REVISTA CARIBEÑA DE PSICOLOGÍA

Fundación para el Desarrollo de la Psicología en el Caribe

Original Article

# Review of the Psychometric Properties of the Social Desirability Scale and the Development of a Short-Form

## Revisión de las Propiedades Psicométricas de la Escala de Deseabilidad Social y Desarrollo de una Versión Breve

**Ernesto Rosario-Hernández [1*], Lillian V. Rovira-Millán [2], & Rafael A. Blanco-Rovira [3]**

1 Ponce Health Sciences University, San Juan University Center, Puerto Rico.
Ponce Research Institute, Ponce Health Sciences University. https://orcid.org/0000-0002-2523-6162
2 Universtiy of Puerto Rico, Cayey Campus, Puerto Rico. https://orcid.org/0000-0001-7479-7401
3 Carlos Albizu University, San Juan Campus, Puerto Rico. https://orcid.org/0000-0002-8153-1967

* Correspondence: School of Behavioral & Brain Sciences, Clinical Psychology Programs, Ponce Health Sciences University, PO Box 7004, Ponce, Puerto Rico, 00732-7004. Email: erosario@psm.edu.

**Received**: January 5, 2025 | **Accepted**: March 25, 2025 | **Published**: October 19, 2025

**WWW.REVISTACARIBENADEPSICOLOGIA.COM**

**Cite as:**

Rosario-Hernández, E., Rovira-Millán, L. V., & Blanco-Rovira, R. A. (2025). Review of the psychometric properties of the Social Desirability Scale and the development of a short-form. *Revista Caribeña de Psicología, 9*(1), e13313. https://doi.org/10.37226/rcp.v9i1.13313

**ABSTRACT**

This study aimed to evaluate the psychometric properties of the Social Desirability Scale (SDS) and develop a short version, the SDS-BV. Using data from 3,855 participants in Puerto Rico, confirmatory factor analyses demonstrated that while the SDS showed excellent incremental fit (CFI = .995, TLI = .994), its RMSEA values were weaker, highlighting potential areas for refinement. The SDS-BV consistently exhibited very strong fit across all indices (CFI = 1.00, TLI = 1.00, SRMR = .012), underscoring its robustness and parsimony. Both scales achieved measurement invariance across demographic and organizational groups, confirming their equivalence for comparative research. High reliability coefficients ($\alpha \geq .926$) and significant correlations with related constructs support the validity of both scales. The SDS-BV, with fewer items and superior psychometric performance, emerges as a practical and reliable tool for assessing social desirability in diverse contexts. Findings highlight their relevance for psychological and organizational studies.

***Keywords***: social desirability, confirmatory factor analysis, internal structure, psychometric properties

**RESUMEN**

Este estudio examinó las propiedades psicométricas de la Escala de Deseabilidad Social (EDS) y desarrolló una versión breve, la EDS-VB. Utilizando datos de 3,855 participantes en Puerto Rico, los análisis factoriales confirmatorios mostraron que, aunque la EDS presentó un ajuste incremental excelente (CFI = .995, TLI = .994), sus valores de RMSEA fueron más débiles, sugiriendo áreas para mejorar. La SDS-BV demostró un ajuste muy fuerte en todos los índices (CFI = 1.00, TLI = 1.00, SRMR = .012), destacando su robustez y parsimonia. Ambas escalas lograron invarianza de medición entre grupos demográficos y organizacionales, confirmando su equivalencia para investigaciones comparativas. Los altos coeficientes de confiabilidad ($\alpha \geq .926$) y las correlaciones significativas con constructos relacionados respaldan la validez de ambas escalas. La EDS-VB, con menos ítems y mejor desempeño psicométrico, surge como una herramienta práctica y confiable para medir la deseabilidad social en diversos contextos. Estos hallazgos destacan su relevancia para estudios psicológicos y organizacionales.

***Palabras Claves***: deseabilidad social, análisis factorial confirmatorio, estructura interna, propiedades psicométricas

## INTRODUCTION

Social desirability bias (SDB) is a significant concern in psychological research, affecting the validity of self-report measures (Grimm, 2010; van de Mortel, 2008). It occurs when participants provide socially desirable responses instead of their true feelings, particularly on sensitive topics (Grimm, 2010). To address this issue, researchers can employ social desirability scales to detect and control for SDB (Nederhof, 1985; van de Mortel, 2008). However, only a small percentage of studies use these scales, despite their potential impact on results (van de Mortel, 2008). Various methods to prevent or reduce SDB have been proposed, including forced-choice items, randomized response techniques, and self-report questionnaires (Nederhof, 1985). Some researchers argue that social desirability is a stable trait rather than a situational response set (Furnham, 1986). Understanding and controlling for SDB is crucial for improving the validity of questionnaire-based research and interpreting results accurately (Grimm, 2010; van de Mortel, 2008).

Thus, the Social Desirability Scale (SDS) was developed by Rosario-Hernández and Rovira-Millán (2002) with the intention to measure this bias in research. These authors define social desirability as the tendency of people to respond to a test in the way they consider to be socially acceptable. In other words, it refers to the tendency of people to respond in a way that presents them favorably to others, rather than responding with complete sincerity. As a result, the SDS was created mainly to support the establishment of new measurement tools' divergent construct validity. As a result, Rosario-Hernández and Rovira-Millán (2002) were the first to address it. They designed the SDS and validated a scale to evaluate attitudes toward retirement, although the SDS was not their primary focus in that paper.

In most situations, when developing a new measurement instrument, the psychological constructs being developed and measured should have no relationship with the construct of social desirability, and in other cases, if they did, it would be inversely so. Although constructs such as positive impression, for example, could have a positive relationship with social desirability, when a new measurement instrument is being developed, the relationship of this new instrument with social desirability can be used to provide evidence of divergent validity. However, SDS scores could also be used to assess the levels of social desirability at which people respond to instruments that measure sensitive constructs, such as stigma, unfavorable attitudes, among others, or the influence of this on these constructs (e.g., Parmač et al., 2014). In summary, measuring social desirability in psychological research is essential to ensure the validity and reliability of the results, as well as to obtain a more complete understanding of human behavior, psychological processes, and the presentation of divergent-type construct validity.

The SDS has been used in several published studies to help establish divergent validity of new psychological measures (Hernández et al., 2020; Pons et al., 2009; Quiñones et al., 2015; Rivera et al., 2008; Rosario-Hernández & Rovira-Millán, 2004; Rosario-Hernández & Rovira-Millán, 2007; Rosario-Hernández & Rovira-Millán, 2008a; Rosario-Hernández & Rovira-Millán, 2008b; Rosario-Hernández & Rovira-Millán, 2008c; Rosario-Hernández & Rovira-Millán, 2011; Rosario-Hernández & Rovira-Millán, 2012; Rosario-Hernández & Rovira-Millán, 2014; Rosario-Hernández et al., 2011; Rosario-Hernández et al., 2004; Rosario-Hernández et al., 2021; Rosario-Hernández et al., 2022; Rovira-Millán et al., 2023; Rovira-Millán et al., 2018); however, the psychometric properties of the SDS have never been examined since its development and validation. Thus, the purpose of this research is to examine the psychometric properties of the SDS in a sample of people in Puerto Rico. We also intend to create an abbreviated version of the SDS.

### Description and Original Psychometric Properties of the Social Desirability Scale

The SDS's eleven items were projectively written in response to Kline's (1993) suggestions. He clarifies that when writing something like this directly— *"I would steal a million dollars if I were sure that I would not get caught"*—no person would answer affirmatively. A more likely useful item would be: "*Most of the people I know would steal a million dollars if they were absolutely sure that they would not get caught."* Thus, the SDS items were written projectively to make it much more useful. The response format of the SDS is a Likert-scale format that ranges from totally disagree (1) to totally agree (6). In terms of the previous psychometric properties of the SDS reported by Rosario-

Hernández and Rovira-Millán (2002), these authors found that the internal structure of the SDS was unidimensional using exploratory factor analysis; that is, all the SDS items were grouped under a single factor in the original study by Rosario-Hernández and Rovira-Millán (2002), supporting the construct validity of the SDS. As can be appreciated, the internal structure of the SDS has not been examined using confirmatory factor analysis (CFA). Regarding reliability, the SDS obtained a Cronbach's alpha coefficient of $\alpha = .86$.

### Objectives of the Study

Therefore, the aims that guided the current study were as follows: (1) to examine the internal structure of the SDS using CFA; (2) to create a short form of the SDS; (3) to examine the measurement invariance by group (gender, age, job position, employment type, tenure, and education); (4) to examine validity evidence on the relationship with other variables; (5) to examine the reliability of the SDS and its brief version via internal consistency using Cronbach's alpha and McDonald's omega with their respective confidence intervals and their descriptive statistics, such as the mean and standard deviation, among others.

## METHOD

### Participants

A total of 4,100 protocols from different research conducted by the principal authors in Puerto Rico (Rosario-Hernández et al., 2018; Rosario-Hernández et al., 2015; Rosario-Hernández et al., 2024; Rosario-Hernández et al., 2019; Rosario-Hernández et al., 2018) were each selected through a non-probabilistic sample for this instrumental research design (Ato et al., 2013; González-Rivera, 2025). They were selected based on availability, and their voluntariness, anonymity, and the right to abandon the investigation were guaranteed when they considered it necessary. Detection of multivariate outliers was made in the response to all items of the SDS using the squared Mahalanobis distance ($D^2$) value, an efficient and sensitive measure for outliers derived from random responses (Zijlstra et al., 2011). The cut-off point for $D^2$ was 3.57 ($df = 15$); therefore, 245 protocols were eliminated, and the final total sample was 3,855. Table 1 shows the description of the sociodemographic characteristics of this revised sample.

**Table 1**

*Sociodemographic characteristics of the sample.*

| Variable | *f* | % |
|---|---|---|
| **Gender** | | |
| Male | 1556 | 40.4 |
| Female | 2182 | 56.6 |
| **Age** | | |
| 21-30 | 1047 | 27.2 |
| 31-50 | 2050 | 53.2 |
| ≥ 51 | 758 | 19.7 |
| **Position** | | |
| Managerial | 845 | 21.9 |
| Non-Managerial | 2889 | 74.9 |
| **Employment Type** | | |
| Tenure | 3017 | 78.3 |
| Temporary | 776 | 20.1 |
| **Organization Type** | | |
| Public | 1299 | 33.7 |
| Private | 2485 | 64.5 |
| **Education** | | |
| ≤ High School | 52 | 1.3 |
| Undergraduate Studies | 2089 | 54.2 |
| Graduate Studies | 1416 | 36.7 |
| | **Mean** | **SD** |
| | 16.74 | 2.00 |
| **Useful Source of Data** | | |
| Sample 1 | 756 | 19.6 |
| Sample 2 | 921 | 23.9 |
| Sample 3 | 956 | 24.8 |
| Sample 4 | 495 | 12.8 |
| Sample 5 | 727 | 18.9 |

Nota. $n = 3{,}855$; *SD* = Standard Deviation.

Of the total sample, 56.6% (2,182) of the people were female, and 53.2% (2,050) were between 31–50 years of age, which can be considered to be in the prime of their working careers. The average education achieved was equal to 16.74 with a standard deviation equal to 2.00, which is equivalent to four years of undergraduate to near one year of graduate university studies. On the other hand, 21.9% (845) of the participants held a managerial position, 78.3% (3,017) had tenure, and 64.5% (2,485) of the sample worked for the private sector.

### Measures

**Social Desirability Scale (SDS).** We used the SDS developed by Rosario-Hernández and Rovira-Millán (2002). This is an 11-item instrument in a Likert-agreement response format ranging from "1" *(Totally Disagree)* to "6" *(Totally Agree)*, which is intended to measure a response bias in which people respond to

a test thinking about what is socially acceptable. Authors report its internal consistency through Cronbach's alpha to be .86, which is an excellent reliability coefficient. Factor analysis results suggest that the Social Desirability Scale's internal structure has only one factor.

**Demand-Control-Support Model Inventory (DCSMI)**. We used the DCSMI to measure work demands, job control, and job support (Rosario-Hernández & Rovira-Millán, 2014). The DCSMI contains three second-order factors, which are Work Demands, Job Control, and Job Support; it has 29 items and is in a Likert-agreement response format ranging from "1" *(Totally Disagree)* to "6" *(Totally Agree)*. The Job Demands factor has three dimensions, which are Psychological Demands, Emotional Demands, and Physical Demands, with six, four, and five items, respectively. The Job Control factor has two dimensions, which are Autonomy and Skills, which have three items each. Whereas the Job Support factor has two dimensions, which are Co-Worker Support and Supervisory Support, with four items each. Factor analyses support the second-order hierarchical three-dimensional internal structure with seven first-order factors. Reliability coefficients using Cronbach's alpha ranged from .63 to .95.

**Patient Health Questionnaire-9 (PHQ-9).** The PHQ-9, developed by Kroenke et al. (2001), was used to assess depression. The PHQ-9 is a nine-item questionnaire used in primary care settings to detect depressive symptoms. This questionnaire assesses the existence of depressive symptoms in the two weeks preceding the completion of the test. Each item is graded on a scale of 0 *(not at all)* to 3 *(very; nearly every day)*. Its diagnostic validity and reliability, as well as its utility in assessing depression severity and monitoring treatment response, have all been established (Löwe et al., 2004; Kroenke et al., 2001; Löwe et al., 2006; Löwe et al., 2004). Moreover, there are studies that have reviewed its psychometric properties with adults in Puerto Rico (González-Rivera, 2019; Rosario-Hernández et al., 2023). These two studies found that the PHQ-9 has excellent reliability coefficients above .89 using Cronbach's alpha and McDonald's omega, and results from the confirmatory factor analyses suggest that the one-factor structure is the most parsimonious interpretation, as some researchers have argued (Arrieta et al., 2017; Aslan et al., 2020; Merz et al., 2011; Quiñones-Freire et al., 2021; Rosario-Hernández et al., 2023; Saldivia et al., 2019; Villarreal-Zegarra et al., 2019).

**Generalized Anxiety Disorder-7 (GAD-7).** To measure anxiety, we used the GAD-7 (Spitzer et al., 2006). The GAD-7 is a seven-item questionnaire that measures general anxiety symptomatology and asks patients how often, during the last two weeks, they were bothered by each symptom. Response options are *"not at all," "several days," "more than half the days," and "nearly every day,"* scored as 0, 1, 2, and 3, respectively. In addition, an item to assess duration of anxiety symptoms was included. Authors of the scale reported a Cronbach's alpha coefficient of .93. In terms of its construct validity, internal structure was supported by factor analysis and convergent validity with its association to similar measures such as the Beck Anxiety Inventory and the anxiety subscale of the Symptom Checklist-90. An item example is: *"Feeling nervous, anxious, or on edge."* The psychometric properties of the GAD-7 have been reviewed with adults in Puerto Rico (Pagán-Torres et al., 2020; Merino-Soto et al., 2023). These studies found excellent reliability coefficients over .90 and found a one-factor internal structure.

**Utrecht Work Engagement Scale (UWES).** We used the UWES (Schaufeli & Bakker, 2003; Schaufeli et al., 2002). The UWES is comprised of 17 items measured on a seven-point Likert scale anchored by the response options *"0" = never and "6" = always*. Six items comprised the vigor subscale (e.g., "*At my work, I feel bursting with energy"*). The dedication subscale was measured with five items (e.g., *"I find the work that I do full of meaning and purpose"*). Finally, the remaining six items comprised the absorption subscale. Reliability, using Cronbach's alpha, of the UWES and its subscales has been reported between .82 and .93 (Schaufeli & Bakker, 2003). Several studies carried out in Puerto Rico have used it with samples of employed people, and their results support the internal structure; its reliability coefficients fluctuated between .81 and .95 using the Cronbach's alpha and omega techniques (e.g., Martínez-Avarado et al., 2017; Rodríguez Montalban et al., 2014; Rosario-Hernández et al., 2021).

**Maslach Burnout Inventory – General Survey (MBI-GS).** To measure burnout, we used the MBI-GS. The MBI-GS was developed from the original MBI

designed originally for human service occupations (Maslach et al., 1996). The MBI-GS consists of 16 items rated on a 7-point scale ranging from 0 (*never*) to 6 (*daily*). The scale is composed of three subscales: Emotional Exhaustion (5 items), Cynicism (5 items), and Professional Efficacy (6 items). Burnout is reflected in higher scores on Emotional Exhaustion and Cynicism and lower scores on Professional Efficacy. The validity and reliability of the scores produced by the MBI-GS have been supported by some studies in Latin American countries (e.g., Bravo et al., 2021; Flores Jiménez et al., 2015; Fernández et al., 2015a, 2015b; Tomás et al., 2016). In a study in Puerto Rico, reliability coefficients fluctuated between .84 and .93 using Cronbach's alpha and McDonald's omega, and the three-factor structure was supported with a sample of workers (Rosario-Hernández et al., 2024).

**Work-Related Rumination Scale (WRRS).** We used the WRRS (Cropley et al., 2012), specifically the Spanish version (WRRS-SV) translated and validated by Rosario-Hernández et al. (2021) in Puerto Rico. The WRRS originally has 15 questions using a 5-point Likert scale *(1 = very seldom or never, 2 = seldom, 3 = sometimes, 4 = often, and 5 = very often or always)*. Rosario-Hernández and collaborators retained 11 of the 15 items in the WRRS-SV: four items of the Affective Rumination and Problem-Solving Pondering subscales and three items of the Detachment subscale. As conceptualized by Cropley and collaborators, results with a Puerto Rican sample of workers support a three-factor internal structure of the WRRS-SV (Rosario-Hernández et al., 2021), and authors reported that reliability ranged between .72 and .87 via Cronbach's alpha and McDonald's omega.

## General Procedures

The authorization to carry out the research was requested from the Institutional Review Board of Ponce Health Sciences University. It was approved on June 17, 2020, with the protocol number 2006040219. The data were analyzed, first, with the IBM SPSS version 29.0 program, and with-it descriptive statistics, correlation, item analysis, and reliability analysis were performed. In addition, we used the "lavaan" package of the R 3.6.3 program (Rosseel, 2012) to perform the confirmatory factor analyses and invariance testing by gender, age, education, job position, organization type, and employment type.

First, we performed descriptive statistics analyses to obtain sociodemographic characteristics of the sample. Also, we conducted descriptive analyses of the scale's items, such as the mean, standard deviation, skewness, and kurtosis. An item analysis was also performed to obtain the discrimination index, which is also known as *"item-total correlation"* or *"$r_{bis}$."* We used the whole sample to perform these descriptives and item analyses. Second, the total sample was randomly split into two samples, calibration and validation. The intention of having a calibration sample and a validation sample was to be able to replicate the CFA results obtained in the first and see if they are repeated in the second, as recommended in some literature (Fabrigar et al., 1999).

Third, SDS items were subjected to CFA using structural equation modeling to confirm the one-factor internal structure of the SDS using the weighted least squares mean and variance adjusted (WLSMV) estimator with the "lavaan" package of the R 3.6.3 program, which robustly deals with potentially non-normal data, and items are treated as ordinal (Li, 2016a, 2016b). To evaluate the fit of our measurement models, we applied a percentile-based approach to interpret the fit indices, as recommended by recent methodological advancements (e.g., Howard et al., 2025), given that dichotomous cutoff values are overly simplistic and do not account for nuanced model quality. This approach provides a deeper understanding of model performance by classifying fit indices into percentile-based categories of Very Weak, Weak, Moderate, Strong, and Very Strong fit. Kline (2016) recommends the use of at least four fit indices, although more can be reported. One of the indices that is reported is chi-square ($\chi^2$); however, given that $\chi^2$ is sensitive to the sample size and therefore the probability of rejecting the hypothesized model increases when the sample size increases, it is recommended to take into account other indices (Marsh et al., 1996), and for this reason, it was reported but not taken into consideration as a fit index. Thus, we assessed the fit of the models using commonly recommended fit indices: Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and Tucker–Lewis Index (TLI). These indices were compared against empirically derived

percentile-based ranges to facilitate contextualized interpretation (see Table 2).

We chose the four items of the SDS with the highest factor loadings to construct a brief version of the SDS. The internal structure of the SDS-BV was therefore investigated using CFA using both the calibration and validation samples, just as was done with the SDS. After recombining both samples, we assessed measuring invariance of the SDS and SDS-BV among gender, age, education level, job position, organization type, and employment type. We tested configural invariance, metric invariance, and scalar invariance as suggested by some of the literature (e.g., Byrne, 2016; Muthén & Muthén, 1998–2012; Wang & Wang, 2012). We conducted hierarchical tests for invariance of measurement parameters. First, we examined the configural invariance model or pattern invariance, which imposes no equality restrictions on model parameters. This is a necessary condition for testing invariance by comparing it with other invariance models based on fit indices. Second, we examined the weak invariance model or metric invariance. In this model, the factor loadings are treated as invariant across groups. This ensures that the measures are on the same scale across groups for making valid comparisons. Third, we examined the strong invariance model. This model imposes invariance on both factor loadings and item intercepts across groups. This is to ensure the underlying factors can be compared across groups. Given that the RMSEA showed limitations in assessing the fit of our model, we decided to use the CFI, TLI, and SRMR indices to evaluate measurement invariance. These indices have been shown to be more stable and less sensitive to factors such as sample size or model complexity, making them ideal for this purpose (Howard et al., 2025; Hu & Bentler, 1999). Furthermore, their interpretation based on percentiles allows for a more nuanced evaluation of the model compared to traditional cutoff-based approaches. We capitalized on fit index differences for CFI, TLI, and SRMR (i.e., $\Delta$CFI & $\Delta$TLI $\leq -.01$, $\Delta$SRMR $\geq .03$ for metric invariance and $\geq .015$ for scalar invariance) reference points as recommended by Cheung and Rensvold (2002), who found in a Monte Carlo study that these indices were equally sensitive to all types of invariance. Notably, as $\chi^2$ is known to be highly influenced by the sample size (e.g., Rigdon, 1995), it was reported but not considered as a fit index for the invariance testing.

Fourth, we examined the convergent and divergent validity of the SDS and the SDS-BV by correlating scores on these with each other and with other supposedly different constructs. Finally, we recombined the samples and performed descriptive, reliability, and correlation analyses for the SDS and SDS-BV to estimate means, standard deviation, internal consistency via Cronbach's alpha and McDonald's omega, standard error of measurement, and 95% confidence intervals for the scale.

## RESULTS

### Descriptive Statistics and Discrimination Index

We conducted descriptive statistics and an item analysis of the SDS. Table 3 shows the mean, standard deviation, skewness, kurtosis, and discrimination index through item–total correlation (rbis). Item 8 obtained the highest mean (3.51) and standard deviation (1.881). All items have skewness and kurtosis values within the ±1.5 range (George & Mallery, 2016; Kline, 2016; Tabachnick & Fidell, 2019). All items were well above the threshold rbis of .30, which indicates high homogeneity (DeVellis, 2017; Spector, 1992).

**Table 2**
Guidelines for model fit interpretation using percentile ranges following Howard et al. (2025) recommendations.

| Interpretation (Percentile) | Fit Index | | | |
|---|---|---|---|---|
| | SRMR | RMSEA | CFI | TLI |
| Very Weak (<10) | > .100 | > .100 | < .900 | < .900 |
| Weak (10 - 33) | .081 - .100 | .081 - .100 | .900 - .920 | .900 - .920 |
| Moderate (34 - 66) | .061 - .080 | .061 - .080 | .921 - .950 | .921 - .950 |
| Strong (67 - 90) | .030 - .060 | .030 - .060 | .951 - .980 | .951 - .980 |
| Very Strong (>90) | < .030 | < .030 | > .980 | > .980 |

### Confirmatory Factor Analyses (CFA)

Table 4 shows the results of the CFAs for the SDS, and its brief version, SDS-BV, were evaluated using percentile-based guidelines for model fit interpretation (Howard et al., 2025). For the SDS, the calibration sample showed a moderate fit for the SRMR (.060, 34th–66th percentile) and a very weak fit for the RMSEA (.121, <10th percentile). However, incremental fit indices demonstrated very strong fit, with *CFI* = .995 and *TLI* = .993, both exceeding the 90th percentile. The validation sample yielded similar results, with the *SRMR* (.057) falling in the moderate fit range and the *RMSEA* (.112) remaining in the very weak fit range. Incremental indices (*CFI* = .995, *TLI* = .994) again indicated very strong fit. These results suggest that while the SDS demonstrates excellent fit relative to baseline models (as shown by CFI & TLI), there are areas for improvement, particularly in absolute fit indices like RMSEA.

In contrast, the SDS-BV consistently exhibited very strong fit across most indices in both the calibration and validation samples. The SRMR values (.009 for calibration; .012 for validation) exceeded the 90th percentile, indicating minimal residual discrepancies. The RMSEA for the calibration sample (.047, 67th–90th percentile) reflected strong fit, while the validation sample (.066, 34th–66th percentile) suggested moderate fit. Incremental indices for the SDS-BV were exceptional, with CFI = 1.00 and TLI = 1.00 in the calibration sample and nearly identical values in the validation sample. These results highlight the robustness and parsimony of the SDS-BV, making it a highly reliable alternative to the original scale.

To evaluate convergent validity of a reflective construct as social desirability, we checked that the average variance extracted (AVE) values were ≥ .50 for the 11 items of the SDS and the four items of the SDS-BV, which indicates that the items converge or share a high proportion of the variance (Fornell & Larcker, 1981). We calculated the AVE for the calibration, validation, and total samples, and all were well above the threshold of 0.50. Therefore, the indicators of the SDS and SDS-BV share a high proportion of variance (see Tables 5 & 6).

**Table 3**

*Descriptive statistics and discrimination index ($r_{bis}$) of items of the Social Desirability Scale.*

| # | Item | Mean | SD | Skewness | Kurtosis | $r_{bis}$ |
|---|---|---|---|---|---|---|
| 1. | La mayoría de las personas se han copiado en un examen, aunque haya sido una sola vez en sus vidas. *"Most people have cheated on an exam, even if it was once in their lives."* | 2.15 | 1.572 | 1.253 | 0.387 | .639 |
| 2. | Creo que la mayoría de las personas han tomado algo que no les pertenecía en algún momento dado en sus vidas. *"I think most people have taken something that did not belong to them at some point in their lives."* | 2.68 | 1.684 | 0.678 | -0.780 | .752 |
| 3. | Creo que la mayoría de las personas se han sentido inseguras de sí mismas en algún momento dado en sus vidas. *"I think most people have felt insecure of themselves at some point in their lives."* | 2.07 | 1.386 | 1.265 | 0.780 | .733 |
| 4. | La mayoría de las personas les han deseado mal a otras. *"Most people have wished bad things to others."* | 2.66 | 1.590 | 0.630 | -0.701 | .813 |
| 5. | La mayoría de las personas se han comportado hipócritamente con otras, aunque sea una vez. *"Most people have behaved hypocritically with others, even once."* | 2.26 | 1.476 | 1.016 | 0.020 | .832 |
| 6. | La mayoría de las personas han hablado mal de otras, aunque sea una sola vez. *"Most people have spoken bad of others, even once."* | 2.20 | 1.460 | 1.124 | 0.307 | .825 |
| 7. | Creo que la mayoría de las personas han mentido alguna vez para salir de algún problema. *"I think most people have ever lied to get out of a problem."* | 2.07 | 1.422 | 1.305 | 0.787 | .799 |
| 8. | Creo que la mayoría de las personas se robarían una gran suma de dinero si estuvieran completamente seguras de que nunca se les atraparía. *"I think most people stole a large sum of money if they were fully confident that they will never catch."* | 3.51 | 1.881 | 0.005 | -1.443 | .582 |
| 9. | La mayoría de las personas han sentido el deseo de romper algo cuando se han sentido disgustadas. *"Most people have felt the desire to break something when they have felt upset."* | 2.63 | 1.606 | 0.687 | -0.657 | .748 |
| 10. | La mayoría de las personas se han aprovechado de otras en algún momento. *"Most people have taken advantage of others at some point."* | 2.73 | 1.615 | 0.604 | -0.764 | .778 |
| 11. | A la mayoría de las personas le gusta chismear a veces. *"Most people like to gossip sometimes."* | 2.32 | 1.492 | 0.992 | 0.015 | .759 |

Note. n = 3,855; SD = Standard Deviation.

### Validity Evidence Based on Relations to Other Variables

In order to provide validity evidence based on relation to other variables of the SDS and SDS-BV, we correlated SDS and SDS-BV scores with each other and with scores on other scales measuring different constructs to establish convergent and divergent validity of the SDS and SDS-BV. As can be seen in Table 7, the scores of the SDS and SDS-BV correlated strongly with each other ($r = .941$, $p < .01$), thus supporting the convergent validity of both. On the other hand, it can also be seen that the correlations of the SDS and SDS-BV with the scores of other psychological instruments that measure different constructs fluctuated between $r = -.165$ and $r = .050$, supporting the divergent validity of the SDS and SDS-BV.

### Measurement Invariance

The measurement invariance of the SDS and SDS-BV were assessed across multiple grouping variables, including gender, age, education level, job position, employment type, and organization type. The analyses tested three levels of invariance: configural, metric, and scalar, using a variety of fit indices (SRMR, CFI, TLI) and changes in fit indices (ΔSRMR, ΔCFI, ΔTLI) to evaluate the robustness of the model across groups. Across all group comparisons, the fit indices remained stable for the SDS, with SRMR, CFI, and TLI consistently indicating strong to very strong fit at all levels of invariance (see Table 8). The changes in fit indices ($\Delta SRMR \leq 0.002$, $\Delta CFI \leq 0.000$, $\Delta TLI \leq 0.003$) were well within recommended thresholds, supporting the equivalence of the SDS measurement model across all tested groups. These results provide strong evidence for the generalizability of the SDS, confirming that the scale operates consistently across diverse demographic and organizational subgroups.

Regarding the measurement invariance of the SDS-BV, it was assessed across key demographic and organizational variables as well. Also, three levels of invariance were evaluated: configural, metric, and scalar. The fit indices (SRMR, CFI, TLI) consistently indicated strong to very strong fit across all models, with SRMR values remaining below .012 and CFI and TLI maintaining values of 1.00 or close to it throughout (see Table 8). Changes in fit indices ($\Delta SRMR \leq .002$, $\Delta CFI = .000$, $\Delta TLI \leq .001$) were well within established thresholds, providing robust evidence for invariance across all groups. These findings suggest that the SDS-BV maintains its structural integrity and measurement consistency across diverse populations and organizational settings, underscoring its utility as a reliable and generalizable tool for assessing social desirability.

### Reliability and Descriptive Statistics of the SDS and SDS-BV

We estimated the mean, standard deviation, standard error of measurement, and 95% confidence interval for the scores of the SDS and SDS-BV (see Table 10). Moreover, we estimated the reliability using Cronbach's alpha and McDonald's omega with their respective confidence intervals, and all reliability coefficients were well above .70 as suggested by some of the literature (e.g., DeVellis, 2017; Spector, 1992).

**Table 4**

Fit Indices obtained by the Social Desirability Scale (SDS) and Social Desirability Scale-Brief Version (SDS-BV)

| Index | SDS | | SDS-BV | |
|---|---|---|---|---|
| | Calibration Sample (n = 1,924) | Validation Sample (n = 1,931) | Calibration Sample (n = 1,924) | Validation Sample (n = 1,931) |
| $\chi^2$ (df) | 1,279.736* (44) | 1,106.269* (44) | 10.466* (2) | 18.618* (2) |
| SRMR | .060 | .057 | .009 | .012 |
| RMSEA | .121 (.115 - .127) | .112 (.106 -.118) | .047 (.022 - .077) | .066 (.041 - .094) |
| CFI | .995 | .995 | 1.00 | 1.00 |
| TLI | .993 | .994 | 1.00 | .999 |

Note. *p < .05.

**Table 5**

*Factor loadings (λ), explained variance by item ($\lambda^2$) and average variance extracted (AVE) by sample of the Social Desirability Scale (SDS)*

| Sample/Item # | Calibration (n =1,924) | | Validation (n = 1,931) | | Total (n= 3,855) | |
|---|---|---|---|---|---|---|
| | λ | $\lambda^2$ | λ | $\lambda^2$ | λ | $\lambda^2$ |
| 1 | .759 | .576 | .757 | .572 | .758 | .574 |
| 2 | .815 | .664 | .831 | .691 | .823 | .677 |
| 3 | .818 | .670 | .831 | .691 | .825 | .680 |
| 4* | .857 | .735 | .875 | .766 | .865 | .749 |
| 5* | .930 | .864 | .934 | .873 | .932 | .868 |
| 6* | .943 | .889 | .940 | .884 | .941 | .885 |
| 7* | .899 | .808 | .892 | .796 | .896 | .802 |
| 8 | .710 | .504 | .675 | .455 | .692 | .479 |
| 9 | .810 | .657 | .821 | .674 | .815 | .664 |
| 10 | .854 | .729 | .841 | .708 | .847 | .718 |
| 11 | .844 | .712 | .825 | .680 | .833 | .694 |
| AVE | .710 | | .708 | | .708 | |

Note. *Items selected for the SDS-BV.

**Table 6**

*Factor loadings (λ), explained variance by item (λ²) and average variance extracted (AVE) by sample of the Social Desirability Scale – Brief Version (SDS-BV)*

| Sample/ # Ítem | Calibration (n =1,924) | | Validation (n = 1,931) | | Total (n = 3,855) | |
|---|---|---|---|---|---|---|
| | λ | λ² | λ | λ² | λ | λ² |
| 4 | .827 | .684 | .862 | .743 | .844 | .712 |
| 5 | .943 | .890 | .947 | .897 | .945 | .893 |
| 6 | .967 | .936 | .961 | .923 | .964 | .929 |
| 7 | .902 | .813 | .893 | .797 | .897 | .805 |
| AVE | .831 | | .840 | | .835 | |

**Table 7**

*Correlation matrix between SDS and SDS-BV scores with other measures measuring different constructs*

| Variable | EDS | EDS-VB |
|---|---|---|
| SDS | 1 | .941** |
| SDS-BV | .941** | 1 |
| Work Demands | -.154** | -.112** |
| Job Control | -.050 | -.038 |
| Job Support | .051 | .057* |
| Depression | -.165** | -.136** |
| Anxiety | -.157** | -.126** |
| Vigor | .010 | -.001 |
| Dedication | -.003 | -.009 |
| Absorption | -.004 | -.090** |
| Professional Efficacy | -.076** | -.129** |
| Exhaustion | -.155** | -.119** |
| Cynicism | -.156** | -.070** |
| Affective Rumination | -.098** | -.063** |
| Problem-Solving Problem | -.076** | .023 |
| Detachment | .024 | -.112** |

Nota. n = 3,855; *p < .05, **p < .01.

**Table 8**

*Measurement invariance analysis of the Social Desirability Scale by gender, age, education, job position, employment type and organization type*

| Model | X² (df) | SRMR | CFI | TLI | Model Ref. | ΔX² | ΔSRMR | ΔCFI | ΔTLI |
|---|---|---|---|---|---|---|---|---|---|
| **Gender (Male/Female)** | | | | | | | | | |
| 1. Configural | 2,329 (88) | .058 | .995 | .994 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 2,369 (98) | .058 | .995 | .994 | 1 | +40 | .000 | .000 | .000 |
| 3. Scalar | 2,502 (141) | .058 | .995 | .994 | 2 | +133 | .000 | .000 | +.002 |
| **Age (21-30 /31-50/ ≥51)** | | | | | | | | | |
| 1. Configural | 2,493 (132) | .060 | .995 | .994 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 2,667 (152) | .062 | .995 | .994 | 1 | +174 | +.002 | .000 | .000 |
| 3. Scalar | 2,669 (238) | .060 | .995 | .996 | 2 | +2 | -.002 | .000 | +.002 |
| **Education (≤HS /UGS/ GS)** | | | | | | | | | |
| 1. Configural | 2,209 (132) | .058 | .995 | .994 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 2,308 (152) | .060 | .995 | .994 | 1 | +99 | +.002 | .000 | .000 |
| 3. Scalar | 2,289 (238) | .059 | .995 | .997 | 2 | -19 | -.001 | .000 | +.003 |
| **Job Position (Managerial/Non-Managerial)** | | | | | | | | | |
| 1. Configural | 2,269 (88) | .058 | .995 | .994 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 2,371 (98) | .059 | .995 | .994 | 1 | +102 | +.001 | .000 | .000 |
| 1. Scalar | 2,311 (141) | .058 | .995 | .996 | 2 | -60 | -.001 | .000 | +.002 |
| **Employment Type (Tenure/Temporary)** | | | | | | | | | |
| 1. Configural | 2,357 (88) | .058 | .995 | .994 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 2,383 (98) | .059 | .995 | .994 | 1 | +26 | +.001 | .000 | .000 |
| 3. Scalar | 2,421 (141) | .058 | .995 | .996 | 2 | +38 | -.001 | .000 | +.002 |
| **Organization Type (Public/Private)** | | | | | | | | | |
| 1. Configural | 2,404 (88) | .059 | .995 | .994 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 2,523 (98) | .060 | .995 | .994 | 1 | +119 | +.001 | .000 | .000 |
| 3. Scalar | 2,454 (141) | .059 | .995 | .996 | 2 | -69 | -.001 | .000 | +.002 |

Nota. df = degree of freedom; HS =High School, UGS = Undergraduate Studies, GS = Graduate Studies.

**Table 9**

*Measurement invariance analysis of the Social Desirability Scale – Brief Version by gender, age, education, job position, employment type and organization type*

| Model | $X^2$ (df) | SRMR | CFI | TLI | Model Ref. | $\Delta X^2$ | ΔSRMR | ΔCFI | ΔTLI |
|---|---|---|---|---|---|---|---|---|---|
| **Gender (Male/Female)** | | | | | | | | | |
| 1. Configural | 33.784 (4) | .011 | 1.00 | .999 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 38.062 (7) | .011 | 1.00 | 1.00 | 1 | +4.28 | .000 | .000 | +.001 |
| 3. Scalar | 50.912 (22) | .011 | 1.00 | 1.00 | 2 | +12.85 | .000 | .000 | .000 |
| **Age (21-30 /31-50/ ≥51)** | | | | | | | | | |
| 1. Configural | 27.549 (6) | .010 | 1.00 | 1.00 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 37.336 (12) | .012 | 1.00 | 1.00 | 1 | +9.79 | +.002 | .000 | .000 |
| 3. Scalar | 47.625 (42) | .010 | 1.00 | 1.00 | 2 | +10.29 | -.002 | .000 | .000 |
| **Education (≤HS /UGS/ GS)** | | | | | | | | | |
| 1. Configural | 26.106 (6) | .010 | 1.00 | 1.00 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 25.669 (12) | .010 | 1.00 | 1.00 | 1 | -0.437 | .000 | .000 | .000 |
| 3. Scalar | 25.669 (42) | .010 | 1.00 | 1.00 | 2 | 0.000 | .000 | .000 | .000 |
| **Job Position (Managerial/Non-Managerial)** | | | | | | | | | |
| 1. Configural | 26.719 (4) | .010 | 1.00 | 1.00 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 44.323 (7) | .012 | 1.00 | 1.00 | 1 | +17.60 | +.002 | .000 | .000 |
| 3. Scalar | 30.734 (22) | .010 | 1.00 | 1.00 | 2 | -13.59 | -.002 | .000 | .000 |
| **Employment Type (Tenure/Temporary)** | | | | | | | | | |
| 1. Configural | 27.776 (4) | .010 | 1.00 | 1.00 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 28.769 (7) | .010 | 1.00 | 1.00 | 1 | +0.993 | .000 | .000 | .000 |
| 3. Scalar | 37.408 (22) | .010 | 1.00 | 1.00 | 2 | +8.64 | .000 | .000 | .000 |
| **Organization Type (Public/Private)** | | | | | | | | | |
| 1. Configural | 26.903 (4) | .010 | 1.00 | 1.00 | ----- | ----- | ----- | ----- | ----- |
| 2. Metric | 31.461 (7) | .011 | 1.00 | 1.00 | 1 | +4.56 | +.001 | .000 | .000 |
| 3. Scalar | 33.127 (22) | .010 | 1.00 | 1.00 | 2 | +1.67 | -.001 | .000 | .000 |

Nota. df = degree of freedom; HS =High School, UGS = Undergraduate Studies, GS = Graduate Studies.

**Table 10**

*Reliability and descriptive statistics of the Social Desirability Scale (SDS) and Social Desirability Scale-Brief Version (SDS-BV)*

| Statistic | SDS | SDS-BV |
|---|---|---|
| Number of Items | 11 | 4 |
| Mean | 27.28 | 9.20 |
| Standard Deviation | 13.67 | 5.38 |
| Reliability | | |
| Cronbach's Alpha | .941 (.937 - .944) | .926 (.920 - .931) |
| McDonald's Omega | .940 (.937 - .943) | .927 (.921 - .932) |
| Stand. Error of Measurement | 3.35 | 1.46 |
| 95% CI | ± 7 | ± 3 |
| Min./Max. Values | 11 - 66 | 4 - 24 |

Note. n = 3,855; CI = Confidence Intervals.

## DISCUSSION

The aim of this study was to review the psychometric properties of the SDS and to create a brief version of the SDS. The results of the CFAs provide important insights into the psychometric properties of the SDS and SDS-BV. The findings indicate that while the SDS demonstrated strong fit based on incremental indices, its performance on absolute fit indices was less satisfactory. This discrepancy suggests that while the SDS aligns well with the theoretical structure when compared to a null model, some aspects of the model, such as residual discrepancies or complexity, may require refinement (Steiger, 2007). These findings align with prior literature indicating that RMSEA can be particularly sensitive to model complexity and sample size in large datasets (Chen, 2007; Cheung & Rensvold, 2002).

In contrast, the SDS-BV demonstrated consistently strong performance across both incremental and absolute fit indices, suggesting it is a robust and parsimonious alternative to the original SDS. The SDS-BV's superior performance, particularly in terms of residual discrepancy and overall fit, indicates a strong theoretical–data alignment. These results are consistent with previous research highlighting the benefits of shorter scales in reducing noise and improving fit without sacrificing reliability or validity (Smith et al., 2000).

The stronger fit of the SDS-BV may be attributed to its reduced complexity and the careful selection of

items, which likely minimized measurement error and multicollinearity (Smith et al., 2000). Furthermore, the excellent performance of incremental indices underscores the appropriateness of the SDS-BV for use in studies requiring a concise and psychometrically sound measure of social desirability.

In terms of measurement invariance, the results for the SDS and SDS-BV provide robust evidence supporting their applicability across diverse demographic and organizational groups. Both scales demonstrated strong measurement invariance across gender, age, education level, job position, employment type, and organization type. This invariance ensures that the scales measure the construct of social desirability equivalently across different groups, allowing meaningful comparisons. For the SDS, configural, metric, and scalar invariance were achieved with consistently strong fit across all group comparisons. Changes in model fit with added constraints were minimal, affirming the stability of the SDS across diverse subgroups and ensuring its utility for cross-group comparisons in research and practice.

Similarly, the SDS-BV exhibited exceptional measurement invariance, with indices remaining at their upper limits across all levels of invariance. Residual-based indices further supported this robustness, consistently demonstrating strong to very strong fit. The minimal changes in fit across all groups reinforce the scale's ability to maintain its measurement properties even with reduced complexity. These results highlight the SDS-BV as a psychometrically sound alternative to the SDS, offering enhanced parsimony without compromising reliability or validity. Establishing measurement invariance ensures that a scale measures the same construct equivalently across different groups, allowing for meaningful comparisons (Chen, 2007; Cheung & Rensvold, 2002).

The correlation matrix between the SDS and SDS-BV with other constructs provides valuable insights into the scales' convergent and discriminant validity. As expected, the SDS and SDS-BV are highly correlated, reflecting their shared measurement of the same construct. This strong relationship confirms the equivalence of the two scales and underscores the SDS-BV's utility as a concise alternative to the longer SDS without significant loss of validity. The correlations between the SDS and SDS-BV and other variables offer further evidence of the scales' construct validity. Both scales demonstrated small but consistent negative relations with constructs such as work demands, depression, anxiety, exhaustion, and cynicism, suggesting that individuals with higher social desirability scores tend to report lower levels of these negative states or perceptions. These findings align with the theoretical understanding of social desirability, which often involves a tendency to portray oneself in a more favorable light, potentially leading to underreporting of negative emotions and stress-related constructs (e.g., Latkin et al., 2017). Interestingly, relations between the scales and positive constructs, such as vigor, dedication, absorption, and professional efficacy, were negligible or slightly negative. This suggests that social desirability does not strongly influence the reporting of positive states, further supporting the specificity of the scales in capturing socially desirable responding rather than general tendencies toward positivity. These results concur with some studies that attempted to relate social desirability with subjective happiness (e.g., Caputo, 2017; Kozma & Stones, 1987; Veenhoven, 1991). Those studies did not find evidence of social desirability bias in the literature on positive traits or emotions, such as happiness measures. Additionally, Konow and Earley (2008) provided experimental evidence showing no meaningful association between overall happiness measures and social desirability scores, likely because global happiness may be less susceptible to social approval bias than other positive emotions like gratitude. Therefore, since the three dimensions of work engagement are measures of positive organizational psychology, they appear not to be affected by social desirability bias in the way that admitting behaviors that might be socially undesirable can be, such as being depressed or stressed.

The SDS-BV, despite its brevity, exhibited slightly weaker relations with some constructs compared to the SDS. For example, its associations with depression and exhaustion were marginally lower. This slight reduction is expected due to the fewer items in the SDS-BV, which may result in reduced sensitivity to certain constructs (Smith et al., 2000). Nonetheless, the relations remained meaningful, indicating that the SDS-BV retains its capacity to assess the relevant aspects of social desirability effectively. The positive association between job support and both scales also

warrants attention. This suggests that individuals with higher social desirability scores might report higher levels of perceived support in their work environments, possibly reflecting a tendency to present their circumstances in a more favorable light. However, this relationship was weak, indicating that while social desirability may influence reporting, the effect is not substantial.

Finally, the SDS and SDS-BV obtained excellent reliability coefficients using Cronbach's alpha and McDonald's omega. This serves as a useful description of the degree of coherence of the constituent parts of a whole; in other words, how the items of the SDS and SDS-BV relate to each other (McCrae et al., 2011). Moreover, the levels achieved are excellent, especially considering the interplay between the number of items on the scale and the sample size (Ponterotto & Ruckdeschel, 2007). Because the coefficients are high, the possibility of measurement error can be considered low. These levels indicate the suitability of the SDS and SDS-BV for both group applications and contexts where decisions on individual subjects are needed (Ponterotto & Ruckdeschel, 2007). Given the similarity of $\alpha$ and $\omega$, it is presumed that any differences between the factor loadings were trivial (Hayes & Coutts, 2020) and did not have a meaningful effect on the distance between one coefficient and the other. The degree of equality of the factor loadings of the items is frequently connected with this distance, a condition known as tau-equivalence to validate the coefficient (Green & Yang, 2009; Hayes & Coutts, 2020). This closeness implies that the assessment of internal consistency may be done well with the coefficient, without the need for SEM modeling to estimate it (Rosario-Hernández et al., 2021). Also, descriptive statistics were calculated, such as mean and standard deviation, to facilitate researchers in comparing their results with the sample used as a reference.

### Implications for Research and Practice

These results highlight the psychometric strengths of the SDS and SDS-BV as reliable measures of social desirability with broad applicability in research and practice. The SDS-BV offers a more concise alternative that maintains comparable validity, making it especially valuable in large-scale surveys or contexts where brevity is critical. Furthermore, the modest relations with negative constructs underscore the importance of considering social desirability bias in studies involving self-reported measures of psychological distress or workplace stress (Caputo, 2017). The slight reductions for the SDS-BV suggest that researchers should carefully consider their choice of scale based on the specific context and purpose of their study (Smith et al., 2000). While the SDS-BV offers efficiency, the SDS may provide greater sensitivity to certain constructs due to its larger number of items.

### Limitations and Recommendations

Although these findings provide robust evidence for the validity of the SDS and SDS-BV, there are limitations that warrant consideration. First, the population representativeness is not guaranteed, because the convenient sample selection of the workers did not corroborate the population similarity of employees in organizations in Puerto Rico. However, the sample was sufficiently large and heterogeneous. Second, the stability of the scores was not evaluated for reliability; to complete the evaluation of this element, it should investigate the reproducibility of scores over time using a test–retest approach. The reliance on self-reported data may amplify the influence of social desirability bias, potentially inflating or deflating relations with other constructs (Caputo, 2017). On the other hand, future research should explore the invariance and validity of these scales across cultural contexts and different populations to ensure their generalizability. Additionally, exploring the utility of the SDS-BV in more nuanced or domain-specific constructs may help further establish its applicability.

## CONCLUSION

The SDS and SDS-BV demonstrated strong validity in their relationships with a variety of psychological and organizational constructs, confirming their utility in measuring social desirability across diverse contexts. The SDS-BV, with its brevity and robust psychometric properties, emerges as a practical tool for researchers and practitioners, enabling efficient yet reliable assessments of socially desirable responding. These findings reinforce the relevance of social desirability scales in psychological research while underscoring the need for careful interpretation of their influence on self-reported data.

**Financing:** This research was not funded by any entity or sponsor.

**Conflict of Interest:** The authors declare that there are no conflicts of interest related to this research.

**Approval of the Institutional Board for the Protection of Human Subjects in Research:** Institutional Review Board of Ponce Health Sciences University (protocol number 2006040219).

**Informed Consent:** All participants completed an informed consent form.

**Review Process:** This study has been reviewed by external peers in double-blind mode.

**Statement of the use of Generative Artificial Intelligence:** No generative artificial intelligence was used in the drafting, analysis, or preparation of the manuscript.

## REFERENCES

Arrieta, J., Aguerrebere, M., Raviola, G., Flores, H., Elliott, P., Espinosa, A., Reyes, A., Ortiz-Panozo, E., Rodriguez-Gutierrez, E. G., Mukherjee, J., Palazuelos, D., & Franke, M. F. (2017). Validity and utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: A cross-sectional study. *Journal of Clinical Psychology, 73*(9), 1076-1090. https://doi.org/10.1002/jclp.22390

Aslan, J., Cova, F., Saldivia, S., Bustos, C., Inostroza, C., Rincón, P., Ortiz, C., & Bühring, V. (2020). Psychometric Properties of the Patient Health Questionnaire-9 in Elderly Chilean Primary Care Users. *Frontiers in psychiatry*, *11*, 555011. https://doi.org/10.3389/fpsyt.2020.555011

Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología, 29*(3), 1038-1059.

Bravo, D.M., Súarez-Falcón, J.C., Bianchi, J.M., Segura-Vargas, M.A., & Ruz, F.J. (2021). Psychometric properties and measurement invariance of the Maslach Burnout Inventory-General Survey in Colombia. *International Journal of Environmental Research and Public Health, 18*(10), 5118. https://doi.org/10.3390/ijerph18105118

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.

Caputo, A. (2017). Social desirability bias in self-reported well-being measures: evidence from an online survey. *Universitas Psychologica, 16*(2). https://doi.org/10.11144/Javeriana.upsy16-2.sdsw

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504. https://psycnet.apa.org/doi/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cropley, M., Michalianou, G., Pravettoni, G., and Millward, L. J. (2012). The relation of post-work ruminative thinking with eating behaviour. *Stress Health 28*, 23-30. https://doi.org/10.1002/smi.1397

DeVellis, R. F. (2017). *Scale Development: Theory and Applications*. SAGE Publications.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Fernández Arata, M., Merino Soto, C., & Guimet Castro, M. (2015a). Propiedades psicométricas del Maslach Burnout Inventory-General Survey en una muestra de docents de Lima (Perú). *Homenaje a Reynaldo Alarcón*, 371-391. Editorial Universal.

Fernández Arata, M., Juárez-García, A., & Merino Soto, C. (2015b). Análisis structural e invarianza de medición del MBI-GS en trabajadores peruanos. *Liberabit, 21*(1), 9-20.

Flores Jiménez, Merino Soto, C., Camacho Ávila, , A., Juárez-García, A., Placencia Reyes, O. (2015). Síndrome de burnout en instructors comunitarios: Propiedades psicométricas del Maslach Burnout Inventory- General Survey (MBI-GS). En A. Juárez-García (Eds.), *Investigaciones psicométricas de escalas psicosociales en trabajadores mexicanos* (pp. 279-308). Plaza y Valdés Editores.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.2307/3151312

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385-400. https://doi.org/10.1016/0191-8869(86)90014-0

George, D., & Mallery, P. (2016). *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference*. Routledge.

Green, S.B. & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika, 74*, 155-167. https://doi.org/10.1007/s11336-008-9099-3

Grimm, P. (2010). Social Desirability Bias. In J. Sheth & N. Malhotra (Eds.), *Wiley International Encyclopedia of Marketing*. John Wiley & Sons. https://doi.org/10.1002/9781444316568.wiem02057

González-Rivera, J. A. (2019). Validation and dimensionality of Patient Health Questionnaire for Depression (PHQ-8 and PHQ-9) in Hispanic LGBT+ Community. *International Journal of Recent Scientific Research, 10*(12), 36670-36676. http://recentscientific.com/sites/default/files/15519-A-2019.pdf

González-Rivera, J. A. (2025). *Metodología de la Investigación Aplicada a la Psicología: Una Guía para Estudiantes y Principiantes* (2da edición). Innova Psychological Institute.

Hayes, A. F. & Coutts, J. J. (2020). Use omega rather than cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1-24. https://doi.org/10.1080/19312458.2020.1718629

Hernández, C.R., Rosario-Hernández, E., & Lorenzo-Ruíz, A. (2020). Propiedades psicométricas de la Escala para una Práctica Profesional Basada en la Evidencia en una muestra de

psicólogos clínicos en la República Dominicana. *Revista Caribeña de Psicología, 4*(3), 204-216. https://doi.org/10.37226/rcp.v.4i3.5291

Howard, M.C., Boudreaux, M., Cogswell, J., Manix, K.G., & Oglesby, M. (2024). A literature review of model fit and model comparison with confirmatory factor analysis: Formalizing the informal in organizational science. Applied Psychology, 74(1), e12592. https://dor.org/10.1111/apps.12592

Hu, L.T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Kline, R.B. (2016). *Principles and practice of structural equation modeling*. The Guilford Press.

Konow, J. & Earley, J. (2008). The Hedonistic Paradox: Is Homo Economicus Happier? *Journal of Public Economics, 92*, 1-33. http://dx.doi.org/10.1016/j.jpubeco.2007.04.006

Kozma, A., & Stones, M. J. (1987). Social desirability in measures of subjective well-being: a systematic evaluation. *Journal of gerontology*, *42*(1), 56-59. https://doi.org/10.1093/geronj/42.1.56

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive behaviors*, *73*, 133–136. https://doi.org/10.1016/j.addbeh.2017.05.005

Li, C.H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavioral Research Methods, 48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21*(3), 369-387. https://doi.org/10.1037/met0000093

Martínez Alvarado, L.Y., Rosario-Hernández, E., & Rovira-Millán, L.V. (2017). La relación entre la inseguridad laboral y el bienestar psicológico en una muestra de asistentes de vuelo: El papel moderador del engagement en el trabajo. *Ciencias de la Conducta, 32*(1), 99-127.

Maslach, C., Jackson, S. E., and Leiter, M. P. (1996). Maslach Burnout Inventory manual. Consulting Psychologists Press.

Maslach, C., & Schaufeli, W. (1993) Historical and conceptual development of burnout. In Schaufeli, W., Maslach, C., & Marek, T. (Ed.), *Professional Burnout: Recent Developments in Theory and Practice*. Taylor & Francis.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28-50. https://doi.org/10.1177/1088868310366253

Merino-Soto, C., Angulo-Ramos, M., Rovira-Millán, L.V., & Rosario-Hernández, E. (2023). Psychometric properties of the generalized anxiety disorder-7 (GAD-7) in a sample of workers. *Frontiers in Psychiatry*, *14*, 999242. https://www.frontiersin.org/articles/10.3389/fpsyt.2023.999242/full

Merz, E. L., Malcarne, V. L., Roesch, S. C., Riley, N., & Sadler, G. R. (2011). A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural diversity & ethnic minority psychology*, *17*(3), 309-316. https://doi.org/10.1037/a0023883

Muthén, L. K., and Muthén, B. O. (1998-2012). *Mplus User's Guide*. Muthén & Muthén.

Nederhof,A.J. (1985). Methods of coping with social desirabilitybias: A review. European *Journal of Social Psychology, 15*(3), 263-280. https://doi.org/10.1002/ejsp.2420150303

Pagán-Torres, O.M., González-Rivera, J.A., & Rosario-Hernández, E. (2020). Reviewing the psychometric properties and factor structurre of the Generalized Anxiety Disorder-7 (GAD-7) in a sample of Puerto Rican Adults. *International Journal of Recent Scientific Research, 11*(1), 36885-36888. https://doi.org/10.24327/IJRSR

Parmač Kovačić, M., Galić, Z., & Jerneić, Ž. (2014). Social desirability scales as indicators of self-enhancement and impression management. *Journal of Personality Assessment*, *96*(5), 532-543. https://doi.org/10.1080/00223891.2014.916714

Pons, J.I, Rosario-Hernández, E. & Mañón, S. (2009). Inteligencia Emocional: Hacia el Desarrollo y validación de un cuestionario para la evaluación de dimensiones de inteligencia emocional en adultos puertorriqueños. *Revista Puertorriqueña de Medicina y Salud Pública, XVII*, 55-62.

Ponterotto, J. G. & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perception and Motor Skills, 105*, 997-1014. https://doi.org/10.2466/pms.105.3.997-1014

Quiñones, M.M., Haddock, R.M. & Rosario-Hernández, E. (2015). Development and validation of an emotional intelligence scale for Puerto Rican students. *Revista de Salud y Conducta Humana, 2(1)*, 23-34.

Quiñones-Freire, C., Vara, M.D., Tomás, J.M., & Baños, R.M. (2021). Psychometric properties of the Spanish version of the Patient Health Questionnaire-9 in users of the Ecuadorian public health care system. *Revista Latinoamericana de Psicología, 53*, 210-217. https://doi.org/10.14349/rlp.2021.v53.23

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research, 30*(3), 359-383. https://doi.org/10.1207/s15327906mbr3003_4

Rivera, E., Pons, J.I., Rosario-Hernández, E. & Ortiz, N. (2008). Traducción y adaptación puertorriqueña del Inventario Bar-On de Cociente Emocional (Bar-On EQ-i): Análisis de propiedades psicométricas. *Revista Puertorriqueña de Psicología, 19*, 148-182.

Rodríguez Montalbán, R., Martínez Lugo, M.E., & Sánchez-Cardona, I. (2014). Análisis de las Propiedades de la Escala de Engagement en el Trabajo de Utrecht en una muestra de trabajadores en Puerto Rico. *Universitas Psychologica, 13*(4), 1255-1266. https://dx.doi.org/10.11144/Javerina.UPSY13-4.appu

Rosario-Hernández, E. & Rovira-Millán, L.V. (2002). Desarrollo y validación de una escala para medir actitudes hacia el retiro. *Revista Puertorriqueña de Psicología, 13*, 45-60.

Rosario-Hernández, E. & Rovira Millán, L.V. (2004). Desarrollo y validación de la Escala de Ciudadanía Organizacional. *Revista Puertorriqueña de Psicología, 15*, 1-25.

Rosario-Hernández, E. & Rovira Millán, L.V. (2007). Desarrollo y validación de la Escala de Justicia Organizacional. *Revista Ciencias de la Conducta, 22(1), 61-78*.

Rosario-Hernández, E. & Rovira Millán, L.V. (2008a). Desarrollo y validación del Índice de Conductas Laborales Contraproducentes. *Revista Interamericana de Psicología Ocupacional*, 2008, *27(1)*, 16-27.

Rosario-Hernández, E. & Rovira Millán, L.V. (2008b). Desarrollo y validación de la Escala de Contrato Psicológico Organizacional. *Revista Caribeña de Psicología*, *1(1)*, 16-23.

Rosario-Hernández, E. & Rovira Millán, L.V. (2008c). Desarrollo y validación del Inventario de Estilos de Personalidad - Rosario Rovira (IEP-RR). *Revista Interamericana de Psicología Ocupacional, 27(2)*, 103-118.

Rosario-Hernández, E. & Rovira Millán, L.V. (2011). Desarrollo y validación de la Escala de Incivilidad Laboral. *Revista Caribeña de Psicología*, *2 (1)*, 1-14.

Rosario-Hernández, E. & Rovira Millán, L.V. (2012). Desarrollo y validación de la Escala de Clima de Seguridad y Salud Ocupacional Percibida. *Revista Interamericana dePsicología Ocupacional*, *31 (1)*, 46-58.

Rosario-Hernández, E. & Rovira Millán, L.V. (2014). Desarrollo y validación del Inventario del Modelo Demandas-Control-Apoyo. *Revista Interamericana dePsicología Ocupacional*, *33 (2)*, 108-127.

Rosario-Hernández, E., Rovira Millán, L. V., Báez Tellado, E. M. & Soto Franceschini, J. A. (2011). Desarrollo y validación de la Escala de Actitudes Hacia las Personas Viejas en el Trabajo. *Revista Ciencias de la Conducta, 26 (1)*, 29-48.

Rosario-Hernández, E., Rovira Millán, L.V., & Merino-Soto, C. (2021). Review of the Internal Structure, Psychometric Properties, and Measurement Invariance of the Work-Related Rumination Scale – Spanish Version. *Frontiers in Psychology, 12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.774472

Rosario-Hernández, E., Rovira Millán, L.V., Vélez Ramos, J., Cruz, M., Vélez, E., Torres, G., Alomar, G., Caldera, J., Santiago, M., Arroyo, Y., Sánchez, I., & Rodríguez, N. (2018). Effect of the exposure to workplace bullying on turnover intention and the mediating role of job satisfaction, work engagement, and burnout. *Revista Interamericana de Psicología Ocupacional, 37*(1), 26-51.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1-36. https://www.jstatsoft.org/v48/i02/

Rovira-Millán, L.V., Blanco-Rovira, R.A., López-Iglesias, A.C., & Rosario-Hernández, E. (2023). Suicidal behavior at work scales: Development and validation of the Work-Related Suicidal Ideation, Defeat, and Entrapment Brief Scales. *Revista Evaluar, 23*(2), 1-18.

Rovira-Millán, L.V. & Rosario-Hernández, E. (2018). Desarrollo y validación del Indicador del Bienestar del Sueño. *Revista Puertorriqueña de Psicología, 29*(2), 348-362.

Schaufeli, W., & Bakker, A. (2003). *Utrecht work engagement scale: Preliminary manual*. Occupational Health Psychology Unit, Utrecht University.

Schaufeli, W.B., Salanova, M., González-Romá, V., Bakker, A.B. (2002). The measurement of engagement and burnout: A confirmative analytic approach. *Journal of Happiness Studies, 3*, 7-92. https://doi.org/10.1023/A:1015630930326

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological assessment*, *12*(1), 102–111. https://doi.org/10.1037//1040-3590.12.1.102

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. SAGE Publications.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893-898. https://doi.org/10.1016/j.paid.2006.09.017

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69. https://doi.org/10.1177/109442810031002

van de Mortel, T.F. (2008) Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*, 40-48.

Veenhoven, R. (1991). Questions on happiness: Classical topics, modem answers, blind spots. In E. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective well-being: An interdisciplinary perspective* (pp. 7-26). Pergamon.

Wang, J. & Wang, X. (2012). *Structural Equation Modeling Applications Using Mplus.* Chichester: John Wiley & Sons Ltd. http://dx.doi.org/10.1002/9781118356258

Zijlstra, W.P., van der Ark, L.A., & Sijtsma, K. (2011). Outliers in test and questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral statistics, 36*(2), 186-212. https://doi.org/10.3102/1076998610366263