Running Head: Multitrait-multimethod SIS-C

Are type, frequency, and daily time equally valid estimators of support needs in children with intellectual disability? A multitrait-multimethod analysis of the Supports Intensity Scale for Children (SIS-C)

Miguel A. Verdugo[1] , Verónica Guillén[2] & Victor B. Arias[3]

[1] INICO-University of Salamanca, Salamanca, Spain

[2] INICO-University of Cantabria, Santander, Spain

[3] INICO-University of Talca, Talca, Chile

Abstract

Support needs represent the intensity of support required by a person with a disability in order to take part in the activities related to normative human functioning. The Supports Intensity Scale for Children (SIS-C; Thompson et al., 2016) is possibly the most promising tool for assessing and designing individualized support programs in children with intellectual disability. The SIS-C measures support needs across 61 activities, each one assessed along three methods: type of support, frequency, and daily time during which support is to be given. We investigated the impact of method effects in the SIS-C through a bifactor approach to the analysis of multitrait-multimethod matrices. The results suggest that neither intensity nor frequency scales produced method effects that significantly distorted the measurement of support needs. However, the daily support time method had substantial undesirable effects on five of the seven subscales of support needs. Considerations about support needs assessment and future modifications of the scale are discussed.

Key words: SIS-C, support needs, validity, reliability, intellectual disability.

Introduction

Conceiving intellectual disability (ID) as a state of human functioning rather than as an intrinsic trait has important consequences on our understanding of the role played by the environment in determining the outcomes achieved by persons with ID. Schalock et al. (2010) argued that needs that arise from the gap between the individual and the environment should be approached through individualized support. An interdisciplinary, holistic planning of support and intervention would permit the issues related to ID to be successfully tackled (Schalock, Luckasson, Tassé, & Verdugo, in press). Despite the fact that there is a mutual relationship between deficiencies and support needs, and that, accordingly, a higher degree of personal limitations is associated with a higher intensity in the needed support, an approach centered on reducing the gap between personal competencies and environmental demands, rather than an approach exclusively focusing on the deficiencies, is more likely to bring about a more precise identification of the support needed to improve human functioning and personal outcomes (Thompson et al., 2009; Thompson & Viriyangkura, 2013).

Support needs (SN) have been defined as a psychological construct referring to the intensity of support required by an individual in order to take part in the activities related to normative human functioning (Thompson et al., 2009). Therefore, they should be identified by integrating information from the individual with an ID and from other knowledgeable informants (Schalock et al., 2010; Thompson et al., 2002; Thompson, Wehmeyer, & Hughes, 2010; Thompson & Viriyangkura, 2013). Currently, it is understood that individuals with ID have the same needs as individuals without ID, and that they may reveal other special needs and require specific support (in terms of type, frequency, and daily support time) for the purpose of identifying, expressing, and meeting those needs (Thompson & Viriyangkura, 2013). From this perspective, the possible differences between individuals with and without ID would mainly depend on the intensity of the support to be provided by the society in order to guarantee their complete participation in social activities and interactions.

Thompson et al. (2009) suggested that not all individuals with ID are equally in need of all the available support or the same type of support in different aspects of their daily functioning or at different times of their lives. Similarly, it can be argued that the support given to individuals with ID does not always correspond to their actual needs, because there is a somewhat traditional tendency towards overprotecting these individuals, which has hindered their development and learning, limiting the achievement of the expected results, as well as their independent functioning in their communities (Stancliffe & Keane, 2000). For this reason, in the last few years, professionals in the field of disability have consistently underlined the importance of objectively identifying the types of needed support for the purpose of establishing systematic interventions capable of delivering personalized and contextualized support. Individualized interventions are expected to improve the physical and psychological wellbeing of individuals with ID, enhancing their abilities, participation in the life of the community, and overall quality of life (Schalock et al., 2010; Schalock, Verdugo, Gómez, & Reinders, 2016).

In 2002, Thompson et al. called for the development of reliable and valid measurement tools capable of delivering (through support needs indexes and profiles) useful information to planning teams regarding the support needs of people with ID. To answer this call, Thompson et al. (2004) developed the Supports Intensity Scale (SIS) as a tool capable of accurately measuring the intensity of the SN in adults with ID (in terms of type, frequency, and daily amount of time required). Their scale has quickly become the most employed tool for designing individualized support programs (Buntinx, Van Unen, Speth, & Groot, 2006; Chou, Lee, Chang, & Yu, 2013; Fortune et al., 2008; Giné et al., 2006; Giné et al., 2007; Giné et al., 2014; Schalock, Thompson, & Tassé, 2008; van Loon, 2009; Verdugo, Arias, & Ibáñez, 2007; Verdugo, Arias, Ibáñez, & Schalock, 2010; Wehmeyer et al., 2009).

Nevertheless, in spite of several attempts to rigorously develop the SIS and other scales to assess SN for people with ID (e.g., Gould, 1998; Hennike et al., 2002; Llewellyn et al., 2005), there is still a clear lack of reliable instruments to assess and measure this construct. In more detail, infancy and adolescence are stages for which the concept of SN has not yet been consensually developed, and there is a widespread demand for assessment tools capable of guiding and committing resources to evidence-based practices from early childhood. A scale with these characteristics would permit: (1) the development of individualized support planning specifically for childhood, promoting the participation of children and adolescents with ID in daily activities, encouraging their adequate integration into the school system and facilitating an appropriate transition into adulthood; (2) the development of coherent collaborations with other lines of research directed at assessing this population in our context, and especially related to the assessment of adaptive behavior (Verdugo et al., 2014) and quality of life (Gómez et al., 2014; Schalock et al., 2016); and (3) the promotion of the achievement of the specific objectives of the United Nations (1989, 2006) in relation to the rights of children with ID.

All these things considered, a new Supports Intensity Scale for Children (SIS-C) was recently developed, originally in English (Thompson et al., 2016), and it has subsequently been adapted, with similar characteristics, for other countries, including Spain (Guillén, Verdugo, Adam-Alcocer, & Giné, 2017; Guillén, Verdugo, Arias, &

Vicente, 2015; Verdugo et al., 2016; Verdugo, Arias, Guillén, & Vicente, 2014; Verdugo, Guillén, Arias, Vicente, & Badia, 2016).

The present study

The SIS-C measures support needs in children with ID across 61 activities, divided into seven environmental contexts (Home Life, Community and Neighborhood, School Participation, School Learning, Health and Safety, Social, and Advocacy). Each activity is assessed along three methods based on the intensity of support: type of support, frequency, and daily time during which support is to be given (see table 1). In theory, type, frequency, and time contribute substantive, rather than redundant, information to the SN measurement. The way in which individual SIS-C scores are obtained for each of these three methods (through the unweighted mean of the scores of type, frequency, and time in each SN domain; Thompson et al., 2016) implies that the three methods of assessment are equally valid and accurate estimators of the SN. Nevertheless, considering the simultaneous resort to different forms of assessment of the same construct, it would be expected that in all measurements, there would be specific effects associated with the method (Campbell & Fiske, 1959).

The non-modeled presence of excessive artifactual variance can lead to several undesirable effects on the validity of the measure (Podsakoff, McKenzie, & Podsakoff, 2012) such as, for instance, altering the correlations between substantive factors, biasing the estimate of certain structural parameters in a SEM context or, crucially, leading to a misrepresentation of the scores in the applied assessment, which are generally of the utmost importance to the assessed individual. In summary, the method effects are a potential threat to the validity with which a certain tool is supposed to evaluate the objective construct. It is therefore important for both research purposes and professional practice to further investigate the size and impact of the method effects.

So far, only one study has tested the properties of the SIS-C taking into account the effect of the various methods of evaluation of the SN (Seo et al., 2016). This study considered seven bifactor models, each comprising a substantive factor (the SN domain) and three specific factors representing the three methods of evaluation. Seo et al. (2016) found that, in general, the factorial loadings of the traits were significantly higher compared to those of the method factors ($p<.01$). An exception was found in the case of the items of two of the subscales ("Community" and "Advocacy"), whose average factorial loadings were significantly higher in one of the method factors ("Daily time support") compared to the substantive dimensions. Seo and colleagues concluded that the multitrait-multimethod model of the SIS-C showed enough evidence of convergent validity to support using the unweighted mean of the scores obtained in the three methods to assess the SN.

Nevertheless, two important observations are necessary regarding Seo et al.'s (2016) study. First, the fact that in two subscales, the mean of the loadings in the method factor was significantly higher than that of the SN factor leads to the conclusion that, in those subscales, the method factor could explain the variance in the items better than the substantive factor. This represents a threat to the convergent validity of the items that should hardly be ignored. Secondly, the statistical significance of the differences in factorial loadings in favor of the trait does not provide further information regarding the size of these differences. Therefore, with that information alone, it is difficult to arrive at a decision regarding the practical relevance of the method effects (e.g. to what extent the method effect affects the reliability with which direct scores represent the individual's position in the pertinent latent variable). In order to facilitate the interpretation of the model, it would be convenient to quantify the size of the factors in terms of explained variance and ratio of reliable variance (Arias, Ponce, & Núñez, 2016; Reise, 2012; Revelle & Wilt, 2013; Rodriguez, Reise, & Haviland, 2016).

In the case of bifactor models in which the specific factors show systematic method variance, their interpretation requires the quantification of their impact on the unidimensionality of the model and on the reliability with which the substantive factor reproduces the target construct. In summary, in addition to focusing on whether or not the SIS-C comprises method effects, research should investigate (a) to what extent these effects hinder a precise assessment of SN, and (b) if they are relevant enough to bring about a change in the ways individual scores are produced or even to promote modifications of the scale as a whole.

In the present study, we aimed to investigate the relevance of the method effects in the SIS-C by utilizing the bifactor model as an approximation to the analysis of multitrait-multimethod and monotrait-heteromethod matrices. In particular, the following questions will be addressed: (a) Is it possible to replicate Seo et al.'s results with a different sample selected in a different cultural and linguistic context? That is, are the method effects for the SIS-C a stable phenomenon, or do they depend on certain assessment conditions? (b) In what proportion are the SIS-C scores due to SN and in what proportion are these scores associated with the specific artefacts of each evaluation method? (c) To what extent does the presence of method variance affect the precision with which the SIS-C measures each substantive dimension of SN? (d) Are the method factors in the SIS-C sufficiently important to bring about modifications in the scale?

Procedure

Participants

The initial sample comprised 833 children from ten Regional Districts (Autonomous Communities) in Spain. Those cases not complying with the three following inclusion criteria were excluded (n = 19): (a) being between 5 and 16 years old at the moment of the evaluation; (b) being in possession of a professional report certifying the presence of ID; (c) having answered all questions in the central axis of the instrument.

The final sample comprised 814 children (35.1% girls) between 5 and 16 years old (M = 11.5; SD = 3.44). Although there was no specific measurement of intellectual and adaptive functioning available for most of the children, professional reports judged that 24% had mild intellectual disability, 40.8% moderate and 35.2% severe/profound.A high percentage of the children presented a form of ID associated with Autistic Spectrum Disorder (30.5%), 13.6% with Down Syndrome, 12.4% with Cerebral Palsy, and 4.3% with other conditions. In 39% of the sampled population, no specific syndrome was detected. Of the children, 60.6% were attending special education programs, 22% regular schools, and 17.4% other educational programs. In general, the assessed children were not using any kind of assistive technologies in their daily lives (80.7%), and the great majority of them were living in their family household (95.6%). Of the children, 97.8% were native speakers of Spanish.

The participants were recruited from the institutions that voluntarily agreed to collaborate with the research team. A letter was sent to several centers and associations providing support in all the Autonomous Communities in Spain. Moreover, an advertisement requesting collaboration was published on the web page of the Institute for Community Inclusion (INICO, University of Salamanca, Spain), so many institutions and professionals interested in participating could establish direct contact with the research team. After the first contact, all the institutions that had shown interest and agreed to participate received a letter of instructions and an informed consent form for the families of all children between five and 16 years old with an ID informing them of the confidentiality of the data. Several institutions participated in the study (n = 48), with an average of 15 participants per institution, and the number of participants in each institution ranged between three and 63.

In order to be selected as an informant, professionals had to comply with the following criteria: (a) being acquainted with the child with ID for at least three months; and (b) having had the opportunity to observe the child in more than one typical context of his/her daily life. In general, the main informant was a professional (teachers in 63% of cases, followed by other support professionals such as educators/instructors [10%], psychologists [8%], logopedists/speech and language therapists [6%], and therapeutic pedagogy teachers [5%]). Moreover, the research team tried to assign two evaluators to each child/adolescent and to obtain a commitment from a member of their families, while maintaining the option to get an assessment from a second professional for those cases in which family collaboration was not possible. The participation of a second informant was obtained for 732 of the applications, of which 460 were relatives (the child's mother in 84% of cases).

Instrument

The SIS-C (Thompson et al., 2016) is a tool designed to evaluate the pattern and intensity of support in children and adolescents with ID (see introduction). It comprises two distinct sections: (1) Exceptional Medical and Behavioral Needs; and (2) Support Needs Scale. The first section considers 32 items through a progression of three scores

(0 = no support needed; 1 = some support needed; 2 = extensive support needed) for the purpose of assessing whether there are specific medical conditions and/or disruptive behaviors that could directly affect the intensity of support needed by the individual. The second section constitutes the central axis of the instrument and measures support needs across 61 activities divided into seven daily contexts (Home Living; Community and Neighborhood; School Participation; School Learning; Health and Safety; Social Activities; and Advocacy). Each dimension is evaluated through a set of eight to nine items, and each item must be evaluated on the basis of three indexes (type, frequency, and daily support time), on a five-point scale (0-4) in which a higher score indicates a higher intensity of support needed (see table 1).

(please insert table 1 approximately here)

The sum of the scores obtained in each index leads to an overall score for each item, obtained by adding up the three indexes. At the same time, the sum of the items of each sub-scale will form the direct score of the corresponding sub-scale, which, through indicators, will be transformed into standardized scores allowing the generation, in turn, of a global standardized score (support needs index). The Spanish adaptation of the SIS-C was carried out in accordance with Tassé and Craig's (1999) recommendations for the purpose of adapting measurement tools to a different context and to the guidelines of the International Test Commission (Muñiz & Hambleton, 1996; Muñiz et al., 2013), including the publication of a pilot study prior to the validation of the instrument (Guillén et al., 2015; Verdugo et al., 2012).

Data Analysis

This study analyzed the structure of the SIS-C from a multitrait-multimethod perspective (MTMM)[1]. Considering a number of traits measured with different methods, the MTMM framework (Campbell & Fiske, 1959) allows the evaluation of convergent validity (i.e. which evaluation methods may present concurrent validity while measuring the same construct) and discriminant validity (i.e. the extent to which different constructs measured with the same method are empirically separable). MTMM approximation has been integrated in confirmatory factorial analysis (Jöreskog, 1971; Marsh & Hocevar, 1988; Widaman 1985) by developing taxonomies of nested models. In addition to this, the CFA models based on the MTMM matrix allow the variance of each indicator to be split into independent sources (substantive variance due to trait and systematic variance - shared between several traits - due to the use of a common evaluation method).

Taking into account that in the SIS-C, each of the items gets three answers, a complete factorial model would require the inclusion of 183 items with seven substantive dimensions and three method factors; in a confirmatory factorial model with categorical variables, this corresponds to the estimation of at least 1029 parameters. Given the practical difficulty of fitting such a parameterized model (Marsh et al., 2010; Morin, Arens, & Marsh, 2016), two alternative approaches were chosen: (a) estimating the complete model using parcels instead of items, and (b) separately estimating seven bifactor models, one for each substantive dimension, using the responses to items (via pseudo MTMM models (Little, 2013), replicating the approach followed by Seo et al., 2016). Finally, the degree of convergence of the results of both approaches was verified.

[1] We refer to the dimensions of support needs with the generic term "traits" (as substantive dimensions) in order to maintain the terminology usually adopted in MTMM analysis.

Fit and model properties (parcels)

In the case of the complete model, three confirmatory factorial models were assessed (figure 1 shows a conceptual representation of the measurement models).

In model 1 (M1: correlated traits, correlated methods), seven dimensions of SN and three method factors (type, frequency, and time) were specified. Each SN domain was measured using three indicators corresponding to the means in the sub-scale for each of the three methods used. For its part, each method factor was measured using the indicators referring to the method used. For example, the "frequency of support" method factor was measured by the mean of the responses to all frequency items, regardless of their substantive dimension. Finally, we allowed correlations between traits and between method factors, but not between traits and method factors. Model 2 (M2: Correlated traits, uncorrelated methods) was identical to M1 except that the correlations of the method factors were set to zero (i.e. it was specified that the sources of variance of method were entirely independent of each other). In model 3 (M3: Correlated traits, no methods), all the loadings of the method factors were set to zero. M3 is equivalent to a model that hypothesizes the nonexistence of method factors (i.e., there is no systematic variance associated with the method, and the common variance is entirely explained by the traits).

(please insert figure 1 approximately here)

Considering these models, we investigated the internal structure of the SIS-C in two steps: first by (1) comparing the fit between pairs of nested models (M1 vs. M2, and M1 vs. M3), and secondly (2) by evaluating the parameters obtained from M1, with special attention on the distribution and size of the factorial loadings and the total variance explained by each of the factors. To do this, we calculated the proportion of total variance (ETV) associated with each substantive and method factor, and captured by each indicator.

Fit and model properties (items)

In the case of analysis at the item level, seven bifactor models composed of a general factor representing the substantive dimension, and three method factors correlated with each other, but orthogonal to the general factor (type, frequency, and daily time), were assessed. Several of the confirmatory bifactor models did not reach convergence, presumably because there was not enough systematic variance to define a stable method factor relative to the support frequency. As a consequence, we decided to make the analysis more flexible by using exploratory structural equation models (ESEM, Asparouhov & Muthén, 2009; Marsh, Morin, Parker, & Kaur, 2014).

The confirmatory bifactor models were replaced by seven ESEM bifactor models with target rotation (M4 to M10). The ESEM allows the estimation of all possible cross-loadings in the model; moreover, target rotation allows the nearest rotated solution to a matrix of cross-loadings and primary loadings specified a priori to be obtained, which enables the use of ESEM in a confirmatory manner (Asparouhov & Muthen, 2009). Figure 2 shows a conceptual representation of the estimated ESEM model for each SN factor.

(please insert figure 2 approximately here)

Taking into account the factor loadings obtained in each ESEM bifactor model, we estimated the extent to which the method variance affected the measurement of the primary construct by calculating the explained common variance (ECV; ten Berge & Socan, 2004) and the hierarchical omega coefficient (ωh; Zinbarg, Revelle, Yovel, &

Li, 2005; Zinbarg, Yovel, Revelle, & McDonald, 2006) for each of the seven SN factors. ECV is the proportion of common variance explained by the general dimension, isolating the effect of the other factors (in ESEM, as well as the effect of cross-loadings). In a bifactor model, the ECV of the general factor can be interpreted as an index of the one-dimensionality of the model, so that values greater than .70 suggest that the measure is highly one-dimensional; consequently, as ECV increases, the scores on the general factor will be progressively more similar to those obtained in a one-dimensional model (Reise, 2012). Logically, high unidimensionality is to be expected in SN models (i.e. it can be expected that the substantive factor, rather than the methods, is the main source of common variance).

ωh is the reliable systematic variance ratio in unit-weighted composite raw scores that can be attributed to the general factor. Consequently, ωh is an estimator of the accuracy with which the scores in the general factor (i.e. each SN dimension) reflect the position of the subject in that same latent variable once the effect of the method factors, the cross-loadings, and the residual variance has been controlled. A high ωh (> .70; Reise, 2012) is necessary to ensure adequate reliability of the substantive factor. A low ωh (<50; Gignac & Watkins, 2013) implies that the total scores on the scale are not effectively measuring the substantive dimension.

In order to evaluate the fit of the models, we followed the suggestions made by Hu and Bentler (1999), Marsh and Hau (1996), Yu (2002), and Browne and Cudeck (1992) according to which CFI and TLI values above .90 and .95 and RMSEA values below .08 and .05 indicate acceptable and good levels of fit, respectively.

Also, in the models M1 to M3, the BIC and AIC indexes were taken into consideration (smaller values are to be preferred) along with the Satorra-Bentler test of chi-square differences for the contrast of nested models. The analysis of the parcel models was performed using Maximum Likelihood Robust (MLR) to deviations of normality. In the case of the models with items, we opted for weighted least squares with an adjusted mean and variance (WLSMV) given the ordinal nature of the input data (Beauducel & Herzberg, 2006). All the analyses were carried out with Mplus v.7.3 (Muthen & Muthen, 2014). Finally, all the models included a multilevel component (TYPE = COMPLEX) for the purpose of controlling the possible non-independence in the observations made in the same school (48 clusters).

Results

Fit and model properties (parcels)

The MTMM correlation matrix (raw scores), along with the Cronbach's alpha for each sub-scale, can be found in table S1 of the supplementary material. Table 2 shows the fits obtained for all models. The first contrast (M1, correlated methods vs. M3, method variance ignored) allowed an evaluation of the relevance of the method effects. The M3 fit was unacceptable (RMSEA = .21; CFI = .72; TLI = .65) and substantially worse than the M1 fit (RMSEA = .059; CFI = .98; TLI = .97; chi-square= 4320 (24), p <.000, ΔCFI = -.26, ΔRMSEA = .15). This implies that it was necessary to model the systematic variance of method to achieve a good fit to the data, and consequently, to conclude that the method factors in the SIS-C are not ignorable from a model fit point of view.

In the second contrast, M1 (correlated methods) and M2 (non-correlated methods) were involved. M1 fit significantly better than M2 (chi-sq diff = 32.2 (3), p <.01), but not substantially better with regard to the absolute and incremental fit indexes (M2: RMSEA = .061; CFI = .98; TLI = .97; ΔCFI = -.0001; ΔRMSEA = .002). This implies that zeroing the correlations between the methods did not produce a relevant divergence in the model, so that the actual values of the parameters are probably close to zero. At the same time, the correlations between methods in M1 were non-significant except between daily time and intensity ($r = .23$; $p < .05$). These results suggest that each evaluation method had a different and independent impact on the measure of the SN, and that there is no evidence of a common source of method variance.

(please insert table 2 approximately here)

Table 3 shows a summary of the parameters obtained in M1. Figure 3 is a visual representation of the distribution of the variance in the indicators. The factorial loadings of the traits were generally high (range = .99-.64; M = .85; SD = .10). In the model, the total variance explained for each trait was around 10% (range = .13-.10), with no trait with an ETV substantially lower than the rest. Overall, substantive factors accounted for 78% of the total variance, suggesting a high convergent validity of the dimensions of SN, taken as a whole, measured with the SIS-C. We then investigated the size of the method effects in the complete model and in each indicator. The factorial loadings were low for “support intensity” and very low for “support frequency”, with each factor accounting for a very small amount of variance in the complete model (3% and 0.7%, respectively).

However, the loadings of the “daily time support” factor were moderately high (range = .67-.34; M = .54; SD = .11) and their ETV paired with that captured by the traits (10%). For an easier visual inspection, each indicator in Figure 3 is fractionated according to the ratio of variance explained by the trait (white), by the method (grey), and by the uniqueness (black). As can be observed, in all cases in which the SN were evaluated through the “daily time”, the method effect was appreciable (except perhaps in the sub-scale "Home"), surpassing the variance explained by the trait in the case of the "Health" sub-scale and practically matching it in the "Advocacy" sub-scale. In the indicators related to intensity and frequency, we also observed method effects, but in all cases, considerably less than the effect associated with trait. These results point to the existence of relevant method effects in the SIS-C scores. Even though the intensity and frequency scales introduced specific method variance that could to some extent be ignored (because most of the systematic variance was located in the substantive factors), the opposite occurred in the indicators measured by daily time of support. The correlations between factors (M1) and factor loadings of M2 can be found in tables S2 and S3 of the supplementary material.

(please insert table 3 and figure 3 approximately here)

Fit and model properties (items)

All models (M4 to M10) presented fit indexes within acceptable limits. The correlations between method factors were generally low, ranging from .01 to .41 (M = .13, SD = .10). Seven of the 21 correlations were not significantly different from zero ($p > .01$). Table 4 shows the distribution of the total variance, the common variance attributable to the substantive factor (ECV), and the hierarchical omega of each of the seven dimensions depending on the method used for its estimation. In the case of SN measured by means of support intensity, in all sub-scales, the substantive dimension explained most of the total variance in the model, from 67% ("Community") to 80% ("Advocacy"). The systematic variance captured by the method was generally low (range: 12% to 4%). All substantive factors acquired ECV values greater than .80, suggesting that all the scales were highly one-dimensional, despite the inclusion of the method factor. Finally, all the scales acquired elevated hierarchical omega values (>

.85), so the overall scores were rather accurate indicators of the position of the subjects in each latent variable of SN. Very similar results were obtained for the SN measured by frequency: In this case, the method barely explained systematic variance, as the substantive factors were clearly the main source of common variance (ECV> .90 in all cases) and substantive factors were highly reliable (wh> .94 in all cases). In the case of SN measured by daily support time, the results were substantially different. Two subscales ("Home" and "Participation") were less affected by the method factor, acquiring values of hierarchical omega and ECV in an acceptable range. In the other scales, the daily support time factor explained a high percentage of common variance in relation to that captured by the substantive factor, producing a substantial deterioration in the unidimensionality of the model. In one case ("Community"), the variance explained by the method was slightly higher than that explained by the SN factor. The dimensions of SN acquired values of hierarchical omega between the limit (.69, "Social" scale) and unacceptable (.47, "Community" scale). In general terms, these results are the same as those obtained in the parcels model, except for the scale most affected by the method variance ("Health" in the parcel model, and "Community" in the item model, although with little difference in general terms).

(please insert table 4 approximately here)

Discussion

The present study has been devoted to the analysis of the impact of the evaluation method on the estimation of SN in children with ID using the SIS-C scale (Thompson et al., 2016). To this end, we have estimated successive factorial models in which the effect of the three methods (type, frequency, and daily support time) was quantified. Our results indicate that neither type nor frequency produced method effects that significantly distorted the measurement of SN. However, the daily support time evaluation method had undesirable effects on five of the seven subscales of SN, especially in the “Health” and “Community” scales, where the method effect produced a significant deterioration in the reliability and validity with which the direct scores represented the position of the children in the substantive dimensions.

Our results essentially replicate (at least regarding the location of the main sources of method variance) those obtained by Seo et al. (2016): The “Community” and “Health” subscales were the most affected, and the method variance was concentrated in the daily support time items. This constitutes evidence in favor of the stability of the SIS-C method factors, which do not seem to be dependent on the sample, culture, language, or other aspects external to the instrument itself.

Direct SN scores measured through daily support time seem to be substantially contaminated by variance unrelated to the construct of interest, so that these direct scores are distorted estimations of the level of SN in the child. As a consequence, estimating SN by means of the unweighted mean of the raw scores in frequency, intensity, and daily support time, assuming that the three modes of evaluation are equally reliable and valid, does not seem to be an optimal decision. A more valid way of scoring would possibly be by means of scales elaborated from estimations of the position of the normative sample in the substantive factors of bifactor models, isolating the effect of the method.

In preparation for a possible refinement of the scale, an aspect that requires attention is the nature of the strong method effect found in the items evaluated through the daily time scale. Both the results of Seo et al. (2016) and those presented here give an account of the existence and size of the effect, but determining its causes (and therefore their possible solutions) requires further investigation. A first solution would be to eliminate the daily support time subscale. However, time is undoubtedly a relevant aspect for assessing SN (Thompson et al., 2009), and it may be impossible to drop it without adversely affecting the validity of the measure. However, it is necessary to take into account that a method factor is no more than the result of the modeling of the systematic variance empirically associated with the use of a particular method; therefore, the presence of method variance does not necessarily imply that the method is lacking *per se*, because the mere existence of the method factor does not provide information regarding its causes (Podsakoff, McKenzie, Lee, & Podsakoff, 2003). In the case of the daily time spurious factor of the SIS-C, several hypotheses concerning its origin could be attempted. The first has to do with the response scale itself, which is the only aspect in which the items answered through different methods differ. In both frequency and type, SIS-C items have response scales composed of ordinal vague quantifiers (monitoring, verbal / gestural prompting, physical assistance, etc.; infrequently, frequently, very frequently, etc.). However, the response scale of daily time is comprised of ordinal numerical quantifiers (less than 30 minutes, 30 minutes to less than two hours, etc.). Given that the ways in which people answer in the two types of response scales tend to be different (Al Baghal, 2014; Griffin, 2013; Wright, Gaskell, & O'Muircheartaigh, 1997), it is possible that the method factor obeys different ways of processing the classification scale, and that this would lead to different patterns of response to the item. If this were the case, a possible solution would be to homogenize the type of response scale used for the three methods (for example, by transforming the numerical scale of daily time to a scale of vague quantifiers similar to that used in frequency). On the other hand, as suggested by an anonymous reviewer, it is possible that the daily time response categories are insufficient and need to be revised to reflect shorter increments of time, or even transformed into continuous open-ended scales.

However, this hypothesis addresses the fact that the daily time method variance was not homogeneously distributed across all subscales, so that some substantive dimensions were affected relatively little (e.g. "Home"), while others were more strongly affected (e.g. "Community"). This points to a possible interaction between the rating scale and the content of the item. Consider the three items most affected by the method in the "Community" scale ("shopping", "using public services", "using basic community services") and the three least affected on the "Home" scale ("eating", "dressing", "sleeping"). Eating, dressing, and sleeping are daily activities that, especially in the case of children, usually follow a stable routine. As a consequence, it is reasonable to expect that the caregiver will not have special difficulties in quantifying quite accurately the daily time that is usually spent to provide support in these activities. On the contrary, activities such as shopping or using public services, while important, are not usually routine, especially for younger children. It is therefore logical to expect the respondent to be unable to provide an accurate estimation of the daily time spent on activities performed only occasionally, thereby introducing a certain proportion of noise in the responses. Assuming the previous hypothesis to be accurate, the solution would be to maintain the time scale only for those items where it provides information with incremental validity (with a consequent reduction in the time of administration, without significant loss of precision). Such items could be identified by a rational analysis of their content, in combination with certain techniques for analyzing the properties of the item (e.g. retaining those items on the time scale that significantly increase the test information in an item response theory model).

A third possibility may address the distribution of the responses to the individual items (these data can be obtained from the first author). In a closer inspection of the distribution of the responses to each category, it was observed that in the items of type and frequency, the endorsement ratio tended to be higher in the higher categories (3 and 4), as opposed to the items measured through daily time, where the highest response frequencies were observed in the middle and lower part of the scale (categories 2 and 1). This suggests that, direct scores being the same, daily support time items tend to evoke levels of SN that are higher than the frequency and time items. This phenomenon could be giving rise to spurious difficulty factors related to certain clusters of items possessing similar distributional properties, even though the structure underlying each subscale is basically one-dimensional (Green et al., 1997; McDonald, 1967).

Overall, the results of this study contribute to research concerned with the measurement of SN in children with ID through a detailed analysis of the method effects in the SIS-C. From a general perspective, SIS-C has good base properties and is a promising instrument. However, this consideration should not represent an obstacle to continued research regarding the metric properties of the scale in order to progressively improve its validity and precision. Considering the recent elaboration of the SIS-C, more research is needed concerning its psychometric properties, as well as the size and nature of its method effects. The presence of a considerable amount of method variance associated with the daily time rating scale suggests the need for modifications. However, because (a) the size of the method effect varies between SN dimensions, and (b) the daily time scale continues to provide more substantive than artifactual information in most cases, a relevant research objective is to determine which items are most affected by bias. As suggested above, the daily time scale is probably not optimal for items referring to less frequent and / or structured support situations. In future research, psychometric models such as those based on the Item Response Theory could serve to detect those indicators in which the daily time does not add relevant information beyond that provided by the measures of type and frequency. Adequate debugging of the instrument could lead to reduced bias present in SN scores, in addition to a reduction in the length of the instrument without relevant loss of information.

This study has several limitations. First, although the sample size may be considered sufficient for the accurate estimation of the model parameters, the non-probabilistic nature of the sample limits the degree of inference at the population level. Second, the impossibility of correctly estimating the MTMM model on the complete set of items (given the excessive number of parameters) prevented the evaluation of the method effects in the full measurement model. Third, in this study, we have assumed that the specific factors of type, frequency, and time exclusively represent method variance. As an anonymous reviewer has suggested, the three approaches could also be understood as different indicators or attributes of the primary construct. Consequently, it would be necessary to verify that method factors contain only artifactual variance and have not captured substantive information by including relevant predictors (such as the level of adaptive behavior) or criteria (such as personal outcomes) in the model. In the case that method factors (especially daily time) contain mainly artifactual variance, its relation with external variables theoretically related to SN should be close to zero.

References

Al Baghal, T. (2014). Is Vague Valid? The Comparative Predictive Validity of Vague Quantifiers and Numeric Response Options. *Survey research methods, 8* (3), 169-179. doi: 10.18148/srm/2014.v8i3.5813#sthash.rw4Q78c9.dpuf

Arias, V.B., Ponce, F.P., & Núñez, D. E. (2016). Bifactor Models of Attention-Deficit/Hyperactivity Disorder (ADHD) An Evaluation of Three Necessary but Underused Psychometric Indexes. *Assessment*. doi: 10.1177/1073191116679260

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438. doi: 10.1080/10705510903008204

Beauducel, A., & Herzberg, P.Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186-203. doi: 10.1207/s15328007sem1302_2

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258.

Buntinx, W., Van Unen, F., Speth, W., & Groot, W. (2006). The Supports Intensity Scale in the Netherlands: Psychometric properties and applications in practice. *Journal of Applied Research in Intellectual disabilities, 19*, 245-246.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81. doi: 10.1037/h0046016

Chou, Y.C., Lee, Y.C., Chang, S.C., & Yu, A.P. (2013). Evaluating the supports intensity scale as a potential assessment instrument for resource allocation for persons with intellectual disability. *Research on Developmental Disabilities, 34*(6), 2056-2063. doi: 10.1016/j.ridd.2013.03.013

Fortune, J. Levelle, J., Meche, S., Severance, D., Smith, G., & Stern, J. (2008). SIS resource allocation: Four papers on issues and approaches. En R. L. Schalock, J. R. Thompson y M. J. Tassé, (Eds.) *AAIDD SIS White Paper Series.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48*(5), 639-662. doi: 0.1080/00273171.2013.804398

Giné, C., Alomar, E., Balcells, A., Borràs, E., Carasa, P., Carbó, M.,Font, J.,...Vidal, X. (2006). L'Escala d'Intensitat de Suports (EIS). Un instrument per promoure el benestar de les persones amb discapacitat intel·lectual [Supports Intensity Scale (SIS). An instrument for promote the wellbeing of people with intelelctual disabilities]. *Suports, 10* (2), 66-72.

Giné, C., Alomar, E., Carasa, P., Font, J., García, M., Gordillo, …González, M. (2007). *Escala d'intensitat dels suports. Manual de l'usuari. Traducció i adaptació de Support Intensity Scale* [Supports Intensity Scale. Users' Manual. Translation and Adaptation]. Barcelona: Federació Catalana Pro Persones amb Retard Mental.

Giné, C., Font, J., Guárdia-Olmos, Balcells-Balcells, A., Valls, J., & Carbó-Carrete, M. (2014). Using the SIS to better align the funding of residential services to assessed support needs. *Research in Developmental Disabilities, 35*(5), 1144-1151.doi: 10.1016-/j.ridd.2014.01.028

Gómez, L.E., Peña. E., Alcedo, M.A., Monsalve, A., Fontanil, Y. Verdugo, M.A., …Arias, B. (2014). El constructo de calidad de vida en niños y adolescents con discapacidades multiples y profundas: propuestas para su evaluación. *Siglo Cero, 45*(1), 57-69.

Gould, A. (1998). *The Service Need Assessment Profile. Versión 1.0*.[Computer Software and Manual]. Sydney, New South Wales: ATG and associates Pty Ltd.

Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling, 4,* 108-120. doi: 10.1080/10705519709540064.

Griffin, J. (2013). On the Use of Latent Variable Models to Detect Differences in the Interpretation of Vague Quantifiers. *Public opinion quarterly*, *77*(S1), 124-144. doi: 10.1093/poq/nfs072

Guillén, V., Verdugo, M.A, Adam-Alcocer, A., & Giné, C. (2017). Comparison between the spanish and catalan versions of the Supports Intensity Scale for Children (SIS-C). *Psicothema, 29*(1), 126-132. doi: 10.7334/psicothema2016.200

Guillén, V., Verdugo, M.A., Arias B., & Vicente, E. (2015). Desarrollo de una escala de evaluación de necesidades de apoyo de niños y adolescentes con discapacidad intelectual [Development of a support needs assessment scale for children and adolescents with intellectual disabilities]. *Anales de Psicología, 31*(1), 137-144. doi:10.6018/analesps.31.1.166491

Hennike, J. M. (2002). NC-SNAP. *Exceptional Parent, 32*(9), 56-58.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55. doi: 10.1080/10705519909540118

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*(2), 109-133.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York, NY: Guilford Press.

Llewellyn, G., Parmenter, T., Chan, J., Riches, V., & Hindmarsh, G. (2005). *I-CAN: Instrument to classify support needs for people with disability*. Sydney: University of Sydney

Marsh, H. W., & Hau, K. T. (1996). Assessing goodness of fit: Is parsimony always desirable?. *The Journal of Experimental Education, 64*(4), 364-390. doi: 10.1080/00220973.1996.10806604

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*(1), 107. doi: 10.1037/0021-9010.73.1.107

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*(3), 471. doi: 10.1037/a0019227.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology, 10,* 85-110. doi: 10.1146/annurev-clinpsy-032813-153700

McDonald, R. P. (1967). *Nonlinear factor analysis* (Psychometric Monographs, No. 15). Richmond, VA: Psychometric Corporation.

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 116-139. doi: 10.1080/10705511.2014.961800

Muñiz, J., Elousa, P., & Hambleton, R. K.(2013). Directrices para la traducción y adaptación de los test (2nd ed.). *Psicothema, 25*(2), 151-157. doi: 10.7334/psicothema2013.24

Muñiz, J., & Hambleton, R. K.(1996). Directrices para la traducción y adaptación de los tests [Guidelines for the translation and adaptation of tests]. *Papeles del Psicólogo, 66,* 63-70. doi: 10.7334/psicothema2013.24

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology, 63,* 539-569. doi: 10.1146/annurev-psych-120710-100452

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879-903. doi: 10.1037/0021-9010.88.5.879

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate behavioral research, 47*(5), 667-696. doi: 10.1080/00273171.2012.715555

Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of research in personality, 47*(5), 493-504. doi: 10.1016/j.jrp.2013.04.012

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods, 21*(2), 137-150. 10.1037/met0000045.

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H. E, Coulter, D., Craig, E. M., …Yeager, M. H. (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schalock, R. L., Luckasson, R., Tassé, M. J., & Verdugo, M. A. (in press). A holistic theoretical approach to intellectual disability: going beyond the four current perspectives. *Intellectual and Developmental Disabilities.*

Schalock, R. L., Thompson, J. R., & Tassé, M. J. (Eds.) (2008*). International Implementation of the Supports Intensity Scale.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schalock, R. L., Verdugo, M. A., Gómez, L. E., & Reinders, H. S. (2016). Moving us toward a theory of individual quality of life. *American Journal on Intellectual and Developmental Disabilities, 121*(1), 1-12.

Seo, H., Shogren, K. A., Little, T. D., Thompson, J. R., & Wehmeyer, M. L. (2016). Construct Validation of the Supports Intensity Scale-Children and Adult

Versions: An Application of a Pseudo Multitrait-Multimethod Approach. *American Journal on Intellectual and Developmental Disabilities, 121*(6), 550-563. doi: 10.1352/1944-7558-121.6.550

Stancliffe, R.J., & Keane, S. (2000). Outcomes and costs of community living: A matched comparison of group homes and semi-independent living. *Journal of Intellectual and developmental disability, 25,* 281-305. doi:10.1080/13668250020019584

Tassé, M. J., & Craig, E. M. (1999). Critical issues in the cross-cultural assessment of adaptive behavior. En R.L. Schalock (Ed.), *Adaptive behavior and its measurement: Implications for the field of mental retardation.* Washington, DC: American Association on Mental Retardation.

ten Berge, J.M., & Socan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69,* 613-625. doi:10.1007/BF02289858

Thompson, J.R., Bradley, V.J., Buntinx, W.H.E., Schalock, R.L., Shogren, K.A., Snell, M.E., ... Wehmeyer, M.L. (2009). Conceptualizing supports and the support needs of people with intellectual disability. *Intellectual and Developmental Disabilities, 47,* 135-146. doi: 10.1352/1934-9556-47.2.135

Thompson, J.R., Bryant, B., Campbell, E.M., Craig, E.M., Hughes, C., Rotholz, D. A.,…Wehmeyer, M. L. (2004). *Supports Intensity Scale user's manual.* Washington, DC: American Association on Mental Retardation.

Thompson, J.R., Hughes, C., Schalock, R.L., Silverman, W., Tassé, M.J., Bryant, B., …Campbell, E. M. (2002). Integrating supports in assessment and planning.

Mental Retardation, 40 (5), 390–405. doi: 10.1352/0047-6765(2002)040<0390:-ISIAAP>2.0.CO;2

Thompson, J.R., Wehmeyer, M.L., & Hughes, C. (2010). Mind the gap! Implications of person-environment fit models of intellectual disability for students, educators, and schools. *Exceptionality, 18,* 168-181. doi: 10.1080/09362835.2010.513919

Thompson, J. R., Wehmeyer, M. L., Hughes, C., Shogren, K. A., Seo, H., Little, T. D., … Schalock, R. (2016). *Supports Intensity Scale - Children's Version Users Manual.* Washington, D.C.: American Association on Intellectual and Developmental Disabilities.

Thompson, J. R., & Viriyangkura, Y. (2013). Supports and support needs. En Michael L. Wehmeyer (Ed.),*The Oxford Handbook of Positive Psychology and Disability* (pp.317-337). Nueva York: Oxford University Press.

United Nations (1989). Convention on the Rigths of the Child. Retrieved from: http://www.un.org/documents/instruments/docs_en.asp

United Nations (2006). Convention on the Rights of People with Disabilities. Retrieved from: http://www.un.org/documents/instruments/docs_en.asp

van Loon, J. (2009). Uso de la escala de intensidad de apoyos en la práctica: asignación de recursos y planificación de apoyos individuales basados en la escala de intensidad de apoyos [Using the Supports Intensity Scale in Practice. Resource Allocation and Planning Individual Supports based on the Supports Intensity Scale]. *Siglo Cero, 40*(1), 54-66.

Verdugo, M.A, Arias, B., Guillén, V., Seo, H., Shogren, K.A., Shaw, L.A. …Thompson, J.R. (2016). Examining age-related differences in support needs on the Supports Intensity Scale-Children's Version-Spanish translation. *International Journal of Clinical Health Psychology, 16,* 306-314. http://dx.doi.org/10.1016/j.ijchp.2016.06.002

Verdugo, M.A., Arias B., Guillén, V., & Vicente, E. (2014). Escala de Intensidad de Apoyos para Niños y Adolescentes (SIS-C) en el contexto español [The Supports Intensity Scale for children (SIS-C) in the Spanish context]. *Siglo Cero, 45*(1), 24-40.

Verdugo, M. A., Arias, B., & Ibáñez, A. (2007). *Escala de Intensidad de Apoyos (SIS). Manual. Adaptación española* [Supports Intensity Scale-SIS Manual. Soanis Version] Madrid: TEA.

Verdugo, M. A., Arias, B., Ibáñez, A., & Schalock, R. L. (2010). Adaptation and psychometric properties of the Spanish version of the Supports Intensity Scale (SIS). *American Journal of Intellectual and Developmental Disabilities, 115*(6), 496-503. doi: 10.1352/1944-7558-115.6.496

Verdugo, M.A., Guillén, V., Arias, M., Vicente, E., & Badia, M. (2016). Confirmatory factor Analysis of the Supports Intensity Scale for Children. *Research in Developmental Disabilities, 49-50,* 140-152. doi: 10.1016/j.ridd.2015.11.022

Verdugo, M.A., Arias, B., & Navas, P. (2014). La Escala de Diagnóstico de la Conducta Adaptativa (DABS): Aplicaciones prácticas [The Diagnostic Adaptive Behavior Scale (DABS): Practical Usefulness]. *Siglo Cero, 45*(1), 8-23.

Wehmeyer, M. L., Chapman, T. E., Little, T. D., Thompson, J. R., Schalock, R. L., & Tasse, M. J. (2009). The efficacy of the Supports Intensity Scale (SIS) to predict extraordinary support needs. *American Journal on Mental Retardation, 114,* 3–14. doi: 10.1352/2009.114:3-14

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*(1), 1-26.

Wright, D. B., Gaskell, G. D., & O'Muircheartaigh, C. A. (1997). How response alternatives affect different kinds of behavioural frequency questions. British *Journal of Social Psychology, 36(4),* 443-456. doi: 10.1111/j.2044-8309.1997.tb01143.x

Yu, C.Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Doctoral dissertation, University of California Los Angeles).

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's a, Revel- le's b, and McDonald's v H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123!133. doi:10.1007/s11336-003-0974-7

Zinbarg, R., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for vh. *Applied Psychological Measurement, 30,* 121-144. doi:10.1177/0146621605278814

Table 1. SIS-C rating metric

Type of support	Frequency of support	Daily support time
0 = None 1 = Monitoring 2 = Verbal/gestual prompting 3 = Partial physical assistance 4 = Full physical assistance	0 = Negligible 1 = Infrequently 2 = Frequently 3 = Very frequently 4 = Always	0 = None 1 = Less than 30 minutes 2 = 30 minutes to less than 2 hours 3 = 2 hours to less than 4 hours 4 = 4 hours or more

Table 2. Fit of the estimated models

Model	RMSEA	CFI	TLI	Chi-sq	df	Chi-sq/df
M1 (CTCM)	.059	.982	.974	554.3	144	3.8
M2 (CTUM)	.061	.981	.973	587.5	147	4.0
M3 (CTNM)	.217	.722	.652	6599.4	168	39.3
M4 (Home)	.078	.986	.980	1473.3	249	5.9
M5 (Community)	.067	.989	.984	852.3	186	4.6
M6 (Participation)	.066	.990	.986	1114.3	249	4.5
M7 (Learning)	.062	.990	.986	1014.3	249	4.1
M8 (Health)	.067	.985	.978	855.4	186	4.6
M9 (Social)	.073	.985	.979	1324.5	249	5.3
M10 (Advocacy)	.071	.990	.986	1252.6	249	5.0

Note. CTCM=Correlated traits, correlated methods; CTUM=Correlated traits, uncorrelated methods; CTNM=Correlated traits, no methods; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; df = Degrees of freedom;

Table 3. Results for model 1 (correlated traits-correlated methods)

	Loadings		Variance		
Sub-scale	Trait	Method	Trait	Method	Residual
Home type	.95 (.00)	.20 (.02)	.91	.04	.04
Home frequency	.97 (.00)	.16 (.03)	.94	.02	.03
Home daily time	.84 (.01)	.34 (.02)	.71	.12	.16
Community type	.93 (.00)	.27 (.02)	.87	.07	.05
Community frequency	.95 (.01)	.20 (.04)	.91	.04	.03
Community daily time	.72 (.01)	.53 (.02)	.52	.28	.19
Participation type	.93 (.00)	.30 (.02)	.87	.09	.03
Participation frequency	.97 (.01)	.18 (.05)	.94	.03	.02
Participation daily time	.81 (.01)	.46 (.02)	.67	.21	.11
Learning type	.87 (.01)	.41 (.02)	.76	.16	.06
Learning frequency	.95 (.01)	.15 (.06)	.90	.02	.07
Learning daily time	.70 (.02)	.55 (.02)	.50	.30	.19
Health type	.92 (.01)	.33 (.02)	.85	.11	.03
Health frequency	.97 (.00)	.09 (.06)	.95	.01	.03
Health daily time	.64 (.01)	.67 (.01)	.41	.45	.12
Social type	.91 (.01)	.35 (.02)	.83	.12	.03
Social frequency	.99 (.00)	.02 (.06)	.98	.00	.01
Social daily time	.76 (.01)	.57 (.02)	.58	.33	.07
Advocacy type	.89 (.01)	.39 (.02)	.79	.15	.04
Advocacy frequency	.98 (.00)	.03 (.06)	.97	.00	.02
Advocacy daily time	.68 (.02)	.65 (.02)	.46	.43	.10
ETV			.78	.15	.07

Note: All loadings are fully standardized; standard errors are in parenthesis; ETV = proportion of total variance explained in the whole model. Note: In all cases, the sum of the trait, method and residual variances would be 1 taking into account the third and subsequent decimals, not included in the table.

Table 4. Explained variance and reliability of substantive factors

Sub-scale	Factor	Trait	Method	ECV	ωh
Type	Home	.774	.049	.88	.981
	Community	.668	.072	.84	.944
	Participation	.769	.041	.90	.975
	Learning	.728	.117	.84	.863
	Health	.784	.026	.91	.976
	Social	.772	.053	.88	.979
	Advocacy	.799	.053	.91	.962
Frequency	Home	.793	.045	.87	.985
	Community	.709	.063	.84	.956
	Participation	.778	.041	.90	.975
	Learning	.768	.075	.89	.920
	Health	.804	.048	.91	.980
	Social	.784	.048	.88	.979
	Advocacy	.790	.059	.88	.941
Daily time	Home	.634	.114	.77	.849
	Community	.347	.356	.44	.474
	Participation	.542	.150	.69	.763
	Learning	.492	.278	.59	.624
	Health	.452	.323	.53	.570
	Social	.533	.221	.63	.692
	Advocacy	.485	.349	.55	.581

Note: ECV = Explained Common Variance of the general (substantive factor); ωh = Omega hierarchical of the substantive factor