Research in Psychology and Behavioral Sciences, 2025, Vol. 13, No. 1, 9-15
Available online at http://pubs.sciepub.com/rpbs/13/1/2
Published by Science and Education Publishing
DOI:10.12691/rpbs-13-1-2

Validation of A Physical Activity Scale for Older Adults Participating in the Health and Retirement Study

Peter D. Hart[1,2,*]

[1]Health Promotion Research, Havre, Montana, USA
[2]Kinesmetrics Lab, Tallahassee, Florida, USA
*Corresponding author: pdhart@outlook.com

Received November 16, 2025; Revised December 18, 2025; Accepted December 24, 2025

Abstract **Background:** The surveillance of physical activity (PA) at the population level generally involves administering a small number of survey questions. The combining of multiple related items to create a scale that yields a score has many psychometric benefits. **Purpose:** The aim of this research was to validate a new scale measuring physical activity (PA) using items contained in a large national survey of older adults. **Methods:** Data from 12,145 adults 50+ years of age participating in the 2022 Health and Retirement Study were used. The assessment strategy involved six steps: 1) defining the PA scale (PAS) items and categories, 2) factor analysis, 3) internal consistency reliability, 4) item response theory (IRT) analysis, 5) construct validity correlations, and 6) modeling PAS scores with a general health (GH) outcome. Polychoric correlations between items were used for the analyses. A graded response model (GRM) for polytomous items was employed for the IRT analysis. Multinomial logistic regression was used to model GH categories with both categorical and numeric PAS scores. **Results:** The PAS included three items of vigorous (VPA), moderate (MPA), and light PA (LPA), each with the same rating scale consisting of "inactive," "low/moderately active," and "highly active." Factor analysis retained a single factor with 70% explained variance, whilst the reliability coefficient for items was 0.79. IRT calibration showed category thresholds ranging from -1.89 to 1.07 and item discrimination parameters between 1.37 and 5.27. IRT theta scores correlated with the PAS sum score (r=0.96), age (r=-0.22), GH (r=0.41), and timed walk performance (r=0.38). Modeling showed that for each point increase in the numeric PAS score, odds of poor (OR=0.31, 0.27-0.34), fair (OR=0.47, 0.43-0.52), good (OR=0.60, 0.56-0.65), and very good (OR=0.79, 0.73-0.85) GH, as compared to excellent (reference), decreased. **Conclusion:** These results support the use of a simple 3-item PAS to measure PA in older adults.

Keywords: *Physical activity, Self-rated health, Measurement, Gerontology*

Cite This Article: Peter D. Hart, "Validation of A Physical Activity Scale for Older Adults Participating in the Health and Retirement Study." *Research in Psychology and Behavioral Sciences*, vol. 13, no. 1 (2025): 9-15. doi: 10.12691/rpbs-13-1-2.

1. Introduction

Several large national surveys repeatedly collect data to assess and examine physical activity (PA) in older U.S. adult populations [1]. Most of these surveys use questionnaire items for participants to self-report their PA at various amounts and intensities. Some surveys are able to use questionnaire responses to calculate specific variables related to PA, such as moderate-to-vigorous PA (MVPA) and meeting PA guidelines (MPAG) status [2]. Single questions have also been used in some surveys to assess PA participation [3]. Another approach, however, is to consider these survey questions as potential items for a behavioral scale targeting a PA construct.

There are many psychometric advantages to using multiple items for the assessment of a single construct [4,5]. First, multiple items forming a single trait can be assessed for study-level item reliability using measures such as Cronbach's alpha. Second, because items from a scale can be summed to provide a score, any random error associated with each item will likely average out, thus creating a more reliable score. Third, including multiple items in a scale can improve the scale's ability to target the full range of the true latent construct. Fourth, multiple items increase the amount of information collected regarding a trait, thus improving the score's ability to differentiate and categorize respondents. In sum, the advantages of combining survey questions into a multi-item PA scale include the ability to assess the measure for reliability, improve its reliability, enhance its construct validity, and increase its sensitivity. With these benefits in mind, the purpose of this study was to develop and validate a new multi-item scale to measure PA in older adults using items contained in a large national survey.

2. Methods

Study design and data

This cross-sectional research used data from the publicly available 2022 Health and Retirement Study (HRS). The HRS is a longitudinal survey that represents U.S. adults over 50 years of age [6]. This research specifically used the HRS core survey, which is available biennially and contains sociodemographic factors, health behaviors, physical measures, and psychosocial constructs. The HRS began collecting data in 1992 from adults 51 to 61 years of age and their spouses/partners of any age. HRS has since added participants to its core survey every two years, with the 2022 dataset including adults from Generation X. All HRS core respondent datasets were merged for this study with the cross-wave tracker file. Participants were included in this research if they were 50+ years of age and had complete PA data.

Physical activity (PA) items

Three different PA items targeting different intensities were available in the HRS physical health dataset and thus used to develop the PA scale (PAS). The first question asked participants how often they participated in sports or activities that were vigorous and gave examples such as jogging, cycling, and digging with a shovel. The second question asked participants how often they participated in sports or activities that were moderately energetic and gave examples such as gardening, walking, and floor exercises. The third question asked participants how often they participated in sports or activities that were mildly energetic and gave examples such as vacuuming, laundry, and home repairs. The response options given to each respondent were the same for the three questions and included "more than once a week," "once a week," "one to three times a month," or "hardly ever or never." An additional option was given to a special web-based subsample of respondents and included "every day." A new 3-category rating scale was developed from the above response options and included "inactive (0)," "low/moderately active (1)," and "highly active (2)" (Table 1). Thus, the PAS included three items of vigorous (VPA), moderate (MPA), and light PA (LPA). The PAS can provide a measure of PA by summing across the three items for participants with complete data.

Table 1. Physical activity (PA) scale (PAS) development

Original item categories	New scale value	New ratings
Hardly ever or never	0	Inactive
One to three times a month	1	Low/moderately active
Once a week	1	Low/moderately active
More than once a week	2	Highly active
Every day	2	Highly active

Note. The proposed scoring for the PAS is a simple summing across the 3 items for participants with complete data. Each item's stem: How often do you take part in sports or activities that are [vigorous] or [moderate] or [mild] ….

Self-rated general health (GH)

A self-rated general health (GH) variable was used to establish construct validity for the new PAS scores. GH was assessed using a single question that asked participants to rate their health. There were five response options that included "excellent," "very good," "good," "fair," or "poor."

Physical measures

Five (5) physical measures were used to for additional construct validity evidence and included body mass index (BMI), waist circumference (WC), grip strength (GS), timed walk speed (TW), and full tandem balance test time (BT). BMI (in kg/m^2) was assessed using objectively measured height (in inches) and weight (in pounds) and dividing weight by the square of height and multiplying by 703. WC (in inches) was assessed using a tape measure at the level of the participant's navel. GS (in kilograms) was assessed using a Smedley spring-type hand dynamometer, with two measurements taken for each hand in alternating fashion. The maximum grip strength test value for the participant's dominant hand was used for GS. TW (in seconds) was assessed by measuring the time it took the participant to walk a 98.5-inch straight course twice at a normal pace. The average of two timed trials was used for TW. Finally, BT (in seconds) was assessed by having the participant stand with the heel of one foot in front of the other foot, heel touching toes, for 30 to 60 seconds. The maximum time, up to the 30-60 second limit, was used for BT.

Sociodemographic covariates

Six different sociodemographic variables were used to describe the sample as well as adjust for potential confounding and included age, sex, race/ethnicity, employment, education, and marital status. Age was used as a categorical variable with group ranges of 50 to 59, 60 to 69, 70 to 79, and 80+ years. Sex included male and female groups. Race/ethnicity was used as a categorical variable and included White, Black, other, and not obtained groups. Employment was used as a categorical variable, with participants considered currently working, not currently working, or retired. Education was a categorical variable, with participants considered as either not having a college degree, having a college degree, or having an advanced degree. Finally, marital status was categorical and placed participants into one of two groups of either married or not married.

Statistical analyses

The study population was described by computing weighted estimates with 95% confidence intervals (CIs) of the percentage of adults considered active at three different PA intensities across sociodemographic characteristics. The development and validation strategy for the PAS involved six steps. First, item analyses, including means, standard deviations (SDs), frequencies, and item-total correlations, were computed to descriptively examine the functioning of the new scale items. Second, a factor analysis, using the PAS item polychoric correlation matrix, was performed to ensure a unidimensional construct using the eigenvalue greater than 1.00 and minimum explained variance of 60% criteria [7,8]. Third, an internal consistency reliability analysis was performed by computing ordinal alpha and alphas with each item deleted [9]. Fourth, an item response theory (IRT) graded response model (GRM) was employed using PAS items. The GRM is appropriate for ordered polytomous response data, and for this analysis, used a logit function with marginal maximum likelihood

estimation [10,11]. Fifth, construct validity correlation coefficients were computed between the IRT person scores (theta, θ) and variables that have a theoretical association with a PA construct. Sixth, regression modeling, using the PAS sum scores as a predictor and general health (GH) as an outcome, was performed for further construct validity evidence. Two sets of multinomial logistic regression models were used for the modeling, one with PAS scores in categorical form (i.e., 0-1, 2-3, 4-5, and 6) and one in numeric form (i.e., 0 thru 6) [12]. SAS version 9.4 was used for all analyses [13].

Table 2. Percentages of adults considered active at various physical activity (PA) intensities across sociodemographic characteristics, 2022 Health and Retirement Study (HRS)

	Active in LPA			Active in MPA			Active in VPA		
Characteristic	%	LL	UL	%	LL	UL	%	LL	UL
Overall (N=12,145)	82.5	81.3	83.7	68.5	67.2	69.9	36.3	34.9	37.7
Sex									
Female	85.8	84.7	87.0	65.1	63.2	67.0	29.8	27.9	31.8
Male	78.7	76.8	80.6	72.5	70.8	74.2	43.8	42.0	45.5
Age group (yr)									
50 to 59	89.7	87.3	92.1	71.9	69.0	74.9	37.5	34.6	40.4
60 to 69	86.2	84.7	87.7	73.7	71.8	75.6	41.9	40.0	43.8
70 to 79	80.9	79.1	82.8	67.4	65.2	69.5	34.0	31.7	36.3
80+	66.4	64.0	68.9	50.8	48.4	53.2	22.0	20.1	23.8
p for trend		<.0001			<.0001			<.0001	
Race/Ethnicity									
White	84.2	82.9	85.5	69.1	67.7	70.6	37.2	35.5	38.9
Black	73.4	70.8	76.1	59.6	56.8	62.4	30.2	27.7	32.7
Other	79.2	75.8	82.7	73.1	69.5	76.7	36.4	32.8	40.1
Not obtained	82.8	71.6	94.0	70.0	53.7	86.3	28.6	11.3	45.9
Employment									
Currently works	89.6	88.0	91.2	76.0	73.8	78.3	42.4	40.3	44.5
Not currently working	67.8	64.1	71.5	49.8	45.7	54.0	19.5	16.7	22.4
Retired	80.8	79.5	82.1	67.2	65.6	68.8	35.9	34.1	37.7
Education									
Has no college degree	80.6	79.2	81.9	63.3	62.0	64.6	31.5	30.0	33.0
Has college degree	85.8	83.8	87.7	77.7	75.0	80.5	44.2	41.6	46.8
Has advanced degree	87.1	84.7	89.5	80.3	77.2	83.4	48.1	44.1	52.1
p for trend		<.0001			<.0001			<.0001	
Marital status									
Married	79.3	77.6	81.0	63.0	61.0	65.1	30.5	28.7	32.2
Not married	84.7	83.3	86.1	72.2	70.5	73.9	40.2	38.3	42.1

Note. LPA is light PA. MPA is moderate PA. VPA is vigorous PA. Participants were considered "active" if they reported engaging in PA at the respective intensity at least once a week.

3. Results

The percentage of older adults considered active in LPA, MPA, and VPA was 82.5%, 68.5%, and 36.3%, respectively (Table 1). PA was more prevalent in male, employed, and unmarried populations. Younger age groups were more active than their older counterparts (p<.0001 for trend). Prevalence of PA increased with advancing college education (p<.0001 for trend). Finally, older Black adults were the least active among the race/ethnicity groups.

The item analyses indicated advancing item difficulty for higher PA intensity with item means for LPA, MPA, and VPA of 1.42, 1.25, and 0.65, respectively (Table 3). Item-total correlations indicated adequate item associations, with all uncorrected and corrected correlations greater than 0.72 and 0.44, respectively. Finally, item categories for each of the three items saw acceptable endorsement rates of at least 12%. The PAS polychoric correlation matrix was used for the exploratory factor analysis and showed correlations between 0.42 and 0.64 (Table 4). The factor analysis retained a single PA factor with a first eigenvalue of 2.10, all factor loadings greater than 0.79, and 70.2% explained variance (Table 5). Additionally, ordinal alpha indicated acceptable internal consistency (α=0.79) with all alpha deleted values in acceptable range.

Results from the GRM analysis confirmed that all PAS items fit a unidimensional construct (ps<.0001) (Table 6). Item calibration showed category thresholds ranging from -1.89 to 1.07 and item discrimination parameters between 1.37 and 5.27. The item characteristic curves (ICCs) provide a visual representation of these results (Figure 1). Finally, the test information curve indicated adequate discrimination across the PAS trait scale, with the

majority of information covering a theta (θ) range of ± 2.0 (Figure 1). IRT factor scores converged with the PAS sum score (r=0.96) and correlated with age (r=-0.22), sex (r=0.09), education (r=0.22), GH (r=0.41), BMI (r=-0.23), WC (r=-0.24), GS (r=0.24), TW (r=0.38), and BT (r=0.27) (Table 7). Similar results were observed with the PAS sum score correlating with variables theoretically known to be associated with PA in older adult populations.

Table 3. New PA scale (PAS) item statistics

	Statistics		Item-total correlations		Category frequencies		
Item	Mean	SD	Uncorrected	Corrected	0	1	2
VPA	0.65	0.83	.770	.440	7,100	2,244	2,801
MPA	1.25	0.81	.832	.571	2,874	3,399	5,872
LPA	1.42	0.69	.721	.438	1,452	4,142	6,551

Note. N=12,145. PAS item scales range from 0 (inactive) to 2 (highly active). Category frequencies are unweighted.

Table 4. New PA scale (PAS) item polychoric correlation matrix

Item	VPA	MPA	LPA
VPA	1	.644	.417
MPA	.644	1	.588
LPA	.417	.588	1

Note. N=12,145.

Table 5. New PA scale (PAS) factor analysis and reliability using polychoric correlation matrix

	Factor Analysis		Reliability	
Item	Loadings	Communality	$\alpha_{Deleted}$	α
VPA	.822	.675	.741	.785
MPA	.899	.808	.589	
LPA	.789	.622	.783	
% explained variance		70.2		

Note. N=12,145. All α values are ordinal level using the polychoric correlation coefficients [9].

Table 6. Graded response model item parameter estimates for the new PA scale (PAS)

Item	Parameter	Estimate	*SE*	*p*
VPA	*b1*	0.308	0.017	<.0001
	b2	1.074	0.026	<.0001
	a	1.583	0.049	<.0001
MPA	*b1*	-0.749	0.016	<.0001
	b2	0.049	0.012	<.0001
	a	5.270	0.637	<.0001
LPA	*b1*	-1.886	0.042	<.0001
	b2	-0.146	0.018	<.0001
	a	1.372	0.039	<.0001

Note. N=12,145. Graded response model used a logit function and marginal maximum likelihood estimation [13].

Table 7. Validity coefficients for the IRT-derived PA scale (PAS) score

Item	PAS	Age	Sex	Educ	GH	BMI	WC	GS	TW	BT
PAS IRT score - Pearson	.961	-.221	.093	.223	.405	-.228	-.244	.241	.375	.268
PAS sum score - Pearson	1.000	-.240	.089	.227	.418	-.213	-.239	.253	.381	.275
PAS IRT score - Spearman	.968	-.188	.093	.232	.381	-.156	-.185	.258	.354	.254
PAS sum score - Spearman	1.000	-.203	.089	.233	.387	-.139	-.178	.263	.358	.259
N	12,145	12,145	12,145	12,020	12,136	4,304	4,451	4,276	2,592	4,042

Note. Larger IRT scores represent greater amounts of PA. All coefficients are significant (p<.0001). PAS is the PAS summed score ranging from 0 to 6 where larger values represent greater amounts of PA. Sex is coded 1 for males and 0 for females. Educ is years of education ranging from 0 to 17. GH is general health ranging from 1 to 5 where larger values represent better health. BMI is body mass index. WC is waist circumference. GS is grip strength. TW is timed walk test speed. BT is full tandem balance test time. Pearson coefficients are weighted. Spearman coefficients are not weighted.

Table 8. Multinomial logistic regression analyses for raw PAS sum scores and GH outcome

	PAS sum score groups								
	PAS: 0-1 vs. 6			PAS: 2-3 vs. 6			PAS: 4-5 vs. 6		
Outcome Variable	OR	LL	UL	OR	LL	UL	OR	LL	UL
GH - Unadjusted									
1 vs. 5	300.23	120.30	749.27	36.00	16.85	76.94	7.14	3.00	16.99
2 vs. 5	42.12	24.80	71.52	13.71	9.80	19.16	4.50	3.25	6.24
3 vs. 5	10.82	6.94	16.87	5.59	4.25	7.34	2.97	2.33	3.79
4 vs. 5	2.71	1.80	4.08	2.30	1.77	2.98	2.01	1.58	2.56
GH - Adjusted									
1 vs. 5	158.24	58.78	425.98	27.50	12.09	62.59	6.15	2.49	15.18
2 vs. 5	27.67	15.68	48.83	11.27	7.75	16.38	4.26	2.99	6.05
3 vs. 5	9.01	5.65	14.35	5.11	3.81	6.85	2.89	2.26	3.71
4 vs. 5	2.65	1.70	4.14	2.28	1.72	3.02	2.00	1.57	2.55

Note. N=12,136 (unadjusted). N=11,820 (adjusted). Larger PAS scores represent greater amounts of PA. GH is general health where 1=poor, 2=fair, 3=good, 4=very good, and 5=excellent. OR is odds ratio. LL is lower limit of the 95% OR confidence interval. UL is upper limit of the 95% OR confidence interval. Adjusted model is adjusted for age, sex, race, employment, education, and marital status.

Table 9. Multinomial logistic regression analyses for PAS sum score and GH outcome

	PAS sum score		
Outcome Variable	OR	LL	UL
GH - Unadjusted			
1 vs. 5	0.31	0.27	0.34
2 vs. 5	0.47	0.43	0.52
3 vs. 5	0.60	0.56	0.65
4 vs. 5	0.79	0.73	0.85
GH - Adjusted			
1 vs. 5	0.34	0.30	0.39
2 vs. 5	0.50	0.46	0.55
3 vs. 5	0.62	0.57	0.67
4 vs. 5	0.79	0.73	0.85

Note. N=12,136 (unadjusted). N=11,820 (adjusted). Larger PAS scores represent greater amounts of PA. GH is general health where 1=poor, 2=fair, 3=good, 4=very good, and 5=excellent. OR is odds ratio. LL is lower limit of the 95% OR confidence interval. UL is upper limit of the 95% OR confidence interval. Adjusted model is adjusted for age, sex, race, employment, education, and marital status.

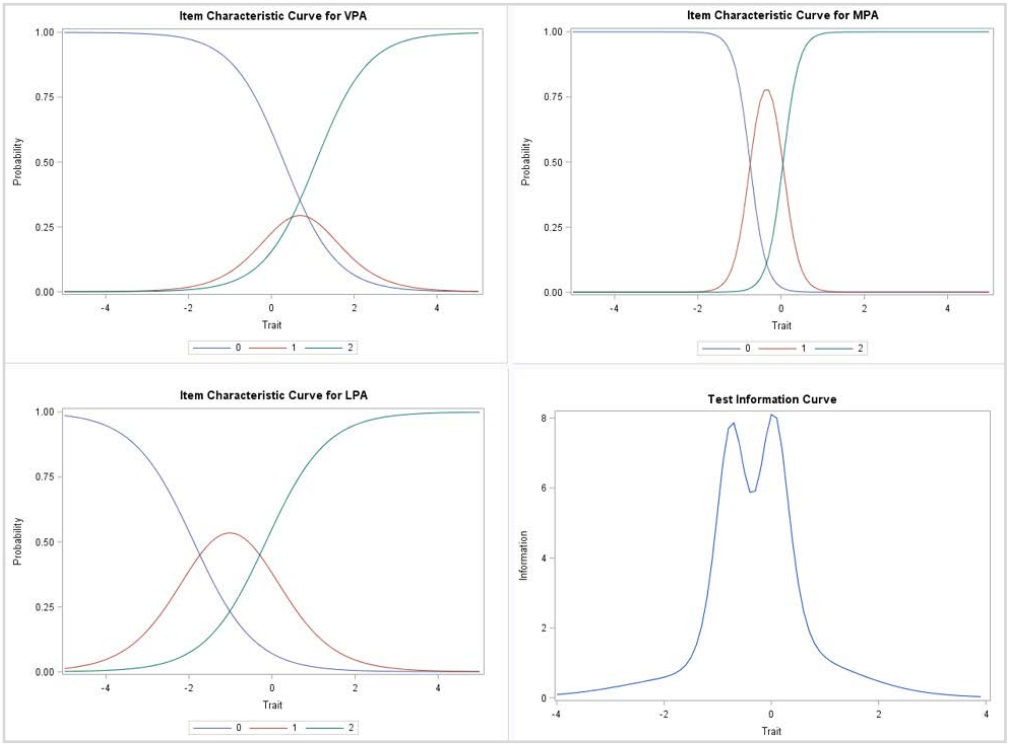

Figure 1. Item characteristic curves (ICCs) and test information curve from the graded response model (GRM) with PAS items

In the adjusted model using categorical PAS score groups, older adults in the lowest PAS score (0-1) group had significantly greater odds of reporting poor (OR=158.24; 95% CI: 58.78-425.98), fair (OR=27.67; 95% CI: 15.68-48.83), good (OR=9.01; 95% CI: 5.65-14.35), and very good (OR=2.65; 95% CI: 1.70-4.14) GH (vs. excellent) as compared to those in the highest PAS score (6) group (Table 9). Similar trends in general health odds were observed for the other PAS score groups in adjusted and unadjusted models. In the unadjusted model using numeric PAS scores, results showed that for each point increase in PAS score, odds of poor (OR=0.31, 0.27-0.34), fair (OR=0.47, 0.43-0.52), good (OR=0.60, 0.56-0.65), and very good (OR=0.79, 0.73-0.85) GH (vs. excellent) decreased. In the same adjusted model, for each point increase in PAS score, odds of poor (OR=0.34, 0.30-0.39), fair (OR=0.50, 0.46-0.55), good (OR=0.62, 0.57-0.67), and very good (OR=0.79, 0.73-0.85) GH (vs. excellent) decreased.

4. Discussion

This study developed a new multi-item scale to measure PA in older adults using items contained in the HRS. The PAS includes three items of VPA, MPA, and LPA, each with three category ratings of "inactive (0)," "low/moderately active (1)," and "highly active (2)." The item analyses indicated adequate functioning with acceptable item-total correlations and category frequencies. Additionally, factor analysis supported the PAS as a unidimensional construct, and the internal consistency reliability analysis indicated acceptable reliability. Finally, the IRT analysis further supports the PAS as a measure that discriminates well, targets a broad trait range, and measures a substantial amount of information. Therefore, the PAS developed here can be considered a valid scale that measures a unidimensional PA construct in older adults. This study also provided validity evidence for the PAS scores by firstly demonstrating correlations with IRT factor scores and other measures known to be associated with PA in older adults. Validity evidence was secondly provided by establishing the PAS sum score as a strong categorical and numerical predictor of GH in unadjusted and adjusted regression models. Thus, the simple sum score provided by the PAS can be considered a valid measure with the ability to discriminate among older adults with varying and incremental degrees of self-reported health status.

The PAS can be used in future studies in several ways using HRS data. In research that examines the prevalence and trends of PA among older adults, the PAS can offer an additional assessment option [14]. For example, scores from the PAS can be used to compute numeric descriptive statistics similar to the HRS scales of life satisfaction, happiness, loneliness, and stress [15]. Additionally, HRS research that examines the relationship between PA and other psychosocial variables can use the numeric capabilities of PAS scores for modeling purposes [16]. For example, instead of using categorical data analysis techniques to model traditional PA variables, the PAS can be used in general linear models, allowing for the examination of more in-depth research questions [17]. Lastly, research examining PA in any context using HRS PAS data can include reliability data that would otherwise be impossible. For example, researchers examining associations between PA and other health behaviors across different demographic populations could compute an internal consistency reliability coefficient (i.e., alpha) for the overall sample, for each subpopulation, and even across different waves of data [18,19,20].

This research has some limitations. One limitation is that the PAS developed in this study used only three simple questions regarding PA. Specifically, each of the three intensity questions (i.e., mild, moderate, and vigorous) was phrased to assess a frequency regarding sport or activities. Therefore, unlike other national surveys, it was not feasible to separate PA related to work, recreation, or transportation [21]. Another limitation is that this study validated the PAS using the entire sample of adults 50 years of age and older. Therefore, neither item nor test bias was examined using techniques such as IRT differential item functioning (DIF) or multi-group factor analysis measurement invariance [22]. Therefore, future research should be directed at examining the PAS for any potential measurement bias across subpopulations. In summary, the above limitations should be considered prior to using the PAS in future HRS research.

5. Conclusions

This study presents evidence for a new simple scale to measure PA in older adult populations. The PAS was shown to represent a unidimensional PA trait, with acceptable reliability, satisfactory construct validity, and valid scores showing high sensitivity. The ability to assess study-level internal consistency reliability when using the PAS should be considered a benefit. These results support the use of the PAS to measure PA in older adults.

References

[1] Keadle SK, McKinnon R, Graubard BI, Troiano RP. Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys. Prev Med. 2016; 89: 37-43.

[2] Liu XY, Yao K. Association between Domain-Specific Physical Activity and Novel Inflammatory Biomarkers Among US Adults: Insights From NHANES 2007-2018. Mediators Inflamm. 2025; 2025: 1989715. Published 2025 Jun 24.

[3] Ammous F, Peterson MD, Mitchell C, Faul JD. Physical Activity Is Associated With Decreased Epigenetic Aging: Findings From the Health and Retirement Study. J Cachexia Sarcopenia Muscle. 2025; 16(3): e13873.

[4] Hart P. Relationship between fitness performance and a newly developed continuous body composition score in U.S. adolescent boys. Int J Adolesc Med Health. 2020; 35(1): 69-79. Published 2020 Sep 23.

[5] Nunnally, J. C. (1978). Psychometric theory. McGraw-Hill Companies.

[6] Sonnega, A. (2025). Using Health and Retirement Study data: A guide for new users. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

[7] Tabachnick, B.G., Fidell, L.S. and Ullman, J.B., 2007. Using multivariate statistics. Boston, MA: Pearson.

[8] Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. Multivariate data analysis. Upper Saddle River, NJ: Prentice hall; 2012 Mar 23.

[9] Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research, and Evaluation. 2012; 17(1): 3.

[10] de Ayala RJ. The theory and practice of item response theory. Guilford Publications; 2013 Oct 15.

[11] SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide. The IRT Procedure. Cary, NC: SAS Institute Inc.

[12] Stokes ME, Davis CS, Koch GG. Categorical data analysis using SAS. SAS institute; 2012 Jul 31.

[13] SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide. The IRT Procedure. Cary, NC: SAS Institute Inc.

[14] Miller MJ, Cenzer I, Barnes DE, Covinsky KE. Physical inactivity in older adults with cognitive impairment without dementia: room for improvement. Aging Clin Exp Res. 2022; 34(4): 837-845.

[15] Lee J, Oh SM, Kim J, Kim J. Different Levels of Leisure Walking and Mental Health Among Older Adults With Mild Cognitive Impairment. J Aging Phys Act. 2023; 31(5): 841-848. Published 2023 Apr 20.

[16] Fiscella AJ, Andel R. The Association between Physical Activity, Obesity, and Cognition in Middle-Aged and Older Adults. J Aging Phys Act. 2024; 32(3): 397-407. Published 2024 Feb 9.

[17] Ruksakulpiwat S, Zhou W, Phianhasin L, et al. Associations between diagnosis with stroke, comorbidities, and activity of daily living among older adults in the United States. Chronic Dis Transl Med. 2023; 9(2): 164-176. Published 2023 Feb 21.

[18] Voss MW, Hung M, Li W, et al. Costs of Forced Retirement: Measuring the Effect of Lost Work Opportunity on Health. J Occup Environ Med. 2024; 66(8): e343-e348.

[19] Howrey BT, Hand CL. Measuring Social Participation in the Health and Retirement Study. Gerontologist. 2019; 9(5): e415-e423.

[20] Blanco LR, Hays RD. Evaluation of the Reliability and Validity of the Retirement Knowledge Scale (RKS). J Retire. 2024; 12(2): 76-94.

[21] Yang Y, Liu H. The association of different types of physical activity and diabetes co-morbid depression: A cross-sectional analysis. PLoS One. 2025; 20(9): e0332719. Published 2025 Sep 26.

[22] Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Front Public Health. 2018; 6: 149. Published 2018 Jun 11.