Can Multi-Item Measures and Single-Item Measures be Trusted to Assess Self-Determination Theory Constructs in the Elderly?

Ariadna Angulo-Brunet, Carme Viladrich, Susana Pallarès, Marta Borrueco, Yago Ramis, Miquel Torregrossa, and Jaume Cruz
Universitat Autònoma de Barcelona

Abstract

Background: Measuring complex constructs, such as those from self-determination theory models, is challenging in the elderly due to the response process, particularly in collective data gathering. In order to examine this construct in physical activity settings we aimed at determining whether single-item measures were as good as pre-existing multi-item measures. For that reason, we developed seven single-item measures targeting perceived interpersonal styles, basic needs satisfaction, and well-being. **Method:** We gathered evidence of validity and reliability for multi-item measures and single-item measures based on a sample of 128 elderly adults. Another sample of 62 elderly adults provided test-retest reliability for the single-item measures. **Results:** Favourable evidence of the expected internal structure, concurrent validity, and reliability was obtained for competence- and relatedness-supportive styles, and for satisfaction of the basic psychological need of relatedness, fairly good evidence was obtained for subjective vitality, whereas mixed evidence was obtained for autonomy-supportive style and the satisfaction of the basic psychological needs of autonomy and competence. **Conclusions:** Single-item measures proved to be psychometrically sound substitutes for their multi-item counterparts, but the autonomy constructs need to be reconsidered. Furthermore the response process and consequences of testing should play a prominent role when devising questionnaires for the elderly.

Keywords: Single-item measure, self-determination theory, elderly, psychometrics, dichotomous items.

Resumen

¿Son fiables las Medidas Multi-Ítem y las Mono-Ítem Para Evaluar los Constructos de la Teoría de la Auto-Determinación en Personas Mayores?
Antecedentes: medir constructos complejos como los de la Teoría de la auto-determinación es un reto en personas mayores debido al proceso de respuesta. Para examinar estos constructos en actividad física nuestro objetivo fue comparar si las medidas mono-ítem eran igual de óptimas que las medidas multi-ítem pre-existentes. Por ello, desarrollamos siete medidas mono-ítem relacionadas con el estilo interpersonal, la satisfacción de las necesidades básicas y la vitalidad. **Método:** obtuvimos evidencias de validez y fiabilidad para las medidas multi-ítem y las medidas mono-ítem en una muestra de 128 personas mayores. Con otra muestra de 62 personas examinamos la fiabilidad test-retest para las medidas mono-ítem. **Resultados:** se obtuvo evidencia favorable relativa a la estructura interna, validez concurrente y fiabilidad para apoyo a la competencia y a la relación, y para satisfacción de la relación y vitalidad, mientras que para apoyo a la autonomía, satisfacción de la autonomía y de la competencia fue no conclusiva. **Conclusiones:** las medidas mono-ítem se han comportado como buenos sustitutos psicométricos para sus medidas multi-ítem homólogas, pero es necesario reconsiderar los constructos de autonomía, y además el proceso de respuesta y las consecuencias de la evaluación deben desempeñar un papel preeminente cuando se crean cuestionarios para personas mayores.

Palabras clave: medidas mono-ítem, teoría de la autodeterminación, personas mayores, psicometría, ítems dicotómicos.

Measurement of the conceptually complex constructs involved in self-determination theory in the elderly can be challenging due to the data collection process, especially in collective data gathering. The use of multi-item measures (MIMs) can lead to negative experiences like fatigue, boredom, apathy or frustration (Alcaraz et al., 2013; Robins et al., 2001) that can impact the validity of the response process. Another, less studied, interference with the response process could come from the maximization paradox:

the elderly seem to prefer satisfiying strategies, which make them happier, over maximizing strategies which require more effort (Bruine de Bruin et al., 2016). As these authors highlight, this is congruent with elderly people's preferences for fewer response options and presentation of less information. Considering this, it seems that using shortened measures could help to overcome these limitations. Nevertheless, this can compromise the construct coverage and the reliability of scores (Kruyen et al., 2013). However, the advantages of short measures are considered to outweigh their limitations in many psychological fields (e.g., clinical, health and organizational psychology), and sports psychology is no exception (Alcaraz et al., 2013).

In the field of sport and exercise psychology it is considered reasonable to develop a single-item measure (SIM) when the construct is concrete and well defined (Konstabel et al., 2017). For

example, SIMs have been comprehensively used for assessing the level of physical activity in the general population (e.g., Milton et al., 2013; Silsbury et al., 2015) and in the elderly population (Portegijs et al., 2016), particularly when physical activity is not the main focus of research (Gill et al., 2012). Nevertheless, there is a need in sport and exercise psychology for measuring complex psychological constructs. Taking one step in this direction, Kwon and Trail (2005) introduced a SIM to measure a highly complex subjective construct, sport loyalty, and advocated for SIMs on the grounds of the simplicity and brevity of their use. In the same vein, Bruton et al. (2016) developed a SIM to measure collective efficacy in sport teams.

Complex constructs are the raw material for self-determination theory, a theoretical framework that targets the social contextual conditions that boost self-determined and healthy psychological development (Ryan & Deci, 2000, 2017). According to self-determination theory, satisfaction of the basic psychological needs of competence (competence satisfaction), relatedness (relatedness satisfaction), and especially autonomy (autonomy satisfaction) predict vitality and can be promoted or thwarted by significant social agents such as exercise class leaders who exhibit autonomy-, competence- and relatedness-supportive or thwarting styles. Additionally, the satisfaction of the three basic psychological needs tend to be positively related, up to the point that it is difficult to find empirically separate factors, with an autonomy-supportive style that plays a preponderant role in fostering all of them. However, concerning the role of autonomy support in cross-cultural contexts, in some societies people engaging in physical activity expect controlling styles from their leaders and do not regard autonomy support as something desirable (Ntoumanis et al., 2017), a result also found when analysing other autonomy related constructs in self-determination theory (Yang et al., 2019). There is a scarcity of sound instruments to assess self-determination theory in the elderly. It is common practice to use MIMs developed for adults without further test of their psychometric properties other than Cronbach's alpha (e.g., Marcos et al., 2014; Gray et al., 2018) and to the best of our knowledge, no SIMs have been attempted.

MIMs psychometric properties have usually been tested on the basis of measurement models and internal consistency reliability coefficients, a method that is not suitable for SIMs. In a literature review published in 2013, Kruyen et al. warned about the limitations of the common practice of reporting Cronbach's alpha coefficient as a reliability indicator, neglecting other coefficients such as test-retest, and not obtaining specific validity evidence for short forms on the assumption that they have the same psychometric properties as the MIMs. More forcefully, the American Psychological Association (2010, 2020) emphasized that, besides reliability, validity evidence should also be provided for all psychological measures. Either way, it is important to describe in detail the psychometric properties when developing a new measure (Muñiz & Fonseca-Pedrero, 2019). Briefly stated, when using SIMs, various sources of reliability and validity of the scores should be considered in their own right, as in all psychological measurements (see American Educational Research Association et al., 1999, 2014). A detailed review and further discussion of how to obtain psychometric properties of SIMs can be found in the work by Angulo-Brunet (2019).

The aim of this study is to determine to what extent pre-existing MIMs developed for adult population and derived SIMs can be trusted as measures to study the self-determination theory constructs of supportive interpersonal style, need satisfaction and subjective vitality in the elderly. An effort is made to report a variety of pieces of evidence of their psychometric quality including the development of SIMs based on qualitative evidence, and gathering qualitative and quantitative reliability and validity evidence for the MIMs and the SIMs.

Method

Participants

In the item development process seven researchers with expertise in sport psychology and in self-determination theory participated in two judgemental tasks, and four elderly adults (two women; Age_{range} = 70 to 76) participated in a cognitive interview to evaluate item comprehension.

We used two non-probability samples of people aged over 60. To prevent presence of frailty or other related problems an inclusion criterion was that they were not living in nursing homes. Sample 1 was comprised of 128 elderly adults (77.6% women; Age: range = 62–85, M [SD] = 69.5 [4.1]) and Sample 2 was comprised of 62 elderly adults (87% women; Age: range = 61–88, M [SD] = 73.7 [7.4]) who were practising physical activity (e.g., water aerobics, yoga, maintenance gymnastics) in 11 centres, with 38 exercise class leaders, in sport clubs or sport public facilities of six different municipalities of three regions of Catalonia.

Instruments

All response scales were homogenised on a 5-point Likert scale ranging from 1 (*Totally disagree*) to 5 (*Totally agree*). Psychometric properties of all instruments are reported in the results section.

Supportive interpersonal style. We used the items of the three supportive dimensions — autonomy-supportive (e.g., "frequently asks about our preferences regarding activities to complete"), competence-supportive (e.g., "helps us learn and improve") and relatedness-supportive (e.g., "helps us to amicably resolve conflicts") — of the Coaches' Interpersonal Style Questionnaire (Pulido et al., 2017). Participants responded to four items per dimension under the stem "During sessions, our exercise class leader...".

Need satisfaction. We used the Spanish adaptation of the Basic Psychological Needs in Exercise Scale (Sánchez & Núñez, 2007; Vlachopoulos & Michailidou, 2006), which has three dimensions: autonomy satisfaction (e.g., "I feel that the way I exercise is definitely an expression of myself "), competence satisfaction (e.g., "I feel that exercise is an activity in which I do very well") and relatedness satisfaction (e.g., "I feel very much at ease with other exercise participants") with four items each under the stem "In physical activity sessions..."

Subjective vitality. We used five out of the six items from the Spanish adaptation of the Subjective Vitality Scale (Castillo et al., 2017; Ryan & Frederick, 1997; e.g., "I have energy and spirit"). We excluded the item "Sometimes I feel so alive I just want to burst" because it was considered potentially offensive to the elderly by expert criteria. The scale was preceded by the stem "I generally feel…"

Procedure

As part of a broader project, ethical approval was obtained by the authors' university ethics committee. We obtained informed

consent from all participants. In the item development process, the first author developed Version 1, considering MIMs dimensions and previous literature including other questionnaires. Version 2 and Version 3 were developed iteratively, asking to five experts to evaluate the sufficiency, relevance and clarity of the SIMs. If the proposed item was not satisfactory, they were asked to suggest an alternative wording. Furthermore, four cognitive interviews were conducted to evaluate comprehension on the part of the elderly participants and with their input we created Version 4. Finally, we asked to another panel of two specialists about the sufficiency, relevance and clarity of the final version.

On Sample 1 we scheduled individual face-to-face interviews, in a private space outside their physical activity schedule, with unlimited time. This administration was made by psychologists who encouraged the participants to do their best to overcome limitations due to reading and comprehension difficulties in the MIMs. For Sample 2, participants answered the SIMs (*test*) in a group setting. After four to six weeks, they answered the SIMs again in the same setting (*retest*).

Data analysis

Evidence based on item content was gathered during SIMs development (i.e., using items from pre-existing questionnaires and assessing the contents in two specialist panels). Validity evidence related to the response process and consequences of testing was first obtained through a cognitive interview in development phase, second based on feasibility of data collection and response time, and finally relying on participants' comments during data collection.

All quantitative analyses were conducted with the R software (R Core Team, 2018). Due to the presence of highly skewed item response distributions both in SIMs and MIMs we collapsed the 5-point Likert scale (see Results section for more details) to *Totally agree* (1) *versus other responses* (0) following the recommendations of Liu et al. (2017). For MIMs (Sample 1),

in line with Viladrich et al.' (2017) recommendations, we used tetrachoric correlations and the unweighted least squares mean and variance adjusted (ULSMV) estimator to test measurement models using confirmatory factor analysis (CFA) through the *lavaan* package (Rosseel, 2012). Comparative fit index (CFI) and Tucker-Lewis index (TLI) values greater than .95, and root mean square error of approximation (RMSEA) less than .06 were considered excellent fit (Hu & Bentler, 1999; Xia & Yang, 2018), while it was considered acceptable when CFI and TLI were greater than .90, and RMSEA was less than .08 (Marsh et al., 2004). We also provided two estimates of internal consistency reliability for MIMs: first, categorical omega (Green & Yang, 2009; Yang & Xia, 2018) obtained with the *reliability* function of the *semTools* package (Jorgensen et al., 2018) and second, Cronbach's alpha obtained with the function *alpha* from the *psych* package (Revelle, 2018). We considered reliability values greater than .70 as adequate in this phase (Nunnally, 1978).

Regarding SIMs reliability, we followed traditional approaches as summarized by Fuchs and Diamantopoulos (2009). First, we obtained three reliability coefficients based on the internal consistency reliability of their corresponding scale in MIM (categorical omega, Sample 1): using Spearman-Brown prophecy; solving for reliability in the correction for attenuation formula; and based on CFA, obtaining the item communality for each SIM that was selected from its original MIM (see below, for the two items that were entirely reworded we determined the item communalities by testing a one-factor CFA model, adding each SIM to its MIM counterpart). Additionally, in order to assess test-retest reliability (Sample 2), we considered the observed proportion of agreement and the prevalence-adjusted bias-adjusted kappa (PABAK; Byrt et al., 1993). In accordance with Landis and Koch (1977), we regarded a range of .21 to .40 as fair agreement, .41 to .60 as moderate agreement, .61 to .80 as substantial agreement, and .81 to 1.00 as almost perfect agreement.

Finally, evidence of validity related to external variables was obtained through the biserial correlation between each SIM and

		Table 1 Single-item development		
Constructs	**Version 1**	**Version 2/Version 3**	**Version 4**
Autonomy-supportive style	I think the exercise class leader values our preferences and opinions regarding the activities.	I think the exercise class leader values the opinions we express during the sessions.	I think the exercise class leader values the contributions (opinions, proposals, etc.) we make during the sessions.
Competence-supportive style	I think the exercise class leader proposes activities that match our level.	I think the exercise class leader gives us confidence in our ability to do the exercises well.	*no change*
Relatedness-supportive style	I think the exercise class leader fosters good relations among classmates and helps us to resolve conflicts in a friendly way.	I think the exercise class leader fosters a healthy climate and good relations among classmates and with him/her.	*no change*
Autonomy satisfaction	I feel that the activities I do are suited to my interests and the way I like to exercise.	I feel that the way of doing the exercises meets my needs. **	I feel I can do the exercises in the way I prefer.
Competence satisfaction	I think I have made great progress and that I perform the exercises effectively.	I feel I can comply with the requirements of the sessions.	I feel I can comply with the requirements of the activities.
Relatedness satisfaction	I feel very comfortable when I exercise with the others.	I feel very comfortable when I exercise with my physical-activity classmates.	I feel very comfortable when I exercise with my companions and my exercise class leader.
Vitality	I generally feel full of energy and enthusiasm.	I generally feel energetic. **	

Note: ** Items newly developed

the sum scores of its MIM counterpart without the item using the *biserial* function from the *psych* package (Revelle, 2018). Moreover, we assessed the tetrachoric correlations between SIMs in each of the three samples. We considered a value between .30 and .50 as moderate and greater than .50 as a high correlation (Cohen, 1988).

Results

Table 1 shows the development of the SIMs. As can be seen, five SIMs were derived from their MIM counterparts with minor changes whereas two SIMs were newly developed in order to measure the seven intended constructs. Changes introduced in Version 2 and Version 3 mainly modified items to adapt them better from a general sport context to that of physical activity in elderly population based on experts' suggestions. SIMs of vitality and autonomy satisfaction were entirely reworded in this phase. In turn, changes suggested by the elderly participants helped to simplify the language in Version 4. As an example of the changes depicted in Table 1, the SIM for relatedness satisfaction ended "... the others" in Version 1, "...my physical activity classmates" in Version 3, and "...my classmates and the exercise class leader" in the last version. The second panel of specialists considered that the final version of the seven developed items were relevant, sufficient and clear.

Individual hetero-administration of MIMs took between 20 and 50 minutes per person, while group data gathering with SIMs took between 5 and 20 minutes.

Table 2 shows descriptive statistics, reliability and validity estimates for all constructs. As depicted in the first column of the Single-item section, at least two out of three participants endorsed the response *Totally agree* in all SIMs. This extreme response pattern was also observed in 69% of the MIMs item distributions, a result reflected in the high mean value and/or standard deviation of the multi-item sum scores. In spite of the accurate development of SIMs, during the questionnaire administration some issues with the response process emerged with SIMs and MIMs. Specifically in hetero-administration, some participants reported having endorsed the *Totally agree* response category in almost all the statements because they really felt that their exercise class leader was

doing a good job, that their needs of autonomy, competence and relatedness were satisfied, and their subjective vitality was good considering their age. Other respondents reported doubts on how to answer the autonomy statements, as it was not clear for them whether their response was "good" or "bad" for their exercise class leader. Moreover, participants shared their worries that their results could impact on their exercise class leader's employment status, which raised an unintended consequence of the test administration. The ordinal nature of the scale was also reported as an issue: some respondents preferred to give a yes/no answer than expressing their level of agreement. Finally, most participants complained about the redundancy of MIMs, expressing the feeling that they were answering the same question over and over again.

Turning to quantitative analysis, generally speaking, all measurement models for the MIMs showed good fit and their derived total scores mainly showed acceptable internal consistency (see ρ_{NL} and α columns in Table 2). When seen in more detail, excellent goodness of fit indices were obtained for the theory-based three-factor model for the three supportive interpersonal styles (χ^2 [df] = 56.13[51], p = .29, CFI = .99, TLI = .99, RMSEA 90% CI = .03[.007]) but the high correlation coefficients observed among factors (range .81–.92) indicated lack of discriminant validity. Internal consistency coefficients were adequate in competence-supportive and in relatedness-supportive style scores while in autonomy-supportive scores they were unacceptable and not attributable to any specific item. Adequate goodness of fit indices were also obtained for the three-factor model for need satisfaction (χ^2 [df] = 77.49[51], p = .01, CFI = .93, TLI = .91, RMSEA 90% CI = .06[.03, .09]). Mirroring the supportive interpersonal style results, we found lack of discriminant validity between dimensions (range .78–.85) and non-adequate reliability coefficients for autonomy satisfaction. In fact, among the three basic needs, only relatedness satisfaction showed adequate reliability coefficients. Finally, the one-factor solution for subjective vitality showed excellent fit (χ^2 [df] = 4.56[5], p = .47, CFI = 1.00, TLI = 1.00, RMSEA 90% CI = .00[.00, .12]) and good internal consistency values.

Additionally, the two one-factor models tested to obtain item communalities for the entirely reworded items in Study 1 had adequate goodness of fit both for autonomy-supportive style (χ^2 [df] = 9.32[5], p = .09, CFI = .95, TLI = .91, RMSEA 90% CI =

Table 2
Quantitative evidence of reliability and validity

| | Multi-item sum scores | | | | | Single-item measures | | | | | | | | |
| | Sample 1 (n = 128) | | | | | | | | | | Sample 2 (n = 62) | | | |
Constructs	M	SD	ρ_{NL}	α	range	%	r_b	ρ^*_{xx}	r_{yy}	h^2	$\%_{test}$	$\%_{ret}$	ag	pa
Autonomy-supportive style	2.1	1.3	.57	.57	0-4	73	.66	.25	.77	.64	82	81	82	.65
Competence-supportive style	3.1	1.2	.75	.74	0-4	84	.86	.43	.98	.82	85	84	82	.65
Relatedness-supportive style	3.3	1.1	.73	.71	0-4	88	.61	.41	.51	.69	90	87	90	.80
Autonomy satisfaction	2.8	1.2	.65	.67	0-4	66	.37	.32	.21	.23	68	47	62	.23
Competence satisfaction	2.6	1.3	.64	.67	0-4	71	.58	.31	.52	.54	75	58	70	.39
Relatedness satisfaction	3.6	1.0	.74	.77	0-4	91	.76	.42	.77	.76	87	90	84	.67
Vitality	3.4	1.9	.88	.87	0-5	82	.77	.60	.67	.58	69	69	68	.35

Note: Sample 1: ρ_{NL} = categorical omega; range = potential range; % = Percentage of endorsement of *Totally agree* category; r_b= biserial correlation between full-form sum score without the item and single-item measure; ρ^*_{xx}= Spearman-Brown prophecy; r_{yy} = reliability derived from correction for attenuation formula; h^2 = communality. Sample 2: $\%_{test}$= Percentage of endorsement of *Totally agree* category in test; $\%_{ret}$= Percentage of endorsement of *Totally agree* category in retest; ag = percentage of agreement between test and retest; pa = PABAK coefficient

.08[.00, .17]) and for subjective vitality (χ^2 [*df*] = 6.06[9], *p* = .73, CFI = 1.00, TLI = 1.00, RMSEA 90% CI = .00[.00, .07]).

The three reliability estimates for SIMs derived from MIMs are presented in the rightmost three columns of the SIMs section of Sample 1 of Table 2. The scores for autonomy satisfaction showed the worst results in all estimates. The three estimates of the reliability coefficients for all other measures were acceptable in general, except for autonomy-supportive style and autonomy and competence satisfaction when using the Spearman-Brown prophecy. This is not surprising taking into account that the MIMs sum score reliability was below acceptable values for these constructs.

Concerning Sample 2, as seen in Table 2, prevalence of *Totally agree* responses was above 2 out of 3 respondents in both test and retest for all SIMs studied except for autonomy and competence satisfaction in retest. Both the percentage of agreement and PABAK showed substantial agreement between the responses from test and retest in all SIMs measuring interpersonal supportive style and relatedness satisfaction, whereas values of agreement for autonomy satisfaction and competence satisfaction and for subjective vitality were only fair.

Considering concurrent validity, biserial correlation between the SIMs and the MIMs (Table 2, column r_b) can be considered fairly good, except for autonomy satisfaction. Again, interpretation of the values obtained using the three constructs with non-adequate reliability values is limited. With respect to relationships between variables, Table 3 depicts the correlation coefficients between SIMs in Sample 1, Sample 2 test and Sample 2 retest. High correlations were found among supportive interpersonal style scores (range .50–.87) and low-high between basic psychological needs satisfaction scores (range .28–.78). Regarding the relationship between groups of variables, not conclusive evidence of relation was found between supportive interpersonal styles and basic psychological needs satisfaction scores, except for the strong correlations between competence- and relatedness-supportive styles with autonomy satisfaction in the three samples. Finally, we found moderated to high correlations between basic psychological needs satisfaction and vitality in all samples except for relatedness satisfaction, which was low in Sample 2 retest.

Discussion

Considering both the scarcity of sound instruments and the challenging collective data collection in the elderly, the aim of this research was to develop SIMs and provide psychometric evidence of MIMs and SIMs suitable to assess self-determination theory constructs in elderly population. The psychometric evidence obtained in this research does indeed suggest that SIMs are at least as valid and reliable as the MIMs in this age group. Moreover, SIMs data collection showed to be more feasible, and was preferred over MIMs, which adds fresh evidence to the preference of the elderly for satisficing rather than maximizing (Bruine de Bruin et al., 2015). For both reasons, SIMs should be prioritized in future research.

Psychometric quality of MIMs is debatable in this group. Although measurement models fit the data adequately, the main issue is the low discriminant validity with correlations between factors ranging between .78 and .85. These values are worse than those published in previous studies with the same questionnaires that ranged between .18 and .79 (Marcos et al., 2014; Pulido et al., 2017; Sánchez & Núñez, 2007; Vlachopoulos & Michailidou, 2006). However, all previous Spanish results were compatible with ours. In fact, similar issues were reported in Sanchez and Nuñez (2007) and Pulido et al. (2017) although these were addressed by allowing correlation between residuals or deleting items and Marcos et al. (2014) did not report the measurement model analysis. Furthermore, all these results, including ours, are consistent with the highly correlated factors previously found in the area of motivational climate (Wilson et al., 2009), and also with the theoretical expectations regarding the three need satisfaction scales being highly correlated among each other (Ryan & Deci, 2017).

There are also some limitations regarding internal consistency reliability that prevented us from further analysing the sum scores using correlational or path analysis. In most of the scales the internal consistency reliability coefficient is lower (values between .56 and .88) than that obtained in previous studies with the same questionnaire (values between .61 and .89; Pulido et al., 2017; Sánchez & Núñez, 2007; Vlachopoulos & Michailidou, 2006), even with elderly adults (values between .81 and .92; Marcos et al., 2014). One explanation would be that we analysed dichotomised data, while the previous studies analysed items with scales of five or more categories, and reliability is known to improve as the number of scale categories increases (e.g., Lozano et al., 2008; Rutkowski et al., 2019). Another compatible explanation is that, in the Basic Psychological Needs in Exercise Scale adaptation, Sánchez and

Table 3
Tetrachoric correlations between single-item measures

Concept	1	2	3	4	5	6
			Sample 1			
1. Autonomy-supportive style						
2. Competence-supportive style	.78**					
3. Relatedness-supportive style	.50**	.65**				
4. Autonomy satisfaction	.53**	.64**	.48**			
5. Competence satisfaction	.13	.28**	.38**	.33**		
6. Relatedness satisfaction	.45**	.46**	.83**	.45**	.62**	
7. Vitality	.38**	.16	.40**	.34**	.34**	78**
			Sample 2 test			
1. Autonomy-supportive style						
2. Competence-supportive style	.67**					
3. Relatedness-supportive style	.71**	.77**				
4. Autonomy satisfaction	.40**	.81**	.7**			
5. Competence satisfaction	.69**	.84**	.56**	.58**		
6. Relatedness satisfaction	.16	.53**	.64**	.78**	.49**	
7. Vitality	.26	.54**	.32**	.63**	.74**	.77**
			Sample 2 retest			
1. Autonomy-supportive style						
2. Competence-supportive style	.71**					
3. Relatedness-supportive style	.53**	.87**				
4. Autonomy satisfaction	.24	.64**	.56**			
5. Competence satisfaction	.02	.42**	.47**	.35*		
6. Relatedness satisfaction	–.12	.26	.35*	.28	.40*	
7. Vitality	.20	.48**	.64**	.43**	.53**	.04

* = *p*< .05. ** = *p*< .01

Núñez (2007) reported Cronbach's alpha coefficients that could be upwardly biased, since they were derived from a correlated errors measurement model (e.g., Viladrich et al., 2017).

Regarding test-retest reliability for SIMs, dichotomisation could again be responsible for lowering the observed values (Lozano et al., 2008; Rutkowski et al., 2019). However, in some cases, like the measurement of subjective vitality, there could be additional reasons, since test-retest instability had already been found in studies with the MIM (Castillo et al., 2017). Standing out in this context of low values are the high stability of the SIMs that refer to interpersonal style, which are in line with the participants' worries that responses questioning excellence could affect the employment status of the exercise class leader. Regarding evidence on validity in relation to external variables, the SIMs show high correlations with their MIM counterparts as expected.

All things considered, the items devised to measure autonomy are the weakest in psychometric terms, both in supportive style and in basic psychological need satisfaction, which is aligned with the doubts about what exactly autonomy statements were about expressed by the participants during data collection. Both results would support the notion that this construct may not be easily applicable to our culture (Ntoumanis et al., 2017) at least in the elderly population.

This study also presents some limitations. First, we worked with two convenience samples. To mitigate its impact, we collected data in three different regions of Catalonia in order to better represent the population of interest. Second, a further increase in sample size is not likely to help to overcome limitations in response variability in SIMs, as the main issues were due to ceiling effects and multicollinearity between variables. Finally, previous research with the elderly has shown that this age group is likely to show response biases (e.g., Vigil-Colet et al., 2015). Again, the low variability of responses prevented the use of statistical bias correction techniques (e.g., Ferrando et al., 2009). Rather, we are inclined to think that our results are attributable to the response process (Padilla & Benítez, 2014) and consequences of evaluation (Lane, 2014), given that, during the hetero-administration of the MIMs, it was clear that the elderly were answering in terms of a dichotomy (i.e., yes or no) regardless of the interviewer's efforts to encourage more detailed answers. Whatever the reasons given for that behaviour, the observed ceiling effects led us to collapse the response scale with consequences on all quantitative analyses.

In conclusion, the short answer to the question raised in the title is that SIMs for measuring self-determination theory constructs in the elderly are easier to answer and psychometrically as sound as their MIM counterparts and that both need further research regarding the constructs related to the concept of autonomy. For that reason, for collective data gathering we recommend using SIM measures and taking into account the preference for satisficing rather than maximizing that the elderly tend to exhibit. In this regard, we recommend considering using a simpler response scale (e.g., less response categories, analogue visual scale) or even choosing from a range of feelings or attributes rather than an agreement scale.

Acknowledgements

References

Alcaraz, S., Viladrich, C., & Torregrosa, M. (2013). Less time, better quality. shortening questionnaires to assess team environment and goal orientation. *The Spanish Journal of Psychology*, *16*, 1-14. https://doi.org/10.1017/sjp.2013.74

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Angulo-Brunet, A. (2019). *Creating synergies between psychometrics for the development and validation of single-item measures and the evaluation of the self-determination theory in the elderly* [Unpublished doctoral dissertation]. Universitat Autònoma de Barcelona.

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). American Psychological Association.

American Psychological Association (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

Bruine de Bruin, W., Parker, A. M., & Strough, J. (2016). Choosing to be happy? Age differences in "maximizing" decision strategies and experienced emotional well-being. *Psychology and Aging*, *31*(3), 295-300. https://doi.org/10.1037/pag0000073

Bruton, A. M., Mellalieu, S. D., & Shearer, D. A. (2016). Validation of a single-item stem for collective efficacy measurement in sports teams. *International Journal of Sport and Exercise Psychology*, *14*(4), 383-401. https://doi.org/10.1080/1612197X.2015.1054853

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*(5), 423-429. https://doi.org/10.1016/0895-4356(93)90018-V

Castillo, I., Tomás, I., & Balaguer, I. (2017). The Spanish-Version of the Subjective vitality scale: Psychometric properties and evidence of validity. *The Spanish Journal of Psychology*, *20*, 1-26. https://doi.org/10.1017/sjp.2017.22

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling*, *16*(2), 364-381. https://doi.org/10.1080/10705510902751374

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195-211.

Gill, D. P., Jones, G. R., Zou, G., & Speechley, M. (2012). Using a single question to assess physical activity in older adults: A reliability and validity study. *BMC Medical Research Methodology*, *12*(1), 1-10. https://doi.org/10.1186/1471-2288-12-20

Gray, S., Wharf, J., & Rhodes, R. E. (2018). Understanding physical activity motivation and behaviour through Self-Determination and Servant Leadership theories in a feasibility study. *Journal of Aging and Physical Activity*, *26*(3), 419-429. https://doi.org/10.1123/japa2017-0066

Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika, 74(1)*, 155-167. https://doi.org/10.1007/s11336-008-9099-3

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.

Structural Equation Modeling: A Multidisciplinary Journal, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools: Useful tools for structural equation modeling (Version 0.5-0.)* [R package]. https://cran.r-project.org/web/packages/semTools

Konstabel, K., Lönnqvist, J. E., Leikas, S., Velázquez, R. G., Qin, H., Verkasalo, M., & Walkowitz, G. (2017). Measuring single constructs by single items: Constructing an even shorter version of the "Short Five" personality inventory. *PLoS ONE*, *12*(8), 1-19. https://doi.org/10.1371/journal.pone.0182714

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, *13*(3), 223-248. https://doi.org/10.1080/15305058.2012.703734

Kwon, H., & Trail, G. (2005). The feasibility of single item measures in sport loyalty research. *Sport Management Review*, *8*, 69-88. https://doi.org/10.1016/S1441-3523(05)70033-4

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, *26*(1), 127-135. https://doi.org/10.7334/psicothema2013.258

Liu, Y., Millsap, R. E., West, S. G., Tein, J., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods, 22*(3), 486-506. https://doi.org/10.1038/nrg3575

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*(2), 73-79. https://doi.org/10.1027/1614-2241.4.2.73

Marcos, P. J., Orquín, F. J., Pedreño, N. B., & Moreno-Murcia, J. A. (2014). Motivación autodeterminada en adultos mayores practicantes de ejercicio físico [Self-determination motivation in elderly practitioners of physical exercise]. *Cuadernos de Psicología del Deporte*, *14*(3), 149-155.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320-341.

Milton, K., Clemes, S., & Bull, F. (2013). Can a single question provide an accurate measure of physical activity? *British Journal of Sports Medicine*, *47*(1), 44-48. https://doi.org/10.1136/bjsports-2011-090899

Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Ten steps for test development]. *Psicothema, 31*(1), 7-16. https://doi.org/10.7334/psicothema2018.291

Ntoumanis, N., Quested, E., Reeve, J., & Cheon, S. H. (2017). Need supportive communication: Implications for motivation in sport exercise and physical activity. In B. Jackson, J. Dimmock & J. Compton (Eds.), *Persuasion and communication in sport, exercise, and physical activity* (pp. 155-169). Routledge.

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, *26*(1), 136-144. https://doi.org/10.7334/psicothema2013.259

Portegijs, E., Sipilä, S., Viljanen, A., Rantakokko, M., & Rantanen, T. (2016). Validity of a single question to assess habitual physical activity of community-dwelling older people. *Scandinavian Journal of Medicine and Science in Sports*, *27*(1), 1423-1430. https://doi.org/10.1111/sms.12782

Pulido, J. J., Sánchez-Oliva, D., Leo, F. M., Sánchez-Cano, J., & García-Calvo, T. (2017). Development and validation of Coaches' interpersonal style questionnaire. *Measurement in Physical Education and Exercise Science*, 1-13. https://doi.org/10.1080/1091367X.2017.1369982

R Core Team (2018*). R: A language and environment for statistical computing (Version 3.5.1)* [Computer software]. https://www.r-project.org/

Revelle, W. (2018). *psych: Procedures for personality and pscychological research (Version 1.8.4)* [R package]. CRAN. https://CRAN.R-project.org/package=psych

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151-161. https://doi.org/10.1177/0146167201272002

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Rutkowski, L., Svetina, D., & Liaw, Y.-L. (2019). Collapsing categorical variables and measurement invariance. *Structural Equation Modeling*, *25*(5), 790-802. https://doi.org/10.1080/10705511.2018.1547640

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R., & Deci, E. (2017). *Self-determination theory: Basic psychological needs in motivation, development*. The Guilford Press.

Ryan, R., & Frederick, C. (1997). On energy, personality and health: Subjective vitality as a dynamic reflection of well-being. *Journal of Personality*, *65*(3), 529-565. https://doi.org/10.1111/j.1467-6494.1997.tb00326.x

Sánchez, J. M., & Núñez, J. L. (2007). Análisis preliminar de las propiedades psicométricas de la versión española de la escala de necesidades psicológicas básicas en el ejercicio físico [Preliminary analysis of the psychometric properties of the Spanish version of the basic psychological needs]. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, *2*, 83-92.

Silsbury, Z., Goldsmith, R., & Rushton, A. (2015). Systematic review of the measurement properties of self-report physical activity questionnaires in healthy adult populations. *BMJ Open*, *5*(9), 1-10. https://doi.org/10.1136/bmjopen-2015-008430

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Annals of Psychology*, *33*(3), 755-782. https://doi.org/10.6018/analesps.33.3.268401

Vigil-Colet, A., Lorenzo-Seva, U., & Morales-Vives, F. (2015). The effects of ageing on self-reported aggression measures are partly explained by response bias. *Psicothema*, *27*(3), 209-215. https://doi.org/10.7334/psicothema2015.32

Vlachopoulos, S. P., & Michailidou, S. (2006). Development and initial validation of a measure of autonomy, competence, and relatedness in exercise: The Basic Psychological Needs in Exercise Scale. *Measurement in Physical Education and Exercise Science*, *10*(3), 179-201. https://doi.org/10.1207/s15327841mpee1003_4

Wilson, P., Paige, J., & Mack, D. (2009). The importance of interpersonal style in competitive sport: A Self- Determination Theory approach. In C. H. Chang (Ed.), *Handbook of Sports Psychology* (pp. 1-21). Nova Science Publishers.

Xia, Y., & Yang, Y. (2018). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*, 409-428. https://doi.org/10.3758/s13428-018-1055-2

Yang, M., Viladrich, C., & Cruz, J. (2019). Measuring motivation in physical education among Chinese and Spanish adolescents: Comparing the psychometric properties of Perceived Locus of Causality Scale. *Revista de Psicología del Deporte*, *28*(2), 87-96.

Yang, Y., & Xia, Y. (2019). Categorical omega with small sample sizes via bayesian estimation: An alternative to frequentist estimators. *Educational and Psychological Measurement*, *79*(1), 19-39. https://doi.org/10.1177/0013164417752008