incrementalism, reactivism, and realism
• Room for improvements in terms of public servants’ creativity lays with “pioneering creativity” dimensions of radicalism, proactivism, and idealism

INTRODUCTION
Public sector creativity—“public servants coming up with novel and useful ideas through various practices” (Houtgraaf et al., 2023, p. 3)—is the front-end of the innovation process (Amabile, 1996; Anderson et al., 2014); ideas that individual public servants generate are the origin of innovations and solutions. Public sector creativity as the aggregate of public servants’ creativity is thereby essential to public sector organizations’ performance in terms of innovativeness and responsiveness and, thus, their ability to optimally serve the public interest. Moreover, creativity is associated with public servants’ well-being (Amabile, 1996; Kaufman & Sternberg, 2010; Mumford, 2012; Ryan & Deci, 2017).

Creativity in the public sector, however, appears atypical in nature. Initially, the public sector was assumed hospitable to stereotypical uncreative public servants (Chen &
Bozeman, 2012) and inhospitable to creativity (Bommert, 2010) due to conditions at odds with—or outright hampering—creativity, such as formalization, centralization, and lack of market mechanisms (see, e.g., Boye et al., 2022; Boyne, 2002; Bozeman & Feeney, 2011; Chen & Bozeman, 2012; Houtgraaf et al., 2022; Rainey, 1999). Recent findings provide a more nuanced assessment; public servants are in fact creative, albeit specific in nature, exemplified by the predominant incremental magnitude of ideas, reactive triggering of creativity, and realistic evaluations of ideas (Houtgraaf, 2022; Houtgraaf et al., 2023; Houtgraaf et al., 2022; Kruyen & Van Genugten, 2017).

Thus, adequate assessments of creativity in the public sector require considering the construct as multidimensional. Academics and practitioners aiming for adequate measurements and—and conclusions on—the presence, causes, effects, and (room for) improvements in terms of public servants’ creativity require a comprehensive multidimensional perspective (Rangarajan, 2008) that is able to capture the idiosyncrasies of public sector creativity, instead of being too narrow or unspecific (Houtgraaf et al., 2023; Kruyen & Van Genugten, 2017). Categorizations of generic creativity are abundant (Beghetto & Kaufman, 2007; Kozbelt et al., 2010; Mumford, 2012; Richards, 2007; Unsworth, 2001), but findings on public sector creativity imply three central aspects to consider when assessing public servants’ creativity: triggers, magnitudes, and perspectives (Houtgraaf, 2022; Houtgraaf et al., 2022; Kruyen & Van Genugten, 2017).

A comprehensive but succinct approach and tool that allows academics and practitioners to adequately assess public sector creativity along these lines, however, remained absent. Existing unidimensional scales are succinct, but too narrow, short, and unspecific, failing to provide a comprehensive multidimensional assessment of public sector creativity (Houtgraaf et al., 2023; Rangarajan, 2008). For example, the frequently applied scale by Tierney and Farmer (2004) features only four items, failing to capture multidimensional complexity of creativity as a construct. Capturing this complexity through administering a battery of items from the available pool of scales on creativity (e.g., Richards, 2007; Tierney & Farmer, 2004) and/or scales or tasks on related concepts such as cognitive fluency or divergent/convergent thinking (e.g., Wronsk et al., 2019) may be more comprehensive but not succinct. To capture all dimensions relevant to public sector creativity, the aggregate list of items would be long and cause respondent burden. Moreover, tasks are labor-intensive to both subjects and researchers. Furthermore, these measurements would still not be tailored to validly capture the central dimensions required for adequate assessment of public sector creativity nor embedded in a theoretical framework aiding structured understanding of the resulting data. These shortcomings may facilitate incorrect estimates and claims on creativity in the public sector, for instance, negative assumptions regarding its absence (Bommert, 2010; Chen & Bozeman, 2012).

Therefore, this study’s aim was to provide a valid and reliable, comprehensive but succinct multidimensional measurement tool for adequate assessment of public servants’ creativity supplemented by a corresponding comprehensive but succinct theoretical framework that aids in understanding the scale, the resulting data, and public sector creativity as a construct. The aim was to analyze the data structure in order to (1) develop a validated multidimensional scale by (2) testing the theoretical multidimensional and multi-order framework formulated based on a compression and structuring of extant theory on creativity, emphasizing the central dimensions indicated as important when assessing public sector creativity and (3) subsequently adjust the framework based on the data structure. To these ends, a scale was developed based on data from three waves of large-scale surveys (n = 2434/ n = 1157/n = 621) and two qualitative datasets stemming from public servants working in a range of Dutch public sector organizations. It answered the research question: How can public sector creativity as a multidimensional construct be adequately measured?

The provided scale has practical use because of its comprehensiveness and succinctness. The scale is comprehensive and tailored to the public sector; the multidimensional nuance and broadness allows for extensive and detailed measurements of the important dimensions of public servants’ creativity. Meanwhile, it is succinct, specifically focusing on dimensions central to public sector creativity by means of a relatively small set of concise items, making it easy to use. This results in an adequate measurement tool, enabling academics, public sector managers, practitioners, and trainers to generate detailed assessments of (1) the status quo of public servants’ creativity, (2) where specific room for improvements may lay, (3) effectiveness of interventions aimed at improving public sector creativity, and (4) correlates, causes, and consequences of public sector creativity (Koopmans, 2014). Such measurements may support amelioration of public organizations’ innovativeness and responsiveness and, thus, their ability to optimally serve the public interest.

The measurement scale and corresponding supplementary framework have theoretical relevance, adding to the emerging body of knowledge on public sector creativity. Extant research on the construct was explorative and qualitative, lacking statistical underpinnings (Houtgraaf, 2022; Houtgraaf et al., 2022; Kruyen & Van Genugten, 2017; Rangarajan, 2008). The analysis of the data structure allowed for statistical falsification/corroboration of previous findings arguing for public sector creativity as a multidimensional construct featuring dimensions of magnitude (incrementalism), trigger (reactivism), and perspective (realism) potentially determined by a second-order construct (Houtgraaf, 2022; Houtgraaf et al., 2023; Houtgraaf et al., 2022; Kruyen & Van Genugten, 2017; Rangarajan, 2008). The corresponding supplementary framework aids academics and practitioners in understanding the scale, the resulting data, and
public sector creativity as a construct. The framework compresses and structures extant theory on creativity (e.g., Beghetto & Kaufman, 2007; Kozbelt et al., 2010; Mumford, 2012; Richards, 2007; Unsworth, 2001), emphasizing what is central according to findings on public sector creativity, leading to a framework that facilitates understanding of the construct and measurement tool. Moreover, the framework was further adjusted by aligning it with the data structure that also underlies the measurement scale. Furthermore, analyses of distributions shed light on the nature of public sector creativity as an aggregate, providing statistical evidence for connotations to pessimistic assumptions regarding public servants’ creativity (Bommer, 2010; Chen & Bozeman, 2012) and corroborating theory on its pragmatic nature (Houtgraaf, 2022). More broadly, the study answers the call for state-of-the-art, transparent, validated, and refined items/questionnaires in quantitative research (Flake & Fried, 2020) and in public administration in particular (Lee et al., 2011; Leeuw et al., 2004; Perry, 2012) required for rigorous research, adequate conclusions (DeVellis, 2016; Lee et al., 2011), and replication (Flake & Fried, 2020).

**PUBLIC SECTOR CREATIVITY IN THEORY**

**Creativity in the public sector**

Houtgraaf et al. (2023, p. 3) provided a specific definition of public sector creativity: “public servants coming up with novel and useful ideas through various practices.” Creativity is not synonymous with innovation; they are analytically distinct concepts, with creativity being an integral part of the innovation process (Anderson et al., 2014). Creativity refers to the initial process of coming up with ideas that takes place on the individual level, whereas the innovation process also encompasses subsequent social evaluation and practical implementation on team/organizational levels (Amabile, 1996). Completion of this process results in organizational innovation (Hon & Lui, 2016). Creativity in the public sector, however, appears atypical because of its pragmatic nature notably featuring dimensions of magnitude, trigger, and perspective that arguably correlate (Houtgraaf, 2022; Houtgraaf et al., 2023, 2022; Kruyen & Van Gennugten, 2017). These dimensions and their intercorrelation form the hypothesized framework.

**Magnitude: Incrementalism versus radicalism**

The first dimension deemed central to the assessment of public sector creativity regards the magnitude of ideas (Houtgraaf, 2022, see also Kaufman & Beghetto, 2009). Ideas come in different sizes, generally considered (Beghetto & Kaufman, 2007, see also Houtgraaf, 2022; Kaufman & Beghetto, 2009; Kozbelt et al., 2010; Richards, 2007) ranging from “Pro-C” ideas referring to expressions of creativity breaking with the current way of working (radicalism) to “little-c” ideas referring to incremental and everyday expressions of creativity in the form of adjustments and improvements (incrementalism). This dimension is deemed important when assessing public sector creativity because of public servants’ assumed emphasis on incremental creativity (Rangarajan, 2008), which can be seen as Lindblom’s (1959) classic “muddling through” analysis that development and progress in the public sector is characterized by continuous small and unplanned changes instead of planned and large leaps. Accordingly, Kruyen and Van Gennugten (2017, p. 838) found that public servants’ creativity generally regards “deviating sensibly from the common repertoire.” Indeed, Houtgraaf (2022) indicated that public servants’ ideas predominantly regard incremental adjustments of existing products and processes as opposed to ideas for more radical products and processes. Thus, public sector creativity appears to particularly provide soil for what is coined exploitative innovative (Barrutia & Echebarria, 2019; Jansen et al., 2006) or optimization (Gieske et al., 2018; Gieske et al., 2020), the implementation of small adjustments based on public servants’ incremental ideas. Correspondingly, Osborne and Brown (2013) argue that government innovation strategies portray innovation as continuous improvement rather than discontinuity. Because of the notable characterization of public servants’ creativity in terms of magnitude of ideas, a scale on public sector creativity ought to cover this dimension in order to adequately measure its nature.

H1. Public sector creativity features a dimension of magnitude of ideas ranging from incrementalism to radicalism.

**Trigger: Reactivism versus proactivism**

A second dimension deemed central to the assessment of public sector creativity regards the trigger of creative processes. Although extensive theorizing exists on initiation of creativity (Puccio & Cabra, 2010), a dichotomic stratification of triggers for (public sector creativity) is commonly applied to differentiative (Houtgraaf, 2022; Houtgraaf et al., 2023; Rangarajan, 2008; Unsworth, 2001) responsively coming up with ideas as solutions to occurring situations (reactivism) or prospectively coming up with ideas for open and undiscovered situations (proactivism). This dimension is deemed important when assessing public sector creativity because of public servants’ assumed emphasis on reactive creativity (Rangarajan, 2008), illustrated by Houtgraaf et al. (2022) findings that approximately 80 percent of public servants’ ideas were triggered reactively. Accordingly, Kruyen and Van Gennugten (2017, p. 837) outline that public servants mainly define creativity as “trying out new things to deal better with specific problems at hand.” Because of the notable characterization of public servants’ creativity in terms of reactive initiation, a scale on public sector creativity ought...
to cover this dimension in order to adequately measure its nature.

**H2.** Public sector creativity can be measured on the dimension of triggers of ideas ranging from reactivism to proactivism.

**Perspective: Realism versus idealism**

A third dimension deemed central to the assessment of public sector creativity regards the perspective for evaluating ideas. Per definition, ideas require aspects of “usefulness” and “novelty” to be creative (Amabile, 1996; Houtgraaf et al., 2023). However, evaluation of ideas may emphasize (Houtgraaf et al., 2022) either the ideas’ originality (idealism) or the ideas’ usefulness and feasibility (realism). This dimension is deemed important when assessing public sector creativity because explorations of public servants’ creativity indicate that public servants’ perception of ideas as useful and feasible is a salient driver for creative processes, whereas originality plays a much less significant role. Thus, whereas creativity is generally defined in terms of ideas that are both “novel” and “useful,” the emphasis in the case of public sector creativity appears to lay in the latter rather than the former, indicating a realistic view (Houtgraaf et al., 2022). Accordingly, Kruyen and Van Genugten’s (2017, p. 837) initial exploration of public sector creativity indicated that public sector practitioners “are not looking for original ideas to innovate, but for useful solutions.” Because of the notable characterization of public servants’ evaluation of creativity, a scale on public sector creativity ought to cover this dimension in order to adequately measure its nature.

**H3.** Public sector creativity can be measured on the dimension of perspective on creativity ranging from realism to idealism.

Findings relating to the hypothesized dimensions of public sector creativity indicate a pattern of interrelatedness implying the potential of an underlying second-order construct named “nature of creativity” (see1 for information on the renaming in comparison with the preregistered name of the second-order construct). Explorative findings consistently point toward an atypical nature of public sector creativity with public servants predominantly coming up with feasible (“perspective”) solutions (“magnitude”) reactively (“trigger”), as opposed to original (“perspective”) innovations (“magnitude”) proactively (“trigger”) (Houtgraaf, 2022; Houtgraaf et al., 2022; Kruyen & Van Genugten, 2017; Rangarajan, 2008). With public sector creativity as an aggregate, the positions on the separate dimensions are presented as correlating. To illustrate, Kruyen and Van Genugten’s (2017) quote that public sector practitioners “are not looking for original ideas to innovate, but for useful solutions” encapsulates “reactivism,” “incrementalism,” and “realism” in a correlated fashion. We argue that this implies the possibility of an overarching dimension with two poles ranging from pragmatic to pioneering creativity, a second-order construct named “nature of creativity” underlying correlations between the scores on separate dimensions of “magnitude,” “trigger,” and “perspective.” This second-order construct would imply interfactorial correlations between the scores on the dimensions that follow a structured pattern as opposed to different configurations. This would mean that if a public servant scores low on one dimension, he/she is likely to also score lower on other dimensions, as scores would be determined by the second-order construct “nature of creativity.” The complete hypothesized theoretical structure of the formative model is outlined in Figure 1.

**H4.** Public sector creativity features three dimensions in terms of “magnitude,” “trigger,” and “perspective” that are caused by a second-order construct “nature of creativity.”

**METHODS**

**Overarching methods**

A multiwave approach is applied, sourcing data from three cross-sectional surveys (n = 2434/n = 1157/n = 621) and two qualitative studies (n = 142/n = 54). Five-point Likert scales were used. If applicable, frequency ratings were preferred over agreement ratings in response options as agreement ratings generally require individuals to rate likelihood of engaging in a behavior and may assess attitude toward the behavior rather than actual behavior (Dalal, 2005; Koopmans, 2014). Moreover, frequency ratings require individuals to recall and mentally calculate how often one engaged in a behavior (Koopmans, 2014; Schwarz & Oyserman, 2001) and were therefore considered to be more valid. Items were randomized within blocks featuring items from multiple dimensions to mitigate context effects such as question order and priming of attitudes (Hjortskov, 2017; Van de Walle & Van Ryzin, 2011).

Items were generated in Dutch, validated in Dutch questionnaires, and finally translated into English. Based on the qualitative data, an initial items pool was generated based on sets of items per dimension. Afterward, the corresponding scale was constructed, validated, and shortened based on quantitative data. The procedure encompassed an exploratory factor analysis and two confirmatory factor analyses as this allowed for stepwise selection and validation of items to eliminate bad items and come to an optimally performing set of good items for the scale (DeVellis, 2016). Moreover, it allowed for assessment of configurational equivalence, testing factorial structure across populations under investigation (Jilke et al., 2015). Summarized, the approach was divided into five phases (see Figure 2).
FIGURE 1  Hypothesized formative data structure model for measuring public sector creativity as a construct.

Phase 1. Exploration and definition of public servants’ creativity
Generating item pool  Sept. 2021

Phase 2. Expert reviews and pilots item pool
Test and refine item pool  Feb. 2022

Phase 3. Binnenlands Bestuur data Explorative Factor Analysis
Explore and determine data structure, refine item pool  Mar. 2022

Phase 4. Flitspanel data Confirmatory Factor Analysis
Test theoretical multidimensional model fit to data, short scaling  Mar. 2022

Phase 5. LISS panel data Confirmatory Factor Analysis
Test configural equivalence. LISS panel data Confirmatory Factor Analysis  Oct. 2022

FIGURE 2  Five-phased process of scale development.
Phase 1: Qualitative exploration and definition of public sector creativity

An initial item pool was generated through exploration of public servants’ diaries on public sector creativity and expert reviews. From March 2020 to March 2021, longitudinal qualitative diary data were collected from 142 public servants working in four Dutch public executive agencies operating at arm’s length of central government (Van Thiel, 2004) facilitating and/or providing diverse public services (e.g., background IT services, public service delivery to citizens) spread across the country, resulting in 562 diaries (see also Houtgraaf, 2022; Houtgraaf et al., 2022). Organizations voluntarily applied to partake, based on a call set out within the Dutch “Network Public Executive Agencies,” whereby respondents were recruited by the organizations themselves from multiple teams through management selection and voluntary enrollment. Demographics of the diary sample are given as follows: gender (57.75 percent male, 42.25 percent female); age ($M = 46.38$, $SD = 7.32$); education (4.26 percent basic education, 25.56 percent intermediate vocational education, 42.60 percent higher education, and 27.58 percent academic education); and types of public servants (84.38 percent policy implementer, 15.62 percent policy maker).

Data were explored with a threefold purpose: (1) identify dimensions and subdimensions of public sector creativity (Houtgraaf, 2022; Houtgraaf et al., 2022; Kruyen & Van Genugten, 2017; Rangarajan, 2008), (2) identify keywords for indicators, and (3) become accustomed to language and terms related to public sector creativity. Basing the formulation of our items on this information improved authenticity and readability, thereby benefitting the validity of the scales (Fowler Jr, 2015). A long list of items was formulated corresponding to this exploration and prior exploratory research on public sector creativity. The first author formulated items, subsequently reviewed by the other authors. Additionally, two external researchers on public sector creativity were asked to fit an unstructured long list of the items with dimensions to check for independent rater reliability (see also Furr, 2011; Smith & Kendall, 1963). Subsequently, these researchers conducted an expert review of the pilot questionnaire for cognitive testing of the items in terms of clarity, readability, and face validity (Fowler, 2015; Koopmans, 2014). A high degree of consensus existed between the first author and other researchers, leading to minor adjustments in terms of formulation, addition, and stratification of items (see Supporting Information Appendix A, operationalization matrices).

Phase 2: Expert reviews and pilot of initial item pool

The initial item pool was refined by a member check and pilot. For the expert review, a convenience sample of 49 public servants from the same four public agencies was asked to test the questionnaire in five separate focus group sessions and four additional public servants from a county administration that were readily available in another focus group session for an additional perspective. Another convenience sample of 21 knowledge workers and academics was asked to pilot the questionnaire with qualitative feedback (Fowler, 2015).

Expert reviewers reported on issues of logical consistency, ease of understanding, uniqueness, selection, sequence, theoretical fit of items, and contextual relevance (Fowler Jr, 2015). Adjustments were made to the formulation of items and excluded items that were regarded as superfluous or vague (see Supporting Information Appendix A, operationalization matrix 3). Feedback per item was noted to serve as background information informing choices for item selection and short-scaling later. For example, this led to items per dimension: magnitude: “In my work I come up with small changes to existing products/services/processes” or “In my work I come up with unconventional ideas for products/services/processes”; trigger: “In my work I come up with solutions for the problems that arise” or “In my work I come up with ideas by proactively looking for points of improvement”; perspective: “Work related ideas have to be useful” or “Work related ideas have to be original.”

Phase 3: Exploratory factor analysis of “Binnenlands Bestuur” data

An exploratory factor analysis was conducted on data from 2434 public servants collected in March 2022. Respondents stemmed from subscriber pool of approximately 6000 public servants of Binnenlands Bestuur, the Dutch online platform/magazine for (mainly higher educated) public servants and administrators featuring news, opinion pieces/background articles, conferences, and job vacancies. Subscribers of Binnenlands Bestuur newsletter were emailed a request to fill out the questionnaire, which they could voluntarily fill out by self-selection without payment. Demographics of the Binnenlands Bestuur sample are given as follows: gender (54.25 percent male, 45.37 percent female, 0.38 percent other); age ($M = 56.12$, $SD = 9.43$); education (2.97 percent intermediate vocational education, 48.87 percent higher education, and 48.17 percent academic education); and types of public servants (21.62 percent policy implementer and 78.38 percent policy maker).

Data accuracy and missing data were checked. Outliers were removed based on Mahalanobis cutoff values. Factor analysis was conducted to determine the data structure. Screeplot eigenvalues were assessed and checked for points of inflection. Moreover, parallel analysis was conducted to check whether the analyses suggested the same number of factors based on the actual and resampled data (DeVellis, 2016). Kaiser’s old and new criteria—respectively, 1 and 0.7—of the eigenvalues were checked but only used this as backdrop information as the use of Kaiser criterion is under debate (DeVellis, 2016). Loading table was checked for simple structure. Item loading should be higher than 0.3 and should not cross load 0.3 or higher on multiple factors;
items violating these rules were deleted accordingly. Fit indices were checked for the simple structure model in terms of chi-square that should be nonsignificant (rarely happens with larger samples), standardized root mean square residual (SRMR) and mean square error of approximation (RMSEA) that should be below .05, and TLI and CFI that should be above .90. Reliability of identified factors was checked using coefficient alpha, for which score 0.70 was deemed an acceptable lower bound (DeVellis, 2016). Sufficient alpha scores indicate possibility to effectively selected/eliminate items based on weak internal consistency/interitem correlations, the theoretical spectrum coverage of the items, and the qualitative feedback to achieve item redundancy (DeVellis, 2016).

**Phase 4: Confirmatory factor analysis of “FlitsPanel” data**

Confirmatory factor analysis was conducted on data from 1157 respondents from the Flitspanel survey collected in March 2022. Respondents stemmed from panel pool of approximately 10,000 public sector workers of Flitspanel, a Dutch panel to which public sector employees can voluntarily apply to receive requests to fill out questionnaires on human resource management (HRM) policies. Approximately 1500 panel members were selected randomly from the total pool and were emailed a request by Flitspanel with the questionnaire, which they could voluntarily fill out by self-selection. Respondents were paid for completing the questionnaire. Demographics of the Flitspanel sample were as follows: gender (67.60 percent male, 32.40 percent female); age (M = 55.91, SD = 8.17); education (9.89 percent basic education, 14.09 percent intermediate vocational education, 40.75 percent higher education, and 35.27 percent academic education); and types of public servants (27.95 percent policy maker, 40.75 percent higher education, and 35.27 percent academic policy implementer, 67.27 percent policy maker, and 4.78 percent both).

Missing data were omitted, and outliers were identified based on Mahalanobis cutoff value. Data were checked for accuracy and additional assumptions, such as additivity, linearity based, normality, homogeneity based, correlation adequacy based on Bartlett’s test of sphericity, and sampling adequacy based on Kaiser–Maier–Olkin values. The correlation data frame was constructed, the model was determined, and latent factors/scales were computed by combining the manifest variables following the proposed simple structure based on the exploratory factor analysis (EFA) results and suggested by our theoretical hypotheses.

Model fit was tested using SEM. Chi-square test is known to be overly sensitive with larger samples (300+), meaning that it reaches unfavorable statistical significance for very trivial model changes (Antonakis et al., 2010; Jilke et al., 2015; Kline, 2011, p. 201). Nevertheless, it remains a test to detect a misspecified model, and the test of fit should be considered in this regard (Antonakis et al., 2010). Other recommended goodness-of-fit measures were applied, namely, the RMSEA that should be at least below .050 with a p value as high as possible, the SRMR that should be at least below .050, the Tucker–Lewis index (TLI) that should be at least above .900, and the comparative fit index (CFI) that should be at least above .900 (Antonakis et al., 2010; Jilke et al., 2015), and also compared models based on expected cross-validation index (ECVI) and Akaike Information Criterion (AIC), for which a lower score indicated a relatively better fit. Correlation matrices and modification indices were checked for additions that lead to improvements in model fit and ran ANOVAs between the base model and modified model to analyze whether these changes were significant improvements. It was checked whether adding the second-order construct led to acceptable outcomes in terms of structure and fit indices. Here too, reliability of the resulting latent factors in the model was checked using coefficient alpha using the same procedure and reasoning.

The nomological network was assessed to test construct validity following guidelines on scale development by DeVellis (2016). We checked for positive correlation with a generic creativity scale because this is a construct that should be most similar (convergent validity, see Tierney & Farmer, 2004), positive correlation with intrinsic motivation as this is deemed the main theoretical antecedent (criterion validity, see Amabile, 1996), negative correlation with formalization as this is deemed a central inhibiting factor (criterion validity, see Hirst et al., 2011), no correlation with gender as assumed unrelated factor (criterion validity), and no correlation with social desirability as important control factor for scale validity (form of discriminant validity, should not correlate—especially not strongly—as this distorts self-report measurements, see Koopmans, 2014; Van der Heijden & Nijhof, 2004) and doing so for multiple dimensions (form of content validity). As such, the nomological network tested construct validity in a variety of ways to analyze whether the scale indeed statistically correlates with other factors as it should, theoretically speaking, which is an indication that the scale measures the concept it claims to represent (see2 for more information on additional constructs for the nomological network).

**Phase 5: Confirmatory factor analysis of “LISS panel” data**

A second confirmatory factor analysis was conducted on data from 621 private sector employees for construct validity through configural equivalence (Jilke et al., 2015) analyzing model fit results for the exact same model on data from another population—in this case, private sector instead of public sector employees. Data were collected in October 2022. Respondents stemmed from the LISS panel pool of approximately 5000 Dutch households across country that are randomly selected by the Dutch Central Bureau of Statistics. Selected respondents may opt in to voluntarily fill out the questionnaire by self-selection as response to the requests sent out by LISS.
Respondents were paid for completed questionnaires. This panel is different from the other panels as it only includes private sector employees. To ensure that all respondents worked for a private sector business or company, filtering public sector employees or unemployed, a control question was implemented at the start of the questionnaire. Demographics of the LISS panel sample were as follows: gender (45.50 percent male, 54.50 percent female); age ($M = 44.67, SD = 12.56$); and education (7.51 percent basic education, 55.50 percent intermediate vocational education, 22.49 percent higher education, and 14.50 percent academic education).

The same procedure as with the first confirmatory factor analysis (CFA) was followed, although without making any adjustments based on modification indices. The fit indices and coefficient alpha values were compared to assess whether the same factorial model also indicates validity and reliability when applied to data stemming from private sector employees.

**Ethics**

Ethical approval was provided by both Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO, funder) and the host university. Informed consent forms were required, following ethical guidelines for academic research as outlined by Emanuel et al. (2000). The design was preregistered to clarify the distinction between planned and unplanned steps in the research and thereby reduce unnoticed flexibility to increase transparency of planned and unplanned steps in the research and thereby reduce unnoticed flexibility to increase transparency of practice, credibility of findings, and calibration of uncertainty (Hansen & Tummers, 2020; Nosek et al., 2019; see link to preregistration). For the analyses, R is applied because of its flexibility and transparency in code for replication (Flake & Fried, 2020).

**RESULTS**

**Initial item pool**

The initial item pool featured 34 items with six to eight items per dimension. Feedback by expert reviewers and practitioners extended the item pool to 43 items. Items were added to cover a broader range of aspects concerning the dimensions, and item formulations were altered based to increase readability (see Supporting Information Appendix A for full procedure per step in matrices).

**Exploratory factor analysis of “Binnenlands Bestuur” data**

EFA was conducted on Binnenlands Bestuur data ($n = 2434$). All 554 respondents with missing data were excluded, and 157 outliers were removed, leaving 1723 respondents. Bartlett’s test indicated correlation adequacy chi-square ($1723$) = 45349.35 with $p < .01$, and Kaiser-Meyer-Olkin (KMO) test indicated sampling adequacy, measurement system analysis (MSA) = 0.96.

Parallel analysis and screeplot examination suggested six-factor model in line with theory. Kaiser criterion suggested a five-factor model. Maximum-likelihood estimation was used with Direct Oblimin (Oblique) rotation because of expected factor (dimension) correlation (DeVellis, 2016; Houtgraaf, 2022). In the six-factor model, after testing all 43 items, one item split across two factors (QC4_12) and one item had mediocre loading (QC4_8); both items were removed. The five-factor model was tested, but this model provided four cross loading items, and three nearly cross loading items and lower loadings. The six-factor model was pursued, where after deletion of the two items mentioned, simple structure was achieved. This model had excellent fit: RMSEA = 0.03, RSMR = 0.01, TLI = 0.97 and CFI = 0.98, and chi-square value $p < .01$ (see Supporting Information Appendix A for factor loading table and R-code).

Factor 1 included six items measuring the dimension “incrementalism” as part of “magnitude” with items as “In my work I come up with ideas in the form of small steps.” Factor 2 included seven items measuring the dimension “radicalism” as part of “magnitude” with questions as “In my work I come up with unconventional ideas”. Factor 3 included seven items measuring the dimension “reactivism” as part of “trigger” with items as “In my work I come up with solutions to problems that arise.” Factor 4 included seven items measuring the dimension “proactivism” as part of “trigger” with items as “In my work I come up with ideas through actively looking forward.” Factor 5 included seven items measuring the dimension “realism” as part of “perspective” with items as “Work-related ideas have to be useful.” Factor 6 included seven items measuring the dimension “idealism” as part of “perspective” with items as “Work-related ideas have to be original.” Exploratory factor analysis indicated that dimensions were unipolar, hence separate dimensions of radical/incremental, reactive/proactive, and realism/idealism. This means that if the dimensions’ (magnitude, trigger, perspective) items are placed on a measurement continuum, their endpoints do not pertain to each other’s opposite—as with a bipolar dimension—but to “different degrees of the presence of the same attribute” (Keulemans & Van de Walle, 2020; Schwarz, 2008, p. 43). Reliability of all six factors was excellent, with raw alpha values of $\alpha = .91$ (ni = 6), $\alpha = .91$ (ni = 7), $\alpha = .89$ (ni = 7), $\alpha = .91$ (ni = 7), $\alpha = .90$ (ni = 7), and $\alpha = .90$ (ni = 7). To achieve item redundancy in context of parsimony, six best performing items for each scale were selected based on qualitative feedback as all quantitative values were excellent. Table 1 indicates the EFA results.

**Confirmatory factors analysis of “Flitspanel” data**

CFA was conducted on Flitspanel data ($n = 1157$). Respondents not classifying as public servants were
**Table 1: EFA factor loadings.**

<table>
<thead>
<tr>
<th>Question</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
<th>Factor 6</th>
<th>H2</th>
<th>u2</th>
<th>Com</th>
<th>Possible reason for exclusion/short-scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work related ideas have to be…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.77</td>
<td>0.02</td>
<td>0.01</td>
<td>0.59</td>
<td>0.41</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Usable</td>
<td>0.01</td>
<td>−0.02</td>
<td>0.02</td>
<td>0.83</td>
<td>0.00</td>
<td>0.00</td>
<td>0.69</td>
<td>0.31</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Executable</td>
<td>−0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.83</td>
<td>−0.04</td>
<td>−0.01</td>
<td>0.68</td>
<td>0.32</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Feasible</td>
<td>0.00</td>
<td>−0.04</td>
<td>0.04</td>
<td>0.78</td>
<td>−0.01</td>
<td>−0.02</td>
<td>0.60</td>
<td>0.40</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Effective</td>
<td>0.02</td>
<td>0.03</td>
<td>−0.03</td>
<td>0.72</td>
<td>0.01</td>
<td>0.01</td>
<td>0.54</td>
<td>0.46</td>
<td>1.0</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Practical</td>
<td>0.05</td>
<td>−0.07</td>
<td>0.00</td>
<td>0.73</td>
<td>0.05</td>
<td>−0.02</td>
<td>0.54</td>
<td>0.46</td>
<td>1.0</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Relevant</td>
<td>−0.03</td>
<td>0.08</td>
<td>−0.03</td>
<td>0.73</td>
<td>−0.02</td>
<td>0.03</td>
<td>0.55</td>
<td>0.45</td>
<td>1.0</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Original</td>
<td>0.02</td>
<td>0.02</td>
<td>−0.06</td>
<td>−0.03</td>
<td>0.79</td>
<td>0.01</td>
<td>0.60</td>
<td>0.40</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Remarkable</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
<td>−0.04</td>
<td>0.76</td>
<td>−0.01</td>
<td>0.58</td>
<td>0.42</td>
<td>1.0</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Ambitious</td>
<td>−0.03</td>
<td>0.17</td>
<td>−0.06</td>
<td>0.06</td>
<td>0.55</td>
<td>0.04</td>
<td>0.36</td>
<td>0.64</td>
<td>1.3</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Unique</td>
<td>0.02</td>
<td>−0.07</td>
<td>0.05</td>
<td>−0.04</td>
<td>0.81</td>
<td>−0.02</td>
<td>0.66</td>
<td>0.34</td>
<td>1.0</td>
<td>Plot low frequency on value 5</td>
</tr>
<tr>
<td>Innovative</td>
<td>−0.08</td>
<td>0.08</td>
<td>−0.05</td>
<td>0.12</td>
<td>0.70</td>
<td>0.07</td>
<td>0.51</td>
<td>0.49</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>0.03</td>
<td>−0.04</td>
<td>−0.02</td>
<td>−0.01</td>
<td>0.74</td>
<td>0.00</td>
<td>0.53</td>
<td>0.47</td>
<td>1.0</td>
<td>Expert feedback on relatively obvious question</td>
</tr>
<tr>
<td>Ground-breaking</td>
<td>−0.02</td>
<td>−0.01</td>
<td>0.10</td>
<td>0.01</td>
<td>0.74</td>
<td>−0.05</td>
<td>0.59</td>
<td>0.41</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>During my work I come up with…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ways for slightly adjusting processes/services/products</td>
<td>0.03</td>
<td>−0.02</td>
<td>0.07</td>
<td>0.01</td>
<td>0.03</td>
<td>0.76</td>
<td>0.65</td>
<td>0.35</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Ways for developing existing processes/services/products step by step</td>
<td>−0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.00</td>
<td>0.01</td>
<td>0.75</td>
<td>0.61</td>
<td>0.39</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Small changes to existing processes/services/products</td>
<td>0.02</td>
<td>−0.04</td>
<td>−0.05</td>
<td>−0.01</td>
<td>0.01</td>
<td>0.84</td>
<td>0.67</td>
<td>0.33</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Adjustments to existing processes/services/products that fit the current way of working</td>
<td>0.07</td>
<td>−0.01</td>
<td>0.10</td>
<td>0.02</td>
<td>−0.03</td>
<td>0.66</td>
<td>0.56</td>
<td>0.44</td>
<td>1.1</td>
<td>Longer question and overlap with QC3_1</td>
</tr>
<tr>
<td>Subtle improvements to existing processes/services/products</td>
<td>0.05</td>
<td>−0.06</td>
<td>−0.01</td>
<td>0.00</td>
<td>−0.03</td>
<td>0.79</td>
<td>0.63</td>
<td>0.37</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Ideas in the form of small steps</td>
<td>0.02</td>
<td>0.15</td>
<td>−0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.66</td>
<td>0.55</td>
<td>0.45</td>
<td>1.1</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Completely new processes/services/products</td>
<td>−0.02</td>
<td>0.01</td>
<td>0.78</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.62</td>
<td>0.38</td>
<td>1.0</td>
<td>Harder to interpret and theoretical overlap with QC3_10</td>
</tr>
<tr>
<td>Unconventional processes/services/products</td>
<td>0.08</td>
<td>0.03</td>
<td>0.72</td>
<td>−0.03</td>
<td>0.01</td>
<td>−0.06</td>
<td>0.57</td>
<td>0.43</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Factor 1</td>
<td>Factor 2</td>
<td>Factor 3</td>
<td>Factor 4</td>
<td>Factor 5</td>
<td>Factor 6</td>
<td>H2</td>
<td>u2</td>
<td>Com</td>
<td>Possible reason for exclusion/short-scaling</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Processes/services/products that break with the current way of working</td>
<td>0.01</td>
<td>0.05</td>
<td>0.67</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.12</td>
<td>0.57</td>
<td>0.43</td>
<td>1.1</td>
<td>Harder to interpret and theoretical overlap with QC3_10</td>
</tr>
<tr>
<td>Processes/services/products that are completely different from previously used</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.80</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
<td>0.66</td>
<td>0.34</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Original processes/services/products</td>
<td>-0.03</td>
<td>0.12</td>
<td>0.59</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.15</td>
<td>0.56</td>
<td>0.44</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Processes/services/products that strongly differ from their predecessors</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.79</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.08</td>
<td>0.68</td>
<td>0.32</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Radical ideas</td>
<td>0.07</td>
<td>0.03</td>
<td>0.75</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.15</td>
<td>0.56</td>
<td>0.44</td>
<td>1.1</td>
<td>Expert feedback on vague question; mean is off; low frequency max; skewed left</td>
</tr>
<tr>
<td>During my work I come up with...</td>
<td>0.81</td>
<td>-0.05</td>
<td>0.01</td>
<td>0.08</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.63</td>
<td>0.37</td>
<td>1.0</td>
<td>Skewed right so, says less</td>
</tr>
<tr>
<td>Solutions to problems that occur</td>
<td>0.57</td>
<td>0.16</td>
<td>0.03</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.51</td>
<td>0.49</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Ways for dealing with obstacles and demands</td>
<td>0.71</td>
<td>0.14</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.64</td>
<td>0.36</td>
<td>1.1</td>
<td>Expert practitioner feedback on relatively vague question</td>
</tr>
<tr>
<td>Ways for solving problems</td>
<td>0.64</td>
<td>0.10</td>
<td>0.00</td>
<td>0.06</td>
<td>0.03</td>
<td>0.09</td>
<td>0.62</td>
<td>0.38</td>
<td>1.1</td>
<td>Theoretical overlap with higher loading QC4_1; skewed right so, says less</td>
</tr>
<tr>
<td>Ideas based on the things that I hear and see around me</td>
<td>0.60</td>
<td>0.19</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.59</td>
<td>0.41</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Ideas based on the situations that occur during work</td>
<td>0.73</td>
<td>0.00</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.60</td>
<td>0.40</td>
<td>1.0</td>
<td>Overlap with higher loading QC4_7</td>
</tr>
<tr>
<td>Ideas as reactions to situations that occur during work</td>
<td>0.78</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.05</td>
<td>0.58</td>
<td>0.42</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Ideas reactively</td>
<td>0.44</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.05</td>
<td>0.24</td>
<td>0.76</td>
<td>1.1</td>
<td>Mediocre loading, off from other items: mean is off and normally distributed</td>
</tr>
<tr>
<td>Ideas by proactively looking forward</td>
<td>0.07</td>
<td>0.76</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.05</td>
<td>0.69</td>
<td>0.31</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Ideas by anticipating on the future</td>
<td>0.11</td>
<td>0.70</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.08</td>
<td>0.66</td>
<td>0.34</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Ideas by proactively looking for points of improvement</td>
<td>0.04</td>
<td>0.61</td>
<td>0.13</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.58</td>
<td>0.42</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Ideas to prevent future problems</td>
<td>0.34</td>
<td>0.47</td>
<td>-0.03</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.06</td>
<td>0.58</td>
<td>0.42</td>
<td>1.9</td>
<td>Cross loading factor 1 and factor 2</td>
</tr>
<tr>
<td>Ideas to create opportunities</td>
<td>0.09</td>
<td>0.68</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.62</td>
<td>0.38</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Ideas to get ahead of other organizations</td>
<td>-0.09</td>
<td>0.55</td>
<td>0.24</td>
<td>-0.02</td>
<td>0.13</td>
<td>-0.06</td>
<td>0.50</td>
<td>0.50</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Ideas to gain an advantage</td>
<td>-0.09</td>
<td>0.65</td>
<td>0.14</td>
<td>-0.03</td>
<td>0.11</td>
<td>-0.02</td>
<td>0.53</td>
<td>0.47</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Ideas proactively</td>
<td>0.12</td>
<td>0.69</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.62</td>
<td>0.38</td>
<td>1.1</td>
<td>Expert practitioner feedback on relatively vague and off question</td>
</tr>
</tbody>
</table>
excluded, leaving a sample of \( n = 950 \). The 66 identified outliers were not removed because they are actual observations with higher degrees of variance that are to be included in further considerations and results pertaining to further scale development. Distributions on latent variables “radicalism,” “idealism,” and “proactiveness” were normally distributed, whereas “realism,” “incrementalism,” and “reactivism” were right-skewed. No issues occurred regarding additivity, linearity, or homoscedasticity. Bartlett’s test indicated correlation adequacy, chi-square (666) = 18812.29 with \( p < .01 \), and the KMO test indicated sampling adequacy, MSA = 0.94.

SEM was conducted. Overall fit of the base model was acceptable: CFI = .920, TLI = .913, RMSEA = .052, SRMR = .046, ECVI = 2.346, AIC = 73264.764, and chi-square \( p < .001 \). Correlation matrix and modification indices indicated three other models to be tested. Model 2 deleted three items that cross loaded. Model 3 deleted these three items and reduced factorial structure to five instead of six factors. Model 4 processed modification indices based on largest significant reductions in chi-square, removing items based on these fit indices while considering theoretically nonredundant items and Cronbach’s alpha values. The optimal model in terms of fit indices was Model 4: CFI = .963, TLI = .956, RMSEA = .042, \( p = 1 \), SRMR = .039, ECVI = .728, AIC = 499088.473, and chi-square \( p < .001 \). No additional parameters, such as correlated error terms, were added to the model as adding extra parameters inflates the fit indices, thereby overfitting the model. Reliability of all items pertaining to the factor scales in Model 4 was very good to excellent, with raw alpha values of \( \alpha = .81 \) (\( \text{ni} = 4 \)), \( \alpha = .82 \) (\( \text{ni} = 4 \)), \( \alpha = .84 \) (\( \text{ni} = 4 \)), \( \alpha = .87 \) (\( \text{ni} = 4 \)), \( \alpha = .74 \) (\( \text{ni} = 4 \)), and \( \alpha = .76 \) (\( \text{ni} = 3 \)). Interfactor Pearson R correlations between dimensions were generally less than .150 and, therefore, did not imply a second-order construct.

Results for the nomological network are presented in Table 2. As expected, results indicate convergent validity of the dimensions with the “general creativity” scale, except from “realism” as this dimension theoretically captures a pragmatic perspective on ideas. Moreover, as expected, results indicate divergent validity in relation to “gender,” although contrary to expectations, this does not hold for “incrementalism” and “reactivism,” which might be explained theoretically by gender differences because these dimensions focus on small, reactive solutions. As expected, nomological results indicate significant correlations with “formalization” as a significant hampering condition, with correlations with the “realism” dimension in opposite fashion; logically being confronted with more stricter rules/procedures correlates positively with a pragmatic perspective focused on feasibility and usability. As expected, nomological results indicate “intrinsic motivation” as a significant antecedent, although for “intrinsic motivation,” this does apply to “realism” for the same reason as with the generic creativity scale. The dimensions indicate no significant correlations with “social desirability,” except from “realism,” possibly because both are related to conformity to external expectations. Notably, all cases of unexpected results regarded dimensions pertaining to “pragmatic creativity,” possibly in part due to the little variation as a result of the severe right-skewness of the distributions on these dimensions.

**Confirmatory factor analysis of “LISS panel” data**

CFA was conducted on LISS panel data (\( n = 621 \)). The 25 identified outliers were not removed from further analyses for reasons consistent with those in the first CFA. Histograms indicated normal distribution with slight right skewness on standardized data, with normal distributions on latent variables “radicalism,” “idealism,” and “proactiveness,” whereas “realism,” “incrementalism,” and “reactivism” were right-skewed. No issues occurred regarding additivity, linearity, or homoscedasticity. Bartlett’s test indicated correlation adequacy chi-square (300) = 9067.29 with \( p < .01 \), and the KMO test indicated sampling adequacy, MSA = 0.93.

SEM was conducted based on the same model as the first SEM, and the fit indices were compared. Overall fit of the model was excellent: CFI = .976, TLI = .971, RMSEA = .040, SRMR = .037, ECVI = .891, AIC = 32693.150, and chi-square \( p < .001 \). Reliability of items pertaining to the factor scales was excellent, with raw alpha values of \( \alpha = .85 \) (\( \text{ni} = 4 \)), \( \alpha = .81 \) (\( \text{ni} = 4 \)), \( \alpha = .91 \) (\( \text{ni} = 4 \)), \( \alpha = .90 \) (\( \text{ni} = 4 \)), \( \alpha = .88 \) (\( \text{ni} = 4 \)), and \( \alpha = .85 \) (\( \text{ni} = 3 \)). The framework of the construct is structured identically in both populations enabling comparisons using the scale. Moreover, preliminary analyses of differences in distributions on the dimensions show differences in the nature of creativity between the public and private sectors indicated by more right-skewed distributions for “proactiveness,” “radicalism,” and “idealism” in the private sector population, which is an avenue for further research.

**DISCUSSION**

**Public sector creativity as a multidimensional construct**

Public sector creativity is a multidimensional construct, whereby its nature can be determined based on scores regarding six dimensions. The dimensions underlying the data fit our theoretical model, albeit differently than expected, corroborating arguments for public sector creativity as a multidimensional construct. The hypothesized framework implied three bipolar dimensions: “magnitude,” “trigger,” and “perspective” (Houtgraaf, 2022; Houtgraaf et al., 2022). The data, however, indicate not three bipolar but six unipolar dimensions underlying the data structure. The theoretically hypothesized bipolar dimensions of “magnitude,” “trigger,” and “perspective” can be split into two dimensions that regard these aspects of public sector
creativity: magnitude encompasses separate unipolar dimensions of “incrementalism” and “radicalism” (see also Houtgraaf, 2022; Kaufman & Beghetto, 2009; Kozbelt et al., 2010; Richards, 2007); trigger separate unipolar dimensions of “reactivism” and “proactivism” (see also Houtgraaf, 2022; Ranganajan, 2008; Unsworth, 2001); and perspective separate unipolar dimensions of “idealism” and “realism” (see also Amabile, 1996; Houtgraaf, 2022; Houtgraaf et al., 2023). Scores on matching poles are not on a continuum or mutually exclusive but can vary independently. To illustrate, public servants scoring higher on “realism” do not per definition score lower on “idealism.” However, this may result from measuring creativity by prompting public servants to reflect on it in general, instead of measuring specific instances where bipolar ratings are possible. The scale enables analysis of public servants’ creativity on six separate dimensions, allowing detailed assessment of the status quo of public servants’ creativity, identification for improvement lay, and detection of improvements in case of interventions.

The scale can also be applied along these lines to measure creativity multidimensionally in private sector contexts as configural equivalence analyses (Jilke et al., 2015) indicate equally valid and reliable measures of the construct. Interestingly, the enabled comparison of distributions on these dimensions indicates differences between the populations, namely, more right-skewed distributions for “proactiveness,” “radicalism,” and “idealism” in the private sector.

### Absence of second-order constructs

There is no evidence that variations on dimensions structurally caused the underlying second-order construct “nature of creativity.” The data indicate some interfactorial correlations but no consistent correlation patterns as theoretically assumed. This means that public servants scoring higher on “radicalism” do not per definition score higher on “proactivism” and “idealism,” nor do public servants scoring higher on “incrementalism” also score higher on “reactivism” and “realism.” Moreover, unipolarity, instead of bipolarity, indicates no evidence for a second-order construct of magnitude, trigger, or perspective; dimensions are to be regarded as statistically separate from one another. This allows for multiple configurations in terms of public servants’ creativity, where their creativity can lack on specific dimensions, whereas others are optimal.

### Pioneering creativity versus pragmatic creativity

Based on theoretical reasoning and statistical evidence, however, we argue a useful distinction between two categories of creativity. Contrary to our hypothesis considering “pioneering creativity” and “pragmatic creativity” as two poles of the “nature of creativity” second-order construct, “pioneering creativity,” and “pragmatic creativity” can be perceived as two categories of creativity. Instead of an underlying dimension in the form of a second-order construct with two poles that causes interfactorial correlations between the scores on the dimensions, the dimensions can be assigned to the two categories: “pioneering creativity” and “pragmatic creativity.” “Pioneering creativity” regards creativity aimed at exploration and disruption (Gieske et al., 2020, 2018; Osborne & Brown, 2013), encompassing the dimensions “radicalism,” “proactivism,” and “idealism,” whereas “pragmatic creativity” regards creativity aimed at problem solving and continuity (Barrutia & Echebarria, 2019; Gieske et al., 2020, 2018; Jansen et al., 2006; Lindblom, 1959), encompassing dimensions of “incrementalism,” “reactivism,” and “realism.” Theoretically, these categories and encapsulated dimensions have different foci, applicability and arguably desirability; “pioneering creativity” in the public sector can be considered risk-taking because of discontinuity, whereas “pragmatic creativity” can be considered risk-mitigation because of continuous improvement (see also Osborne & Brown, 2013). Moreover, “pioneering creativity” is related to divergent forms of creative thinking, wherein wide ranges of ideas with fewer a priori criteria are generated, whereas “pragmatic creativity” is related to convergent thinking, wherein ideas as correct...
solutions are generated (Kaufman & Sternberg, 2010). Although the data structure provides little evidence for interfactorial correlations indicating second-order constructs, there is statistical evidence for this delineation between the two categories and encapsulated dimensions. Distributions on the dimensions “incrementalism,” “reactivism,” and “realism” pertaining to the “pragmatic creativity” category consistently exhibit little variation through severe right skewness, whereas distributions on the dimensions “radicalism,” “proactivism,” and “idealism” pertaining to “pioneering creativity” category consistently exhibit variation and a normal distribution. Thus, the notion that the public sector lacks creativity (Bommert, 2010; Chen & Bozeman, 2012) requires nuance. Public servants indeed appear pragmatically creative to high degrees (see also Houtgraaf, 2022; Kruyen & Van Genugten, 2017), though also pioneers in creativity to varying degrees. When analyzing public sector creativity, focus should lay in the scales’ scores regarding pioneering creativity; these dimensions are where variation occurs and where room for improvement may lay.

Limitations

This study has limitations. First, the scale requires cross-cultural validation in multiple languages, as it concerns a Dutch questionnaire validated with Dutch data. Measurements constructs are only cross-culturally equivalent when the unobserved latent trait shares meaning across different groups and when the examined latent concept is scaled equally across countries (Ilkje et al., 2015). In line with prior recommendations (De Leeuw et al., 2004; DeVellis, 2016; Flake & Fried, 2020; Lee et al., 2011; Perry, 2012), we encourage replication of the validation process in different languages, countries, and cultures based on the procedure and R-code to ensure the application of valid measurements.

Second, the nomological network has limitations. Though testing various forms of construct validity, the nomological network did not include closely related constructs that would shed more light on convergent/divergent validity, such as cognitive fluency and divergent thinking. Therefore, we encourage future studies to measure these variables and conduct additional correlation measures to further ensure construct validity.

Third, the scale is developed based on aggregated data stemming from policy implementers and policy makers, although types and degrees of creativity may differ between these types of public servants. Aggregation of data may cause distortions in distributions and consequently some of this study’s results. However, our intention was to develop a scale for the public sector in general and propose delving deeper into differences between policy makers and implementers in future research.

Fourth, the scale was developed using data from self-selected respondents reached through convenience sampling. Respondents chose to enroll in a panel/subscription and to complete the questionnaire. This may result in selection bias if respondents who make both choices are more prone to be characterized by specific features, such as increased affinity with or interest in the topic. This may lead to overrepresentation of specific groups in the data, leading to statistical distortions in distributions and correlations.

Fifth, limitations related to self-reporting on performance must be acknowledged. Several downsides accompany self-reporting of performance-related aspects such as creativity when compared to peer or managerial ratings. For example, research indicates that managerial rating of performance correlated significantly higher with objective performance measures than self-report ratings (Jaramillo et al., 2005; Koopmans, 2014). Moreover, although we tested social desirability in the nomological network, self-rating shows social desirability and leniency effects (Van der Heijden & Nijhof, 2004), resulting in self-ratings of performance being generally one-half to one standard deviation higher than ratings by peers or managers (Van der Heijden & Nijhof, 2004). Nevertheless, self-reporting scales were chosen for several reasons. Objective measures of performance are not easily obtainable (Jaramillo et al., 2005), and creativity is per definition a construct that is exceedingly difficult to measure (Koopmans, 2014; Mumford, 2012). Moreover, employees often have more opportunity to observe their own behaviors than peers/managers (Dalal, 2005; Van der Heijden & Nijhof, 2004). Furthermore, peers/managers rate an employee’s performance on the basis of his/her general impression (Dalal, 2005; Viswesvaran et al., 2005), rendering ratings prone to distortions through homogeneity in dimensions and overestimated correlations as a result of halo effects. Finally, compared to objective measures or managerial ratings, self-reporting has practical advantages in terms of ease of collection, confidentiality, and complete data (Schoorman & Mayer, 2008).

Sixth, there is potentially some overlap between Flitspanel and Binnenlands Bestuur. Both samples feature Dutch public servants, with the potential of overlap in respondents, which may harm the results of this study to a certain degree. Flitspanel and Binnenlands Bestuur, however, differ in multiple ways, respectively: proactive subscribers to a magazine/platform versus panel members; no payment versus payment; broad set of topics versus HRM only; predominantly higher educated versus representative; and emphasis on administrators and (middle)management versus representative. Additionally, both (identical) questionnaires were sent out within a time window of 2 weeks, reducing chance that respondents would fill out both because they recognize and/or get bored with identical questionnaires. This reduces chance of overlap.

Finally, the scale’s specific focus on general measurements of creativity enables unipolarity of dimensions, whereas focus on specific instances could enable bipolarity of dimensions. The goal was to measure public
servants’ creativity multidimensionally in general, instead of specific instances; thus, general reflections provide most valid insight. However, future research may consider questions/measurements regarding specific instances able to capture bipolarity, for example, presenting vignettes on specific problems/opportunities to public servants and have them 1) generate, pick, or rate ideas or 2) have peers, managers, researchers rate their generated ideas on bipolar dimensional scales regarding magnitude or perspective. The resulting data would be indicative of tendencies in specific cases and contingent on the specific problem/opportunity and context, although this could be useful when researching specific cases. However, considering the goal of this study, these contingencies would result in downsides regarding ecological validity (with vignettes) or external validity (with respondents picking instances) for measurements of creativity in general.

CONCLUSION

This study offers a comprehensive but succinct six-dimensional measurement scale that allows for adequate assessment of public servants’ creativity, supplemented by a corresponding framework that aids in understanding the scale, the resulting data, and public sector creativity as a construct (see Figure 3). Public sector creativity as a multidimensional construct can be characterized based on the magnitude, trigger, and perspective as aspects of public servants’ creativity, whereby each aspect features two unipolar dimensions. Magnitude can be measured based on the degree of incrementalism and radicalism, trigger based on the degree of reactivism and proactivism, and perspective based on the degree of realism and idealism. Although no evidence for a second-order factor was found, based on our theoretical reasoning and statistical results, we argue that these dimensions can be divided into two categories, namely, “pioneering creativity” or “pragmatic creativity.” “Pioneering” creativity encapsulates the dimensions “radicalism,” “proactivism,” and “idealism” as aspects of creativity aimed at exploration and disruption (see also Gieske et al., 2018; Gieske et al., 2020; Osborne & Brown, 2013), whereas “pragmatic creativity” encapsulates the dimensions “incrementalism,” “reactivism,” and “realism” as aspects of creativity aimed at problem solving and continuity (see also Barrutia & Echebarria, 2019; Gieske et al., 2018; Gieske et al., 2020; Jansen et al., 2006; Lindblom, 1959). The scale is comprehensive and can be applied for extensive and detailed assessment of the status quo in terms of public servants’ creativity, indicating on which dimensions specific room for improvements may lay and assessing the effectiveness of interventions aimed at improving public sector creativity. Its succinctness makes it easy to administer, and the supplementary framework makes it easy to understand. Moreover, the scale may also be applied by practitioners/academics interested in measurement and comparisons using private sector data.

CONFLICT OF INTEREST STATEMENT

No potential conflict of interest was reported by the authors.

DATA AVAILABILITY STATEMENT

Data will be made available for verification upon request but are under embargo until January 2028 as it is part of a dissertation and may not be used for additional research until that time.

ENDNOTES

1 The name of the second-order construct was initially formulated as “pragmatism” in the preregistration but has been renamed to “nature of
creativity” in the paper to avoid confusion with the “pragmatic creativity” category as one of the study’s findings. The change in this instance is solely a label/name change to avoid confusion with “pragmatism” as second-order construct label in the preregistration and the “pragmatic creativity” category as one of the study’s findings; there is no discrepancy in terms of contents between the second-order construct as defined and planned in the preregistration and tested in the actual study.

Future studies may consider additional nomological tests. For example, we did not opt for cognitive fluency and divergent thinking, although these constructs offer viable options for additional testing of convergent validity. However, we did not opt for cognitive fluency as this is measured mostly through association tests (e.g., make as many associations in relation to a given word in 1 minute and then the researchers review the quantity and divergence), which we found not to be a viable option in a large-scale survey because of the required resources for both respondents and researchers. Divergent thinking would possibly be an addition to test for convergent validity, but we had limited space in terms of the number of items we could include in the survey and prioritized testing convergent validity by means of the generic creativity scale because this is a construct that should be most similar.

REFERENCES

**AUTHOR BIOGRAPHIES**

Glenn Houtgraaf research studies performance and motivation amidst constraints. His dissertation regards creativity amidst public sector constraints, and his current line of research regards motivation amidst constraints. Glenn is focused on introducing innovative and rigorous new research methods, such as longitudinal digital diary studies and physiological measurements, as well as interdisciplinary integration of Public Administration, Psychology and Neurobiology frameworks and concepts.

Houtgraaf@essb.eur.nl

Peter Kruyen has a background in Public Administration and Organization Studies (BA & MA) and a PhD in psychometrics. He works as assistant professor at the Department of Public Administration, Radboud University. He teaches organizational theory, public management, and statistics to public administration. His research program focuses on civil servants’ behavior, psychological characteristics, and competencies and on how their work is affected by managerial techniques, human resource management, and organizational structures.

p.kruyen@ru.nl

Sandra van Thiel studied psychology and educational sciences at Catholic University Nijmegen and has a PhD in sociology from Utrecht University. She is a professor of public management at Erasmus University Rotterdam. Her research interests regard semiautonomous executive agencies, which have been charged with the implementation of public policies but which operate at a distance from the government. She is a frequent advisor to governments and executive agencies.

vanthiel@essb.eur.nl

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Houtgraaf, Glenn, Peter Kruyen, and Sandra van Thiel. 2023. “Measuring the Construct of Public Sector Creativity: Development of a Validated Scale.” *Public Administration Review* 1–16. https://doi.org/10.1111/puar.13762