

**Table 1 – Table of data extraction and COSMIN methodological quality ranking for the articles found.**

<u>Reference</u>	<u>Outcome Measure</u> <sup>a</sup>	<u>Measurement Property Assessed</u>	<u>Study Design and Aim</u>	<u>Gender and Age Distribution</u> <sup>b</sup>	<u>Study Findings</u> <sup>c</sup>	<u>COSMIN Rank</u> <sup>d</sup> [37]
<u>1</u> Besson et al. (2010),[47]	<b>SRPAQ</b> - Recent Physical Activity Questionnaire (RPAQ)	Reliability	Test Re-test. 14 days apart	n= 131; Age range: 31-57 M 60 F 71	ICC=0.76 ( $p<0.0001$ ) Subscales ranged from 0.32 (transport) to 0.85 (work).	Box B: 'good'
	<b>Comparative</b> i) DLW + Resting Metabolic Rate (RMR) ii) accelerometer + Heart Rate (Acc+HR)	Criterion Validity	Comparison	n= 50; Age range: 21-55 M 25 F 25 Age 34.3+- 8.8 35.2+- 9.9 BMI - 25.1 +- 3.1 23.9 +- 3.8 kg/m <sup>2</sup> Body fat 22% +-7.9 32.1% +-7.6	-RPAQ significantly associated TEE ( $r=0.67$ $p<0.0001$ ) and PAEE ( $r=0.39$ $p=0.0004$ ) with DLW+RMR. -Estimated self-reported time spent exercising vigorously significantly associated with Acc + HR ( $r=0.70$ $p<0.0001$ ). -Estimated sedentary time was not ( $r=0.27$ $p=0.06$ ).	Box H: 'good'
<u>2</u> Boon et al. (2010),[46]	<b>SRPAQ</b> International Physical Activity Questionnaire -Usual week Long Form (IPAQ-LUS) <b>Comparative</b> Actigraph accelerometer	Criterion Validity	Comparison	n= 64; Age = 18-65 years old Mean(SD)= 35.2 (14.4) M:F = 28 : 36 Height (SD)= 1.73m (0.09) Weight (SD)= 72.1kg (13.6) BMI (SD)= 24.5kg/m <sup>2</sup> (4.1)	-The IPAQ-LUS showed moderate correlation with the Actigraph for combined moderate and vigorous activity ( $r=0.30$ ). -A strong correlation for vigorous activity ( $r=0.42$ ) and a small correlation with moderate activity ( $r=0.19$ ). -Overall the IPAQ-LUS tended to over-estimate mean PA by approximately 165%	Box H: 'good'

<sup>a</sup>DLW = Doubly Labelled Water; RMR = Resting Metabolic Rate; HR = Heart Rate;

<sup>b</sup> Gender = M = Male; F = Female; SD = Standard Deviation; BMI = Body Mass Index;

<sup>c</sup>ICC = IntraClass Coefficient; TEE = Total Energy Expenditure; PAEE = Physical Activity Energy Expenditure;  $r$  = Spearman's Correlation;  $p$  = Significance value; AEE = Activity Related Energy Expenditure; MVPA = Moderate to Vigorous Physical Activity

<sup>d</sup> COSMIN = Consensus-based Standards for selection of health Measurement INstruments;

3	Craig et al. (2003),[41]	<b>SRPAQ</b> - IPAQ 7-day Long Form (IPAQ-L7S)	Reliability	Test Re-test <10 days apart	<b>UK1</b> n= 149; M:F = 68:81; Age= 35.2yrs(SD10.6); Education= 16.3yrs (SD 3.6); Employed= 128 Hours worked= 35.6/wk (SD 13.4); Live in large city= 117; Good Health= 137.	-All English forms of the IPAQ demonstrated good to very good Spearman's reliability correlation coefficients ranging from 0.66 to 0.91 depending on the version and population used:	Box B: 'fair'
		IPAQ-LUS			Employed= 128 Hours worked= 35.6/wk (SD 13.4); Live in large city= 117; Good Health= 137.	<b>IPAQ-L7S:</b> UK1 $p=0.70$ ; USA1 $p=0.87$ ; USA2 $p=0.72$ ; USA2 $p=0.88$	
		IPAQ 7-day Short Form (IPAQ-S7S)			<b>USA1</b> n=28; M:F = 7:21; Age= 48.9 (SD6.1); Employed= 28; Height= 168.3cm (SD10.0); Weight= 78.8kg (SD22.7).	<b>IPAQ-LUS:</b> USA1 $p=0.91$ ; <b>IPAQ-S7S:</b> UK1 $p=0.69$ ; USA1 $p=0.81$ ; USA2 $p=0.88$ ; USA2 $p=0.66$ ;	
		IPAQ Usual Week Short Form (IPAQ-SUS)			<b>USA2</b> <b>V1-V2</b> and <b>V2-V3</b> n=30; M:F= 14:16; Age= 36.1 (SD12.8); Education= 17.4(SD 3) Employed= 27 Hours worked= 35/wk (SD12.7); Live in large city= 25; Good Health= 28; Height = 171.4cm (SD9.7); Weight= 77.3cm (SD17.3)	<b>IPAQ-SUS:</b> USA1 $p=0.84$ .	
		IPAQ-Short Form (SF) vs. Long Form (LF)	Internal Consistency	Comparison		The IPAQ-LF and IPAQ-SF showed reasonable correlation ( $p=0.67$ ; 95%CI 0.64-0.70)	Box A: 'poor'
		<b>Comparative</b> CSA accelerometer	Criterion Validity	Comparison		All IPAQ-LFs showed moderate ( $p=0.33$ ) and IPAQ-SFs fair correlation ( $p=0.30$ ) with the CSA accelerometer.	Box H: 'good'
						English forms of the IPAQ demonstrated Spearman's correlation coefficients from - 0.02 to 0.43 with the CSA accelerometer: <b>IPAQ-L7S:</b> UK1 $p=0.43$ ; USA1 $p=0.05$ ; USA2 $p=0.32$ .	
						<b>IPAQ-LUS:</b> USA1 $p=-0.02$ .	
						<b>IPAQ-S7S:</b> UK1 $p=0.40$ ; USA1 $p=0.26$ ; USA2 $p=0.27$ .	
						<b>IPAQ-SUS:</b> USA1 $p=0.13$ .	

4	Levy and Readdy (2009),[45]	<b>SRPAQ -</b> IPAQ-L7S	Reliability	Test Re-Test 4 days apart	n= 151 M:F= 52:99 Mean Age(SD)=24.2 (5.0) Marital status	-IPAQ-L7S scores for total PA was excellent ( $r=0.84$ ), good for vigorous activity ( $r=0.77$ ) and moderate activity ( $r=0.78$ ), and good for total walking activity ( $r=0.74$ ).	Box B: 'good'
		<b>Comparative</b> Physical Activation Measure (PAM)	Hypothesis Testing	Comparison	Partnered/Married 43%; Single 56.2%; Has job = 80.1%	PAM: weak correlation between PA motivation and total PA ( $p=0.17$ )	Box F: 'good'
		Perceptions of Competence			Ethnicity Caucasian 64.2%; Hispanic 13.2%; Asian-American 5.3%; Pacific Islanders 2.6%; African American 1.3%; Middle Eastern 1.4%; Other 10.6%.	Competence: weak correlation between PA competence and total PA ( $p=0.17$ )	
5	Maddison et al. (2007),[44]	<b>SRPAQ -</b> IPAQ-L7S	Reliability	Test-Retest: 0, 8 and 15 days	n= 36 M:F= 16:20; Age= 18-64yrs, Mean 39, SD 10; Height= 171.5cm (159-185), SD 7.4; Weight= 75.9kg (55.2-122), SD 14.8; BMI= 25.7 (19.9-37.2), SD 4; Fat Free Mass= 54.74 (40.66-85.54), SD 10.3;  Ethnicity: NZ Euro 27; NZ Mauri 4; Tongan 1; Asian 2; Other 5.	IPAQ-L7S demonstrated acceptable reliability at two different intervals with Spearman correlation coefficients of 0.74-0.79 ( $p<0.0001$ ).	Box B: 'fair'
		<b>Comparative</b> DLW	Criterion Validity	Comparison		-The IPAQ-L7S showed moderate correlation with DLW for AEE ( $r=0.31$ $p=0.06$ ) and metabolic energy equivalent per week ( $r=0.33$ $p<0.05$ ).	Box H: 'Fair'
						-Overall the IPAQ-L7S tended to underestimate mean PA by approximately 27%, with a more marked under-estimation at higher activity levels.	

<u>6</u>	Meriwether et al. (2006),[43]	<u>SRPAQ</u> - Physical Activity Assessment Tool (PAAT)	Reliability	Test Re-test, 7 days apart	n=67; M:F= 9:58; Age Mean = 37.5 (20-61); BMI = 25.7 (18.3-37.6)	-PAAT demonstrated significant test-retest reliability with Spearman correlation coefficients ( $r=0.618$ ; $p<0.001$ )	Box B: 'good'
		IPAQ-L7S	Reliability			IPAQ-L7S demonstrated significant reliability with Spearman correlation coefficients ( $r=0.627$ $p<0.001$ )	
		<u>Comparative</u> MTI accelerometer	Criterion Validity	Comparison		-PAAT correlated significantly with the Accelerometer for combined moderate and vigorous PA ( $r=0.392$ $p<0.01$ ) No data for the correlation of IPAQ-L7S with the MTI Accelerometer data was reported.	Box H: 'good'
<u>7</u>	Milton et al. (2013),[48]	<u>SRPAQ</u> – Single item measure [27] <u>Comparative</u> Actigraph Accelerometer	Criterion Validity	Comparison	n=66; M:F=23:43; Age Mean = 39 (11) Age Range = 21 – 62 years old	-The single item measure showed a significant correlation with accelerometer data. When all data assessed, $r=0.46$ . When only MVPA accumulated in bouts of $\geq 10$ mins was assessed $r=0.57$ .  -Participants underreported activity on the single item measure (-1.59 days) when compared with all objectively measured MVPA, but stronger congruence was observed when compared with MVPA accumulated in bouts of $\geq 10$ mins (0.38 days).	Box H: 'good'
<u>8</u>	Sirard et al (2013),[49]	<u>SRPAQ</u> - Godlin-Shephard (G-S) recall (1 week)	Hypothesis Testing	Comparison	n=121; M:F = 61:60; Age Mean = 24 (1.7) years; 50.8% White; 53.4% Overweight or obese evenly distributed within each subgroup	-SRPAQ derived PA measure demonstrated females less active than males, overweight individuals less active than non-overweight individuals and no differences detected for PA levels between race.  -Gender, Ethnicity or Weight did not significantly alter the correlation between the G-S recall and the Accelerometer data.	Box F: 'poor'

		<b>Comparative</b> Accelerometer	Criterion Validity	Comparison		-Reported correlations for moderate (r=0.3), moderate-vigorous (r=0.4) and vigorous (r=0.5) physical activity were reported as 'adequate'.	Box H: 'good'
<u>9</u>	Van der Ploeg et al. (2010),[42]	<b>SRPAQ</b> - IPAQ-S7S- Walking component only	Reliability	Test Re-test, 3 days apart	<b>S7S</b> n= 278; M:F= 124:154; Age (SD)= 36(11); Education (SD)= 16(4); Employed= 250 (90%).	-The IPAQ-S7S estimate of total time walked demonstrated excellent reliability over 3 days n=240 (spearman correlation r=0.77).	Box B: 'good'
		IPAQ-SUS Walking component only	Reliability	Test Re-test, 3 or seven days apart	<b>SUS</b> n= 763; M:F= 324:421; Age(SD)= 39(13); Education (SD)= 14(5); Employed= 430(76%).	-The IPAQ-SUS estimate of total walking time demonstrated very good (r=0.91; n=78) to excellent (r=0.72; n=684) reliability over 3 and 7 days, respectively	
		<b>Comparative</b> CSA accelerometer vs. IPAQ-S7S	Criterion Validity	Comparison	<b>Validity</b> <b>S7S</b> n= 446; M:F= 203:243; Age(SD)= 40(13); Education (SD)= 15(4); Employed= 376 (84%).	-Using individuals who reported moderate activity as walking only, the IPAQ-S7S estimate of total walking demonstrated a moderate correlation (r=0.39; n=137) with the accelerometer data of moderate PA.	Box H: 'good'
		CSA accelerometer vs. IPAQ-SUS			<b>SUS</b> n= 202; M:F= 93:109; Age (SD)= 37(12); Education (SD)= 16(3); Employed= 161(80%).	-Using individuals who reported no other moderate activity other than walking, the IPAQ-SUS estimate of total time walked demonstrated a poor correlation coefficient (r=0.26; n=84) with accelerometer data of moderate-intensity PA.	
<u>10</u>	Webster et al. (2011),[40]	<b>SRPAQ</b> a) IPAQ-L7S b) 6-point scale c) Human Activity Profile (HAP)	Hypothesis Testing	Comparison a)Age b)Gender c)Smoking status d)Marital status e)Co- morbidity f)Occupation	n=100; Age=18-80 years, Mean = 49.3; M:F= 22:78; Married= 61; Completed High School= 72; Smoker= 10; 1+ Co-morbidities= 25; Occupation: Sedentary= 49, Very heavy= 0.	-Age significantly different between the PA categories of all three SRPAQs (p<0.05).  -Gender and smoking status did not vary between the PA categories in any of the three SRPAQs (p>0.05).  -Marital status, co-morbidity and occupation did not have a significant effect on the PA categories of the IPAQ-LF (p>0.05).	Box F: 'fair'

-Marital status varied significantly between the PA categories on the HAP ( $p < 0.05$ ) and the 6-point scale ( $p < 0.01$ ).

-The presence of one or more co-morbidities had a significant effect on the PA category for the HAP ( $p < 0.000$ ) and the 6-point scale ( $p < 0.000$ ).

-Having a physical occupation had a significant effect on the PA categories of the HAP ( $p < 0.000$ ).

---

a) IPAQ-L7S vs. HAP	Structural Validity	Comparison
b) IPAQ-L7S vs. 6 point scale		
c) 6 point scale vs. HAP		

---

a) IPAQ-L7S vs. HAP: poor agreement, kappa value of  $k=0.38$

Box E:  
'poor'

b) IPAQ-L7S vs. 6 point scale: fair to good agreement, kappa=0.46

c) 6 point scale vs. HAP: fair to good agreement, kappa value of  $k=0.57$

**Table 2** – Summary of data extraction and methodological quality for each SRPAQ.

SRPAQ	Measurement Property	Article	Methodological COSMIN Rank	Results	Conclusions
<b><u>RPAQ</u></b>	Test-Re-test Reliability	Besson et al. (2010),[47]	Good	ICC = 0.76; $p < 0.0001$	-High test re-test reliability when measuring PAEE which can be accepted due to good methodological quality.
	Criterion Validity against DLW and RMR		Good	PAEE – $r = 0.39$ ; $p = 0.0004$ Total Energy Expenditure – $r = 0.67$ ; $p < 0.0001$ .	-Significant associations with total PA measured by RPAQ and DLW plus RMR.
<b><u>PAAT</u></b>	Test-Re-test Reliability	Meriwether et al. (2006),[43]	Good	$r = 0.618$ ; $p < 0.001$ .	-Significant reliability for the PAAT measurement of total PA.
	Criterion Validity against MTI accelerometer		Good	For combined vigorous and moderate PA - $r = 0.392$ ; $p < 0.01$	-Low but significant values linking the PAAT with accelerometer. PAAT underestimates 15.6% of active participants compared to the accelerometer and misclassified 68.4% as under-active but this was not significant ( $p = 0.169$ ).
<b><u>6-Point Scale</u></b>	Structural Validity against HAP and IPAQ-L7S	Webster et al. (2011),[40]	Poor	With HAP – $k = 0.57$ With IPAQ-L7S – $k = 0.46$	-Moderate agreement of both the HAP and the IPAQ-L7S and the 6-point scale.
	Hypothesis Testing		Fair	<u>Age</u> – $p < 0.05$ ; <u>Presence of co-morbidities</u> – $p < 0.000$ ; <u>Marital Status</u> – $p < 0.01$ ; <u>Gender</u> – $p > 0.05$ ; <u>Smoking</u> – $p > 0.05$ .	-Age, marital status and the presence of co-morbidities had a significant effect on PA measured. Gender and smoking did not.
<b><u>HAP</u></b>	Structural Validity against IPAQ-L7S	Webster et al. (2011),[40]	Poor	With IPAQ-L7S – $k = 0.38$	-Fair level of agreement of the HAP and the IPAQ-L7S.
	Hypothesis Testing		Fair	<u>Age</u> – $p < 0.05$ ; <u>Presence of co-morbidities</u> – $p < 0.000$ ; <u>Marital Status</u> – $p < 0.05$ ; <u>Physical Occupation</u> – $p < 0.000$ ; <u>Gender</u> – $p > 0.05$ ; <u>Smoking</u> – $p > 0.05$ .	-Age, marital status, presence of co-morbidities and occupation significantly affected PA therefore the HAP can distinguish PA between these groups.

Abbreviations: ICC = Intra-Class Coefficient;  $r$  = Spearman Correlation Coefficient;  $p$  = Levels of significance;  $k$  = Cohen's kappa

<b><u>Single Item Measure</u></b>	Criterion Validity against Actigraph Accelerometer	Milton et al. (2013),[48]	Good	All MVPA data assessed $r=0.46$ ( $p<0.001$ ); Overall agreement of MVPA between measures 58% with $k$ statistic 0.23 (95% CI 0.05 to 0.41). In objectively measured PA in $\geq 10$ min bouts $r=0.57$ ( $p<0.001$ ); Overall agreement 76% with $k$ statistic 0.39 (95% CI 0.14 to 0.64).	-Using all MVPA data the single item measure correctly identified 88% of participants meeting PA recommendations (specificity) and 48% of those not (sensitivity). Using PA data in $\geq 10$ min sessions specificity was 56% and sensitivity was 83%. Fair agreement shown with kappa statistic.
<b><u>Godlin-Shephard Recall (1 week)</u></b>	Criterion Validity against Accelerometer	Sirard et al. (2013),[49]	Good	For moderate PA $r=0.3$ ; For moderate-vigorous PA $r=0.4$ ; For vigorous PA $r=0.5$ ; Combined PA $r=0.43$ .	-Reported correlations between stratified PA levels described as 'moderate in magnitude'. Scatterplots demonstrated that the SRPAQ did not systematically over- or under-estimate MVPA compared to the Accelerometer.
	Hypothesis Testing		Poor	<u>Gender</u> – Male 3.5(3.65): Female 1.6(1.94) hours/week MVPA; <u>Weight</u> – Overweight and obese 2.2(3.00): Non-overweight 3.0(3.10) hours/week MVPA; - Overweight and obese 1.4: Non-overweight 2.2 (2.01) hours/week moderate PA	-SRPAQ demonstrated females completed less MVPA than males and overweight individuals reported less moderate PA and MVPA than non-overweight individuals. No significant difference between the correlation of SRPAQ and Accelerometer data for Gender, Ethnicity or Weight (all $p>0.05$ ).
<b><u>IPAQ-SUS</u></b>	Test-Re-test Reliability	Craig et al. (2003),[41]	Fair	For USA1 $r = 0.84$ .	-Excellent reliability with repeated testing over three and seven days for both the walking component and entire IPAQ-SUS.
		Van der Ploeg et al. (2010),[42]	Good	For walking component: Over three days – $r = 0.91$ ; Over seven days - $r = 0.72$ .	
	Criterion Validity against accelerometer	Craig et al. (2003),[41]	Good	For USA1 $r = 0.13$ .	-Poor correlation was found between the walking component of, and the total, IPAQ-SUS and the Accelerometer.
		Van der Ploeg et al. (2010),[42]	Good	For walking component $r = 0.26$ .	



<b><u>IPAQ-S7S</u></b>	Test-Re-test Reliability	Craig et al. (2003),[41]	Fair	UK1 $r = 0.69$ ; USA1 $r = 0.81$ ; USA2 (V1-2) $r = 0.88$ ; USA2 (V2-V3) $r = 0.66$ .	-Excellent test re-test reliability demonstrated over three days for total and 'walking only' IPAQ-S7S.
		Van der Ploeg et al. (2010),[42]	Good	For walking component $r = 0.77$ .	
	Criterion Validity against accelerometer	Craig et al. (2003),[41]	Good	UK1 $r = 0.4$ ; USA1 $r = 0.26$ ; USA2 $r = 0.27$ .	-Moderate correlation between the walking component and the Accelerometer. There was a small to moderate correlation for the full IPAQ-S7S.
		Van der Ploeg et al. (2010),[42]	Good	For walking component $r = 0.39$ .	
<b><u>IPAQ-LUS</u></b>	Test-Re-test Reliability	Craig et al. (2003),[41]	Fair	USA1 $r = 0.91$ .	-Excellent reliability demonstrated albeit with fair methodological quality.
		Boon et al. (2010),[46]	Good	For combined moderate and vigorous PA $r = 0.3$ . Overall IPAQ-LUS over-estimated by approximately 165%	
		Craig et al. (2003),[41]	Good	USA1 $r = -0.02$ .	
<b><u>IPAQ-L7S</u></b>	Test-Re-test Reliability	Meriwether et al. (2006),[43]	Good	$r = 0.627$ ; $p < 0.001$ .	-Very good to excellent reliability found for the IPAQ-L7S on repeated testing.
		Maddison et al. (2007),[44]	Fair	Between 0 and 8 days $r = 0.79$ ; $p < 0.0001$ . Between 8 and 15 days $r = 0.74$ ; $p < 0.0001$ .	
		Craig et al. (2003),[41]	Fair	UK1 $r = 0.7$ ; USA1 $r = 0.87$ ; USA2 (V1-V2) $r = 0.72$ ; USA2 (V1-V2) $r = 0.88$ .	
		Levy and Readdy (2009),[45]	Good	Total PA ICC = 0.84.	

Criterion Validity against accelerometer and DLW	Meriwether et al. (2006),[43]	Good	IPAQ-L7S significantly classified participants as “active” compared to the accelerometer ( $p<0.001$ ).	-Poor to moderate validity found for the IPAQ-L7S against an accelerometer.
	Craig et al. (2003),[41]	Good	UK1 $r = 0.43$ ; USA1 $r = 0.05$ ; USA2 $r = 0.32$ against accelerometer.	-The IPAQ-L7S marked 80% of participants as “active” corresponding to 71.4% by the accelerometer showing a significant level of agreement.
	Maddison et al. (2007),[44]	Fair	PAEE $r = 0.31$ ; $p=0.06$ ; Metabolic equivalents $r = 0.33$ ; $p<0.05$ . Tends to under-estimate mean PA by 27% compared to DLW.	-Moderate correlation for PAEE and significant values for metabolic equivalents against DLW.  -However, comparative to DLW the IPAQ-L7S tended to under-estimate mean PA by 27%.
Hypothesis Testing	Webster et al. (2011),[40]	Fair	Age – $p<0.05$ ; Presence of co-morbidities – $p>0.05$ ; Marital Status – $p>0.05$ ; Gender – $p>0.05$ ; Smoking – $p>0.05$ .	-Age significantly affected PA categorisation by IPAQ-L7S. The other sub-categories did not.