



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

Creando sinergias entre la psicometría para el desarrollo y validación de medidas de un solo ítem y la evaluación de la teoría motivacional de la autodeterminación en personas mayores

Autora: Ariadna Angulo Brunet

Directora: Carme Viladrich

Tesis doctoral

Doctorado en Psicología de la Salud y del Deporte

Departamento de Psicología Básica, Evolutiva y de la Educación

Facultad de Psicología

Universitat Autònoma de Barcelona

Bellaterra, septiembre 2019



Doctorado en Psicología de la Salud y del Deporte
Departamento de Psicología Básica, Evolutiva y de la Educación

**Creando sinergias entre la psicometría para el desarrollo y
validación de medidas de un solo ítem y la evaluación de la teoría
motivacional de la autodeterminación en personas mayores**

Ariadna Angulo Brunet
Tesis doctoral dirigida por
Carme Viladrich Segués

AGRADECIMIENTOS

M'agradaria començar donant les gràcies a la meva directora, Carme Viladrich, per ser sempre allà. Per donar-me molt més del que es pot esperar d'una tutora. Pels bons moments, els ànims, i per fer de guia en els moments de desesperació, que mai m'ha deixat passar sola. Per voler fer amb mi aquest camí.

També vull agrair a tota la meva família el seu suport en aquest procés. Especialment a la mare i el pare, i a la iaia Antonia i l'avi per la paciència i els ànims en els darrers anys. A la Xancla, la meva gata, per ser gairebé co-autora d'aquesta tesis i acompanyar-me incondicionalment. I a la Cara, per donar-me escalfor en els moments de tranquil·litat.

Continuar agraint als amics que no han sortit corrents durant aquesta tesis i als que ara formen part de la meva vida. Gràcies Xevi Cayuela per poder comprendre la importància que té aquest projecte i saber trobar les paraules en cada moment. Gràcies Roger per deixar-me compartir amb tu aquesta part de la meva vida. Gràcies Núria, Clàudia, Cristina i Eduardo per donar-me ànims quan creia que sortir corrents era una alternativa. Gràcies Joel i Jorge per escoltar-me durant l'esprint final.

M'agradaria agrair també a les companyes i companys del GEPE per haver-me acollit a l'equip de recerca i donar-me les eines per afrontar un doctorat en psicologia de l'activitat física. Entre d'altres gràcies per convidar-me a participar en el projecte ACTIVATICS amb el

qual es desenvolupa aquest treball. Gràcies a totes les persones que van acceptar participar en la validació de les mesures d'un sol ítem: a les persones expertes; a les 128 persones que voluntàriament van contestar els qüestionaris; al Javier i la Marta per acompanyar-me en alguna part de les recollides de dades.

Gràcies a tots els ens que heu fet possible aquesta tesi i que heu aconseguit, d'una manera o d'una altra, acompanyar-me en aquesta travessada.

Per últim, agrair al *Ministerio de Economía y Competitividad* per poder fer possible econòmicament aquesta tesi mitjançant el projecte ACTIVATICS (DEP-2014-52481).

ABSTRACT

The evaluation of complex constructs, such as those contemplated in self-determination theory of motivation (SDT), through questionnaires is a field that has not been widely explored in the elderly. The feasibility of self-administering 65 items (24 of coaches' interpersonal style, 12 of basic psychological needs satisfaction, 24 of behavioral regulation and 5 of subjective vitality) necessary to evaluate the chain of relationships of the SDT is limited in this population. Using single-item measures could lead to greater feasibility and therefore to obtain more knowledge about the elderly.

In this thesis it is proposed to look for a substantive-methodological solution to this question. The first aim is to describe the chain of relationships proposed by the SDT in people over sixty who practice directed physical activity. The second is to obtain evidence of psychometric quality for questionnaires already used in other adult populations (i.e., Coaches' Interpersonal Style Questionnaire [CIS-Q], Basic Psychological Need Exercise Scale [BPNES], Behavioral Regulation in Exercise Questionnaire [BREQ-3] and Subjective Vitality Scale [SVS]). The third is to develop single-item measures to examine the same dimensions as in multi-item measures. Specifically, we want to study the feasibility and response process of single-item measures in comparison to multi-item measures, to examine the psychometric properties of single-item measures with the same sample as multi-item measures, and to study with a new sample the temporal stability.

The multi-item measures and the measures created were administered to a sample of 128 elderly people and the responses of other 62 people were used for the study of temporal stability. The results obtained with the two measures point that people over sixty years of age who practice directed physical activity perceive a great support and satisfaction of their basic psychological needs, practice autonomously motivated physical activity and in general feel vital. For

multi-item measures, favorable evidence of internal structure was found for the interpersonal style of support to basic psychological needs, for satisfaction of the basic psychological needs and for subjective vitality. From the study of internal consistency only favorable evidence was found for the interpersonal style of support to competence and support to relationships, for satisfaction of the basic psychological need of relationship and for subjective vitality. The 16 single-item measures developed were considered as relevant, sufficient and clear to measure the concepts of SDT in this population by the panel of specialists. Field research showed on one hand that the collection of data with single-item measures was more feasible than with multi-item measures. On the other hand, response process related to the desire not to harm the leader or the difficulty of differentiating between the concepts were observed. Additionally, favorable validity evidence has been provided based on the relationship with external variables for all single-item measures. Regarding reliability, the results have been less conclusive since only three measures (competence support, relationship support and relationship satisfaction) have been acceptable in all indicators.

RESUMEN

La evaluación de constructos complejos, como los contemplados en la teoría motivacional de la autodeterminación (SDT), mediante cuestionarios es un campo que no ha sido ampliamente explorado en personas mayores. La factibilidad de auto-administrarse los 65 ítems (24 de estilo interpersonal del monitor/a, 12 de satisfacción de las necesidades psicológicas básicas, 24 de regulación conductual y 5 de vitalidad subjetiva) necesarios para evaluar la cadena de relaciones de la SDT es limitada en esta población. Utilizar medidas de un solo ítem podría conducir a una mayor factibilidad y por lo tanto a obtener más conocimiento sobre las personas mayores.

En esta tesis se busca una solución sustantivo-metodológica a esta cuestión. El primer objetivo es describir la cadena de relaciones propuesta por la SDT en personas mayores de 60 años que practican actividad física. El segundo objetivo es obtener evidencia de calidad psicométrica para cuestionarios ya utilizados en otras poblaciones de adultos (i.e., Coaches' Interpersonal Style Questionnaire [CIS-Q], Basic Psychological Need Exercise Scale [BPNES], Behavioral Regulation in Exercise Questionnaire [BREQ-3] y Subjective Vitality Scale [SVS]). El tercero es desarrollar medidas de un solo ítem para examinar las mismas dimensiones que en las medidas de múltiples ítems. Concretamente se quiere estudiar la factibilidad y el proceso de respuesta de las medidas de un solo ítem en comparación de las medidas de múltiples ítems, examinar las propiedades psicométricas de las medidas de un solo ítem con la misma muestra que las medidas de múltiples ítems y estudiar con una nueva muestra la estabilidad temporal.

Se ha contado con una muestra de 128 personas que han contestado las medidas de múltiples ítems y las medidas creadas, y también se contó con las respuestas de 62 personas para el estudio de la estabilidad temporal. Los resultados obtenidos con las dos medidas

apuntan a que las personas mayores de sesenta años que practican actividad física dirigida perciben un gran apoyo y satisfacción de sus necesidades psicológicas básicas, practican actividad física motivadas autónomamente y en general se sienten vitales. Para las medidas de múltiples ítems, se ha obtenido evidencia favorable de estructura interna para el estilo interpersonal de apoyo a las necesidades psicológicas básicas, para satisfacción a las necesidades psicológicas básicas y para vitalidad subjetiva. El estudio de la consistencia interna solo aportó evidencia favorable para los estilos interpersonales de apoyo a la competencia y apoyo a la relación, para la satisfacción de la necesidad psicológica básica de relación y para la vitalidad subjetiva. Las 16 medidas de un solo ítem desarrolladas fueron consideradas por el panel de especialistas como relevantes, suficientes y claras para medir los conceptos de la SDT en esta población. El estudio de campo evidenció por un lado que la recogida de datos con medidas de un solo ítem era más factible que con las medidas de múltiples ítems. Por otro lado, se observaron procesos de respuesta relacionados con la voluntad de no perjudicar al líder o bien con la dificultad de diferenciar entre los conceptos. Adicionalmente se ha aportado evidencia favorable de validez basada en la relación con variables externas para todas las medidas de un solo ítem. En referencia a la fiabilidad los resultados han sido menos concluyentes ya que solo tres medidas (apoyo a la competencia, apoyo a la relación y satisfacción de la relación) han resultado aceptables en todos los indicadores.

RESUM

L'avaluació de constructes complexes, com els contemplats a la teoria motivacional de l'autodeterminació (SDT), mitjançant qüestionaris és un camp que no ha sigut àmpliament explorat en persones grans. La factibilitat d'auto-administrar-se els 65 ítems (24 d'estil interpersonal del monitor/a, 12 de satisfacció de les necessitats psicològiques bàsiques, 24 de regulació conductual i 5 de vitalitat subjectiva) necessaris per avaluar la cadena de relacions de la SDT és limitada en aquesta població. Utilitzar mesures d'un sol ítem podria conduir a una major factibilitat i per tant a poder obtenir més coneixement sobre les persones grans.

En aquesta tesis es busca una solució substantiu-metodològica a aquesta qüestió. El primer objectiu és descriure la cadena de relacions proposada per la SDT en persones majors de 60 anys que practiquen activitat física. El segon es obtenir evidència de qualitat psicomètrica per a qüestionaris ja utilitzats en altres poblacions d'adults (i.e., Coaches' Interpersonal Style Questionnaire [CIS-Q], Basic Psychological Need Exercise Scale [BPNES], Behavioral Regulation in Exercise Questionnaire [BREQ-3] i Subjective Vitality Scale [SVS]). El tercer es desenvolupar mesures d'un sol ítem per a examinar les mateixes dimensions que en les mesures de múltiples ítems. Concretament es vol estudiar la factibilitat i el procés de resposta de les mesures d'un sol ítem en comparació a les mesures de múltiples ítems, examinar les propietats psicomètriques de les mesures de un sol ítem amb la mateixa mostra que les mesures de múltiples ítems i estudiar amb una nova mostra la estabilitat temporal.

S'ha comptat amb una mostra de 128 persones que han contestat les mesures de múltiples ítems i les mesures creades, i també s'ha comptat amb les respostes de 62 persones per a l'estudi de l'estabilitat temporal. Els resultats obtinguts amb les dues mesures apunten a que les persones majors de seixanta anys que practiquen

activitat física dirigida perceben un gran suport i satisfacció de les seves necessitats psicològiques bàsiques, practiquen activitat física motivades autònomament i en general es senten vitals. Per les mesures de múltiples ítems, s'ha obtingut evidència favorable d'estructura per a l'estil interpersonal de suport a les necessitats psicològiques bàsiques, per a satisfacció de les necessitats psicològiques bàsiques i per a vitalitat subjectiva. L'estudi de la consistència interna només ha aportat evidència favorable per als estils interpersonals de suport a la competència i suport a la relació, per a la satisfacció de la necessitat psicològica bàsica de relació i per la vitalitat subjectiva. Les 16 mesures d'un sol ítem desenvolupades van ser considerades pel panell d'especialistes com a rellevants, suficients i clares per a mesurar els conceptes de la SDT en aquesta població. L'estudi de camp va evidenciar per una banda que la recollida de dades amb mesures d'un sol ítem era més factible que amb mesures de múltiples ítems. Per altre banda, es van observar processos de resposta relacionats amb la voluntat de no perjudicar al líder o bé amb la dificultat de diferenciar entre conceptes. Addicionalment s'ha aportat evidència favorable de validesa basada en la relació amb variables externes per a totes les mesures d'un sol ítem. En referència a la fiabilitat els resultats han sigut menys concloents ja que només tres mesures (suport a la competència, suport a la relació i satisfacció de la relació) han resultat acceptables en tots els indicadors.

ÍNDICE DE CONTENIDOS

AGRADECIMIENTOS	I
ABSTRACT	III
RESUMEN	V
RESUM	VII
ÍNDICE DE CONTENIDOS	IX
ÍNDICE DE TABLAS	XI
ÍNDICE DE FIGURAS	XIII
ABREVIATURAS	XV
CONSIDERACIONES LINGÜÍSTICAS	XIX
1. EL PROBLEMA APLICADO Y LA SOLUCIÓN METODOLÓGICA	1
Los constructos de la teoría de la autodeterminación	5
Aplicación a la realización de actividad física en personas mayores	10
Las medidas de un solo ítem como una forma de reducción de cuestionarios	16
Evidencias de validez para medidas de un solo ítem	18
Evidencias de fiabilidad para medidas de un solo ítem	29
2. OBJETIVOS DE ESTA TESIS	35
3. MÉTODO	37
Participantes	37
Instrumentos	39
Procedimiento	43
Análisis de datos	47
4. RESULTADOS	53
Propiedades psicométricas de las medidas de múltiples ítems	53
Análisis preparatorio de datos	53
Evidencias basadas en la estructura interna del test y la fiabilidad	69
Propiedades psicométricas de las medidas de un solo ítem	76
Desarrollo de las medidas de un solo ítem y evidencias de validez basadas en su contenido	76
Análisis preparatorio de datos	82
Validez	86
Evidencias basadas en el proceso de respuesta a los ítems	86
Evidencias basadas en la relación con otras variables	89
Evidencias basadas en las consecuencias de la evaluación	95
Fiabilidad	96
5. DISCUSIÓN	101
6. LIMITACIONES Y FORTALEZAS	113
7. CONCLUSIONES	119
8. REFERENCIAS	127

APÉNDICES	II
Apéndice A. Material utilizado en la recogida de datos	IV
Apéndice A1. Consentimiento informado	IV
Muestra 1	IV
Muestra 2	VI
Apéndice A2. Medidas de un solo ítem	VIII
Apéndice A3. Medidas de múltiples ítems	IX
Estilo interpersonal	X
Satisfacción necesidades psicológicas básicas	XI
Regulación conductual	XII
Vitalidad subjetiva	XIII
Apéndice B. SubscaleExplorer: un paquete para la gestión y descripción de datos psicométricos	XV
Instalación y activación de SubscaleExplorer	XVI
El caso práctico	XXVIII
Gestión de datos: <i>getscale()</i>	XX
Descripción de escalas Likert	XXV
Funciones <i>boxPlot()</i> y <i>violinPlot()</i>	XXVI
Función <i>frequencies()</i>	XXVII
Función <i>freqbubble()</i>	XXIX
Función <i>missingness()</i>	XXX
Función <i>uniqueValues()</i>	XXXIII
Función <i>designEffect()</i>	XXXIV
Usabilidad del paquete	XXXV
Apéndice C Trabajos presentados en congresos nacionales e internacionales	XXXVII
Apéndice C1. SubscaleExplorer: Un paquete de R para la gestión y descripción de subescalas psicométricas	XXXVIII
Apéndice C2. ¿Limitaciones o cambio de objetivos? Lecciones aprendidas comparando dos opciones para reportar medidas con baja consistencia interna	XXXIX
Apéndice C3. How to handle ordered responses with floor and ceiling effects in SEM using small samples?	XL
Apéndice C4. Desarrollo de medidas de un ítem para evaluar los resultados de una intervención percibidos por personas mayores	XLI
Apéndice C5. Single-item measures validity evidence based on correlation network analysis: motives for physical activity in the elderly	XLII
Apéndice C6. Validity of single-item measures to assess self-determination theory constructs in the elderly	XLIII

ÍNDICE DE TABLAS

TABLA 1 DESCRIPCIÓN DE LAS MUESTRAS OBTENIDAS EN EL ESTUDIO	38
TABLA 2 BONDAD DE AJUSTE DE LAS MEDIDAS CON MÚLTIPLES ÍTEMS	71
TABLA 3 ESTADÍSTICOS DESCRIPTIVOS PARA LA PUNTUACIÓN DE LAS MEDIDAS CON MÚLTIPLES ÍTEMS.....	74
TABLA 4 DESARROLLO DE LAS MEDIDAS DE UN SOLO ÍTEM	78
TABLA 5 FIABILIDAD DE LAS MEDIDAS DE UN SOLO ÍTEM	97

ÍNDICE DE FIGURAS

FIGURA 1 CADENA DE RELACIONES PROPUESTA POR LA TEORÍA DE LA AUTODETERMINACIÓN	6
FIGURA 2 DISTRIBUCIÓN DE RESPUESTAS DE ESTILO INTERPERSONAL	54
FIGURA 3 CORRELOGRAMA DE ESTILO INTERPERSONAL	56
FIGURA 4 DISTRIBUCIÓN DE RESPUESTAS DE SATISFACCIÓN DE LAS NECESIDADES PSICOLÓGICAS BÁSICAS	59
FIGURA 5 CORRELOGRAMA DE SATISFACCIÓN DE LAS NECESIDADES PSICOLÓGICAS BÁSICAS	60
FIGURA 6 DISTRIBUCIÓN DE RESPUESTAS DE REGULACIÓN CONDUCTUAL	62
FIGURA 7 CORRELOGRAMA DE REGULACIÓN CONDUCTUAL.....	64
FIGURA 9 CORRELOGRAMA DE RESPUESTAS DE VITALIDAD SUBJETIVA.....	65
FIGURA 10 CORRELATIONAL NETWORK ANALYSIS	66
FIGURA 11 MODELO FACTORIAL DE APOYO A LAS NECESIDADES PSICOLÓGICAS BÁSICAS .	71
FIGURA 12 MODELO FACTORIAL DE SATISFACCIÓN DE LAS NECESIDADES PSICOLÓGICAS BÁSICAS.....	72
FIGURA 13 MODELO FACTORIAL DE VITALIDAD SUBJETIVA	73
FIGURA 14 DISTRIBUCIÓN DE RESPUESTAS DE LAS MEDIDAS DE UN SOLO ÍTEM MUESTRA 1	83
FIGURA 15 DISTRIBUCIÓN DE RESPUESTAS DE LAS MEDIDAS DE UN SOLO ÍTEM MUESTRA 2	84
FIGURA 16 DISTRIBUCIÓN DE RESPUESTAS DICOTOMIZADAS DE LAS MEDIDAS DE UN SOLO ÍTEM MUESTRA 2	85
FIGURA 17 CORRELOGRAMA DE LAS MEDIDAS DE UN SOLO ÍTEM MUESTRA 1	91
FIGURA 18 CORRELOGRAMA DE LAS MEDIDAS DE UN SOLO ÍTEM MUESTRA 1 FASE TEST.	92
FIGURA 19 CORRELOGRAMA DE LAS MEDIDAS DE UN SOLO ÍTEM MUESTRA 2 FASE RETEST	94

ABREVIATURAS

AA	Medida de apoyo a la autonomía (nombre breve para el constructo: apoyo a la necesidad psicológica básica de autonomía). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
AC	Medida de apoyo a la competencia (nombre breve para el constructo: apoyo a la necesidad psicológica básica de competencia). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
AERA	<i>American Educational Research Association</i>
AMO	Medida de amotivación (no regulación). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
APA	<i>American Psychological Association</i>
AR	Medida de apoyo a la relación (nombre breve para el constructo: apoyo a la necesidad psicológica básica de relación). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
BPNES	<i>Basic Psychological Need Exercise Scale</i> ; Escala de necesidades psicológicas básicas en ejercicio
C-OAR-SE	<i>Construct definition- Object classificattion –Attribute classification -Rater Identification- Scale formation and enumeration and reporting</i> ; Definición del constructo - Clasificación de objetos - Clasificación de atributos - Identificación del evaluador - Formación de escala, enumeración e informe

CEEAH	<i>Comissió d'Ètica en l'Experimentació Animal i Humana</i> ; Comisión de Ética en la Experimentación Animal y Humana
CFA	<i>Confirmatory factor analysis</i> ; Análisis factorial confirmatorio
CFI	<i>Comparative fit index</i> ; Índice de ajuste comparativo
CIS-Q	<i>Coaches Interpersonal Style – Questionnaire</i> ; Cuestionario de Estilo Interpersonal de los Entrenadores
CSI	<i>Comprehensive single item</i>
EXT	Medida de regulación externa. En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
FA	Medida de frustración a la autonomía (nombre breve para el constructo: frustración a la necesidad psicológica básica de autonomía). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
FC	Medida de frustración a la competencia (nombre breve para el constructo: frustración a la necesidad psicológica básica de competencia). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
FR	Medida de frustración a la relación (nombre breve para el constructo: frustración a la necesidad psicológica básica de relación). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
GEPE	<i>Grup d'estudis de Psicologia de l'Activitat Física i l'Esport</i> ; Grupo de estudios de Psicología de la Actividad Física y del Deporte
h ²	Comunalidad

IDE	Medida de regulación identificada. En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
INT	Medida de regulación intrínseca. En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítem
ITG	Medida de regulación integrada. En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
ITY	Medida de regulación introyectada. En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
NCME	<i>National Council on Measurement in Education</i>
PABAK	<i>Prevalence-adjusted bias-adjusted kappa</i> ; Kappa ajustada por sesgo y prevalencia
RMSEA	<i>Root mean error of approximation</i> ; Error cuadrático medio de aproximación
R_{yy}	Fiabilidad obtenida mediante fórmula de corrección por atenuación
SA	Medida de satisfacción de la autonomía (nombre breve para el constructo: satisfacción de la necesidad psicológica básica de autonomía). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
SC	Medida de satisfacción de la competencia (nombre breve para el constructo: satisfacción de la necesidad psicológica básica de competencia). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems.
SDT	<i>Self-determination theory</i> ; Teoría de la autodeterminación

SEM	<i>Structural equation modeling</i> ; Modelos de ecuaciones estructurales
SR	Medida de satisfacción de la relación (nombre breve para el constructo: satisfacción de la necesidad psicológica básica de relación). En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
SVS	Subjective Vitality Scale; Escala de Vitalidad Subjetiva
TLI	<i>Tucker-Lewis Index</i> ; Índice de Tucker-Lewis
UAB	Universitat Autònoma de Barcelona
ULSMV	<i>Unweighted least squares mean and variance adjusted</i> ; Mínimos cuadrados no ponderados corregidos por media y varianza
VI	Medida de vitalidad subjetiva. En mayúscula hace referencia a la medida de un solo ítem. En minúscula, seguido de un número, hace referencia a los ítems de las medidas de múltiples ítems
α	Alfa de Cronbach
ω_c	Omega categórica
r_{xy}	Correlación entre las medidas de un solo ítem y las medidas de múltiples ítems
ρ_{xx}^*	Fiabilidad aplicando profecía de Spearman-Brown

CONSIDERACIONES LINGÜÍSTICAS

En este trabajo se han seguido las consideraciones lingüísticas de la Universitat Autònoma de Barcelona (UAB) para un uso no sexista del lenguaje.

Para facilitar la comparación con la literatura científica actual, el uso de siglas y acrónimos en este trabajo se ha conservado en inglés. En aquellos casos en los que existe también una traducción al castellano conocida se proporciona también su equivalente en castellano en la sección de abreviaturas.

De la misma forma se ha considerado el uso de extranjerismos como la palabra *software*. Al tratarse de palabras que conservan la grafía y la pronunciación originaria en este texto se encontrarán escritas en cursiva tal y como recomienda la Real Academia Española de la lengua.

Aunque tradicionalmente en el uso del castellano se ha utilizado la coma como separador decimal viniendo de la tradición francoalemana, en este trabajo se ha utilizado como separador decimal el punto proveniente de la tradición anglosajona aceptado en el Sistema Internacional de Unidades y el ISO 80000.

1

EL PROBLEMA APLICADO Y LA SOLUCIÓN METODOLÓGICA

Una de las necesidades encontradas dentro del proyecto de investigación “ACTIVATICS: Promoción de la salud y la adhesión a la actividad física a través de la teoría de la autodeterminación y con el apoyo de aplicaciones móviles”, en el que se enmarca esta tesis, fue evaluar constructos de la teoría motivacional de la autodeterminación (SDT) en personas mayores que practicaban actividad física con la finalidad de desarrollar hacer intervenciones personalizadas dirigidas a promover la actividad física en esta población. En un inicio, con la finalidad de tener medidas comparables a las de otros grupos participantes en el proyecto (i.e., adolescentes y deportistas de élite), se pretendió realizar una recogida de datos colectiva dentro del contexto de sesiones de actividad física dirigidas con el mismo protocolo y medidas que el grupo de adolescentes. Después de una primera prueba piloto se vio rápidamente que no era factible utilizar las mismas medidas en personas mayores que con adolescentes, ya

que en un contexto grupal con personas mayores no era posible administrar 65 ítems y obtener respuestas.

Esta necesidad en seguida se convirtió en un reto: ¿Cómo podemos evaluar los mismos constructos que en los otros grupos? En general, para medir constructos complejos, como pueden ser los de la SDT, se suelen utilizar medidas de múltiples ítems. La idea es reflejar mejor la complejidad pero también que es conocido que las medidas de múltiples ítems pueden llevar a una tasa de respuesta baja, a la interrupción en la administración del cuestionario o a obtener muchos valores perdidos (Dillman, Sinclair, y Clark, 1993), y si consideramos que las personas mayores necesitan más tiempo para procesar la información que las personas más jóvenes (Li, Banerjee, y Zumbo, 2017) queda claro que es necesario buscar una solución que simplifique el proceso de respuesta en nuestro grupo de interés. La simplificación más elemental consiste en reducir el número de ítems que deberá contestar cada persona. Para no omitir ninguno de los constructos contemplados en la SDT propusimos que las medidas de un solo ítem podían ser una alternativa a las medidas de múltiples ítems comúnmente utilizadas en SDT hasta el momento. Estas medidas son discutidas psicométricamente hablando, pero hay al menos dos buenos motivos en su favor. Como indican Wanous y Reichers (1996), el principal argumento para no utilizar medidas de un solo ítem suele ser que no es posible obtener un coeficiente de consistencia interna, pero en cambio sí es posible obtener otros tipos de evidencia de calidad psicométrica tal y como demuestran en su trabajo. Como indican Fuchs

y Diamantopoulos (2009) la inclusión de medidas de un solo ítem puede ser preferible a ignorar algunos de los constructos contemplados por la teoría.

Una vez tomada la decisión de desarrollar medidas de un solo ítem para medir los constructos motivacionales en personas mayores se puso en evidencia la necesidad de profundizar y sistematizar el conocimiento en dos ámbitos: el psicométrico y el sustantivo. De esta forma se pretendía crear la sinergia para dar respuesta al reto que nos habíamos planteado.

En el ámbito de la psicometría, aunque actualmente las medidas de un solo ítem ya no son una herramienta nueva en investigación, el conocimiento sobre ellas no se encuentra centralizado. Es necesario el acceso a múltiples fuentes primarias para ver de forma global el estado de la cuestión. La literatura que trata sobre medidas de un solo ítem se ha centrado en la obtención de evidencias de calidad psicométrica de diferente tipo, pero no integrando toda la metodología desarrollada en un mismo trabajo. Es por ello por lo que antes del desarrollo de las medidas de un solo ítem para constructos complejos creímos necesario integrar el conocimiento obtenido hasta el momento, de manera que fuese posible tener una visión holística del tipo de evidencia psicométrica necesaria para el desarrollo y la interpretación de las puntuaciones de las medidas de un solo ítem.

Puesto que como ya se ha dicho el reto a cubrir fue el desarrollo de medidas de un solo ítem para los constructos de la SDT,

este capítulo se divide en dos secciones. En primer lugar, se describen los conceptos de esta teoría, así como los resultados obtenidos en su aplicación a la evaluación de la motivación para realizar actividad física en personas mayores.

En segundo lugar, siguiendo la estructura de clasificación de los indicadores de calidad psicométrica que se promueve en el texto normativo *Standards for Educational and Psychological Testing* (AERA [American Educational Research Association], APA [American Psychological Association], y NCME [National Council on Measurement in Education], 1999, 2014) se presenta y discute de manera integrada el estatus de las medidas de un solo ítem en relación con las evidencias de validez y de fiabilidad.

Los constructos de la teoría de la autodeterminación

La SDT es una teoría motivacional que hace énfasis en la importancia de los recursos innatos que poseen las personas para hacer frente al desarrollo de la personalidad, así como a su autorregulación (Deci y Ryan, 1985; Ryan y Deci, 2000). Esta teoría ha recibido atención en diversos campos de estudio, además de la psicología del deporte y de la actividad física, como pueden ser la psicología de la educación (e.g., Deci, Vallerand, Pelletier, y Ryan, 1991), la psicología de las organizaciones (e.g., Gagné y Deci, 2005) o la psicoterapia (e.g., Ryan y Deci, 2008).

En la Figura 1 se encuentra de forma simplificada la cadena de relaciones entre constructos propuesta por la SDT. En términos generales, tal y como indican Ryan y Deci (2000; 2017), la SDT postula que existen tres necesidades psicológicas básicas innatas. La primera, la necesidad psicológica básica de autonomía (necesidad de autonomía) se refiere a la percepción de autodirección en lugar de dirección por fuerzas externas. Dicho de otra forma, guarda relación con la percepción de elección de los comportamientos y de hacerlos por voluntad propia. La segunda, la necesidad psicológica básica de competencia (necesidad de competencia) se define como la percepción de que se dan las condiciones para que la persona pueda expresar sus capacidades o talentos. La tercera, la necesidad psicológica básica de relación (necesidad de relación) se conceptualiza por un lado como la capacidad de experimentar a las otras personas como receptivas y sensibles, y por otro lado como la capacidad de ser

receptivas y sensibles a ellas. Simplificándolo podemos decir que tiene que ver con el sentimiento de sentirse conectada e involucrada con las otras personas.

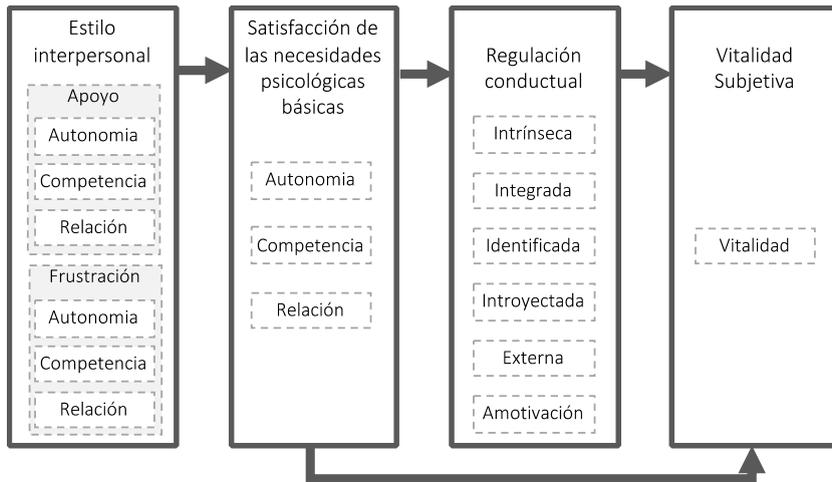


Figura 1 Cadena de Relaciones Propuesta por la Teoría de la Autodeterminación

Un elemento clave de esta teoría es la idea de que satisfacer estas necesidades psicológicas básicas es esencial para el crecimiento psicológico, la integridad, el bienestar, las experiencias de vitalidad, así como la auto-congruencia. La SDT postula que las personas buscan situaciones en las cuales se puedan ver satisfechas estas necesidades y también que podrán percibir la satisfacción de las tres necesidades psicológicas básicas por separado (Véase el cuadro 2 de la Figura 1): satisfacción de la necesidad psicológica básica de autonomía (satisfacción de la autonomía), satisfacción de la necesidad psicológica básica de competencia (satisfacción de la competencia) y satisfacción

de la necesidad psicológica básica de relación (satisfacción de la relación). La satisfacción de las necesidades psicológicas básicas se facilita por el clima generado en el contexto que se desarrolla una actividad. En dicho contexto participan distintas variables, pero este trabajo se centra en el estilo interpersonal de personas con autoridad, concretamente de las sesiones de actividad física. Este concepto está representado en el primer cuadro de la Figura 1. Concretamente se distingue el apoyo y la frustración (*thwarting*) de las necesidades psicológicas básicas. De esta forma es necesario distinguir entre seis estilos interpersonales: apoyo a la necesidad psicológica básica de autonomía (apoyo a la autonomía), apoyo a la necesidad psicológica básica de competencia (apoyo a la competencia), apoyo a la necesidad psicológica básica de relación (apoyo a la relación), frustración a la necesidad psicológica básica de autonomía (frustración a la autonomía), frustración a la necesidad psicológica básica de competencia (frustración a la competencia), y frustración a la necesidad psicológica básica de relación (frustración a la relación). El apoyo a las necesidades psicológicas básicas, por parte de agentes promotores podrá facilitar una mayor satisfacción de estas, mayor motivación y bienestar. De la misma forma si en lugar de satisfechas estas necesidades se perciben como frustradas también se esperará un cambio negativo en motivación y bienestar. Aunque se trata de seis dimensiones distintas, se prevé una alta correlación entre las dimensiones de estilo interpersonal (Wilson, Paige, y Mack, 1982). Concretamente se espera una relación positiva entre las tres

dimensiones de apoyo y también entre las tres de frustración, y negativa entre los dos grupos.

En esta teoría, la motivación no debe entenderse como un solo constructo. Tal y como explican Ryan y Deci (2000) la SDT ofrece un marco en el cual se diferencian diversos tipos de motivación. Cada tipo de motivación tendrá consecuencias diferentes en los procesos de aprendizaje, rendimiento y bienestar, entre otros. El tipo de motivación vendrá modulado por la regulación conductual que se dé. Tal y como se observa en el tercer cuadro de la Figura 1, se distinguen seis tipos de regulación conductual que forman parte de un continuo. En un extremo se encuentra el comportamiento autodeterminado regulado intrínsecamente (regulación intrínseca) y caracterizado por realizar una actividad por puro interés, disfrute y satisfacción inherente. La regulación integrada, siendo la segunda regulación más autodeterminada de todas después de la regulación intrínseca, describe el hecho de hacer las actividades en congruencia con los valores que tiene una persona y con su ser. La regulación identificada, aún considerada autodeterminada hace referencia a la importancia que tiene la actividad para la persona. Estas tres regulaciones constituyen la motivación autónoma, es decir la motivación que se da por mero disfrute de la actividad, porque es congruente con los valores de cada persona o por su importancia para esta persona. En contraposición con la motivación autónoma se encuentra la motivación controlada, es decir, cuando existe motivación, pero está basada en algún tipo de presión o contingencia ya sea de carácter interno o

externo. De esta forma distinguimos la regulación introyectada, que hace referencia a la realización de una actividad con el fin de obtener recompensas o evitar castigos internos. Sería el caso, por ejemplo, de realizar una actividad para que desaparezca la percepción de ansiedad. Por otro lado, identificamos la regulación externa, que ha sido el tipo de motivación extrínseca más estudiada, que hace referencia a la obtención de recompensas o castigos externos. En el otro extremo del continuo encontramos la no regulación (*amotivación*) que consiste en estar realizando una actividad sin ningún tipo de motivación. Este sería el caso, por ejemplo, de una persona que tras estar realizando una actividad durante un tiempo y ver que no cumple sus expectativas, o que se aburre, sigue realizando la actividad, pero sin saber porqué la está realizando, maximizando la probabilidad de abandonarla.

Al final de la cadena de relaciones se encuentra el bienestar representado en el cuarto cuadro de la Figura 1. El bienestar “hace referencia al funcionamiento y experiencia psicológica óptima” (Ryan y Deci, 2001, pp. 142). En general la SDT distingue dos tipos de bienestar: el hedónico y el eudaimónico. El primero equipara el bienestar con la experiencia de placer y felicidad. El segundo, en el que se centra este trabajo, tiene en cuenta la autorrealización de las personas. Mientras que la felicidad puede representar al bienestar hedónico no es así con el bienestar eudaimónico. En este trabajo se toma como indicador de bienestar eudaimónico la vitalidad subjetiva. Tal y como se ha indicado, una de las formas de promocionar el bienestar será mediante la satisfacción de las necesidades psicológicas básicas. Otro camino que

también está representado en la Figura 1, mediante una flecha, será promoviendo una motivación más autodeterminada.

Aplicación a la realización de actividad física en personas mayores

La SDT se ha aplicado a diversos grupos y ámbitos de actividad humana. Dentro del ámbito de la actividad física, encontramos aplicaciones en grupos humanos de diversas edades. Por ejemplo, infantes que practican actividad física en áreas con una situación económica desaventajada (e.g., Breslin, Shannon, Fitzpatrick, Hanna, Belton, y Brennan, 2017), adolescentes escolarizados que practican Educación física (e.g., Yang, Viladrich, y Cruz, 2019) o mayores de 18 años que practican ejercicio (e.g., Sibley y Bergman, 2017). Tal y como indican algunos autores (e.g., Dacey y Newcomer, 2005; Kirkland, Karlin, Stellino, y Pulos, 2011) la SDT deviene también un marco teórico adecuado para comprender las motivaciones de las personas mayores para realizar actividad física.

Sin embargo, este último grupo presenta características particulares. Por una parte, basándonos en evidencia científica podemos decir que la satisfacción de la autonomía es el mayor predictor de los comportamientos autodeterminados en adultos entre 16 y 64 años (Vlachopoulos y Michailidou, 2006). No obstante, Marcos, Orquín, Pedreño y Moreno-Murcia (2014), en el campo de la actividad física, sugieren que en personas mayores de entre 65 y 85 años, solo la satisfacción de la competencia es un predictor de estos

comportamientos. Estos hallazgos son también compatibles con los de Couto, Antunes, Monteiro, Moutão, Marinho y Cid (2017), en personas mayores de entre 60 y 90 años, que establecen la satisfacción de competencia como un predictor de la actividad física, y a la vez identifican que aquellos que tienen mayores niveles de actividad física expresan tener mayor bienestar. En la misma línea, en un estudio longitudinal a lo largo de cuatro años con personas de entre 87 y 97 años al inicio del estudio (Neubauer, Schilling, y Wahl, 2017), se vio como solo la satisfacción de la competencia emergía como un predictor significativo de la variabilidad intraindividual del bienestar. Tal y como indican los mismos autores esto puede tener una doble lectura: por un lado, puede ser explicado porque a medida que las personas se hacen mayores pierde importancia la satisfacción de la autonomía y dan más importancia a otros factores como por ejemplo la satisfacción de la competencia; por otro lado, podría explicarse por un problema con el instrumento de medida de la satisfacción de la autonomía que emplean, ya que tal y como indican, no está formulado desde la SDT. Por el contrario, otros autores (Henning, Bjälkebring, Stenling, Thorvaldsson, Johansson, y Lindwall, 2019) encontraron que la satisfacción de las tres necesidades psicológicas básicas impactaba en el bienestar, mientras que a nivel intraindividual la satisfacción de la competencia no. Así pues, aunque en general el enfoque de las intervenciones suele estar encaminado a promover la satisfacción de la autonomía (Alcaraz, Pons, Angulo-Brunet, y Cruz, 2017) este papel no está tan claro en personas mayores. En este grupo humano es

importante clarificar el rol de la satisfacción de las necesidades psicológicas básicas para entender que papel juega en la motivación autónoma y/o la vitalidad subjetiva.

Aunque se han utilizado diversas escalas para evaluar la satisfacción de las necesidades psicológicas básicas en personas mayores, las propiedades psicométricas en muestras de estas características todavía son poco conocidas. Por ejemplo, Henning et al. (2019) utilizan el promedio de dos ítems como indicador de cada dimensión de la escala de satisfacción y frustración a las necesidades psicológicas básicas extraídos de la escala de Chen et al. (2015). O Marcos et al. (2014) con la escala de satisfacción de las necesidades psicológicas básicas en el ejercicio (Sanchez y Nuñez, 2007) admiten no hacer análisis más detallados, aportando solamente evidencia psicométrica con alfa de Cronbach, debido al tamaño de la muestra.

Respecto a los agentes promotores que pueden facilitar o frustrar la satisfacción de las necesidades psicológicas básicas, el interés por el estudio de los estilos interpersonales tiene ya larga tradición. En 2011, Su y Reeve (2011) realizaron un meta-análisis incluyendo un total de 19 trabajos sobre programas de apoyo a la autonomía. En los últimos años desde la psicología del deporte y de la actividad física, tal y como indican Ntoumanis, Quested, Reeve y Cheon (2017) ha crecido el interés en el estudio de los estilos interpersonales. Una clara evidencia de dicho interés es la reciente creación de diferentes instrumentos para el propósito de medir los estilos interpersonales ya sea con cuestionarios (e.g., Appleton, Ntoumanis,

Quested, Viladrich, y Duda, 2016; Bhavsar et al., 2019; Pulido, Sánchez-Oliva, Leo, Sánchez-Cano, y García-Calvo, 2017; Rocchi, Pelletier, y Desmarais, 2017) o con plantillas de observación (e.g., Cheon, Reeve, y Moon, 2012; Fabra, Balaguer, Tomas, Smith, y Duda, 2018; Pulido, Sánchez-Oliva, Silva, Palmeira, y García-Calvo, 2019). De forma similar a lo que ha sucedido con las escalas de satisfacción de las necesidades psicológicas básicas, a pesar de todos estos esfuerzos en los últimos años para tener instrumentos para evaluar la percepción del estilo interpersonal de personas con autoridad aún no se ha obtenido evidencia de calidad psicométrica relativa a su percepción por parte de las personas mayores.

También existe escasa literatura que examine las relaciones de las regulaciones conductuales con otras variables en personas mayores. Aunque en los últimos años en algunos estudios se ha utilizado el cuestionario de regulación conductual que contempla las seis dimensiones (González-Cutre, Sicilia, y Fernández, 2010; Markland y Tobin, 2004; Wilson, Rodgers, Loitz, y Scime, 2006) en personas mayores (e.g., Maillot, Becquard, y Perrot, 2018; Marcos et al., 2014) tampoco existe evidencia de calidad psicométrica que avale su uso en esta población. Aún así, algunos autores, por ejemplo Solberg, Halvari y Ommundsen (2013) examinaron los moderadores y los mediadores entre la satisfacción de las necesidades psicológicas básicas y la vitalidad subjetiva. En este estudio los autores mediante un *path analysis* observan como las personas que acababan abandonando la actividad que estaban realizando manifestaban tener menor vitalidad

subjetiva y menor satisfacción de la competencia que las personas que continuaban haciendo actividad física. A su vez, una mayor motivación autónoma estaba relacionada con cambios positivos en la vitalidad subjetiva mientras que aumentos en la motivación controlada estaban negativamente asociados con cambios en la vitalidad subjetiva. Otro ejemplo puede ser el de Gray, Wharf y Rhodes (2017) en el que reportan el uso de medidas para evaluar las necesidades psicológicas básicas, así como la regulación conductual, pero en ningún momento contemplan ninguna propiedades psicométrica en su estudio, examinado directamente los efectos mediante ANCOVA.

En resumen podemos decir que la SDT ha sido ampliamente probada en grupos de personas jóvenes y adultas, mostrándose que la satisfacción de la autonomía es el principal predictor de los resultados positivos de manera que las intervenciones basadas en esta tesis se han encaminado a promover la satisfacción de esta necesidad psicológica básica. Al mismo tiempo los cuestionarios necesarios para evaluar los constructos de la SDT se han desarrollado y validado también en estos grupos humanos. Lo sorprendente es que cuando se ha intentado verificar los postulados de la SDT en personas mayores estos esfuerzos no han venido acompañados de las comprobaciones psicométricas necesarias, pasando directamente a estudiar las relaciones entre los constructos. Otro dato llamativo en este grupo humano es que la satisfacción de la competencia es un predictor más importante de los resultados positivos que la satisfacción de la autonomía, lo que ha

llevado a la conclusión que es necesario seguir estudiando el rol de las tres necesidades psicológicas básicas en personas mayores.

Es decir, para desarrollar intervenciones que promuevan la motivación para la realización de actividad física por parte de personas mayores en el marco de la SDT es necesario realizar dos pasos previos. En primer lugar hay que cubrir la falta de datos sobre validez y fiabilidad para medir los constructos implicados en esta población específica. En segundo lugar hay que estudiar si en esta población se establecen las relaciones previstas por la SDT en la misma dirección e intensidad que en las poblaciones de adolescentes y adultos, atendiendo a que los estudios empíricos sugieren que podría no ser así.

Las medidas de un solo ítem como una forma de reducción de cuestionarios

Utilizar cuestionarios reducidos es una práctica extendida tanto en investigación como en psicología aplicada. Podemos discutir las ventajas prácticas para las personas que contestan cuestionarios reducidos en comparación a cuestionarios largos y estaremos de acuerdo en que tienen aspectos positivos como reducir la fatiga, aumentar la atención y en general hacer que las personas destinatarias tengan mayor predisposición a contestarlos. Ahora bien, probablemente nos costaría más aceptar las consecuencias que puede tener tomar decisiones basadas en esos cuestionarios reducidos. Dicho de otra manera, si decidimos administrar un cuestionario con el que vamos a tomar decisiones importantes necesitaremos tener el convencimiento de que el cuestionario es válido y fiable. Ambas propiedades son más fáciles de obtener con cuestionarios formados por múltiples ítems. Encontrar el equilibrio entre longitud del cuestionario y calidad psicométrica óptima debería ser una prioridad para cualquier persona que utilice instrumentos de evaluación psicológica.

Aunque reducir cuestionarios es una práctica habitual en muchos campos de la psicología, tal y como indican Alcaraz, Viladrich y Torregrosa (2013) no es una práctica muy extendida en psicología del deporte. Sin embargo, hay algunos ejemplos, sin ir más lejos estos mismos autores desarrollaron tres cuestionarios reducidos para unas escalas relacionadas con el ambiente en los equipos y la motivación

orientada a la tarea obteniendo buenas propiedades psicométricas en las muestras en las que se aplicó el cuestionario.

En el presente trabajo se considera un caso particular de la reducción de cuestionarios: las medidas de un solo ítem. Las medidas de un solo ítem son, si cabe, más controvertidas que el resto de los cuestionarios reducidos. Aunque en el campo de la salud han sido ampliamente utilizadas, su uso en el campo de la psicología del deporte y de la actividad física para medir constructos complejos, como pueden ser los de la SDT, ha sido hasta la fecha limitado. Sin embargo, hay excepciones, cabe destacar el trabajo de Bruton, Mellalieu y Shearer (2016) que desarrollaron una medida de un solo ítem para la evaluación de la eficacia colectiva, o el trabajo de Kwon y Trail (2005) con una medida de lealtad en el deporte. Aunque hay pocos antecedentes claros para los constructos complejos, el uso de medidas de un solo ítem para evaluar constructos más simples si que ha sido bien establecido en psicología del deporte y la actividad física. Es así como la actividad física global ha sido evaluada en múltiples estudios con medidas de un solo ítem (e.g., Milton, Cledes, y Bull, 2013; Wanner, Probst-Hensch, Kriemler, Meier, Bauman, y Martin, 2014; Zwolinsky, McKenna, Pringle, Widdop, y Griffiths, 2015). Teniendo en cuenta que las medidas de un solo ítem para constructos simples han sido bien aceptadas en este campo, es de esperar, que igual que en otras disciplinas afines se han desarrollado medidas de un solo ítem para constructos complejos, también puedan tener una buena aceptación en el marco de la SDT.

En la última década algunos autores han dado recomendaciones para la reducción de cuestionarios (e.g., Hayduk y Littvay, 2012; Kruey, Emons, y Sijtsma, 2013, 2014; Raykov, Rodenberg, y Narayanan, 2015). Son de especial interés las recomendaciones dadas por Kruey et al. (2013) a partir de la revisión sistemática de los artículos sobre reducción de cuestionarios publicados entre 2005 y 2010 en seis revistas líderes en psicología. Estos autores dan diversas recomendaciones generales a considerar cuando se reducen cuestionarios, como dejar de confiar en reglas generales (*rules of thumb*) basadas en un umbral de fiabilidad asumiendo que serán seguras para la aplicación de tests, o como no transferir las propiedades de validez del cuestionario largo al cuestionario reducido. Si bien las recomendaciones son claras, la metodología para llevarlas a la práctica está publicada de forma dispersa tal y como veremos en los apartados siguientes.

Evidencias de validez para medidas de un solo ítem

Las últimas recomendaciones para la publicación de estudios cuantitativos promovidas por la APA (Appelbaum, Cooper, Kline, Mayo-Wilson, Nezy, y Rao, 2018) establecen como punto de partida los *Standards for educational and psychological testing* en sus últimas versiones (AERA, APA, y NCME, 1999, 2014) y hacen especial énfasis en la obtención y reporte de diferentes evidencias de validez. Ello solo constituye una vuelta de tuerca más a lo que ya decía el manual de publicaciones de la APA (2010), que instaba a reportar evidencia de

fiabilidad y validez de las puntuaciones. Sin embargo, el mensaje aún no ha llegado a la mayoría de las investigaciones. Tal y como explican Viladrich, Angulo-Brunet y Doval (2017) la anterior recomendación de APA (2010) en muchas ocasiones se había reducido a reportar el coeficiente alfa de Cronbach antes de realizar cualquier tipo de análisis excluyendo otros tipos de evidencia de validez o fiabilidad.

En el caso de las medidas de un solo ítem estas recomendaciones cobran especial importancia teniendo en cuenta la imposibilidad de obtener un coeficiente de consistencia interna para un solo ítem. Por lo tanto, disponer de métodos para proveer del máximo de evidencias de validez y fiabilidad que sea posible, teniendo en cuenta las características de las medidas y su diseño, y utilizando las técnicas estadísticas apropiadas, será un punto esencial para el desarrollo de estas medidas.

Las medidas de un solo ítem, aunque no exentas de crítica, emergieron en investigación aplicada, siendo una práctica común en el campo de la psicología industrial y en el de la gestión de recursos humanos (Wanous y Reichers, 1996). En sus inicios la evaluación de las propiedades psicométricas de estas medidas radicaba principalmente en la obtención de evidencias de validez.

Es así como en disciplinas afines, como el máquetin, se desarrollaron procedimientos como C-OAR-SE (*Construct definition-Object classification –Attribute classification -Rater identification- Scale formation and enumeration and reporting*; Rossiter, 2002) centrados

en la obtención de evidencias de validez de contenido, exponiendo que era innecesaria la introducción de pruebas estadísticas.

Precisamente, una crítica evidente a las medidas de un solo ítem puede ser la pérdida de validez de contenido en el caso de constructos complejos ya que posiblemente se necesita más de un ítem para cubrir todo el dominio. En este sentido, tal y como Jordan y Turner (2008) indican, las medidas de un solo ítem pueden clasificarse en dos categorías: atributos observables y medidas cognitivas o afectivas. En el caso de las medidas cognitivas o afectivas, constructos complejos, se tomarán consideraciones especiales y al menos será necesaria la creación de una medida de un ítem para cada dimensión del constructo.

Tal y como indica Bergkvist (2016), para seleccionar el mejor ítem es necesario confiar en el juicio de personas expertas. Para la selección o creación del mejor indicador a menudo lo más adecuado será seleccionar el reactivo que sea semánticamente parecido o igual al descriptor del atributo, o a la definición del constructo. Tras un estudio de simulación, Diamantopoulos, Sarstedt, Fuchs, Wilczynski y Kaiser (2012) llegaron a la conclusión que tan solo era recomendable crear medidas de un solo ítem cuando el constructo se podía definir en una sola palabra, había muestras pequeñas ($n < 50$), se esperaban tamaños de efecto bajos (correlaciones entre ítems de diferentes dimensiones $< .30$), cuando los ítems de la escala original eran homogéneos (correlaciones entre ítems de la misma dimensión $> .80$, fiabilidad $> .90$) y cuando los ítems eran semánticamente redundantes.

Aunque parece sensato en estos casos utilizar evidencias basadas en la validez de contenido para el desarrollo de medidas de un solo ítem, tras analizar un total de 108 publicaciones que citaban la publicación inicial del procedimiento C-OAR-SE, a Rossiter (2002), Sarstedt, Diamantopoulos y Salzberger (2016) concluyeron que tan solo se hacía énfasis en la validez de contenido en 10 publicaciones (9.26%) y que el 11.11% si que hacían una definición del constructo. Sin embargo, un 35.10% usaban las medidas de un solo ítem sin añadir muchas explicaciones.

Uno de los argumentos centrales para el uso de estas medidas suele ser la mejora en el proceso de respuesta. En el caso de poblaciones específicas, como las personas mayores, en las que el proceso de respuesta aún es más costoso que en adultos más jóvenes, este tipo de evidencia puede adquirir especial importancia. Si la factibilidad con medidas de múltiples ítems es tan limitada que no puede desarrollarse una recogida de datos, obtener información a partir de medidas de un solo ítem puede dar más información que no obtener ningún tipo de medida. En la misma línea del procedimiento C-OAR-SE, encontramos el procedimiento CSI (*comprehensive single ítem*; Konstabel, Lönnqvist, Walkowitz, Konstabel, y Verkasalo, 2012). En este método el énfasis no recae tanto en el hecho de que el ítem pueda definirse en una sola palabra sino en el proceso de respuesta (e.g., mantener la atención, ausencia de fatiga, ...). Un ejemplo de medidas de un solo ítem desarrolladas con este procedimiento es la escala de 30 ítems sobre las 30 facetas del *Big Five* (Konstabel, et al.,

2012). Estos autores defienden que el hecho de reducir la escala hará que las personas participantes estén más motivadas y presten mayor atención a las preguntas que se plantean. Además de este procedimiento, otras investigaciones también aluden al proceso de respuesta para justificar el uso de medidas de un solo ítem. Este es el caso de los ya mencionados Kwon y Trail (2005) que tras examinar varias propiedades psicométricas concluyeron que las medidas de un solo ítem tienen como ventaja la brevedad y simplicidad de su uso, mientras que las medidas de múltiples ítems cubren mejor la amplitud del constructo y tienen mejor fiabilidad.

Robins, Hendin y Trzesniewski (2001) señalan que las medidas de un solo ítem pueden eliminar la redundancia y el aburrimiento producido por contestar preguntas similares repetidamente. Otro aspecto que también destacan algunos autores (e.g., Bruton et al., 2016) es la disposición de las personas participantes a contestar los cuestionarios. Dicho de otra manera, el hecho de utilizar medidas de un solo ítem maximiza la adherencia de las participantes, facilitando que terminen la recogida de datos y por otro lado reduce el hecho de que las participantes contesten los cuestionarios sin prestar atención lo que podría conducir a sesgo de respuesta. Otro argumento a favor del uso de escalas de un solo ítem viene dado por Fuchs y Diamantopoulos (2009) que sugieren que en escalas de múltiples ítems las participantes no leen cuidadosamente los enunciados y hacen inferencias a partir de otros ítems reduciendo pues la calidad de las respuestas.

En cuanto a datos cuantitativos para apoyar la calidad psicométrica de las medidas de un solo ítem, las pruebas más utilizadas son la capacidad predictiva sobre otras variables y la relación con una medida homóloga de múltiples ítems. De esta manera Bergkvist y Rossiter (2007) examinaron la validez predictiva de una de las medidas más utilizadas en márketing: la actitud hacia una marca. Aunque concluyeron que no había diferencias entre la validez predictiva de las medidas de un solo ítem respecto a su homólogo con múltiples ítems su estudio fue criticado posteriormente por Kamakura (2015) aludiendo que ese tipo de evidencia no consistía evidencia de validez predictiva, sino de validez concurrente. La crítica de Kamakura fue respondida por Bergkvist (2015) con nuevos análisis y determinó que las conclusiones de Bergkvist y Rossiter (2007) seguían siendo admisibles tras haber testeado las hipótesis en las condiciones oportunas. A pesar de las críticas que ha recibido el artículo, en septiembre de 2013 el artículo de Bergkvist y Rossiter (2007), tal y como señala Kamakura, había recibido más de 600 citas. Seis años después (septiembre 2019) el artículo, en Scopus, ya tiene más de 1180 citas. Aunque las medidas de un solo ítem tienen sus puntos débiles y sus puntos fuertes, tal y como vemos, siguen siendo utilizadas.

Ya en el campo de la psicología del deporte los autores que desarrollaron la medida de eficacia colectiva (Bruton et al., 2016) hicieron hincapié en este tipo de evidencias. El estudio recoge tanto evidencias de validez concurrente como predictiva concluyendo que las

medidas de un solo ítem pueden ser igual de válidas y fiables que las medidas de múltiples ítems.

Un recurso ampliamente utilizado para la validación concurrente ha sido correlacionar la medida de un solo ítem con la medida de múltiples ítems que mide el mismo concepto. Este es el caso por ejemplo de escalas de satisfacción con la vida (Cheung y Lucas, 2014; Gnams y Buntins, 2017; Jovanović y Lazić, 2018) donde observan valores superiores a .60 en este tipo de coeficientes de correlación, considerándolo una medida de asociación fuerte.

Al igual que con las medidas de múltiples ítems, un tipo de evidencia de validez de especial interés es el estudio de la red nomológica de relaciones de las variables estudiadas en el modelo teórico que se plantea. Tal y como apuntaron ya en 1955 Cronbach y Meehl, para que un constructo sea científicamente admisible se debe demostrar que se da dentro de una red nomológica de relaciones. En este sentido, y tal y como lo hacen Gogol et al. (2014), se puede obtener evidencia de calidad psicométrica de las medidas de un solo ítem evaluando el grado en el cual estas medidas reproducen la misma red nomológica que las medidas de múltiples ítems. En su investigación, por ejemplo, desarrollan tres medidas de un solo ítem para medir constructos emotivo-afectivos. Tal y como indican, hasta esa fecha no había trabajos que hubieran estudiado la red nomológica de relaciones entre las medidas de un solo ítem y también entre las de múltiples ítems para establecer si se reproducía la misma estructura. En un primer paso examinaron la correlación entre las medidas de un

solo ítem y las de múltiples ítems. Posteriormente examinaron la correlación de las medidas de un solo ítem y las de múltiples ítems con variables externas relevantes para la teoría. Aunque el estudio mediante correlaciones es un buen punto de partida, Teniendo en cuenta las posibilidades de la metodología actual, un paso más sería el estudio de las redes nomológicas mediante métodos que permitan comprobar el conjunto de todas las relaciones (e.g., modelos de ecuaciones estructurales [SEM], *path analysis*). En este sentido, aunque comprobar la red nomológica de relaciones para las medidas de un solo ítem podría ser suficiente, otra opción sería comprobar la misma red mediante las medidas de un solo ítem y también mediante las medidas de múltiples ítems provenientes de una misma muestra. De esta forma además de comprobar la red nomológica de relaciones de la medida de un solo ítem estaríamos aportando evidencia mediante las medidas de múltiples ítems de la réplica de la misma estructura, aportando una clara evidencia de equivalencia psicométrica entre ambas medidas. Para que esto suceda es importante que ambas redes se especifiquen con los mismos parámetros (e.g., correlaciones entre dimensiones, regresiones...). Claro que para poder obtener este tipo de evidencia es necesario un paso previo: obtener evidencia de la calidad psicométrica basada en la estructura interna y en la consistencia interna de las medidas de múltiples ítems. Si este paso previo no es favorable, no será posible comprobar la red nomológica de relaciones con las medidas de múltiples ítems.

Un tipo de evidencia de validez que no ha recibido especial atención es la validez basada en las consecuencias intencionadas y no intencionadas de la aplicación de un test. La incorporación de este tipo de evidencias en las últimas ediciones de los *Standards for Educational and Psychological Testing* (AERA, APA, y NCME, 1999, 2014) fue el producto de la discusión de las consecuencias que podía tener la evaluación en los años noventa (e.g., Padilla, Gómez, Hidalgo y Muñiz, 2006). Tal y como se desarrolla en el trabajo de Hubley y Zumbo (2011), basado en anteriores trabajos de Messick (1980, 1998), hay diversos aspectos a tener en cuenta cuando se valoran las consecuencias de los tests. Tal y como exponen, la mayoría de los cuestionarios se utilizan con una finalidad que tendrá consecuencias: ya sea para evaluar una intervención, tener retroalimentación, para políticas públicas o para procesos de decisión. Estas consecuencias son de dos tipos: implicaciones en los valores e implicaciones en las consecuencias sociales. El solo hecho de nombrar el concepto que se quiere medir puede estar influenciando la validez de la prueba ya que la persona que contesta valorará las consecuencias de contestar ese cuestionario y puede que sesgue su respuesta. Pongamos un ejemplo dentro de la SDT en personas mayores. Si se pregunta a las personas mayores sobre su satisfacción de las necesidades psicológicas básicas, siendo una pregunta sobre ellos mismos probablemente no tengan ningún problema en mostrarse sinceros. Ahora bien, si se pregunta si el/la monitor/a apoya estas necesidades puede que las personas modulen

su respuesta para que no tenga consecuencias negativas sobre su monitor/a.

Por otra parte, siempre de acuerdo con Zumbo y Hubley (2016), a partir de la aplicación del test pueden surgir algunas consecuencias sociales que pueden influenciar en la interpretación de las puntuaciones de un test. Siguiendo con el mismo ejemplo, ¿Qué pasaría si después de administrar el cuestionario sobre la percepción del estilo interpersonal de el/la monitor/a, se evidencia un problema de comunicación? ¿Qué consecuencias tendría? Aunque no explica como hacerlo, la propuesta de Messick (1980) para examinar este punto es estudiar que pasaría en este caso si se administra o no se administra un cuestionario. Si el objetivo es establecer una línea base sobre la que trabajar en una intervención, seguramente se quiere trabajar para mejorar el estilo comunicativo y por lo tanto estas consecuencias sociales están contempladas y serán atendidas. Pero quizá no está contemplado si se cuenta con un estudio transversal en el que se quieren estudiar las características de una población.

Además, si no se tienen en cuenta este tipo de evidencias, su interferencia podría conducir a conclusiones erróneas al interpretar otros tipos de evidencias. Por ejemplo, puede ser que una red nomológica de relaciones no se reproduzca debido a que no estemos midiendo realmente el constructo, y no porque esa cadena de relaciones no se de en realidad.

En pocas palabras las medidas de un solo ítem se han sustentado básicamente en evidencias de validez de tipo cualitativo: las

relativas al contenido y al proceso de respuesta a las que habría que añadir las relacionadas con las consecuencias de la evaluación. Respecto a las evidencias de tipo cuantitativo al estar descartado el estudio de la estructura interna, la evidencia se ha centrado en la relación del ítem con una medida homóloga de múltiples ítems, y su capacidad predictiva sobre otras variables. Solo últimamente se ha propuesto el estudio de la red nomológica de relaciones.

Evidencias de fiabilidad para medidas de un solo ítem

El estudio de las evidencias de fiabilidad en la literatura ha sido más sistematizado que el estudio de la validez. Por ejemplo, existen revisiones sistemáticas (e.g., Spörrle y Bekk, 2014) que afrontan los coeficientes de fiabilidad disponibles para las medidas de un solo ítem. Aunque las medidas de un solo ítem no dejan de ser un caso particular de reducción de cuestionarios, para estudiar su fiabilidad será necesario tomar enfoques distintos. Como ya advertía Loo (2002) las medidas de un solo ítem permiten la estimación de la fiabilidad mediante test-retest pero este tipo de medidas no permiten estimar la fiabilidad de consistencia interna que suele ser una evidencia central en el estudio de las propiedades psicométricas. Es importante tener en cuenta que evaluar la fiabilidad test-retest solo tendrá sentido cuando se trate con constructos estables (e.g., personalidad) mientras que si se trata con actitudes o estados de ánimo no será una buena aproximación (e.g., Jordan y Turner, 2008).

En la misma línea, otra forma de estimar la fiabilidad en un diseño longitudinal, cuando al menos se tienen tres recogidas de datos, será utilizar modelos cuasi-simplex para modelar un rasgo latente con una estructura autorregresiva (Kenny y Zautra, 1995, 2001; Lucas y Donnellan, 2012). En este caso la fiabilidad se estima como la ratio entre la suma de la varianza del rasgo estable y la varianza del rasgo autoregresivo entre la varianza total. Una aplicación de este método se encuentra en Gnambis y Buntins (2017). En el estudio comprobaron tanto con medidas de un solo ítem como con medidas de múltiples

ítems la fiabilidad de medidas de satisfacción con la vida. Encontraron que, aunque con otros tipos de evidencia de fiabilidad y validez esas medidas aportan suficiente evidencia en comparación a las medidas de múltiples ítems, las medidas de un solo ítem capturaban menor diferencia interindividual en cambios intraindividuales. Más recientemente, Alessandri, Zuffianò y Perinelli (2017) han extendido el modelo para medidas de múltiples ítems con la posibilidad de poder evaluarlo con solo dos recogidas de datos, aunque tal y como apuntan no es posible realizarlo con medidas de un solo ítem porque necesita al menos dos variables observadas en cada ocasión.

Hasta hoy todos los enfoques para estudiar la fiabilidad de las medidas de un solo ítem, más allá de la fiabilidad test-retest, requieren de la administración del cuestionario homólogo completo. Aunque es una aproximación muy utilizada hay que tener en cuenta dos consideraciones. La primera, esta idea parece contradictoria con Krueger et al. (2013) que recomienda no transferir las propiedades psicométricas de un cuestionario completo a su homólogo reducido. La segunda es que, igual que con las evidencias obtenidas a partir de comprobar la red nomológica de relaciones no era posible hacerlo a partir de medidas de múltiples ítems con una validez o fiabilidad no adecuada, calcular los índices de fiabilidad de las medidas de ítems únicos a partir de la fiabilidad de medidas de múltiples ítems que no han demostrado un mínimo de fiabilidad tampoco será adecuado. Bajo este diseño existen diversas alternativas de análisis a considerar. Una opción clásica es utilizar la profecía de Spearman-Brown (e.g., Nunnally,

1978). Como indican Kruyen et al. (2013), tomando como referencia la fiabilidad del cuestionario completo esta práctica no está exenta de crítica ya que incluso cuando se utiliza para predecir la fiabilidad de escalas reducidas que no son medidas de un solo ítem, en la práctica, los ítems no son paralelos y por lo tanto la fiabilidad que se intenta predecir es diferente que la que se está prediciendo con la fórmula de Spearman-Brown.

Otra opción especialmente interesante en fases iniciales de una investigación es el uso de la fórmula de corrección por atenuación (Wanous y Reichers, 1996). De esta forma, asumiendo que la correlación verdadera entre la medida de un solo ítem y la de múltiples ítems es 1 la fórmula queda simplificada como: $r_{xy} = \sqrt{r_{xx}} \cdot \sqrt{r_{yy}}$, donde r_{xy} es la correlación entre la medida de un solo ítem y la medida completa, $\sqrt{r_{xx}}$ es la fiabilidad de la medida de un solo ítem y $\sqrt{r_{yy}}$ es la fiabilidad de la medida completa. Así, si aislamos $\sqrt{r_{xx}}$ obtenemos la fiabilidad de la medida de un solo ítem. Aunque es un enfoque muy utilizado, en ocasiones puede conducir a conclusiones demasiado optimistas. Pongamos el caso que una medida de múltiples ítems ha mostrado un bajo coeficiente de fiabilidad, pero la correlación entre la medida de un solo ítem y las medidas de múltiples ítems es elevada. Mediante la obtención de fiabilidad por la fórmula de corrección por atenuación se puede llegar a pensar que la fiabilidad de la medida de un solo ítem es adecuada, mientras que en realidad teniendo en cuenta el bajo valor de la medida de múltiples ítems se estaría obteniendo

evidencia contradictoria debida al alto coeficiente de correlación entre ambas medidas.

Recientemente Zijlmans, van der Ark, Tijmstra y Sijtsma (2018) han adaptado otros métodos como el de Molenaar-Sijtsma (Mokken, 1971), el coeficiente λ_6 de Guttman (Guttman, 1945) o el coeficiente de fiabilidad de clases latentes (Andersen, Hagenars, y McCutcheon, 2006; Lazarsfeld, 1950; McCutcheon, 1987) para estimar la fiabilidad de un solo ítem y así poder elegir el ítem óptimo. A partir de datos simulados, así como también de datos reales, aplican esta metodología para estudiar la fiabilidad de las medidas de un solo ítem. En este trabajo los autores concluyen que estas propuestas no son una mejor aproximación que la obtención de la fiabilidad mediante la fórmula de corrección por atenuación. Tal y como indican los autores aún hace falta mucha investigación para determinar valores realistas de la fiabilidad de las medidas de un solo ítem. En este trabajo concluyen que el método de Molenaar-Sijtsma y la fórmula de corrección por atenuación son los más precisos. Más recientemente los mismos autores (Zijlmans, Tijmstra, van der Ark, y Sijtsma, 2019) han comparado el método de Molenaar-Sijtsma, el método λ_6 de Guttman, la fórmula de corrección por atenuación y la correlación ítem-total. Han llegado a la conclusión que utilizar la fórmula de corrección por atenuación o la correlación ítem-total lleva prácticamente a los mismos resultados. Esto es debido a que la fiabilidad obtenida a partir de la fórmula de corrección por atenuación también incluye la correlación ítem-total. Por lo tanto, destacan que quizá sería suficiente reportar la

correlación ítem-total en lugar de aplicar la fórmula de corrección por atenuación, por su simplicidad en el cálculo y sus resultados parecidos. Cambiar de enfoque, tal y como indican los mismos autores, también implica un cambio de filosofía ya que la interpretación de los dos no es estrictamente igual. En este sentido parece sensato aportar ambos coeficientes para poder hacer una mejor comparación con otros estudios según el enfoque que hayan tomado.

Tal y como indican Wanous y Hudy (2001) otra forma de estimar la fiabilidad de las medidas de un solo ítem es considerar las comunalidades a partir de un análisis factorial. Aunque se ha avanzado que la fiabilidad de consistencia interna no es posible con medidas de un solo ítem, esta puede ser una aproximación. Para poder usar este enfoque será necesario aplicar un análisis factorial a las medidas de múltiples ítems incluyendo también las correspondientes medidas de un solo ítem. Tal y como indican Wanous y Hudy asumiendo que la fiabilidad es la suma de la varianza común y de la varianza específica entre la varianza total, la comunalidad puede devenir una estimación de la consistencia interna de las medidas de un solo ítem.

En conclusión, se puede decir que hay dos tipos de indicadores de fiabilidad de las medidas de un solo ítem: los que se basan en la estabilidad temporal y son muy adecuados en estudios longitudinales y los que se derivan de la fiabilidad de medidas homólogas de múltiples ítems que requieren de la aplicación de ambas medidas, las de múltiples ítems y las de un solo ítem, a un mismo grupo de personas. Mientras que los primeros solo son adecuados para constructos

estables, estos últimos son más discutibles por el hecho de transferir las propiedades del cuestionario completo a su homólogo reducido y también por la falta de consenso sobre cual sería el mejor indicador entre la diversidad que están disponibles.

Además, en este segundo grupo de indicadores es importante considerar como se calcula previamente el coeficiente de fiabilidad de las medidas de múltiples ítems. En este sentido, son escasas las publicaciones que no utilizan el coeficiente alfa de Cronbach como coeficiente de fiabilidad (e.g., Gogol et al., 2014). Sin embargo, es conocido que el coeficiente alfa de Cronbach no siempre constituye el límite inferior de la fiabilidad de las puntuaciones (Viladrich et al., 2017) ya que en algunos modelos (e.g., errores correlacionados/bifactor) alfa de Cronbach puede sobrestimar la fiabilidad. Para obtener una buena estimación de la fiabilidad de las medidas de un solo ítem sería conveniente obtener la mejor estimación de fiabilidad para sus medidas homólogas de múltiples ítems teniendo en cuenta el modelo de medida que tienen (Savalei y Reise, 2019; Viladrich et al., 2017).

2

OBJETIVOS DE ESTA TESIS

En la formulación inicial del proyecto ACTIVATICS se proponía conocer el estatus de la motivación hacia la actividad física de personas mayores que practicaban actividad física con la finalidad de poder realizar intervenciones personalizadas dirigidas a promover la actividad física en esta población.

Tras examinar la literatura científica sobre SDT aplicada a personas mayores que realizan actividad física se vio que, aunque algunas investigaciones habían utilizado cuestionarios que median constructos de la SDT en esta población, las propiedades psicométricas de las puntuaciones todavía eran desconocidas. Además, algunos de los hallazgos con personas mayores no reproducían exactamente los contemplados por la SDT y encontrados en adolescentes y adultos más jóvenes especialmente por lo que se refiere a la importancia relativa de la satisfacción de la autonomía.

Los constructos de la SDT son conceptos complejos que se miden de manera autopercebida y, como tales, suelen registrarse mediante cuestionarios de múltiples ítems, sin embargo, en personas mayores la factibilidad de una recogida de datos en un contexto grupal

es limitada y por ello se propuso el uso de medidas de un solo ítem para poder examinar la cadena de relaciones de la SDT.

Tras conocer el estado de la cuestión y para poder aproximarnos a la cuestión inicial planteada, conocer el estatus de la motivación de las personas mayores para realizar actividad física dirigida en grupos, se plantean tres objetivos:

- a. Describir a las personas mayores de 60 años que practican actividad física en grupo. Concretamente se quiere describir la percepción de apoyo, frustración y satisfacción de las necesidades psicológicas básicas, la motivación hacia la actividad física y la vitalidad subjetiva.
- b. Conocer, en personas mayores de 60 años, las propiedades psicométricas de las puntuaciones de cuestionarios ya utilizados en otras poblaciones de adultos. Concretamente se plantea conocer las propiedades psicométricas de cuatro cuestionarios, uno para cada uno de los bloques teóricos de la SDT:
 - i. Estilo interpersonal
 - ii. Satisfacción de las necesidades psicológicas básicas
 - iii. Regulación conductual
 - iv. Vitalidad subjetiva
- c. Bajo la hipótesis de que utilizar medidas de un solo ítem puede contribuir a mejorar el proceso de respuesta, y en consecuencia mejorar la factibilidad de la recogida de datos, se propone proveer de cuestionarios de un solo ítem para medir los constructos fundamentales de la SDT en esta población. Para ello se propone:
 - i. Examinar la factibilidad de la recogida de datos y el proceso de respuesta a los ítems de las medidas de un solo ítem en comparación de las medidas de múltiples ítems.
 - ii. Obtener evidencia de validez y fiabilidad cuantitativa de las medidas de un solo ítem con la misma muestra que se han examinado las propiedades psicométricas de las medidas de múltiples ítems.
 - iii. Examinar la estabilidad temporal de las medidas de un solo ítem.

3

MÉTODO

Participantes

Para el desarrollo de las medidas de un solo ítem se contó, además de la propia autora de esta tesis, con un panel de ocho personas expertas, todas ellas integrantes del *Grup d'estudis de Psicologia de l'Activitat Física i l'Esport* (GEPE), que constaba de una metodóloga, una psicóloga de las organizaciones y seis psicólogos del deporte. También se contó con una muestra de cuatro personas mayores de 60 años a las cuales se les realizó una entrevista cognitiva para examinar la idoneidad de las medidas creadas.

En el estudio cuantitativo se analizaron los datos de 190 personas mayores de 60 años (80.75% mujeres; Edad: rango = 60 a 88, $M[DE] = 71.39 [6.19]$) que realizaban actividad física en grupos dirigidos por monitoras o monitores. Concretamente 128 personas (*Muestra 1*; 77.6% mujeres; Edad: rango = 60 a 87, $M[DE] = 70.21 [5.12]$) contestaron los cuestionarios largos de forma heteroadministrada y los ítems únicos de forma autoadministrada y 62 personas (*Muestra 2*; 87.10% mujeres; Edad: rango = 61 a 88, $M[DE] = 73.74 [7.39]$)

contestaron en dos ocasiones las medidas de un solo ítem de forma autoadministrada en la medida de lo posible. En la Tabla 1 puede encontrarse el resumen de las características de los centros y las actividades realizadas en ambas muestras.

Tabla 1 Descripción de Las Muestras Obtenidas en el Estudio

	Muestra 1 (n = 128)	Muestra 2 (n = 62)
Número de comarcas	3	2
Número de ciudades	4	4
Número de centros	7	4
Número de monitores/as	30	8
Actividades		
Acuáticas	66 (51.56 %)	
Mantenimiento	44 (34.38%)	62 (100 %)
Explosivas	12 (9.38%)	
Desconocido	6 (10.94%)	

Nota. Actividades acuáticas: aquagym, aquapower y natación. Actividades de mantenimiento: gimnasia de mantenimiento, estiramientos, yoga, bodymind, taichí, ... Actividades explosivas: spinning, zumba, aerobic, cardio, ...

Instrumentos

A continuación, se describen las medidas de múltiples ítems. Las medidas de un solo ítem no se describen ya que, tal como se ha dicho, se desarrollaron durante este trabajo y, en consecuencia, forman parte de los resultados. La solución final puede encontrarse también en el Apéndice A2. Los diferentes cuestionarios se describen en orden de presentación en el protocolo de obtención de datos. En el Apéndice A3 se encuentra el contenido de las medidas de múltiples ítems utilizadas para la recogida de datos. Para facilitar la interpretación de los datos de este trabajo se ha modificado la plantilla original identificando los ítems con nombres psicométricos que no fueron mostrados a las personas participantes. Para facilitar el proceso de respuesta se tomaron distintas precauciones. En primer lugar, se homogenizó el formato de respuesta entre los cuestionarios y se utilizó en todos los casos una escala de respuesta tipo Likert de 5 puntos (*1 = Totalmente en desacuerdo a 5 = Totalmente de acuerdo*). En segundo lugar, teniendo en cuenta las dificultades asociadas a la visión en esta población para los cuestionarios autoadministrados, se utilizó un tamaño de letra de 14 puntos y la tipografía Arial. En tercer lugar, se adaptaron los cuestionarios al sexo de las monitoras o monitores de la sesión. Finalmente, las personas participantes elegían si querían contestar el cuestionario en castellano o bien en catalán. Las propiedades psicométricas de estos cuestionarios en la presente muestra se proporcionan en el apartado de resultados ya que son un objetivo específico de este trabajo.

Estilo interpersonal. Se utilizó el cuestionario *Coaches' Interpersonal Style questionnaire* (CIS-Q; Pulido, Sánchez-Oliva, Leo, Sánchez-Cano, y García-Calvo, 2017). La versión utilizada fue adaptada por miembros del GEPE al contexto de la actividad física y va precedida por "Pienso que mi monitor/a..." y consta de 24 ítems en seis dimensiones con cuatro ítems cada una. Concretamente se trata de la dimensión de apoyo a la autonomía (e.g., " ...nos pregunta a menudo sobre nuestras preferencias con respecto a las actividades/tareas a realizar."), la de apoyo a la competencia (e.g., " ...nos anima a que confiemos en nuestra capacidad para hacer bien las tareas."), la de apoyo a la relación (e.g., " ...fomenta en todo momento las buenas relaciones entre los participantes del programa."), la de frustración a la autonomía (e.g., "... nos exige hacer las cosas de una manera concreta."), la de frustración a la competencia (e.g., "...nos propone tareas y situaciones que nos hacen sentir torpes.") y la de frustración a la relación (e.g., "... hace que nos sintamos poco aceptados/as en este grupo."). La puntuación de las escalas es directa: una puntuación elevada en el promedio o suma de los ítems de una dimensión indica una gran posesión de ese estilo interpersonal.

Satisfacción de las necesidades psicológicas básicas. Se utilizó la versión española de *Basic Psychological Need Exercise Scale* (BPNES; Sanchez y Nuñez, 2007; Vlachopoulos y Michailidou, 2006). La escala va precedida de "En las sesiones de Actividad Física..." y está compuesta por 12 ítems que miden la satisfacción de la de autonomía (4 ítems; e.g., "...los ejercicios que realizo se ajustan a mis intereses."),

la satisfacción de la competencia (4 ítems; e.g., "...siento que he tenido una gran progresión.") y la satisfacción de la relación (4 ítems; e.g., "...me siento muy cómodo/a con los compañeros/as."). Obtener una puntuación alta en esta escala es indicativo de percibir una alta satisfacción de las necesidades psicológicas básicas.

Regulación conductual. Se utilizó la versión española del *Behavioral Regulation in Exercise Questionnaire* (BREQ-3; González-Cutre, et al., 2010; Markland y Tobin, 2004; Wilson, et al., 2006). El cuestionario tiene 24 ítems agrupados en seis subescalas y va precedido por "Participo en las sesiones de Actividad Física...". La primera subescala hace referencia a la regulación intrínseca (4 ítems; e.g., "...porque creo que el ejercicio es divertido."), la segunda hace referencia a la regulación integrada (4 ítems; e.g., "...porque considero que el ejercicio forma parte de mí."), la tercera es la regulación identificada (4 ítems; e.g., "...porque valoro los beneficios que tiene el ejercicio físico."), la cuarta es la regulación introyectada (4 ítems; e.g., "...porque me siento avergonzado si falto a la sesión."), la quinta es regulación externa (4 ítems; e.g., "...porque los demás dicen que debo hacerlo.") y finalmente amotivación (4 ítems; e.g., "...pero no veo el sentido de hacer ejercicio."). La obtención de la puntuación de las subescalas es directa.

Vitalidad subjetiva. Se utilizaron cinco ítems de la versión española de la escala unidimensional *Subjective Vitality Scale* (SVS; Castillo, Tomás, y Balaguer, 2017; Ryan y Frederick, 1997). Se excluyó un ítem, "me siento tan vivo que quiero saltar" de la versión de seis

ítems ya que mediante un panel de personas expertas fue considerado como inadecuado para las personas mayores. Los ítems estaban precedidos por “Generalmente me siento...”. Una puntuación alta en esta escala, obtenida mediante la suma o promedio, indica una elevada percepción de vitalidad subjetiva.

Procedimiento

Tal y como se ha comentado en el Capítulo 1, esta tesis está enmarcada dentro del proyecto ACTIVATICS. Como parte del proyecto ACTIVATICS se contó con la aprobación de los procedimientos por parte de la *Comissió d'Ètica en l'Experimentació Animal i Humana* (CEEAH) de la UAB.

En todas las fases de esta investigación se ha informado a las personas participantes de los objetivos e implicaciones de participar en este estudio, así como de sus derechos en relación con la confidencialidad y la voluntariedad de su participación. Previo a las recogidas de datos con personas mayores se facilitó un consentimiento informado que las participantes devolvían firmado antes de contestar cualquier tipo de cuestionario o realizar las entrevistas. Los documentos para proporcionar el consentimiento informado pueden encontrarse en el Apéndice A1.

Al inicio del estudio, tal como se ha dicho en el Capítulo 1, con la finalidad de disponer de medidas comparables entre diversos grupos se realizó un estudio piloto para determinar la factibilidad de una recogida de datos con las medidas de múltiples ítems. El proceso duró en torno a 30 minutos, dado que las dos personas investigadoras dieron por finalizada la obtención de datos ya que no era posible obtener respuestas por parte de las personas participantes. Las barreras eran variadas: dificultades por falta de corrección de la visión, comprensión del texto, ...

La creación de las medidas de un solo ítem se desarrolló en cuatro fases. En primer lugar (Versión 1) redacté una primera versión de las medidas de un solo ítem atendiendo a las recomendaciones clásicas en la reducción de escalas, a los resultados de los grupos focales que se habían obtenido en fases previas del proyecto ACTIVATICS (Pallarès et al., en prensa), y teniendo en cuenta la formulación escalas ya existentes construidas para medir los mismos constructos (Véase Capítulo 1 de esta tesis). En segundo lugar, se propuso a un grupo de seis personas expertas que elaboraran un nuevo ítem a partir de lo que ellas creían que debía representar cada constructo y en una misma sesión, se pidió a las mismas personas que dieran argumentos en base a la relevancia, la suficiencia y la claridad de los ítems de la Versión 1. La conciliación de estas dos evaluaciones dio lugar a la Versión 2. A partir de entrevistas cognitivas a cuatro personas mayores de sesenta años se desarrolló la Versión 3. En la cuarta fase un nuevo grupo de cuatro especialistas revisó su relevancia, suficiencia y claridad. Cuando los ítems se consideraron totalmente satisfactorios, se tradujeron al idioma catalán por parte de dos personas, una de ellas la autora de esta tesis. Tal y como se verá en el Capítulo 4 en el apartado de desarrollo de las medidas de un solo ítem, 12 ítems resultaron seleccionados de los preexistentes y cuatro ítems fueron de nueva redacción.

Con la finalidad de obtener respuestas de las medidas de un solo ítem y de las medidas de múltiples ítems (Muestra 1) se concertaron entrevistas con la administración de los diversos centros

implicados en la recogida de datos para heteroadministrar los cuestionarios. Las medidas de múltiples ítems se heteroadministraron y en la medida de lo posible se intentó que las personas participantes se autoadministraran los cuestionarios reducidos. En todos los centros se ofreció un certificado de participación a las personas que colaboraron. Además, en uno de los centros también se ofreció bebida y un obsequio por parte de la administración del centro.

Para la Muestra 1, siempre que las capacidades tecnológicas de los centros lo permitían, las personas que realizaban la entrevista recogían los datos mediante un ordenador portátil con acceso a internet mediante *Google Forms*. En caso contrario lo hacían rellenando un cuadernillo de cuestionarios en papel. Todas las medidas de un solo ítem que fueron contestadas por personas mayores (Muestra 1 y Muestra 2) fueron contestadas en papel.

Para la segunda recogida de datos (Muestra 2) se contactó con otros centros diferentes a los de la Muestra 1 para maximizar la independencia de las fuentes. Se realizaron dos recogidas de datos con un intervalo de tiempo de entre tres y cuatro semanas entre ellas (*test* y *retest*) en las que participaron un total de 100 personas. En este trabajo solo se presentan los datos de las personas que estuvieron presentes en las dos ocasiones ($n = 62$). Un total de 27 personas solo estuvieron presentes en la fase *test*, y un total de 11 personas en la fase *retest* por lo que se excluyeron del estudio de la estabilidad temporal.

Como retorno por su participación, se ofreció un informe a todos los centros u organizaciones que participaron en la investigación con estadísticos descriptivos del estudio en el centro.

Análisis de datos

Todos los análisis estadísticos se han realizado con el *software* R y la hoja de cálculo Microsoft Excel™. El análisis descriptivo y preparatorio de los datos en las distintas fases de este se ha realizado mediante el paquete *psych* (Revelle, 2018) y con el paquete *SubscaleExplorer* (Apéndice B; Apéndice C1; Angulo-Brunet y Viladrich, 2017a). Los gráficos descriptivos se han realizado con el paquete *SubscaleExplorer*, con *ggplot2* (Wickham, 2016) y con el paquete *corrplot* (Wei y Simko, 2017). Para los análisis se ha tenido en cuenta que menos de un 5% de datos perdidos no tienen consecuencias graves en sesgo o potencia (Graham, 2009).

Para examinar las relaciones entre las variables de los diferentes cuestionarios se ha realizado un *correlational network analysis* (Epskamp, Borsboom, y Fried, 2018) con la finalidad de facilitar la interpretación de la tabla de correlaciones tetracóricas utilizando el paquete *qgraph* (Epskamp, Angelique, Cramer, Waldorp, y Chmittman, 2012). Para interpretar las correlaciones se ha tenido en cuenta el criterio clásico de Cohen (1988) considerando correlaciones inferiores a .10 nulas, correlaciones de .11 a .29 débiles, correlaciones de .30 a .49 moderadas y correlaciones mayores a .50 fuertes.

Al crear las tablas de contingencia de los ítems dos a dos para calcular los coeficientes de correlación policórica es relativamente frecuente que se encuentren celdas vacías. Puesto que ello puede comprometer la estimación de los coeficientes de correlación policórica, algunos autores (e.g., Liu et al. 2017) recomiendan colapsar

categorías hasta obtener tablas de contingencia sin celdas con frecuencias igual a cero (e.g., Rey, Abad, Barrada, Garrido, y Ponsoda, 2014; Penelo, De la Osa, Navarro, Domènech, y Ezpeleta, 2017). Considerando el proceso de respuesta y viendo la distribución de las respuestas, en detalle en el Capítulo 4, se dicotomizó la escala de respuesta original ($1 = \text{Totalmente en desacuerdo}$; $5 = \text{Totalmente de acuerdo}$) a una escala de respuestas extremas teniendo en cuenta la direccionalidad de las subescalas siguiendo el procedimiento sugerido por Jöreskog (2005). Por ejemplo, en la escala dicotómica de los ítems con contenido positivo, como pueden ser los de soporte a la autonomía, se ha asignado el valor “1” a la respuesta “Totalmente de acuerdo” y el valor “0” al resto de categorías de respuesta (0,0,0,0,1). Por el contrario, en los ítems de contenido negativo, como pueden ser los de amotivación, se ha asignado el valor “0” a la respuesta “Totalmente en desacuerdo” y el valor 1 al resto de categorías de respuesta (1,0,0,0,0).

El ajuste de los modelos de medida para las medidas de múltiples ítems se realizó según las recomendaciones de Viladrich, et al. (2017). Teniendo en cuenta que los datos eran medidas de dos categorías ordenadas y el tamaño de la muestra era pequeño ($n = 128$) se utilizaron correlaciones tetracóricas y el estimador de mínimos cuadrados no ponderados ajustado por media y varianza (ULSMV) para realizar el análisis factorial confirmatorio (CFA) mediante el paquete *lavaan* (Rosseel, 2012). Puesto que actualmente no existe ningún punto de corte satisfactorio para los índices de bondad de ajuste derivados

de un estimador para datos categóricos (Xia y Yang, 2018), se ha considerado que, como mínimo, debían cumplirse los criterios clásicos de bondad de ajuste: se consideró una bondad de ajuste excelente (Hu y Bentler, 1999) valores de CFI (*Comparative fit index*) y TLI (*Tucker-Lewis index*) mayores de .95 y de RMSEA (*root mean error of approximation*) menores de .06, mientras que valores de CFI y TLI mayores que valores mayores a .90 e inferiores a .08 en RMSEA fueron considerados aceptables (Marsh, Hau, y Wen, 2004).

De forma coherente con el modelo de medida se ha utilizado el coeficiente de fiabilidad no lineal (ω_c , omega categórica; Green y Yang, 2009; Yang y Xia, 2019), obtenido mediante la función *reliability* del paquete *semTools* (Jorgensen, Pornprasertmanit, Schoemann, y Rosseel, 2018). También se ha calculado el coeficiente alfa de Cronbach (α), obtenido mediante la función *alpha* del paquete *psych* (Revelle, 2018) así como su intervalo de confianza. Este coeficiente es equivalente al coeficiente de Kuder-Richardson [KR-20] cuando los datos son dicotómicos (Nunnally, 1978) y se incluye como estimación de la fiabilidad para proveer datos de consistencia interna comparables con investigaciones previas que han utilizado estos cuestionarios. Siguiendo el criterio de Nunnally (1978) valores mayores a .70 pueden verse suficientes en la fase de desarrollo de los cuestionarios. Después de analizar los modelos de medida y la fiabilidad de los cuestionarios se calculó la puntuación total de las subescalas por sumatorio de los ítems, así como sus estadísticos descriptivos.

Para evaluar la validez de las medidas de un solo ítem con evidencias cuantitativas, en primer lugar se consideró la correlación biserial entre las medidas de un solo ítem y las medidas de múltiples ítems, obtenida con la función *biserial* del paquete *psych* (Revelle, 2018). En segundo lugar, se consideró la correlación tetracórica entre las medidas de un solo ítem. Por ello, se obtuvo mediante el paquete *corrplot* (Wei y Simko, 2017) un correlograma para las medidas de un solo ítem de la Muestra 1, un correlograma para las medidas de un solo ítem de la fase test de la Muestra 2 y un correlograma para la fase retest de la Muestra 2.

Para evaluar la fiabilidad de las medidas de un solo ítem se tuvieron en cuenta los diferentes coeficientes de fiabilidad explicados en el Capítulo 1 de esta tesis. Para todos los cálculos que requerían estimación a partir de la fiabilidad de medida de múltiples ítems se utilizó el coeficiente omega categórico obtenido con la misma muestra. En primer lugar, se aplicó la profecía de Spearman-Brown. En segundo lugar, se aplicó la fórmula de corrección por atenuación. También se obtuvieron las comunalidades de los ítems a partir de un CFA. En el caso de las medidas de un solo ítem que habían sido elegidas entre los ítems que ya existían en los cuestionarios se utilizó la comunalidad del ítem dentro del cuestionario largo. Para los ítems que fueron desarrollados de nuevo (i.e., soporte a la autonomía y vitalidad subjetiva) se realizaron CFA unidimensionales incluyendo el nuevo ítem junto con los preexistentes, se evaluó de nuevo la bondad de ajuste, y finalmente

se utilizó la comunalidad obtenida como fiabilidad de la medida de un solo ítem.

Finalmente evaluar la fiabilidad test-retest en la Muestra 2 se consideró, además del porcentaje de acuerdo entre el *test* y el *retest* el coeficiente PABAK (*Prevalence - adjusted bias - adjusted kappa*; Byrt, Bishop, y Carlin, 1993). Siguiendo las recomendaciones de Landis y Koch (1977) se consideró acuerdo bajo el rango de .21 a .40, acuerdo moderado de .41 a .60, acuerdo bueno el rango de .61 a .80 y acuerdo excelente valores mayores a .81.

4

RESULTADOS

Propiedades psicométricas de las medidas de múltiples ítems

Previo al estudio de la estructura y consistencia interna de las medidas con múltiples ítems se realizó un análisis preparatorio de datos. Se examinaron las distribuciones de respuesta antes y después de dicotomizar los datos y se analizaron las correlaciones tetracóricas entre las variables de un mismo cuestionario. Finalmente se realizó un *correlational network analysis* para examinar las relaciones entre todas las variables involucradas en la cadena motivacional de la SDT.

Análisis preparatorio de datos

En la Figura 2 se encuentra la distribución de respuestas para los ítems de Estilo interpersonal, representados ordenadamente agrupados en las seis subescalas de apoyo y frustración a las necesidades psicológicas básicas, identificados con su nombre psicométrico seguido de un número tal como constan también en el Apéndice A3. El 4% de las celdas son valores perdidos y no están

representados en la tabla. A la izquierda se encuentra la distribución de respuestas inicial en la que se observa, con

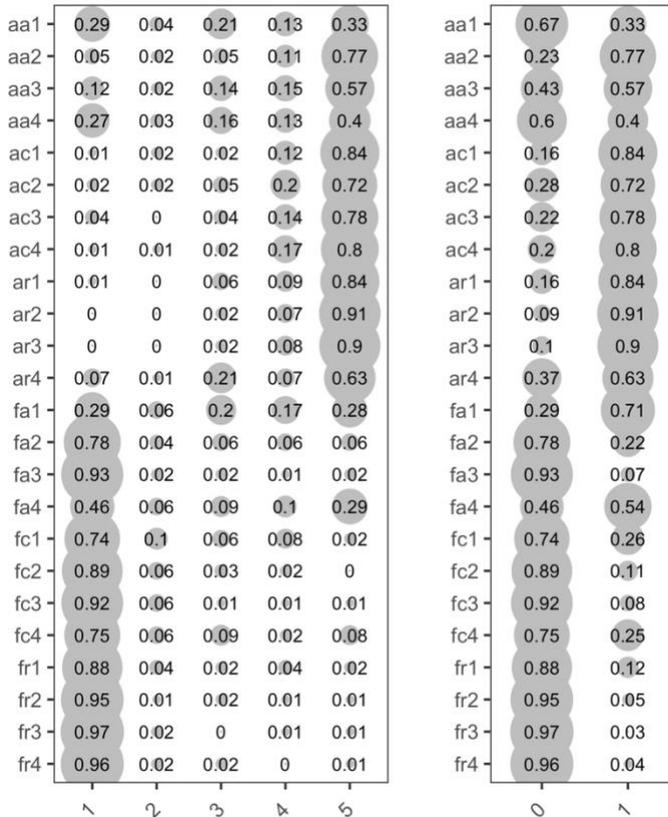


Figura 2 Distribución de Respuestas de Estilo Interpersonal

Nota. A la izquierda, proporción de respuestas en cada categoría de los datos originales (1 = Totalmente de acuerdo; 5 = Totalmente en desacuerdo). A la derecha, proporción de respuestas después de dicotomizar a respuestas extremas (0 = En desacuerdo; 1 = De acuerdo). aa = apoyo a la autonomía; ac = apoyo a la competencia; ar = apoyo a la relación; fa = frustración a la autonomía; fc = frustración a la competencia; fr = frustración a la relación.

algunas excepciones, un marcado efecto techo o suelo con más del 55% de las respuestas en una misma categoría de respuesta. Totalmente de acuerdo en los ítems con contenido positivo, y totalmente en desacuerdo en los ítems con contenido negativo. A la derecha se encuentran las mismas respuestas después de dicotomizar la escala a respuestas extremas. De los 24 ítems del cuestionario de Estilo interpersonal cuatro ítems relacionados con el concepto de autonomía no presentaron efecto techo o suelo. Concretamente dos ítems de apoyo a la autonomía (i.e., aa1[“...nos pregunta a menudo sobre nuestras preferencias con respecto a las actividades a realizar”] y aa4[“... nos deja tomar decisiones durante el desarrollo de las tareas.”]) y dos ítems de frustración a la autonomía (i.e., fa1[“...nos exige hacer las cosas de una forma concreta”] y fa4[“... nos fuerza a seguir una forma determinada de hacer los ejercicios]). Estos cuatro ítems presentaron una proporción de respuestas más equilibrada entre las categorías. Por lo tanto, podemos decir que las personas que han contestado los cuestionarios perciben que su monitor o monitora apoya y no frustra las necesidades psicológicas básicas de competencia y relación mientras que la percepción de apoyo y frustración a la necesidad psicológica de autonomía presenta más variabilidad.

En la Figura 3 están representados los coeficientes de correlación tetracórica entre los ítems del cuestionario de Estilo interpersonal. Los ítems se presentan de nuevo ordenadamente agrupados en las seis subescalas de los estilos de apoyo y frustración

de las necesidades psicológicas básicas. Tal y como puede observarse en la leyenda situada en la parte inferior del correlograma, el color azul más oscuro representa coeficientes de correlación cercanos a 1, el color rojo más oscuro representa coeficientes de correlación cercanos a -1 y los colores más claros coeficientes cercanos a 0.

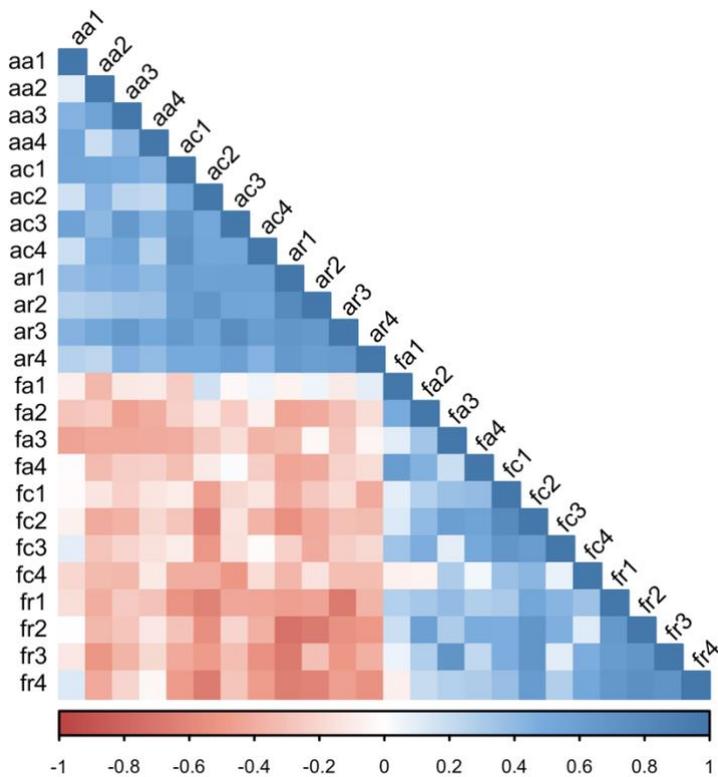


Figura 3 Correlograma de Estilo Interpersonal

Nota. aa = apoyo a la autonomía; ac = apoyo a la competencia; ar = apoyo a la relación; fa = frustración a la autonomía; fc = frustración a la competencia; fr = frustración a la relación.

A grandes rasgos, la figura muestra los resultados esperados. Las dos zonas de coloración azul en el gráfico reflejan las relaciones

positivas entre los ítems de apoyo a las necesidades psicológicas básicas y de los ítems que miden frustración a las necesidades psicológicas básicas entre sí. La zona de coloración roja refleja correlaciones negativas entre ambos grupos.

Dentro de cada grupo de ítems las cosas no resultan tan claras. En referencia a los ítems de apoyo a las necesidades psicológicas básicas, las correlaciones entre los ítems de apoyo a la autonomía son entre débiles y fuertes (rango .11 a .57), siendo en algunos casos más elevadas con ítems que no forman parte de esta subescala sino de las de apoyo a la competencia y apoyo a la relación. Se establecen correlaciones altas en los ítems de apoyo a la competencia (rango .52 a .74) y apoyo a la relación (rango .61 a .81). Cabe destacar que algunas de estas correlaciones son muy elevadas como por ejemplo ar1 (“...fomenta en todo momento las buenas relaciones entre los participantes del programa.”) con ar2 (“...favorece el buen ambiente entre los compañeros/as del grupo.”) con un valor de .81 mostrando redundancia entre los pares de ítems. Este también es el caso entre ítems que no forman parte de la misma dimensión teórica, como por ejemplo ac3 (“...siempre intenta que consigamos los objetivos que se plantean en las actividades.”) con ar3 (“...promueve que todos los participantes nos sintamos integrados.”) con un coeficiente de correlación tetracórica igual a .78. En los ítems de frustración a las necesidades psicológicas básicas sucede un fenómeno similar.

Las correlaciones entre los ítems de frustración a la autonomía son poco homogéneas con valores entre .10 y .62. En cambio, los ítems

de frustración a la competencia están todos fuertemente correlacionados entre ellos (rango .63 a .79), excepto el ítem fc4 (“...no nos da oportunidades para demostrar nuestro potencial”) que tiene correlaciones entre .07 y .36 con los ítems de la subescala. Los ítems de frustración a la relación forman un claro grupo de ítems diferenciado con correlaciones entre sí que toman valores entre .64 y .75. Por lo tanto, tal y como se puede observar en la Figura 3 mientras que los ítems de apoyo a las necesidades psicológicas básicas forman una agrupación de ítems correlacionados positivamente entre ellos representada por una intensa coloración azul, esta agrupación no es tan evidente con los ítems de frustración a las necesidades psicológicas básicas donde se observan muchos matices en la coloración.

En cuanto a la satisfacción de las necesidades psicológicas básicas, el 5% de las celdas son valores perdidos. A la izquierda de la Figura 4, se encuentra la distribución de respuestas original a los ítems de satisfacción de las necesidades psicológicas básicas. A la derecha la distribución de respuestas después de dicotomizar valores extremos. Todos los ítems tienen un marcado efecto techo, con más del 59% de las respuestas siendo totalmente de acuerdo, a excepción del ítem sa4 (i.e., “tengo la oportunidad de elegir cómo realizar los ejercicios.”) con un porcentaje un poco menor. Se puede decir que las personas que han participado en el estudio ven sus necesidades psicológicas básicas satisfechas.

Como se puede observar el correlograma de la Figura 5 solo existen relaciones positivas, representadas por el color azul, entre los ítems de satisfacción de las necesidades psicológicas básicas. En los

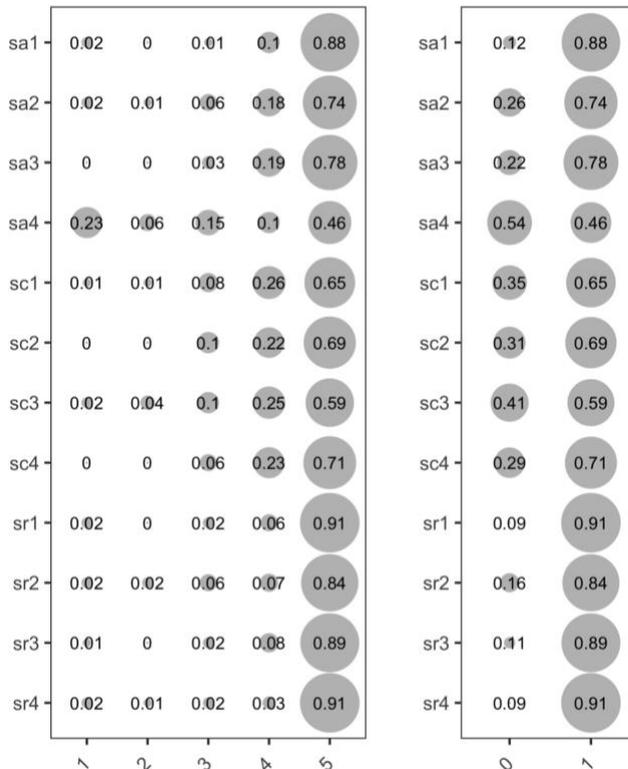


Figura 4 Distribución de Respuestas de Satisfacción de las Necesidades Psicológicas Básicas

Nota. A la izquierda, proporción de respuestas en cada categoría de los datos originales (1 = Totalmente de acuerdo; 5 = Totalmente en desacuerdo). A la derecha, proporción de respuestas después de dicotomizar a respuestas extremas (0 = En desacuerdo; 1 = De acuerdo). sa = satisfacción de la autonomía; sc = satisfacción de la competencia; sr = satisfacción de la relación.

ítems de satisfacción de la autonomía se observan correlaciones elevadas (rango .61 a .83) excepto el ítem sa4 (“... tengo la oportunidad de elegir cómo realizar los ejercicios.”) que tiene una correlación

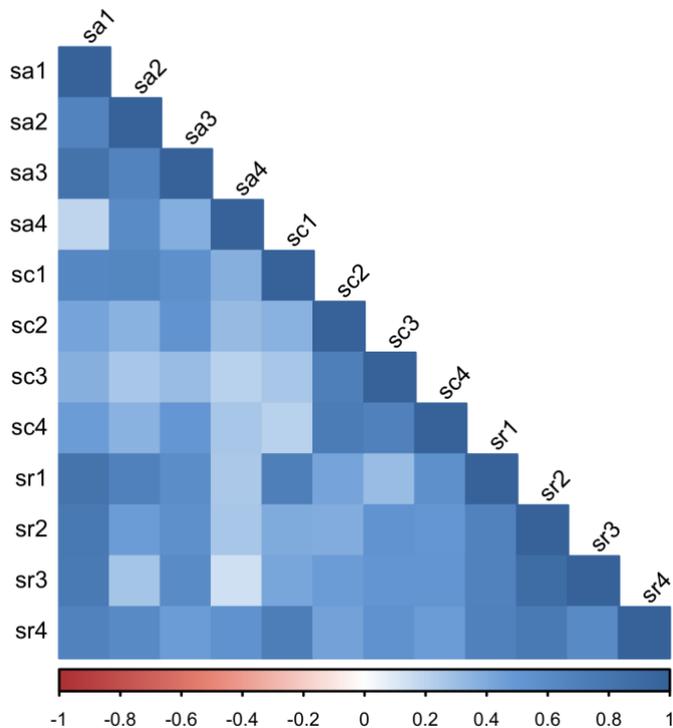


Figura 5 Correlograma de Satisfacción de las Necesidades Psicológicas Básicas

Nota. sa = satisfacción de la autonomía; sc = satisfacción de la competencia; sr = satisfacción de la relación.

tetracórica de .19 con sa1 (“... los ejercicios que realizo se ajustan a mis intereses.”) y una correlación de .39 con sa3 (“... la forma de realizar los ejercicios responde a mis deseos.”). El ítem sa1 presenta correlaciones más elevadas con ítems de satisfacción de la relación (rango .68 a .82)

que con ítems de la subescala de satisfacción de la autonomía. Si observamos los ítems de la dimensión teórica satisfacción de la competencia se observan altas correlaciones entre ítems (rango .70 a .75) a excepción de el ítem sc1 (i.e., "...siento que he tenido una gran progresión.") con correlaciones bajas y moderadas con el resto de los ítems (rango .21 a .36). En el caso de los ítems de satisfacción de la relación se puede observar una clara agrupación (rango .62 a .89).

Las respuestas al cuestionario de regulación conductual mostraron datos perdidos en el 1% de las celdas. Tal y como puede observarse a la izquierda de la Figura 6 las respuestas sugieren un claro efecto techo o suelo, con más del 55% de las respuestas en una categoría extrema, a excepción de los ítems de la dimensión teórica de regulación introyectada y ide4["...me siento inquieto/a si no hago ejercicio regularmente"]. El efecto suelo es especialmente relevante en los ítems de la dimensión de amotivación donde más del 94% de las respuestas han sido Totalmente en desacuerdo. Este efecto también es muy marcado en regulación identificada donde excepto el ítem ide4, los otros ítems han presentado más del 92% de las respuestas en Totalmente de acuerdo. Por lo tanto, las personas que han contestado los cuestionarios dicen que practican actividad física porque disfrutan de la actividad, porque es importante para ellas y porque es congruente con sus valores, es decir autónomamente. Además en general no están reguladas por contingencias externas y que son personas que no están desmotivadas con la actividad que están realizando.

Las correlaciones tetracóricas entre los ítems del cuestionario con ítems en el sentido esperado, siendo algunas agrupaciones más claras que otras. Se puede observar una clara relación positiva, representada por la coloración azul fuerte y homogénea en la zona

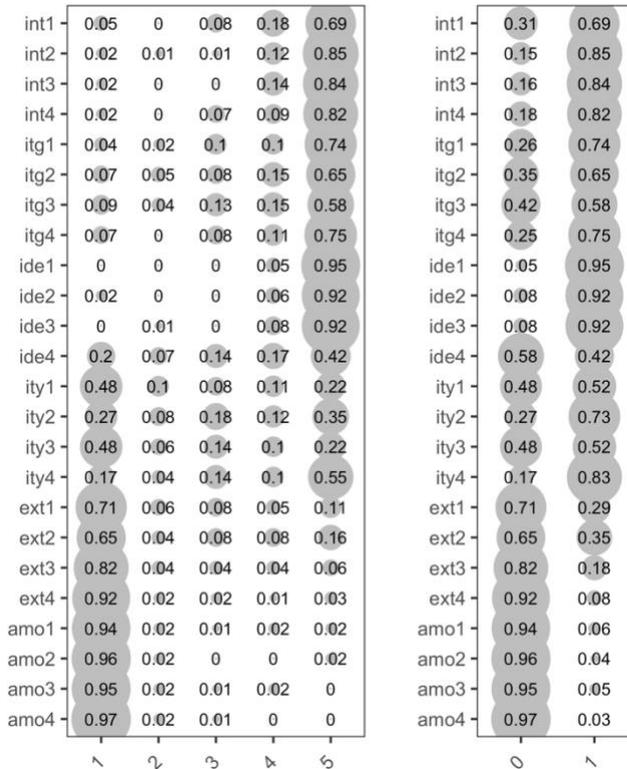


Figura 6 Distribución de Respuestas de Regulación Conductual

Nota. A la izquierda, proporción de respuestas en cada categoría de los datos originales (1 = Totalmente de acuerdo; 5 = Totalmente en desacuerdo). A la derecha, proporción de respuestas después de dicotomizar a respuestas extremas (0 = En desacuerdo; 1 = De acuerdo). int = regulación intrínseca; itg = regulación integrada; ide = regulación identificada; ity = regulación introyectada; ext = regulación extrínseca; amo = amotivación.

correspondiente del gráfico, entre los ítems que conforman la dimensión teórica de motivación autónoma: regulación intrínseca, regulación integrada y regulación identificada. De la misma forma puede observarse también una relación positiva, aunque más débil, entre los ítems de la dimensión teórica de motivación controlada: regulación introyectada y regulación externa. También se observa claramente la relación inversa entre las regulaciones extremas del continuo teórico de regulación reflejada en la coloración roja de la zona del gráfico donde se representa el cruce entre ambas regulaciones extremas.

A la vez, también se observan en general relaciones más intensas entre los ítems de cada regulación, aunque también se dan correlaciones elevadas entre ítems que no son de la misma dimensión. Los ítems de regulación intrínseca presentan entre si valores de correlación entre .57 y .80, los ítems de regulación integrada por su parte presentan valores entre .47 y .81, los de regulación identificada valores entre .37 y .71, los de regulación introyectada valores entre -.25 y .68, los de regulación externa valores entre .67 y .86 y finalmente los ítems de amotivación presentan valores entre .32 y .82.

Finalmente, la distribución de respuestas de los ítems de vitalidad subjetiva se encuentra en la Figura 8. Se encontró menos del 1% de celdas con valores perdidos. Como puede observarse en la figura más del 60% de las personas estuvieron totalmente de acuerdo con los reactivos, indicando en general, una alta vitalidad subjetiva.

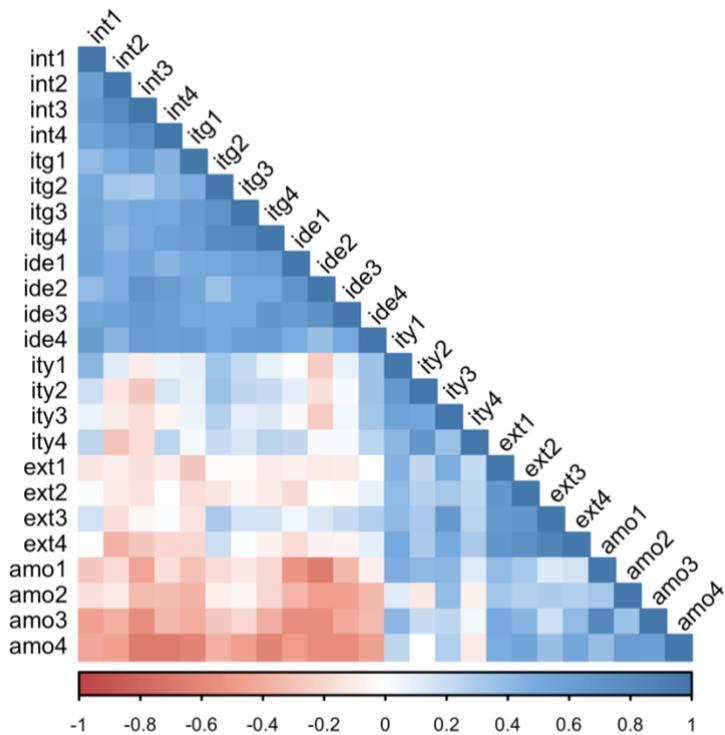


Figura 7 Correlograma de Regulación Conductual

Nota. int = regulación intrínseca; itg = regulación integrada; ide = regulación identificada; ity = regulación introyectada; ext = regulación extrínseca; amo = amotivación.

Como se puede observar en la Figura 9, representado por el color azul fuerte, todas las correlaciones fueron muy elevadas (rango .79 a .95) indicando posible redundancia entre los ítems.

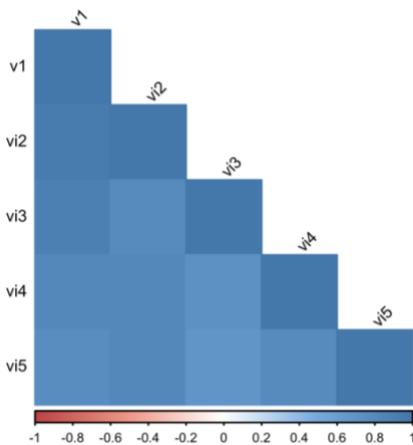


Figura 8 Correlograma de Respuestas de Vitalidad Subjetiva

Nota. vi = vitalidad subjetiva.

Con la finalidad de estudiar las relaciones entre los ítems de distintos cuestionarios se llevo a cabo un *correlation network analysis* (Figura 10) incluyendo todos los ítems de las distintas medidas. En este gráfico los ítems están representados por círculos a los que se les ha asignado color y posición ordenadamente por subescalas. Para facilitar la interpretación se ha fijado la posición de los ítems de manera que empezando por arriba y siguiendo el sentido de las agujas de un reloj reproduce la cadena de la SDT tal y como se ha presentado en la Figura 1 de este trabajo. En primer lugar, se encuentran los ítems de estilo interpersonal, siguen en la parte baja del gráfico los ítems de satisfacción de las necesidades psicológicas básicas, y el círculo se cierra con los ítems de regulación conductual y de vitalidad.

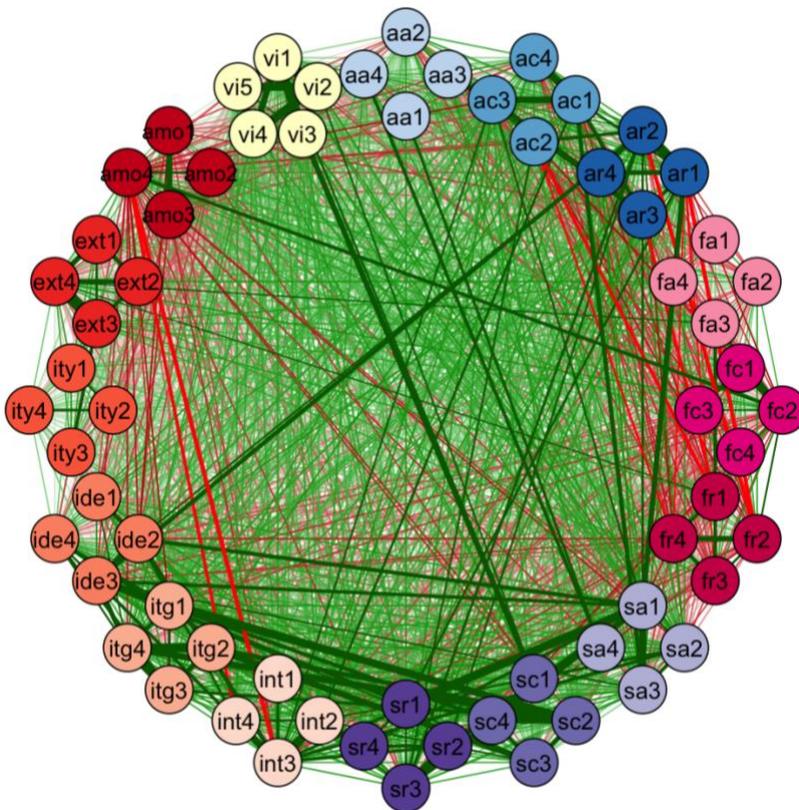


Figura 9 Correlational Network Analysis

Nota. aa = apoyo a la autonomía; ac = apoyo a la competencia; ar = apoyo a la relación; fa = frustración a la autonomía; fc = frustración a la competencia; fr = frustración a la relación; sa = satisfacción autonomía; sc = satisfacción competencia; sr = satisfacción relación; int = regulación intrínseca; itg = regulación integrada; ide = regulación identificada; ity = regulación introyectada; ext = regulación extrínseca; amo = amotivación; vi = vitalidad subjetiva.

Las líneas que conectan los diferentes ítems representan las correlaciones tetracóricas entre ellos. Las líneas de color verde representan correlaciones positivas y las líneas de color rojo

representan correlaciones negativas. Su grosor indica la intensidad de la correlación: a más grosor más elevada es la correlación.

Si examinamos los clústeres formados por los ítems de un mismo cuestionario, tal y como es de esperar ya que se analizan de forma visual las correlaciones tetracóricas, se observan las agrupaciones ya comentadas en los correlogramas. De forma general, y como sería de esperar, los ítems que forman parte de las escalas con contenido positivo correlacionan positivamente entre si. Ello se refleja en la coloración verde de las líneas que unen los ítems de apoyo de las necesidades psicológicas básicas, los de satisfacción de dichas necesidades, los de motivación autónoma y los de vitalidad. También de forma consistente con lo esperado las escalas de frustración de las necesidades psicológicas básicas, y las de motivación controlada y la de amotivación correlacionan positivamente (en verde) entre si y negativamente (en rojo) con el resto de las escalas. La observación más pormenorizada pone de manifiesto que varios ítems presentan correlaciones más fuertes con ítems de otros cuestionarios que con los propios ítems de la escala de la que forman parte.

En relación con los ítems de apoyo a las necesidades psicológicas básicas se establecen relaciones puntuales con algunos ítems de satisfacción de las necesidades psicológicas básicas como por ejemplo aa1 con sc1 (.58) o aa4 con sc1 (.60). También se establecen correlaciones intensas entre los ítems de apoyo a la competencia con los ítems de satisfacción de la autonomía y satisfacción a la relación con correlaciones positivas alrededor de .50 a excepción del ítem sa4 y sr4.

En referencia a los ítems de frustración a las necesidades psicológicas básicas no se establecen relaciones muy marcadas con ítems de otros cuestionarios. Hay dos excepciones: el ítem fc2 correlaciona de forma negativa con el ítem sc4(-.51) y el ítem fr4 correlaciona negativamente con sa1(-.51).

Los ítems de satisfacción de las necesidades psicológicas básicas presentan correlaciones entre moderadas y altas (rango .30 a .69) con las escalas relacionadas con la motivación autónoma (i.e., regulación intrínseca, regulación integrada y regulación identificada) siendo una excepción el ítem sa4 que tampoco mantiene relación con los ítems de su cuestionario. Si consideramos la relación de los ítems de satisfacción de las necesidades psicológicas básicas con vitalidad subjetiva las correlaciones son entre bajas y moderadas (rango .11 a .54) a excepción del ítem sc1 con todos los ítems de vitalidad (rango .46 a .65).

En cuanto a las regulaciones conductuales además de las correlaciones ya comentadas cabe destacar que los ítems de amo4 tienen correlaciones negativas entre moderadas y altas con los ítems de apoyo a las necesidades psicológicas básicas, satisfacción de las necesidades psicológicas básicas y motivación autónoma. En cambio, para los ítems de motivación controlada no se establecen relaciones intensas (rango entre .001 a .39).

Es decir, aunque en general se cumplen las expectativas teóricas en cuanto al signo de los coeficientes de correlación entre ítems, también se observa que hay desigualdades muy visibles entre

ítems de una misma escala por lo que respecta a la intensidad de sus relaciones con los demás ítems.

Evidencias basadas en la estructura interna del test y la fiabilidad

Se probaron los distintos modelos compatibles con la SDT mediante CFA. En primer lugar, se comentan los modelos cuyos parámetros no pudieron ser estimados debido a la falta de variabilidad de los datos o a la colinealidad entre ítems para centrarnos después en los resultados de aquellos modelos que si pudieron ser evaluados.

Para el cuestionario de Estilo interpersonal se consideraron inicialmente los diversos modelos propuestos por los autores que desarrollaron el CIS-Q. Ninguno de los modelos probados mostró soluciones satisfactorias debido a los ítems de frustración a las necesidades psicológicas básicas: varianzas de los ítems negativas, soluciones no convergentes, ... Es por ello por lo que se decidió examinar por separado dos modelos distintos: por un lado, un modelo de tres factores correlacionados para el apoyo a las necesidades psicológicas básicas y por otro lado un modelo de tres factores correlacionados para la frustración a las necesidades psicológicas básicas.

El modelo de tres factores correlacionados para el apoyo a las necesidades psicológicas básicas se pudo ajustar sin problemas y se comenta más adelante, en cambio, el modelo de tres factores correlacionados para la frustración a las necesidades psicológicas básicas mostró varianzas negativas en dos ítems ($ct_2 = .21$; $rt_2 = -.05$)

llevando a una solución no admisible y por ello no se muestran los resultados de este modelo. En consecuencia, tampoco se obtendrán los índices de fiabilidad ni la puntuación mediante suma de estas variables.

Para estudiar la estructura interna el cuestionario de regulación conductual se consideraron diversos modelos compatibles con la teoría: un modelo de seis factores correlacionados, un modelo de tres factores (motivación autónoma, motivación controlada y amotivación) y un modelo de dos factores (motivación autónoma y motivación controlada). Los dos primeros modelos presentaron problemas de estimación: la matriz de covarianzas de las variables latentes no estaba definida positivamente y la varianza residual de amo4 era negativa (-.001 y -.017 respectivamente). El tercer modelo presentó índices de ajuste no adecuados ($\chi^2[df] = 203.87[169]$, $p = .03$, CFI = .90, TLI = .89, RMSEA[90%IC] = .04[.01-.06]) por lo que tampoco se presentan más resultados en relación con este cuestionario.

En la Tabla 2 se presentan los índices de bondad de ajuste para los modelos de estructura interna de cada cuestionario. Puede observarse que todos los modelos muestran índices adecuados de acuerdo con los criterios clásicos, aunque no hay que perder de vista que se están aplicados a datos categóricos.

En la Figura 11 se encuentran los parámetros estimados del modelo para el apoyo a las necesidades psicológicas básicas. Las cargas

Tabla 2 Bondad de Ajuste de las Medidas con Múltiples Ítems

Modelo/Subescala	χ^2	gl	P	CFI	TLI	RMSEA [90%IC]
Apoyo a las necesidades	56.13	51	.29	.99	.99	.03 [0.00-.07]
Satisfacción de las necesidades	77.49	51	.01	.93	.91	.07 [0.03-.09]
Vitalidad subjetiva	4.56	5	.47	1.00	1.00	0 [0.00-.11]

Nota. Apoyo a las necesidades = Apoyo a las necesidades psicológicas básicas; Satisfacción de las necesidades = Satisfacción a las necesidades psicológicas básicas; gl = grados de libertad, CFI = Comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean error of approximation.

factoriales de cada uno de los ítems en el factor previsto son elevadas estando todas en un rango entre .58 y .91. Sin embargo, las altas correlaciones entre los factores, en un rango entre .81 y .92, sugieren dificultades con la validez discriminante entre ellos.

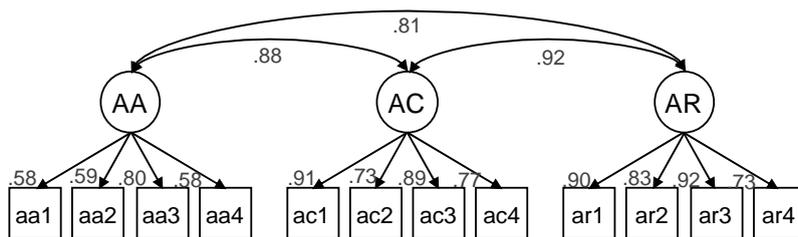


Figura 10 Modelo Factorial de Apoyo a las Necesidades Psicológicas Básicas

Nota. aa = apoyo a la autonomía; ac = apoyo a la competencia; ar = apoyo a la relación

La estructura interna del cuestionario de satisfacción de las necesidades psicológicas básicas (Figura 12) también está conformada por tres factores correlacionados. Igual que sucedió con el apoyo a las necesidades psicológicas básicas también existe falta de discriminación entre los factores, siendo las correlaciones entre los factores superiores a .78. Las cargas factoriales entre ítems de cada factor

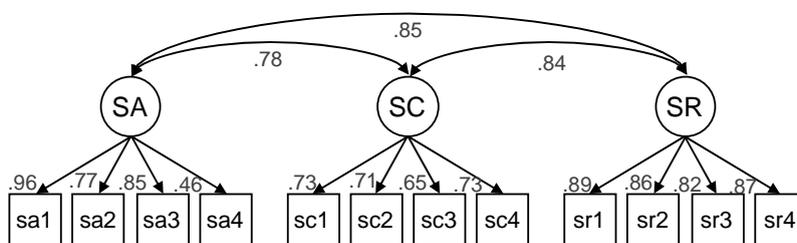


Figura 11 Modelo Factorial de Satisfacción de las Necesidades Psicológicas Básicas

Nota. sa = satisfacción de la autonomía; sc = satisfacción de la competencia; sr = satisfacción de la relación

fueron homogéneas, a excepción del ítem sa4, tal como se ha comentado, que ya mostraba bajas correlaciones con el resto de los ítems del factor y que tiene una carga de .46 en su factor mientras que las de otros ítems toman valores entre .77 y .96.

Finalmente, el modelo unifactorial de la escala de vitalidad subjetiva (Figura 13) presenta cargas factoriales muy elevadas y homogéneas.

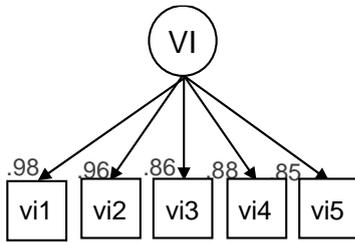


Figura 12 Modelo Factorial de Vitalidad Subjetiva

Nota. vi = vitalidad subjetiva.

En la Tabla 3 pueden encontrarse los estadísticos descriptivos así como los coeficientes de fiabilidad obtenidos para la suma de los ítems de cada una de las medidas de múltiples ítems. En negrita se ven destacadas las variables que han alcanzado el umbral mínimo de omega categórica de al menos .70 Los coeficientes de fiabilidad, resultaron adecuados para apoyo a la competencia, apoyo a la relación, satisfacción de la relación y vitalidad subjetiva.

Como puede observarse, los sumatorios fiables que son por un lado el apoyo a la competencia y el apoyo a la relación, con una media superior a tres y una mediana de cuatro y considerando los altos valores de asimetría y curtosis podemos decir que las personas que han contestado perciben un gran apoyo a la competencia y un gran apoyo a la relación.

También, con otro sumatorio fiable, la satisfacción de las necesidades psicológicas básicas, los estadísticos descriptivos obtenidos presentan también una media y una mediana muy elevadas, así como una asimetría y curtosis muy altas en valor absoluto. Estos resultados son un buen resumen de los que se comentaron en la Figura

4: las personas que han participado en el estudio tienen la percepción de que tienen la necesidad psicológica básica de relación satisfecha.

Tabla 3 Estadísticos Descriptivos para la Puntuación de las Medidas con Múltiples Ítems

Variable	M	DE	Mdn	rango	a	c	ω_c	α [95%IC]
Apoyo a la autonomía	2.06	1.25	2	0-4	-0.11	-1.08	.57	.57 [.45- .69]
Apoyo a la competencia	3.13	1.22	4	0-4	-1.42	0.99	.75	.74 [.66- .81]
Apoyo a la relación	3.28	1.09	4	0-4	-1.62	1.92	.74	.73 [.65- .81]
Satisfacción de la autonomía	2.85	1.21	3	0-4	-0.97	0.11	.67	.67 [.57- .76]
Satisfacción de la competencia	2.64	1.34	3	0-4	-0.53	-1.09	.67	.67 [.58- .76]
Satisfacción de la relación	3.55	0.97	4	0-4	-2.32	4.67	.74	.68 [.59- .76]
Vitalidad subjetiva	3.38	1.87	4	0-5	-0.81	-0.89	.87	.87 [.83- .90]

Nota. Mdn = Mediana; rango = rango de respuestas; a = asimetría; c = curtosis. ω_c = omega categórica; α = alfa de Cronbach; En negrita subescalas que han resultado satisfactorios en el estudio de la estructura interna y fiabilidad

Finalmente, con la variable vitalidad subjetiva, teniendo en cuenta que el rango de puntuaciones de la escala de vitalidad subjetiva es de entre 0 a 5 puntos, y presentando en esta muestra una media de 3.55 y una mediana de cuatro puntos, así como unos valores de asimetría y curtosis inferiores 1 en valor absoluto podemos decir que también resumen perfectamente los resultados presentados en la Figura 8: las personas que han contestado en general se sienten vitales.

Podemos resumir los resultados del estudio de la estructura interna y de la fiabilidad de consistencia interna de la siguiente forma. Aún partiendo de la poca variabilidad de las respuestas en este grupo de personas mayores, de los cuatro cuestionarios sometidos a análisis ha sido posible ajustar modelos de medida compatibles con la teoría en tres casos, con lo que se ha proporcionado evidencia de validez en relación con la estructura interna para siete de los 16 constructos teóricos sometidos a análisis. Son el apoyo a las tres necesidades psicológicas básicas, la satisfacción de las tres necesidades psicológicas básicas y la vitalidad subjetiva. Por otra parte, tal y como hemos visto, cuatro de las siete dimensiones de las que se han obtenido los coeficientes de consistencia interna tienen índices de fiabilidad adecuados.

Propiedades psicométricas de las medidas de un solo ítem

Desarrollo de las medidas de un solo ítem y evidencias de validez basadas en su contenido

En la Tabla 4 se encuentran las características de cada una de las tres versiones del cuestionario que se obtuvieron en el proceso de desarrollo de las medidas de un solo ítem. La tabla presenta el desarrollo de cada uno de los ítems. En negrita está remarcada la versión final de cada medida de un solo ítem y también se encuentra en cursiva su traducción al catalán.

Tal como es usual en investigación cualitativa, a continuación, se comentan los resultados mediante ejemplos detallados de los casos más relevantes. Por ejemplo, dos ítems resultaron satisfactorios en la Versión 1: regulación identificada (IDE) y regulación externa (EXT) sin introducir ningún cambio posterior. Un ítem, apoyo a la autonomía (AA) sufrió cambios menores en la Versión 2 con la finalidad de simplificarlo para adaptarse mejor a la población a la que iba dirigido y finalmente a partir de las entrevistas cognitivas se volvió a ajustar (Versión3). Los cambios afectaron principalmente a la forma de explicar las posibles aportaciones de las personas en las actividades realizadas durante la sesión de actividad física. Un ítem, satisfacción de la relación (SR), fue ajustado gracias al panel de personas expertas para representar mejor el constructo detallando que las demás personas a las que se refiere el ítem son los compañeros y compañeras y el monitor o la monitora.

Un total de cinco ítems tuvieron que ser ajustados en la segunda versión ya que en un inicio eran la adición de dos ítems de la versión de múltiples ítems del constructo. Este es el caso de frustración a la autonomía (FA), satisfacción de la competencia (SC) regulación intrínseca (INT) y regulación introyectada (ITY). Todos ellos se redactaron de manera que expresaran una sola idea. Además, en el caso de satisfacción de la competencia también se realizaron ajustes en la Versión 3 de acuerdo con las aportaciones de las personas destinatarias.

En otros casos, atendiendo al juicio de las personas expertas se decidió elegir un ítem diferente al de la Versión 1. Esto es el caso de tres ítems: apoyo a la competencia (AC), apoyo a la relación (AR) y amotivación (AMO).

Finalmente, también a partir del juicio de personas expertas, se desarrollaron cuatro ítems totalmente nuevos. Este es el caso de frustración a la competencia (FC), frustración a la relación (FR), satisfacción de la autonomía (SA) y vitalidad subjetiva (VI)

El panel de especialistas estuvo de acuerdo en que los ítems finalmente seleccionados eran relevantes y suficientes para medir los 16 constructos, así como suficientemente claros en su redacción teniendo en cuenta el grupo humano al que iban dirigidos.

Tabla 4 Desarrollo de las Medidas de un Solo Ítem

Medida	Versión 1	Versión 2	Versión 3
AA	<i>Pienso que el/la monitor/a tiene en cuenta nuestras preferencias y opiniones para realizar las actividades.</i>	<i>Pienso que el/la monitor/a tiene en cuenta nuestras opiniones en el desarrollo de las sesiones.</i>	<i>Pienso que el/la monitor/a tiene en cuenta nuestras aportaciones (opiniones, propuestas, ...) en el desarrollo de las sesiones.</i> [<i>Penso que el/la monitor/a té en compte les nostres aportacions (opinions, propostes,...) en el desenvolupament de les sessions.</i>]
AC	<i>Pienso que el/la monitor/a propone actividades ajustadas a nuestro nivel.</i>	<i>Pienso que el/la monitor/a nos anima a que confiemos en nuestra capacidad para hacer bien los ejercicios.</i> [<i>Penso que el/la monitor/a ens anima a que confiem en la nostra capacitat per fer bé els exercicis.</i>]	
AR	<i>Pienso que el/la monitor/a fomenta las buenas relaciones entre los compañeros y compañeras y nos ayuda a resolver los conflictos amistosamente.</i>	<i>Pienso que el/la monitor/a favorece el buen ambiente tanto entre los compañeros y compañeras de clase como con él/ella.</i> [<i>Penso que el/la monitor/a fomenta el bon ambient i les bones relacions tant entre els companys/es de classe com amb ell/a.</i>]	
FA	<i>Pienso que el/la monitor/a nos obliga a hacer las actividades de una</i>	<i>Pienso que el/la monitor/a nos exige hacer las cosas de una manera concreta.</i>	

Medida	Versión 1	Versión 2	Versión 3
	determinada manera y nos impide tomar decisiones.	[Penso que el/la monitor/a ens exigeix fer les coses d'una forma concreta.]	
FC	Pienso que el/la monitor/a en ocasiones nos propone tareas y situaciones que nos hacen sentir torpes o incapaces.	Pienso que el/la monitor/a nos propone actividades demasiado difíciles. [Penso que el/la monitor/a ens proposa activitats massa difícils.]	
FR	Pienso que el/la monitor/a en ocasiones es indiferente con nosotros.	Pienso que el/la monitor/a genera un ambiente que dificulta las buenas relaciones tanto entre los/las compañeros/as de clase como con él/lla. [Penso que el/la monitor/a genera un ambient que dificulta les bones relacions tant entre els companys/es de classe com amb ell/a.]	
SA	En las sesiones de actividad física siento que las actividades que realizo se ajustan a mis intereses y a la forma como me gustaría hacer los ejercicios.	En las sesiones de actividad física siento que la forma de realizar los ejercicios se ajusta a mis necesidades.	En las sesiones de actividad física siento que puedo realizar los ejercicios de la forma en que prefiero hacerlos. [A les sessions d'activitat física sento que puc realitzar els exercicis de la manera en que prefereixo fer-los.]
SC	En las sesiones de actividad física siento que he hecho un gran progreso y que realizo	En las sesiones de actividad física siento que puedo cumplir con las exigencias de las sesiones.	En las sesiones de actividad física siento que puedo cumplir con las exigencias de las

Medida	Versión 1	Versión 2	Versión 3
	los ejercicios eficazmente.		actividades. [A les sessions d'activitat física sento que puc complir amb les exigències de les activitats.]
SR	En las sesiones de actividad física me siento muy cómodo/a cuando hago ejercicio con los demás.	En las sesiones de actividad física me siento muy cómodo/a cuando hago ejercicio con mis compañeros y compañeras y con el monitor/a. [A les sessions d'activitat física em sento molt còmode quan faig exercici amb els companys i companyes i amb el monitor/a.]	
INT	Participo en las sesiones de Actividad física porque me resulta placentero y satisfactorio el hacer ejercicio.	Participo en las sesiones de Actividad física porque disfruto con las actividades que realizo. [Participo a les sessions d'Activitat física perquè gaudeixo amb les activitats que realitzo.]	
ITG	Participo en las sesiones de Actividad física porque está de acuerdo con mi forma de vida y valores.	Participo en las sesiones de Actividad física porque considero que el ejercicio físico está de acuerdo con mis valores. [Participo a les sessions d'Activitat física perquè considero que l'exercici físic està d'acord amb els meus valors.]	
IDE	Participo en las sesiones de Actividad física porque valoro los beneficios del ejercicio físico. [Participo a les sessions		

Medida	Versión 1	Versión 2	Versión 3
ITY	d'Activitat física perquè valoro els beneficis de l'exercici físic.] Participo en las sesiones de Actividad física porque me siento culpable, avergonzado/a o nervioso/a cuando no hago ejercicio.	Participo en las sesiones de Actividad física porque me siento culpable cuando no practico ejercicio físico. [Participo a les sessions d'Activitat física perquè em sento culpable si no practico exercici físic.]	
EXT	Participo en las sesiones de Actividad física porque los demás dicen que debo hacerlo. [Participo a les sessions d'Activitat física perquè els altres diuen que ho he de fer.]		
AMO	Participo en las sesiones de Actividad física pero no veo por qué tengo que hacerlo.	Participo en las sesiones de Actividad física pero pienso que es una pérdida de tiempo. [Participo a les sessions d'Activitat física però pensó que és una pèrdua de temps.]	
VI	Generalmente me siento lleno/a de entusiasmo y energía.	Generalmente me siento vital. [Generalment em sento vital.]	

Nota. Los ítems marcados en negrita hacen referencia a la versión final. AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva.

Análisis preparatorio de datos

En la Figura 14 se encuentra la distribución de respuestas para las medidas de un solo ítem de la Muestra 1. A la izquierda antes de dicotomizar y a la derecha después de dicotomizar. Al igual que con las medidas de múltiples ítems, se observa una clara tendencia a dar una respuesta extrema (totalmente de acuerdo o totalmente en desacuerdo) con todos los ítems a excepción de frustración a la autonomía y regulación introyectada.

En la Figura 15 se encuentran representadas las respuestas para las medidas de un solo ítem antes de dicotomizar para la fase test y para la fase retest. El color del círculo representa la fase (test o retest). Al igual que con los otros gráficos el tamaño del círculo representa la proporción de respuestas. En este gráfico dado que no se puede solapar el valor numérico porque dificultaría la lectura, se acompaña con una leyenda para poder ver el valor aproximado de las proporciones. En la Figura 16, complementariamente se encuentran representadas las respuestas después de dicotomizar. A la izquierda para la fase test y a la derecha para la fase retest. Tal y como puede observarse en ambas figuras, la distribución de respuestas es similar en la fase test y en la fase retest. Igual que con la Muestra 1, el ítem de frustración a la autonomía y el ítem de regulación introyectada no presentan una distribución tan asimétrica. Los resultados estos dos ítems son compatibles con la descripción de las medidas de múltiples ítems. El ítem de frustración a la autonomía fue elegido de los ítems

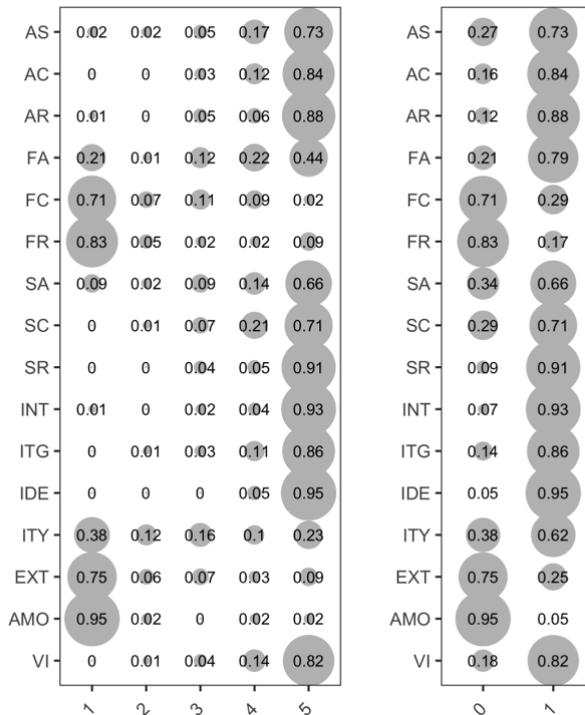


Figura 13 Distribución de Respuestas de las Medidas de un Solo Ítem Muestra 1

Nota. $n = 128$. A la izquierda, proporción de respuestas en cada categoría de los datos originales (1 = Totalmente de acuerdo; 5 = Totalmente en desacuerdo). A la derecha, proporción de respuestas después de dicotomizar a respuestas extremas (0 = En desacuerdo; 1 = De acuerdo). AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva

ya existentes del cuestionario homólogo (fa1) y tiene la misma distribución que este. De la misma forma, los ítems de regulación

introyectada también mostraron mayor variabilidad con las medidas de múltiples ítems. Hechas estas excepciones, en general, de nuevo se observa que las personas participantes están totalmente de acuerdo con las frases que tienen contenido positivo y totalmente en desacuerdo con las frases cuyo contenido es negativo.

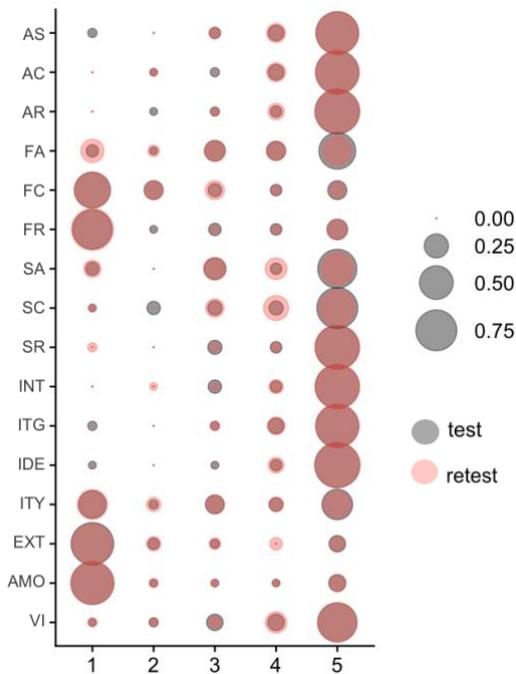


Figura 14 Distribución de Respuestas de las Medidas de un Solo Ítem Muestra 2

Nota. $n = 62$. El tamaño del círculo representa la proporción de respuestas. El color la fase (test o retest). AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva

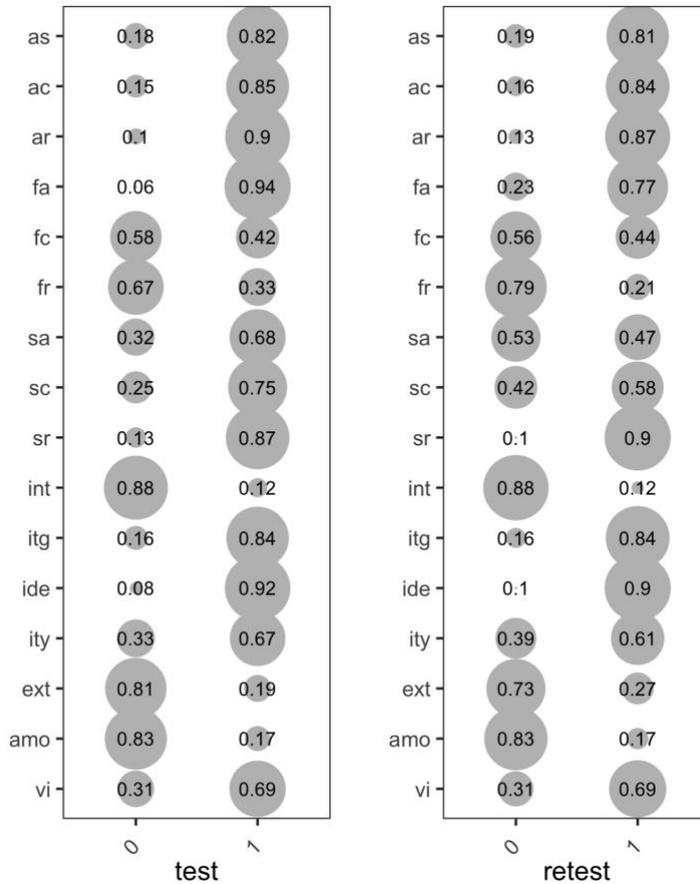


Figura 15 Distribución de Respuestas Dicotomizadas de las Medidas de un Solo Ítem Muestra 2

Nota. n = 62. El tamaño del círculo representa la proporción de respuestas. El color la fase (test o retest). AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva.

Validez

Evidencias basadas en el proceso de respuesta a los ítems

Tal como se ha mencionado con anterioridad, en una fase inicial de el proyecto ACTIVATICS se intentó obtener respuestas de forma autoadministrada de las medidas de ítems múltiples a un grupo de 17 personas que practicaban yoga en instalaciones municipales. Tras media hora intentando llevar a cabo la recogida de datos se suspendió la sesión debido a que ninguna de las personas presentes había contestado más de 10 ítems. Al examinar los cuadernillos de respuestas se observó que, más allá del tiempo que pudiera tomar la recogida de datos autoadministrada, la calidad de estos no era adecuada (e.g., casillas en blanco, solo se había contestado el primer ítem de cada página, ...). En algún cuadernillo se encontraron comentarios cualitativos como “La profesora lo hace muy bien” sin obtener ninguna respuesta en los ítems propuestos.

En cambio, al pasar a la heteroadministración de las medidas de múltiples ítems (Muestra 1) con un formato de entrevista la obtención de datos se facilitó enormemente. Estas entrevistas tomaron entre 20 y 50 minutos por persona participante. Durante la administración colectiva de los cuestionarios de un solo ítem (Muestra 2) la mayoría de los cuestionarios fueron autoadministrados, siendo una excepción las personas que no podían leer (i.e., no sabían leer, no llevaban las gafas en la sesión de actividad física) que fueron heteroadministrados por una de las investigadoras presentes. La

recogida de datos tomó entre 10 y 20 minutos en la fase *test*, y entre 6 y 15 minutos en la fase *retest*.

En este sentido cabe destacar la factibilidad de la recogida de datos colectiva con las medidas de un solo ítem respecto a las medidas de múltiples ítems. Durante la heteroadministración de ambas medidas se vieron reflejados algunos problemas de proceso de respuesta en algunos de los reactivos, especialmente en los relacionados con la autonomía. En algunos casos las participantes expresaban no saber que contestar. Este fue el caso del ítem de apoyo a la autonomía "Pienso que el/la monitor/a trata que tengamos libertad a la hora de realizar las actividades" o del de satisfacción de la autonomía "En las sesiones de actividad física la forma de realizar los ejercicios coincide perfectamente con la forma en que yo quiero hacerlos." donde las participantes exponían no saber que contestar porque precisamente venían a actividades dirigidas para que las dirigieran ya que consideraban que ellas no eran las expertas en la materia. Los ítems de frustración a la autonomía tenían especial dificultad ya que las personas participantes tenían la sensación de que si expresaban su opinión real podía tener consecuencias negativas para su monitor/a. Este es el caso por ejemplo del ítem "nos exige hacer las cosas de una manera concreta". Al contestar este reactivo las personas que participaban contestaban que evidentemente tenían que hacer las cosas de una manera concreta ya que por eso estaban en una actividad dirigida, pero que si marcaban que estaban de acuerdo quizá eso sería malo para el monitor/a ya que tenían la percepción de que el ítem

quería medir alguna cosa negativa por el hecho de usar la palabra “exige”.

En referencia a los ítems relacionados con el apoyo o la satisfacción de competencia, más allá de los resultados cuantitativos obtenidos, las participantes dijeron reportar su percepción con base en la edad que tenían, no comparándolo con épocas pasadas en las cuales sus competencias habían sido mejores. También destacaron que si no se sintieran competentes para realizar esas actividades probablemente no asistirían a las sesiones.

Especialmente en los ítems relativos a la necesidad psicológica básica de relación, ya sea de apoyo o de satisfacción, las participantes manifestaron la sensación de haber contestado repetidamente a la misma pregunta. Eso también sucedió con las medidas de múltiples ítems para el cuestionario de regulación conductual. A menudo las participantes se mostraban aburridas de contestar y decían si no podían contestar simplemente que participaban porque les gustaba. En referencia a los ítems nuevos desarrollados, el ítem de regulación integrada “porque está de acuerdo con mis valores” en ocasiones resultó difícil de comprender ya que las personas consideraban que ese ítem era demasiado pretencioso para el contexto en el que se daban las sesiones y que quizá tendría sentido si fueran deportistas profesionales.

En la autoadministración de los cuestionarios, más allá de la comprensión de alguno de los reactivos, destacan cuestiones relacionadas con la escala de respuesta, que habían sido detectadas

previamente en la heteroadministración. Aunque las personas investigadoras insistían en que tenían que expresar su grado de acuerdo en una escala del 1 (*Totalmente en desacuerdo*) al 5 (*Totalmente de acuerdo*), las personas participantes insistían en contestar Sí (*Estoy de acuerdo*) y No (*No estoy de acuerdo*). En varios cuestionarios autoadministrados algunas participantes han tachado "*Totalmente en desacuerdo*" y "*Totalmente de acuerdo*" y lo han substituido por un "*No*" y por un "*Sí*" precisando exactamente lo que estaban contestando. Hay dos tipos de excepciones a este fenómeno: una excepción a estos casos fueron personas que reportaron haber trabajado anteriormente como monitores/as o profesores/as que intentaban ser más críticas, sobre todo con el apoyo a las necesidades psicológicas básicas; otra excepción fueron los ítems relacionados con el apoyo y la satisfacción de la autonomía que cuando las participantes no sabían que contestar decían a las personas que administraban los cuestionarios que marcaran un "3" que eso significaba "*Ni sí ni no*" y no querían dejar la pregunta en blanco.

Evidencias basadas en la relación con otras variables

La primera evidencia de validez aportada basada en la relación con otras variables viene del estudio de las correlaciones entre las medidas de un solo ítem y sus medidas homólogas de múltiples ítems. Los tres ítems de apoyo a las necesidades psicológicas básicas mostraron valores elevados (apoyo a la autonomía [$r_{xy} = .81$], apoyo a la competencia [$r_{xy} = .96$] y apoyo a la relación [$r_{xy} = .67$]). Dos ítems de satisfacción de las necesidades psicológicas básicas mostraron

valores elevados (satisfacción de la competencia [$r_{xy} = .70$], satisfacción de la relación [$r_{xy} = .78$]) mientras que la satisfacción de la autonomía fue una excepción [$r_{xy} = .37$]. En referencia al ítem de vitalidad subjetiva mostró una alta correlación entre ambas medidas [$r_{xy} = .77$].

Para examinar la relación entre las variables implicadas en la cadena de relaciones de la SDT, y dado que por falta de potencia no se ha podido examinar con un *path analysis* se han examinado las correlaciones tetracóricas entre las medidas de un solo ítem. Para ello, en primer lugar, se han examinado las correlaciones para la Muestra 1 ($n = 128$). Para la Muestra 2 ($n = 62$), se han examinado las correlaciones en la fase test y en la fase retest.

En la Figura 17, se encuentra el correlograma para las medidas de un solo ítem para la Muestra 1. En color azul se ven representadas las correlaciones positivas entre sí de los ítems de apoyo a las necesidades psicológicas básicas, de los de frustración a las necesidades psicológicas básicas, de los de satisfacción de las necesidades psicológicas básicas, de los de motivación autónoma y de los de motivación controlada. También se encuentran representadas las relaciones esperadas entre los diversos conceptos implicados en la SDT. De esta forma se observa una relación coloreada en azul por ser positiva entre los ítems de apoyo a las necesidades psicológicas básicas con los de satisfacción de las necesidades psicológicas básicas, y los de motivación autónoma y de vitalidad subjetiva. También se observan relaciones negativas, representadas con el color rojo, de los ítems de

frustración a las necesidades psicológicas básicas y amotivación con todos los ítems positivos.

Cabe destacar que en la Muestra 1 el ítem de apoyo a la relación tiene una elevada correlación ($r = .67$) con el ítem de vitalidad mientras que los otros ítems de satisfacción de las necesidades

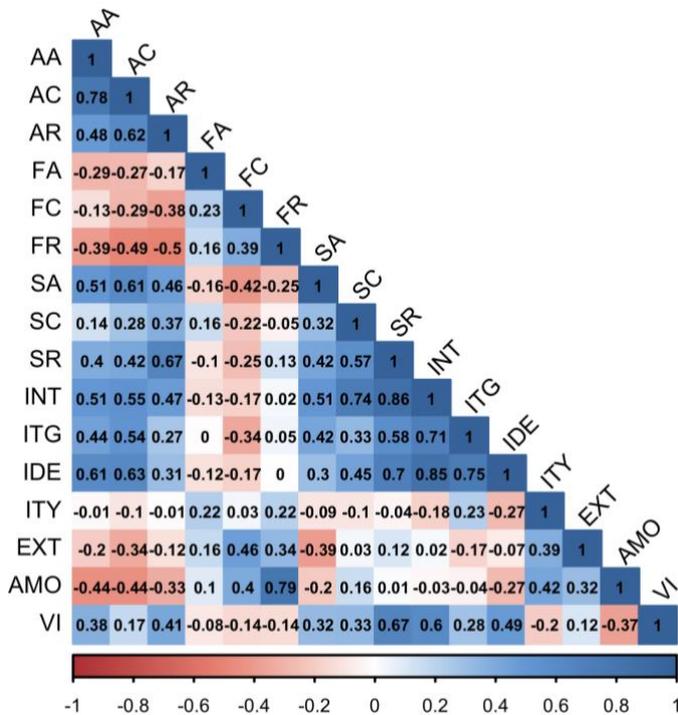


Figura 16 Correlograma de las Medidas de un Solo Ítem Muestra 1

$n = 128$; AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva.

psicológicas básicas presentan correlaciones moderadas. El ítem regulación intrínseca también presenta una correlación elevada con el ítem de vitalidad subjetiva ($r = .60$).

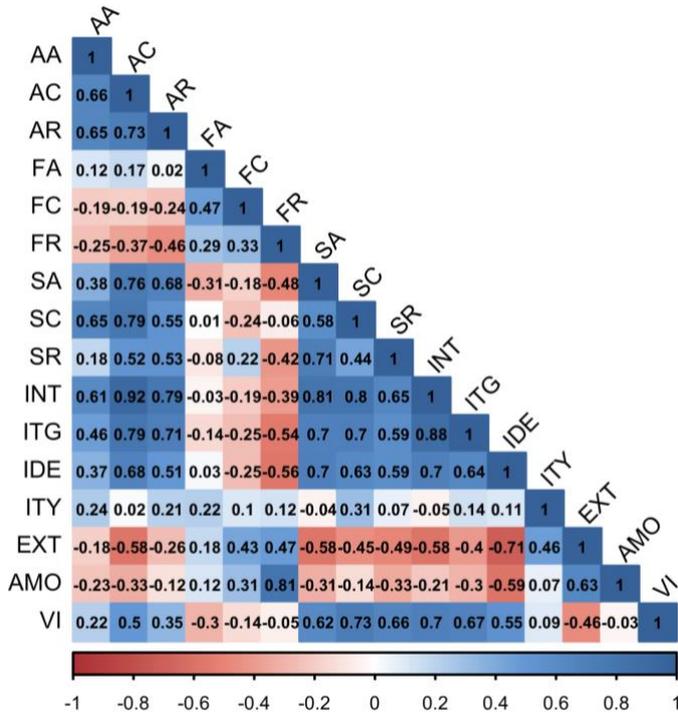


Figura 17 Correlograma de las Medidas de un Solo Ítem Muestra 1 Fase Test

$n = 128$; AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva.

En la Figura 18, se encuentra el correlograma para las respuestas a las medidas de un solo ítem de la fase test de la Muestra 2. Al igual que con la Muestra 1, se observan las relaciones positivas y negativas que postula la SDT. Una excepción es el ítem de frustración a la autonomía que establece relaciones nulas o positivas débiles con los ítems de apoyo a las necesidades psicológicas básicas.

En referencia a la relación con vitalidad subjetiva, a diferencia de la Muestra 1, todos los ítems de satisfacción de las necesidades psicológicas básicas y todos los ítems de motivación autónoma tienen correlaciones elevadas con el ítem de vitalidad subjetiva (rango .55 a .73).

Finalmente, en la Figura 19 se encuentra el correlograma para la fase retest de la Muestra 2. Aunque en general se reproduce la misma direccionalidad de las relaciones que en la Muestra 1 y en la fase test de la Muestra 2 estas relaciones son en general más débiles. En referencia a las relaciones con vitalidad subjetiva, cabe destacar que en esta muestra y fase se observan correlaciones moderadas con satisfacción de la autonomía y satisfacción de la competencia y también entre los ítems de motivación autónoma.

En conclusión, se puede decir que, por un lado, de las medidas de un solo ítem que se ha podido obtener evidencia en relación con su medida de múltiples ítems homóloga, la única medida que no ha mostrado evidencia favorable es la de satisfacción de la autonomía. Es importante considerar que solo cuatro de las siete medidas de un solo

ítem habían llegado a un umbral mínimo de fiabilidad en las medidas de múltiples ítems homólogas.

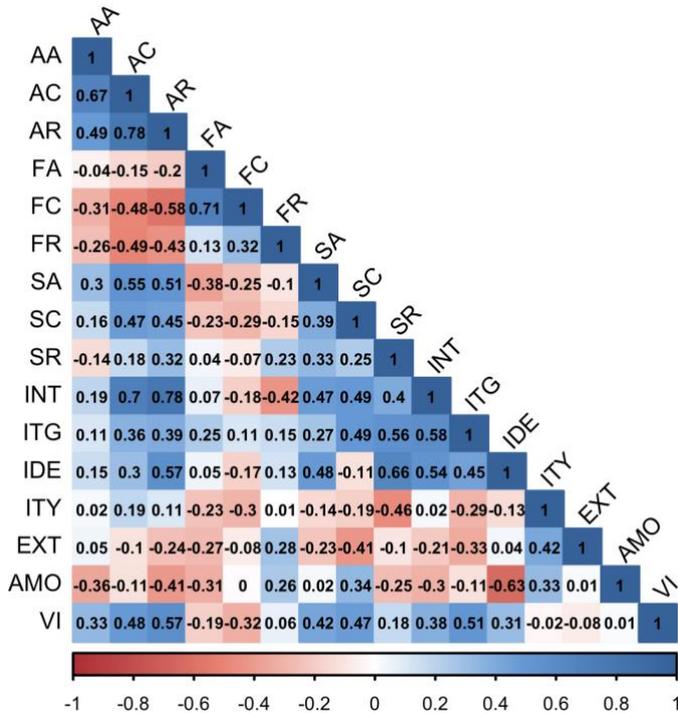


Figura 18 Correlograma de las Medidas de un Solo Ítem Muestra 2 Fase Retest

n = 62; AA = apoyo a la autonomía; AC = apoyo a la competencia; AR = apoyo a la relación; FA = frustración a la autonomía; FC = frustración a la competencia; FR = frustración a la relación; SA = satisfacción autonomía; SC = satisfacción competencia; SR = satisfacción relación; INT = regulación intrínseca; ITG = regulación integrada; IDE = regulación identificada; ITY = regulación introyectada; EXT = regulación extrínseca; AMO = amotivación; VI = vitalidad subjetiva.

Por otra parte, con el estudio de las correlaciones a partir de los tres correlogramas (i.e., Muestra 1, Muestra 2 test, Muestra 2

retest) se puede decir que en general se reproducen las relaciones positivas y negativas contempladas en la SDT. Con la evidencia aportada se puede concluir que en general existe una relación positiva entre las variables implicadas en la motivación autónoma con la vitalidad subjetiva pero no hay evidencias concluyentes sobre el rol de cada satisfacción de las necesidades psicológicas básicas en el impacto de la vitalidad puesto que en cada correlograma presentan un patrón de correlaciones diferente.

Evidencias basadas en las consecuencias de la evaluación

Es importante tener en cuenta las consecuencias de la aplicación de las medidas reportadas en este estudio concretamente las consecuencias no intencionadas. Como parte del proceso de obtención del consentimiento informado, las personas investigadoras explicaron a las participantes el objetivo de la investigación y sus aplicaciones, así como que al terminar el proyecto también se presentaría un informe a los centros con un resumen de los resultados principales.

Algunas participantes percibieron este hecho como que una respuesta negativa hacia los monitores podría perjudicar a su estatus laboral. Esta percepción puede haber afectado a que algunas de las respuestas. En particular, algunos de los ítems relacionados con la autonomía donde las participantes expresaron que no sabían que contestar, uno de los motivos podría ser que no sabían que contestar para no perjudicar a sus monitores/as.

Fiabilidad

Todos los coeficientes de fiabilidad obtenidos para las medidas de un solo ítem, tanto los derivados de las medidas de múltiples ítems como del test-retest, pueden encontrarse en la Tabla 5.

Utilizando como indicador la profecía de Spearman-Brown (ρ_{xx}^*) la medida con un valor más elevado de fiabilidad es la de vitalidad subjetiva ($\rho_{xx}^* = .60$) mientras que la más limitada es apoyo a la autonomía ($\rho_{xx}^* = .25$). Estos resultados eran de esperar puesto que este era el caso también del indicador de referencia, el coeficiente omega categórico de consistencia interna de sus medidas homólogas de múltiples ítems. El resto de los valores se sitúan entre .31 y .43. Considerando que estamos aplicando la profecía a una escala de un ítem no se esperan valores muy elevados de fiabilidad en esta estimación.

De la aplicación de la fórmula de corrección por atenuación (R_{yy}) podemos destacar que cuatro de los ítems pueden ser considerados adecuados (i.e., apoyo a la autonomía, apoyo a la competencia, satisfacción de la competencia y satisfacción de la relación). Apoyo a la relación y vitalidad no llegan al umbral de .70 pero su coeficiente es más optimista que el de satisfacción a la autonomía, el cual es especialmente bajo (.21).

Tabla 5 Fiabilidad de las Medidas de un Solo Ítem

Variable	Muestra 1 (n = 128)			Muestra 2 (n = 62)	
	ρ_{xx}^*	R_{yy}	h^2	acu	pa
Apoyo a la autonomía	.25	1.00	.64	82	.65
Apoyo a la competencia	.43	1.00	.82	82	.65
Apoyo a la relación	.41	.61	.69	90	.80
Frustración a la autonomía				74	.48
Frustración a la competencia				58	.15
Frustración a la relación				77	.55
Satisfacción de la autonomía	.32	.21	.23	62	.23
Satisfacción de la competencia	.31	.77	.54	70	.39
Satisfacción de la relación	.42	.81	.76	84	.67
Regulación intrínseca				83	.66
Regulación integrada				74	.48
Regulación identificada				88	.77
Regulación introyectada				61	.21
Regulación externa				73	.45
Amotivación				73	.47
Vitalidad subjetiva	.60	.67	.58	68	.35

Nota.; ρ_{xx}^* = fiabilidad obtenida por profecía de Spearman-Brown; R_{yy} = fiabilidad obtenida por fórmula de corrección por atenuación; h^2 = fiabilidad como comunalidad del ítem en un análisis factorial confirmatorio; acu = Porcentaje de acuerdo entre la fase test y la fase retest; pa = coeficiente PABAK.

Tal como ya se ha explicado en la sección de Análisis de datos, para calcular las comunalidades (h^2) se utilizó el valor obtenido en estudio de la estructura interna de la medida de múltiples ítems.

Excepcionalmente, para los dos ítems que se redactaron de nuevo, se realizó con un CFA unidimensional incluyendo el nuevo ítem. Para satisfacción de la autonomía se obtuvo un ajuste adecuado (χ^2 [df] = 9.32[5], $p = .09$, CFI = .95, TLI = .91, RMSEA [90%IC] = .08[.00-.17]) y también para vitalidad subjetiva (χ^2 [df] = 6.06[9], $p = .73$, CFI = 1.00, TLI = 1.00, RMSEA[90%IC] = .09[.00-.07]). Igual que con el estudio de la fiabilidad derivado de la fórmula de corrección por atenuación, el ítem apoyo a la autonomía ($h^2 = .23$) mostró el coeficiente más bajo de fiabilidad de todos los estudiados y en este caso fue apoyo a la competencia ($h^2 = .82$) el indicador con mejor fiabilidad. Para el resto de los ítems se obtuvieron valores entre .54 y .76.

Finalmente, en el estudio de la fiabilidad test-retest se valoró el porcentaje de acuerdo, así como el coeficiente PABAK. Si tenemos en cuenta el porcentaje de acuerdo, todos los ítems tienen al menos un acuerdo moderado, a excepción de frustración a la competencia (.58), siendo incluso excelente en cinco ítems (los tres ítems de apoyo a las necesidades psicológicas básicas, el ítem de satisfacción de la relación y el de regulación intrínseca).

Si valoramos el coeficiente PABAK, la visión es algo menos optimista. Los mismos ítems que mostraron un acuerdo excelente, muestran un buen acuerdo con el coeficiente PABAK (rango .65 a .80). En cambio, los ítems de frustración a la autonomía, frustración a la relación, regulación integrada, regulación externa y amotivación muestran un acuerdo moderado (rango .45 a .55). Tres ítems (satisfacción de la autonomía, satisfacción de la competencia y vitalidad

subjetiva) mostraron un acuerdo bajo (rango .23 a .39) y dos ítems mostraron un acuerdo muy débil (frustración a la competencia y regulación introyectada).

Resumiendo, la fiabilidad de las medidas de un solo ítem examinada a partir de las medidas de múltiples ítems ha mostrado evidencia favorable en cuatro de las siete medidas de un solo ítem de las que se ha podido obtener este tipo de evidencia. En referencia a la fiabilidad test-retest se ha mostrado evidencia favorable para las medidas de apoyo a las necesidades psicológicas básicas y evidencia menos concluyente en los otros ítems.

5

DISCUSIÓN

Con el objetivo inicial de conocer el estado motivacional que manifiestan las personas mayores que realizan ejercicio físico, se han examinado las propiedades psicométricas de cuatro cuestionarios con los conceptos principales de la SDT en actividad física y se han desarrollado medidas de un solo ítem para las mismas dimensiones.

Tal y como hemos visto con las frecuencias tanto de las medidas de un solo ítem como de las medidas de múltiples ítems, las personas que han contestado los cuestionarios en este trabajo son, personas que manifiestan recibir apoyo y no frustración por parte de sus líderes, tener satisfechas sus necesidades psicológicas básicas, que practican actividad física motivadas autónomamente y que en general se sienten vitales. Los datos obtenidos se caracterizan por su poca variabilidad observándose en todos los ítems una gran propensión de las personas participantes a seleccionar respuestas en los extremos favorables de la escala. Por ello, todos los análisis subsiguientes se han basado en las respuestas dicotomizadas.

Para conocer las propiedades psicométricas de las puntuaciones de cuestionarios ya utilizados en otras poblaciones de adultos se examinó la estructura interna y la consistencia interna utilizando CFA, para facilitar su comparación con investigaciones previas. Como es conocido, con variables binarias este análisis proporciona reescalados los parámetros de dificultad y discriminación que se obtendrían utilizando los modelos de teoría de respuesta al ítem (e.g., Raykov y Marcoulides, 2016). Si se tienen en cuenta los criterios clásicos establecidos para valorar la bondad de ajuste se puede considerar que la estructura interna de los cuestionarios de apoyo a las necesidades psicológicas básicas, satisfacción de las necesidades psicológicas básicas y vitalidad subjetiva encaja con las expectativas basadas en la SDT.

Para el cuestionario de estilo interpersonal no se ha podido reproducir la estructura de seis factores propuesta por los autores del cuestionario original aplicado a adolescentes (Pulido et al., 2017). El modelo propuesto de tres factores en apoyo a las necesidades psicológicas básicas, aunque presenta baja discriminación entre factores, muestra resultados compatibles con los mostrados en adolescentes como se discute más adelante.

Por su parte, el cuestionario de regulación conductual no ha mostrado ninguna solución factorial satisfactoria en esta muestra. Aunque existen estudios que utilizan el mismo cuestionario con personas mayores los datos psicométricos aportados son solo en parte comparables con los resultados de esta tesis. Investigaciones previas,

por ejemplo, Maillot et al. (2018) o Marcos et al. (2014) con muestras de personas mayores aplicaron el mismo cuestionario sin examinar su estructura interna y además el coeficiente alfa de Cronbach aportado para alguna de las subescalas era inferior al umbral comúnmente admitido de .70. En línea con los resultados de esta tesis, los estadísticos descriptivos que muestran en Maillot et al. (2018) sugieren también presencia de efectos suelo y techo. Por ejemplo, el promedio de regulación intrínseca, siendo el mínimo 0 y el máximo 4, es de 3.52 (DE = 0.68). En otro estudio con personas entre 18 y 84 años (Zazo y Moreno-Murcia, 2015) tampoco examinaron la estructura interna y también obtuvieron valores alfa de Cronbach por debajo del umbral. O más recientemente, Cid et al. (2018), en una muestra entre 16 y 78 años en la cual si que han examinado la estructura interna, no han obtenido resultados favorables con la estructura propuesta y han terminado presentando un modelo en el cual excluyeron seis ítems, uno por dimensión. Esta solución dejaba cada factor exactamente identificado con solo tres indicadores además de constituir una reducción del cuestionario no suficientemente justificada basada únicamente en la estructura interna y no en otros argumentos como la validez de contenido. Se puede concluir que los resultados psicométricos discutibles no son infrecuentes cuando se pretende medir regulaciones conductuales mediante cuestionarios. Uno de los más claros es la observación de efectos techo y suelo en las respuestas a los ítems.

Otro de ellos es el problema de discriminación entre factores encontrados en el modelo de apoyo a las necesidades psicológicas básicas y en el modelo de satisfacción de las necesidades psicológicas o la falta de discriminación entre algunos ítems de diferentes cuestionarios como se ha visto en el *correlation network analysis*. Por una parte, el hecho de que las dimensiones de apoyo a las necesidades psicológicas básicas tengan altas correlaciones entre ellas está contemplado desde la teoría (Wilson et al., 1982). Por otra parte, finalmente los autores del cuestionario original (Pulido et al., 2017) decidieron eliminar algún ítem debido a las altas cargas cruzadas con otros factores. En un estudio con adolescentes dentro del proyecto en el que se desarrolla esta tesis, con el mismo cuestionario también se encontró falta de discriminación entre estos factores con adolescentes (Angulo-Brunet, Sanchez-Oliva, y Viladrich, 2017). En la misma línea también está contemplado que las dimensiones de satisfacción de las necesidades psicológicas básicas estén relacionadas entre ellas tal y como indican Ryan y Deci (2017). En la adaptación española del cuestionario, Sanchez y Nuñez (2007) permitieron la correlación entre los residuales de algunos de los ítems de satisfacción de la autonomía y satisfacción de las relaciones. También Couto et al. (2017) en un estudio con personas mayores que practicaban actividad física detectaron la falta de discriminación entre las tres necesidades psicológicas básicas, concretamente entre la necesidad de competencia y la necesidad de relación. Es decir, la falta de discriminación entre los factores propuestos por la SDT es un resultado

que ha sido replicado en el ámbito de la actividad física en grupos humanos de edades diversas.

Respecto a las correlaciones entre los ítems de cuestionarios distintos, teniendo en cuenta que tanto el apoyo como la satisfacción de las necesidades psicológicas básicas son conceptos que examinan un mismo fenómeno es de esperar que en cierta medida estén relacionados, aunque se espere que uno preceda a otro.

En referencia a la fiabilidad es de esperar que los valores sean sistemáticamente más bajos que los de los cuestionarios originales ya que es bien conocido que cuanto más reducida es la escala de respuesta, peor es la fiabilidad (Lozano, García-Cueto, y Muñiz, 2008; Simms, Zelazny, Williams, y Bernstein, 2019). Hasta la fecha no se conocen estudios en los cuales se hayan utilizado estos cuestionarios con esta escala de respuesta y por lo tanto las posibilidades de comparación son limitadas. A pesar de este hecho, también cabe destacar algunos puntos relevantes que podrían explicar, en parte, estos valores bajos en los coeficientes de fiabilidad.

En primer lugar, la falta de fiabilidad de las escalas relacionadas con la autonomía (i.e., apoyo a la autonomía y satisfacción de la autonomía) podría ser explicada por el proceso de respuesta tal y como se discute más adelante. No obstante, los resultados en la escala de apoyo a la autonomía son compatibles con los del coeficiente obtenido en un estudio previo con una muestra con adolescentes (Angulo-Brunet, Sanchez-Oliva, et al., 2017). En particular la fiabilidad de la suma de los ítems de apoyo a la competencia es incluso mayor que el

obtenido con muestras de adolescentes (Angulo-Brunet, Sanchez-Oliva, et al., 2017; Pulido et al., 2017).

En relación con la fiabilidad de la satisfacción de las necesidades psicológicas básicas, y teniendo en cuenta que Sanchez y Nuñez (2007) contemplaron un modelo con errores correlacionados, el coeficiente que publican (alfa de Cronbach) podría estar sesgado, sobrestimando la fiabilidad (Viladrich et al., 2017).

Finalmente, el coeficiente de fiabilidad de consistencia interna para vitalidad subjetiva obtenido en la muestra de este estudio está en la línea del coeficiente obtenido en Castillo et al. (2017) donde obtienen valores del coeficiente alfa de Cronbach similares (rango .82 a .89).

Para las medidas de un solo ítem, si examinamos la validez en relación con las variables externas, la evidencia más clara aportada son las altas correlaciones entre las medidas de un solo ítem y las medidas de múltiples ítems. En la mayoría de los casos, se puede pensar que las medidas de un solo ítem sustituirían sin gran pérdida de información a las medidas de múltiples ítems homólogas. Mediante el estudio de las correlaciones, en los tres correlogramas, se ha visto como se forman clústeres entre los conjuntos de ítems que deberían estar relacionados positivamente de acuerdo con las previsiones de la SDT. Se ha visto que hay una relación positiva entre la satisfacción de las necesidades psicológicas básicas y la vitalidad subjetiva, entre la satisfacción de las necesidades psicológicas básicas y la motivación autónoma, y entre la motivación autónoma y la vitalidad subjetiva. Puesto que se ha

obtenido evidencia exploratoria, las técnicas utilizadas no permiten dar luz a la discusión de que satisfacción de las necesidades psicológicas básicas es más importante en personas mayores (e.g., Henning et al., 2019; Neubauer, et al., 2017).

En referencia a la fiabilidad de las medidas de un solo ítem, concretamente las evidencias obtenidas con los coeficientes que dependen de las medidas de múltiples ítems, se puede decir que excepto el ítem de satisfacción de la autonomía, los otros ítems tienen unos índices de fiabilidad adecuados. Es importante recordar que de estos siete constructos solo cuatro resultaron fiables en su medida de múltiples ítems.

Respecto a la fiabilidad test-retest, hay más particularidades. Cabe destacar que los ítems de apoyo a las necesidades psicológicas básicas tienen un valor elevado. En cambio, tanto en la frustración como en la satisfacción de las necesidades psicológicas básicas se obtienen valores bajos o moderados. El comportamiento de la fiabilidad test-retest en regulación conductual es más heterogéneo: la regulación intrínseca y la identificada obtienen coeficientes especialmente buenos, y en el caso de la regulación introyectada se observa un valor muy bajo. Un ítem preocupante dado el buen desempeño mostrado con las otras evidencias de validez y fiabilidad es el de vitalidad subjetiva que tiene un valor extremadamente bajo. Con todo, este resultado, es congruente con el valor de fiabilidad test-retest

obtenido en la última obtención de evidencias de validez y fiabilidad del cuestionario en muestra general española (Castillo et al., 2017).

Si nos centramos en las evidencias cuantitativas de calidad psicométrica aportadas podemos decir que para seis de las siete medidas de un solo ítem se ha obtenido evidencia favorable de validez mientras que para cuatro de ellas (i.e., apoyo a la competencia, apoyo a la relación, satisfacción de la relación y vitalidad subjetiva) se han obtenido evidencias favorables tanto de validez como de fiabilidad lo que apoyaría su uso en esta población. Es importante destacar que las evidencias de calidad derivadas de las medidas de múltiples ítems son favorables siempre y cuando se confíe en su medida múltiple homóloga. Estos no serían pues argumentos muy fuertes frente a la crítica presentada por Krueger et al. (2013) que insta a la comunidad científica a no transferir las propiedades psicométricas de una medida larga a su medida reducida.

Sin embargo, hay otros argumentos a tener en cuenta. Siguiendo las recomendaciones actuales que promueve la APA (Appelbaum et al., 2018; AERA, APA y NCME, 2014) cuando se reportan las propiedades psicométricas de los instrumentos es importante, además atender a las evidencias de calidad psicométrica de tipo cualitativo. A continuación se discuten los datos presentados en esta tesis a la luz de las evidencias de validez relacionadas con el contenido de los ítems, el proceso de respuesta y las consecuencias de evaluación.

Aunque el panel de expertos y las entrevistas cognitivas sugirieron que los ítems eran suficientemente claros en el estudio de

campo tomaron especial interés las evidencias aportadas en relación con el proceso de respuesta y con las consecuencias de evaluación. Tal y como se ha reportado en el capítulo de resultados, algunas personas tenían dudas sobre que contestar para no perjudicar a su monitor/a. En este sentido interpretar entonces que una puntuación concreta de apoyo a la autonomía puede significar que un monitor/a apoya en mayor o menor grado la autonomía de sus pupilos sería discutible. Por ejemplo, algunas personas indicaron que preferían la elección de la respuesta neutra antes que dejar en blanco un ítem. A la luz de todo ello serían comprensibles pues los valores comprometidos de fiabilidad encontrados en estas dimensiones. Además, y concretamente sobre el concepto de autonomía, en estudios anteriores se han proporcionado explicaciones que afectarían más bien a la validez de contenido e incluso al papel del propio concepto dentro de la SDT. Tal y como indican Marcos et al. (2014) quizá sería necesario revisar la definición del concepto de satisfacción de la autonomía en esta población en concreto. O desde otra perspectiva, tal y como indican Ntoumanis et al. (2017) en algunas culturas las personas ven como más deseable un monitor/a que promueva el apoyo a la autonomía mientras que en otras culturas esto no es socialmente tan deseable. Es posible que esta conclusión basada en la cultura pueda extenderse a distintos grupos humanos en función de su edad. Si así fuera, los valores asociados a este concepto podrían estar haciendo que las personas modulen su respuesta y por lo tanto que la interpretación de estas puntuaciones se vea comprometida. Puesto que la mayoría de las intervenciones

basadas en la SDT radican en fomentar el apoyo y la satisfacción de la autonomía (Alcaraz et al., 2017) a la luz de esta discusión consideramos que sigue siendo importante trabajar en tres líneas. En primer lugar, es necesario comprender como se conceptualiza la necesidad de autonomía en personas mayores para poder diseñar e implementar intervenciones basadas en este componente, si fuera necesario. En segundo lugar, es crucial comprender los procesos de respuesta antes de utilizar la puntuación de estas escalas con el fin de utilizarla para evaluar el efecto de las intervenciones que se puedan realizar. Y en tercer lugar, tendría sentido replantear el foco básico de las intervenciones y comprobar si una intervención dirigida principalmente a satisfacer la necesidad de relación o de competencia es más efectiva en grupos humanos como el que nos ocupa.

En el aspecto metodológico también hay algunos puntos a destacar. Primero, tal como se ha visto repetidamente en esta discusión, la falta de transparencia o alguna de las prácticas cuestionables a la hora de obtener evidencias de validez o fiabilidad comportan limitaciones en la comparación de los resultados entre diferentes estudios. Sería conveniente que, especialmente en este tipo de población, los estudios contemplen no solo los análisis encaminados a examinar sus hipótesis principales sino también las propiedades psicométricas de los instrumentos. Tal como hemos visto en la revisión de la literatura, son numerosos los estudios que han obviado el paso de presentar evidencias de validez u fiabilidad de los instrumentos que utilizan (e.g., Gray et al., 2017; Henning et al., 2019; Solberg et al.,

2013). De acuerdo con las recomendaciones de la APA aportar datos de validez psicométrica para la muestra concreta que se está estudiando es un paso insoslayable en toda investigación. Estas recomendaciones son completamente explícitas en la publicación más reciente. En esta tesis se defiende que es importante publicar cuando una medida no funciona y porqué. La discusión de este punto se ha llevado al XVI Congreso de Metodología de las Ciencias Sociales y de la Salud (Apéndice C2; Angulo-Brunet y Viladrich, 2019) donde se ha puesto en relieve la discutible utilidad de realizar un *path analysis* cuando hay problemas de validez o fiabilidad en las medidas. En esta tesis se plantea un doble dilema: por un lado, no se ha obtenido evidencia de validez basada en la estructura interna para uno de los cuestionarios, de manera que si a pesar de todo quisiéramos estudiar la cadena de relaciones prevista por los teóricos de la SDT deberíamos simplificar la cadena eliminando los constructos no válidos del análisis; por otro lado, de los cuestionarios en los que sí que se ha obtenido evidencia favorable de estructura interna, tres de los siete constructos no han obtenido un valor adecuado de fiabilidad. Aunque algunos investigadores podrían estar de acuerdo en que dos de estos valores son limítrofes (entre .60 y .70), el valor de apoyo a la autonomía está muy por debajo del umbral mínimo de .70. Teniendo en cuenta que el apoyo a la autonomía es uno de los conceptos principales de la SDT, incluir una medida en un *path analysis* de la cual no tenemos evidencia de fiabilidad para poder interpretar su puntuación solo añade confusión a la descripción del fenómeno que se quiere estudiar.

Segundo, la práctica de dicotomizar la escala de respuesta es el resultado de tratar los datos teniendo en cuenta su distribución y se ha adoptado después de discutirla con profesionales. Durante el congreso *Small Sample Size Solutions*, se presentaron resultados parciales de las medidas de un solo ítem y de las medidas de múltiples ítems con la finalidad de que los asistentes aportaran soluciones para la poca variabilidad encontrada en las respuestas (Apéndice C3; Angulo-Brunet y Viladrich, 2018). Es así como, por ejemplo, Yves Rosseel en comunicación personal sugirió dicotomizar la escala. El coste de la dicotomización en pérdida de fiabilidad puede ser considerado muy elevado a priori (Lozano et al., 2008). Sin embargo, en este caso, los efectos techo y suelo observados hubieran comprometido igualmente todo el estudio psicométrico. Otra opción de análisis contemplada por algunos de los asistentes fue utilizar *parcelling* (Little, Cunningham, Shahar, y Widaman, 2002). Esta opción de análisis fue finalmente descartada ya que, por definición no era posible hacerlo con medidas de un solo ítem y tampoco en este caso con las medidas de múltiples ítems puesto que cada constructo tenía únicamente cuatro indicadores.

Se debe avanzar en el conocimiento de los procesos internos de las personas mayores ya sea mediante la administración de cuestionarios o mediante otro tipo de metodología. Fuera como fuese es importante que la metodología sea rigurosa y transparente, y que se hagan públicos tanto resultados positivos como negativos con la finalidad de poder aumentar el conocimiento de esta población.

6

LIMITACIONES Y FORTALEZAS

Como todo trabajo esta tesis también presenta limitaciones. En este trabajo es necesario reconocer las limitaciones ligadas al desarrollo del trabajo, y por otro lado examinar las limitaciones encontradas al crear una sinergia entre la evaluación de la teoría de la autodeterminación en personas mayores y la psicometría para el desarrollo y validación de medidas de un solo ítem.

Si nos centramos en las limitaciones propias del trabajo, por un lado, se han utilizado muestras de conveniencia para examinar las propiedades psicométricas de las medidas desarrolladas. Para que la muestra fuera representativa, al menos de la provincia de Barcelona, se intentó contar con muestras de cuatro comarcas diferentes.

Cabe destacar también el tamaño limitado de la muestra, sobre todo teniendo en cuenta que en algunos de los análisis se ha utilizado una técnica de estimación para datos categóricos. Tal y como indican Lloret-Segura, Ferreres-Traver, Hernández-Baeza y Tomás-Marco (2014) en presencia de asimetría, especialmente cuando el tamaño de la muestra es limitado o hay muchos ítems por factor, hay que ser más exigentes con el tamaño de la muestra. Aunque esto es una clara

limitación, teniendo en cuenta la baja variabilidad de las respuestas y las continuas afirmaciones de las personas participantes de que realizan esta actividad porque les gusta, no se espera que aumentando el tamaño de la muestra este problema cambie. Parece difícil aumentar la variabilidad en las respuestas sin modificar la población diana. En esta línea, aunque se dispone de datos sociodemográficos de los cuales podría ser interesante desglosar algunos de los análisis, dada la baja variabilidad se ha considerado que no cabe lugar a su desglose o estratificación. En este sentido hubiera sido pertinente poder analizar si existía algún efecto entre las diferentes formas de administración de las medidas de un solo ítem (i.e., autoadministrado o heteroadministrado) o entre los dos idiomas utilizados en la recogida de datos.

Tanto el tamaño de la muestra como la distribución de las respuestas ha condicionado la elección final de las técnicas estadísticas a utilizar en este trabajo. Aunque un objetivo inicial fue examinar mediante modelos de ecuaciones estructurales y/o *path analysis* la red nomológica de relaciones entre los constructos, ateniéndonos a los resultados obtenidos en el estudio de la estructura interna de las medidas de múltiples ítems y a la falta de variabilidad de las medidas de un solo ítem, finalmente no ha sido posible. Aún así la recomendación sigue siendo que en la medida de lo posible futuros estudios aporten este tipo de evidencia de validez.

Otra limitación se deriva del hecho de usar medidas de múltiples ítems con garantías psicométricas limitadas. Tanto el

cuestionario de satisfacción de las necesidades psicológicas básicas, el de regulación conductual, como el de vitalidad subjetiva habían sido previamente utilizados en otros grupos etarios. En referencia a la elección del cuestionario estilo interpersonal desde el enfoque de la SDT, al inicio de este proyecto no existía ninguna herramienta disponible en castellano que hubiera aportado evidencia suficiente de validez y fiabilidad. La elección del cuestionario de estilo interpersonal se hizo a partir del conocimiento informal de que se estaba desarrollando ese cuestionario. Cuando el estudio de campo ya había finalizado, paralelamente a la publicación definitiva del CIS-Q (Pulido et al., 2017) se publicó el IBQ (Rocchi, Pelletier, Cheung, Baxter, y Beaudry, 2017; Rocchi, Pelletier, y Desmarais, 2017) con el mismo número de dimensiones, aunque distintos ítems. El cuestionario desarrollado para ser aplicado en el campo del deporte por Rocchi, Pelletier y Desmarais (2017), tiene como punto de partida el cuestionario más general sobre estilo interpersonal de Rocchi, Pelletier, Cheung et al. (2017) aplicado a otros contextos no deportivos. Más recientemente se ha desarrollado otro instrumento que además del apoyo y la frustración de estas necesidades también contempla un estilo interpersonal diferente: la indiferencia (Bhavsar et al., 2019). No obstante, la evidencia aportada en el trabajo tampoco es muy prometedora.

En referencia a las limitaciones encontradas al intentar buscar soluciones metodológicas para poder examinar la teoría de la SDT en

personas mayores que practican actividad física, cabe destacar que no se han encontrado soluciones para todas las cuestiones planteadas.

En este sentido, aunque se proporciona el coeficiente omega categórico como estimación de fiabilidad de consistencia interna, no se ha podido calcular el intervalo de confianza para el coeficiente omega categórico ya que, con los paquetes disponibles actualmente en R, ni en otro software, no se dispone de este intervalo si la escala no es unidimensional (Viladrich, Angulo-Brunet, 2019; Viladrich, Angulo-Brunet y Doval, 2019).

Debido a la dicotomización, tampoco se han podido examinar posibles efectos multinivel ya que las técnicas estadísticas disponibles para ello, como puede ser el estudio de la correlación intraclase y posterior aplicación del efecto de diseño multinivel (Hox, Moerbeek, y Schoot, 2018)

En los últimos años muchos han sido los esfuerzos para el desarrollo de metodología en los casos donde no existe una distribución Normal o aproximadamente Normal. Es importante que estos esfuerzos no disminuyan ya que mucha de la investigación aplicada, ya sea en psicología de la actividad física y del deporte como en otros campos afines, trabajan a menudo con escalas ordinales de este tipo constituyendo en muchas ocasiones un fuerte condicionante en la metodología a utilizar.

En referencia a las fortalezas de este estudio es importante tener en cuenta que la mayoría de las investigaciones donde se desarrollan medidas de un solo ítem se centran en el desarrollo de un

solo reactivo. En esta investigación se ha pretendido desarrollar múltiples ítems teniendo en cuenta que un punto central en el desarrollo de medidas es la relación con otras variables y su red nomológica de relaciones. En el contexto de personas mayores que desarrollan actividades en grupo es importante obtener medidas de un solo ítem que cubran todas las dimensiones de interés.

En este estudio se han examinado de forma exhaustiva las diversas opciones analíticas disponibles para poder describir a la población de interés con la máxima garantía de calidad psicométrica. Para ello, el diseño y resultados parciales de este trabajo han sido divulgados en diversos congresos nacionales e internacionales. Durante el desarrollo de las medidas se presentaron los resultados en forma de poster en el XV Congreso de Metodología de las Ciencias Sociales y de la Salud (Apéndice C4; Angulo-Brunet, Viladrich, y Pallarès, 2017). También se ha examinado la posibilidad de tratar la escala de respuestas original de cinco puntos para el análisis. Este es el caso del trabajo presentado en el congreso de la *International Test Commission* (Apéndice C5; Angulo-Brunet, Viladrich, Ramis, y Pallarès, 2018) donde se examinaron las similitudes entre dos modelos de *correlation network analysis* llevando a resultados no concluyentes. O también es el caso del trabajo presentado en el congreso de la *European Association of Methodology* (Apéndice C6; Angulo-Brunet, Viladrich, Pallarès, Ramis, y Borrueco, 2018) donde se presentaron resultados de la estabilidad temporal de las medidas de un solo ítem, examinando los *correlation network analysis* en la fase test y en la fase retest y

examinando también su fiabilidad test-retest. En estos dos últimos casos, se puso en evidencia la falta de potencia para hacer análisis multivariados más complejos.

7

CONCLUSIONES

Las conclusiones de esta tesis se han organizado siguiendo los objetivos, por ello en primer lugar se describirán las características de las personas mayores de sesenta años que practican actividad física dirigida. En segundo lugar, se mostrarán las principales evidencias cuantitativas de calidad psicométrica para cuestionarios ya utilizados previamente en otras poblaciones (i.e., CIS-Q, BPNES, BREQ y SVS). Finalmente se concluye con los principales hallazgos cualitativos y cuantitativos durante el desarrollo y obtención de evidencias de calidad psicométrica de las medidas de un solo ítem.

Se puede decir que las personas mayores de sesenta años que han participado en esta investigación están caracterizadas por percibir un gran apoyo y satisfacción, y baja frustración, de sus necesidades psicológicas básicas de competencia y relación. Perciben también un buen apoyo y satisfacción de la autonomía y en cambio la percepción de la frustración a la autonomía es más variable. Las personas que han participado practican actividad física motivadas autónomamente y en general se sienten vitales.

En referencia al objetivo planteado con las medidas de múltiples ítems, a continuación, se concluye sobre las evidencias cuantitativas basadas en la estructura interna y consistencia interna de cada cuestionario, uno para cada uno de los bloques de la SDT.

Estilo interpersonal (CIS-Q). Se ha proporcionado evidencia favorable relacionada con la estructura interna para el apoyo a las tres necesidades psicológicas básicas, aunque los tres factores muestran elevadas correlaciones entre sí. Los coeficientes de fiabilidad son adecuados para apoyo a la competencia y apoyo a la relación. En cambio, no existe evidencia favorable de fiabilidad para apoyo a la autonomía. En relación con la frustración a las necesidades psicológicas básicas no se ha podido estimar los parámetros del modelo de medida por lo que tampoco se ha examinado la fiabilidad.

Satisfacción de las necesidades psicológicas básicas (BPNES). El modelo propuesto de tres factores correlacionados ha resultado adecuado. Igual que el modelo de apoyo a las necesidades psicológicas básicas, también ha mostrado problemas de discriminación entre los factores. De las tres dimensiones, satisfacción de la relación ha mostrado un coeficiente de fiabilidad de consistencia interna adecuado, mientras que la fiabilidad de las otras escalas variables no ha sido aceptable.

Regulación conductual (BREQ-3). Ninguno de los modelos compatibles con la SDT ha sido corroborado por los datos. Por ello tampoco se ha obtenido fiabilidad de consistencia interna.

Vitalidad subjetiva (SVS). Se ha proporcionado evidencia favorable para un factor único de vitalidad subjetiva y el coeficiente de fiabilidad de consistencia interna es adecuado.

Para cubrir el objetivo de proveer medidas de un solo ítem para los constructos motivacionales de la SDT se proponían tres objetivos específicos sobre los que se concluye a continuación.

El primer objetivo fue examinar la factibilidad de la recogida de datos y el proceso de respuesta a los ítems de las medidas de un solo ítem en comparación con las medidas de múltiples ítems. Una conclusión a destacar es que mientras con las medidas de múltiples ítems no ha sido posible realizar una recogida de datos grupal si que ha sido posible hacerlo con las medidas de un solo ítem. Otro punto importante ha sido el hecho de que independientemente de si se utilizaban las medidas de múltiples ítems o las medidas de un solo ítem, la escala propuesta tipo Likert de cinco puntos no ha resultado adecuada para esta población, por lo que finalmente se ha dicotomizado la escala de respuesta.

En relación con el segundo y tercer objetivo de obtener evidencias de calidad psicométrica para las medidas de un solo ítem, se comentan las conclusiones a continuación. Para ello se ha conservado el orden de los conceptos en la cadena de relaciones de la SDT.

Estilo interpersonal. Se han desarrollado seis ítems, uno por cada subescala del CIS-Q. Todos los ítems han sido elegidos de ítems ya existentes del CIS-Q, excepto frustración a la competencia y frustración a la relación, que han sido desarrollados totalmente de nuevo.

Mediante las evidencias de validez y fiabilidad aportadas se puede decir que las medidas de apoyo a las necesidades psicológicas básicas han mostrado buen desempeño. Todos estos ítems han sido considerados relevantes, suficientes y claros por parte del panel de especialistas. A pesar de ello, estos ítems referidos al monitor o a la monitora de las sesiones son los que han presentado resultados más comprometidos al analizar el proceso de respuesta debido a la voluntad manifestada por las personas participantes de no perjudicar a su líder. Los coeficientes de correlación observados entre los seis ítems y también con el resto de las variables estudiadas han sido congruentes con las expectativas teóricas: los tres ítems de apoyo a las necesidades psicológicas básicas correlacionaron positivamente entre sí, y también con los ítems que medían motivación autónoma y vitalidad subjetiva. Por otra parte, los tres ítems de frustración de las necesidades psicológicas básicas han correlacionado positivamente entre sí y con las medidas de motivación controlada y amotivación y negativamente con el resto de los ítems. Ello constituye un primer paso de la obtención de evidencia favorable derivada del estudio de la red nomológica de relaciones entre estos ítems y el resto de los constructos de la SDT analizados en esta tesis.

La validez concurrente de los tres ítems de apoyo a las necesidades psicológicas básicas se ha sustentado en los coeficientes de correlación elevados observados con sus medidas homólogas de múltiples ítems. En cuanto a la fiabilidad, los tres ítems de apoyo a las necesidades psicológicas básicas pueden considerarse fiables con la

excepción de los indicadores derivado de la medida de múltiples ítems de apoyo a la autonomía que tampoco había resultado fiable.

Satisfacción de las necesidades psicológicas básicas. De los tres ítems desarrollados dos fueron elegidos del cuestionario BPNES y el ítem de satisfacción de la autonomía fue totalmente nuevo. Al igual que los ítems de estilo interpersonal, estos ítems han sido considerados relevantes, suficientes y claros por parte del panel de personas expertas. Durante la recogida de datos se ha puesto de manifiesto, que en muchas ocasiones no se distinguía entre apoyo y satisfacción de las necesidades psicológicas básicas ya que las personas expresaban tener la sensación de haber contestado reiteradamente a lo mismo. En referencia a la validez concurrente, tanto satisfacción de la competencia como satisfacción de la relación han mostrado una alta correlación con su medida de múltiples ítems homóloga. En cambio, para la medida de satisfacción a la autonomía esta relación ha sido débil.

En la misma línea que apoyo a las necesidades psicológicas básicas, las relaciones entre los ítems de satisfacción han sido congruentes con las expectativas planteadas en la SDT. Además de las relaciones ya comentadas, se han observado relaciones positivas con motivación autónoma y vitalidad, y relaciones negativas o nulas con motivación controlada y amotivación. Todos los coeficientes de fiabilidad obtenidos para satisfacción de la relación han sido favorables. En cambio, para satisfacción de la competencia solo ha sido favorable el coeficiente de fiabilidad test-retest. Por el contrario, el ítem de

satisfacción de la autonomía no mostró evidencia de fiabilidad favorable para su uso.

Regulación conductual. Los seis ítems seleccionados han sido considerados claros, relevantes y suficientes por el panel de especialistas. Durante el proceso de respuesta se ha evidenciado la dificultad de diferenciar entre conceptos, particularmente el de regulación introyectada.

En referencia a la red nomológica de relaciones, tal y como contempla la SDT se han observado relaciones positivas entre los ítems de la dimensión teórica de motivación autónoma y relaciones negativas o nulas entre los ítems de motivación controlada. Además de las relaciones que ya han sido comentadas se observa una relación positiva entre los ítems de motivación autónoma y vitalidad y relaciones negativas o nulas con los de motivación controlada y amotivación con vitalidad.

Vitalidad subjetiva. La medida desarrollada fue creada por el primer panel de especialistas y fue considerada adecuada por el segundo panel. Tal y como se ha visto en los párrafos anteriores las relaciones del ítem de vitalidad subjetiva con el resto de las medidas de un solo ítem ha sido congruente con las expectativas teóricas de la SDT. Además, se mostró evidencia de validez concurrente positiva mediante la correlación con su medida de múltiples ítems homóloga. La fiabilidad derivada de la fórmula de Spearman-Brown, de la aplicación de la fórmula de corrección por atenuación y de la comunalidad fue adecuada. A pesar de que el porcentaje de acuerdo fue elevado entre

el test y el retest, la evidencia de fiabilidad aportada mediante el coeficiente PABAK es débil.

8

REFERENCIAS

- AERA, APA, y NCME. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- AERA, APA, y NCME. (2014). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
- Alcaraz, S., Pons, J., Angulo-Brunet, A., y Cruz, J. (2017). Intervenciones para la promoción de actividad física basadas en la teoría de la autodeterminación: Una revisión narrativa (2011-2016). *Revista de Psicología del Deporte*, 26, 116-123.
- Alcaraz, S., Viladrich, C., y Torregrosa, M. (2013). Less time, better quality. Shortening questionnaires to assess team environment and goal orientation. *The Spanish Journal of Psychology*, 16, 1-14. doi: 10.1017/sjp.2013.74
- Alessandri, G., Zuffianò, A., y Perinelli, E. (2017). Evaluating intervention programs with a pretest-posttest design: A structural equation modeling approach. *Frontiers in Psychology*, 8, 1-12. doi: 10.3389/fpsyg.2017.00223

- American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC, DC: American Psychological Association.
- Andersen, R., Hageaars, J. A., y McCutcheon, A. L. (2006). Applied latent class analysis. *Canadian Journal of Sociology / Cahiers canadiens de sociologie*, 28(4), 584. doi: 10.2307/3341848
- Angulo-Brunet, A., Sanchez-Oliva, D., y Viladrich, C. (2017). How Coaches' Interpersonal Style Questionnaire (CIS-Q) became Physical Education Teachers' Interpersonal Style Questionnaire (PETIS-Q). ISSP Sevilla.
- Angulo-Brunet, A., y Viladrich, C. (2017a). SubscaleExplorer: Psychometric scale explorer and management. R package version 0.1.1. Recuperado de www.github.com/AnguloB/SubscaleExplorer
- Angulo-Brunet, A., y Viladrich, C. (2017b). Un paquete de R para la gestión y descripción de subescalas psicométricas. En *XV Congreso de Metodología de las Ciencias Sociales y de la Salud*. Barcelona.
- Angulo-Brunet, A., y Viladrich, C. (2018). How to handle ordered responses with floor and ceiling effects in SEM using small samples? En *Small Sample Size Solutions*. Utrecht, The Netherlands.
- Angulo-Brunet, A., y Viladrich, C. (2019). ¿Limitaciones o cambio de objetivos? Lecciones aprendidas comparando dos opciones para reportar medidas con baja consistencia interna. En *XVI Congreso*

- de Metodología de las Ciencias Sociales y de la Salud*. Madrid.
- Angulo-Brunet, A., Viladrich, C., y Pallarès, S. (2017). Desarrollo de medidas de un ítem para evaluar los resultados de una intervención percibidos por personas mayores. En *XV Congreso de Metodología de las Ciencias Sociales y de la Salud*. Barcelona.
- Angulo-Brunet, A., Viladrich, C., Pallarès, S., Ramis, Y., y Borrueco, M. (2018). Validity of single-item measures to assess self-determination theory constructs in the elderly. En *European Association of Methodology*. Jena, Alemania.
- Angulo-Brunet, A., Viladrich, C., Ramis, Y., y Pallarès, S. (2018). Single-item measures validity evidence based on correlation network analysis: motives for physical activity in the elderly. En *International Test Commission*. Montreal, Canada.
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., y Rao, S. M. (2018). Reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist*, *1*(2), 26-46. doi: 10.1037/amp0000151
- Appleton, P. R., Ntoumanis, N., Quested, E., Viladrich, C., y Duda, J. L. (2016). Initial validation of the coach-created Empowering and Disempowering Motivational Climate Questionnaire (EDMCQ-C). *Psychology of Sport y Exercise*, *22*, 53-65. doi: 10.1016/j.psychsport.2015.05.008
- Bergkvist, L. (2015). Appropriate use of single-item measures is here to stay. *Marketing Letters*, *26*(3), 245-255. doi: 10.1007/s11002-

014-9325-y

- Bergkvist, L. (2016). The nature of doubly concrete constructs and how to identify them. *Journal of Business Research*, 69(9), 3-4. doi: 10.1016/j.jbusres.2016.02.001
- Bergkvist, L., y Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44(2), 175-184. doi: 10.1509/jmkr.44.2.175
- Bhavsar, N., Ntoumanis, N., Quested, E., Gucciardi, D. F., Thøgersen-Ntoumani, C., Ryan, R. M., ... Bartholomew, K. J. (2019). Conceptualizing and testing a new tripartite measure of coach interpersonal behaviors. *Psychology of Sport and Exercise*. doi: 10.1016/j.psychsport.2019.05.006
- Breslin, G., Shannon, S., Fitzpatrick, B., Hanna, D., Belton, S., y Brennan, D. (2017). Physical activity, well-being and needs satisfaction in eight and nine-year-old children from areas of socio-economic disadvantage. *Child Care in Practice*, 23(3), 275-291. doi: 10.1080/13575279.2017.1299108
- Bruton, A. M., Mellalieu, S. D., y 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. doi: 10.1080/1612197X.2015.1054853
- Buuren, S. van, y Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1-68.

- Byrt, T., Bishop, J., y Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5), 423-429. doi: 10.1016/0895-4356(93)90018-V
- Castillo, I., Tomás, I., y 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. doi: 10.1017/sjp.2017.22
- Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., ... Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, 39(2), 216-236. doi: 10.1007/s11031-014-9450-1
- Cheon, S. H., Reeve, J., y Moon, I. S. (2012). Experimentally based, longitudinally designed, teacher-focused intervention to help physical education teachers be more autonomy supportive toward their students. *Journal of Sport and Exercise Psychologyexercise psychology*, 34(3), 365-396.
- Cheung, F., y Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: results from three large samples. *Quality of Life Research*, 23(10), 2809-2818. doi: 10.1007/s11136-014-0726-4
- Cid, L., Monteiro, D., Teixeira, D., Teques, P., Alves, S., Moutão, J., ... Palmeira, A. (2018). The Behavioral Regulation in Exercise Questionnaire (BREQ-3) Portuguese-Version: Evidence of Reliability, Validity and Invariance Across Gender. *Frontiers in*

psychology, 9. doi: 10.3389/fpsyg.2018.01940

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*.

New York, NY: NY: Routledge Academic.

Couto, N., Antunes, R., Monteiro, D., Moutão, J., Marinho, D. A., y Cid,

L. (2017). Impact of the basic psychological needs in subjective

happiness, subjective vitality and physical activity in an elderly

Portuguese population. *Motricidade*, 13(2), 58-70. doi:

10.6063/motricidade.9746

Cronbach, L. J., y Meehl, P. E. (1955). Construct validity in

psychological tests. *Psychological Bulletin*, 52(4), 281.

Dacey, M. L., y Newcomer, A. R. (2005). A client-centered counseling

adults toward physical activity. *Topics in Geriatric Rehabilitation*,

21(3), 194-205.

Deci, E. L., y Ryan, R. M. (1985). *Intrinsic motivation and Self-*

determination in human behavior. New York: Plenum.

Deci, E. L., Vallerand, R. J., Pelletier, L., y Ryan, R. M. (1991).

Motivation and education: The Self-determination perspective.

Educational Psychologist, 26(3), 325-346. doi:

10.1207/s15326985ep2603

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., y Kaiser, S.

(2012). Guidelines for choosing between multi-item and single-

item scales for construct measurement: A predictive validity

perspective. *Journal of the Academy of Marketing Science*, 40(3),

434-449. doi: 10.1007/s11747-011-0300-3

Dillman, D. A., Sinclair, M. D., y Clark, J. R. (1993). Effects of

- Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. *Public Opinion Quarterly*, 57(3), 289-304.
- Epskamp, S., Angelique, O. J., Cramer, L. ., Waldorp, V. D., y Chmittleman, D. B. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, 48(4), 1-18.
- Epskamp, S., Borsboom, D., y Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195-212. doi: 10.3758/s13428-017-0862-1
- Fabra, P., Balaguer, I., Tomás, I., Smith, N., y Duda, J. L. (2018). Versión española del Sistema de Observación del Clima Motivacional Multidimensional (MMCOS): fiabilidad y evidencias de validez. *Revista de Psicología del Deporte*, 27(1), 11-22.
- Fuchs, C., y Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, 69(2), 195-211.
- Gagné, M., y Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4), 331-362. doi: 10.1002/job.322
- Gnambs, T., y Buntins, K. (2017). The measurement of variability and change in life satisfaction: A comparison of single-item and multi-item instruments. *European Journal of Psychological*

- Assessment*, 33(4), 224-238. doi: 10.1027/1015-5759/a000414
- Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., ...
Preckel, F. (2014). "My Questionnaire is Too Long!" The
assessments of motivational-affective constructs with three-item
and single-item measures. *Contemporary Educational
Psychology*, 39(3), 188-205. doi:
10.1016/j.cedpsych.2014.04.002
- González-Cutre, D., Sicilia, Á., y Fernández, A. (2010). Hacia una mayor
comprensión de la motivación en el ejercicio físico : medición de
la regulación integrada en el contexto español. *Psicothema*, 22,
841-847.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real
world. *Annual Review of Psychology*, 60, 549-576. doi:
10.1146/annurev.psych.58.110405.085530
- Gray, S., Wharf, J., y 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.
- Green, S. B., y Yang, Y. (2009). Reliability of summed item scores using
structural equation modeling: An alternative to coefficient alpha.
Psychometrika, 74(1), 155-167. doi: 10.1007/s11336-008-9099-3
- Guttman, L. (1945). A basis for analyzing test-retest reliability.
Psychometrika, 10(4), 255-282. doi: 10.1007/BF02288892
- Hayduk, L. A., y Littvay, L. (2012). Should researchers use single
indicators, best indicators, or multiple indicators in structural

- equation models? *BMC Medical Research Methodology*, 12(1), 159. doi: 10.1186/1471-2288-12-159
- Henning, G., Bjälkebring, P., Stenling, A., Thorvaldsson, V., Johansson, B., y Lindwall, M. (2019). Changes in within- and between-person associations between basic psychological need satisfaction and well-being after retirement. *Journal of Research in Personality*, 79, 151-160. doi: 10.1016/j.jrp.2019.03.008
- Hox, J. J., Moerbeek, M., y Schoot, R. (2018). *Multilevel analysis: Techniques and Applications*. Third edition. Routledge, New York.
- Hu, L., y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi: 10.1080/10705519909540118
- Hubley, A. M., y Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, 103(2), 219-230. doi: 10.1007/s11205-011-9843-4
- Jordan, J. S., y Turner, B. A. (2008). The Feasibility of Single-Item Measures for Organizational Justice. *Measurement in Physical Education and Exercise Science*, 12(4), 237-257. doi: 10.1080/10913670802349790
- Jöreskog, K. (2005). Structural equation modeling with ordinal variables using LISREL, 1, 1-77. Recuperado de http://www.ams.sunysb.edu/~zhu/tmp/Yue/SEM_brain/covariate/covariate2/SEM with ordinal variables using LISREL.pdf
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., y Rosseel, Y.

- (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-0.
- Jovanović, V., y Lazić, M. (2018). Is longer always better? A comparison of the validity of single-item versus multiple-item measures of life satisfaction. *Applied Research in Quality of Life*. doi: 10.1007/s11482-018-9680-6
- Kamakura, W. A. (2015). Measure twice and cut once: the carpenter's rule still applies. *Marketing Letters*, 26(3), 237-243. doi: 10.1007/s11002-014-9298-x
- Kenny, D. A., y Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*. US: American Psychological Association. doi: 10.1037/0022-006X.63.1.52
- Kenny, D. A., y Zautra, A. (2001). Trait–state models for longitudinal data. *New methods for the analysis of change*. Washington, DC, US: American Psychological Association. doi: 10.1037/10409-008
- Kirkland, R. A., Karlin, N. J., Stellino, M. B., y Pulos, S. (2011). Basic psychological needs satisfaction, motivation, and exercise in older adults. *Activities, Adaptation and Aging*, 35(3), 181-196. doi: 10.1080/01924788.2011.596764
- Konstabel, K., Lönnqvist, J.-E., Walkowitz, G., Konstabel, K., y Verkasalo, M. (2012). The 'Short Five' (S5): Measuring personality traits using comprehensive single items. *Journal of Personality*, 26(1), 68-84. doi: 10.1002/per
- Kruyen, P. M., Emons, W. H. M., y Sijtsma, K. (2013). On the

- shortcomings of shortened tests: a literature review. *International Journal of Testing*, 13(3), 223-248. doi: 10.1080/15305058.2012.703734
- Kruyen, P. M., Emons, W. H. M., y Sijtsma, K. (2014). Assessing individual change using short tests and questionnaires. *Applied Psychological Measurement*, 38(3), 201-216. doi: 10.1177/0146621613510061
- Kwon, H., y Trail, G. (2005). The feasibility of single item measures in sport loyalty research. *Sport Management Review*, 8, 69-88. doi: 10.1016/S1441-3523(05)70033-4
- Landis, J. R., y Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. doi: 10.2307/2529310
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction*, 362-412.
- Li, Z., Banerjee, J., y Zumbo, B. D. (2017). Response time data as validity evidence: Has it lived up to its promise and, if not, what would it take to do so. En *Understanding and Investigating Response Processes in Validation Research* (pp. 159-177). Springer.
- Little, T. D., Cunningham, W. A., Shahar, G., y Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151-173. (doi:10.1207/S15328007SEM0902_1).

- Liu, Y., Millsap, R. E., West, S. G., Tein, J., Tanaka, R., y Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486-506. doi: 10.1038/nrg3575.Systems
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. doi: 10.6018/analesps.30.3.199361
- Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of Managerial Psychology*, 17(1), 68-75. doi: 10.1108/02683940210415933
- Lozano, L. M., García-Cueto, E., y Muñoz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79. doi: 10.1027/1614-2241.4.2.73
- Lucas, R. E., y Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: results from four national panel studies. *Social Indicators Research*, 105(3), 323-331. doi: 10.1007/s11205-011-9783-z
- Maillot, P., Becquard, C., y Perrot, A. (2018). Motivation des personnes âgées à pratiquer une activité physique. *Gérontologie et société*, 40(2), 127. doi: 10.3917/ges1.156.0127
- Marcos, P. J., Orquín, F. J., Pedreño, N. B., y Moreno-Murcia, J. A. (2014). Motivación autodeterminada en adultos mayores

- practicantes de ejercicio físico. *Cuadernos de Psicología del Deporte*, 14(3), 149-155.
- Markland, D., y Tobin, V. (2004). A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *Journal of Sport and Exercise Psychology*, 26(2), 191-196.
- Marsh, H. W., Hau, K.-T., y Wen, Z. (2004). In search 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, 320-341.
- McCutcheon, A. L. (1987). *Latent Class Analysis*. University of Nebraska, Lincoln, USA: SAGE.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45(1-3), 35-44. doi: 10.2307/27522333
- Milton, K., Clemes, S., y Bull, F. (2013). Can a single question provide an accurate measure of physical activity? *British Journal of Sports Medicine*, 47(1), 44-48. doi: 10.1136/bjsports-2011-090899
- Mokken, R. J. (1971). A Theory And Procedure Of Scale Analysis: White applications in Political Research (Vol. 1). *Netherlands: Mouton and Co*.
- Neubauer, A. B., Schilling, O. K., y Wahl, H. W. (2017). What do we

need at the end of life? Competence, but not autonomy, predicts intraindividual fluctuations in subjective well-being in very old age. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 72(3), 425-435. doi: 10.1093/geronb/gbv052

Ntoumanis, N., Quested, E., Reeve, J., y Cheon, S. H. (2017). Need supportive communication: Implications for motivation in sport exercise and physical activity. En B. Jackson y J. C. James Dimmock (Eds.), *Persuasion and Communication in Sport, Exercise, and Physical Activity* (Routledge, pp. 155-169). Abingdon, UK.

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

Padilla, J. L., Gómez, J., Hidalgo, M. D., y Muñiz Fernández, J. (2006). La evaluación de las consecuencias del uso de los tests en la teoría de la validez. *Psicothema*, 18 (2).

Pallarès, S., Miró, S., Perez-Rivases, A., Torregrossa, M., Ramis, Y., Cruz, J., y Viladrich, C. (en prensa). Creación de programas motivacionales para la actividad física de adolescentes, exdeportistas y adultos mayores. *Pensamiento Psicológico*

Penelo, E., De la Osa, N., Navarro, J. B., Domènech, J. M., y Ezpeleta, L. (2017). The Brief Problem Monitor-Parent Form (BPM-P), a short version of the child behavior checklist: Psychometric properties in spanish 6- to 8-year-old children. *Psychological Assessment*, 29(11), 1309-1320. doi: 10.1037/pas0000428

Pulido, J. J., Sánchez-Oliva, D., Leo, F. M., Sánchez-Cano, J., y García-

- Calvo, T. (2017). Development and validation of Coaches' interpersonal style questionnaire. *Measurement in Physical Education and Exercise Science*, 22(1) 25-37. doi: 10.1080/1091367X.2017.1369982
- Pulido, J. J., Sánchez-Oliva, D., Silva, M. N., Palmeira, A. L., y García-Calvo, T. (2019). Development and preliminary validation of the Coach Interpersonal Style Observational System. *International Journal of Sports Science and Coaching*, 0(0), 174795411985896. doi: 10.1177/1747954119858965
- Raykov, T., y Marcoulides, G. A. (2016). On the relationship between classical test theory and item response theory: From one to the other and back. *Educational and psychological measurement*, 76(2), 325-338.
- Raykov, T., Rodenberg, C., y Narayanan, A. (2015). Optimal shortening of multiple-component measuring instruments: A latent variable modeling procedure. *Structural Equation Modeling*, 22(2), 227-235. doi: 10.1080/10705511.2014.935927
- Revelle, W. (2018). psych: Procedures for personality and psychological research Version 1.8.4. Northwestern University, Evanston, Illinois, USA.
- Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., y Ponsoda, V. (2014). The impact of ambiguous response categories on the factor structure of the GHQ-12. *Psychological assessment*, 26(3), 1021.
- Robins, R. W., Hendin, H. M., y 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. doi: 10.1177/0146167201272002
- Rocchi, M., Pelletier, L., Cheung, S., Baxter, D., y Beaudry, S. (2017). Assessing need-supportive and need-thwarting interpersonal behaviours: The Interpersonal behaviours questionnaire (IBQ). *Personality and Individual Differences*, 104(January), 423-433. doi: 10.1016/j.paid.2016.08.034
- Rocchi, M., Pelletier, L., y Desmarais, P. (2017). The validity of the Interpersonal behaviors questionnaire (IBQ) in Sport. *Measurement in Physical Education and Exercise Science*, 21(1), 15-25. doi: 10.1080/1091367X.2016.1242488
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. doi: 10.18637/jss.v048.i02
- Rossiter, J. (2002). The C-OA-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 30. doi: 10.1016/S0167-8116(02)00097-6
- Ryan, R.M., y Deci, E.L. (2017). *Self-determination Theory: Basic Psychological Needs In Motivation, Development*. The Guildord Press, New York.
- Ryan, R.M., y Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, 55(1), 68-78. doi: 10.1037/0003-066X.55.1.68

- Ryan, R M, y Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annual Reviews*, 52, 141-166.
- Ryan, R. M, y Deci, E. L. (2008). A self-determination theory approach to psychotherapy: The motivational basis for effective change. En *Canadian Psychology* (Vol. 49, pp. 186-193). doi: 10.1037/a0012753
- Ryan, R.M, y Frederick, C. (1997). On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. *Journal of Personality*, 65(3), 529-565. doi: 10.1111/j.1467-6494.1997.tb00326.x
- Sanchez, J. M., y Nuñez, J. L. (2007). Análisis preliminar de las propiedades psicométricas de la versión española de la escala de necesidades psicológicas básicas en el ejercicio físico. *Revista Iberomericana de Psicología del Ejercicio y el Deporte*, 2, 83-92.
- Sarstedt, M., Diamantopoulos, A., y Salzberger, T. (2016). Should we use single items? Better not. *Journal of Business Research*, 69(8), 3199-3203. doi: 10.1016/j.jbusres.2016.02.040
- Savalei, V., y Reise, S. P. (2019). Don't Forget the Model in Your Model-based Reliability Coefficients: A Reply to McNeish (2018). *Collabra: Psychology*, 5(1), 36. doi: 10.1525/collabra.247
- Sibley, B. A., y Bergman, S. M. (2017). What keeps athletes in the gym? Goals, psychological needs, and motivation of CrossFit™ participants. *International Journal of Sport and Exercise Psychology*, 16(5), 555-574. doi:

10.1080/1612197X.2017.1280835

- Simms, L. J., Zelazny, K., Williams, T. F., y Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*. doi: 10.1037/pas0000648
- Solberg, P. A., Halvari, H., y Ommundsen, Y. (2013). Linking exercise and causality orientations to change in well-being among older adults: Does change in motivational variables play a role? *Journal of Applied Social Psychology*, 43(6), 1259-1272. doi: 10.1111/jasp.12088
- Spörrle, M., y Bekk, M. (2014). Meta-analytic guidelines for evaluating single-item reliabilities of personality instruments. *Assessment*, 21(3), 272-285. doi: 10.1177/1073191113498267
- Su, Y. L., y Reeve, J. (2011). A Meta-analysis of the Effectiveness of Intervention Programs Designed to Support Autonomy. *Educational Psychology Review*, 23(1), 159-188. doi: 10.1007/s10648-010-9142-7
- Viladrich, C., Angulo-Brunet, A. (2019). Reliability of essentially unidimensional measures derived from bifactor modeling with R, Mplus and Stata. En *15Th European Congress Of Sport and Exercise Psychology*, Münster, Alemania
- Viladrich, C., Angulo-Brunet, A., y Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Annals of Psychology*, 33(3), 755-782. doi: 10.6018/analesps.33.3.268401

- Viladrich, C., Angulo-Brunet y Doval, E. (2019). Recursos para calcular los coeficientes alfa y omega de fiabilidad de consistencia interna. En *XVI Congreso de Metodología de las Ciencias Sociales y de la Salud*. Madrid.
- Vlachopoulos, S. P., y 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. doi: 10.1207/s15327841mpee1003_4
- Wanner, M., Probst-Hensch, N., Kriemler, S., Meier, F., Bauman, A., y Martin, B. W. (2014). What physical activity surveillance needs: validity of a single-item questionnaire. *British Journal of Sports Medicine*, 48(21), 1570-1576. doi: 10.1136/bjsports-2012-092122
- Wanous, J. P., y Hudy, M. J. (2001). Single-item reliability: a replication and extension. *Organizational Research Methods*, 4(4), 361-375. doi: 10.1177/109442810144003
- Wanous, J. P., y Reichers, A. E. (1996). Estimating the reliability of a single-item measure. *Psychological Reports*, 78(1), 631-634.
- Wei, T., y Simko, V. (2017). R package «corrplot»: Visualization of a Correlation Matrix.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wilson, P., Rodgers, W. M., Loitz, C. C., y Scime, G. (2006). "It's Who I Am ... Really! The importance of Integrated regulation in

- exercise contexts. *Journal of Applied Biobehavioral Research*, 11(2), 79-104. doi: 10.1111/j.1751-9861.2006.tb00021.x
- Wilson, P., Paige, J., y Mack, D. (1982). The importance of interpersonal style in competitive sport: A Self- Determination Theory approach. *Handbook of Sports Psychology*, 1-21.
- Xia, Y., 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*, 1-20. doi: 10.3758/s13428-018-1055-2
- Yang, M., Viladrich, C., y 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).
- Yang, Y., 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. doi: 10.1177/0013164417752008
- Zazo, R., y Moreno-Murcia, J. A. (2015). Hacia el bienestar psicológico en el ejercicio físico acuático. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 10(1), 33-39.
- Zijlmans, E. A. O., Tijmstra, J., van der Ark, L. A., y Sijtsma, K. (2019). Item-score reliability as a selection tool in test construction. *Frontiers in Psychology*, 9(JAN), 1-12. doi: 10.3389/fpsyg.2018.02298
- Zijlmans, E. A. O., van der Ark, L. A., Tijmstra, J., y Sijtsma, K. (2018).

Methods for estimating item-score reliability. *Applied Psychological Measurement*, 42(7), 553-570. doi: 10.1177/0146621618758290

Zumbo, B. D., y Hubley, A. M. (2016). Bringing consequences and side effects of testing and assessment to the foreground. *Assessment in Education: Principles, Policy and Practice*, 23(2), 299-303. doi: 10.1080/0969594X.2016.1141169

Zwolinsky, S., McKenna, J., Pringle, A., Widdop, P., y Griffiths, C. (2015). Physical activity assessment for public health: Efficacious use of the single-item measure. *Public Health*, 129(12), 1630-1636. doi: 10.1016/j.puhe.2015.07.015

Apéndices

Apéndice A. Material utilizado en la recogida de datos

Apéndice A1. Consentimiento informado

Muestra 1

INFORMACION DE LA INVESTIGACION SOBRE MOTIVACION PERSONALIZADA PARA LA ADHESIÓN A LA ACTIVIDAD FÍSICA

Estimado/a,

Nuestro equipo de investigación del *Grup d'Estudis en Psicologia de l'Esport* de la *Universitat Autònoma de Barcelona* con colaboración del XXXXXXXX está llevando a cabo una investigación con el objetivo de *promocionar la salud mediante programas de actividad física basados en las necesidades de las personas (jóvenes, mayores y deportistas de élite...)*.

Este estudio nos ayudará a *desarrollar programas de formación para profesores y profesoras de Educación Física y para monitores y monitoras de Actividad Física, para que aumenten la motivación a participar en estas actividades a un mayor número de personas.*

Nos gustaría invitarte a participar en esta investigación. El estudio requerirá participar en una sesión menos de 1 hora de duración para hablar y contestar unos cuestionarios sobre estos temas.

Tu participación en este estudio es voluntaria y la información que aportes será estrictamente confidencial, así que, nadie verá tus respuestas o opiniones emitidas en situación de entrevista.

Si decides participar, queremos que sepas que podrás dejar de hacerlo en cualquier momento del proyecto, informando a la Dirección del Proyecto y sin que ello tenga ninguna consecuencia negativa para ti.

Una vez finalizado el estudio, el *Club Natació Sabadell* recibirá un informe con los resultados obtenidos y las recomendaciones adecuadas sin que figuren los nombres de los participantes. Serán datos generales referidos al colectivo al que tú perteneces de forma general. Agradeceríamos muchísimo tu participación en esta investigación.

Por favor, firma y entrega al responsable la hoja adjunta donde se expresa tu consentimiento. Si tienes alguna pregunta al respecto, puedes comunicárnoslo al investigador que está en la entrevista o al responsable que figura a continuación.

Atentamente,

Carme Viladrich Segués (Responsable del Proyecto)

Mail: xxxxxx@uab.cat

Tel.: XXXXXXXX

Consentimiento Informado

Título del proyecto: Promoción de la salud y de la adhesión a la actividad física a través de la Teoría de la Autodeterminación con el apoyo de aplicaciones móviles.

Yo, _____ (nombre y apellidos),

- He leído la hoja de información que se me ha entregado.
- He podido hacer preguntas sobre el estudio.
- He recibido suficiente información sobre el estudio.
- He hablado con el investigador

_____ (nombre del investigador)

- Comprendo que mi participación es voluntaria.
- Comprendo que puedo retirarme del estudio:

1. Cuando quiera.
2. Sin tener que dar explicaciones.
3. Sin que esto tenga para mí ninguna repercusión.

Presto libremente mi conformidad para participar en el estudio.

Firma del participante

Firma del investigador

Muestra 2

Estimado/a,

Nuestro equipo de investigación del *Grup d'Estudis en Psicologia de l'Esport* de la *Universitat Autònoma de Barcelona* con colaboración de XXXXXXXX está llevando a cabo una investigación con el objetivo de *promocionar la salud mediante programas de actividad física basados en las necesidades de las personas (jóvenes, mayores y deportistas de élite...)*.

Este estudio nos ayudará a *desarrollar programas de formación para profesores y profesoras de Educación Física y para monitores y monitoras de Actividad Física, para que aumenten la motivación a participar en estas actividades a un mayor número de personas*.

Nos gustaría invitarte a participar en esta investigación. El estudio requerirá que contestes un cuestionario hoy y otro de aquí aproximadamente un mes.

Tu participación en este estudio es voluntaria y la información que aportes será estrictamente confidencial, así que, nadie verá tus respuestas o opiniones emitidas en el cuestionario.

Si decides participar, queremos que sepas que podrás dejar de hacerlo en cualquier momento, informando al responsable del Proyecto y sin que ello tenga ninguna consecuencia negativa para ti.

Una vez finalizado el estudio, XXXXXXXXX recibirá un informe con los resultados obtenidos y las recomendaciones adecuadas sin que figuren los nombres de los participantes. Serán datos generales referidos al colectivo al que tú perteneces de forma general. Agradeceríamos muchísimo tu participación en esta investigación.

Por favor, firma y entrega al responsable la hoja adjunta donde se expresa tu consentimiento. Si tienes alguna pregunta al respecto, puedes comunicárnoslo al investigador que está en la entrevista o al responsable que figura a continuación.

Atentamente,

Carme Viladrich Segués (Responsable del Proyecto)

Mail: xxxxxxx@uab.cat

Tel.: XXXXXXXXX

**Consentimiento
Informado**

Título del proyecto: Promoción de la salud y de la adhesión a la actividad física a través de la Teoría de la Autodeterminación con el apoyo de aplicaciones móviles.

Yo, _____ (nombre y apellidos),

- He leído la hoja de información que se me ha entregado.
- He podido hacer preguntas sobre el estudio.
- He recibido suficiente información sobre el estudio.
- He hablado con el investigador

_____ (nombre del investigador)

- Comprendo que mi participación es voluntaria.
- Comprendo que puedo retirarme del estudio:

1. Cuando quiera.
2. Sin tener que dar explicaciones.
3. Sin que esto tenga para mí ninguna repercusión.

Presto libremente mi conformidad para participar en el estudio.

Firma del participante

Firma del investigador

Apéndice A2. Medidas de un solo ítem

La versión presentada en este Apéndice es la versión en castellano teniendo en cuenta que la persona que dirige la sesión es una mujer monitora. La versión en catalán y la versión para monitores está disponible bajo demanda.

NOMBRE PSICOMÉTRICO	CONTENIDO DEL ÍTEM
AA	Pienso que la monitora de Actividad Física tiene en cuenta nuestras aportaciones (opiniones, propuestas...) en el desarrollo de las sesiones.
AC	Pienso que la monitora de Actividad Física nos anima a que confiemos en nuestra capacidad para hacer bien los ejercicios.
AR	Pienso que la monitora de Actividad Física favorece el buen ambiente y las relaciones tanto entre los/las compañero/as de clase como con ella.
FA	Pienso que la monitora de Actividad Física nos exige hacer las cosas de una manera concreta.
FC	Pienso que la monitora de Actividad Física nos propone actividades demasiado difíciles.
FR	Pienso que la monitora de Actividad Física genera un ambiente que dificulta las buenas relaciones tanto entre los/las compañeros/as de clase como con ella.
AUT	En las sesiones de Actividad Física siento que puedo realizar los ejercicios de la manera en la que prefiero hacerlos.
COM	En las sesiones de Actividad Física pienso que puedo cumplir con las exigencias de las actividades.
REL	En las sesiones de Actividad Física me siento a gusto con los/las compañeros/as y con la monitora.
INT	Participo en las sesiones de Actividad Física porque disfruto con las actividades que realizo.
ITG	Participo en las sesiones de Actividad Física porque considero que el ejercicio físico está de acuerdo con mis valores.
IDE	Participo en las sesiones de Actividad Física porque valoro los beneficios del ejercicio físico.
ITY	Participo en las sesiones de Actividad Física porque me siento culpable cuando no practico ejercicio físico.
EXT	Participo en las sesiones de Actividad Física porque los demás dicen que debo hacerlo.
AMO	Participo en las sesiones de Actividad Física, pero pienso que es una pérdida de tiempo.
VI	Generalmente me siento vital .

Apéndice A3. Medidas de múltiples ítems

Las medidas de múltiples ítems utilizadas en este trabajo son adaptaciones de los cuestionarios que se detallan a continuación, teniendo en cuenta el contexto y la edad de las personas destinatarias. En las siguientes páginas puede encontrarse el contenido de los cuestionarios junto al nombre psicométrico asignado a los ítems. La versión presentada en este Apéndice es la versión en castellano teniendo en cuenta que la persona que dirige la sesión es una mujer (monitora). La versión en catalán y la versión para monitores está disponible bajo demanda.

Estilo interpersonal

Pulido, J. J., Sánchez-Oliva, D., Leo, F. M., Sánchez-Cano, J., y García-Calvo, T. (2017). Development and validation of Coaches' interpersonal style questionnaire. *Measurement in Physical Education and Exercise Science*, 22(1) 25-37. doi: 10.1080/1091367X.2017.1369982

Satisfacción de las necesidades psicológicas básicas

Sanchez, J. M., y Nuñez, J. L. (2007). Análisis preliminar de las propiedades psicométricas de la versión española de la escala de necesidades psicológicas básicas en el ejercicio físico. *Revista Iberomericana de Psicología del Ejercicio y el Deporte*, 2, 83-92.

Regulación conductual

González-Cutre, D., Sicilia, Á., y Fernández, A. (2010). Hacia una mayor comprensión de la motivación en el ejercicio físico: medición de la regulación integrada en el contexto español. *Psicothema*, 22, 841-847.

Vitalidad

Castillo, I., Tomás, I., y 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. doi: 10.1017/sjp.2017.22

Estilo interpersonal

Pienso que mi monitora...

NOMBRE PSICOMÉTRICO	CONTENIDO DEL ÍTEM
ac4	... fomenta el aprendizaje y la mejora de los contenidos del programa.
ac1	... nos anima a que confiemos en nuestra capacidad para hacer bien las tareas.
fc1	... nos propone tareas y situaciones que nos hacen sentir torpes.
fc2	... nos propone situaciones que nos hacen sentir incapaces.
aa2	... trata de que tengamos libertad a la hora de realizar las actividades.
fc4	...no nos da oportunidades para demostrar nuestro potencial.
ac2	... nos propone actividades ajustadas a nuestro nivel.
fc3	... a veces nos hace sentir incompetentes.
aa3	... tiene en cuenta nuestra opinión en el desarrollo de las sesiones
ac3	... siempre intenta que consigamos los objetivos que se plantean en las actividades.
ar2	... favorece el buen ambiente entre los compañeros/as del grupo.
fr1	... en ocasiones es indiferente con nosotros.
aa1	... nos pregunta a menudo sobre nuestras preferencias con respecto a las actividades a realizar.
fa1	... nos exige hacer las cosas de una manera concreta.
fa2	... nos impide tomar decisiones respecto al modo en que hacemos los ejercicios.
fa3	... nos obliga a aceptar una forma de hacer los ejercicios que no compartimos.
ar1	... fomenta en todo momento las buenas relaciones entre los participantes del programa.
fa4	... fuerza a seguir una determinada forma de hacer los ejercicios.
aa4	... nos deja tomar decisiones durante el desarrollo de las tareas.
ar3	... promueve que todos los participantes nos sintamos integrados.
fr2	... hace que nos sintamos poco aceptados/as en este grupo.
fr3	... en algunas ocasiones nos hace sentir rechazados/as.
ar4	... nos ayuda a resolver los conflictos amistosamente.
fr4	... crea un ambiente que no nos agrada.

Satisfacción necesidades psicológicas básicas

En las sesiones de actividad física...

NOMBRE PSICOMÉTRICO	CONTENIDO DEL ÍTEM
sa1	... los ejercicios que realizo se ajustan a mis intereses.
sc1	... siento que he tenido una gran progresión.
sr1	... me siento muy cómodo/a cuando hago ejercicio con mis compañeros/as.
sa2	... la forma de realizar los ejercicios coincide perfectamente con la forma en que yo quiero hacerlos.
sc2	... realizo los ejercicios eficazmente.
sr2	... me relaciono de forma muy amistosa con el resto de compañeros/as.
sa3	... la forma de realizar los ejercicios responde a mis deseos.
sc3	... el ejercicio es una actividad que hago muy bien.
sr3	... siento que me puedo comunicar abiertamente con mis compañeros/as.
sa4	... tengo la oportunidad de elegir cómo realizar los ejercicios.
sc4	... pienso que puedo cumplir con las exigencias de la clase.
sr4	... me siento muy cómodo/a con los/as compañeros/as.

Regulación conductual

Participo en las sesiones de Actividad física...

NOMBRE PSICOMÉTRICO	CONTENIDO DEL ÍTEM
ide1	...porque hacer ejercicio regularmente es importante para mí.
amo1	...pero no veo por qué debería hacer ejercicio.
int1	...porque hacer ejercicio es divertido.
ity1	...porque me siento culpable cuando no hago ejercicio.
itg1	...porque es consistente con mis objetivos vitales.
ext1	...porque los demás dicen que tengo que hacerlo.
ide2	...porque valoro los beneficios del ejercicio.
amo2	...pero no veo por qué habría de molestarme en hacer ejercicio.
int2	...porque disfruto de mis sesiones de ejercicio.
ity2	...porque me siento mal cuando me pierdo una sesión de ejercicio.
itg2	...porque considero el hacer ejercicio como parte de mi identidad.
ext2	...porque mis amistades/familia/pareja me dicen que he de hacer ejercicio.
ide3	...porque creo que es importante hacer el esfuerzo de hacer ejercicio regularmente.
amo3	...pero no veo el sentido de hacer ejercicio.
int3	...porque hacer ejercicio me parece una actividad agradable.
ity3	...porque tengo sensación de fracaso cuando no he hecho ejercicio durante un tiempo.
itg3	...porque considero el ejercicio una parte fundamental de quién soy.
ext3	...porque hay personas que no estarían satisfechas conmigo si no hiciera ejercicio.
ide4	...porque me siento inquieto/a si no hago ejercicio regularmente.
amo4	...pero pienso que hacer ejercicio es una pérdida de tiempo.
int4	...porque obtengo placer y satisfacción al hacer ejercicio.
ity4	...porque me sentiría mal conmigo mismo/a si no reservo tiempo para el ejercicio.
itg4	...porque considero que el ejercicio encaja con mis valores.
ext4	...porque me siento presionado/a por mis amigos/familia para hacer ejercicio.

Vitalidad subjetiva

Generalmente...

NOMBRE PSICOMÉTRICO	CONTENIDO DEL ÍTEM
vi1	... me siento vivo/a y vital.
vi2	... tengo energía y ánimo.
vi3	... me ilusiono con cada nuevo día.
vi4	...casi siempre me siento alerta y despierto/a.
vi5	...me siento activo/a o siento que tengo mucha energía.

Apéndice B. *SubscaleExplorer*: un paquete para la gestión y descripción de datos psicométricos

El paquete *SubscaleExplorer* (Angulo-Brunet y Viladrich, 2017a) se desarrolló con la idea de facilitar la gestión de datos psicométricos, así como de obtener con pocas líneas de código descripciones (i.e., recuento de datos perdidos, frecuencias, valores únicos, tamaño del efecto) y gráficos (i.e., bubbleplot, boxplot y violinplot) con R. Las funciones iniciales que se habían desarrollado juntamente con Carme Viladrich se unificaron en el paquete *SubscaleExplorer* que actualmente se encuentra alojado en GitHub (www.github.com/AnguloB/SubscaleExplorer). GitHub es un repositorio que permite compartir código, y obtener *feedback* de otros usuarios sin perder la trazabilidad de los cambios que se realizan. Todas las funciones de *SubscaleExplorer* están preparadas para realizar los análisis y los gráficos por grupo. La excepción es *getscale()* que sirve para la gestión inicial de datos.

Las siguientes páginas de este apéndice serán dedicadas a explicar las posibilidades y el funcionamiento de *SubscaleExplorer*. Para ello se ha elegido un formato de tutorial con el fin de facilitar su uso por parte de investigadores no necesariamente familiarizados con R.

Instalación y activación de *SubscaleExplorer*

En primer lugar, para poder instalar el paquete *SubscaleExplorer* será necesario tener instalado R que puede

descargarse en *The Comprehensive R Archive Network* (<https://cran.r-project.org>).

En segundo lugar, dependiendo del sistema operativo que utilices deberás instalar un programa u otro. Si eres usuario de Windows es necesario instalar Rtools (<https://cran.r-project.org/bin/windows/Rtools/>), si eres usuario de Mac es necesario que instales Xcode desde la *App Store* y si eres usuario de Linux es necesario escribir “r-base-dev” o “r-devel” en la consola de R.

En tercer lugar, ya en la consola de R, será necesario instalar el paquete *devtools*, que entre otras funcionalidades, permite interactuar con la plataforma GitHub e instalar el paquete *SubscaleExplorer*,

```
install.packages("devtools")
```

Una vez se han realizado estos pasos previos, que solo deben realizarse una sola vez, es posible instalar el paquete *SubscaleExplorer* con las siguientes líneas,

```
library(devtools)  
install_github("AnguloB/SubscaleExplorer")
```

Para activar el paquete *SubscaleExplorer*, porque se empieza una sesión de trabajo nueva activaremos la librería como con cualquier otro paquete de R,

```
library(SubscaleExplorer)
```

El caso práctico

Para ilustrar el funcionamiento del paquete `SubscaleExplorer` vamos a utilizar un conjunto de datos simulados que pueden descargarse manualmente en `SubscaleExplorerExamples` (<https://github.com/AnguloB/SubscaleExplorerExamples/>) junto con otros archivos que contienen la definición de las variables, la sintaxis para resolver el caso y el output que se obtiene. Concretamente se trata de cuatro materiales:

- **Data.txt** Archivo con los datos simulados
- **reference.xlsx** Libro de referencias utilizado para la resolución del caso práctico
- **SubscaleExplorer example.Rmd** Archivo Rmd que incluye la sintaxis utilizada en la resolución del caso práctico
- **SubscaleExplorer__example.html** Output del archivo Rmd con el mismo nombre

Se trata de un ejemplo simple que ilustra un caso muy frecuente en psicometría. El conjunto de datos contiene para cada persona información sociodemográfica y las respuestas a uno o más cuestionarios en los que los ítems de las distintas subescalas suelen presentarse mezclados unos con los otros bien sean de forma aleatoria o no. El conjunto de datos llamado *Data*, ilustrado en la Figura AC1, contiene los datos simulados, de 800 personas entre 60 y 80 años que participan en clases de actividad física con un monitor o una monitora. De izquierda a derecha, se encuentra una variable identificativa (*id*), la variable sexo (*sexo*), la variable cuantitativa edad (*edad*), la variable que asigna un número al monitor/a que realiza la sesión

(*monitor*) y diez variables (*v01* a *v10*) que contienen respuestas a ítems de vitalidad subjetiva y afecto positivo. Concretamente, *v01*, *v02*, *v06*, *v08* y *v10* pertenecen a la escala de vitalidad subjetiva mientras que *v03*, *v04*, *v05*, *v07* y *v09* son ítems de una escala de afecto positivo. Las respuestas son en una escala tipo Likert de 1 a 5 (1 = *Totalmente en desacuerdo*; 5 = *Totalmente de acuerdo*).

	id	sexo	edad	monitor	v01	v02	v03	v04	v05	v06
1	1	Hombre	81	14	NA	1	2	3	3	3
2	2	Hombre	80	1	2	2	1	4	4	4
3	3	Mujer	75	18	4	2	5	4	2	2
4	4	Hombre	77	5	5	5	1	1	4	1
5	5	Mujer	74	3	NA	5	5	1	4	1
6	6	Mujer	61	12	3	4	2	4	3	3
7	7	Hombre	75	7	NA	5	1	3	5	1
8	8	Mujer	78	18	4	1	3	4	3	1
9	9	Mujer	68	3	5	4	1	3	2	5
10	10	Hombre	85	5	2	4	4	2	5	5

Figura AC1 Representación del Conjunto de Datos de Ejemplo para *SubscaleExplorer*

En primer lugar, vamos a solucionar la gestión inicial de datos: hacer agrupaciones sin tener que definir manualmente el nombre de los ítems o su posición. En segundo lugar, tomaremos una decisión sobre la naturaleza de los datos: teniendo en cuenta que tenemos una escala tipo Likert de cinco puntos, ¿podemos tratarlos como cuantitativos o bien es más adecuado considerarlos como ordinales o categóricos?

El primer paso será abrir el conjunto de datos en R. Para ello existen múltiples funciones pero en este caso vamos a utilizar la

instrucción *read.csv()* que viene incorporada en R sin necesidad de instalar nada más.

```
Data <- read.csv("Data.txt", sep="")
```

Con “<-” asignamos al conjunto de datos externo “Data.txt” a un conjunto de datos con el mismo nombre “Data” en R.

Gestión de datos: *getscale()*

Una de las necesidades encontradas cuando se analizan datos psicométricos con R es trabajar con diversos subconjuntos de variables que forman las diferentes dimensiones. Para afrontarlo, se incorporan estas variables a una función llamándolas por su nombre o por su posición en el conjunto de datos. Llamar a las variables por su nombre o posición puede ser fácil cuando las subescalas tienen nombres intuitivos y una numeración correlativa. En la práctica no suele ser así, en muchas ocasiones el protocolo de creación del conjunto de datos requiere que los ítems de las diferentes subescalas tengan nombres poco intuitivos y con números no correlativos. En este caso la instrucción para utilizar solo las variables de la escala de vitalidad subjetiva es compleja ya que requiere que escribamos uno por uno el nombre de los ítems que conforman esta dimensión,

```
Data[c("v01", "v02", "v06", "v08", "v10")]
```

o el lugar que ocupan en el conjunto de datos (el número de columna),

```
Data[c(5, 6, 10, 12, 14)]
```

Ambas formas son complicadas y están sujetas a múltiples errores. Aunque esto es un paso que hay que realizar en todos los *softwares*, ya sea con la selección de las variables con menús o con sintaxis parecidas a las de R, en el caso del tratamiento de datos de cuestionarios es interesante poderlo simplificar para reducir posibles errores en la definición de las subescalas. Para hacer este proceso más ágil se desarrolló la función *getscale()*.

Esta función tiene como objetivo hacer estas agrupaciones de forma automática a partir de un libro de referencias construido en una hoja de Microsoft Excel™. Para que este libro de referencias sea compatible con la función *getscale()* debe tener unas características concretas (Figura AC2): tiene que tener una variable denominada `VariablesInput` que contenga los mismos nombres y en el mismo orden que las columnas en el conjunto de datos, y una variable de agrupación en la que se indicará el subconjunto de datos al que pertenece cada columna. La variable de agrupación puede tener un nombre cualquiera. En este caso se llama “Subscales”.

	VariablesInput	Subscales
1	id	NA
2	sexo	NA
3	edad	NA
4	monitor	NA
5	v01	vitalidad
6	v02	vitalidad
7	v03	afecto
8	v04	afecto
9	v05	afecto
10	v06	vitalidad
11	v07	afecto
12	v08	vitalidad
13	v09	afecto
14	v10	vitalidad

Figura AC2 Representación de la Hoja de Referencias para *getscale()* de *SubscaleExplorer*

Para que la gestión de datos funcione con *getscale()* una vez el conjunto de datos (*Data*) está abierto es necesario ejecutar *getscale()*,

```
getscale(Data, "reference.xlsx", column="Subscales")
```

La función *getscale()* devuelve un objeto llamado “output” que contiene una lista con vectores con las subescalas definidas. Esta instrucción genera tantos vectores de valores como agrupaciones se han definido en el libro de referencias. La lista tiene el mismo número de valores que el número de variables del conjunto de datos. Cada

vector, especifica si cada variable forma parte o no de esa agrupación. De esta forma tal y como se puede ver en la Figura AC3, los valores que constan como TRUE son los que coinciden con la posición de las variables del conjunto de datos.

```
Subscales: afecto
[1] FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE FALSE
-----
Subscales: vitalidad
[1] FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE TRUE
```

Figura AC 3 Objeto “output” de la Función getscale()

Una vez definidas las listas, si se quiere utilizar una de las agrupaciones, por ejemplo la denominada vitalidad subjetiva ejecutaríamos,

```
Data[output$vitalidad]
```

o en el caso de afecto positivo,

```
Data[output$afecto]
```

Para simplificar todavía más la llamada de las subescalas, se puede ejecutar la siguiente instrucción,

```
a<-unlist(output)
a<-data.frame((vapply(output,unlist,unlist(output[[1]]))))
for(i in 1:ncol(a)){
  assign(names(output)[i], a[[i]])
}
```

De forma que, por ejemplo, para hacer la llamada a los ítems de la subescala de vitalidad el código se reduce a,

```
Data[vitalidad]
```

El objeto “output” generado por la función *getscale()* es de especial utilidad cuando se tiene que trabajar con muchos conjuntos de datos. A partir de la lista de vectores podemos crear listas con conjuntos de datos tal y como podemos ver en la Figura AC4 mediante la siguiente instrucción,

```
datos<-lapply(output, function(x) Data[x])
```

```
datos      List of 2
afecto : 'data.frame': 800 obs. of 5 variables:
 ..$ v03: int [1:800] 2 1 5 1 5 2 1 3 1 4 ...
 ..$ v04: int [1:800] 3 4 4 1 1 4 3 4 3 2 ...
 ..$ v05: int [1:800] 3 4 2 4 4 3 5 3 2 5 ...
 ..$ v07: int [1:800] 5 5 3 3 5 3 2 4 5 2 ...
 ..$ v09: int [1:800] 4 4 2 2 1 2 2 5 2 4 ...
vitalidad: 'data.frame': 800 obs. of 5 variables:
 ..$ v01: int [1:800] NA 2 4 5 NA 3 NA 4 5 2 ...
 ..$ v02: int [1:800] 1 2 2 5 5 4 5 1 4 4 ...
 ..$ v06: int [1:800] 3 4 2 1 1 3 1 1 5 5 ...
 ..$ v08: int [1:800] 5 2 4 5 4 3 1 2 1 5 ...
 ..$ v10: int [1:800] 4 1 1 4 3 2 5 1 4 4 ...
```

Figura AC4 Listado "datos" con los Conjuntos de Datos Generados a Partir de "output"

donde “datos” es el nombre que se ha asignado manualmente con los caracteres “<-” el resultado de aplicar la función “lapply”. La función *lapply* es una función que aplica a cada elemento de una lista (en este caso output) una función. En el ejemplo donde “x” representa cada elemento de la lista “output” se aplica una función que permite

que esta lista de vectores sea convertida en una lista de conjuntos de datos.

Descripción de escalas Likert

Siguiendo la misma lógica podemos realizar cualquier tipo de análisis para todos los subconjuntos de datos a la vez. De esta forma, por ejemplo, si queremos calcular unos estadísticos descriptivos utilizando la función *describeBy()* del paquete *psych* (Revelle, 2018) escribiríamos,

```
lapply(lapply(output, function(x) Data[x]), describeBy)
```

obtendríamos los resultados que se observan en la Figura 6.

\$afecto													
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
v03	1	800	3.00	1.43	3	3.00	1.48	1	5	4	0.00	-1.34	0.05
v04	2	800	2.98	1.42	3	2.98	1.48	1	5	4	0.03	-1.33	0.05
v05	3	800	2.97	1.44	3	2.96	1.48	1	5	4	0.02	-1.32	0.05
v07	4	800	2.97	1.41	3	2.97	1.48	1	5	4	0.00	-1.29	0.05
v09	5	800	2.93	1.42	3	2.91	1.48	1	5	4	0.04	-1.35	0.05
\$vitalidad													
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
v01	1	645	3.02	1.40	3	3.03	1.48	1	5	4	-0.06	-1.29	0.06
v02	2	800	2.95	1.40	3	2.93	1.48	1	5	4	0.01	-1.27	0.05
v06	3	800	2.94	1.43	3	2.92	1.48	1	5	4	0.01	-1.34	0.05
v08	4	800	2.94	1.44	3	2.92	1.48	1	5	4	0.10	-1.33	0.05
v10	5	800	2.91	1.41	3	2.88	1.48	1	5	4	0.08	-1.31	0.05

Figura AC5 Estadísticos Descriptivos Univariabes Obtenidos para Múltiples Subescalas

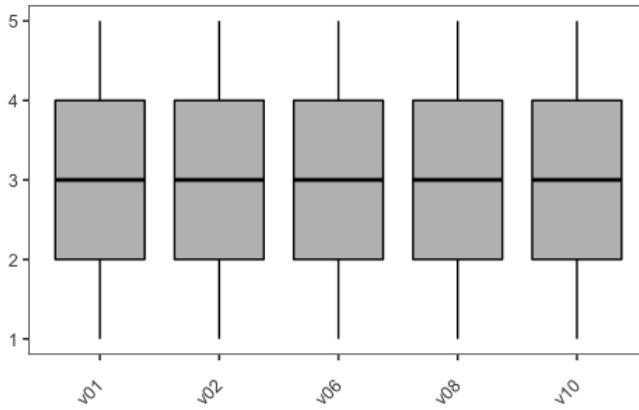
Una vez obtenidos los estadísticos descriptivos vamos a interpretar los resultados. Cada escala está formada por cinco ítems y la totalidad de las 800 personas participantes ha contestado a todos ellos con la excepción del ítem v01 de la subescala de vitalidad subjetiva

en el que faltan muchas respuestas. Los valores medios son alrededor de 3 puntos ligeramente elevados para una escala de 5 puntos y las desviaciones estándar están alrededor de 1.4, también elevadas para el rango de la escala. Los valores de asimetría y curtosis nos ayudarán a imaginar mejor la distribución de respuestas. Podemos ver que, aunque los valores de asimetría son cercanos a 0, los valores de curtosis para todas las variables son todos mayores que 1 en valor absoluto sugiriendo que la distribución de los datos no es Normal (e.g., Lloret-Segura et al., 2014). En este punto, para seguir examinando la distribución de respuestas teniendo en cuenta que estamos tratando con una escala tipo Likert de cinco puntos puede ser interesante hacer una descripción de tipo ordinal.

Funciones `boxPlot()` y `violinPlot()`

Para facilitar la descripción gráfica de los ítems cuando son o pueden ser considerados con una escala ordinal disponemos de la función `boxPlot()` y `violinPlot()` que proporcionan información ordinal sobre la tendencia central y la dispersión de estas variables. Tal y como se observa en la Figura AC6, en el caso de la gráfica de violines (`violinplot`) también se incluye el `boxplot` en su interior.

A



B

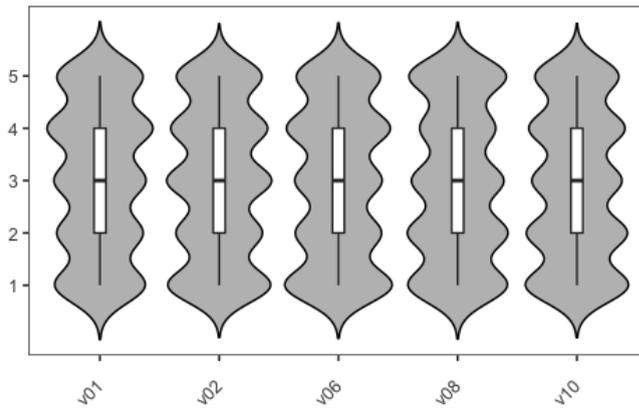


Figura AC6 Boxplot y violinplot Obtenidos con las Funciones *boxPlot()* y *violinPlot()* de *SubscaleExplorer*

Función frecuencias()

Para la descripción de escalas tipo Likert podemos hacer uso de paquetes de R como *psych*, que acabamos de utilizar, para obtener

estadísticos descriptivos. En muchas ocasiones será necesario también el uso de tablas de frecuencias o de proporciones. Hacer una tabla de frecuencias para una sola variable no es difícil en ningún *software* y tampoco en R pero cuando se requiere de una tabla de frecuencias para múltiples variables no es tan sencillo ya que requiere manipular el formato de los datos. Es por ello por lo que en *SubscaleExplorer* se propone la función *frecuencias()* para obtener una tabla de frecuencias.

```
frecuencias(Data[vitalidad])
```

	value					
variable	1	2	3	4	5	<NA>
v01	128	120	128	148	121	155
v02	174	144	176	162	144	0
v06	185	146	148	177	144	0
v08	172	172	158	129	169	0
v10	175	170	151	162	142	0

Figura AC7 *Tabla de Frecuencias Obtenida con la Función *frecuencias()* de *SubscaleExplorer**

En la Figura AC7 podemos ver las frecuencias para la escala de vitalidad subjetiva. Si se quieren ocultar los datos perdidos se puede añadir `missing = FALSE` o si se quieren obtener proporciones se debe ejecutar `prop = TRUE`. Aunque por defecto no podemos obtener el porcentaje de respuestas, para hacerlo podríamos manipular la tabla multiplicando por 100 tal y como se muestra a continuación,

```
frecuencias(Data[vitalidad], prop = TRUE)*100
```

Función freqbubble()

Una representación todavía más general de las escalas Likert es la dada por los gráficos de “burbuja” (bubbleplots). Estas gráficas se generan a partir de una tabla de frecuencias o proporciones, de esta si se integra la función *frecuencias()* con la función *freqbubble()* se obtiene un gráfico que permite visualizar la distribución de respuestas (Figura AC8).

```
freqbubble(frecuencias(Data[vitalidad], prop = TRUE) , legend = FALSE  
)
```

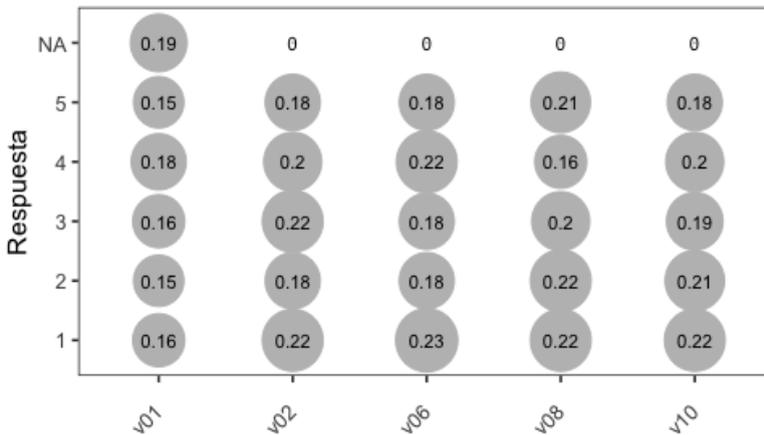


Figura AC8 Gráfica de Burbujas (bubbleplot) Obtenida con la Función *freqbubble()* de *SubscaleExplorer*

Estos gráficos son de especial utilidad para detectar presencia de efectos suelo o techo u otras desviaciones de la distribución Normal. Tal y como se puede observar, en los datos presentados en este ejemplo la falta de normalidad no es debida a la presencia de efecto techo y/o suelo, sino a la equiprobabilidad de las respuestas a lo largo de todos los valores posibles de la escala de respuestas. Además, tal y como se ha explicado en el apartado de la función *freqbubble()* se puede controlar la presentación de los valores faltantes añadiendo la opción `missing = FALSE`.

Función missingness()

La función *missingness()* fue desarrollada con la finalidad de estudiar los datos perdidos. Con el paquete `SubscaleExplorer`, las ideas principales son que el número y la distribución de datos perdidos condicionan los análisis de datos. Mediante una simple línea podemos obtener un resumen exhaustivo de los casos perdidos calculados por casos, por variables y por celdas,

```
missingness(Data[vitalidad])
```

Como podemos ver en la Figura AC9 la función *missingness()* proporciona las frecuencias absolutas de los datos perdidos así como sus proporciones. En primer lugar, en el resumen general (*General summary*) se encuentra el número de personas que tienen todos los ítems de la escala de vitalidad contestados (*Cases with all items answered*) y complementariamente el número de personas que tienen

valores perdidos en alguna de las variables de vitalidad (*Cases with incomplete data*). Son los casos con algún valor perdido, los que se eliminarían si se usara la estrategia de eliminación *listwise*. Llegados a este punto se puede considerar que hay 155 (19%) personas con valores perdidos y siguiendo la recomendación de Graham (2009), que considera que si existen más de un 5% de valores perdidos hay consecuencias importantes de sesgo o pérdida de potencia, se deben tomar precauciones especiales para trabajar con nuestros datos. Sin embargo, en las siguientes líneas de la tabla se da otra información relevante que puede afectar a la conclusión. La totalidad de las 800 personas ha contestado algún ítem (*Cases with some or all items answered*) de manera que los 155 casos con valores perdidos no corresponden a registros completamente en blanco. De hecho, una forma más precisa de calcular la proporción de datos faltantes es obtener el número de celdas vacías. En este caso el número de celdas vacías es precisamente 155. Por lo tanto, en realidad, la proporción de celdas perdidas es únicamente de 3.8% (*Proportion of cell missingness*) siendo esta una visión mucho más optimista que la anterior. Por otra parte, el resumen por variables (*Variable summary*) informa de que solo tenemos datos perdidos en la variable *v01*.

Como conclusión, en este ejemplo se encuentran 155 personas que no han contestado a la variable *v01*, una información que en este ejemplo se deducía también de la tabla de estadísticos descriptivos y del análisis de frecuencias. Además de los datos faltantes por variable,

con el resultado de la función *missingness()* proporciona el cálculo de valores faltantes por casos y por celdas.

Aunque en este ejemplo no se muestra, si se hubiera encontrado un caso con datos perdidos en todas las variables, es decir una persona que no contestó a ninguno de los ítems de la subescala que se analiza, saldría reflejado en “*Cases with completely missing data*” y justo debajo del resumen por variable saldría un vector indicando que caso o casos son los problemáticos.

```

$`General Summary`
      Description      Freq   Prop
1      CASE MISSINGNESS
2
3      Cases with all items answered      645   0.81
4      Cases with incomplete data       155   0.19
5
6      Cases with some or all items answered      800   1
7      Cases with completely missing data        0   0
8      -----
9      CELL MISSINGNESS
10     n variables          5
11     n cases             800
12     n missing cells     155
13     Total cases x Total variables      4000
14     Proportion of cell missingness      0.03875

$`Variable summary`
variable  na prop.na
1      v01 155  0.194
2      v02  0  0.000
3      v06  0  0.000
4      v08  0  0.000
5      v10  0  0.000

```

Figura AC9 Resumen de Valores Perdidos Obtenidos con la Función *missingness()* de *SubscaleExplorer*

Si modificamos los argumentos por defecto que tiene la función *missingness()* también conseguiremos tener un análisis del patrón de valores perdidos, extraído del paquete *mice* (Buuren y Groothuis-Oudshoorn, 2010),

```
missingness(Data[vitalidad], pattern = TRUE)
```

El patrón de datos perdidos es un resumen multivariado del número de datos perdidos que contiene el conjunto de datos.

Función uniqueValues()

Una función que tiene especial interés si se quiere realizar un análisis de la invariancia y se dispone de datos ordinales, que no pueden ser tratados como cuantitativos, es la función *uniqueValues*. En este tipo de análisis se facilita mucho si los grupos tienen el mismo número de categorías de respuesta con algún valor (e.g., Liu, Millsap, West, Tein, Tanaka y Grimm, 2017). Siguiendo el mismo ejemplo de la escala de vitalidad, pero ahora haciendo el análisis por la variable sexo,

```
uniqueValues(Data[vitalidad], group=Data$sexo)
```

tal y como podemos ver en la Figura AC10, distinguimos dos tipos de valores únicos: *uniqueValid* que representa los valores únicos sin tener en cuenta los valores perdidos y *unique* que incluye también los valores perdidos.

```

group: Hombre
  variable uniqueValid unique
1      v01           5      6
2      v02           5      5
3      v06           5      5
4      v08           5      5
5      v10           5      5
-----
group: Mujer
  variable uniqueValid unique
1      v01           5      6
2      v02           5      5
3      v06           5      5
4      v08           5      5
5      v10           5      5

```

Figura AC10 Valores Únicos Obtenidos con la Función *uniqueValues()* de *SubscaleExplorer*

En este ejemplo podemos ver como tanto hombres como mujeres han utilizado los cinco puntos de la escala de respuesta. Por otra parte, tal y como ya sabíamos por los análisis previos existen valores perdidos en v01.

Función `designEffect()`

Finalmente la función *designEffect()* permite evaluar el coeficiente de correlación intraclase, que representa la variancia a nivel individual que puede ser explicada por la pertenencia a un grupo. A partir de este coeficiente se obtiene el efecto de diseño multinivel siguiendo la formulación de (Hox, Moerbeek, y Schoot, 2018).

```
designEffect(Data[vitalidad], cluster=Data$monitor)
```

```
[1] "n (mean cluster size when groups are different size)"  
[1] 40  
      ICC1 designEffect  
v01  0.011          1.428  
v02 -0.005          0.820  
v06  0.007          1.255  
v08  0.010          1.398  
v10  0.013          1.522
```

Figura AC11 Coeficiente de Correlación Intraclase y Efecto de Diseño Multinivel Obtenido con `designEffect ()` de `SubscaleExplorer`

Teniendo en cuenta que el efecto de diseño multinivel es inferior a dos puntos, podemos considerar que no hay efecto multinivel.

Usabilidad del paquete

Hasta el momento el paquete *SubscaleExplorer* ha mostrado una buena aceptación entre el público que lo ha probado. Desde su primera presentación en el XV congreso de la Asociación Española de Metodología de las Ciencias del Comportamiento (Apéndice C6; Angulo-Brunet y Viladrich, 2017b) el paquete ha sido utilizado para la descripción y gestión de datos en diversos cursos y sesiones prácticas. Este es el caso del curso impartido en el *Servei d'Estadística Aplicada* de la UAB entre el 30 de enero y el 2 de febrero de 2017 (*Disseny i adaptació de qüestionaris: estudi de l'estructura interna i de la fiabilitat amb R*). O los cursos de veinte horas cada uno impartidos en el centro CuSur de la Universidad de Guadalajara en marzo de 2019 (Descripción

y gestión de datos psicométricos con R y Generalización de la estimación de fiabilidad de consistencia interna basada en el coeficiente omega). En los cursos, los usuarios de R mayormente noveles conseguían hacer un análisis descriptivo de sus propios datos mediante el conjunto de funciones proporcionadas por el paquete *SubscaleExplorer* con la finalidad de poder analizar el modelo de medida y obtener el coeficiente de fiabilidad adecuado.

Apéndice C Trabajos presentados en congresos nacionales e internacionales

Apéndice C1. SubscaleExplorer: Un paquete de R para la gestión y descripción de subescalas psicométricas

Ariadna Angulo-Brunet y Carme Viladrich

La correcta gestión de datos previa a cualquier análisis psicométrico suele ser costosa con cualquier programa estadístico. Ello se debe a que se manejan múltiples variables que suelen ser ordinales, estar agrupadas por motivos teóricos en diversas subescalas, con datos que pueden derivarse de diseños complejos con varios grupos, longitudinales, con datos faltantes o multinivel. El paquete SubscaleExplorer ofrece al usuario de R un conjunto de funciones dirigidas a facilitar la gestión de datos así como a realizar un análisis descriptivo a fin de tomar decisiones sobre como abordar análisis psicométricos más complejos. El objetivo de este trabajo es hacer una demostración de las funciones del paquete a partir de datos reales proporcionados por 593 estudiantes de ESO de 32 grupos clase que contestaron el cuestionario (63.24% contestó en catalán y el resto en castellano) de estilo interpersonal del profesorado de educación física (PETIS-Q). Este cuestionario tiene un total de 24 ítems y seis subescalas. En primer lugar, con la función `getscale`, se definen agrupaciones de variables con significado psicométrico, bien sea por subescalas, por cuestionarios, por ocasiones, o cualquier otra agrupación de interés. En unas sencillas líneas y para cada agrupación de variables, las diversas funciones descriptivas permiten: (a) una evaluación del proceso de respuesta mediante el estudio de la proporción de valores faltantes (`missingness`) y del uso de los valores disponibles en la escala de respuesta (`uniqueValues`), (b) la valoración de la forma de la distribución de las respuestas mediante el cálculo de las frecuencias de respuesta y su representación gráfica (`freqbubble` y `boxPlot/violinPlot`), y (c) el estudio del efecto del diseño basado en el coeficiente de correlación intraclase para examinar una posible estructura multinivel de los datos (`designEffect`). Todas estas funciones pueden aplicarse al conjunto de la muestra o a los diferentes grupos de casos. Compartiendo la filosofía de R, recomendamos utilizar las funciones de otros paquetes cuando éstas son simples, como por ejemplo, los estadísticos descriptivos del paquete `psych`. La aportación de SubscaleExplorer es la programación de funciones inexistentes en otros paquetes, como `getscale` o `designEffect` y la simplificación de funciones complejas, como las gráficas.

keywords: *software*, cuestionarios, escalas, gestión de datos, gestión de escalas, descripción de datos, gráficos

Apéndice C2. ¿Limitaciones o cambio de objetivos? Lecciones aprendidas comparando dos opciones para reportar medidas con baja consistencia interna

Ariadna Angulo-Brunet y Carme Viladrich

Tal como se recomienda en los textos normativos, actualmente está consolidada la práctica de estudiar la consistencia interna de las medidas basadas en cuestionarios antes de usarlas para comprobar la hipótesis principal de un estudio. En este trabajo nos planteamos la disyuntiva de qué hacer cuando, una vez obtenidos todos los datos del estudio principal, se observa que la consistencia interna de algunas medidas no es aceptable. En algunas publicaciones, la baja consistencia interna se trata como una limitación del trabajo y se invita a tener precaución al interpretar los resultados principales; mientras que otra opción consiste en modificar los objetivos del trabajo para aportar datos a la clarificación del problema psicométrico detectado. El dilema se ilustra con un caso práctico cuyo objetivo inicial fue comparar, utilizando path analysis, la red nomológica entre los constructos registrados mediante dos cuestionarios, uno formado por medidas de un solo ítem y otro formado por medidas de múltiples ítems. Para ello se dispuso de respuestas dicotómicas obtenidas en una muestra de 128 mujeres y hombres mayores de 60 años que contestaron ambos cuestionarios. Aunque el ajuste del path analysis con las medidas de múltiples ítems resultó adecuado (χ^2 [df] = 4.15[3], $p = .25$, CFI = 1.00, TLI = .97, RMSEA[90%IC] = .06[0-.17]) y con las medidas de un solo ítem fue parecido aunque más discutible (χ^2 [df] = 6.44[3], $p = .09$, CFI = .99, TLI = .89, RMSEA [90%IC] = .08[.00-.17]), la consistencia interna de tres de las siete medidas de múltiples ítems no fue aceptable (omega categórica; rango .57 a .88). A la luz de estos resultados se discuten las implicaciones de utilizar variables calculadas a partir de la suma o promedio de otras variables en un path analysis cuando no se ha alcanzado un umbral mínimo de fiabilidad.

keywords: fiabilidad, red nomológica, psicometría, path analysis

Apéndice C3. How to handle ordered responses with floor and ceiling effects in SEM using small samples?

How to handle ordered responses with floor and ceiling effects in SEM using small samples?

Angulo-Brunet, A.^{1*}, Viladrich, C.²

¹ Dept. Psychobiology and Methodology of Health Sciences, Universitat Autònoma de Barcelona, Spain

* Presenting author, PhD-student, Carme Viladrich supervisor

Briefly Explain Your Question (max. 100 words)

In order to decrease the high response costs related to the assessment of multiple constructs in the elderly population, we developed single-item measures for 17 constructs related to motivation for physical activity. Our aim is to provide validity evidence for the new measures comparing two nomological networks via SEM: one based on the single-item measures and the other based on the full questionnaire (69 items). Seventy elderly people responded using a 5-point Likert-Scale (1 = Totally disagree, 5 = Totally agree). Response distributions showed ceiling and floor effects in the short and the long form questionnaires.

Scientific field(s) of the author(s)

Psychometrics, motivation

Relevance to conference theme (max. 50 words)

This study is motivated by the difficulties in assessment with questionnaires of our target group: the elderly who practice physical activity with a trainer. Data collection has high response and administrative costs and sample sizes are typically small. Testing relations between multiple constructs in SEM is pestered with lack of statistical power, non-convergent solutions and parameter instability.

Keywords (max. 3)

Structural equation modeling, nomological network, the elderly

Acknowledgments: Funding was provided by grant MINECO/DEP2014-52481-C3-1R from Spanish Government

Apéndice C4. Desarrollo de medidas de un ítem para evaluar los resultados de una intervención percibidos por personas mayores

Ariadna Angulo-Brunet, Carme Viladrich y Susana Pallarés

Las dificultades para contestar cuestionarios largos en los estudios con grupos y longitudinales son bien conocidas. A la posible falta de atención y de interés, y la fatiga y el aburrimiento se añaden problemas de visión y de lectoescritura cuando el cuestionario se dirige a personas mayores. El objetivo de este trabajo es desarrollar medidas de un solo ítem para cuantificar los constructos utilizados en la teoría motivacional de la autodeterminación de forma que puedan ser utilizados como medidas de resultados de una intervención en monitores de actividad física para personas mayores. Estos son el estilo interpersonal del monitor (CIS-Q), la satisfacción de las necesidades psicológicas básicas (BPNES), la regulación conductual (BREQ-3), la vitalidad (SVS) y la intención de continuar. Para ello, la primera autora desarrolló las medidas de un solo ítem a partir del contenido de las escalas originales y de las recomendaciones de la literatura científica (versión 1). Posteriormente, se creó una nueva versión del cuestionario a partir de tres fuentes de información. Primero, cinco especialistas evaluaron los ítems creados en términos de relevancia, suficiencia y claridad de forma independiente. También se les pidió que seleccionaran un ítem del cuestionario o propusieran uno nuevo que pudiera representar mejor el constructo que se quería medir. Como tercera fuente se utilizaron los índices de discriminación de los ítems de los cuestionarios originales contestados por 160 personas de la población diana con edades entre 65 y 92 años (versión 2). Las aportaciones de cuatro nuevas personas de la población diana en términos de comprensión dieron pie a nuevas modificaciones (versión 3). Finalmente, otras nueve personas contestaron la versión larga del cuestionario en formato heteroadministrado y la versión corta del cuestionario en formato autoadministrado. La distribución de respuestas a las medidas de un ítem fue similar a la distribución de las medias de las subescalas de la versión larga. La tercera versión del cuestionario reducido mostró ser factible con personas de la población diana y los resultados preliminares fueron prometedores por lo que nos planteamos en futuras fases de esta investigación la obtención evidencias cuantitativas de validez y fiabilidad para estas medidas.

keywords: medidas de un ítem, desarrollo de cuestionarios, cuestionario reducido, teoría de la autodeterminación.

Apéndice C5. Single-item measures validity evidence based on correlation network analysis: motives for physical activity in the elderly

Ariadna Angulo-Brunet, Carme Viladrich, Yago Ramis and Susana Pallarès

Introduction: Due to the importance of physical activity in frailty prevention among the elderly, it is important to assess their motives for physical activity. In the elderly population, reading difficulties must be added to those negative feelings also experimented in general population, as fatigue, boredom or frustration in data collection. Using single-item measures could overcome those difficulties and increase feasibility when multidimensional constructs are assessed, which is the case of Self-determination theory of motivation. This theory establishes that coaches' interpersonal style can impact on basic psychological needs satisfaction, which can predict motives for physical activity which in turn can anticipate feelings of vitality and engagement in an activity. **Objectives:** The aim of this study is to gather validity evidence for the single-item measures, developed in early stages of this study, based on an empirical test of the expected nomological network. **Design/Methodology:** Seventy elderly adults who attended to physical activity programs participated in this study. It was heteroadministered the full-form of Coaches' interpersonal style questionnaire, the Basic psychological needs satisfaction scale, the Behavioral regulation exercise questionnaire, the Subjective vitality scale and three engagement items. Short-form was self-reported. We performed a correlation network analysis for each questionnaire form with qgraph package of R *software* using polychoric correlations due to ceiling and floor effects. **Results:** Both network analyses reproduce a similar correlation pattern showing differences in the strength of the relationship between variables. Due to the small sample size, some items of the long form were deleted due to non-variance, indicating low discrimination of those items. **Conclusions:** More validity evidences should be provided for the assessment of motives for physical activity in the elderly with short and long measures. Decision-making consequences related to feasibility and psychometric properties are discussed.

Keywords: Correlation network, short-measures, elderly, motivation, self-determination theory

Apéndice C6. Validity of single-item measures to assess self-determination theory constructs in the elderly

Ariadna Angulo-Brunet, Carme Viladrich, Susana Pallarès,
Yago Ramis and Marta Borrueco

According to Self-determination theory, supportive interpersonal styles are determinant on the satisfaction of basic psychological needs for autonomy, competence and relatedness, which in turn can impact on subjective vitality, forming a three-concept chain. In order to support the use of single-item measures for these concepts, the aim of this study is to provide validity evidence through testing the replicability of the expected chain and also to provide test-retest reliability for each measure. Embedded in a larger study, seventy-one elderly adults who were attending to group physical activities answered the relevant seven single-item questions in a five-points Likert scale for two times in a lapse of five weeks. Considering small sample size and floor and ceiling effects, descriptive correlation networks based on polychoric correlations were used to test the expected three-concept chain. Data gathered in wave 1 yielded favorable evidence for the chain showing moderate correlations (.37-.68) between adjacent points and weaker correlations (.26-.41) between distant points. In contrast, mixed evidence was obtained in wave 2, mainly due to unexpectedly low correlations (.13-.43) between basic psychological needs and the other variables in the chain. Regarding test-retest reliability, supportive interpersonal style single-items showed higher values (.50 -.58) than basic psychological need satisfaction (.31 - .45) and vitality (.37). On the basis of the evidence currently available, single-item measures of supportive interpersonal styles and vitality are promising measures whereas more validity and reliability evidence should be gathered for basic psychological need satisfaction constructs. This research was supported by the Spanish Ministerio de Economía y Competitividad (DEP2014-52481-C3-1R).

