Validity and reliability of three self-report instruments for assessing attainment of physical activity guidelines in university students.

Abstract

Purpose: The purpose of this study was to compare the validity and reliability of three short physical activity self-report instruments to determine their potential for use with university student populations.

Methods: Participants (N = 155; 44.5% male; 22.9 ± 5.13 years) wore an accelerometer for nine consecutive days and completed a single item measure (SIM), the PACE+ and the IPAQ-SF questionnaires on day 1 and 9.

Results: Correlations between self-reported and accelerometer derived moderate-to-vigorous physical activity levels were moderate for the IPAQ-SF, while poor for the SIM and the PACE+. The agreement level was high with the IPAQ-SF (77.4%) and moderate for both the SIM (45.2 %) and PACE+ (44.5 %). The Intraclass Correlations between the two administrations were moderate to strong across all measures (0.52 – 0.70) in 133 participants.

Conclusions: The IPAQ-SF is the most suitable of these three self-report instruments for use with this population due to higher correlations and levels of agreement with accelerometry.

Keywords: Measurement, Validity, Reliability, Third level students.

Introduction

University or tertiary level students comprise a large portion of the population and may wield a sizable degree of future influence in society through their post-graduation roles (Hussain, Guppy, Robertson, & Temple, 2013). Globally tertiary education enrolments reached 170 million in 2009, and have been forecast to grow by an additional 21 million by 2020 (British Council, 2012). This makes the tertiary level sector an important setting for specific population monitoring, surveillance and intervention.

The transition from school to university brings greater independence in lifestyle choices, allowing students to become involved in more healthy or unhealthy behaviours (Dinger, Brittain, & Hutchinson, 2014). University students spend a considerable amount of time in educational environments which promote sedentary behaviour and in addition are largely being educated for sedentary occupations (Fotheringham, Wonnacott, & Owen, 2000), which may contribute to shaping persistent and potentially long-term physical inactivity patterns (Lesliephillip, Owen, Salmon, Sallis, & Lo, 1999; Owen, Lesliephillip, Salmon, & Fotheringham, 2000; Sallis, Bauman, & Pratt, 1998). In Ireland, the physical activity guidelines (PAGL) state that adults should engage in at least 150 minutes of moderate-intensity aerobic physical activity or 75 minutes of vigorous-intensity aerobic physical activity each week (Department of Health, 2009). Meeting these PAGL is associated with positive physical and mental health benefits (Reiner, Niermann, Jekauc, & Woll, 2013), while a high level of inactivity is a recognised risk factor for cardiovascular disease, diabetes, and some forms of cancer (Hallal, Andersen, Bull, Guthold, & Haskell, 2012). The regular monitoring and surveillance of population physical activity (PA) is of paramount importance (Hallal et al., 2012), but the challenges are with establishing a universal measurement tool, one that is psychometrically valid and specifically applies to this young adult population.

The measurement of PA can be challenging due to its varied nature (Janz, 2006), with a range of measurement tools available. Subjective measures include questionnaires, surveys and diaries, whereas objective methods include doubly-labelled water and motion sensors such as accelerometers

(Strath et al., 2013). Selecting the most appropriate measurement tool depends on a range of factors including the population of interest, the purpose of the study, the required outcome variables (Chinapaw, Mokkink, van Poppel, van Mechelen, & Terwee, 2010; Ridgers, Timperio, Crawford, & Salmon, 2012), and of prime importance the instrument's validity and reliability (Warren et al., 2010). Self-report questionnaires, due to their feasibility and convenience, are the most commonly used method of assessing populations PA levels (Helmerhorst, Brage, Warren, Besson, & Ekelund, 2012), with a diversity of questionnaires available for this purpose (Dyrstad, Hansen, Holme, & Anderssen, 2014). However, the use of different measures for assessing PA often results in findings which are inconsistent and incomparable across studies. For example, the reported prevalence of physical inactivity in undergraduate students has ranged from 22-81 % in 23 countries (Pengpid et al., 2015) to between 23-39% for an earlier study of 23 countries (Haase, Steptoe, Sallis, & Wardle, 2004). Although these studies looked at different samples, they both assessed PA using two different self-report methods. The use of one valid and reliable measurement tool, which is simple and effective for assessing PA at a population level (Ridgers et al., 2012), would allow comparability of findings.

Three questionnaires frequently used to assess populations levels of PA are the single item measure (SIM) (Milton, Bull, & Bauman, 2011), the PACE two item measure (Hardie Murphy, Rowe, Belton, & Woods, 2015; Prochaska, Sallis & Long, 2001), and the International Physical Activity Questionnaire- Short Form (IPAQ-SF) (Craig et al., 2003). Validity and reliability has only been established for the IPAQ-SF in this population (Dinger, Behrens, & Han, 2006) but each questionnaire has been validated against accelerometer derived moderate to vigorous physical activity (MVPA). The SIM demonstrated moderate validity (Cohen, 1988) ($r= 0.46$, $p<0.01$) in adults with the ActiGraph GT3X accelerometer (Milton, Clemes, & Bull, 2013). Hardie-Murphy and colleagues (2015) found the PACE had moderate validity ($r = 0.34 – 0.49$, $p<0.01$) with ActiGraph GT1M and GT3X accelerometers in children, however this measure has not yet been validated in adults. In university students, the IPAQ-SF demonstrated acceptable validity for accelerometer (ActiGraph Monitor Model 7164) derived MVPA with moderate ($r = 0.45$, $p<0.01$) and vigorous PA ($r = 0.20$, $p<0.05$) (Dinger et al., 2006). Research has reported the test-retest reliability of each measure in various populations

across different studies. The SIM demonstrated strong 2-5 day test-retest reliability (r= 0.72 - 0.82) using a Spearman's rank correlation coefficient in adults (Milton et al., 2011). Using Intraclass Correlation Coefficients (ICC), the PACE and IPAQ-SF reported strong test-retest reliability with the PACE reporting scores of 0.74 – 0.82 in children (Liu et al., 2010) and with the IPAQ-SF reporting scores of 0.71 – 0.89 in university students (Dinger et al., 2006).

There is a need to assess the validity and reliability of the SIM, the PACE and the IPAQ-SF for measuring adherence to PAGL across populations, such as the university population (Bobakova et al., 2015; Helmerhorst et al., 2012; Lee, Macfarlane, Lam, & Stewart, 2011). The purpose of this study was to assess the SIM, PACE and IPAQ-SF among a population of university students.

Methods

A convenience sample was recruited from 5 tertiary level institutions in Ireland (N = 463, 53% male, mean age = 22.2 ± 4.5). All participants were aged 18 years and provided written informed consent to take part in the study.

The three self-report measurement tools were presented to the participants in a questionnaire. Participants were provided with definitions of walking, moderate and vigorous PA and instructed to only include activities of this intensity when completing the questionnaire. The SIM asked participants to report the number of days they were physically active at a moderate to vigorous level for at least 30 minutes in the past 7 days (Milton et al., 2011). The PACE instrument was adapted from a 60 to a 30 minute timeframe to reflect the adult PAGL and renamed the PACE+ (Hardie Murphy et al., 2015). It used two items to assess PA. Item one of the PACE+ was a replica of the single item measure, while item two of the PACE+ asked the same question with respect to a usual week (Hardie Murphy et al., 2015). An average of the two items produced a score of days per week that the participants accumulated at least 30 minutes of MVPA. The IPAQ-SF included 9 items and required each participant to report the frequency and duration they were physically active at a walking, moderate and vigorous intensity. Total minutes MVPA was generated for the IPAQ-SF by accumulating each participants weekly moderate and vigorous PA. For the purpose of this study and to make each measurement comparable, minutes of PA at a moderate and vigorous intensity were combined and considered as minutes of MVPA. Compliance with the aerobic component of the PAGL was defined in two ways depending on the measurement tool used; 1) 30 minutes MVPA on 5 or more days a week (30 mins MVPA/day; SIM and PACE+) and 2) 150 minutes of MVPA over 7 days (150 mins MVPA/week; IPAQ-SF).

PA was also objectively measured using the ActiGraph (GT1M and GT3X) accelerometer. This monitor is an acceptable measure for evaluating questionnaire validity (Welk, 2005) and is widely used for this purpose (Craig et al., 2003; Dinger et al., 2014; Hardie Murphy et al., 2015; Milton et al., 2013). Participants were instructed to wear the device for nine consecutive days on their right hip

during all waking hours, except for when in water. The first and last days of wear time were excluded from analysis to give seven full wear days. The epoch length was set at ten seconds with data being downloaded and cleaned using the ActiLife software (Hardie Murphy et al., 2015). Consecutive zero counts of sixty minutes or more (Choi, Liu, Mattws, & Buchowski, 2011) were eliminated from total wear time and participants who did not meet the wear time criteria of at least 10 hours per day (Troiano et al., 2008) on seven days were excluded from the analysis. Accelerometer data were then analysed using the Troiano Adult cut-points (Troiano et al., 2008). A summary score of counts per minute (CPM) represented total PA. Participant responses were dichotomised into meeting or not meeting the PAGL for each measurement tool.

Researcher training across all institutional testing sites was conducted to ensure that standardized procedures were adopted and used. Participants completed a supervised self-report questionnaire which included demographic information (sex, age and year of study) and each of the three PA measures. An all days method (AD) (Ridgers et al., 2012) was used to determine compliance over 7 individual days, compared to accelerometry, to the PAGL with the SIM and the PACE+. A total minutes MVPA method (TM) was used to determine compliance over a total 7 days, compared to accelerometry, to the PAGL with the IPAQ-SF. A second questionnaire, containing each of the PA measure was given to the participants to complete nine days following the first. This allowed for the test-retest reliability to be assessed with each of the self-report measurement tools.

Statistical Analysis

Descriptive statistics were calculated for demographic, self-report and accelerometer data. For inclusion in the study, participants were required to have completed all the self-report measures and meet the accelerometer wear time criteria. The sample that met the inclusion criteria was compared to the full sample for sex and age. All statistical analyses were performed for the sample and stratified by sex, allowing any differences to be reported. Spearman Rho correlation coefficients were calculated between accelerometry (mins of MVPA/ day; CPM) and the SIM, PACE+ (mins of MVPA/day), IPAQ-SF (minutes of MVPA/ day). The strength of the Spearman Rho correlations were

ranked as poor (>0.1), moderate (>0.3), and strong (>0.5) (Cohen, 1988). Percentage agreement between each measure and accelerometer data was established by assessing the consistency of classification of achieving the PAGL. Sensitivity (defined as proportion of participants meeting PAGL that were correctly identified) and specificity (defined as the proportion of participants correctly identified as not meeting the PAGL) were determined using the accelerometry derived average MVPA/ day and the AD method for 7 valid days (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008) or by using the total MVPA/week and the TM method for 7 valid days. The percentage who self-reported meeting the PAGL and who met the guideline via accelerometer data is represented by the positive predictive value (PPV) and the percentage who self-reported not meeting the PAGL who did not meet them, as measured by accelerometer data, by the negative predictive value (NPV) (Parikh et al., 2008). Reliability analysis was available for all participants who completed the questionnaire on both occasions, nine days apart. An ICC, using a two way mixed average method, was recorded for each measure to determine its test-retest reliability, with scores being ranked as poor (0.0 – 0.2), fair (0.3 – 0.4), moderate (0.5 – 0.6), strong (0.7 – 0.8), and almost perfect (>0.8) (Landis & Koch, 1977).

Results

155 (44.5% male; 22.93 ± 5.13) students met the inclusion criteria and could be used in the analysis. Participants were excluded from the analysis if they were missing one of the self-report measurement tools (N= 48) or if they did not meet accelerometer wear time criteria (N = 260). The final sample were significantly older ($t(386) = 2.36$, $p<0.05$) and more likely to be female ($X^2(1, N= 434) = 6.41$, $p<0.05$) than those excluded. Participants included were undergraduate (88.8%) and postgraduate students spread across different years including 1st (30.5%), 2nd (38.1%), 3rd (11.4%), and 4th (20.0%).

Table 1 shows PA levels and compliance with PAGL for all measures used. Across all participants the proportion meeting the PAGL was 29.0% using the SIM and the 29.7 % using PACE+, but was higher with accelerometry using the AD method (68.4%). A higher proportion met the PAGL with the IPAQ-SF (76.8%) and accelerometry (94.8%) using the TM method. Males had significantly ($p<0.05$) higher values than females for self-reported PA using the SIM and IPAQ-SF, which are presented in Table 1.

Insert Table 1 about here

Correlation coefficients (**Table 2**) were poor to moderate ($r= 0.29 – 0.37$, $p<0.01$) between each self-report measurement of MVPA and accelerometer data in terms of minutes of MVPA per day and total PA in the whole sample. Correlations were significant ($r= 0.29 – 0.47$, $p<0.01$) for females and the total sample for each of the self-report measures with accelerometer derived MVPA and total PA. Significant scores were reported for males only between the IPAQ-SF and accelerometer derived MVPA ($r =0.31$, $p<0.05$) and total PA ($r= 0.27$, $p<0.05$).

Insert Table 2 about here

Details of agreement, sensitivity, specificity, PPV, and NPV between each of the self-report measures and accelerometer data are displayed in **Table 3**. There was a moderate level of agreement with both

the SIM (45.2%) and the PACE+ (44.5%) measures with accelerometer data using the AD method. IPAQ-SF demonstrated high levels of agreement with accelerometer data using the TM method (77.4%). Overall, the accuracy of classifying those achieving the guidelines (sensitivity) was poor with the SIM (31.1%) and the PACE+ (31.1%) but was high for the IPAQ-SF (78.2%). The percentage of participants who self-reported meeting the PAGL, who actually met (PPV) was high across all measures (71.7 - 96.6%). The accuracy of those not meeting the guidelines (specificity) was high with the SIM (75.5%) and the PACE+ (73.5%), while moderate for the IPAQ-SF (50.0 %). The percentage of participants who self-reported not meeting the guidelines who actually did not meet (NPV) the guidelines was poor for the SIM (33.6%), PACE+ (33.0%), and the IPAQ-SF (11.1%).

Table 4 shows the ICC scores for each of the self-report measures. These scores indicated moderate reliability with the SIM (0.67) and the IPAQ-SF (0.52) but stronger with the PACE+ (0.70) in 133 of the students (22 students were excluded from the analysis as they failed to complete the retest measure).

Insert Table 3 about here

Insert Table 4 about here

Discussion

Few studies have been conducted to examine the validity of PA questionnaires in university students using objective measures of PA such as accelerometers (Dinger et al., 2006). Additionally, few have explored the associations between self-report PA measurements and accelerometer measured MVPA using the recommended PAGL as the cut-points (Milton et al., 2013). The IPAQ-SF was the only measure found to have a significant association with accelerometer derived MVPA and total PA for males ($r = 0.27 – 0.31$, $p<0.05$) and females ($r = 0.29 – 0.33$, $p<0.01$). Similar results were reported in a publication by Craig and colleagues (2003), which found the validity of the IPAQ-SF in adults to be 0.30 (CI = 0.23 – 0.36) across 12 countries. A significant association between accelerometry and both the SIM and PACE+ was found in females only. Differences among sex have not been shown with regards to the validity of measures in university students, but have been reported in adolescents (Hardie Murphy et al., 2015; Rangul, Holmen, Kurtze, Cuypers, & Midthjell, 2008). Rangul and colleagues (2008) suggested that self-report instruments may become better measures if sex differences are taken into account.

The IPAQ-SF reported a strong level of agreement (77.4%) which was lower than previous findings (66.0%), but similar results for sensitivity (78.2% vs. 77.0%) and specificity (78.2% vs. 77.0%) (Ekelund et al., 2006). The SIM had a lower level of agreement (45.2%) and sensitivity (31.1%) with accelerometry, with higher levels of both being reported in a previous study (Milton et al., 2013). The PACE+ achieved similar results as the SIM, showing that it may be useful in adults but both of these measures achieved poor overall validity with this population, when compared to the results produced by the IPAQ-SF. This may be simply due to the fact that the IPAQ-SF contains more dimensions of PA (i.e. walking, moderate and vigorous) and also asks about the duration of PA on each day. The inability of the two shorter questionnaires to capture the same levels of information, as the IPAQ-SF, may lead to their poorer validity.

Test-retest reliability showed the PACE+ score a strong ICC (0.70), followed by the SIM (0.67) and finally the IPAQ-SF (0.52). Reliability scores reported in this study were lower than research suggests

for both the SIM (ICC = 0.86) (Milton et al., 2011) and the IPAQ-SF (ICC = 0.71 – 0.89) (Dinger et al., 2006). The number of days between the first and second administration of each questionnaire was longer in this study compared to previous research which may account for lower ICC scores for the SIM and the IPAQ-SF in adults. Reliability scores were still moderate (SIM and IPAQ-SF) to strong (PACE+) in this study suggesting that each of the measures has suitable reliability for use in this population.

Overall, objectively measured PA showed that a high proportion of this sub-population of students achieved the PAGL using the AD method (68.4%) and using the TM method (94.8%). A higher number of participants achieving the PAGL using the TM method is due to participants' accumulated minutes of MVPA reaching 150 minutes over a week but may not be spread over five or more days, which is needed to achieve the PAGL using the AD method. The IPAQ-SF reported a high proportion of students meeting the PAGL (76.8%), while the SIM and PACE+ reported much lower figures (29.0-29.7%). Research has found that students reported being very physically active when using the IPAQ-SF, engaging in 589 ± 405 minutes of total PA in the previous week (Dinger et al., 2006). Although the IPAQ-SF typically overestimates when compared with objective measures (Lee et al., 2011), it has underestimated in this study along with the other self-report measures. Other studies have reported underestimating in self-report measures when compared to accelerometry (Ekelund et al., 2006; Lim, Wyker, Bartley, & Eisenhower, 2015). Lim and colleagues (2015) reported that participants with higher accelerometer values were more likely to underestimate PA levels using the Global Physical Activity Questionnaire (GPAQ) in a sample of adults from New York City. This study suggested that underestimation may have been due to the built environment and widespread public transport in the participant setting, which led to more active body movement, thus potentially leading to people being more physically active than perceived (Lim et al., 2015). All of the students in the current study were in a university setting which could be considered as being built up, with widespread active and public transport opportunities when compared to rural areas of Ireland. Like Lim and colleagues' conclusion, this may have led to the current participants not considering their

active transport and occupational movements as being physically active, in turn causing the self-report measures to underestimate when compared to accelerometry.

This study had a number of limitations which should be noted. A convenience sample was used to recruit students across all faculties within each institution, however, a higher proportion of highly active students took part. Research has suggested that the difference between self-report and accelerometer measured MVPA may increase with higher activity and intensity levels (Dyrstad et al., 2014). Rowe and Mahar (2006) have also stated that the validity of such tools is an ongoing process and that when using a measure to validate against, it should be the most accurate measure of the construct, bringing into question activity monitors as a measure to validate against. As this study was being used as a precursor for future student surveys it is still important to use these findings to aid with the selection of self-report measures for use in future studies and interventions. Future studies should use representative samples, varying in PA levels in order to establish if these measures can be used across all university students. Another limitation is that the self–report measures were given to the participants before they wore the accelerometer meaning the same seven days were not being reported, which is also important as PA is not a stable behaviour itself. The measures selected give an indication of 'general or usual' physical activity levels, categorising population groups into meeting versus not meeting the physical activity guidelines. As such their sensitivity should allow for an objective measure to be administered over the same general time period.

The approach used for test-retest reliability may be questionable due to a behaviour such as PA not being stable from day to day, meaning that the measure may seem like it is not repeatable when in fact is measuring the correct PA levels. Using this approach can lead to measures having a low to moderate reliability, rather than acknowledging that the behaviour itself might have low reliability or stability (Kelly, Fitzsimons, & Baker, 2016).

Conclusion

This paper would recommend that when assessing levels of high active university students achieving the PAGL, the IPAQ-SF is the most suitable of these three self-report measures. This concurs with Dinger et al. (2014) who also found the IPAQ-SF to be a suitable PA measurement tool for university students. Another recommendation would be that other tools are available for PA measurement, especially for measuring the number of days university students are achieving the PAGL. Although validity for the SIM and the PACE+ were low, the overall results suggest that both tools may be useful for this population in the future. Finally, it is important to ensure that suitable measures are selected in future studies, depending on the population, aims and outcome measure of the studies.

References

Bobakova, D., Hamrik, Z., Badura, P., Sigmundova, D., Nalecz, H., & Kalman, M. (2015). Test-retest reliability of selected physical activity and sedentary behaviour HBSC items in the Czech Republic, Slovakia and Poland. *International Journal of Public Health*, *60*(1), 59–67. doi:10.1007/s00038-014-0628-9

British Council. (2012). *The shape of things to come: higher education global trends and emerging opportunities to 2020*. Retrieved from: https://www.britishcouncil.org/sites/default/files/the_shape_of_things_to_come_-_higher_education_global_trends_and_emerging_opportunities_to_2020.pdf

Chinapaw, M., Mokkink, L., van Poppel, M., van Mechelen, W., & Terwee, C. (2010). Physical Activity Questionnaires for Youth. *Sports Medicine*, *40*(7), 539–63. doi:10.2165/11530770-000000000-00000

Choi, L., Liu, Z., Mattws, C., & Buchowski, M. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and Science in Sports and Exercise*, *43*(2), 357. doi:10.1249/MSS.0b013e3181ed61a3

Cohen, J. (1988). The analysis of variance statistical power analysis for the behavioural sciences. 2nd ed. Erlbaum Associates: Hillsdale.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., … Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, *35*(8), 1381–95. doi:10.1249/01.MSS.0000078924.61453.FB

Department of Health (2009). National Guidelines on Physical Activity for Ireland. Retrieved from http://health.gov.ie/healthy-ireland/physical-activity/recommended-guidelines/

Dinger, M. K., Behrens, T. K., & Han, J. L. (2006). Validity and Reliability of the International Physical Activity Questionnaire in College Students. *American Journal of Health Education*, *37*(6), 337–343. doi:10.1080/19325037.2006.10598924

Dinger, M. K., Brittain, D. R., & Hutchinson, S. R. (2014). Associations between physical activity and health-related factors in a national sample of college students. *Journal of American College Health*, *62*(1), 67–74. doi:10.1080/07448481.2013.849710

Dyrstad, S. M., Hansen, B. H., Holme, I. M., & Anderssen, S. a. (2014). Comparison of self-reported versus accelerometer-measured physical activity. *Medicine and Science in Sports and Exercise*, *46*(1), 99–106. doi:10.1249/MSS.0b013e3182a0595f

Ekelund, U., Sepp, H., Brage, S., Becker, W., Jakes, R., Hennings, M., & Wareham, N. J. (2006). Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. *Public Health Nutrition*, *9*(2), 258–265. doi:10.1079/PHN2005840

Fotheringham, M., Wonnacott, R., & Owen, N. (2000). Computer use and physical inactivity in young adults: public health perils and potentials of new information technologies. *Annals of Behavioral Medicine*, *22*(4), 269–75. doi:10.1007/BF02895662

Haase, A., Steptoe, A., Sallis, J. F., & Wardle, J. (2004). Leisure-time physical activity in university students from 23 countries: associations with health beliefs, risk awareness, and national economic development. *Preventive Medicine*, *39*(1), 182–90. doi:10.1016/j.ypmed.2004.01.028

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., & Haskell, W. (2012). Physical activity levels of the world ' s population Surveillance progress , gaps and prospects, *6736*, 247–257. doi:10.1016/S0140-6736(12)60646-1

Hardie Murphy, M., Rowe, D. A, Belton, S., & Woods, C. B. (2015). Validity of a two-item physical activity questionnaire for assessing attainment of physical activity guidelines in youth. *BMC*

381 *Public Health*, *15*(1), 1080. doi:10.1186/s12889-015-2418-6

382 Helmerhorst, H. J. F., Brage, S., Warren, J., Besson, H., & Ekelund, U. (2012). A systematic review
383 of reliability and objective criterion-related validity of physical activity questionnaires. *The*
384 *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 103. doi:10.1186/1479-
385 5868-9-103

386 Hussain, R., Guppy, M., Robertson, S., & Temple, E. (2013). Physical and mental health perspectives
387 of first year undergraduate rural university students. *BMC Public Health*, *13*(1), 848.
388 doi:10.1186/1471-2458-13-848

389 Janz, K. (2006). Physical activity in epidemiology: moving from questionnaire to objective
390 measurement. *British Journal of Sports Medicine*, *40*(3), 191–2. doi: 10.1136/bjsm.2005.023036

391 Kelly, P., Fitzsimons, C., & Baker, G. (2016). Should we reframe how we think about physical
392 activity and sedentary behaviour measurement? Validity and reliability reconsidered.
393 *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 32.
394 doi:10.1186/s12966-016-0351-4

395 Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data.
396 *Biometrics*, *33*, 159–74.

397 Lee, P. H., Macfarlane, D. J., Lam, T. H., & Stewart, S. M. (2011). Validity of the International
398 Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. *The International*
399 *Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 115. doi:10.1186/1479-5868-8-115

400 Lesliephillip, E., Owen, N., Salmon, J., Sallis, J. F., & Lo, S. (1999). Insufficiently active Australian
401 college students: perceived personal, social, and environmental influences. *Preventive Medicine*,
402 *28*(1), 20–7. doi:10.1006/pmed.1998.0375

403 Lim, S., Wyker, B., Bartley, K., & Eisenhower, D. (2015). Measurement error of self-reported
404 physical activity levels in new York City: Assessment and correction. *American Journal of*
405 *Epidemiology*, *181*(9), 648–655. doi:10.1093/aje/kwu470

406 Liu, Y., Wang, M., Tynjälä, J., Lv, Y., Villberg, J., Zhang, Z., & Kannas, L. (2010). Test-retest
407 reliability of selected items of Health Behaviour in School-aged Children (HBSC) survey
408 questionnaire in Beijing, China. *BMC Medical Research Methodology*, *10*, 73.
409 doi:10.1186/1471-2288-10-73

410 Milton, K., Bull, F. C., & Bauman, a. (2011). Reliability and validity testing of a single-item physical
411 activity measure. *British Journal of Sports Medicine*, *45*(3), 203–8.
412 doi:10.1136/bjsm.2009.068395

413 Milton, K., Clemes, S., & Bull, F. (2013). Can a single question provide an accurate measure of
414 physical activity? *British Journal of Sports Medicine*, *47*(1), 44–8. doi:10.1136/bjsports-2011-
415 090899

416 Owen, N., Lesliephillip, E., Salmon, J., & Fotheringham, M. (2000). Environmental determinants of
417 physical activity and sedentary behavior. *Exercise and Sport Sciences Reviews*, *28*(4), 153–8.

418 Parikh, R., Mathai, A., Parikh, S., Sekhar, G., & Thomas, R. (2008). Understanding and using
419 sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, *56*(1), 45.
420 doi:10.4103/0301-4738.37595

421 Pengpid, S., Peltzer, K., Kassean, H. K., Tsala Tsala, J. P., Sychareun, V., & Müller-
422 Riemenschneider, F. (2015). Physical inactivity and associated factors among university
423 students in 23 low-, middle- and high-income countries. *International Journal of Public Health*.
424 doi:10.1007/s00038-015-0680-0

425 Prochaska, J. J., Sallis, J. F., & Long, B. (2001). A physical activity screening measure for use with
426 adolescents in primary care. Archives of Pediatrics & Adolescent Medicine, 155(5), 554-559.
427 doi:10.1001/archpedi.155.5.554

428 Rangul, V., Holmen, T. L., Kurtze, N., Cuypers, K., & Midthjell, K. (2008). Reliability and validity of

two frequently used self-administered physical activity questionnaires in adolescents. *BMC Medical Research Methodology, 8*, 47. doi:10.1186/1471-2288-8-47

Reiner, M., Niermann, C., Jekauc, D., & Woll, A. (2013). Long-term health benefits of physical activity--a systematic review of longitudinal studies. *BMC Public Health, 13*, 813. doi:10.1186/1471-2458-13-813

Ridgers, N., Timperio, A., Crawford, D., & Salmon, J. (2012). Validity of a brief self-report instrument for assessing compliance with physical activity guidelines amongst adolescents., *15*(2), 136–141. doi: 10.1016/j.jsams.2011.09.003

Rowe, D. A., & Mahar, M. T. (2006). Construct validity. In T. Wood & W. Zhu (Eds.), *Measurement theory and practice in kinesiology* (pp. 9–26). Human Kinetics.

Sallis, J. F., Bauman, A., & Pratt, M. (1998). Environmental and policy interventions to promote physical activity. *American Journal of Preventive Medicine, 15*(4), 379–97.

Strath, S., Kaminsky, L., Ainsworth, B., Ekelund, U., Freedson, P. S., Gary, R., Swartz, A. (2013). Guide to the assessment of physical activity: Clinical and research applications A scientific statement from the American heart association. *Circulation, 128*(20), 2259–79. doi:10.1161/01.cir.0000435708.67487.da

Troiano, R. P., Berrigani, D., Dodd, K. W., Masse, L. C., Tilert, T., & Mcdowell, M. (2008). Physical Activity in the United States Measured by Accelerometer. *Medicine & Science in Sports & Exercise, 40*(1), 181–188. doi:10.1249/mss.0b013e31815a51b3

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity - a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *European Journal of Cardiovascular Prevention and Rehabilitation, 17*(2), 127–39. doi:10.1097/HJR.0b013e32832ed875

Welk, G. J. (2005). Principles of Design and Analyses for the Calibration of Accelerometry-Based Activity Monitors. *Medicine & Science in Sports & Exercise*, *37*(Supplement), 501–511. doi:10.1249/01.mss.0000185660.38335.de