Learning and Individual Differences xxx (2016) xxx-xxx

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com

Assessing perfectionism in children and adolescents: Psychometric properties of the Almost Perfect Scale Revised

Sylvia Sastre-Riba,* Alicia Pérez de Albéniz, Eduardo Fonseca-Pedrero

Department of Educational Sciences, University of La Rioja, Spain

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 18 March 2016
Accepted 24 June 2016
Available online xxx

Keywords:
Perfectionism
Assessment
Educational
Psychometric properties
APS-R
Validity

ABSTRACT

The main goal of the present study was to study perfectionism through the psychometric properties of the Almost Perfect Scale-Revised (APS-R) in a representative sample of children and adolescents. The sample encompassed $n = 1476$ students from 9 to 16 years-old ($M = 12.29$ years; $SD = 2.17$). Analysis of the internal structure by means of exploratory factor analysis, yielded a three-dimensional solution (Discrepancy, Order, and Standards). Confirmatory factor analyses (CFAs) showed that the three-factor model displayed better goodness-of-fit indices than the competing models tested. Multigroup CFAs showed that the three-factor model had strong measurement invariance across gender and partial strong invariance across age. Significant statistical differences in the mean scores of the APS-R were found by gender and age. The level of internal consistency for the APS-R scores ranged from 0.81 to 0.89. The study of the psychometric properties of the APS-R scores supports the notion that it is a useful tool for the assessment of perfectionism in children and adolescents. The results have clear implications for the understanding of the expression of perfectionism and provide new sources of validity evidence for the APS-R in educational settings.

1. Introduction

Perfectionism is an important psychological construct. It is related to the concept of excellence and performance, which has been defined and measured by investigators in many ways, from a unidimensional focus towards a multidimensional one. The turning point in its conceptualization was set by Hamacheck's postulates (1978) based on a pioneering vision by Adler (1956). Perfectionism can be healthy whenever the pressure to achieve excellence includes a social interest to maximize one's own potential, and unhealthy if it involves strong neuroticism. These postulates changed their consideration as a unidimensional concept to include a distinction between a healthy perfectionism and an unhealthy or neurotic perfectionism (Neihart, Pfeiffer, & Cross, 2016; Sirois & Molnar, 2015).

This distinction is the one in force nowadays, and is considered as a multidimensional construct (Hewitt, Flett, Besser, Sherry, & McGee, 2003) that both researchers and professionals are trying to get to grips with; particularly when taking into consideration the many differences in the components that configure each type of perfectionism. These stances are influential in the development of many measuring tools aimed towards this goal. Its study is aimed both towards its relations and consequences in the configuration of the personality and as a cognitive function pattern which is related in particular with high intellectual ability as a potentiality in a place of privilege for a possible consecution of excellence (Pyryt, 2007).

One operative criterion that could be used to understand high intellectual ability is excellence. Authors such as Sternberg, Jarvin, and Grigorenko (2011) propose the point of view that high intellectual ability is composed by five criterions: a) Excellence, because of its higher intellectual ability; b) Rarity, because high intellectual ability is not common; c) Productivity, as the acquisitive performance or the numerous products obtained by the person during adulthood; d) Evidence of its existence, through an objective and multidimensional evaluation of high ability; and e) Worth, because the exceptional products obtained must be valued by society and other people. We would expect excellence to be present in high intellectual abilities because of its structural neurobiological potentiality. Excellence, however, is not always manifested, either with high intellectual ability (Subotnik, Olszewski-Kubilius, & Worrell, 2011) or typical intellectual ability. Perfectionism would be one of the variables that influence this.

From this conceptualization as a cognitive functioning pattern, perfectionism is related to motivation in school and other signs such as test anxiety or satisfaction and academic achievement (DiBartolo & Rendón, 2012; Eum & Rice, 2011; Fletcher & Neumeister, 2012). Thus, according to its performance, it could have a negative impact that could weaken their resolutive capacity, metacognitive regulation, and excellence relating it with motivation and academic performance, or anxiety before an evaluation (Kristie & Neumeister, 2012; Mobley, Slaney, & Rice, 2005; Rice, Richardson, & Tueller, 2014; Sastre-Riba, 2012). Moreover, perfectionism has been associated as a risk factor for mental disorders and symptoms (e.g., depression, eating disorders) (DiBartolo & Rendón, 2012; Flett & Hewitt, 2002; Rice et al., 2014) as well as psychological well-being (DiBartolo & Rendón, 2012). Taking that into account, and given its consequences,

* Corresponding author at: Department of Educational Sciences, University of La Rioja, C/ Luis de Ulloa, s/n (Edificio Vives), 26005 Logroño, Spain.

Email address: silvia.sastre@unirioja.es (S. Sastre-Riba)

http://dx.doi.org/10.1016/j.lindif.2016.06.022
1041-6080/© 2016 Published by Elsevier Ltd.

the investigation tries to discern which composition would result in its optimal contribution as a force in positive achievement and well-being.

Currently, authors do agree on the existence of many traits that lead to a perfectionist behaviour (Frost, Marten, Lahart, & Rosenblate, 1990; Hewitt et al., 2003; Stairs, Smith, Zapolski, Combs, & Settles, 2012). For instance, high personal standards (Frost et al., 1990), auto-oriented perfectionism (Hewitt et al., 2003), fear to err (Frost et al., 1990), or discrepancy (Slaney, Rice, Mobley, Trippi, & Ashby, 2001) between what one expects to achieve and the real achievement (Flett & Hewitt, 2002), and up to nine components Empirical support to its multidimensional composition starts to converge from two stances that, trying to grasp it, ended up building the first instruments for its measurement.

On the one hand, the Multidimensional Perfectionism Scale (MPS) (Hewitt & Flett, 1991), with 45 items organized in three subscales: 1) auto-oriented perfectionism, referred to the personal tendency of high-standard achieving, a strict evaluation of behaviour and motivation to strive for perfection; 2) perfectionism oriented towards others, that is, towards the expectation to achieve high standards by evaluating them strictly; and 3) socially-prescribed perfectionism, directed by the perception other people have of one's own standards, waiting to achieve excellence and through a strict evaluation.

On the other hand, the Frost Multidimensional Perfectionism Scale (FMPS) (Frost et al., 1990) is made up of 35 items and grouped in six subscales, and its psychometric properties are well established (e.g., Gelabert et al., 2011). Its authors remark on the importance of the existence of high standards followed by a demanding self-evaluation, as well as a special sensitivity towards parental criticism, with tendencies towards order and organization. These characteristics lead to the six subscales of the FMPS: 1) concern towards errors, as a tendency to take them as failures; 2) high personal standards as self-efficacy measurement; 3) doubt before an action, as a tendency to evaluate the non-adequate result of a task; 4) parental expectations, as a personal perception that parents have high expectations that need to be met; 5) parental criticism, as an excessive critique subjective feeling on their behalf; and 6) organization, referring to the tendency to heighten and prefer order.

Given the different starting approaches, the question is knowing whether each perspective's components and the given measurement instruments are related to each other, with a motivation towards performance or not. Authors such as Shafran and Mansell (2001) have studied its covariation, proposing that the self-oriented Perfectionism (Hewitt & flett, 1991) seems similar to the “Personal standards” and “organization” FMPS subscales, with shows a good correlation with the first ones (0.61 and 0.62) but scarce with “organization” (0.26–0.29); on the other hand, the correlation is lesser with “concern towards errors” (0.38–0.53), and scarce with: “doubt before an action”, “parental expectations” and “parental criticism” (0.16–0.27): Socially-prescribed perfectionism seems similar and with a correlation, with FMPS' “parental expectations” and “parental criticism” (0.49–0.57) but also with “concern towards errors” (0.49–59), and low with “doubt before an action” (0.28–0.37) and “personal standards” (0.16–28). Finally, the perfectionism oriented to others does not seem to be conceptually related to any of the FMPS' subscales, even when having a moderate correlation with “concern with errors” and “personal standards”, and low correlation with “parental expectations” and “organization”. In this manner, the investigation begins to establish some type of relationship between the possible components of perfectionism and the results from the created measuring instruments.

More recently, Slaney et al. (2001) revised the Almost Perfect Scale (APS-R) with a similar conceptual and measurement goal. It is formed of three subscales: a) Order, referring to the tendency to prefer one's own work; b) High standards, referring to the tendency towards high self-achievement; and c) Discrepancy, referring to the subjective perspective of the non-accomplishment of personal goals and objectives (Slaney, Mobley, Trippi, Ashby, & Johnson, 1996; Slaney et al., 2001). The APS-R scores have shown strong psychometric properties in previous studies and has been used in a range of samples and researches (Rice et al., 2014; Slaney et al., 1996, 2001; Stoeber & Otto, 2006).

For instance, the internal consistency values ranges between 0.91/0.92 for the Discrepancy subscale, 0.85 for the Standard subscale, and between 0.82/0.86 for the Order subscale (Slaney et al., 2001). The three-factor structure of the APS-R (Standards, Order and discrepancy) has been supported in several exploratory and confirmatory factor analyses (CFAs) (Mobley et al., 2005; Slaney et al., 2001; Suddarth & Slaney, 2001; Ulu, Tezer, & Slaney, 2012; Vandiver & Worrell, 2002; Wang, Yuen, & Slaney, 2009). Moreover, the APS-R showed factorial equivalence across gender (Rice et al., 2014) and cultural groups (Mobley et al., 2005). While true that this tool has shown adequate psychometric properties in previous research, it is beneficial and interesting to conduct new studies in different samples and settings (e.g., schools), for instance, children and adolescents from representative samples of the general population.

The main purpose of the present study was to study the construct of perfectionism, through the APS-R (Slaney et al., 2001), testing its psychometric properties in a large sample of children and adolescents. From this general goal four specific objectives have been formulated to: a) analyze the internal structure of the APS-R scores using exploratory and CFAs; b) test the measurement invariance of the APS-R scores across gender and age; c) examine the reliability of the APS-R scores through McDonald's Omega (McDonald, 1999) as well as the information functions from Item Response Theory (IRT) framework (Hambleton, Swaminathan, & Rogers, 1991); and d) compare APS-R mean scores by gender and age. Based on previous research, it is hypothesized that sound reliability will be established, and that the proposed three-factor dimensional (Order, Standards, and Discrepancy) model will be supported for this measure. In addition, we hypothesized that the three-factor model would be equivalent across gender and age. Moreover, differences in the means scores of the APS-R according gender or/and age will be found.

2. Method

2.1. Participants

Pupils were selected from different types of secondary schools – public, grant-assisted private, and private – and from vocational/technical schools of La Rioja (a region situated in the north of Spain). The sample comprised a total of 1476 students, of which 740 were male (50.1%) and 736 were female (49.9), belonging to eight schools and 20 classrooms. The age of the participants ranged from 9 to 16 years old (M = 12.29 years old; SD = 2.17). The age distribution of the sample was the following: 9 years (n = 195; 13.2%), 10 years (n = 195; 13.2%), 11 years (n = 193; 13.1%), 12 years (n = 189; 12.8%), 13 years (n = 191; 12.9%), 14 years (n = 216; 14.6%), 15 years (n = 210; 14.2%), and 16 years (n = 87; 5.9%). With the aim of conducting pertinent statistical analyses, a cross-validation study was performed where the total sample was randomly split into two subsamples. The first sub-sample consisted of 738 participants (374 male and 364 female), with a mean age of 12.24 (SD = 2.13). The second sub-sample consisted of 738 participants (366 male and 372 female), mean age of 12.34 ($SD = 2.21$). Neither gender ($\chi^2 = 0.173$; $p = 0.677$) nor age rates ($t = -0.900$; $p = 0.368$) differed across subsamples.

2.2. Instrument

The Almost Perfect Scale-Revised (APS-R) (Slaney et al., 1996, 2001). The APS-R was developed to assess the adaptive and the maladaptive components of perfectionism. It consists of 23 items and three subscales: a) High Standards (7 items) subscale measures the high personal standards one sets for oneself (e.g., *I expect the best from myself*); b) Discrepancy subscale (12 items) assesses respondents' perceived inadequacy in meeting personal standards (e.g., *I am never satisfied with my accomplishments*); and c) Order (4 items), refers to one's preference for neatness and orderliness (e.g., *I am an orderly person*). Participants responded to each item using a seven-point Likert scale ranging from 1 (I strongly disagree) to 7 (I strongly agree). The Spanish adaptation of the APS-R was made in accordance with the international guidelines for test translation and adaptation (Muñiz, Elosua, & Hambleton, 2013).

2.3. Procedure

The measurement instrument was administered collectively, in groups of 10 to 35 students, during normal school hours and in a classroom specially prepared for this purpose. For participants under 18, parents were asked to provide written informed consent in order for their child to participate in the study. Participants were informed of the confidentiality of their responses and of the voluntary nature of the study. No incentive was provided for their participation. Administration took place under the supervision of researchers. The study was approved by the research and ethics committee at the University of La Rioja.

2.4. Data analyses

First, we calculated descriptive statistics of the APS-R items (mean, standard deviation, skewness, and kurtosis) for the overall sample.

Second, in order to analyze the internal structure of APS-R scores, we conducted a cross-validation study randomly dividing the total sample into two subsamples. In the first subsample, given that this study is the first validation in Spanish children and adolescents, exploratory factor analyses were performed using the Minimum Rank Factor Analysis with Promin rotation. The procedure employed for determining the number of dimensions was optimal implementation of Parallel analysis (Timmerman & Lorenzo-Seva, 2011). This procedure is an implementation of Parallel Analysis where it is computed based on the same type of correlation matrix (i.e., Pearson correlation).

In the second subsample, several confirmatory factor analyses (CFAs) were conducted. The parameters were obtained from the Muthen's quasi-likelihood estimator (Muthén & Muthén, 1998a,b). The following goodness-of-fit indices were used: Chi-square (χ^2), Confirmatory Factor Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) (and 90% Confidence Interval), and Standardized Root Mean Square Residual (SRMR). Hu and Bentler (1999) suggested that RMSEA should be 0.06 or less for a good model fit and CFI and TLI should be 0.95 or more, though any value over 0.90 tends to be considered acceptable. Furthermore, SRMR cut-off values close to 0.95 or 1.00 were suggested as adequate for models with dichotomous outcomes (Yu & Muthén, 2002).

Third, measurement invariance across gender and age was tested via multigroup comparisons using structural equation modeling within the framework of CFA. Basically, a hierarchical set of steps are followed when testing measurement invariance, typically starting with the determination of a well-fitting multigroup, baseline model and continuing with the establishment of successive equivalence constraints in the model parameters across groups (Byrne, 2008; Meredith, 1993). The baseline model is called the configural model, which is the first and least restrictive model specified and is important because it represents the baseline model against which all subsequent specified invariance models are compared. The configural model is established by specifying and testing the CFA model for each group, separately. Once the theoretical model has been validated in both groups, configural invariance is then examined requiring that the same pattern of fixed and freely estimated model parameters is equivalent across groups. Configural invariance is tested by assessing the model fit. The next step is to impose equality constraints on the factor loadings across the groups to test metric or weak invariance. If the model fit with the constrained parameters is significantly and practically worse than the baseline or configural model, then weak invariance is not supported. The final step is to impose constraints on the item intercepts and factor loadings to test strong or scalar invariance model across groups.

The analyzed models are nested in that the imposed constraints are progressively added. The fit of nested models may be assessed by comparing the respective chi-square fit statistic or goodness-of-fit indices between the model with additional constraints to the less restricted model. Due to the limitations of the $\Delta\chi^2$ regarding its sensitivity to sample size, Cheung and Rensvold (2002) have proposed a more practical criterion, the ΔCFI, to determine if nested models are practically equivalent. In this study, when ΔCFI is > 0.01 between two nested models, the more constrained model is rejected since the additional constraints have produced a practically worse fit. However, if the change in CFI is less than or equal to 0.01, it is considered that all specified equal constraints are tenable; and, therefore, we can continue with the next step in the analysis of measurement invariance. However, when this criterion is not met and some of the parameters (e.g., factorial loadings or intercepts) are not specified to be equal across groups, partial measurement invariance can be considered (Byrne, Shavelson, & Muthén, 1989).

Fourth, we examined the reliability of the APS-R scores, using McDonald's Omega (McDonald, 1999) for the overall sample. This index is better than Cronbach's alpha to compute the reliability of the scores (Dunn, Baguley, & Brunsden, 2014). Moreover, the information functions of the APS-R subscale scores were estimated. The information function is an extension of the precision of measurement (e.g., reliability) in Classical Test Theory, within the IRT framework. It allows for the estimation of the contribution of each item or dimension to the assessment of each level of the latent construct or theta (e.g., perfectionism). Theta scores are measured on an interval scale ($M = 0$; $S^2 = 1$). Test information functions are related to the measurement precision (or standard error of measurement) and show the degree of precision at different levels of theta or latent trait.

Fifth, the effect of gender and age on the APS-R subscales was analyzed. In order to do this, a Multivariate Analysis of Variance was conducted, taking the APS-R subscales as the dependent variables, and gender and age groups as the fixed factors. As an estimate of effect size, partial eta squared was employed. SPSS 15.0 (Statistical Package for the Social Sciences, 2006), Mplus 7.0

(Muthén & Muthén, 1998a,b), and FACTOR 9.2 (Lorenzo-Seva & Ferrando, 2013) were used for data analyses.

3. Results

3.1. Descriptive statistics for the APS-R items

Descriptive statistics for the APS-R items for the total sample are shown in Table 1.

3.2. Validity evidence based on internal structure of the APS-R scores: exploratory and confirmatory factor analysis

Exploratory factor analysis was conducted using the first subsample. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.88, and the Bartlett test of sphericity was 5163.6 ($p < 0.001$). The results suggested a three-factor solution as the most adequate and parsimonious. Table 2 shows the factor loadings for this factorial structure that explained of 45.22% of the total variance. The first factor grouped items related to *Discrepancy* (21.53% of explained variance). The second factor grouped items related to *Standards* (17.21% of explained variance). The third factor grouped items related to *Order* (6.49% of explained variance). As shown in Table 2, the item distribution was entirely homogeneous and some overlaps were found between factors on items 6 and item 12. Only factorial loading of the item 5 was lower than 0.30. The correlations between factors ranged from − 0.26 (FI-FII) to 0.33 (FII-FIII) ($p < 0.01$).

Several CFAs were conducted using the second subsample. Different hypothetical models were tested: a) one dimensional model; b) the two-factor model (Discrepancy and Order plus Standards); and c) the three-factor model (Discrepancy, Order, and Standards). CFAs showed that the three-factor model displayed better goodness of-fit indices than the other hypothetical models tested. However, as shown in Table 3, goodness-of-fit indices for the three-factor baseline model did not reach the cut-offs recommended. For this model, substantial modification indices were found, so the correlation between error terms was allowed for those items that have similar content. For this model correlation between errors for the following items were allowed: 11 with 6; 19 with 16; 9 with 16; 3 with 15; 20 with 19; and 9 with 19. The goodness-of-fit indices for the three-factor model with modifications did reach the cut-offs recommended in the psychometric literature (CFI ≥ 0.90 and RMSEA ≤ 0.05). The standardized factorial loadings for the three-factor model allowing correlation between the error terms for these items are shown in Table 4. The correlations between latent factors ranged from − 0.18 (FI-FII) to 0.60 (FII-FIII) ($p < 0.01$).

Table 1
Descriptive statistics of the Almost Perfect Scale Revised (APS-R) items for the overall sample.

Items	*M*	*SD*	Skewness	Kurtosis
1	5.16	1.38	− 0.70	0.19
2	5.19	1.58	− 0.76	− 0.22
3	4.06	1.88	− 0.03	− 1.12
4	5.65	1.46	− 1.06	0.61
5	5.64	1.73	− 1.32	0.84
6	4.36	2.08	− 0.25	− 1.24
7	5.87	1.40	− 1.23	0.90
8	5.75	1.31	− 1.14	1.09
9	3.78	1.75	0.14	− 0.91
10	5.50	1.52	− 0.95	0.31
11	4.02	2.01	− 0.05	− 1.22
12	4.30	1.95	− 0.15	− 1.11
13	2.79	1.86	0.80	− 0.54
14	6.10	1.23	− 1.60	2.48
15	4.58	1.84	− 0.42	− 0.83
16	3.66	1.95	0.24	− 1.14
17	3.19	2.03	0.51	− 1.05
18	6.04	1.25	− 1.42	1.63
19	3.59	1.80	0.22	− 0.94
20	3.23	1.87	0.46	− 0.89
21	3.70	1.94	0.20	− 1.15
22	4.57	1.89	− 0.38	− 0.87
23	4.34	1.99	− 0.21	− 1.16

Table 2
Exploratory factor analysis of the Almost Perfect Scale Revised items (first subsample).

Items	Factors		
	I	II	III
1		0.68	
2			0.73
3	0.45		
4			0.70
5	–	–	–
6	0.68	0.35	
7			0.68
8		0.44	
9	0.43		
10			0.70
11	0.73		
12	0.37	0.51	
13	0.67		
14		0.69	
15	0.59		
16	0.54		
17	0.69		
18		0.57	
19	0.62		
20	0.61		
21	0.61		
22		0.35	
23	0.48		

Note. Factor loadings under 0.30 have been omitted.

3.3. Testing measurement invariance of the APS-R scores across gender and age

Subsequently, measurement invariance of the APS-R across gender and age were analyzed. The goodness-of-fit indices for both male and female are shown in Table 3. The configural model in which no equality constraints were imposed, provided adequate fit to the data. As can be observed, when the equivalence of the factorial loadings and intercept values were incorporated, the difference in the ΔCFI between the configural and the strong invariance model did not exceed 0.01. Therefore, we concluded that the factorial structure of the APS-R was operating equivalently across the two gender groups.

Then, measurement invariance of the APS-R scores across age was tested. Prior to the analysis of measurement invariance across age, we tested whether the three-factor model of the APS-R showed a reasonably good fit to the data in each group separately. The goodness-of-fit indices for the sample of participants from 9 to 10 years of age, form 11 to 13 years of age, and from 14 to 16 years of age are shown in Table 3. The configural model in which no equality constraints were imposed, provided a modest fit to the data (close to the recommended cut-off points).

As can be observed, when the equivalence of the factorial loadings and intercept parameters were incorporated, the difference in the ΔCFI between models exceeds 0.01. Therefore, several intercept across age groups were relaxed (6, 12, 13, 18, 1, 5, 6, 7, 3, 8, 9, 20, 21, 4, 16, 2, 14, 10, and 22). Nineteen items were non-invariant

Table 3
Goodness-of-fit indices of the models tested in the confirmatory factor analysis and measurement invariance across gender and age (second subsample).

Model	χ^2	*df*	CFI	TLI	RMSEA (CI 90%)	SRMR	CFI
One factor	2685.87	230	0.459	0.405	0.120 (0.105–0.113)	0.129	
Two factor	1074.82	229	0.780	0.757	0.071 (0.067–0.075)	0.074	
Three factor	822.54	227	0.845	0.827	0.060 (0.055–0.064)	0.067	
Three factor with modifications[a]	607.06	221	0.900	0.885	0.049 (0.044–0.053)	0.063	
Measurement invariance: gender							
Male ($n = 366$)	415.43	221	0.895	0.880	0.049 (0.042–0.056)	0.068	
Female ($n = 372$)	415.43	221	0.906	0.892	0.049 (0.042–0.056)	0.070	
Configural invariance	832.57	442	0.900	0.886	0.049 (0.044–0.054)	0.069	
Metric invariance	865.45	465	0.899	0.889	0.048 (0.043–0.053)	0.071	− 0.01
Strong invariance	915.92	488	0.891	0.887	0.049 (0.044–0.054)	0.073	− 0.01
Measurement invariance: age							
9–10 years-old ($n = 200$)	317.49	221	0.855	0.835	0.047 (0.035–0.058)	0.080	
11–13 years-old ($n = 269$)	353.24	221	0.913	0.901	0.047 (0.038–0.056)	0.068	
14–16 years old ($n = 269$)	422.81	221	0.880	0.861	0.058 (0.050–0.067)	0.084	
Configural invariance	1091.78	663	0.888	0.872	0.051 (0.046–0.057)	0.078	
Metric invariance	1173.74	709	0.879	0.870	0.052 (0.046–0.057)	0.091	− 0.01
Strong invariance	1507.39	755	0.804	0.803	0.064 (0.059–0.068)	0.119	+ 0.01
Partial strong invariance	1187.30	720	0.878	0.871	0.051 (0.046–0.057)	0.092	− 0.01

Note. χ^2 = Chi square; *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Square Residual.

[a] For this model correlated error terms were allowed: 11 with 6; 19 with 16; 9 with 15; 3 with 15; 20 with 19; and 9 with 19.

Table 4
Standardized factorial loadings estimated for the three-factor model with modifications (second subsample).

	Factor		
Items	I	II	III
1		0.62	
2			0.73
3	0.41		
4			0.78
5		0.01	
6	0.44		
7			0.65
8		0.45	
9	0.45		
10			0.76
11	0.58		
12		0.36	
13	0.66		
14		0.68	
15	0.49		
16	0.64		
17	0.63		
18		0.71	
19	0.56		
20	0.68		
21	0.58		
22		0.26	
23	0.44		

Note. All standardized factorial loadings estimated were statistically significant ($p < 0.01$), except item 5.

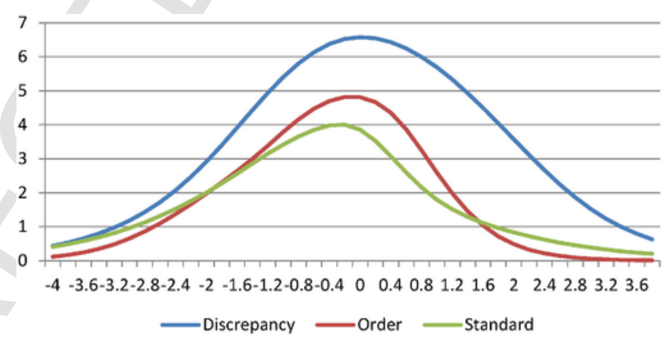

Fig. 1. Information functions for each subscale of the Almost Perfect Scale Revised (APS-R) in the overall sample.

across groups. After these parameters were freed, the ΔCFI between the constrained and the unconstrained model was under 0.01, indicating that partial strong measurement invariance by gender was supported. Hence, the results support strong measurement invariance of the APS-R scores by gender and partial strong invariance by age.

3.4. Estimation of reliability and information functions of APS-R scores

Reliability of the APS-R scores estimated by means of McDonald's Omega coefficient were 0.85 (Discrepancy), 0.67 (Standards), 0.82 (Order), and APS-S total score 0.73.

To further explore the measurement precision of the APS-R, the Information Function, was estimated for each dimension. All three information functions exhibit maximum information between − 1.2 and + 1.2 trait levels, showing the best measurement precision around the mean of the trait levels (see Fig. 1). The amount of information explained by the Discrepancy subscale was greater than that explained by the other two subscales. This mean shows that this subscale explains a greater amount of information (less standard error of measurement) for individuals with latent trait levels ranged between − 1.2 and + 1.2, compared with the others three subscales. The three subscales reduce their accuracy around the highest levels of the trait, especially in individuals with a latent trait level above − 2 and + 2.

3.5. Differences according to gender and age in the APS-R mean scores

The λ Wilks revealed statistically significant differences by gender ($\lambda = 0.988$, $F_{(3,1458)} = 6.050$, $p \leq 0.001$, partial $\eta^2 = 0.012$) and age ($\lambda = 0.746$, $F_{(21,4187.1)} = 21.380$, $p \leq 0.001$, partial $\eta^2 = 0.093$). By gender statistically significant differences were found in Order ($M_{\text{male}} = 21.8$ (5.02) $M_{\text{female}} = 22.6$ (4.60); $F_{(1,1460)} = 11.598$, $p \leq 0.001$, partial $\eta^2 = 0.008$) and Discrepancy ($M_{\text{male}} = 46.9$ (12.72) $M_{\text{female}} = 48.6$ (13.04); $F_{(1,1460)} = 6.772$, $p = 0.009$, partial $\eta^2 = 0.005$), but not in Standards ($M_{\text{male}} = 37.3$ (6.01) $M_{\text{female}} = 37.8$ (5.9); $F_{(1,1460)} = 2.725$, $p = 0.099$, partial $\eta^2 = 0.002$). The results showed that the females obtained higher mean scores than the males in both subscales. The effect-size estimates showed small effects in all subscales.

By age statistically significant differences were found in Standard ($F_{(7,1460)} = 49.681$, $p \leq 0.001$, partial $\eta^2 = 0.192$), Order ($F_{(7,1460)} = 35.572$, $p \leq 0.001$, partial $\eta^2 = 0.146$), and Discrepancy ($F_{(7,1460)} = 4.765$, ≤ 0.001, partial $\eta^2 = 0.022$). The effect-size estimates in Standards and Order subscales were large. Mean scores and standard deviation on the APS-R subscales for each age group are shown in Table 5. No statistically significant interactions gender × age were found.

4. Discussion

The main purpose was to analyze the perfectionism construct in a large sample of children and adolescents from the general population. To this end, the psychometric properties of the Almost Perfect Scale Revised (APS-R) (Slaney et al., 1996, 2001) was tested. We examined the internal structure of the APS-R scores through exploratory and CFAs; we tested the measurement invariance of the three-factor model across gender and age; we estimated the reliability of the scores; considered Omega coefficient and information functions from IRT framework, and compared the raw scores of the APS-R subscales by gender and age. The findings support the idea that the APS-R is a useful and brief tool for the assessment and screening of perfectionism traits during childhood and adolescence.

Table 5
Means scores according age group for the subscales of the Almost Perfect Scale Revised (APS-R) (overall sample).

APS-R	Age (years-old)	*M*	*SD*	Post-hoc comparison
Standard	9	41.59	5.08	9 > 11, 12, 13,14, 15, 16
	10	40.55	5.12	10 > 11, 12, 13,14, 15, 16
	11	39.39	5.39	11 < 9.10; 11 > 12, 13,14, 15, 16
	12	37.38	4.94	12 < 9.10.11; 12 > 14, 15, 16
	13	36.46	5.59	13 < 9.10.11; 13 > 14, 15, 16
	14	34.63	5.77	14 < 9, 10, 11, 12, 13
	15	34.67	5.69	15 < 9, 10, 11, 12, 13
	16	34.94	5.65	16 < 9, 10, 11, 12, 13
Order	9	24.84	3.58	9 > 11, 12, 13,14, 15, 16
	10	24.37	3.96	10 > 12, 13,14, 15, 16
	11	23.61	4.00	11 < 9; 11 > 12, 13,14, 15, 16
	12	22.10	4.92	12 < 9.10.11; 12 > 14, 15, 16
	13	21.54	4.69	13 < 9.10.11; 13 > 14, 15, 16
	14	20.24	4.68	14 < 9, 10, 11, 12, 13
	15	19.98	5.02	15 < 9, 10, 11, 12, 13
	16	20.26	5.04	16 < 9, 10, 11, 12, 13
Discrepancy	9	52.07	12.84	9 > 10, 11, 12, 13,14, 15, 16
	10	48.32	12.38	10 < 9; 10 > 11
	11	45.38	13.73	11 < 9, 10, 16
	12	46.39	13.28	12 < 9
	13	46.90	13.22	13 < 9
	14	47.50	12.46	14 < 9
	15	47.24	12.04	15 < 9
	16	48.70	11.82	16 < 9; 16 > 11

Analysis of the internal structure of the APS-R by means of exploratory factor analysis, yielded a three-dimensional solution composed by the factors: Discrepancy, Order, and Standards. CFAs showed that this three-factor model (with modifications) displayed better goodness-of-fit indices than the competing models tested. It is worth mentioning that optimal levels of goodness-of-fit indices were found after adding error correlation between items, indicating discrete values in the three-factor baseline model. Similar results were found in previous factorial studies (Mobley et al., 2005; Slaney et al., 2001; Suddarth & Slaney, 2001; Ulu et al., 2012; Vandiver & Worrell, 2002; Wang et al., 2009). For instance, Mobley et al. (2005) using CFA, found a three-factor solution of the APS-R (adding correlated errors) as the solution that better fits the data. In another study, Vandiver and Worrell (2002), in a sample of talented middle schools students, found the three dimensions of the APS-R in the exploratory factor analysis and acceptable goodness-of-fit indices (just below for the cut-off criteria) for the three-factor model in the CFA. These results would appear to indicate that the underlying dimensional structure of the APS-R scores is composed by three factors.

Results also supported the hypothesis of strong measurement invariance of the three-factor model of the APS-R by gender and partial measurement invariance by age. These results showed that nineteen items were non-equivalent across age groups, i.e., showed differential items functioning (DIF) by age. Another possibility to explain the presence of these non-invariant items is across age groups that differences are rooted in the complexity of the tested factor model, the psychometric properties of the tool, the method of assessment (e.g., self-report instruments) or sampling bias. The review of the literature shows that there are few studies of measurement invariance in the revised version of the APS-R. Recent studies have found measurement equivalence of the APS-R scores, across different demographic variables including gender (Rice et al., 2014) and cultural groups (Mobley et al., 2005) in samples of young adults. For instance, Mobley et al. (2005) showed that APS-R had partial measurement equivalence by cultural groups (e.g., African and American College students). In another study conducted by Rice et al. (2014) the measurement invariance across gender was supported. These findings suggest that the three-factor model of the APS-R seems to operate functionally in the same manner across the groups compared. In addition, the finding of measurement equivalence across gender and age provides new sources of validity evidence of the APS-R scores in this sector of the population. In this regard, the study of the measurement invariance across groups is relevant in order to assure the comparability of scores and for determining the generalizability of latent constructs across groups compared (Byrne, 2008, 2012; Meredith, 1993). Due the fact that partial measurement invariance of the APS-R by age was found, the mean comparison by age must be interpreted cautiously.

The levels of reliability of the APS-R scores were adequate, ranging from 0.67 to 0.85. In this sense, APS-R scores showed good reliability levels to estimate the three dimensions of perfectionism construct. These results are in line with the internal consistency values reported in previous studies using this measure (Mobley et al., 2005; Slaney et al., 1996, 2001). For instance, Mobley et al. (2005) found the following Cronbach's alpha: 0.75 (Standards), 0.81 (Order), and 0.88 (Discrepancy). However, in this study we have used Omega coefficient that has been shown by many researchers to be a more sensible index of internal consistency. The impact of poor measurement reliability can compromise a researcher's ability to make inferences of the results found (Dunn et al., 2014). When an IRT framework was used, the results showed that ASP-R scores provide more accurate information at the medium level of the each latent trait (e.g., perfectionism). This data is essential, because the IRT framework provides a modern approach to study the precision of perfectionism (and it facets) across each level of the latent construct. That is, the ASP-R provides greater accuracy of measurement of those individuals with medium-high levels of the latent construct. This point might be relevant in order to improve our accuracy in detecting individuals with healthy or unhealthy perfectionism.

Statistically significant differences by gender and age in mean scores on the APS-R were found. The results showed that the females obtained higher mean scores than the males in Order and Discrepancy subscales. By age, the results have shown a negative association between age and APS-R scores. There are few studies that analyze the effect of gender and age in representative samples of children and adolescents from the general population. In previous research, inconsistencies regarding the effect of gender on the APS-R scores have also been found (Ashby, Rice, & Kutchins, 2008; Rice & Ashby, 2007; Stoeber & Stoeber, 2009). For instance, Rice and Ashby (2007) found that woman scored higher than men on Order and High Standards dimensions of the APS-R; however, in another study, Ashby et al. (2008) did not find statistical differences in the mean scores of APS-R by gender. In this regard, it could be interesting to conduct new studies in order to improve our knowledge of the effect of gender and age on the perfectionism domains in this age group.

The study of the psychometric properties of the APS-R scores supports that it is a useful tool for assessing perfectionism in children and adolescents. The results have clear implications for the understanding of the phenotypic expression of perfectionism and provide new sources of validity evidence in the APS-R scores. However, the results of the present study should be interpreted in the light of the following limitations: first, one possible limitation of this study is that in spite of having a representative sample of children and adolescents, we focused on a particular Spanish region. Given the peculiarities, diversity and plurality of the nation, future studies should examine the psychometric properties of the APS-R in samples in other regions or countries. Second, in this study, information was gathered based solely on self-reports during childhood and adolescence, for which, we consider that it would have been interesting to complete this information with a clinical interview or with a hetero-report administered to the participants' parents. It would be interesting to add new instruments to test the relationship with other external variables (e.g., school achievement). Previous research have associated the perfectionism traits, amongst others, with motivation, psychological well-being, and academic performance (DiBartolo & Rendón, 2012; Elion, Wang, Slaney, & French, 2012; Rice et al., 2014; Wang et al., 2009).

Future studies could test the measurement invariance of the APS-R scores across cultures with look to analyzing the perfectionism in samples of children and adolescents with high intellectual ability or to conduct follow-up studies to test the predictive validity of this measure in both clinical and academic settings.

References

Adler, A., 1956. Striving for superiority. In: Ansbacherm, H.K., Ansbacher, R. (Eds.), The individual psychology of Alfred Adler: A systematic presentation in selections from his writings. Harper & Row, New York, NY, pp. 101–125.

Ashby, J.S., Rice, K.G., Kutchins, C.B., 2008. Matches and mismatches: Partners, perfectionism, and premarital adjustment. Journal of Counseling Psychology 55 (1), 125–132. http://dx.doi.org/10.1037/0022-0167.55.1.125.

Byrne, B., 2008. Testing for multigroup equivalence of a measuring instrument: A walk through the process. Psicothema 20, 872–882.

Byrne, B., 2012. Structural equation modeling with Mplus: Basic concepts, applications, and programming. Routledge Taylor & Francis Group, New York.

Byrne, B.M., Shavelson, R.J., Muthén, B., 1989. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin 105, 456–466.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233-255.DiBartolo, P.M., Rendón, M.J., 2012. A critical examination of the construct of perfectionism and its relationship to mental health in Asian and African Americans using a cross-cultural framework. Clinical Psychology Review 32, 139–152.

Dunn, T.J., Baguley, T., Brunsden, V., 2014. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology 105, 399–412.

Elion, A.A., Wang, K.T., Slaney, R.B., French, B.H., 2012. Perfectionism in African American students: Relationship to racial identity, GPA, self-esteem, and depression. Cultural Diversity and Ethnic Minority Psychology 18 (2), 118–127. http://dx.doi.org/10.1037/a0026491.

Eum, K., Rice, K.G., 2011. Test anxiety, perfectionism, goal orientation, and academic performance. Anxiety, Stress and Coping 23, 1–12.

Fletcher, K.L., Neumeister, S., 2012. Research on perfectionism and achievement motivation: Implication for gifted students. Psychology in the Schools 49, 668–676.

Flett, G.L., Hewitt, P.L., 2002. Perfectionism: Theory, research, and treatment. APA, Washington, DC.

Frost, R.O., Marten, P.A., Lahart, C., Rosenblate, R., 1990. The dimensions of perfectionism. Cognitive Therapy and Research 14, 449–468.

Gelabert, E., García-Esteve, L., Martín-Santos, R., Gutiérrez, F., Torres, A., Subirá, S., 2011. Psychometric properties of the Spanish version of the Frost Multidimensional Perfectionism Scale in women. Psicothema 23, 133–139.

Hamacheck, D.E., 1978. Psychodynamics of normal and neurotic perfectionsim. Psychology 15, 27–33.

Hambleton, R.K., Swaminathan, H., Rogers, H.J., 1991. Fundamentals of item response theory. Sage Publications, Inc., Beverly Hills, CA.

Hewitt, P.L., Flett, G.L., 1991. Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. Journal of Personality and Social Psychology 60, 456–470.

Hewitt, P.L., Flett, G.L., Besser, A., Sherry, S.B., McGee, B., 2003. Perfectionism is muldimensional: A reply to Shafran, Cooper and Fairbum (2002). Behaviour Research and Therapy 41, 1221–1236.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.Kristie, K.L., Neumeister, K.L.S., 2012. Research on perfectionism and achievement motivation: Implications for gifted students. Psychology in the Schools 497, 668–676.

Lorenzo-Seva, U., Ferrando, P.J., 2013. FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. Applied Psychological Measurement 37, 497–498.

McDonald, R.P., 1999. Test theory: A unified treatment. Lawrence Erlbaum, Mahwah, NJ.

Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58, 525–543.

Mobley, M., Slaney, R.B., Rice, K.G., 2005. Cultural validity of the almost perfect scale-revised for African American college students. Journal of Counseling Psychology 52 (4), 629–639. http://dx.doi.org/10.1037/0022-0167.52.4.629.

Muñiz, J., Elosua, P., Hambleton, R.K., 2013. Directrices para la traducción y adaptación de los tests: segunda edición [International Test Commission Guidelines for test translation and adaptation: Second edition]. Psicothema 25, 151–157.

Muthén, L.K., Muthén, B.O., 1998. Mplus user's guide, 5th ed., , Muthén & Muthén, Los Angeles, CA.

Muthén, L.K., Muthén, B.O., 1998. Mplus user's guide, 7th ed., , Muthén & Muthén, Los Angeles, CA.

Neihart, M., Pfeiffer, S., Cross, T., 2016. The social and emotional development of gifted children: What do we know?. Prufrock Press, Waco, TX.

Pyryt, M.C., 2007. The giftedness/perfectionism: Recent research and implications. Gifted Education International 23, 141–147.

Rice, K.G., Ashby, J.S., 2007. An efficient method for classifying perfectionists. Journal of Counseling Psychology 54 (1), 72–85. http://dx.doi.org/10.1037/0022-0167.54.1.72.

Rice, K.G., Richardson, C.M.E., Tueller, S., 2014. The short form of the revised almost perfect scale. Journal of Personality Assessment 96, 368–379.

Sastre-Riba, S., 2012. Alta Capacidad intelectual, perfeccionismo y regulación metacognitiva. Revista de Neurologia 54, 21–29.

Shafran, R., & Mansell, W. (2001). Perfectionism and psychopathology: A review of research and treatment. Clinical Psychology Review, 21, 879–906.Sirois, F.M., Molnar, D.S., 2015. Perfectionism, health, and well-being. Springer, New York.
Slaney, R. B., Mobley, M., Trippi, J., Ashby, J. S., & Johnson, D. G. (1996). The Almost Perfect Scale-Revised Unpublished manuscript, Pennsylvania State University at University Park.
Slaney, R.B., Rice, K.G., Mobley, M., Trippi, J., Ashby, J.S., 2001. The Revised Almost Perfect Scale. Measurement and Evaluation in Counseling and Development 34, 130–145.
Stairs, A.M., Smith, G.T., Zapolski, T.C., Combs, J.L., Settles, R.E., 2012. Clarifying the construct of perfectionism. Assessment 19, 146–166.
Statistical Package for the Social Sciences, 2006. SPSS Base 15.0 user's guide. SPSS Inc., Chicago, IL.
Sternberg, R.J., Jarvin, L., Grigorenko, E.L., 2011. Explorations in giftedness. Cambridge University Press, New York.
Stoeber, J., Otto, K., 2006. Positive conceptions of perfectionism: Approaches, evidence, challenges. Personality and Social Psychology Review 10 (4), 295–319.
Stoeber, J., Stoeber, F.S., 2009. Domains of perfectionism: Prevalence and relationships with perfectionism, gender, age, and satisfaction with life. Personality and Individual Differences 46, 530–535.
Subotnik, R.F., Olszewski-Kubilius, P., Worrell, F.C., 2011. Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. Psychological Science 12, 3–54.
Suddarth, B.H., Slaney, R.B., 2001. An investigation of the dimensions of perfectionism in college students. Measurement and Evaluation in Counseling and Development 34 (3), 157–165.
Timmerman, M.E., Lorenzo-Seva, U., 2011. Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods 16, 209–220.
Ulu, I.P., Tezer, E., Slaney, R.B., 2012. Investigation of adaptive and maladaptive perfectionism with Turkish Almost Perfect Scale-Revised. Psychological Reports 110, 1007–1020.
Vandiver, B.J., Worrell, F.C., 2002. The reliability and validity of scores on the Almost Perfect Scale-Revised with academically talented middle school students. The Journal of Secondary Gifted Education 13, 108–119.
Wang, K.T., Yuen, M., Slaney, R.B., 2009. Perfectionism, depression, loneliness, and life satisfaction: A study of high school students in Hong Kong. The Counseling Psychologist 37 (2), 249–274. http://dx.doi.org/10.1177/0011000008315975.
Yu, C.Y., Muthén, B.O., 2002. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes (technical report). UCLA, Graduate School of Education and Information Studies, Los Angeles.