MEASUREMENT IN PHYSICAL EDUCATION AND EXERCISE SCIENCE, 4(1), 57–67

Objectivity and Reliability of the 90° Push-Up Test

Beth G. McManis
Institutional Research and Evaluation
Samuel Merritt College

Ted A. Baumgartner
Department of Exercise Science
University of Georgia

Deborah A. Wuest
Department of Health Promotion and Human Movement
Ithaca College

The purpose of this study was to determine the objectivity and stability reliability of the 90° push-up (90PU) test for elementary, high school (HS), and college-age students. One hundred fifty-six elementary and 70 HS students performed the 90PU test on 2 different days. Of those students, 49 elementary students were videotaped for the objectivity portion of the study. Eighty-four college students performed the 90PU on 2 different days, and each day they were scored by 2 to 4 judges. For the elementary girls and boys, the objectivity coefficients were .46 and .75, respectively, and the stability reliability coefficients ranged from .50 to .86 for the elementary and HS students. For the college students, the objectivity coefficients were between .16 and .91, with 6 of the 16 coefficients above .70. The reliability coefficients were between .22 and .87, with 5 of the 7 coefficients greater than .70. Several administration problems with the 90PU test were identified during the testing.

Key words: push-up test, reliability, objectivity, upper-arm and shoulder-girdle strength and endurance

---

Requests for reprints should be sent to Ted A. Baumgartner. Department of Exercise Science, University of Georgia, 115 Ramsey Center, Athens, GA 30602–6554. E-mail: tbaumgar@coe.uga.edu

Muscular strength and endurance is an essential component of physical fitness. The level of muscular strength and endurance affects an individual's ability to perform daily functions and various physical activities throughout the life span. Assessments of abdominal, upper-arm, and shoulder-girdle strength and endurance have been included in national physical fitness tests. Tests that measure upper-arm and shoulder-girdle strength and endurance include the pull-up (PU), flexed-arm hang (FAH), traditional push-up (TPU), and the 90° push-up (90PU).

Researchers and practitioners have criticized the PU and FAH tests. These two tests discriminate against students with greater body weight or body fat or both (Considine, 1973; Hensley, East, & Stillwell, 1982). However, Engleman and Morrow (1991) found that the discrimination against heavier students is similar in both the modified PU, where the body weight is supported, and the TPU. In addition, the PU test is plagued with a large proportion of students receiving zero scores. The FAH test does not test muscular strength and endurance throughout the full range of motion and may be difficult to administer because the student must be lifted to the bar and the clock started simultaneously.

Various features of the TPU limit its utility as a measure of upper-arm and shoulder-girdle strength and endurance in national physical fitness tests. People often have a difficult time maintaining correct form (body straight) while performing the TPU. Also, determining whether people assumed the proper down position for the TPU has been hard to judge.

In an attempt to rectify these problems, the 90PU was adopted for use in FITNESSGRAM (Cooper Institute for Aerobics Research, 1992) based on a report by Massicotte (1990) concerning various fitness assessments and norms for children and youth. The 90PU was adopted because most students can perform at least one, eliminating zero scores. Second, many students may be tested at one time by allowing students to test each other.

First, when performing the 90PU, the participant lowers himself or herself until there is a 90° angle at the elbow (down position). Second, the 90PU is performed at the rate of 20 push-ups per minute or 1 push-up every 3 sec. Like the TPU test, the participant continues performing 90PUs until he or she can no longer continue, rests between repetitions, or consistently performs the 90PUs with poor form despite corrections (with a maximum of three corrections allowed).

Before an instrument is used in a testing or research situation, the validity, reliability, and objectivity of the instrument should be established. Only a few studies concerning the reliability and validity of the 90PU test have been conducted. Tomson (1992) and Zorn (1992) conducted reliability but not validity studies on the 90PU for elementary and middle school children, respectively. However, both authors used small sample sizes and reported interclass correlation coefficients (r) rather than intraclass correlation coefficients (R). Tomson (1992) reported a reliability coefficient of .76 for 16 boys and .78 for 7 girls ages 6 to 9. In Zorn's study, the 10- and 11-year-old students ($N = 28$) were tested on 2 different days. The reli-

ability coefficient was .85 for the boys and .64 for the girls. Objectivity findings for the 90PU were not reported.

In 1994, *The Prudential FITNESGRAM Technical Reference Manual* (Morrow, Falls, & Kohl, 1994) was published to document the reliability and validity of the test items used in the Prudential FITNESSGRAM. Some additional studies are reported for tests of upper-arm and shoulder-girdle strength and endurance for a variety of age groups. The size of these groups varied from small ($N = 8–21$) to several hundred. Reliability coefficients for a single day ranged from .56 to .96 depending on the test, sex and age of the participants, and size of the group tested.

More psychometric evidence for the 90PU is needed. Only two studies regarding the reliability of the 90PU have been conducted, and these studies had methodology problems (incorrect statistic and small sample sizes). Therefore, the purpose of this study was to determine the stability reliability of the 90PU test for elementary, high school (HS), and college students and objectivity for elementary and college students.

METHODS

This study consisted of two parts: a field portion and a laboratory portion. The objectives of both portions were to determine the objectivity and reliability of the 90PU test. The specific objective of the field portion was to determine the reliability and objectivity during a typical physical fitness testing situation while mass testing was conducted. Elementary and HS students participated in the field portion of the study. The scores reported by the students were used to calculate the reliability coefficients. Judges' scores, obtained from watching a videotape of the students performing the push-ups, were used for determining the objectivity coefficients.

For the laboratory portion, the specific objective was to determine the reliability and objectivity under controlled conditions. Two to four judges counted the 90PUs performed by college students, and these values were used to calculate the reliability and objectivity coefficients. Each student was tested one at a time. The college students did not partake in counting the number of 90PUs performed.

FIELD STUDY

Participants

Seventy-three female and 83 male elementary students and 34 female and 36 male HS students participated in the stability reliability portion of the study. The majority of the elementary students were in the 3rd grade ($n = 76$). Forty-three 4th graders and forty-two 5th graders participated in the study. The mean age of the elementary

students was 9.0 years ($SD = 0.83$). Forty-one of the HS students were 9th graders, 27 were 10th graders, one was an 11th grader, and one was a 12th grader. The mean age of the HS students was 15.0 years ($SD = 0.76$). Informed consent was obtained from the students and their parents for the objectivity portion of the study because students were videotaped. The Institutional Research Board did not require written consent to be obtained for the reliability portion because the 90PU test was part of the regular testing curriculum. The physical education teachers and the principals approved the reliability and objectivity portions of the field study.

Twenty-four female and 25 male elementary students and 12 female and 8 male HS students agreed to participate in the objectivity study. Reliability and objectivity coefficients for samples less than 20 are not reported because they may be associated with an unacceptable level of measurement error (Morrow & Jackson, 1993).

Procedures

The researchers met with the students during their regularly scheduled physical education class or during lunchtime. Prior to the first testing session, the physical education teacher described the study to the students and sent a letter and informed consent form home with the students. All of the students in the class participated in the reliability study. The students who returned the informed consent forms participated in the reliability and objectivity studies.

During a practice session, the researchers described and demonstrated the 90PU test to the students using the protocol described in *The Prudential FITNESSGRAM: Test Administration Manual* (Cooper Institute for Aerobics Research, 1992). The importance of counting only properly performed 90PUs was emphasized, and students were taught how to count 90PUs. The students then practiced the test, and the researchers and physical education teacher corrected improper form or push-up counting techniques.

Three to 7 days later, at the first testing session, the 90PU test was administered. The students were paired, and the test was described and demonstrated again. One half the students were tested at a time with one student in each pair performing the test while the other student in the pair counted the number of properly performed 90PUs. The students then switched places and the test was administered a second time. The number of properly performed 90PUs was recorded for each student.

Three to 7 days after the first testing session, the students were retested. The same testing procedures used on the 1st testing day were followed. The scores reported in this session and the previous session were used to estimate the reliability for the field portion of the study. During each of the testing sessions, the students who had returned consent forms were videotaped for the objectivity portion of the study. One to three students were taped simultaneously, and each student was taped only once. The video camera was set up with the lens pointing toward the

students' heads. The camera was elevated so that, in most cases, the student's entire body could be seen in the videotape.

After the testing sessions were completed, the tapes were analyzed. A researcher and graduate student (judges) viewed the tape. Judges had been trained to identify properly performed 90PUs before any data collection occurred. The judges viewed the tapes independently and recorded the number of 90PUs each student performed properly. Because the camera was facing the students, it was easy to see the angle of the elbow. Excessive swaying of the upper or lower back also could be detected.

Data Analyses

Objectivity estimates for a single judge and stability reliability estimates for a single day were calculated with an *R* and a one-way analysis of variance (ANOVA) model with the formulas presented by Baumgartner and Jackson (1995). Stability reliability for a single day was estimated for each subgroup (i.e., elementary girls, elementary boys, HS boys, and HS girls). The stability reliability coefficients were calculated using only the scores reported by the students. Objectivity coefficients were calculated for each subgroup using the scores reported by the two judges.

Results

The means for the test and retest sessions are reported in Table 1. The *R* values for stability reliability of scores collected on a single day also are reported in Table 1. The HS girls had the highest stability reliability estimate ($R = .86$). The *R* values ranged from .50 to .86.

The means for the objectivity portion of the study are reported in Table 2. Means for scores reported by the students always exceeded the means for judges' scores. Judge 2 always had a higher mean score than Judge 1.

TABLE 1
Means and Stability Reliability Coefficients for the 90° Push-Up Test for Elementary and High School Girls and Boys

Group	*n*	*Test*[a] *(M)*	*Retest*[a] *(M)*	*1-Day Reliability (R)*
Elementary girls	73	16.4	17.9	.64
Elementary boys	83	20.0	20.4	.71
High school girls	34	14.7	15.4	.86
High school boys	36	17.3	14.8	.50

[a]Mean scores were calculated from scores reported by the students.

TABLE 2
Means for the Objectivity Data for the 90° Push-Up Test for Elementary Girls and Boys

Group	*n*	*Student (M)*	*Judge 1[a] (M)*	*Judge 2[a](M)*
Elementary girls	24	15.6	3.0	6.1
Elementary boys	25	17.9	6.4	8.7

[a]Judges scores are from observing a videotape.

The objectivity estimates calculated from scores reported by the two judges for a criterion score based on a single judge were .46 for the elementary girls and .75 for the elementary boys.

LABORATORY STUDY

Participants

Forty male and 44 female college students in fitness classes who were accustomed to doing push-ups or modified knee push-ups participated in the study. Informed consent forms were completed by all students prior to the beginning of data collection.

Procedures

Judges included a faculty member and three graduate students who were experienced in administering physical performance tests. They spent considerable time discussing the test directions and practicing administering the 90PU test prior to collecting data.

On the 1st day, the 90PU test was explained and demonstrated to the students, who then practiced the test and were coached on correct form by the judges. Five to 7 days later, the students were tested for the maximum number of 90PUs possible. Seven days later they were retested. During the testing, two to four judges watched the test and independently recorded the number of 90PUs the student had properly performed. All judges could not score all students due to schedule conflicts.

Data Analyses

Objectivity estimates for a single judge and stability reliability estimates for a single day were calculated using an *R* and a one-way ANOVA model with formulas presented by Baumgartner and Jackson (1995). Two to four judges recorded the

number of 90PUs that the student performed. It is possible that the agreement between the pairs of judges could differ if one judge was consistently different from the other judges. Therefore, objectivity estimates were calculated using the scores of two judges at a time. For example, if three judges (A, B, and C) recorded the scores, objectivity estimates were calculated for A versus B, A versus C, and B versus C. Reliability estimates were calculated for each judge so that any differences in scoring among judges would not influence the reliability estimates.

Results

Objectivity and reliability estimates were calculated when the number of students tested was at least 20. The mean scores for each judge and the objectivity coefficients are reported in Table 3. The objectivity coefficients ranged from .16 to .91. The lowest (.16) and highest (.91) coefficients were obtained for agreement between Judges 2 and 3. The difference between two judges in mean scores was usually three to four 90PUs. Judge 1 tended to score lower than the other three judges. Judges 2 and 3 were in agreement the majority of the time with three objectivity coefficients of .70 or higher.

Means for Days 1 and 2 for each judge and the stability reliability coefficients are reported in Table 4. Reliability coefficients were between .22 and .87.

TABLE 3
Judge Means and Objectivity Coefficients for a Single Judge Using the Scores of Two Judges for College Women and Men

Judges	*Student Sex*	*Day*	*n*	*First Judge (M)*	*Second Judge (M)*	*One-Judge Objectivity (R)*
1 and 2	M	1	27	25.07	28.03	.61
1 and 2	F	1	38	3.53	6.26	.69
1 and 2	M	2	27	22.96	27.37	.71
1 and 2	F	2	38	5.08	7.00	.85
1 and 3	M	1	34	24.79	28.00	.62
1 and 3	F	1	28	4.89	6.89	.63
1 and 3	M	2	34	24.85	24.00	.48
1 and 3	F	2	28	6.36	4.79	.68
1 and 4	F	1	29	3.72	10.14	.31
1 and 4	F	2	29	5.93	9.28	.64
2 and 3	M	1	21	27.33	28.33	.91
2 and 3	F	1	22	6.73	6.55	.88
2 and 3	M	2	21	25.95	21.29	.16
2 and 3	F	2	22	7.77	5.00	.70
2 and 4	F	1	29	7.38	10.14	.68
2 and 4	F	2	29	8.00	9.28	.79

TABLE 4
College Student Day Means by Judge and Stability Reliability Coefficients

Judge	*Student Sex*	*n*	*Day 1 (M)*	*Day 2 (M)*	*1-Day Reliability (R)*
1	M	40	25.48	24.95	.68
1	F	44	4.00	5.61	.84
2	M	27	28.04	27.37	.75
2	F	38	6.26	7.00	.84
3	M	34	28.00	24.00	.22
3	F	28	6.89	4.79	.75
4	F	29	10.14	9.28	.87

Seventy-one percent of the time the Day 1 mean was greater than the Day 2 mean for a judge.

DISCUSSION

Recommendations for acceptable reliability coefficients vary. Baumgartner and Jackson (1995) recommended that an acceptable objectivity or reliability coefficient for a psychomotor test must be at least .80. This is a slightly lower than the recommendation of .85 by Safrit (1986). Portney and Watkins (1993) suggested that reliability coefficients above .75 should be considered good to excellent and coefficients below .75 should be considered moderate to poor. However, the final interpretation of the coefficient must be left to the reader, test user, or both. The test user needs to assess the practical significance of the reliability coefficient (Morrow & Jackson, 1993).

Very little research has been conducted concerning the reliability and objectivity properties of the 90PU test. Tomson (1992) and Zorn (1992) conducted reliability studies on the 90PU test and found the reliability coefficients ranged from .64 to .85. However, Tomson used small sample sizes, and both Tomson and Zorn reported interclass correlation coefficients (r) instead of intraclass correlation coefficients (R).

In this study, the stability reliability Rs based on a single day for the HS and elementary boys and girls ranged from .50 to .86. These coefficients are similar to what Tomson (1992) and Zorn (1992) found.

The objectivity estimates for the elementary students were .46 (girls) and .75 (boys). The scores reported by the elementary students were, on average, two to three times higher than the scores reported by the independent judges. The mean scores for Judge 2 were higher than the mean scores for Judge 1 (see Table 2). Based on these results, elementary students do not appear to be capable of properly

counting 90PUs. Thus, testing with students counting each other's push-ups does not seem appropriate.

Stability reliability coefficients for the college students were higher than the objectivity coefficients. The stability coefficients ranged from .22 to .87. The lowest coefficient (.22) was far below the others and may be an outlier. The next lowest coefficient was .68. Five of the seven coefficients were higher than .70.

The objectivity coefficients for the college students ranged from .16 to .91. Thirteen of the 16 coefficients were above .60; however, only 6 of the coefficients were above .70. It is unknown how the college students would have scored the 90PUs because they were not required to count push-ups performed by other students.

Several problems with the 90PU were encountered during the testing sessions. It was difficult to discriminate between a properly and improperly performed 90PU. The mean scores of the trained judges usually differed by three or more 90PUs. The elementary and HS student's scores exceeded the judges scores by 3 to 11 90PUs. The students seemed to count every 90PU performed whether it was done correctly or not.

Counting the number of correctly performed 90PUs was difficult. Numerous times, the judge(s) could not see if the elementary and HS student's arms reached the required 90° angle because the student was wearing a baggy shirt. Even with the college students, who were required to wear tight, short-sleeved shirts, judging whether students' arms reached the 90° angle was difficult. For all students, it was difficult to determine how much back sway up or down was unacceptable.

The elementary students had a difficult time understanding how to properly perform the 90PU. Although they practiced the test and their arm form was corrected, many of the students were unable to lift their hips off of the ground. They lay on the ground and extended their elbows but their hips remained on the ground. Many of the students were not strong enough to maintain the correct form, and often their elbows did not reach the required 90° angle. This test may be too difficult for elementary students.

Performance in time to the cadence presented additional difficulties for the elementary students. The cadence seemed too slow for many of the students. Some of the students had a difficult time generating strength to push up from the down position after being there for several seconds. Others appeared to collapse after having to support themselves in the up position between push-ups. Increasing the cadence of the test or eliminating it may be beneficial.

Low-strength college women had form problems similar to the elementary students. They had difficulties maintaining correct form and often did not have sufficient strength to control their bodies as they tried to come down to the required 90° angle or tried to come up from the down position. They were more successful performing push-ups on their knees as an exercise in their class.

The placement of the hands can affect the results of the test. As presented in the FITNESSGRAM manual, the hands are supposed to be placed underneath the shoulders. However, the figure in the manual does not seem to show the same placement. If the hands are moved beyond shoulder width, the 90PUs are easier to perform. Although the students were told where to place their hands each testing session, they could have placed their hands differently for the retest than for the test, which would have affected the stability reliability.

Most of the objectivity and reliability coefficients for the 90PU were above .60. However, the researchers encountered several problems, such as excessive swaying and insufficient strength to complete the push-up, while administering the test. Many of these problems were similar to those with the TPU test. The only differences between the 90PU protocol and the TPU protocol are the 90° angle at the elbow as the down position and the cadence. Researchers need to determine how the 90PU protocol can be modified to eliminate many of these problems and, hopefully, improve the reliability and objectivity of the 90PU test. Some possible alterations may include increasing the cadence, clarifying the hand placement in the manual, and changing the protocol to a push-up from the knees and requiring the students to wear tight-fitting shirts or to tuck their shirts into their shorts.

There is a need for an easily administered test of upper-arm and shoulder-girdle strength and endurance. Modifications of the 90PU test may provide practitioners with a useful, objective, and reliable measure of this important fitness component.

ACKNOWLEDGMENT

We wish to express our appreciation to Libby Bolton, Fred Caliel, Bob Congdon, Cathy Haight, Lisa Kanter, Roseanne Morosco, Vic Mancini, Scott Murr, David Rowe, and Kent Wagoner.

REFERENCES

Baumgartner, T. A., & Jackson, A. S. (1995). *Measurement for evaluation in physical education and exercise science* (5th ed.). Dubuque, IA: Brown & Benchmark.

Considine, W. J. (1973). *An analysis of selected upper body tasks as measures of strength* [abstract]. AAHPER Research Section, Houston, Texas.

Cooper Institute for Aerobics Research. (1992). *The Prudential FITNESSGRAM: Test administration manual.* Dallas, TX: Author.

Engelman, M. E., & Morrow, J. R., Jr. (1991). Reliability and skinfold correlates for traditional and modified pull-ups in children Grades 3–5. *Research Quarterly for Exercise and Sport, 62,* 88–91.

Hensley, L. D., East, W. B., & Stillwell, J. L. (1982). Body fatness and motor performance during preadolescence. *Research Quarterly for Exercise and Sport, 53,* 133–140.

Massicotte, D. (1990). *Partial curl-ups, push-ups, and multistage 20 meter shuttle run, national norms for 6 to 17 year olds* (Final report submitted to the Canadian Association for Health, Physi-

cal Education, and Recreation [CAHPER], and Fitness and Amateur Sport Canada). University of Quebec, Montreal, Canada.

Morrow, J. R., Jr., Falls, H. B., & Kohl, H. W. (Eds.). (1994). *The Prudential FITNESSGRAM technical reference manual.* Dallas, TX: Cooper Institute for Aerobics Research.

Morrow, J. R., Jr., & Jackson, A. W. (1993). How "significant" is your reliability? *Research Quarterly for Exercise and Sport, 64,* 352–355.

Portney, L. G., & Watkins, M. P. (1993). *Foundations of clinical research: Applications to practice.* Norwalk, CT: Appleton & Lange.

Safrit, M. J. (1986). *Introduction to measurement in physical education and exercise science.* St. Louis, MO: Times Mirror/Mosby.

Tomson, L. M. (1992). *A field trial of strength and endurance tests for primary grade children.* Unpublished master's thesis, Arizona State University, Tempe.

Zorn, R. L., Jr. (1992). *Selected tests of muscular strength and endurance for children.* Unpublished master's thesis, Arizona State University, Tempe.