Quantifying and explaining trainer variation in fitness assessments using multilevel modeling

Peter D. Hart

Health Promotion Research, Havre, MT 59501

Corresponding Author: Peter D. Hart, Ph.D.**,** Health Promotion Research, Havre, MT 59501-7751, 406.301.1043 (Phone) pdhart@outlook.com

ABSTRACT

Background: A characteristic of a good fitnesstestis its ability to create variation between participants. However, variation in fitness scores between different trainers is not often considered.

Purpose: The purpose of this study was to use multilevel linear models to quantify and explain variation in fitness scores.

Methods: Data for this research came from N=131 college students attending a rural public university. Five (5) different fitness assessments were administered to each participant and included measures of percent body fat (PBF, %), muscular endurance (ME, reps), muscular strength (MS, kg), flexibility (FL, cm), and cardiorespiratory fitness (CRF, ml/kg/min). Various random intercept and random slope multilevel models were evaluated to quantify and explain variance in participants (level 1) nested within trainers (level 2). Participant physical activity rating (PAR) and trainers' percentage of athlete clients (Athletes) were used as level 1 and level 2 predictors, respectively, while controlling for age and sex.

Results: Results from the unconditional means (variance components) modelsindicated significant trainer (intercept) variation in the CRF model only (ICC = .35, p= .020). A CRF outcome model with random intercept and fixed slopeswas the most parsimonious model found (b_{PAR} = 1.45, b_{Athletes} = 10.33, b_{Age} = -0.21, b_{Sex} = 2.54, p= .001, R^2_{Pseudo} = .57). The final model explained 32% and 68% of level 1 and level 2 variance, respectively.

Conclusion: Results from this study showedthat trainer variation in fitness assessment scoresis only considerable in CRF testing. Furthermore, athlete status of the trainers' participants explains a substantial amount of trainer variation in CRF.

Keywords: Multilevel linear models, Fitness assessment, Cardiorespiratory fitness.

INTRODUCTION

Multilevel linear modeling is a set of statistical analyses that can be used to identify and account for multilevel structure in a dataset [1]. In a cross-sectional study, multilevel data may consist of participants nested within larger units, such as schools, doctors, or clinics. Similarly, in a cross-sectional health-related physical fitness study, a sample of participants may be drawn from different fitness centers or different trainers. Multilevel linear modeling then can account for the correlated nature of nested data, typically considered a violation of independent observations (i.e., participant fitness scores from the same trainer are more likely to be correlated than scores for participants from other trainers).

Multilevel linear modeling allows for the examination of four (4)research questions [2]. **First**, is there significant variation in fitness scores across different trainers (level 2)? **Second**, can a participant-level (level 1) variable and trainer-level (level 2) variable predict fitness test scores? **Third**, does the effect of the participant-level variable on fitness scores vary significantly across trainers? **Fourth**, is the effect of the participant-level variable on fitness scores moderated by the trainer-level variable? The purpose of this study is to address these four questions and explain trainer variation in fitness scores.

METHODS

Study Variables

Participants for this study were college students enrolled at a rural public university.Five (5) different fitness assessments were administered to each participant and included measures of percent body fat (PBF, %) [3], muscular endurance (ME, reps) [4], muscular strength (MS, kg)[5], flexibility (FL, cm) [6], and cardiorespiratory fitness (CRF, ml/kg/min) [7]. Each fitness test served as the outcome variable for its own multilevel model. Physical activity rating (PAR) was measured at the participant level and was therefore used as a level 1 predictor. PAR ranged from 0 to 10, where higher scores represent greater amounts of physical activity [8]. Athletes was a variable measured at the trainer-level and therefore a level 2 predictor. Athletes was computed as the percentage of the trainers' participants who were current collegiate athletes. Two covariates were used in this study, Sex and Age. For purposes of the multilevel modeling and to make intercepts easier to interpret, PAR was group mean centered, Sex was coded 1 = male and 0 = female, Athletes was grand mean centered, and Age was group mean centered.

Statistical Analysis

Descriptive statistics were computed for study variables by sex and across high and low levels of self -reported PAR. Multilevel linear modeling unconditional means (variance components) models were first computed for each of the five (5) fitness score outcomes to determine the extent of trainer-level variance and the subsequent need to account for correlated data. From these models, the intraclass correlation coefficient (ICC) was computed as a measure of the proportion of participant fitness score variancethat can be explained by mean fitness scores across trainers. The ICC was computed as: $ICC = \tau_{00}/(\tau_{00}+ \sigma^2)$, where τ_{00} is intercept variance and $\sigma^2$is residual variance. The Wald Z statistic was also used to test for significant intercept variance in each variance components model and used as objective criteria for correlated data. Multilevel linear modeling was then used to address four (4) main questions using four (4) different models. Model 1 is the unconditional means model and used as the baseline model to judge subsequent models, $Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij}$.Model 2 is a level 1 and level 2 random intercept with fixed slopes model, $Y_{ij} = \gamma_{00} + \gamma_{10}(PAR) + \gamma_{20}(Sex) + \gamma_{30}(Age) + \gamma_{01}(Athletes) + \mu_{0j} + r_{ij}$. Model 3 is a level 1 and level 2 random intercept with fixed and random slopes model, $Y_{ij} = \gamma_{00} + \gamma_{10}(PAR) + \gamma_{20}(Sex) + \gamma_{30}(Age) + \gamma_{01}(Athletes) + \mu_{0j} + \mu_{1j}(PAR) + r_{ij}$. Model 4 is a level 1 and level 2 random intercept with fixed slopes and cross-level interaction model, $Y_{ij} = \gamma_{00} + \gamma_{10}(PAR) + \gamma_{20}(Sex) + \gamma_{30}(Age) + \gamma_{01}(Athletes) + \gamma_{11}(PAR*Athletes) + \mu_{0j} + r_{ij}$. In these models, σ^2 represents the variance of the r_{ij} (level 1) residuals, τ_{00} represents the variance of the μ_{0j} (level 2) residuals, τ_{11} represents the variance of the μ_{1j}(slope) residuals, and τ_{11} represents the random intercept and random slope covariance.

From the models, the proportion reduction in variance (PRV) was computed at each level. Specifically, PRV is the proportion reduction in intercept (level 2) or residual (level 1) variance, PRV = (τ_{00} No.Predictors - τ_{00}Model.Predictors)/τ_{00}No.Predictors. The PRV.L2 (level 2) is more important to this study because it directly relates to explained variance across different trainers.Additionally, model pseudo R^2 values were computed from observed and predicted values and represent the percentage of fitness score variance explained by the model predictor variables.SAS PROC Mixed was used for the multilevel modeling using maximum likelihood estimation, unstructured covariances, and Kenward degrees of freedom for coefficient t tests [9].

RESULTS

Table 1 displays simple statistics on the study variables by sex and across PAR groups. Significant (ps < .05) PAR group differences were seen across all variables except MS and FL in females and ME, MS, and FL in males. Table 2 contains the variance components statistics for each health-related physical fitness model.Of the five models, only CRF showed significant intercept variation (ICC = .35, p = .020) and hence the only fitness outcome requiring multilevel modeling. Table 3 contains the estimates from multilevel linear modeling predicting CRF.Model 1, the unconditional means and baseline model, indicated an overall fixed grand mean (34.3 ml/kg/min, p < .05). Model 2, thelevel 1 and level 2 random intercept with fixed slopes model, indicated fixed grand mean (32.9 ml/kg/min, p < .05), fixed slopes (b_{PAR} = 1.45, b_{Athletes} = 10.33, b_{Age} = -0.21, b_{Sex} = 2.54, all ps < .05 less Age), and variance components (σ^2 = 32.1 and τ_{00}= 7.9, ps < .05). Model 2 was also improved from the baseline model (Model 2 AIC = 856.7 vs Model 1 AIC = 907.2. Model 3, the level 1 and level 2 random intercept with fixed and random slopes model, indicated non-significant PAR slope variance. Additionally, model 3 fit statistics were not improved relative to model 2 and therefore was not considered a final model in the analysis. Model 4, the level 1 and level 2 random intercept with fixed slopes and cross-level interaction model, showed a non-significant PAR-by-Athletes interaction term, indicating no moderation effect, was also no considered a final model in the analysis. Therefore, model 2 is considered the most parsimonious multilevel linear model predicting CRF.

Model 2 indicates that approximately 57% (R^2_{Pseudo} = .57) of CRF variation is explained by participant PAR and trainer Athletes, while controlling for participant age and sex.Model 2 also reduces 68% (PRV.L2 = .68) of intercept (level 2) and

32% (PRV.L1 = .32) of residual variance by including trainer Athletes and participant PAR, respectively. Figure 1 displays fixed effects as depicted from model 2 with random intercepts only. Figure 2 makes it easy to envision the difference in predicted mean CRF scores by sex.

DISCUSSION

The purpose of this study was to employ multilevel linear modeling to explain trainer variation in fitness scores. Initial results found that CRF was the only health-related fitness test with significant trainer variation requiring multilevel modeling. Of the different multilevel models examined on CRF scores, the random intercept with fixed slopes model was the most parsimonious. This model clearly showed a large percentage of CRF variation was explained by the physical activity levels of the participants and the proportion of athletes that the trainer tested, after controlling for participant age and sex. Additionally, this model clearly reduced a considerable percentage of intercept variance – that is, CRF variance between trainers. These findings are valuable to the fitness profession because they validate the differences seen across various trainers by showing that the majority of the variation is logically due to physical activity and fitness levels of the participants.

Results from this multilevel linear modeling analysis should be considered along with the limitations of the study. One limitation is the population from which the sample was drawn. This study was conducted on college students attending a relatively smalland rural public university. Given this fact, the findings from this study may not necessarily generalize to larger universities.A second limitation concerning this study was the use of field-based physical fitness tests. The use of field-based tests may have allowed for less precision in terms of fitness trait measurement. Although criterion-based tests (i.e., hydrostatic weighing, indirect calorimetry, etc.) may have provided more precision in terms of fitness trait measurement, they also would have used more time and had been less familiar to and less efficient for the participants.

CONCLUSIONS

Results from this study showed that trainer variation in fitness assessment scores is only considerable in CRF testing. Furthermore, athlete status of the trainers' participants explains a substantial amount of trainer variation in CRF. Variation is CRF scores across different trainers is likely due to obvious and logical differences in the participants.

REFERENCES

1. Peugh JL. A practical guide to multilevel modeling. Journal of school psychology. 2010 Feb 1;48(1):85-112.
2. Bell BA, Ene M, Smiley W, Schoeneberger JA. A multilevel model primer using SAS PROC MIXED. InSAS global forum 2013 (Vol. 433, pp. 1-19).
3. Omron Fat Loss Monitor. (2012). Model HBF-306C. Omron Healthcare Co., Ltd.
4. Haff G, Triplett NT. Essentials of strength training and conditioning. 4th edition.Human Kinetics. 2016.
5. Suni J, Husu P, Rinne M. Fitness for Health: The ALPHA-FIT Test Batteryfor Adults Aged 18–69. UKK Institute for Health Promotion Research. 2009;1- 29.
6. ACSM's Guidelines For Exercise Testing And Prescription 9thEd 2014.Lippincott Williams & Wilkins.
7. Ramsbottom R, Brewer J, Williams C. A progressive shuttle run test to estimate maximal oxygen uptake. British Journal of Sports Medicine. 1988 Dec 1;22(4):141-4.
8. George JD, Stone WJ, Burkett LN. Non-exercise VO2max estimation for physically active college students. Medicine and science in sports and exercise. 1997 Mar;29(3):415-23.
9. Wang J, Xie H, Fisher JF. Multilevel models: applications using SAS®. Walter de Gruyter; 2011 Dec 23.

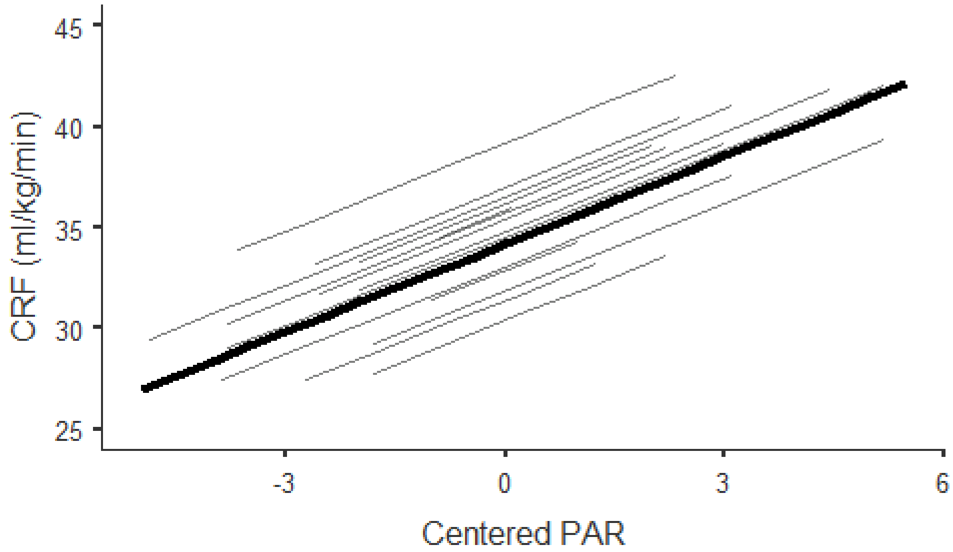

Figure 1. Plot of CRF scores regressed on centered PAR by trainer (level 2).

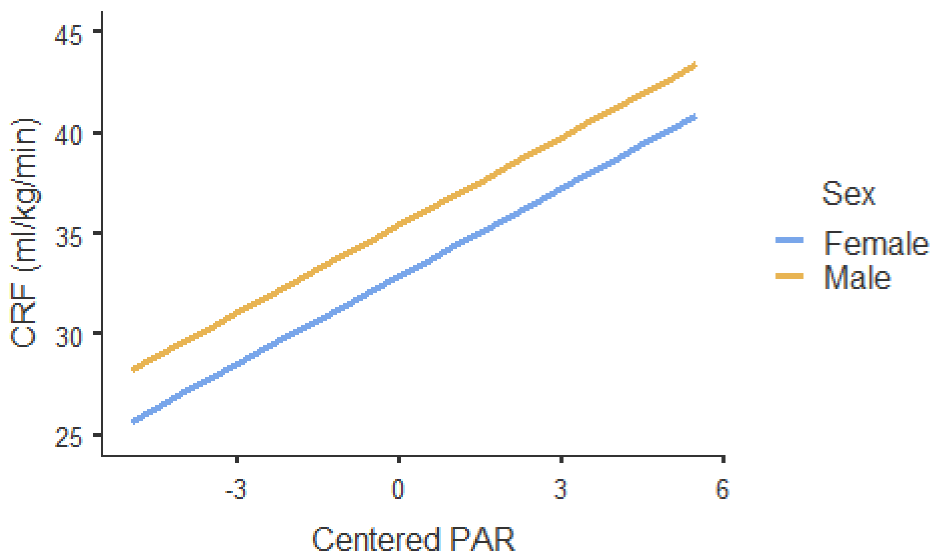

Figure 2. Plot of CRF scores regressed on centered PAR by sex.

Table1. Descriptive statistics on study variables by physical activity rating (PAR) and sex.

		Low PAR		High PAR		*t*
Sex	Variable	*Mean*	*SD*	*Mean*	*SD*	*p*
Female (N=44)	Age (yr)	24.8	8.6	20.0	1.0	.007
	PBF (%)	25.9	7.4	22.1	3.1	.023
	ME (reps)	29.9	18.0	44.3	17.5	.013
	MS (kg)	34.6	8.2	34.1	5.4	.849
	FL (cm)	33.8	7.7	35.4	8.5	.506
	CRF (ml/kg/min)	26.7	5.4	37.4	5.1	<.001

Male (N=87)	Age (yr)	21.8	4.0	20.3	2.0	.034
	PBF (%)	20.0	6.8	14.7	5.7	<.001
	ME (reps)	31.9	13.4	34.9	13.6	.311
	MS (kg)	54.3	8.4	55.4	9.2	.569
	FL (cm)	27.8	7.9	30.8	10.0	.128
	CRF (ml/kg/min)	34.4	7.6	40.5	7.4	<.001

Note. PAR is physical activity rating. Low PAR assessed as self-report PAR below sex-specific median. High PAR assessed as self-report PAR at or above sex-specific median. PBF is percent body fat (%) assessed by handheld bioelectrical impedance analysis. ME is muscular endurance assessed by YMCA bench press maximal repetitions. MS is muscular strength (kg) assessed by handheld grip dynamometer. FL is flexibility (cm) assessed by sit-and-reach box test. CRF is cardiorespiratory fitness (ml/kg/min) assessed by multistage run and beep test.

Table 2. Variance components model statistics for health-related physical fitness outcomes.

Outcome	τ_{00}	σ^2	*Z*	*p*	*ICC*
PBF (%)	6.299	49.295	1.01	.157	.11
ME (reps)	18.186	219.280	1.15	.126	.08
MS (kg)	2.308	156.590	0.26	.396	.01
FL (cm)	5.844	73.305	1.10	.135	.07
CRF (ml/kg/min)	25.192	47.029	2.06	.020	**.35**

Note. Bold ICC statistics are significant ($p<.05$). τ_{00} is the estimate of L2 intercept variance. σ^2 is estimate of L1 within group variance. Z is Wald statistic for τ_{00}. p is p-value for Z. ICC is intraclass correlation coefficient.

Table 3. Estimates from multilevel linear modeling predicting cardiorespiratory fitness (CRF).

	Model 1		Model 2		Model 3		Model 4	
Participants								
Intercept ($\gamma 00$)	**34.265**	1.455	**32.915**	1.198	**32.807**	1.194	**32.981**	1.202
PAR ($\gamma 10$)			**1.451**	0.243	**1.304**	0.377	**1.398**	0.257
Sex ($\gamma 20$)			**2.539**	1.212	**2.695**	1.205	**2.437**	1.222
Age ($\gamma 30$)			-0.212	0.132	-0.214	0.129	-0.210	0.132
Testers								
Athletes ($\gamma 01$)			**10.332**	2.484	**10.255**	2.616	**10.387**	2.485
Interaction								
PAR x Athletes ($\gamma 11$)							-0.437	0.697
Variance components								
Within-tester variance (σ^2)	**47.029**	6.218	**32.075**	4.213	**29.002**	4.148	**31.966**	4.199
Intercept variance (τ_{00})	**25.192**	12.221	**7.939**	4.485	**8.278**	4.481	**7.956**	4.489
PAR Slope variance (τ_{11})					0.827	0.790		
Intercept-slope covariance (τ_{01})					-0.294	1.270		
Intercept-slope correlation (*r*)					-.11			
Model Statistics								

ICC	.35	.20	.22	.20
DE	3.70	2.53	2.72	2.54
AIC	907.2	856.7	858.5	858.3
BIC	909.4	861.6	864.9	863.9
-2LL	901.2	842.7	840.5	842.3
Parameters	3	7	9	8
p	<.001	.001	.006	.001
R^2_{Pseudo}	.40	.57	.64	.57
PRV.L2	-	.68	.67	.68
PRV.L1	-	.32	.38	.32

Note. Estimates in bold are significant ($p < .05$). Model 1 is the unconditional means model, Yij = γ00 + μ0j + rij. Model 2 is a level 1 and level 2 random intercept with fixed slopes model, Yij = γ00 + γ10(PAR) + γ20(Sex) + γ30(Age) + γ01(Athletes) + μ0j + rij. Model 3 is a level 1 and level 2 random intercept with fixed and random slopes model, Yij = γ00 + γ10(PAR) + γ20(Sex) + γ30(Age) + γ01(Athletes) + μ0j + μ1j(PAR) + rij. Model 4 is a level 1 and level 2 random intercept with fixed slopes and cross-level interaction model, Yij = γ00 + γ10(PAR) + γ20(Sex) + γ30(Age) + γ01(Athletes) + γ11(PAR*Athletes) + μ0j + rij. σ2 represents the variance of the rij (level 1) residuals. τ00 represents the variance of the μ0j (level 2) residuals. ICC is the intraclass correlation coefficient, ICC = τ00/(τ00 + σ2). DE is the design effect statistic, DE = 1 + (n_c − 1) × ICC, where n_c = 131/15 = 8.7. R^2_{psuedo} is the percentage of CRF variance explained by the model predictor variables. PRV is the proportion reduction in intercept (L2) or residual (L1) variance, PRV = (τ00 No.Predictors - τ00 Model.Predictors)/τ00 No.Predictors. PAR is group mean centered. Sex is coded 1 = male and 0 = female. Athletes is grand mean centered. Age is group mean centered.