Boise State University
ScholarWorks

Kinesiology Faculty Publications and Presentations | Department of Kinesiology

5-1-2011

Identifying Group-Sensitive Physical Activities: A Differential Item Functioning Analysis of NHANES Data

Yong Gao
Boise State University

Weimo Zhu
University of Illinois at Urbana-Champaign

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Identifying Group Sensitive Physical Activities: A DIF Analysis of NHANES Data

Yong Gao
Boise State University

Weimo Zhu
University of Illinois at Urbana-Champaign

Abstract

Purpose: To identify subgroup sensitive physical activities using differential item functioning (DIF) analysis.
Methods: A sub-unweighted sample of 1,857 (males = 923 and females = 934) from the 2003-2004 National Health and Nutrition Examination Survey Physical Activity (PA) questionnaire data was used for the analyses. Using Mantel-Haenszel, SIBTEST and ANOVA DIF methods, 33 specific leisure-time moderate and/or vigorous PA (MVPA) items were analyzed for DIF across race/ethnicity, gender, education, income and age groups.
Results: Many leisure-time MVPA items were identified as large DIF items. When participating in the same amount of leisure-time MVPAs, non-Hispanic Blacks (NHB) were more likely to participate in basketball and dance activities than non-Hispanic Whites (NHW); NHW were more likely to participated in golf and hiking than NHB; Hispanics were more likely to participate in dance, hiking and soccer than NHW whereas NHW were more likely to engage in bicycling, golf, swimming, and walking than Hispanics; females were more likely to participate in aerobics, dance, stretching and walking than males whereas males were more likely to engage in basketball, fishing, golf, running, soccer, weightlifting and hunting than females; educated persons were more likely to participate in jogging and treadmill exercise than less educated persons; persons with higher incomes were more likely to engage in golf than those with lower incomes; and adults (20-59 yrs) were more likely to participate in basketball, dance, jogging, running, and weightlifting than older adults (60+ yrs), whereas older adults were more likely to participate in walking and golf than younger adults.
Conclusions: DIF methods are able to identify subgroup sensitive PAs, and thus provide useful information to help design group sensitive, targeted interventions for disadvantaged PA subgroups.

Keywords: differential item functioning, physical activity intervention, group sensitive components, physical activity questionnaire, Mantel-Haenszel method, SIBTEST

Introduction

Promoting physical activity (PA) among traditionally disadvantaged PA subgroups (e.g., racial/ethnic minorities, women, people with low incomes and/or low formal education levels, older adults, etc.) as a means to reduce PA disparities is very important to help eliminate health disparities. However, the success of PA interventions targeting disadvantaged subpopulations has been limited (30, 38). Lack of subpopulation sensitive components (e.g., group sensitive PAs) could be one of the factors affecting the effectiveness of PA interventions with these subpopulations (11, 38). Many PA interventions have used "popular" activities, such as walking, aerobic dance/activities, stretching, weight lifting, resistance exercise, stair climbing, and bicycling to promote PA or target behavior change outcomes (30). Since group sensitive PAs of targeted subpopulations are not taken into consideration, the selected "popular activities" might not be desirable to a specific subpopulation. For example, since non-Hispanic Black males may like playing basketball more than other types of activities, an intervention promoting walking only will likely be ineffective for this subpopulation.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Recently, the importance of including subpopulation sensitive components in health and PA interventions has gained recognition (21, 24, 37). For example, a study conducted by Anderson et al. (3) showed that the use of a culturally sensitive sign could significantly increase the use of stairs by African American commuters. Another study conducted by Van Duyn and colleagues (37) indicated that minority participants found culturally appropriate information and advice for physical activities appealing in PA interventions. However, efforts to identify subpopulation sensitive activities have been mainly based on the results of qualitative studies, e.g., to have focus-group participants name their favorite activities/ideas (3, 24, 37). Modern quantitative techniques have not been used to identify group sensitive components in the field of physical activity research.

Among modern quantitative techniques, differential item functioning (DIF) analysis has great potential. DIF refers to the different probabilities of getting an item correct/endorsed between two groups of respondents at the same ability. DIF analysis was originally developed to detect a potentially biased item/task (10). The presence of DIF indicates that an item or question may be biased to a subgroup, which could be a serious threat to the validity of a test or an instrument (4). Because a DIF item may measure an additional trait or culturally specific factor (36, 39), the presence of DIF could also imply that one group is more knowledgeable or comfortable with an item within a certain context. Therefore, persons in that group are more likely to endorse (or respond positively to) that item than those at the comparable ability levels in another group (13, 32). This second characteristic of DIF is actually desirable in identifying group sensitive activities among different subpopulations. In PA research, specific leisure-time moderate and/or vigorous physical activity (MVPA) questions (i.e., a list of MVPAs, e.g., basketball, dance, walking, etc.) are commonly included in PA questionnaires for PA assessment or used to facilitate name recognition of various activities (e.g., correct answering of questions) (1, 5, 17, 19, 29). If a MVPA question in a questionnaire is identified as a DIF item, it implies that activity is a subgroup sensitive activity.

Many statistical procedures have been developed for DIF detection. Mantel-Haenszel (MH), item response theory (IRT) based approaches, simultaneous item bias test (SIBTEST) and ANOVA methods are the most commonly used procedures (6, 14, 23, 26, 27, 31, 35). All DIF analysis approaches need to match two groups of respondents by the same ability levels first, then examine whether they differ from each other in the probabilities of getting an item correct/endorsed (4, 10). The same ability levels can be determined by either an internal measure derived from the test itself (e.g., total test score) or an external measure that measures the same construct, but is not directly related to the test. These internal and external measures often refer to the internal matching criterion and the external matching criterion, respectively (7). The compared two groups in DIF analysis are called the focal group (typically the group of interest) and the reference group (the group that the focal group is compared with; 25). In practice, several DIF methods are often used simultaneously to cross-examine the findings (4).

Although the concept of DIF has been introduced to Kinesiology (8, 40), the capacity of DIF methods to identify group sensitive activities has not been recognized and applied in the design of group-tailored PA interventions. The purpose of this study was to apply the MH, SIBTEST and ANOVA methods to identify subgroup sensitive PAs using 2003-2004 NHANES PA data.

Methods

Data Source

A sub-set of adult data, including respondents' demographic information, PA questionnaire, PA individual activities and PA monitor (i.e., ActiGraph accelerometer) data from the 2003-2004 NHANES data were used for the study (18). The PA questionnaire data consisted of participants' responses to whether or not they participated in leisure-time moderate and/or vigorous physical activities and activities in other PA domains (e.g., transportation related or domestic PAs); and if yes, what were the type, frequency and duration of the specific activities participants performed in the past 30 days (19). The NHANES Actigraph data were collected to objectively assess PA in free-living conditions at the population level. Respondents were asked to wear the Actigraph device on an elastic waist belt over the right hip for 7 consecutive days during normal waking hours except when swimming and bathing (18, 33). These data are publicly available through the Centers of Disease Control and Prevention (CDC) website. The current study was approved and received exempt status from the institutional review board of Boise State University. Respondents (unweighted N =1857, males = 923 and females = 934) age 20 years and older, who

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

reported participating in one or more types of moderate or vigorous activities for at least 10 minutes in their leisure-time, were included in the data analyses. Thirty-three commonly reported leisure-time MVPAs listed in the NHANES PA questionnaire (19), e.g., "aerobics, baseball, basketball, dancing, golfing, walking, hiking, jogging, etc.", were analyzed for DIF. The "item" and "a specific MVPA" therefore are used interchangeably in the rest of the text.

Data Analysis

DIF methods. DIF analyses were conducted using the MH, SIBTEST and ANOVA methods. MH and SIBTEST were chosen because they are routine methods in educational and psychological measurement practice for DIF detection, and have been described as statistically powerful, practically inexpensive, and do not require very large sample sizes (4, 7, 22). ANOVA is usually appropriate when the dependent variable is a continuous variable (35). The MH procedure involves matching the focal group and the reference group on their total or subtotal test scores. Then, it must be determined if the odds of getting an item correct/endorsed at each test score level are the same for both groups, across all levels of the matching scores. SIBTEST detects DIF by identifying a secondary dimension in an item that is not part of what the test intends to measure (26, 27). SIBTEST splits test items into two subsets: the matching items and the studied items. Items for DIF investigation are called "studied items" (28) and the rest of the items in the test are referred to as the "matching items". The scores from the matching items are summed and used to place test takers into different performance levels. Within each performance level, test takers in the reference and focal groups are considered to have the equivalent ability of being measured. Then DIF can be detected by comparing differences in performance between the reference and focal groups on the studied item whereas the two groups of respondents are matched at the same matching scores (25). In cross-cultural studies, ANOVA is usually used for DIF detection with continuous data (35). A two-way (i.e., grouping variable and matching variable) ANOVA is conducted to examine if there is a group difference across the matching scores. Significant group difference across the matching scores indicates the presence of DIF.

The continuous PA data were recoded into categorical data when conducting MH and SIBTEST DIF analyses since these two methods are designed to analyze categorical data (6, 10). A MET-Minutes per day score was summed from the questionnaire for each respondent. Items that were below 60.7 were coded as "1", those between 60.7 and 180 were coded as "2", and those above 180 were coded as "3", where "1" represented less physically active and "3" indicated more physically active. A MET-Minutes per day score for an item was calculated as [duration of activity] X [frequency of activity during the last 30 days] X [MET value of activity], and then divided by 30 days. The suggested metabolic equivalent (MET) values (1, 2) for the reported leisure-time MVPAs were used for the calculation according to the NHANES 2006 analytic guidelines (20). The cutoff points used above were based on tertiles from MET-Minutes per day scores. Log transformation was used to normalize the skewed distributions of all related variables for the ANOVA method (15).

Matching criteria. Both internal and external matching criteria were used to match ability levels of the focal and reference groups for this study. The "ability" in this case refers to the amount of reported leisure-time MVPA participation. The internal matching criterion was the total leisure-time MET-Minutes score during the last 30 days for each person calculated from their reported specific leisure-time MVPAs. That is, a person's total leisure-time MET-Minutes score during the last 30 days was calculated as [duration of activity] X [frequency of activity during the last 30 days] X [MET value of activity] and was summed across all reported leisure-time activities during the last 30 days (9). The total leisure-time MET-Minutes scores were then divided into 10 PA levels according to the percentiles (i.e., P_{10}, P_{20}, P_{30}, P_{40}, P_{50}, P_{60}, P_{70}, P_{80} and P_{90}). Respondents at the same PA levels were considered to be those reporting the same amount of MVPA participation. Actigraph data were processed and analyzed as suggested by Triano et al. (33) and the National Cancer Institute (http://riskfactor.cancer.gov/tools/nhanes_pam/; 16). The average PA counts per day during the wear time collected by the ActiGraph were used for the external matching criterion. The average daily activity counts were calculated as follows: first summed the product of [average counts per minute] and [minutes/hours the device was worn per valid day], and then divided the sum by [the number of valid days the device was worn]. The valid day means a day that a participant wore an ActiGraph unit for at least 10 hours (16). The average PA counts per day during wear time were divided into 10 PA categories according to the nine cut-off percentiles. Respondents at the same PA levels were determined according to their PA count information.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Reference and focal groups. DIF analyses were conducted across five major demographic characteristics/factors, including race/ethnicity, gender, education, income, and age. Table 1 summarizes these factors.

[Insert Table 1 Here]

Analysis procedures. DIF analyses were conducted for each of the five factors, i.e., race/ethnicity, gender, education, income, and age, separately. By each factor, 5 DIF analyses were conducted to determine if an item was a DIF item: 3 (MH, SIBTEST and ANOVA) with internal matching criterion and 2 (MH and ANOVA) with external matching criterion. For example, when it was desirable to examine whether an activity favored the female subgroup, DIF analyses were conducted with females as the focal group and males as the reference group. MH, SIBTEST and ANOVA analyses were employed when the ability levels of the two groups were matched using the internal matching criterion. The MH and ANOVA analyses were repeated when the ability levels were matched using the external criterion. No external matching criterion was used for SIBTEST because the SIBTEST software does not provide such an option. Similar analyses were conducted to determine if an item may be biased (e.g., favors a specific subgroup of interest: Non-Hispanic Whites vs. Non-Hispanic Blacks, Non-Hispanic Whites vs. Hispanics, higher income vs. low income, higher education vs. low education, and younger adults vs. older adults). To control potential inflation of α from multiple comparisons, significant level was set at 0.0005. SAS 9.0 and SIBTEST software DIFPACK were used for data analysis.

DIF item activity identification. An item was flagged as a DIF item when: (a) it was identified by two or more DIF methods regardless of using internal or external matching criterion, or (b) it was identified by a single DIF method, but confirmed with similar observations using both internal and external matching criteria. DIF direction (i.e., favoring or against a subgroup) and size (i.e., the magnitude of DIF) were also determined for each DIF item. For this study, a positive Beta value from SIBTEST indicated that DIF favored the reference group and a negative Beta value indicated that DIF favored the focal group. In the case that the direction of DIF for an item was inconsistent across the three methods, the final direction was determined by the two DIF methods that showed the same direction. An absolute Beta value greater than 0.088 indicated the presence of large DIF (25). When SIBTEST was insignificant, DIF size would not be reported because MH for polytomous (e.g., Likert scale) items and ANOVA methods do not provide a DIF size index.

An identified DIF item was considered as a group sensitive MVPA for a specific subgroup. The term “preferred” or “favored” is used in the following sections to facilitate the understanding of a DIF item, which indicates one subpopulation was more likely to endorse the item/activity than another subpopulation at the same leisure-time MVPA participation levels. It does not mean that one subpopulation had the higher participation rate in a specific activity over other activities than another subpopulation.

Results

Table 2 presents the prevalence of specific leisure-time MVPAs for respondents during the past 30 days using data from the 2003-2004 NHANES PA questionnaires. Walking was the most popular leisure-time PA (58%) among the general population, followed by bicycling (19%), weightlifting (14%), treadmill exercise (14%), dance (13%), aerobics (11%), stretching (10%), running (9%), swimming (9%), golf (9%), hiking (9%), basketball (8%), jogging (7%), stair-climbing (7%), fishing (5%), and yoga (4%). The prevalence of specific leisure-time MVPAs varies across different demographic characteristics/factors. For example, non-Hispanic Whites, females, respondents with higher incomes, and older adults reported higher participation in walking (ranging from 60% to 67%) whereas Hispanics had the lowest participation (44%) in walking. Of the subpopulations, non-Hispanic Blacks had the highest proportion of respondents playing basketball, and Hispanics reported the highest participation rates in baseball, dance, hiking and soccer (see Table 2).

[Insert Table 2 Here]

A summary of statistics for DIF analyses and DIF evaluation is provided in Table 3. The results are briefly described below by race/ethnicity, gender, education, income, and age group.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

By race/ethnicity. Table 3 presents the results of DIF analyses between non-Hispanic Whites and non-Hispanic Blacks. Four items showed large DIF with the absolute Beta values larger than 0.088, $p < 0.0005$. Basketball and hiking were identified as DIF items by all three DIF methods, and were confirmed with both internal and external matching criteria. Basketball was favored by non-Hispanic Blacks and hiking by non-Hispanic Whites. Dance and golf were identified as DIF items by at least two of the three DIF methods and were confirmed with both internal and external matching criteria. Golf was favored by non-Hispanic Whites and Dance by non-Hispanic Blacks (see Table 3).

For the comparison between non-Hispanic Whites and Hispanics, bicycling, dance, swimming and walking were identified as DIF items by all three DIF methods, and were confirmed with both internal and external matching criteria. Absolute Beta values were larger than 0.088, indicating the presence of large DIF (Table 3). More Hispanic respondents reported higher participation in dance compared with their non-Hispanic White counterparts at each comparable PA level, whereas non-Hispanic Whites were more likely to participate in bicycling, swimming, and walking. Although hiking was not detected as a DIF item by SIBTEST, MH and ANOVA methods using internal and external matching criteria confirmed that Hispanic respondents were more likely to participate in hiking than non-Hispanic Whites . Golf was detected as a DIF item by ANOVA and it was favored by Non-Hispanic Whites. This finding was confirmed with both internal and external matching criteria. Soccer was detected as a DIF item only by MH and ANOVA methods using the internal matching criterion. Soccer was favored by Hispanics (Table 3).

By gender. Thirteen DIF items were identified across gender by at least two of the three DIF methods, with 10 items showing large DIF (Table 3). Running was identified as a DIF item by only internal matching criterion from SIBTEST, MH and ANOVA. Soccer, treadmill exercise and weightlifting were identified as DIF items by all three DIF methods using internal matching criterion and MH using external matching criterion, whereas walking was confirmed as a DIF item by all three DIF methods using internal matching criterion and ANOVA using external matching criterion. All of the other seven items were confirmed as DIF items with both internal and external matching criteria from the three DIF methods. Aerobics, dance, stretching and walking were favored by females whereas basketball, fishing, golf, running, soccer, weightlifting, and hunting were favored by males. Although treadmill exercise and yoga were identified as DIF items, the direction of DIF was not consistent across three DIF methods: SIBTEST indicated these activities were favored by males but MH and ANOVA indicated they were favored by females. (Table 3).

By education. Jogging, treadmill exercise and yoga were identified as DIF items (Table 3). The presence of large DIF in jogging was confirmed by all three DIF methods with both internal and external matching criteria whereas only SIBTEST, MH and ANOVA with external matching criterion confirmed large DIF relative to treadmill exercise. Yoga was recognized as a DIF item by MH and ANOVA with both internal and external matching criteria, but not by SIBTEST. The direction of DIF for these items was inconsistent across three DIF methods: SIBTEST showed that these activities were favored by persons with low education, but MH and ANOVA indicated they were favored by persons with high education.

By income. Only two items, golf and treadmill exercise, were identified as DIF items based on income. Both items were favored by the high income group (Table 3). The presence of large DIF in golf was confirmed by all three DIF methods with both internal and external matching criteria whereas DIF for the treadmill exercise was confirmed by MH and ANOVA methods with both internal and external matching criteria (Table 3).

By age group. When items were examined by age, seven items with large DIF were identified. Basketball, dance, jogging, running, and walking were confirmed by all three DIF methods with both internal and external matching criteria. Golf was a DIF item confirmed by SIBTEST and MH with both internal and external matching criteria. Weightlifting was recognized as a DIF item only by the three methods with the internal matching criterion. Basketball, dance, jogging, running, and weightlifting were favored by adults aged 20 to 59 years whereas walking and golf were favored by older adults (60+ yrs) according to MH and ANOVA methods. SIBTEST again reported an opposite direction showing walking and golf were favored by adults aged 20 to 59 years old (Table 3).

[Insert Table 3 Here]

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Discussion

The results from DIF analyses in this study indicate that many leisure-time MVPA questions used in the 2003-2004 NHANES PA questionnaires are DIF items across each of the five studied factors (i.e., 4 to 7 based on race/ethnicity, 13 based on gender, 3 based on education, 2 based on income, and 7 based on age group). Most of these items present large DIF and they are favored by either a reference group or a focal group. For example, basketball and dance activities were favored by non-Hispanic Blacks whereas golf and hiking were favored by non-Hispanic Whites. Such DIF results were expected since it is well known that one subpopulation may prefer specific kinds of physical activities over other physical activities (e.g., aerobics by females and weightlifting by males). The implications of the DIF findings in this study can be viewed as both negative and positive.

On the negative side, any DIF items may lead to a misleading estimation of, and/or conclusion about, a subpopulation's PA participation. Because most individual activities are favored by a subgroup (typically the advantaged PA groups), questionnaires that consist entirely of specific PA questions will likely lead to a biased estimation of a subgroup's PA level and therefore, should be used cautiously for PA assessment. Moreover, to facilitate a respondent's answer to a general PA question, a questionnaire needs to be very careful about what example of a specific PA is used as an illustration. For example, in the 2003-2004 NHANES PA questionnaires, a question about leisure-time vigorous PA is asked in the following manner: "Over the past 30 days, did you do any vigorous activities for at least 10 minutes that caused heavy sweating, or large increases in breathing or heart rate? Some examples are running, lap swimming, aerobics classes or fast bicycling" (19). All the example activities (running, swimming, aerobics and bicycling) in this question were identified as DIF items in the current study favoring at least one of the traditionally advantaged PA groups such as non-Hispanic Whites. This means that even with the same amount of total leisure PA participation, persons from the other two race/ethnicity groups (non-Hispanic Blacks and Hispanics) may be less likely to answer "yes" to this question than non-Hispanic Whites because they may not consider the vigorous activities they participate in as being akin to the vigorous PA given in the question. As a result, vigorous PA participation for these subgroups can be underestimated. From this example we can see that when assessing PA, group sensitive examples (e.g., aerobics for non-Hispanic White females and basketball for non-Hispanic Black males) should be employed so that respondents will not be impacted by unfamiliar examples when responding to a general PA question. This practice could be difficult for hardcopy questionnaires. For a telephone-based survey, which is the method that many national surveillance systems currently use to collect their data, this practice should not be too difficult to implement. Studies are needed to determine the benefits of such practice.

On the positive side, the identification of DIF items among the specific leisure-time MVPAs in this study provides useful information to help design targeted PA interventions. Low regular PA participation is often associated with poor health outcomes (34). Many PA interventions are designed to improve PA participation among traditionally disadvantaged PA subgroups through promoting "popular" activities such as walking, aerobics, dance, weightlifting, and bicycling (30). However, without taking into consideration group sensitive PAs of a targeted subpopulation, the selected popular activities could be unappealing to that subpopulation. As a result, such PA interventions are less likely to succeed and may not be sustainable after an intervention. It is therefore reasonable for us to believe that incorporating group sensitive activities into a PA intervention will likely increase the probability of success. So far, efforts made to identify subpopulation sensitive components are mainly based on results from qualitative studies where focus groups are used to identify culturally sensitive components (3, 24, 38). Current understanding and identification of subpopulation sensitive PA components would increase if information provided by qualitative studies could be supplemented with similar information from quantitative studies.

As shown in this study, the DIF technique has the capacity to provide such information, and this study is the first effort to use DIF analysis methods and national data to provide subgroup sensitive components for PA interventions. The presence of DIF in this study implies that when participating in the same amount of total leisure-time PA, persons in a subgroup, likely due to a cultural influence, are more likely to participate in a different activity than persons in another subgroup. A DIF item/activity in this context implies a subgroup sensitive PA. The important next step is to design a subgroup targeted intervention to promote PA using that subgroup's sensitive activities and compare it with conventional PA interventions. One may wonder if a simple frequency analysis of these specific activities could lead to a similar finding. The major difference between DIF and the frequency analysis is that the

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

former examines group difference after controlling the overall ability, e.g., total amount of reported MVPA participation in this study. In other words, research questions are now extended from "What activity is a subgroup more likely to endorse?" to "By comparing people at the same PA levels, what activity is a subgroup more likely to endorse?" In practice, information from both analyses should be used. It is expected that, when promoting PA among a subgroup, using popular and subpopulation sensitive activities will be welcomed and the chance for success should be higher.

It should be noted that this study has a few limitations. First, this study used three DIF methods, among which, MH and SIBTEST methods were originally developed in education measurement to analyze categorical data. In order to use MH and SIBTEST, PA continuous data in this study were recoded into categorical data. This procedure might have caused some information loss and could affect the DIF results. Inconsistent findings among different methods were likely caused by the recoding. More simulations studies to better understand the impact should be conducted. Second, although it is generally appropriate to use the ANOVA approach for analyzing continuous variables, some researchers believe ANOVA cannot distinguish DIF and impact (i.e., the real group difference; 10), when there are only a few levels of ability matching. Third, the low correlation between the external matching variable (i.e., ActiGraph counts) and the internal criterion variable (12) could be a problem for DIF analysis. A further investigation is needed to examine if the ActiGraph and PA questionnaires measured the same construct.

Conclusions

Many items used in the 2003-2004 NHANES PA questionnaires were found to be DIF items. Therefore, PA estimations generated based on the questionnaires should be used with caution. Meanwhile, the information resulting from DIF analyses can be used in a very positive way, i.e., to identify group sensitive activities of traditionally disadvantaged social subgroups and design more subgroup targeted, and therefore more effective, PA interventions for these subgroups. DIF methods were demonstrated to be a useful approach for identifying subgroups' sensitive PAs, but consistency of the results from different methods needs to be further investigated.

Acknowledgements

This study was funded by the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) Research Consortium. The results from the present study do not constitute endorsement by the American College of Sports Medicine.

Conflict of Interest

The authors have no conflicts of interest to disclose.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

References

1. Ainsworth BE, Leon AS, Richardson MT, Jacobs DR, Paffenbarger RS. Accuracy of the college alumnus physical activity questionnaire. *J Clin Epidemiol.*1993; 46: 1403-11.
2. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activity: An update of activity codes and MET intensities. *Med Sci Sports Exerc.* 2000; 32(9 suppl): S498-16.
3. Anderson RE, Franckowiak SC, Zuzak KB, Cummings ES, Bartlett SJ, Crespo CL. Effects of a culturally sensitive sign on the use of stairs in African American Commuters. *Soz Praventivmed*. 2006; 51: 373-80.
4. Camilli G, Shepard LA. Methods for identifying biased test items. Thousand Oaks (CA): Sage; 1994. 181 p.
5. Centers for Disease Control and Prevention (CDC). An explanation of U.S. physical activity surveys. (2007). [cited 2007 July 23]. Available from http://www.cdc.gov/nccdphp/dnpa/physical/health_professionals/data/physical_surveys.htm#Surveys.
6. Chang, HH, Mazzeo, J, Roussos, L. Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *J Educ Meas, 1996; 33*(3): 333-353.
7. Clauser, BE, Mazor, KM. Using statistical procedures to identify differentially functioning test items. *Educational Measurement.* 1998; 17(1): 31-44.
8. Cohen AS. Item bias and differential item functioning. In: Wood TM, Zhu W, editors. *Measurement Issues and Practice in Physical Activity.* Champaign: Human Kinetics; 2006. p. 113-26.
9. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003; 35: 1381-95.
10. Dorans NJ, Holland PW. DIF detection and description: Mantel-Haenszel and standardization. In: Holland PW, Wainer H, editors. *Differential item functioning.* Hillsdale: Lawrence Erlbaum; 1993. p. 35-66.
11. Freimuth VS, Mettger W. Is there a hard-to-reach audience? *Public Health Rep.1990; 105*(3): 232-38.
12. Gao Y. Differential Item Functioning Analysis of Physical Activity Disparities in the U.S [dissertation]. Urbana-Champaign (IL): University of Illinois at Urbana-Champaign; 2008. 188 p.
13. Hambleton RK, Patsula L. Adapting tests for use in multiple languages and cultures. *Soc Indic Res. 1998; 45:* 153-71.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

14. Holland PW, Thayer DT. Differential item functioning and the Mantel-Haenszel procedures. In: Wainer H, Brain H, editors. *Test validity*. Hillsdale (NJ): Lawrence Erlbaum; 1988. P. 128-45.
15. Kirk RE. Experimental design, Procedures for the behavioral science (2nd ed.). Monterey (CA): Brooks/Cole; 1982. 911 p.
16. National Cancer Institute. Risk factor monitoring and methods: SAS programs for analyzing NHANES 2003-2004 accelerometer data. (2007). [cited 2008 July 29]. Available from: http://riskfactor.cancer.gov/tools/nhanes_pam.
17. National Center for Health Statistics (NCHS). National Health Interview Survey, *2006*. Hyattsville (MD): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; (2006). [cited 2007 July 29]. Available from ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2006/English.
18. National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey, *2003-2004* [Data file]. Hyattsville (MD): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2006. [cited 2007 July 29]. Available from http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/nhanes03_04.htm.
19. National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire, *2003-2004*. Hyattsville (MD): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2006. [cited 2007 July 29]. Available from http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/nhanes03_04.htm.
20. National Center for Health Statistics (NCHS). NHANES analytic guidelines. Hyattsville (MD): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2006. [cited 2007 Aug 10]. Available from http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf.
21. Pekmezi D, Jennings E. Physical activity interventions in minority populations. Curr Cardiovasc Risk Rep. 2009; 3: 275-80.
22. Penfield, RD, Camilli, G. Differential item functioning and item bias. In Rao CR, & Sinharay S, editors. Handbook of Statistics Psychometrics. Amsterdam: Elsevier; 2007. Vol. 26, p. 125-167.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

23. Raju NS. The area between two item characteristic curves. Psychometrika. 1988; 53(4): 495-502.
24. Resnicow K, Jackson A, Braithwaite R, Dilorio C, Blisset D, Rahotep S, Periasamy S. Healthy body/healthy spirit: A church-based nutrition and physical activity intervention. Health Educ Res. 2002; 17(5):562-73.
25. Roussos L, Stout W. A multimensionality-Based DIF analysis paradigm. *Appl Psych Meas. 1996; 20*(4): 355-71.
26. Shealy RT, Stout WF. An item response theory model for test bias and differential test functioning. In: Holland P, Wainer H, editors. *Differential item functioning.* Hillsdale: Erlbaum; 1993a. p. 197-240.
27. Shealy R, Stout W. A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika. 1993b; 58*: 159-94.
28. Stout W, Li HH, Nandakumar R, Bolt D. MULTISIB: A procedure to investigate DIF when a test is intentionally two-dimensional. *Appl Psych Meas.1997; 21*(3): 195-213.
29. Taylor HL, Jacobs DR, Shucker B, Knudsen J, Leon AS, DeBacker G. A questionnaire for the assessment of leisure time physical activities. J Chron Dis. 1978; 31: 741-55.
30. Taylor WC, Baranowski T, Young DB. Physical activity interventions in low-income, ethnic minority, and populations with disability. *Am J of Prev Med. 1998; 15*(4): 334-43.
31. Teresi JA, Golden RR, Cross P, Gurland B, Kleinman M, Wilder D. Item bias in cognitive screen measures: Comparisons of elderly White, Afro-American, Hispanic and low and education subgroups. *J Clin Epidemiol.* 1995; 48(4): 473-83.
32. Teresi JA. *Differential item functioning and health assessment*. [Internet]. 2004. [cited 2007 June 18]. Availble from: http://outcomes.cancer.gov/conference/irt/teresi.pdf.
33. Triano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, Mcdowell M. Physical activity in the United States measured by accelerometer. Med. Sci. Sports Exerc. 2008; 40(1): 181-88.
34. U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 1996. p. 278. [cited 2007 May 21]. Available from: www.cdc.gov/nccdphp/sgr/sgr.htm.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

35. Van de Vijver F, Leung K. Methods and data analysis for cross-cultural research (Cross-cultural psychology, vol. 1). Thousand Oaks (CA): Sage Publications; 1997. 200 P.
36. Van de Vijver F, Tanzer NK. Bias and equivalence in cross-cultural assessment: An overview. Revue européenne de psychologie appliqué. 2004; 54: 119-35.
37. Van Duyn MA, McCrae T, Wingrove BK, et al. Adapting evidence-based strategies to increase physical activity among African Americans, Hispanics, Hmong, and Native Hawaiians: A social marketing approach. Prev Chronic Dis [Internet]. 2007 [cited 2010 July 22]; 4(4). Available from: http://www.cdc.gov/pcd/issues/2007/oct/07_0025.htm.
38. Yancey AK, Ory MG, Davis SM. Dissemination of physical activity promotion interventions in underserved populations. *Am J of Prev Med. 2006; 31*(4 Suppl): S82-91.
39. Zhu W. Measurement challenges in assessing international women's physical activity. *Res Q Exerc Sport. 2000; 71*(Suppl 1): A-24.
40. Zhu W, Kurz KA. Graphical DIF analysis for assessing biased motor items/tasks. *Res Q Exerc Sport. 1996; 67*(4 Suppl): A-63 - 4.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 1. Five factors for DIF analysis

Factor	Reference Group	Focal Group
Race	Non-Hispanic Whites	Non-Hispanic Blacks
	Non-Hispanic Whites	Hispanics
Gender	Males	Females
Income	Higher income (i.e., annual family income greater than $34,999)	Low income (i.e., annual family income equal or below $34,999)
Education	Some college or above	High school or less
Age group	20 to 59 yrs	60+ yrs

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 2. Prevalence of Specific Activity Participation by Overall and Subpopulations (% [SEM])

Activity	Overall	Gender		Race/Ethnicity			Education		Income		Age Group	
		Male	Female	Whites	Blacks	Hispanics	High	Low	High	Low	Adult	Older adult
	n=1857	923	934	1211	298	348	1063	792	1189	602	1206	651
Aerobics	11 (1.1)	5 (1)	17 (1.5)	11 (1.3)	15 (2.5)	12 (1.5)	13 (1.3)	8 (1.1)	12 (1.3)	10 (1.3)	12 (1.1)	9 (2.1)
Baseball	2 (0.4)	3 (0.6)	1 (0.4)	2 (0.7)	1 (0.8)	4 (2)	2 (0.7)	2 (0.4)	2 (0.5)	2 (0.7)	2 (0.6)	1 (0.4)
Basketball	8 (1)	14 (1.9)	3 (0.5)	7 1.1)	16 (1.6)	9 (2.1)	8 (1.2)	8 (1.3)	8 (1.1)	8 (1.4)	10 (1.2)	2 (0.7)
Bicycling	19 (1.6)	22 (2)	17 (1.6)	21 (1.9)	15 (2.2)	7 (2.4)	21 (2.2)	16 (1.8)	21 (1.9)	15 (1.8)	20 (1.8)	16 (2.3)
Bowling	3 (0.5)	3 (0.6)	3 (0.6)	3 (0.6)	3 (1.5)	3 (1.1)	3 (0.5)	3 (0.9)	3 (0.5)	3 (1.2)	3 (0.5)	3 (1.1)
Dance	13 (1.4)	9 (1.4)	16 (1.7)	10 (1.3)	19 (2.1)	30 (3.5)	13 (1.5)	13 (1.6)	11 (1.7)	17 (1.6)	15 (1.4)	7 (1.9)
Fishing	5 (0.9)	8 (1.6)	2 (0.6)	6 (1)	2 (1)	1 (0.6)	5 (1)	5 (1.3)	5 (1)	4 (0.8)	5 (1.1)	6 (1.1)
Golf	9 (1)	15 (1.5)	3 (0.7)	10 (1.3)	1 (0.6)	2 (0.7)	10 (1.2)	6 (1.1)	11 (1.3)	2 (0.5)	8 (1.1)	9 (2)
Hiking	9 (1.6)	10 (1.6)	8 (2.1)	10 (1.8)	1 (0.5)	16 (2.5)	10 (1.8)	8 (1.9)	10 (1.9)	8 (1.6)	10 (1.9)	5 (1.1)
Jogging	7 (0.5)	9 (0.7)	5 (1)	6 (0.7)	11 (2)	9 (2.9)	9 (0.9)	4 (0.9)	7 (0.7)	7 (1.3)	9 (0.6)	1 (0.4)

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 2 (continued)

Activity	Overall	Gender		Race/Ethnicity			Education		Income		Age Group	
		Male	Female	Whites	Blacks	Hispanics	High	Low	High	Low	Adult	Older adult
	n=1857	923	934	1211	298	348	1063	792	1189	602	1206	651
Running	9 (0.9)	13 (1.7)	6 (0.9)	8 (0.9)	14 (1.4)	14 (2.5)	11 (1.5)	7 (1)	9 (1)	10 (1.4)	11 (1.1)	1 (0.4)
Soccer	2 (0.5)	4 (0.8)	1 (0.3)	2 (0.5)	1 (0.4)	8 (2)	2 (0.5)	3 (0.7)	2 (0.6)	4 (0.8)	3 (0.6)	1 (0.4)
Stair Climbing	7 (0.6)	6 (0.9)	8 (1.3)	7 (0.6)	8 (1)	7 (2.5)	9 (0.6)	5 (1)	7 (0.8)	6 (1.4)	8 (0.7)	5 (0.8)
Stretching	10 (0.9)	7 (0.9)	14 (1.2)	10 (1.1)	10 (2.6)	12 (1.7)	11 (1.4)	9 (1.1)	10 (1.1)	12 (1.6)	11 (1.2)	9 (1.7)
Swimming	9 (1.3)	9 (1.4)	9 (1.5)	10 (1.5)	4 (1.1)	6 (2)	11 (1.8)	5 (1)	10 (1.8)	6 (1)	9 (1.5)	9 (1.5)
Treadmill	14 (1.3)	10 (1.3)	17 (1.7)	14 (1.6)	16 (2.3)	7 (2.1)	16 (1.9)	8 (1.6)	15 (1.6)	9 (1.7)	14 (1.4)	12 (1.4)
Walking	58 (1.5)	50 (1.7)	65 (2)	60 (1.8)	52 (3.8)	44 (2.7)	59 (2.2)	56 (2.4)	60 (1.8)	53 (1.7)	56 (1.7)	67 (1.8)
Weightlifting	14 (1.2)	17 (1.6)	11 (1)	13 (1.3)	15 (2.2)	16 (2.6)	15 (1.8)	10 (2)	15 (1.3)	9 (1.5)	15 (1.3)	7 (1.3)
Yoga	4 (0.6)	2 (0.6)	6 (1.2)	4 (0.8)	3 (0.9)	3 (1.4)	6 (0.8)	1 (0.4)	4 (0.7)	4 (1.4)	4 (0.7)	2 (0.7)
Football	2 (0.2)	3 (0.5)	1 (0.3)	1 (0.3)	2 (0.7)	2 (0.8)	1 (0.4)	2 (0.5)	1 (0.3)	2 (0.6)	2 (0.3)	0 (0.3)
Hunting	3 (0.6)	6 (1.2)	0 (0.2)	3 (0.7)	0 (0.4)	0 (0.2)	2 (0.4)	5 (1.4)	4 (0.7)	1 (0.6)	3 (0.7)	1 (0.6)

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 2 (continued)

Activity	Overall	Gender		Race/Ethnicity			Education		Income		Age Group	
		Male	Female	Whites	Blacks	Hispanics	High	Low	High	Low	Adult	Older adult
	n=1857	923	934	1211	298	348	1063	792	1189	602	1206	651
Kayaking	1 (0.5)	2 (0.8)	1 (0.3)	2 (0.6)	0 (0.3)	1 (0.5)	2 (0.6)	1 (0.3)	1 (0.5)	2 (1)	2 (0.5)	0 (0.4)
Rollerblading	1 (0.3)	1 (0.5)	1 (0.3)	1 (0.3)	1 (0.6)	2 (0.4)	2 (0.3)	1 (0.5)	1 (0.3)	2 (0.7)	2 (0.3)	0 (0)
Rowing	1 (0.3)	2 (0.5)	1 (0.5)	1 (0.4)	0 (0.4)	2 (1.2)	1 (0.4)	1 (0.3)	1 (0.4)	1 (0.4)	1 (0.3)	0 (0.2)
Skating	1 (0.3)	1 (0.4)	1 (0.3)	1 (0.4)	2 (0.5)	2 (1.1)	2 (0.5)	1 (0.4)	1 (0.4)	1 (0.3)	1 (0.3)	0 (0.3)
Softball	3 (0.6)	4 (0.6)	2 (0.9)	3 (0.7)	1 (0.9)	2 (1.1)	3 (0.9)	2 (0.7)	3 (0.7)	1 (0.5)	3 (0.7)	1 (0.4)
Tennis	2 (0.4)	3 (0.5)	2 (0.4)	2 (0.5)	2 (1)	3 (1.9)	3 (0.5)	1 (0.5)	2 (0.5)	3 (0.7)	3 (0.5)	1 (0.6)
Volleyball	2 (0.5)	2 (0.6)	2 (0.5)	2 (0.6)	2 (0.8)	1 (0.5)	2 (0.6)	1 (0.6)	2 (0.6)	1 (0.5)	2 (0.6)	0 (0.4)
Frisbee	2 (0.4)	2 (0.7)	1 (0.4)	2 (0.5)	0 (0.4)	0 (0)	1 (0.4)	2 (0.9)	2 (0.4)	1 (0.6)	2 (0.5)	0 (0.3)
Martial art	1 (0.2)	1 (0.5)	0 (0.1)	1 (0.3)	0 (0.1)	1 (0.5)	1 (0.3)	1 (0.3)	1 (0.2)	1 (0.6)	1 (0.3)	1 (0.3)
Wrestling	1 (0.3)	1 (0.3)	1 (0.5)	1 (0.3)	1 (0.4)	2 (1.2)	1 (0.5)	1 (0.2)	1 (0.4)	1 (0.5)	1 (0.4)	0 (0)
Yard work	1 (0.2)	1 (0.4)	1 (0.3)	1 (0.3)	2 (0.8)	0 (0)	1 (0.3)	1 (0.4)	1 (0.2)	1 (0.6)	1 (0.2)	1 (0.4)

Note. SEM means standard error of the mean. Due to the constraint of the space, activities with prevalence less than 2% are not shown in the table.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 3. Summary of DIF Analyses by Five Major Factors

Item	Internal						External			
	SIBTEST		MH		ANOVA		MH		ANOVA	
	Beta	p-value	Chi-Square	p-value	F (1, 1538)	p-value	Chi-Square	p-value	F (1, 1539)	p-value
Race/Ethnicity (Non-Hispanic Whites vs. Non-Hispanic Blacks)										
Basketball **- a,b**	-0.198	<0.0001	41.5412	0.0001	36.43	0.0001	38.6899	0.0001	23.94	0.0001
Dance **- a,b**	-0.305	<0.0001	13.1013	0.0003	8.06	0.0046	12.148	0.0005	5.62	0.0179
Golf **+ a,b**	0.146	<0.0001	9.9739	0.0016	14.56	0.0001	11.5128	0.0007	12.93	0.0003
Hiking **+ a,b**	0.203	<0.0001	18.3937	0.0001	19.89	0.0001	18.3771	0.0001	17.44	0.0001
Race/Ethnicity (Non-Hispanic Whites vs. Hispanics)										
Bicycling **+ a,b**	0.316	<0.0001	21.3845	0.0001	29.25	0.0001	25.3779	0.0001	25.97	0.0001
Dance **- a,b**	-0.224	<0.0001	27.2431	0.0001	28.87	0.0001	23.5901	0.0001	16.35	0.0001
Golf **+ a,b**	0.088	0.001	10.4562	0.0012	13.57	0.0002	12.8835	0.0003	14.54	0.0001
Hiking - a,b	-0.032	0.375	60.8121	0.0001	50.59	0.0001	47.6286	0.0001	47.8	0.0001
Soccer - a	0.024	0.29	19.4486	0.0001	23.21	0.0001	11.0189	0.0009	6.17	0.0131
Swimming **+ a,b**	0.228	<0.0001	11.1463	0.0005	12.22	0.0005	11.822	0.0005	12.04	0.0005
Walking **+ a,b**	0.553	<0.0001	28.7935	0.0001	37.87	0.0001	27.1865	0.0001	28.02	0.0001

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 3 (continued)

Item	Internal						External			
	SIBTEST		MH		ANOVA		MH		ANOVA	
	Beta	p-value	Chi-Square	p-value	F (1, 1538)	p-value	Chi-Square	p-value	F (1, 1539)	p-value
Gender										
Aerobics **- a,b**	-0.114	<0.0001	58.0911	0.0001	62.92	0.0001	39.4582	0.0001	38.7	0.0001
Basketball **+ a,b**	0.2	<0.0001	36.8396	0.0001	43.93	0.0001	40.7204	0.0001	35.14	0.0001
Dance - a,b	-0.039	0.069	22.7259	0.0001	27.52	0.0001	13.7752	0.0002	17.99	0.0001
Fishing **+ a,b**	0.099	<0.0001	25.1159	0.0001	25.25	0.0001	30.8566	0.0001	26.72	0.0001
Golf **+ a,b**	0.089	<0.0001	40.6171	0.0001	43.17	0.0001	63.6334	0.0001	58.06	0.0001
Running **+ a,b**	0.176	<0.0001	11.6954	0.0005	15.46	0.0001	9.0634	0.0026	10.27	0.0014
Soccer + a,b	0.064	<0.0001	13.2528	0.0003	15.46	0.0001	12.8034	0.0003	11.85	0.0006
Stretching **- a,b**	-0.132	<0.0001	13.8095	0.0002	17.59	0.0001	12.3372	0.0004	15.45	0.0001
Treadmill ***- a,b**	0.488	<0.0001	19.5304	0.0001	15.14	0.0001	13.2048	0.0003	7.05	0.008
Walking **- a,b**	-0.207	<0.0001	18.0433	0.0001	25.59	0.0001	10.0517	0.0015	15.12	0.0001
Weightlifting **+ a,b**	0.173	<0.0001	16.2447	0.0001	13.08	0.0003	16.1556	0.0001	11.76	0.0006
Yoga * -a,b	0.061	<0.0001	17.327	0.0001	20.58	0.0001	18.1339	0.0001	19.64	0.0001
Hunting **+ a,b**	0.785	<0.0001	16.6381	0.0001	18.75	0.0001	21.1607	0.0001	18.91	0.0001

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Table 3 (continued)

Item	Internal						External			
	SIBTEST		MH		ANOVA		MH		ANOVA	
	Beta	p-value	Chi-Square	p-value	F (1, 1538)	p-value	Chi-Square	p-value	F (1, 1539)	p-value
Education										
Jogging *+ a,b	-0.302	<0.0001	11.979	0.0005	15.33	0.0001	17.4457	0.0001	16.75	0.0001
Treadmill *+ a,b	-0.175	<0.0001	7.7063	0.0055	10.35	0.0013	12.1013	0.0005	14.63	0.0001
Yoga *+ a,b	-0.03	<0.0001	17.0144	0.0001	19.18	0.0001	19.1329	0.0001	18.96	0.0001
Income										
Golf + a,b	0.088	<0.0001	23.8296	0.0001	28.82	0.0001	26.7387	0.0001	30.24	0.0001
Treadmill + a,b	0.051	0.025	18.5506	0.0001	15.32	0.0001	16.3794	0.0001	14.52	0.0001
Age Group										
Basketball + a,b	0.471	<0.0001	43.8064	0.0001	50.55	0.0001	24.1771	0.0001	15.93	0.0001
Dance + a,b	0.088	<0.0001	17.3357	0.0001	24.45	0.0001	17.0422	0.0001	13.21	0.0003
Golf *- a,b	0.172	<0.0001	17.3272	0.0001	9.98	0.0016	16.4798	0.0001	4.77	0.0291
Jogging + a,b	0.359	<0.0001	30.8608	0.0001	37.91	0.0001	14.0073	0.0002	12.97	0.0003
Running + a,b	0.423	<0.0001	42.6228	0.0001	56.42	0.0001	20.7574	0.0001	15.78	0.0001
Walking *- a,b	0.454	<0.0001	33.5627	0.0001	21.62	0.0001	43.4062	0.0001	26.85	0.0001
Weightlifting + a,b	0.28	<0.0001	23.4208	0.0001	23.76	0.0001	9.333	0.0023	5.58	0.0183

Note. Due to the constraint of the space, only items with significant DIF are shown in the table.

This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, Volume 43, Issue 5, 922-929. Copyright restrictions may apply. DOI: 10.1249/MSS.0b013e3181fdcc25

Activities below "Hunting" are not listed because no DIF items were detected.

"+" indicates DIF favoring the reference group.

"-" indicates DIF favoring the focal group.

[a] indicates DIF was confirmed only by internal matching criterion.

[b] indicates DIF was confirmed only by external matching criterion.

[a,b] indicates DIF was confirmed by both internal and external matching criterion.

"*" indicates the direction of DIF was inconsistent across three DIF methods. The final direction was determined by the two DIF methods that showed the same direction.

Bolded text indicates large DIF items.