ACADEMIC PROCRASTINATION SCALE: PSYCHOMETRIC EVIDENCE FOR THE ITALIAN VERSION BASED ON CFA AND IRT MODELS

Rosa Fabbricatore
Department of Economics and Statistics, University of Naples Federico II, Naples, Italy

Anna Parola
Department of Humanities, University of Naples Federico II, Naples, Italy

Giuliana Pepicelli, Francesco Palumbo
Department of Political Sciences, University of Naples Federico II, Naples, Italy

Abstract. *This study aimed to translate and validate the Academic Procrastination Scale – Short Form (APS-SF) in the Italian context. Specifically, this study focused on the psychometric properties of validity and reliability in a sample of N=340 university students. Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT) models were performed. CFA and IRT models showed the unifactorial model proposed in the original version of the scale. The IRT approach also provided evidence for satisfactorily test information and the measurement invariance for gender. These findings confirm that the Italian version of APS-SF is a valid and reliable measure to assess procrastination in Italian academic students.*

Keywords: *Procrastination, Higher education, Confirmatory factor analysis, Item response theory.*

1. INTRODUCTION

Procrastination refers to needlessly task delaying that is followed by the experience of subjective discomfort (Solomon and Rothblum, 1984). From a historical perspective, the term procrastination derives from the Latin adjective crastĭnus, translated as "of tomorrow"; it did not necessarily have a negative connotation until the mid-eighteenth century (Ferrari et al., 1995). Nowadays, it

is considered a dysfunctional tendency with moral connotations (Sabini and Silver, 1982) and psychological consequences as experienced guilt, decreasing performance outcomes, and social impact such as being considered unreliable. In attempting to understand the procrastination construct, researchers generally considered procrastination a stable disposition, a fixed personal trait that does not vary according to tasks, contexts, and time (Ferrari et al., 1995; Schouwenburg, 1995; Van Eerde, 2000).

Everyone may procrastinate sometimes but, according to this perspective, procrastination is defined as an intra-individual process depending by own norms and moral principles and consisting of delaying actions or decisions even if it is not appropriate to a particular situation (Burka and Yuen, 1983; Milgram et al., 1998). However, not every late-performing task is procrastination: indeed, the delay can be used as a wise strategy if purposely planned. Instead, procrastination refers to the act of postponing needless the implementation of what was planned (Van Eerde, 2003). Moreover, other studies presented empirical support for its temporal and situational variability. They demonstrated that procrastination is better described as a situation-specific behavior that changes over time and depends on the tasks and contexts interaction (e.g., Blunt and Pychyl, 2000; Lonergan and Maher, 2000; Milgram et al., 1988).

The amount of procrastination literature originated in studies with college students because they tended to delay writing term papers, preparing for tests, studying for examinations, and keeping up with weekly reading assignments (Burka and Yuen, 1983; Ellis and Knaus, 1977; Onwuegbuzie and Jiao, 2000). Several studies showed that over 70% of college students engage in procrastination (Ellis and Knaus, 1977) and regularly procrastinate (Ferrari, O'Callaghan, and Newbegin, 2005; Rabin et al., 2011). In this field of research, academic procrastination has been defined as the students' tendency to delay academic tasks intentionally, although it will carry out negative consequences (Simpson and Pychyl, 2009; Steel, 2007). Furthermore, other studies found a strong relationship between academic procrastination and academic performance (e.g., Beswick et al., 1988; Steel et al., 2001). Several scholars underlined the adverse outcomes of procrastination in (higher) educational settings such as psychological distress, anxiety, reduced well-being, low academic performance, and avoidance of social relations (Kim and Seo, 2015; Krause and Fruend, 2014). In a ground theory study examining the process by

which academic procrastination occurs, students involved in the research reported that they procrastinate to manage their time better, reduce boredom, and work more efficiently. Moreover, students had a low perception that procrastination can negatively impact learning achievement (Schraw et al., 2007). Several studies investigated the relationship between academic procrastination and perceived self-efficacy. They pinpointed that perceived self-efficacy is negatively related to procrastination (Steel, 2007). Accordingly, procrastinating students presented weak self-efficacy beliefs (Lindblom et al., 2015). Instead, students' perception of a high self-efficacy in an academic context seemed to be strictly linked to a positive quality of studying and academic performance (Caprara et al., 2008).

Recent studies in the Italian academic context have shown that procrastination represented a failure in regulating students' behavior (Limone et al., 2020). Regarding the outcomes of academic procrastination, some studies show a strong relationship between procrastination and students' academic performance in secondary school (De Paola and Scoppa, 2015) and university (De Paola and Scoppa, 2015, Limone et al., 2020). Furthermore, in a recent study, Licata et al. (2024) suggested that the harms of procrastination are not limited to academic performance, but also include effects on health and well-being. In this sense, an assessment of the academic procrastination experienced by students turns out to be a good starting point for conceivable programs to enhance a student's academic success and well-being.

McCloskey (2011) proposed the first theorization of academic procrastination and the Academic Procrastination Scale (APS), consisting of six main dimensions: psychological beliefs about abilities, attention distractions, social factors, time management skills, personal initiative, and laziness. Starting with this scale (25 items). The APS was validated and used in the Italian context by Limone et al. (2020). Yockey has tested the Academic Procrastination Scale - Short Form (APS-SF) that consisted of five items. The original factor analysis showed a single-factor solution, a good factorial load of the five items, and a Cronbach's alpha of .87. The measurement invariance suggests the validity for males and females.

As things stand, the APS-SF has been tested with samples of American (Yockey, 2016) and Spanish-speaking (Brando Garrido et al., 2020) academic students. To our knowledge, there is no Italian validation study of the Short Form of the Academic Procrastination Scale (APS-SF; Yockey, 2016). In

recent years, the need for short and valid assessment instruments has been highlighted in research practice. The 25 items of the long version of the APS, despite their undoubted usefulness, do not always allow for a short and routine assessment of academic procrastination. It can be burdensome to administer, making it difficult to include in longer assessment batteries (Moreira et al., 2020). In line with the validation of the APS-SF in other countries (e.g. Brando-Garrido et al., 2020, Yockey, 2016), the purpose of the short version of the procrastination measure is to provide a measure that can be answered by subjects in a short time, especially in contexts where the administration time cannot be unduly prolonged, in a long psychological battery with several other instruments, and to avoid fatigue or exhaustion among participants.

Therefore, this study aimed to validate the APS-SF in the Italian context by examining the psychometric properties (factor structure, item fit, convergent validity, internal consistency) of the APS-SF using a sample of university students.

2. METHODS

2.1 STUDY DESIGN

Following the guidelines for the cross-cultural adaptation of questionnaires and rating scales (Beaton et al., 2000; Van De Vijver and Hambleton, 1996), the APS-SF was translated from English to Italian by three expert psychologists, psychometry and teaching and back-translated by an independent English translator to ensure cross-cultural equivalence.

The sample size was determined according to the adaptation to the confirmatory factor analysis (CFA), taking into account the recommendations to use between 5 and 20 subjects per item, with a minimum of 100-150 subjects in the case of short scales. For item response theory (IRT) models, authors suggested that a sample of 200 or fewer subjects could be adequate for differential item functioning (DIF) detection and scale properties evaluation, especially when IRT assumptions are met (Edelen and Reeve, 2007).

2.2 PARTICIPANTS AND PROCEDURE

Data collection was conducted in 2021 using a convenience sample of 340 Italian university students enrolled in a bachelor's degree program in

Psychology. The sample included 72 males and 268 females (M_{age}= 20.75, SD_{age}= 3.50), primarily from high schools (12% classical high schools, 30% scientific high schools, 42% other high schools, 12% technical or professional schools, 4% other institutions). The inclusion criteria were the minimum age of 18, the status of university students, and the Italian-language speakers. Before starting the data collection, researchers briefed the participants on the scope of the study, and their agreement was kindly requested. Students who agreed to participate completed an online self-report questionnaire via the MOODLE platform. This study complies in full with the recommendations in the ethical guidelines of the Italian Psychologists Association for research.

2.3 MEASURES

Academic Procrastination Scale Short Form (APS-SF; Yockey, 2016). The instrument consists of five items that assess academic procrastination. Participants responded to each item on a five-point Likert scale from 1 (total agreement) to 5 (complete disagreement). The total score is defined as the sum of the items (range 5-25), where higher scores indicate a greater tendency to postpone academic tasks.

Procrastination Scale (TPS; Tuckman, 1991; Limone et al., 2020). The instrument consists of 16 items that refer to procrastinating behaviors. Participants responded to each item on a four-point Likert scale from 1 (strongly disagree) to 4 (strongly agree). TPS includes two dimensions of procrastination: a general explanation of procrastination and the likelihood to avoid difficult or unpleasant tasks. The score is defined as the sum of the item scores (range 16-64), where higher scores indicate a higher procrastination tendency. For this study, Cronbach's alpha was .78.

Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich and DeGroot, 1990; Bonanomi et al., 2018). The self-regulation learning strategies (SRLS) scale was administered for this study; it consists of 22 items and comprises two subscales assessing Cognitive strategy use (12 items) and Self-regulation (10 items). The participants responded to each item on a 5-point Likert scale from 1 (not at all appropriate for me) to 5 (very appropriate for me). The arithmetic mean of the related items is the score of the subscale. For this study, the overall Cronbach's alpha was .75.

2.3 DATA ANALYSIS

The psychometric assessment of the APS-SF (Italian version) included Confirmatory Factor Analysis (CFA), item response theory (IRT) models for differential item functioning detection and item fit evaluation, internal consistency analysis, and validity. The classical test approach and IRT allow for a more robust psychometric analysis. In particular, the IRT is a model-based approach aiming to estimate the probability of selecting different item response categories according to the individual level of latent trait (academic procrastination in our case) and item characteristics (difficulty and discrimination). This approach is particularly advantageous in educational and psychological assessment because it provides item- and person-level insights, addressing limitations of traditional methods such as Classical Test Theory (CTT), which assumes uniform measurement properties across items (Edelen and Reeve, 2007). Very briefly, the latent trait (usually indicated with θ) can be defined as an unobservable construct measured by means of several manifest variables (e.g., items of a Likert scale); the item difficulty refers to the level of latent trait needed to have a probability $p = .5$ of endorsing the item, and when polytomous items are considered a different difficulty parameter is estimated for each threshold (where the number of thresholds is equal to the number of levels minus 1); the item discrimination expresses the item's ability to distinguish the different levels of the latent trait in the subjects.

The IRT approach provides a collection of techniques for an accurate item- and test-level evaluation, which have proven to be highly useful in scale validation studies. Indeed, it offers robust tools for identifying items that best capture the latent trait and determining the measurement precision across different latent trait levels, which is crucial for both research and practical applications in educational assessment (Edelen and Reeve, 2007). In this theoretical framework, several models have been proposed for both binary and ordinal item responses (see Bartolucci et al. (2015), Tutz (2020), and Van der Linden (2018) for an overview). These models differ primarily in the constraints imposed on item parameters (e.g., free or constrained discrimination parameters across items). Additionally, for ordinal items, the relationship between observed indicators and the latent variable can be modeled using either a global (cumulative) logit or a local (adjacent categories) logit, which

correspond to the graded response model (GRM) and the generalized partial credit model (GPCM), respectively.

In general, traditional IRT models are grounded on three main assumptions required to be met before the model estimation: one-dimensionality, local independence, and monotonicity. Moreover, also the measurement invariance must be ensured to reach comparable measures.

Confirmatory Factor Analysis (CFA) was performed to assess the factorial structure and, thus, the one-dimensionality of the latent trait. According to the response scale, the diagonal weighted least square (DWLS) estimator was used as a robust method for non-normality. For the model fit, the study considered the following indexes and test: the Chi-square test statistics, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMS). As a general rule, a model satisfactorily fits if the CFI and TLI coefficients are .90 or above (Medsker et al., 1994); RMSEA approximately .08 or less (Byrne, 2010), and SRMS .10 or less. The CFA was performed using the R package `lavaan` (Rosseel, 2012, v. 0.6-6).

The residual correlation matrix of the CFA model with one factor allows checking the second IRT assumption of local independence. This assumption implies that the responses to a set of items are independent, conditional to a latent trait (variable). Thus, there should be no significant correlation between item residuals (weak independence condition). About that, correlations below the threshold of .2 are assumed as evidence of independence (Chen and Thissen, 1997). Moreover, a non-parametric IRT model relying on the Mokken scaling was used to assess the monotonicity assumption, consisting of a non-decreasing probability of endorsing higher item response categories as the level of latent trait increases. The scalability coefficient H for each item and the total scale was calculated using the R package `mokken` (Van der Ark, 2007). Values greater than .30 for items and greater than .50 for the entire scale were considered acceptable for monotonicity assurance (Mokken, 1971).

Finally, according to the measurement invariance definition, item parameters are assumed to be invariant across different groups of subjects (e.g., males and females). Violations of this assumption lead to what is known as differential item functioning (DIF; Ackerman, 1992). The DIF occurs when subjects with the same latent trait level but from different groups have a different probability of endorsing the item response categories. The DIF can be

consistent along with all the latent traits (uniform DIF) or vary in direction or magnitude according to different latent trait levels (non-uniform DIF). In this study, both uniform and non-uniform DIF for gender were evaluated by exploiting ordinal logistic regression models and considering a change of McFadden's pseudo R^2 equal to .02 as the critical value. The R package `Lordif` (Choi et al., 2011) was used for this aim. In addition to DIF, differential test functioning (DTF) was also examined to assess whether there is overall bias at the test level with respect to gender. The omnibus measure proposed by Chalmers et al. (2016), referred to as *signed* DTF (sDTF), was employed. This measure identifies whether one group consistently scores higher across a defined range of θ. Values close to 0 indicate negligible or no DTF. The measure is implemented in the R package `mirt` (Chalmers, 2012).

It is worth noting that before testing for measurement invariance, the most appropriate link function (global logit or local logit) was selected by comparing the fit of the two most general models, namely the GRM (global logit) and the GPCM (local logit), through the Bayesian Information Criterion (BIC; Schwarz, 1978). Afterward, the model selected according to the most appropriate link function was compared with its constrained formulations (see Bartolucci et al. (2015) for a review of IRT models) to test if some item constraints should be set to describe the APS-SF properties better. The IRT models at issue were estimated using the R package `mirt` (Chalmers, 2012), and the BIC was considered the selection criterion.

Once the best model was selected, Orlando and Thissen's S-X^2 was used to evaluate polytomous items fitting (Kang and Chen, 2008). An S-X^2 p-value $\leq$ .001 indicates a poor fit.

Besides, also the following graphical diagnostic tools for IRT model assessment were employed to explore deeper the goodness of the items and the scale: Item information curve (IIC), test information curve (TIC), and item response category characteristic curves (IRCCCs). These curves are a function of the item parameters and allow for the evaluation of item and test measurement precision across the entire range of the latent trait, with a particular focus on the critical area of the construct continuum (in this case, high levels of academic procrastination) (Edelen and Reeve, 2007).

The internal consistency of the factor was measured using Cronbach's alpha coefficient. The correlations between APS-SF and TPS and MLSQ were computed to test the validity of the instrument. Following the recommendations

of Clark and Watson (2019), Flake et al. (2017), and testing standards (AERA, APA, and NCME, 2014), the relationships between academic procrastination and other measures, such as procrastination and self-regulated learning strategies, are used to assess the same or similar constructs and provide convergent evidence. In particular, the correlation between the APS-SF and the TPS measures the same psychological dimension. The relationship between self-regulation learning strategies and the APS highlights the relationship between two correlated psychological dimensions (e.g. Limone et al., 2020).

3. RESULTS

Results include the estimated parameters for the APS-SF and fit indices demonstrating that the model adequately fits the data: $\chi 2(5) = 13.467$, $p = .01$; TLI = .996; CFI = .998; RMSEA = .071 [.026 - .117], SMSR = .034. These values confirmed the adequacy of the unidimensional structure of the instrument. Item arithmetic means, standard deviations, skewness, kurtosis, and standardized factor loadings are in Table 1.

The factor loadings varied from .81 to .87 (see also Figure 1), confirming the soundness of the items used to measure the APS-SF. In the estimated model, no correlations between the residuals of different items were allowed. In a subsequent step, the residual correlation matrix was extracted to verify the IRT local independence assumption. All item residual correlations were below the critical threshold of .2, confirming that the assumption was met. The monotonicity assumption was also confirmed: the scalability coefficient H ranged between .657 and .672 for the single items (see Table 2) and was equal to .670 for the total scale. Regarding measurement invariance for gender, McFadden's pseudo-R^2 changes ranged from .0001 to .0170, detecting neither uniform nor non-uniform DIF. Moreover, a negligible DTF emerged, with sDTF = .48 (95% CI: -.28 to 1.19), which was not significantly different from 0 (p-value = .21). At this step, the GRM (global logit) shows a better fit than the GPCM according to the BIC (3672.09 and 3680.86, respectively).

Table 1: Means, standard deviations of each item, and results of CFA in university students

Item	M	SD	SK	K	(λ)	R^2
I put off projects until the last minute (APS1).	2.33	1.17	0.46	-0.73	.83*	.69
I know I should work on schoolwork, but I just don't do it (APS2).	1.79	1.04	1.10	0.24	.85*	.72
I get distracted by other, more fun, things when I am supposed to work on schoolwork (APS3).	2.47	1.14	0.40	-0.54	.81*	.65
When given an assignment, I usually put it away and forget about it until it is almost due (APS4).	1.56	0.88	1.81	3.29	.85*	.72
I frequently find myself putting important deadlines off (APS5).	1.74	1.08	1.41	1.11	.87*	.76

Note. M = mean, SD = standard deviation, SK= Skewness, K= Kurtosis, λ = factor loading.

* $p < .001$

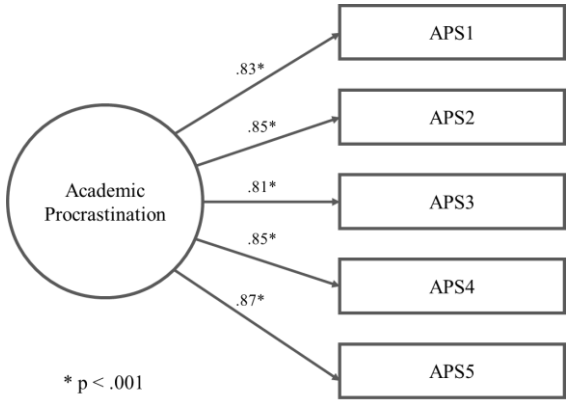

Figure 1: CFA results: Standardized estimates of the factor loadings

Afterward, by comparing GRM with its constrained formulations, namely the Rating Scale GRM (RS-GRM; Muraki, 1990), the one-parameter GRM (1P-GRM; Van der Ark, 2001), and the one-parameter RS-GRM (1P-RS-GRM; Van der Ark, 2001), the BIC indicated that the RS-GRM was the best model. In the RS-GRM, the item discrimination parameter can freely vary among items, whereas the item difficulty parameter is expressed in an additive way so that items have a different general difficulty level but equal threshold difficulty level. Table 2 reports the estimated parameters for the Italian version of the APS-SF.

Table 2: Scalability coefficient (H), RS-GRM item parameters, and model fit (S-X^2 p-value)

Item	*H*	S-X^2 p-value	RS-GRM item parameters					
			a	c	b_1	b_2	b_3	b_4
APS1	.672	.07	2.55	0	1.52	-0.55	-2.93	-5.12
APS2	.664	.02	2.79	-1.81	1.52	-0.55	-2.93	-5.12
APS3	.657	.64	2.21	0.39	1.52	-0.55	-2.93	-5.12
APS4	.666	.05	2.69	-2.67	1.52	-0.55	-2.93	-5.12
APS5	.672	.03	3.24	-2.24	1.52	-0.55	-2.93	-5.12

Note. a = item discrimination; c = item general difficulty; b1, b2, b3, b4 = threshold difficulty; $c_1 = 0$ for model identifiability.

Even for a low level of procrastination, the easiest item to endorse was Item 4 ("When given an assignment, I usually put it away and forget about it until it is almost due"). In contrast, Item 3 ("I get distracted by other, more fun, things when I am supposed to work on schoolwork") was the most difficult item that requires a high level of procrastination to encounter agreement. On the other hand, Item 5 ("I frequently find myself putting important deadlines off") allows discriminating different levels of procrastination better than the other items do.

Orlando and Thissen's S-X^2 index suggested no misfitting items (see Table 2). The item information curves in Figure 2b show that all items are centered on medium-high values of the latent trait and that the most informative one is Item 5, whereas Item 3 is the least informative. Overall, the information provided by the APS-SF scale is satisfactory and maximum for students with high levels of procrastination (the curve in Figure 2a is shifted on the right).

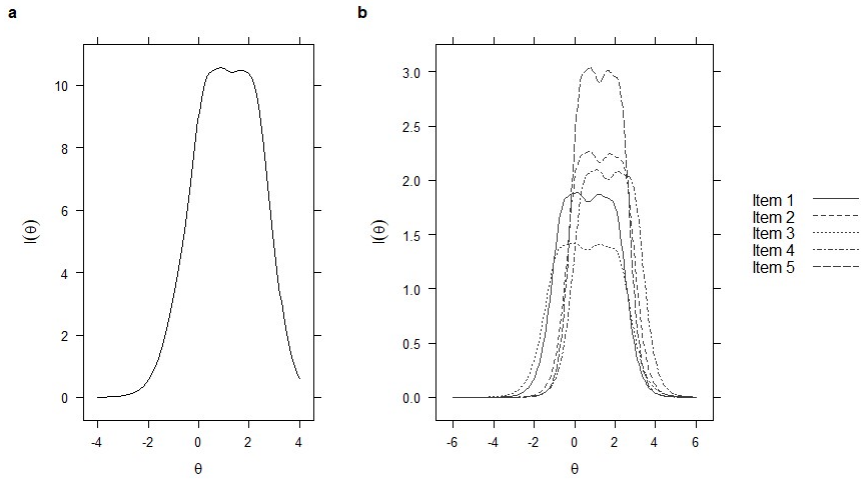

Figure 2: Test (a) and Item (b) information curves

Finally, the item response category characteristic curves (IRCCCs) in Figure 3 are well-shaped and the response categories well-separated, indicating that there are no redundant response categories needed to be merged. Moreover, for Item 4 and Item 5, students tend to strongly disagree also for a medium level of latent trait (see the IRCCC for the first category, represented by the solid line in the figure).

Cronbach's alpha was .88, confirming the reliability of the measure. The Italian form of APS-SF also showed adequate convergent validity correlated with APS (see Table 3). Moreover, APS-SF and APS showed the same negative relationship with SRLS.

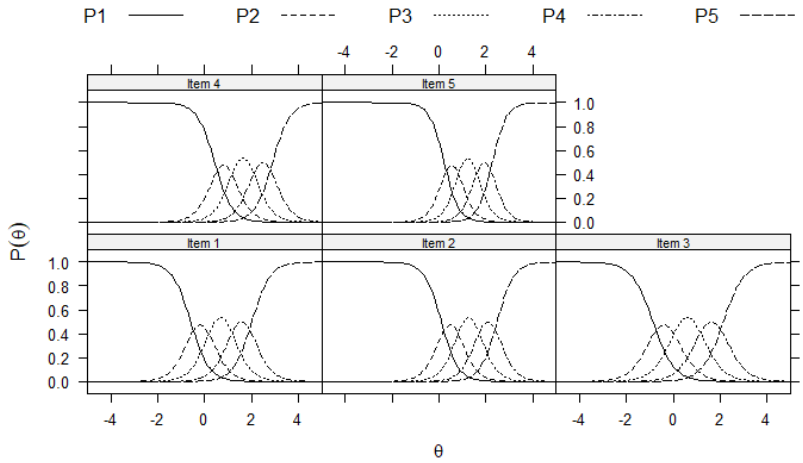

Figure 3: Item response category characteristic curves (IRCCCs): P1 = "Strongly Disagree", P2 = "Disagree", P3 = "Do not disagree or agree", P4 = "Agree", P5 = "Strongly Agree".

Table 3: Means, standard deviation, and bivariate correlations

Measures	*M*	*SD*	1	2
1. APS-SF	9.89	4.37	-	
2. TPS	33.97	7.21	.76**	-
3. SRLS	3.62	.40	-.27**	-.16*

Note. M = means, SD = standard deviations, APS-SF: Academic Procrastination Scale – Short Form, APS: Tuckman Procrastination Scale, SRLS: Self regulation Learning Strategies.

* p < .01, **p < .001

4. DISCUSSION

This study aimed to test the Italian version of the Academic Procrastination Scale – Short Form (APS-SF) psychometric characteristics. The Confirmatory Factor Analyses confirmed the unifactorial structure of the Italian version of the APS-SF, supporting its use in the Italian context. Cronbach's alpha indicated that the internal consistency of the APS-SF was satisfactory. The association between APS-SF, APS, and MLSQ were further investigated. The results confirmed the convergent validity of the APS-SF, showing the correlation between the APS-SF and APS and the negative relationship between APS-SF and MLSQ.

Item and test proprieties investigation through IRT models showed that the RS-GRM was the best model for describing the Italian version of the APS-SF. Both item and test information were satisfactory and maximum for students with high levels of procrastination. All items reported a more than satisfactory fit to the model; the response categories were well-separated and not redundant. Regarding the measurement invariance for gender, results provided evidence for the absence of uniform and non-uniform DIF, ensuring comparable measures for males and females.

Although the current study contributed to the literature by testing the psychometric properties of the Italian form of the APS-SF, some limitations and future perspectives have to be highlighted. A first limitation regards subjects involved in the study that mainly were female-gendered and homogenously aged, so further studies with wider and more varied samples are awaited. Moreover, the study was carried out only through a single-time assessment of students' procrastination. Therefore, future research will focus on longitudinal studies that would explore students' tendency to procrastinate throughout the whole academic year and the entire enrollment time, hopefully.

One of the limitations of this study was not to consider the relationships between academic procrastination and other variables linked to this construct. Further research should also explore these relationships in-depth, taking into account related psychological variables (such as self-efficacy, self-regulated learning strategies, self-esteem, anxiety, and stress), aiming to assess the procrastination construct in different and more varied conditions.

In conclusion, our findings demonstrate that the APS-SF is a psychometrically valid measure of procrastination in university students. In

terms of practical implication, the instrument can be used as a valid alternative to the long questionnaires composed of different scales, in particular, to reduce the risk of fatigue of the participants. Thanks to its length, this scale can be easily integrated into surveys or batteries of instruments.

Finally, the results of this study demonstrate that researchers can reliably use the APS-SF in future studies to evaluate academic procrastination and explore its associated factors, thereby gaining a deeper understanding of this behavior. Moreover, this measure of procrastination could also be beneficial for learners and teachers in educational contexts. Indeed, as highlighted in Martín-Puga et al. (2022), the APS-SF serves as a useful tool to identify students prone to procrastination, guiding the development of targeted interventions to help them manage this behavior and thus enhance their likelihood of achieving positive academic outcomes.

ACKNOWLEDGMENTS

Rosa Fabbricatore acknowledges the financial support provided by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, Inclusive and Sustainable project (GRINS PE00000018 - CUP E63C22002140007). The contribution of Giovanbattista Califano was funded by the European Union - NextGenerationEU (Call PRIN 2022 - Italian Ministry of University and Research - RECIPE - cod. 20224ZHNXE). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

REFERENCES

Ackerman, T.A. (1992). A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective. *Journal of Educational Measurement*. 29: 67-91.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (AERA, APA, & NCME). (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Bartolucci, F., Bacci, S., and Gnaldi, M. (2015). *Statistical Analysis of Questionnaires: A Unified Approach based on R and Stata*. Chapman & Hall, CRC Press, Boca Raton, FL, USA.

Beaton, D. E., Bombardier, C., Guillemin, F., and Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 25: 3186-3191.

Beswick, G., Rothblum, E. D., and Mann, L. (1988). Psychological antecedents of student procrastination. *Australian Psychologist*. 23: 207-217. https://doi.org/10.1080/00050068808255605

Blunt, A. K., and Pychyl, T. A. (2000). Task aversiveness and procrastination: A multidimensional approach to task aversiveness across stages of personal projects. *Personality and Individual Differences*. 28: 153–167. https://doi.org/10.1016/S0191-8869(99)00091-4

Bonanomi, A., Olivari, M. G., Mascheroni, E., Gatti, E., and Confalonieri, E. (2018). Using a multidimensional Rasch analysis to evaluate the psychometric properties of the Motivated Strategies for Learning Questionnaire (MSLQ) among high school students. *Testing, Psychometrics, Methodology in Applied Psychology*. 25: 83-100. https://doi.org/10.4473/TPM25.1.5

Brando-Garrido, C., Montes-Hidalgo, J., Limonero, J. T., Gómez-Romero, M. J., andTomás-Sábado, J. (2020). Academic procrastination in nursing students. Spanish adaptation of the Academic Procrastination Scale-Short Form (APS-SF). *Enfermería Clínica (English Edition)*. 30: 371-376.

Burka, J. B., and Yuen, L. M. (1983). *Procrastination*. Reading, Addison-Wesley, Boston.

Byrne, B. M. (2010). *Structural equation modeling with AMOS. Basic concepts, applications, and programming (2nd ed.)*. Routledge Academic, London.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., and Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*. 100: 525-534. http://dx.doi.org/10.1037/0022-0663.100.3.525.

Chalmers, R.P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*. 48: 1-29.

Chalmers, R. P., Counsell, A., & Flora, D. B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. *Educational and Psychological Measurement*. 76: 114-140.

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

Chen, W.H. and Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*. 22: 265-289.

Choi, S. W., Gibbons, L. E., and Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of statistical software*. 39: 1-30.

Clark, L. A., and Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*. 31: 1412.

De Paola, M., and Scoppa, V. (2015). Procrastination, academic success and the effectiveness of a remedial program. *Journal of Economic Behavior & Organization*. *115: 217-236.*

Edelen, M.O., and Reeve, B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of life research*. 16: 5-18.

Ellis, A., and Knaus, W. J. (1977). Overcoming procrastination (New York, Institute for Rational Living). *Scientific Research*. 1: 1-11.

Ferrari, J. R., Johnson, J. L., and McCown, W. G. (1995). *Procrastination and task avoidance: Theory, research, and treatment*. Plenum Press, New York.

Ferrari, J. R., O'Callaghan, J., and Newbegin, I. (2005). Prevalence of procrastination in the United States, United Kingdom, and Australia: Arousal and avoidance delays among adults. *North American Journal of Psychology*. 7: 1-6.

Flake, J. K., Pek, J., and Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*. 8: 370-378.

Interview study. *Frontline Learning Research*. 3: 47–62. https://doi.org/10.14786/flr.v3i2.154

Kang, T., and Chen, T. T. (2008). Performance of the generalized S-X^2 item fit index for polytomous IRT models. *Journal of Educational Measurement*. 45: 391-406.

Kim, K. R., and Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*. 82: 26-33. doi:10.1016/j.paid.2015.02.038.

Krause, K., and Fruend, A. M. (2014). Delay or procrastination – A comparison of self-report and behavioral measures of procrastination and their impact on effective well-being. *Personality and Individual Differences*. 63: 75–80. doi:10.1016/j.paid.2014.01.050

Licata, F., Citrino, E. A., Maruca, R., Di Gennaro, G., and Bianco, A. (2024). Procrastination and risky health behaviors: a possible way to nurture health promotion among young adults in Italy. *Frontiers in Public Health*. 12: 1432763.

Limone, P., Sinatra, M., Ceglie, F., and Monacis, L. (2020). Examining Procrastination among University Students through the Lens of the Self-Regulated Learning Model. *Behavioral Sciences*. 10: 184. https://doi.org/10.3390/bs10120184

Lindblom-Ylänne, S. A., Saariaho-Räsänen, E. J., Inkinen, M. S., Haarala-Muhonen, A. E., and Hailikari, T. K. (2015). Academic procrastinators, strategic delayers and something betwixt and between: An interview study. *Frontline Learning Research*. 3: 47-62.

Lonergan, J. M., and Maher, K. J. (2000). The relationship between job characteristics and workplace procrastination as moderated by locus of control. *Journal of Social Behavior and Personality*. 15: 213-224.

Martín-Puga, M. E., Pelegrina, S., Gómez-Pérez, M. M., & Justicia-Galiano, M. J. (2022). Psychometric properties and measurement invariance of the academic procrastination scale-short form in Spanish children and adolescents. *Journal of Psychoeducational Assessment*. 40: 880-894.

Medsker, G. J., Williams, L. J., and Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*. 20: 439-464. https://doi.org/10.1177/014920639402000207

Milgram, N. A., Sroloff, B., and Rosenbaum, M. (1988). The procrastination of everyday life. *Journal of Research in Personality*. 22: 197-212. https://doi.org/10.1016/0092-6566(88)90015-3

Milgram, N., Mey-Tal, G., and Levison, Y. (1998). Procrastination, generalized or specific, in college students and their parents. *Personality and Individual Differences*. 25: 297-316. https://doi.org/10.1016/S0191-8869(98)00044-0

Mokken, R.J. (1971). *Theory and procedure of scale analysis: with applications in political research*. The Hague, Mouton.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*. 14: 59-71.

Onwuegbuzie, A. J., and Jiao, Q. G. (2000). I'll go to the library later: The relationship between academic procrastination and library anxiety. *College & Research Libraries*. 61: 45-54. https://doi.org/10.5860/crl.61.1.45.

Pintrich, P. R., and De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*. 82: 33-40. https://doi.org/10.1037/0022-0663.82.1.33

Rabin, L. A., Fogel, J., and Nutter-Upham, K. E. (2011). Academic procrastination in college students: The role of self-reported executive function. *Journal of Clinical and Experimental Neuropsychology*. 33: 344–357. https://doi.org/10.1080/13803395.2010.518597.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*. 48: 1-36.

Sabini, J., and Silver, M. (1982). *Moralities of everyday life*. Oxford University Press, Oxford.

Schouwenburg, H. C. (1995). Academic procrastination: Theoretical notions, measurement, and research. In J. R. Ferrari, J. L. Johnson, and W. G. McCown (Eds.), *Procrastination and task avoidance: Theory, research, and treatment*. Plenum Press, New York: 71-96.

Schraw, G., Wadkins, T., and Olafson, L. (2007). Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational Psychology*. 99: 12-25. https://doi.org/10.1037/0022- 0663.99.1.12.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*. 6: 461-464.

Simpson, W. K., and Pychyl, T. A. (2009). In search of the arousal procrastinator: Investigating the relation between procrastination, arousal-based personality traits and beliefs about procrastination motivations. *Personality and Individual Differences*. 47: 906-911. https://doi.org/10.1016/j.paid.2009.07.013.

Solomon, L. J., and Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*. 31: 503-509. https://doi.org/10.1037/0022-0167.31.4.503

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*. 133: 65-94. https://doi.org/10.1037/0033-2909.133.1.65.

Steel, P., Brothen, T., and Wambach, C. (2001). Procrastination and personality, performance, and mood. *Personality and individual differences*. 30: 95-106. https://doi.org/10.1016/S0191-8869(00)00013-1

Tuckman, B. W. (1991). The development and concurrent validity of the procrastination scale. *Educational and Psychological Measurement*. 51: 473-480. https://doi.org/10.1177/0013164491512022

Tutz, G. (2020). A taxonomy of polytomous item response models. arXiv preprint arXiv:2010.01382.

Van der Ark, L.A. (2007). Mokken scale analysis in R. *Journal of statistical software*. 20: 1-19.

Van der Ark, L.A. (2001). Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement*. 25: 273-282.

Van Eerde, W. (2000). Procrastination: Self-regulation in initiating aversive goals. *Applied Psychology: An International Review*. 49: 372-389. https://doi.org/10.1111/1464-0597.00021

Van Eerde, W. (2003). A meta-analytically derived nomological network of procrastination. *Personality and individual differences*. 35: 1401-1418. https://doi.org/10.1016/S0191-8869(02)00358-6

Van der Linden, W. J. (Ed.). (2018). *Handbook of item response theory: Three volume set*. CRC Press.

Van de Vijver, F., and Hambleton, R. K. (1996). Translating tests. *European Psychologist*. 1: 89-99. https://doi.org/10.1027/1016-9040.1.2.89

Yockey, R. D. (2016). Validation of the short form of the academic procrastination scale. *Psychological Reports*. 118: 171-179. https://doi.org/10.1177/0033294115626825