



# Validation of the HRI-24 on Adolescents and Development of a Short Version of the Instrument

Daiana Colledani<sup>1</sup> · Pasquale Anselmi<sup>1</sup> · Lucia Monacis<sup>2</sup> · Egidio Robusto<sup>1</sup> · Bruno Genetti<sup>3</sup> · Alessandra Andreotti<sup>3</sup> · Luis J. Gómez Pérez<sup>4</sup> · Luisa Mastrobattista<sup>5</sup> · Claudia Mortali<sup>5</sup> · Adele Minutillo<sup>5</sup>

Accepted: 14 June 2023  
© The Author(s) 2023

## Abstract

The Hikikomori phenomenon often starts during adolescence and, once it develops, it tends to persist. Thus, having an instrument specifically validated for detecting it at early ages could play a pivotal role to reduce the chronicity risk. This work aims to validate the 24-item Hikikomori Risk Inventory (HRI-24) on adolescents and to develop a short version of it. In Study 1, an exploratory structural equation model was used to evaluate the functioning of the HRI-24 and to select the items for inclusion in the short version. In Study 2, confirmatory factor analyses were run on the short version, and measurement invariance across gender and school levels was investigated. Structural validity and measurement invariance of the HRI-24 were supported. The psychometric properties of the short version, denoted as HRI-15, were satisfactory and analogous to those of the HRI-24, while accuracy and specificity in identifying at-risk individuals were slightly higher. Measurement invariance of the HRI-15 was supported as well. The validation of the HRI-24 on adolescents would help professionals to screen young people at the first onset of the Hikikomori phenomenon, and the short version could be highly useful in large epidemiological and screening studies.

**Keywords** Hikikomori · HRI-24 · HRI-15 · Short scale · Social withdrawal

The Hikikomori phenomenon, also known as social withdrawal, describes adolescents and young adults who remain locked in their homes, unable to work or go to school for months or years. This phenomenon was initially linked to a specific condition of the Japanese

---

✉ Daiana Colledani  
daiana.colledani@unipd.it

<sup>1</sup> Department of Philosophy, Sociology, Education and Applied Psychology (FISPPA), University of Padova, Padua, Italy

<sup>2</sup> Department of Humanities, University of Foggia, Foggia, Italy

<sup>3</sup> Explora Research and Statistical Analysis, Padova, Italy

<sup>4</sup> Novella Fronda Foundation, Padua, Italy

<sup>5</sup> National Centre on Addiction and Doping, Italian National Institute of Health, Rome, Italy

culture (Teo & Gaw, 2010), but it has recently become object of attention also in Western countries (Malagón-Amor et al., 2015). The current debate on the phenomenon is mainly focused on the absence of a clear definition and consensus as a syndrome or a specific cultural condition (Tajan, 2015; Teo & Gaw, 2010). Indeed, some authors differentiate between two types of Hikikomori. The primary type refers to a condition strictly related to behavioral problems, thus excluding mental illness. In line with this position, social withdrawal, which is considered as a primary symptom (Suwa & Suzuki, 2013), characterizes those individuals who avoid the pressures of society, school, and parents by retiring to their own residence for at least six months (Saito, 1998). Two subtypes have been proposed: the ‘hard core’ subtype that includes individuals who never leave their room and never talk to their families, and the ‘soft’ subtype that characterizes those individuals who go out and talk to others occasionally (Heinze & Thomas, 2014). The secondary type of the Hikikomori condition is related to a pervasive developmental disorder (Heinze & Thomas, 2014; Suwa & Suzuki, 2013; Suwa et al., 2003), that is caused by preexisting psychological issues (Kondo et al., 2013), or to a form of “modern-type depression” (Kato et al., 2016).

Although there are no worldwide epidemiological data on Hikikomori, its prevalence has been found to be approximately 0.87–1.2% in Japan, 1.9% in Hong Kong, and 2.3% in Korea. The prevalence from 12.64% to 63.07% in Eastern countries may be linked to differences in inclusion criteria, assessment tools, or recruitment strategies (Pozza et al., 2019).

Concerning construct measurement, three self-report tools have been proposed: the NEET-Hikikomori Risk (Uchida & Norasakkunkit, 2015), the Hikikomori Questionnaire (Teo et al., 2018), and the Hikikomori Risk Inventory (HRI-24; Loscalzo, et al., 2022). Among these scales, only the HRI-24 was developed through the collaboration of Western (Italian) and Eastern (Japanese) researchers. This scale showed satisfactory psychometric properties in both contexts, and it was found to be invariant across Italian and Japanese respondents. This feature of the HRI-24 is very important for its consideration as a promising tool for the study of the Hikikomori phenomenon at an international level. Concerning gender differences, ANOVAs revealed that the HRI-24 total score did not differ across males and females in both the Italian and Japanese samples. However, the authors did not test measurement invariance across gender. As pointed out in the literature, measurement invariance is a crucial scale property for the comparison between groups to be meaningful (Colledani, 2018; Colledani et al., 2019, 2022; Vandenberg & Lance, 2000).

The first aim of this work was the validation of the HRI-24 in the Italian context with an explicit focus on the adolescent population. Since the Hikikomori phenomenon often starts during adolescence (Koyama et al., 2010) and, once it develops, it tends to persist (Hamasaki et al., 2021), having available an instrument specifically validated for the detection of this condition at early ages could play a pivotal role to reduce the chronicity risk (Koyama et al., 2010). This work took into consideration middle and high school students, who were recruited from schools located throughout the entire national territory.

To date, a short version of the HRI-24 does not exist. However, such an instrument could be useful in large epidemiological and screening studies in which several dimensions are investigated, as well as when the participant burden may be high. The second aim of this work was the development of a short version of the HRI-24.

To reach the objectives, two different studies were carried out. In Study 1, the structural validity of the HRI-24 was verified on a large sample of middle and high school students and measurement invariance across gender and school levels was tested. Moreover, a short version of the instrument was developed and validated. In Study 2, the short form was tested on a new large and independent sample of students.

## Study 1

### Method

#### Participants and Procedure

This study was based on the data collected within the project “Generation Z” (“Generazione Z”), a national survey conducted by the Italian National Institute of Health (Istituto Superiore di Sanità) in 2022. The survey included different sections that evaluated students’ demographic characteristics, lifestyles, habits, and subjective feelings. It was administered through an electronic format, in the classroom during class time and in the presence of a trained experimenter who was instructed to help students if necessary. Students, their parents, and school directors were asked to consent to student participation in the study. A total of 312 middle and high schools were invited to participate. At the end, 10.1% middle schools (2 in North West, 4 in North East, 1 in Central Italy, 3 in the South, and 4 in the Islands) and 8.6% high schools (5 in North West, 3 in North East, 1 in Central Italy, 3 in the South, and 3 in the Islands) participated in this study. Data collection was anonymous, and it took place from 26 March to 13 April 2022. The study was approved by the Ethical Committee of the Italian National Institute of Health (prot. PREBIO CE 01.01, March 28, 2022).

The sample included 2,034 students (mean age = 13.97,  $SD = 2.05$ ; females = 1,001, 49.2%; 56 students, 2.8%, did not report their gender) attending middle (first, second, and third grade) or high schools (first, second, third, and fourth grade). Students attending middle school were 1,037 (51.0%; mean age = 12.30,  $SD = 0.935$ ; females = 517, 47.9%; 23 students, 2.2%, did not report their gender), while those attending high school were the remaining 997 (49.0%; mean age = 15.71,  $SD = 1.30$ ; females = 484, 48.5%; 33 students, 3.3%, did not report their gender).

#### Measures

The HRI-24 (Loscalzo et al., 2022) was administered to all participants. It investigates the typical Hikikomori feelings and behaviors, and includes 24 items which measure five factors: Anthropophobia, Agoraphobia, Lethargy, Paranoia, and Depressive Mood. Anthropophobia is the fear of people and social contacts. Agoraphobia is the avoidance of places where it would be difficult to get assistance in case of a panic attack or a high state of anxiety. Lethargy is the feeling of having little energy or of being unable or unwilling to do anything. Paranoia describes feelings of suspicion and distrust of others. Depressive mood is a feeling of sadness and disinterest, with loss of pleasure in life. Anthropophobia, Agoraphobia, and Lethargy are measured by four items each, Paranoia by seven items, and Depressive Mood by five items. All items are scored on a 5-point Likert scale ranging from 1 (*Strongly disagree*) to 5 (*Strongly agree*). According to the authors, a score can be computed for each subscale as well as a total score representing a Hikikomori risk score. Construct validity, factor structure, and internal consistency were verified in Western (Italy) and Eastern (Japan) countries, and showed satisfactory results (Loscalzo et al., 2022).

Demographic variables such as gender and age were investigated with a few close-ended items. One yes–no item (“Have you ever experienced the tendency to lock yourself in your

room for several months, never going out, not even to eat meals or to entertain social relations?") was used to investigate the Hikikomori risk.

## Data Analysis

A five-factor exploratory structural equation model (ESEM) was run on the total sample to explore the structure of the HRI-24. This model was used to test configural, metric, and scalar invariance across gender (males and females) and school level (middle and high schools). All models were run using the robust maximum likelihood estimator (MLR; Yuan & Bentler, 2000). To evaluate the goodness of fit of the models, several fit indexes were inspected:  $\chi^2$ , comparative fit index (CFI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA). A good fit is indicated by non-significant ( $p \geq 0.05$ )  $\chi^2$  values. Since this statistic is sensitive to sample size, the other fit measures were also inspected. CFI values close to 0.95 (0.90 to 0.95 for reasonable fit), and SRMR and RMSEA smaller than 0.06 (0.06 to 0.08 for reasonable fit) were considered indicative of adequate fit (Marsh et al., 2004). For testing the equivalence of nested models in measurement invariance, the tests of change in CFI, RMSEA, and SRMR ( $\Delta$ CFI,  $\Delta$ RMSEA,  $\Delta$ SRMR; Chen, 2007; Cheung & Rensvold, 2002) were used. Invariance is supported by  $\Delta$ CFI values  $\leq |.01|$ , paired with  $\Delta$ RMSEA values  $\leq |.015|$  and  $\Delta$ SRMR values  $\leq |.030|$  for metric invariance or  $\leq |.015|$  for scalar invariance. Mean differences in the scores of the five subscales and the total score across gender and school levels were tested through *t*-tests. Cohen's *d* measures of effect size were also reported ( $d < 0.2$ ,  $0.2 \leq d < 0.5$ ,  $0.5 \leq d < 0.8$ , and  $d \geq 0.8$  denote very small, small, medium, and large effect size, respectively; Cohen, 1988).

Relying on the results of the ESEM, a short version of the HRI-24 was developed by selecting 15 items, three for each of the five subscales. The choice to select three items for each subscale was motivated by the intention to obtain an instrument shorter than the original one and comprising subscales of equal length. The items to be included in the short version, which is denoted as HRI-15, were selected based on several criteria: the strength of the factor loadings on the target factor (selecting items with substantial loadings, even if not necessarily the largest ones), the absence of cross-loadings, the absence of gender and school-level bias, and the relevance of the item content to the target factor. These criteria were employed to develop simple structured measures, ensuring high content validity and avoiding biases related to gender or age.

Concerning reliability, Cronbach's  $\alpha$  coefficients were computed for the total HRI-15 and its subscales and were compared with those of the HRI-24. The Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910) was used for estimating the internal consistency that was expected for a shortened instrument consisting of only three items for each subscale. In addition, composite reliability (CR; Bagozzi & Yi, 1988) was calculated. It is another measure of internal consistency, which is conceptually similar to Cronbach's  $\alpha$  as it represents the ratio of true variance to total variance. However, compared to Cronbach's  $\alpha$ , CR is often considered a better index of internal consistency (Raykov, 2001).

Validity was evaluated through Pearson's correlation coefficients which were calculated between the scores obtained on the full-length and abbreviated scales, with the correction for common items suggested by Levy (1967). Moreover, the average variance extracted (AVE) was calculated for each subscale. AVE is a measure of the proportion of variance captured by a construct relative to the variance attributed to measurement error. In general, values close to 0.50 are considered acceptable and indicative of convergent validity

(Fornell & Larcker, 1981). The square root of the AVE was employed to evaluate discriminant validity. If this value exceeds the highest correlation with any other latent variable, discriminant validity is established at the construct level (Fornell & Larcker, 1981).

The effectiveness of the HRI-15 in identifying at-risk individuals was explored using receiver operating characteristic (ROC) curve analysis. This method allows the detection of the total score, among all possible scores, that best discriminates the individuals based on the presence or absence of the measured characteristic (Zhou et al., 2011). To compute a ROC curve, two types of data are needed, one being represented by the total scores of a group of individuals on the instrument and the other being an external criterion (named gold standard), which indicates the condition of each individual (e.g., at-risk vs non-at-risk). Starting from these data, the ROC curve allows for computing sensitivity (i.e., the proportion of at-risk individuals who are correctly classified as being at-risk) and specificity (i.e., the proportion of non-at-risk individuals who are correctly classified as being non-at-risk) for each of the possible total scores. Once the total score maximizing both sensitivity and specificity is identified, it can be used as cut-off score to perform future classifications. In this study, ROC curve analyses were run on both the HRI-24 and the HRI-15. The external classification criterion used in the two analyses was the binary answer (Yes, No) to the self-report question asking individuals to indicate their tendency to lock themselves in their room for several months. The performance of the HRI-24 and the HRI-15 was compared in terms of sensitivity, specificity, accuracy, and area under the ROC curve (AUC). Accuracy is the proportion of individuals correctly classified as being at-risk or non-at-risk. AUC is a measure of classification accuracy that indicates how much an instrument is capable of distinguishing between at-risk and non-at-risk individuals. The four measures range from 0 to 1, with higher values indicating higher capability of the instrument to correctly classify individuals.

## Results

The ESEM run on the total sample reached a successful fit (Table 1). All items loaded on the intended factor with large and significant coefficients (Table 2). Configural, metric, and scalar invariance across gender and school levels were also supported (Table 1). The *t*-tests revealed that females scored higher than males on both the subscales and the total score (*t* from -12.185 to -16.998, *df*=1,976, *p*<0.001, Cohen's *d* from 0.547 to 0.766), and that high school students scored higher than middle school students on the total score and all subscales (*t* from -4.045 to -8.137, *df*=2,032, *p*<0.001, Cohen's *d* from 0.180 to 0.361), excluding agoraphobia (*t*=-0.354, *df*=2014,316, *p*=0.723, Cohen's *d*=0.016). The effect sizes showed that the mean score differences were medium for the comparisons between males and females, and from very small to small for the comparisons between middle and high school students.

Based on the results of the ESEM on the total sample and on the subsamples by school level and gender, 15 items were selected to compose the HRI-15 (Table 2). Since no item showed school-level or gender bias, the selection was performed considering the other selection criteria. For the subscales measuring Anthropophobia and Agoraphobia, the items were selected considering the large loadings on the intended factor (that, in these two subscales, turned out to be the largest ones), the simple structure, and the relevance of the item content to the target factor. For the subscale measuring Paranoia, items 10 and 11 were selected for their large loadings on the intended factor (that turned out to be the largest ones), whereas item 15 was selected for its substantial loading on the target factor,

**Table 1** Fit Indices of the ESEM Run on the Total Sample and on Subsamples by School Level and Gender, and Results of Invariance Testing

	$\chi^2$	df	p ≤	RMSEA	CI90% RMSEA	p	CFI	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
Total sample	728.933	166	.001	0.041	0.038, 0.044	1.00	0.97	0.02			
School level invariance											
Middle school	412.313	166	.001	0.038	0.033, 0.042	1.00	0.968	0.021			
High school	451.71	166	.001	0.042	0.037, 0.046	0.999	0.975	0.019			
Configural	863.055	332	.001	0.040	0.036, 0.043	1.00	0.972	0.020			
Metric	1063.955	427	.001	0.038	0.035, 0.041	1.00	0.966	0.033	0.006	0.002	-0.013
Scalar	1144.642	446	.001	0.039	0.036, 0.042	1.00	0.963	0.034	0.003	-0.001	-0.001
Gender invariance											
Males	396.54	166	.001	0.038	0.033, 0.042	1.00	0.963	0.024			
Females	472.536	166	.001	0.043	0.038, 0.048	0.995	0.969	0.020			
Configural	864.469	332	.001	0.040	0.037, 0.044	1.00	0.967	0.022			
Metric	1056.964	427	.001	0.039	0.036, 0.042	1.00	0.96	0.033	0.007	0.001	-0.011
Scalar	1141.566	446	.001	0.040	0.037, 0.043	1.00	0.956	0.035	0.004	-0.001	-0.002

ΔCFI = test of change in CFI; ΔRMSEA = test of change in RMSEA; ΔSRMR = test of change in SRMR

**Table 2** Results of the ESEM Run on the Total Sample

Item	An	Ag	P	L	D	Items selected for the HRI-15 are marked by "✓"
1	0.695***	0.071	-0.016	0.090**	0.000	✓
2	0.829***	0.000	0.027	0.019	-0.007	✓
3	0.633***	0.029	0.014	0.021	0.098*	✓
4	0.549***	0.191***	0.069	-0.03	0.017	
5	0.120	0.546***	-0.007	0.029	0.050	
6	-0.039	0.856***	-0.012	-0.013	0.028	✓
7	0.057	0.599***	0.071	-0.002	0.012	✓
8	0.093	0.674***	0.009	0.10	-0.008	✓
9	0.186***	0.050	0.456***	0.049	0.054	
10	0.002	0.081	0.767***	-0.057	-0.023	✓
11	0.001	0.134*	0.664***	0.052	-0.021	✓
12	0.121*	-0.039	0.648***	0.019	0.061	
13	-0.059	0.012	0.427***	0.046	-0.077	
14	-0.017	-0.053	0.645***	-0.018	0.062	
15	0.133	-0.043	0.455***	0.026	0.092	✓
16	0.052	-0.044	0.121***	0.691***	0.023	
17	0.015	0.053	-0.022	0.798***	0.018	✓
18	0.015	0.005	0.005	0.865***	-0.007	✓
19	-0.043	0.053	0.013	0.557***	0.067	✓
20	-0.103**	0.046	0.161***	0.072	0.614***	
21	0.079	-0.013	0.010	0.020	0.583***	✓
22	0.079	0.057	0.039	-0.064	0.608***	
23	0.081*	-0.010	-0.085**	0.002	0.770***	✓
24	-0.028	0.001	0.030	0.050	0.792***	✓

An = Anthropophobia; Ag = Agoraphobia; P = Paranoia; L = Lethargy; D = Depressive Mood. Correlations between latent factors ranging from .41 to .73

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$

the relevance of its content, and the simple structure. For the subscale measuring Lethargy, items 17 and 18 were selected considering their large loadings on the intended factor (that turned out to be the largest ones) and the simple structure, while item 19 was selected for its substantial loading on the target factor, the simple structure, and the relevance of its content relative to the intended dimension. Finally, for the subscale measuring Depressive Mood, items 23 and 24 were selected for their large loadings on the intended factor, while item 21 was chosen considering its relevant content and simple structure.

The short subscales composed of the selected items showed satisfactory reliability. Cronbach's  $\alpha$  coefficients ranged from 0.72 to 0.91 and were larger than the coefficients for 3-item subscales that were predicted using the Spearman-Brown prophecy formula

**Table 3** Descriptive Statistics, Internal Consistency Coefficients, Correlations and ROC Curve Analyses for Short and Full-Length Scales

	An		Ag		P		L		D		Total scale	
	Full-length		Full-length		Full-length		Full-length		Full-length		Full-length	
	Short	Full-length	Short	Full-length	Short	Full-length	Short	Full-length	Short	Full-length	Short	Full-length
N item	3	4	3	4	3	7	3	4	3	5	5	24
Mean	2.09	2.06	2.08	1.98	2.80	2.84	2.13	2.24	2.17	2.27	2.25	2.35
SD	1.13	1.09	1.16	1.07	1.07	0.95	1.15	1.15	1.16	1.11	0.90	0.86
Cronbach's $\alpha$	0.83	0.86	0.81	0.83	0.72	0.83	0.82	0.86	0.78	0.85	0.91	0.94
Predicted $\alpha$	0.82		0.78		0.68		0.82		0.77		0.90	
CR	0.77	0.77	0.75	0.77	0.67	0.78	0.80	0.82	0.72	0.81	0.94	0.95
AVE	0.54	0.47	0.50	0.46	0.44	0.35	0.57	0.54	0.47	0.46	0.50	0.44
Square root of AVE	0.73	0.68	0.71	0.68	0.66	0.59	0.75	0.74	0.69	0.68	0.71	0.67
Correlation corrected for common items	0.80		0.77		0.60		0.79		0.73		0.88	
ROC-based best cut-off score											$\geq 37$	$\geq 59$
Sensitivity											0.78	0.81
Specificity											0.71	0.68
Accuracy											0.72	0.69
AUC											0.80	0.81

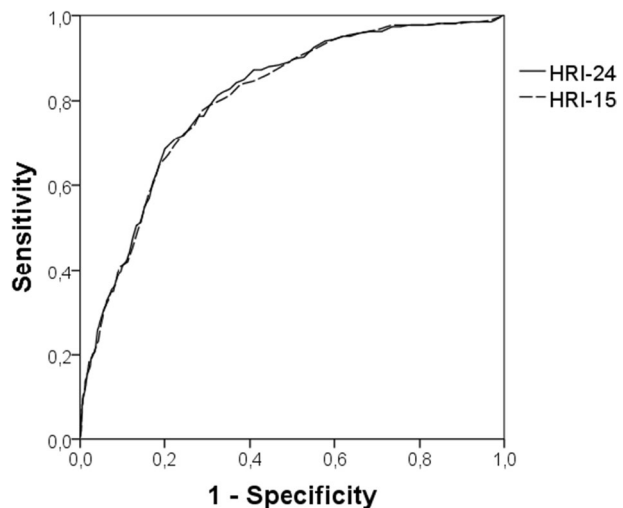
An=Anthrophobia; Ag=Agoraphobia; P=Lethargy; D=Depressive Mood. Predicted  $\alpha$ =Cronbach's  $\alpha$  for 3-item subscales predicted using the Spearman-Brown prophecy formula; CR=composite reliability for the short scales, CR was computed based on the factor loadings of an ESEM model including only the 15 selected items; AVE=average variance extracted for the short scales, AVE was computed based on the factor loadings of an ESEM model including only the 15 selected items); Correlation corrected for common items=correlation between short and full-length scales corrected with the formula suggested by Levy (1967); ROC-based best cut-off score=cut-off score identified using receiver operating characteristic curve analysis on the total score calculated on the full-length and short subscales; AUC=area under the curve

(Table 3). Furthermore, CR coefficients were found to be satisfactory for both the short and full-length scales (Table 3).

The correlation coefficients (corrected for common items) between the short and full-length scales were positive and large (Table 3). A weaker coefficient was observed for the subscale measuring Paranoia ( $r=0.60$ ). This result was expected since this subscale exhibited the lowest Cronbach's  $\alpha$  (a value that enters in Levy's correlation correction formula for common items) and since the full-length and short Paranoia subscales differed by four items (the full-length and short subscales pertaining to the other four factors differed by one or two items only). Concerning convergent validity, the AVE values of the full-length scales were close to 0.50 (ranging from 0.46 to 0.54), excluding those of the subscale measuring Paranoia and the total scale, which were lower (0.35 and 0.44, respectively). Concerning the short scales, all AVE values were close to or higher than 0.50 (ranging from 0.47 to 0.57), excluding that of the subscale measuring Paranoia which was lower (0.44; Table 3; for the short scales, AVE was computed based on the factor loadings of an ESEM model including only the 15 selected items;  $r$ s ranging from 0.44 to 0.71; factor loadings from 0.29 to 0.93). These results suggest sufficient convergent validity for both the short and full-length scales. Concerning discriminant validity, the square roots of AVEs were larger than the correlations with the other latent variables for all the full-length scales, excluding those of Anthropophobia and Agoraphobia, which were weaker than the correlation between them ( $r=0.71$ ), Paranoia, which was weaker than the correlation with Depressive Mood ( $r=0.62$ ), and Depressive Mood, which was weaker than the correlation with Lethargy ( $r=0.73$ ). A better result was observed for the short scales, where all the square roots of AVEs were larger than the correlations with the other latent variables, excluding that of Depressive Mood, which was weaker than the correlation with Lethargy ( $r=0.71$ ). Overall, the results suggest a satisfactory discriminant validity of both the short and full-length scales.

ROC curve analyses revealed that both the HRI-24 and the HRI-15 can be useful instruments to identify at-risk adolescents who express the tendency to lock themselves to avoid interactions and social relationships (Fig. 1). AUCs for both scales were good: 0.81 for the HRI-24 and 0.80 for the HRI-15 (based on the literature, AUCs between 0.80 and 0.90 can

**Fig. 1** ROC Curves for the HRI-24 and the HRI-15. *Note.* The figure depicts the ROC curves for the HRI-24 and the HRI-15. The closer the curves to the upper right corner, the larger the AUC and the more accurate the instrument. AUC = .81 for the HRI-24, AUC = .80 for the HRI-15



be interpreted as good; Safari et al., 2016). These values indicate that the two scales allow for correctly distinguishing between at-risk and non-at-risk individuals in about 80% of cases. The cut-off score of 59, defined on the HRI-24 total score, and the cut-off score of 37, defined on the HRI-15 total score, showed good accuracy, sensitivity, and specificity. The cut-off score defined on the HRI-15 total score slightly outperformed that defined on the HRI-24 total score in accuracy and specificity but fell slightly behind it in sensitivity (Table 3).

## Brief Discussion

The five-factor structure of the HRI-24 was supported and measurement invariance across gender and school levels was confirmed. A short version of the HRI-24 was developed which consists of 15 items, three for each of the five subscales. Both the HRI-24 and the HRI-15 revealed satisfactory reliability and validity coefficients. Compared with the HRI-24, the HRI-15 showed better convergent (i.e., greater AVEs) and discriminant validity.

The HRI-15 showed slightly larger accuracy and specificity in identifying at-risk adolescents, while the HRI-24 was slightly better in terms of sensitivity. This means that the HRI-24 slightly outperforms the HRI-15 in identifying at-risk individuals, while the latter slightly outperforms the former in identifying non-at-risk adolescents.

## Study 2

### Method

#### Participants

This study is based on the data collected within the project “Generation Z” (“Generazione Z”) conducted by the Italian National Institute of Health (*Istituto Superiore di Sanità*) in 2022. This second data collection took place from 6 May to 7 June 2022, with 21.7% of the invited middle schools (4 in North West, 5 in North East, 8 in Central Italy, 9 in the South, and 4 in the Islands) and 13.2% of the invited high schools (7 in North West, 4 in North East, 2 in Central Italy, 5 in the South, and 5 in the Islands) taking part in the study.

The sample included 1,599 students (mean age = 13.79,  $SD = 2.08$ ; females = 726, 45.4%; 48 students, 3.0%, did not report their gender). The students attending middle school were 956 (59.8%; mean age = 12.35,  $SD = 0.98$ ; females = 439, 45.9%; 30 students, 3.1%, did not report their gender), while those attending high school were the remaining 643 (40.2%; mean age = 15.93,  $SD = 1.31$ ; females = 287, 44.6%; 18 students, 2.8%, did not report their gender).

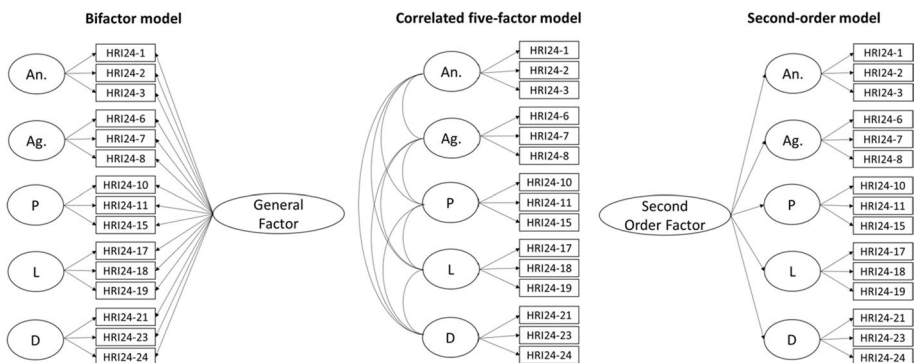
#### Measure

The HRI-15 developed in Study 1 was administered to all participants. The 15 items of the instrument are scored on a 5-point Likert scale (from 1 = “*Strongly disagree*” to 5 = “*Strongly agree*”) and assess the five factors pertaining to Anthropophobia, Agoraphobia, Paranoia, Lethargy, and Depressive Mood (three items for each factor).

## Data Analysis

The factor structure of the HRI-15 was verified through confirmatory factor analyses (CFAs). Three models were tested and compared. In the first model, five correlated factors were defined, each measured by three items (correlated five-factor model). In the second model, the five first-order factors were used as indicators of a second-order factor (second-order model). In the third model, finally, a general factor, measured by all the 15 items of the HRI-15, was modeled together with five non-correlated specific factors, measured by three items each (bifactor model). A graphical representation of the three models is provided in Fig. 2. These models were tested to deeply investigate the scale structure and to determine whether a common underlying construct accounting for the variance in the observed indicators exists, this suggesting that the HRI-15 total score can be used to evaluate the Hikikomori risk. The goodness-of-fit of the three models was evaluated using the same fit indices described in Study 1 and compared using the Akaike information criterion (AIC; Akaike, 1974). A difference in AIC ( $\Delta$ AIC) by 10 or more was considered meaningful (Burnham et al., 2011). All the analyses were run using Mplus7 (Muthén & Muthén, 2012), and the maximum likelihood estimator with adjusted means and variances (MLMV; Muthén & Muthén, 2012) that provides standard errors and statistical tests that are robust to non-normality.

For the bifactor model, a series of indices were also calculated: explained common variance (ECV; Sijtsma, 2009; Ten Berge & Sočan, 2004), percent of uncontaminated correlations (PUC; Rodriguez et al., 2016b), and McDonald's (1999) omega ( $\omega$ ) and omega hierarchical ( $\omega_h$ ) coefficients. ECV is the ratio between the common variance explained by the general factor and the total common variance (Reise et al., 2013a, 2013b; Reise et al., 2013a, 2013b; Rodriguez et al., 2016a). PUC describes the percentage of covariance terms which only reflect the variance from the general factor (Rodriguez et al., 2016b), and measures the biasing effects of forcing bifactor data into a one-dimensional model. According to Rodriguez et al. (2016b), ECV values  $> 0.70$  paired with PUC values  $> 0.70$  can be taken as an indication that a scale, despite the presence of some multidimensionality, can be



**Fig. 2** Graphical Representation of the Bifactor, Correlated Five-Factor, and Second-Order Models. *Note.* An. = Anthropophobia; Ag. = Agoraphobia; P = Paranoia; L = Lethargy; D = Depressive Mood. In the bifactor model, a general factor, measured by all the 15 items of the HRI-15, is modeled together with five non-correlated specific factors, measured by three items each. In the correlated five-factor model, five correlated factors were defined, each measured by three items. In the second-order model, the five first-order factors were used as indicators of a second-order factor

regarded as the measure of an essentially one-dimensional construct. McDonald's (1999)  $\omega$  and  $\omega_h$  coefficients are factor-analytic "model-based" estimates of internal consistency. The former represents the proportion of variance of the scores that can be attributed to all sources of variance (i.e., general and domain-specific factors), whereas the latter quantifies the amount of variance accounted for by the general factor (Revelle & Zinbarg, 2009; Zinbarg et al., 2005, 2007). In the present study,  $\omega$  was computed for the general factor and for each domain-specific factor, whereas  $\omega_h$  was computed for the general factor only. Concerning  $\omega$ , values close to or greater than 0.70 are satisfactory. Concerning  $\omega_h$ , values larger than 0.75-0.80 indicate that the general factor can be interpreted as the measure of a single construct despite multidimensionality (Reise et al., 2013a, 2013b; Reise et al., 2013a, 2013b).

Metric (equality of factor loadings) and scalar (equality of both factor loadings and item intercepts) invariance across gender and school levels were tested. The same indices considered in Study 1 were used to test the equivalence of nested models.

## Results and Brief Discussion

The correlated five-factor model showed an excellent fit ( $\chi^2(80)=219.821$ ,  $p<0.001$ ; RMSEA=0.033 [0.028, 0.038]; CFI=0.985; SRMR=0.025; AIC=66,975.664), items strongly loading on the intended factor (Table 4), and latent factors positively and strongly correlated. The second-order model also reached a good fit ( $\chi^2(85)=377.654$ ,  $p<0.001$ ; RMSEA=0.046 [0.042, 0.051]; CFI=0.968; SRMR=0.038; AIC=67,211.839), with all first-order factors strongly loading on the higher-order dimension (Table 4). However, as indicated by the  $\Delta$ AIC (1236.175), the correlated five-factor model fitted the data better than the second-order model. The bifactor model also showed a good fit ( $\chi^2(75)=297.231$ ,  $p<0.001$ ; RMSEA=0.043 [0.038, 0.048]; CFI=0.976; SRMR=0.031; AIC=67,097.939), with all items significantly and meaningfully loading on both the general and the intended specific factors (Table 4). In this model, the factor loadings pertaining to the general factor were, on the whole, quite similar to those observed on the correlated five-factor model. The bifactor model fitted the data better than the second-order model ( $\Delta$ AIC=122.275), but the correlated five-factor model was superior to both the bifactor model ( $\Delta$ AIC=122.275) and the second-order model ( $\Delta$ AIC=1236.175).

In the bifactor model, PUC was 0.86 and ECV and  $\omega_h$  coefficient of the general factor were 0.73, and 0.89, respectively. Taken together these results suggest that, despite the multidimensional nature of the scale, the common variance accounted for by the general factor can be regarded as essentially one-dimensional (Reise et al., 2013a, 2013b; Rodriguez et al., 2016a).

Having chosen the correlated five-factor model as the best fitting model, it was used to test gender and school-level invariance. The results confirmed full metric and scalar invariance across gender and school levels (Table 5).

## Conclusion

This work aimed to validate the HRI-24 on adolescents and to develop and validate a short version of the instrument. To these aims, two studies were carried out that used two large and nationwide samples of Italian adolescents attending middle and high schools. The results of Study 1 supported the structural, convergent, and discriminant validity of the HRI-24 in adolescent samples. Evidence of configural, metric, and



**Table 5** Invariance of the Correlated Five-Factor Structure Across Gender and School Level

	$\chi^2$	df	$p \leq$	RMSEA	CI90% RMSEA	<i>p</i>	CFI	SRMR	$\Delta$ CFI	$\Delta$ RMSEA	$\Delta$ SRMR	
Gender invariance												
Males	138.134	80	0.001	0.030	0.021, 0.038	1.000	0.980	0.031				
Females	174.642	80	0.001	0.040	0.032, 0.049	0.975	0.979	0.033				
Configural	305.29	160	0.001	0.034	0.028, 0.040	1.000	0.979	0.032				
Metric	329.15	170	0.001	0.035	0.029, 0.040	1.000	0.977	0.037	-0.002	0.001	0.005	
Scalar	372.962	180	0.001	0.037	0.032, 0.043	1.000	0.973	0.042	-0.004	0.002	0.005	
School level invariance												
Middle school	163.054	80	0.001	0.033	0.026, 0.040	1.00	0.982	0.028				
High school	127.275	80	0.001	0.030	0.020, 0.040	1.00	0.987	0.027				
Configural	288.564	160	0.001	0.032	0.026, 0.038	1.00	0.985	0.028				
Metric	301.669	170	0.001	0.031	0.025, 0.037	1.00	0.984	0.031	-0.001	-0.001	0.003	
Scalar	317.493	180	0.001	0.031	0.025, 0.036	1.00	0.984	0.032	0.000	0.000	0.001	

$\Delta$ CFI = test of change in CFI;  $\Delta$ RMSEA = test of change in RMSEA;  $\Delta$ SRMR = test of change in SRMR

scalar invariance was also found between gender and school levels, indicating that the scale has the same functioning across the considered groups. However, mean score differences were observed across gender and school levels. Another relevant contribution provided by this study is the development of a short version of the instrument, which was called HRI-15. Its psychometric properties were satisfactory and analogous to those of the full-length scale. A final merit of this study is that, for the first time, empirical cut-off scores were defined (59 and 37 for the HRI-24 and the HRI-15, respectively), which allow for discriminating against at-risk and non-at-risk adolescents. Both the HRI-24 and the HRI-15 showed satisfactory accuracy, specificity, and sensitivity. In this respect, the performance of the two scales was very similar. However, while the HRI-24 turned out to be slightly better in identifying at-risk individuals, the HRI-15 revealed to be slightly better in identifying non-at-risk adolescents. Researchers should choose between the HRI-15 and the HRI-24 based on their specific needs. If reducing the respondent burden is a crucial consideration, the HRI-15 may be preferred. On the other hand, if greater sensitivity is desired, the HRI-24 could be the preferable option. Study 2 validated the HRI-15 on a different large sample of adolescents. The factor structure of the instrument was verified through CFAs. The results indicated that the correlated five-factor model outperformed the second-order and bifactor models. Moreover, the bifactor model allowed for observing that the general factor, despite the multidimensional nature of the instrument, accounts for a large portion of the variance. This provides empirical evidence that the HRI-15 total score can be used to evaluate the Hikikomori risk, together with the scores on the five subscales. Finally, the results of this study supported measurement invariance of the HRI-15 across gender and school levels.

Although the findings of this work are relevant, some limitations could be highlighted. A single self-report item (“Have you ever experienced the tendency to lock yourself in your room for several months, never going out, not even to eat meals or to entertain social relations?”) was used to investigate the Hikikomori risk. This was done because a consensus gold standard for Hikikomori risk does not yet exist (Teo et al., 2018). Future studies could be devoted to further validate the HRI-24 and the HRI-15, as well as the empirical cut-off scores defined on their total scores, using widely accepted gold standards, if they become available, or individuals from the clinical population. Another limitation is the lack of empirical evidence of validity in comparison to other measures of the same construct or related constructs, such as depression, poor emotional regulation, gaming addiction, or internet addiction (Lin et al., 2022). Future studies are advocated that examine this aspect of validity. In the present work, females were found to score higher than males on all HRI-24 subscales and on the total score, whereas in the work of Loscalzo et al. (2022), such gender differences were not observed. Future studies should further explore gender differences, as well as try to replicate our findings in cross-cultural contexts. Finally, future studies can be aimed to develop even shorter versions of the instrument to use in large epidemiological studies. For instance, it would be interesting to develop a very short scale, including only one item from each subscale. To such purpose, the most suitable items could be selected relying on the I-ECV (item-explained common variance) indices from a bifactor model since they allow for identifying the items that are most adequate to develop an essentially unidimensional measure (Stucky et al., 2013).

Overall, this work provides a relevant contribution to the literature on the Hikikomori phenomenon by validating, on adolescent samples, a scale that has been found to be adequate in both Eastern and Western contexts. This contribution could foster the worldwide expansion of the research on this increasingly alarming phenomenon. Moreover, the validation of the scale

on adolescents would help professionals to screen young people at the first onset of the phenomenon in order to reduce the chronicity risk. To such purpose, this work provides other two relevant contributions by defining a short version of the instrument, which could be highly useful in large epidemiological and screening studies, and by providing cut-off scores that can be used to identify at-risk adolescents. After identifying them, the specific factors that are responsible for the onset of the Hikikomori condition can be singled out and, based on them, policymakers can develop appropriate prevention campaigns.

**Funding** Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement. This work was supported by Presidenza del Consiglio dei Ministri, Dipartimento delle Politiche Antidroga, Italy.

**Data Availability** The data are available from the corresponding author upon reasonable request.

## Declarations

**Ethical Approval** The study was approved by the Ethical Committee of the Italian National Institute of Health (prot. PREBIO CE 01.01, March 28, 2022) and it is conformed to the provisions of the Declaration of Helsinki.

**Informed Consent** All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from the people included in the study, according to the indication of the ethical committee.

**Conflict of Interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. [https://doi.org/10.1007/978-1-4612-1694-0\\_16](https://doi.org/10.1007/978-1-4612-1694-0_16)
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, 74–94. <https://doi.org/10.1007/BF02723327>
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322. <https://doi.org/10.1111/j.2044-8295.1910.tb00207.x>
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1), 23–35. <https://doi.org/10.1007/s00265-010-1029-6>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Colledani, D. (2018). Psychometric properties and gender invariance for the Dickman Impulsivity Inventory. *Testing, Psychometrics, Methodology in Applied Psychology*, 25, 49–61. <https://doi.org/10.4473/TPM25.1.3>
- Colledani, D., Anselmi, P., & Robusto, E. (2019). Using multidimensional item response theory to develop an abbreviated form of the Italian version of Eysenck's IVE questionnaire. *Personality and Individual Differences*, 142, 45–52. <https://doi.org/10.1016/j.paid.2019.01.032>
- Colledani, D., Meneghini, A. M., Mikulincer, M., & Shaver, P. R. (2022). The Caregiving System Scale: Factor structure, gender invariance, and the contribution of attachment orientations. *European Journal of Psychological Assessment*, 38(5), 385–396. <https://doi.org/10.1027/1015-5759/a000673>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
- Hamasaki, Y., Pionnié-Dax, N., Dorard, G., Tajan, N., & Hikida, T. (2021). Identifying social withdrawal (hikikomori) factors in adolescents: Understanding the hikikomori spectrum. *Child Psychiatry & Human Development*, 52, 808–817. <https://doi.org/10.1007/s10578-020-01064-8>
- Heinze, U., & Thomas, P. (2014). Self and salvation: Visions of hikikomori in Japanese manga. *Contemporary Japan*, 26, 151–169. <https://doi.org/10.1515/cj-2014-0007>
- Kato, T. A., Hashimoto, R., Hayakawa, K., Kubo, H., Watabe, M., Teo, A. R., & Kanba, S. (2016). Multidimensional anatomy of 'modern type depression' in Japan: A proposal for a different diagnostic approach to depression beyond the DSM-5. *Psychiatry and Clinical Neurosciences*, 70(1), 7–23. <https://doi.org/10.1111/pcn.12360>
- Kondo, N., Sakai, M., Kuroda, Y., Kiyota, Y., Kitabata, Y., & Kurosawa, M. (2013). General condition of hikikomori (prolonged social withdrawal) in Japan: Psychiatric diagnosis and outcome in mental health welfare centres. *International Journal of Social Psychiatry*, 59(1), 79–86. <https://doi.org/10.1177/0020764011423611>
- Koyama, A., Miyake, Y., Kawakami, N., Tsuchiya, M., Tachimori, H., Takeshima, T., World Mental Health Japan Survey Group. (2010). Lifetime prevalence, psychiatric comorbidity and demographic correlates of "hikikomori" in a community population in Japan. *Psychiatry Research*, 176(1), 69–74. <https://doi.org/10.1016/j.psychres.2008.10.019>
- Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*, 23(1), 84–86. [https://doi.org/10.1002/1097-4679\(196701\)23:1%3c84::aid-jclp2270230123%3e3.0.co;2-2](https://doi.org/10.1002/1097-4679(196701)23:1%3c84::aid-jclp2270230123%3e3.0.co;2-2)
- Lin, P. K., Koh, A. H., & Liew, K. (2022). The relationship between Hikikomori risk factors and social withdrawal tendencies among emerging adults—An exploratory study of Hikikomori in Singapore. *Frontiers in Psychiatry*, 13, 1065304. <https://doi.org/10.3389/fpsy.2022.1065304>
- Loscalzo, Y., Nannicini, C., Huai-Ching Liu, I. T., & Giannini, M. (2022). Hikikomori Risk Inventory (HRI-24): A new instrument for evaluating Hikikomori in both Eastern and Western countries. *International Journal of Social Psychiatry*, 68(1), 90–107. <https://doi.org/10.1177/00207640209758>
- Malagón-Amor, Á., Córcoles-Martínez, D., Martín-López, L. M., & Pérez-Solà, V. (2015). Hikikomori in Spain: A descriptive study. *International Journal of Social Psychiatry*, 61(5), 475–483. <https://doi.org/10.1177/0020764014553003>
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320–341. [https://doi.org/10.1207/s15328007sem1103\\_2](https://doi.org/10.1207/s15328007sem1103_2)
- McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables. User's guide* (7th ed.). Muthén & Muthén.
- Pozza, A., Coluccia, A., Kato, T., Gaetani, M., & Ferretti, F. (2019). The 'Hikikomori' syndrome: worldwide prevalence and co-occurring major psychiatric disorders: a systematic review and meta-analysis protocol. *BMJ Open*, 9(9), e025213. <https://doi.org/10.1136/bmjopen-2018-025213>
- Raykov, T. (2001). Bias of coefficient  $\alpha$  for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, 25(1), 69–76. <https://doi.org/10.1177/01466216010251005>
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>

- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5–26. <https://doi.org/10.1177/0013164412449831>
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154. <https://doi.org/10.1007/s11336-008-9102-z>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine; part 5 receiver operating curve and area under the curve. *Emergency*, 4(2), 111–113. <https://doi.org/10.22037/aaem.v4i2.232>
- Saito, T. (1998). *Shakaiteki hikikomori: Owaranai shishunki* [Societal hikikomori: Unending adolescence]. PHP-Kenkyujo.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120. <https://doi.org/10.1007/S11336-008-9101-0>
- Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Stucky, B. D., Thissen, D., & Orlando Edelean, M. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, 37(1), 41–57. <https://doi.org/10.1177/0146621612462759>
- Suwa, M., Suzuki, K., Hara, K., Watanabe, H., & Takahashi, T. (2003). Family features in primary social withdrawal among young adults. *Psychiatry and Clinical Neurosciences*, 57(6), 586–594. <https://doi.org/10.1046/j.1440-1819.2003.01172.x>
- Suwa, M., & Suzuki, K. (2013). The phenomenon of “hikikomori” (social withdrawal) and the socio-cultural situation in Japan today. *Journal of Psychopathology*, 19, 191–198.
- Tajan, N. (2015). Social Withdrawal and Psychiatry: A Comprehensive Review of Hikikomori. *Neuropsychiatrie de l'enfance et de l'adolescence*, 63, 324–331. <https://doi.org/10.1016/j.neurenf.2015.03.008>
- Ten Berge, J. M., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613–625. <https://doi.org/10.1007/BF02289858>
- Teo, A. R., Chen, J. I., Kubo, H., Katsuki, R., Sato-Kasai, M., Shimokawa, N., & Kato, T. A. (2018). Development and validation of the 25-item Hikikomori Questionnaire (HQ-25). *Psychiatry and Clinical Neurosciences*, 72(10), 780–788. <https://doi.org/10.1111/pcn.12691>
- Teo, A. R., & Gaw, A. C. (2010). Hikikomori, a Japanese culture-bound syndrome of social withdrawal?: A proposal for DSM-5. *The Journal of Nervous and Mental Disease*, 198(6), 444–449.
- Uchida, Y., & Norasakkunkit, V. (2015). The NEET and Hikikomori spectrum: Assessing the risks and consequences of becoming culturally marginalized. *Frontiers in Psychology*, 6, 1117. <https://doi.org/10.3389/fpsyg.2015.01117>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. <https://doi.org/10.1177/109442810031002>
- Yuan, K. H., & Bentler, P. M. (2000). 5. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30(1), 165–200. <https://doi.org/10.1111/0081-1750.000>
- Zhou, X. H., McClish, D. K., & Obuchowski, N. A. (2011). *Statistical Methods in Diagnostic Medicine*. Wiley.
- Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating  $\omega_h$  for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement*, 31(2), 135–157. <https://doi.org/10.1177/0146621606291558>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega$  H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. <https://doi.org/10.1007/s11336-003-0974-7>