Article

Measurement properties of scores from a novel blind football assessment for adolescents with visual impairments

Adam Pennell [1], Emily Gilbert [2], Ali Brian [3], Pamela Beach [4], and Lauren Lieberman [4]

Received: 3rd February 2021; Accepted: 2nd December 2021; 28th April 2022:

Abstract: Football (soccer) is one of the most popular sports for individuals with visual impairments (VI) worldwide but nascent within the United States (US). While there is burgeoning interest toward developing blind football at grassroot and national levels, particularly with the forthcoming 2028 Paralympics, no football-based talent identification or skill assessment has been developed or vetted in the US. Therefore, the purpose of this study was to investigate the immediate measurement properties of scores from a novel football assessment in amateur-skilled adolescents with VI in the US. Adolescents (n = 57) with VI physically completed the Blind Football Skills Test for Adolescents with VI (BFST-AVI) which was comprised of five football-related tasks. Scores were analyzed by item analyses, internal consistency, convergent validity, and construct validity procedures. The results suggested that total and individual assessment scores had acceptable levels of reliability and validity. The BFST-AVI can be adopted immediately by grassroot organizations, national governing bodies, coaches, adapted and general physical educators, and associated practitioners for talent identification or general motor skill assessment purposes for amateur-skilled adolescents with VI.

Keywords: soccer; children; youth; reliability; validity; evaluation; talent identification; motor skill; Paralympics

Introduction

Association football, other known as soccer, is the most popular sport in the world (Fédération Internationale de Football Association, 2018). Subsequently, it is unsurprising that adapted versions of football have become one of the most popular sports for individuals with visual impairments (VI) worldwide (Gamonales et al., 2018; International Blind Sports Association [IBSA], 2021). Blind football and partially sighted football, which are similar to conventional futsal, are adapted varieties of football for individuals with VI that have centralized rules, regulations, and competitions (IBSA, 2017a,b) which have gained popularity in recent years.

In sanctioned partially sighted 5-a-side competitions, four field players must be a combination of individuals who are partially sighted while the goalkeeper may be fully sighted or partially sighted. During partially sighted 5-a-side, none of the players wear eyeshades. While the qualifications for goalkeepers in blind football are the same as in partially sighted football, the four field players must be completely blind and wear eyeshades. These criteria are most useful for official blind or partially sighted football competitions. However, any individual could participate in a version of blind football in less formal settings (e.g., physical education, recreational sport participation). Further, it is important to note that while blind and partially sighted football modalities were established with individuals with VI, individuals with VI may also participate or have interest in non-adapted forms of football.

Although the game of football is popular worldwide and across individuals with VI, adolescents with VI have not been provided with equitable pathways to participate in blind football (Gamonales et al., 2018). This is evident at elite and beginner levels of competition. For example, in the United States (US), neither a national blind football team nor a developmental pathway for blind or partially sighted adolescent players have been supported—financially or otherwise—by a federal governing body (Oh et al., 2021). Yet, the US Olympic and Paralympic Committee hopes to field a national blind football team for the first time at the Los Angeles 2028 Paralympic Games as the country will have an automatic bid as the host nation (United States Association of Blind Athletes [USABA], 2021). Given the immediate and long-term implications of these systemic shortcomings, there is a pressing need to develop general structures and opportunities for individuals with VI to become familiar with, participate in, and develop skills related to the sport of blind football.

Football-related participation for younger individuals with VI is particularly relevant as childhood and adolescence are timepoints in which sport participation is encouraged and fostered. If adolescents with VI play football early and often, it is reasonable to suggest that doing so could lead to improved biopsychosocial outcomes (e.g., increased: fitness levels, autonomy, mobility and orientation, relatedness; Gamonales et al., 2018; Lankhorst et al., 2019) or lead to superior football-related success across time (Bergeron et al., 2015). Further, professional- and elite-level football entities typically rely on long-term childhood and youth development through clubs, academies and farms. Thus, young talent should be identified and nurtured toward elite levels of performance as they age.

To this end, talent identification concerning blind football within the US is one of the most pressing needs for individuals with VI along with hiring coaches, procuring funding, initiating training sessions, and participating in competitions (Oh et al., 2021). However, a significant limitation exists as a population-specific football assessment has not been developed or vetted for individuals with VI, adolescent or otherwise. This is a significant issue, as previous football assessments (e.g., tasks, equipment) may not be appropriate for adolescents with VI. Further, it cannot be assumed that the measurement properties of a previously introduced football-skill assessment would be equivalent in adolescents with VI (American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education, 2014).

Thus, a football assessment created specifically with adolescents with VI in mind that also has satisfactory measurement properties is increasingly needed as such a tool could have utility in talent identification and related settings. For example, if deemed to have acceptable forms and levels of reliability and validity, the proposed assessment scores could be used by grassroot organizations or national governing bodies for talent identification purposes at the amateur level. Further, the assessment scores could have utility in adapted or general physical education and coaching settings. Therefore, the purpose of this study was to investigate the immediate measurement properties of scores from a novel football assessment in amateur-skilled adolescents with VI in the US. It was hypothesized that scores would, in general, have sound measurement properties in adolescents with VI.

Materials and Methods

Participants

A convenience sample of adolescents with VI (9-18 years) was recruited from three separate week-long overnight educational sport camps for adolescents with VI. All camps occurred in two states within the US. Two camps were in New York and one camp was in Florida.

Participants varied in degrees of VI as classified by the USABA (2017) visual classification scale. The USABA scale has four categories: B1, B2, B3, and B4. B1 represents the lowest level of visual acuity: no light perception in either eye up to light perception, and an inability to recognize the shape of a hand at any distance or in any direction. B4 characterizes the highest level of visual acuity: from visual acuity above 20/200 and up to visual acuity of 20/70 and a visual field larger than 20 degrees in the best eye with the best practical eye correction. B2 is demarcated as the ability to recognize the shape of a hand up to visual acuity of 20/600 or a visual field of less than 5 degrees in the best eye with the best practical eye correction while B3 includes visual acuity above 20/600 and up to visual acuity of 20/200 or a visual field of less than 20 degrees and more than 5 degrees in the best eye with the best practical eye correction.

Assessments

Blind Football Skills Test for Adolescents with VI (BFST-AVI). The BFST-AVI includes five sport-specific skills: timed dribbling, push passing with accuracy, shooting zones, timed continuous passing and receiving with feet, and a timed continuous catch and roll with the hands as a goalkeeping assessment. In-depth protocols for the BFST-AVI can be found in a free online manual on the United States Association of Blind Athlete website (Gilbert & Pennell, 2019). Equipment for the BFST-AVI includes official blind football balls (Handi Life Sports, Skibby, Denmark), folding metal football goals (10'W x 5'H x 4'D; Item #15657; FlagHouse, Inc., Hasbrouck Heights, NJ, US), the 'Original 4 Footer' curved pop-up goals (4'W x 2.5'H x 2.5'D, Pugg Company, Inc., Boston, MA, US), cones, and stopwatches.

Timed dribbling (Dribble). The timed dribbling task requires participants to dribble around a 5-meter equilateral triangle-shaped course. Raw scores are measured in seconds, beginning on the verbal "go" command and stopped after the participant reaches the original starting cone. Each participant completes two trials by traveling to the right or left direction. Participants choose the preferred direction, however, both trials must occur in the same direction. Justification for this process will be provided within the Results section. The quickest of the two trial is retained for analysis. Verbal feedback from an evaluator may be used to provide distance and directional cues to cones or the ball during the task.

Push passing with accuracy (Accuracy Pass). Participants attempt twelve passes toward three Pugg goals using a predetermined sequence and limb. Left- and right-sided Pugg goals are positioned in line with a central Pugg goal. The left and right side goals are placed one foot (0.33 m) from the outer edge of the center goal to the outer edge of the left and right goals. While standing five meters away from the center goal, participants pass to the passer's left, right, and center goals using their preferred leg. Participants repeat this sequence for a second time using their preferred leg. Next, participants repeat the passing sequence twice more using their non-preferred leg. Each pass that successfully lands within the correct goal is scored as one point. A maximum of six points for the preferred and non-preferred sides, respectively equates to 12 total points. For this skill, an evaluator should be positioned behind the center of the goal being passed to (in the set sequence) to provide verbal directional cues to the athlete: "[Name]. Pass to me."

Shooting zones (Shooting). Participants complete ten uncontested penalty-style shots on a 10-foot (3.33 m) wide by 5-foot (1.52 m) high football goal from a six meter distance. Two cones are placed within the goal designating the outer 2.5 feet (0.76 m) of the left and right sides of the goal. The outer 'green' zones are worth two points and the inner 5-foot (1.52 m) 'red' zone in the middle of the goal is worth one point. The first five trials are completed using the preferred limb. The last five trials are completed using the non-preferred limb. Based on a maximum of ten points for the preferred and non-preferred sides, this assessment was out of 20 total points. Evaluators tap on each side of the goal, return to the center, and verbally state "center" before participants shoot.

Timed continuous passing and receiving with feet (Continuous Pass). From two metres apart, participants complete as many successful consecutive passes and receptions as possible with an evaluator without VI in 30 seconds (back and forth pass = one point; half points were not permitted). Both parties are instructed to remain in a one-by-one metre coned-off zone across from each other. A successful pass and reception had to occur within the one-metre by one-metre zone to count towards the raw score. A bad pass or reception that resulted with the ball going outside of the respective zone was not counted. Participants complete one trial on their preferred limb followed by one trial on their non-preferred limb. Participants are permitted to trap the ball with either foot; however, to count towards their raw total, participants are required to only pass with the foot being evaluated during that trial. The total score for this task is determined by summing the number of correct preferred and non-preferred passes.

Timed continuous catch and roll with the hands (Goalkeeping). For the goalkeeping assessment, participants catch and roll back an evaluator-tossed blind football at around the chest-level as many times as possible within 30-seconds. A catch is successful if the ball is caught without crossing over a goal line which was directly behind each participant. Each successful catch is equal to one point. Two trials are completed with the highest of the two scores retained for analysis.

Convergent validity measures. In tandem with the BFST-AVI, two additional football-related outcomes were assessed. Specifically, participants were scored based upon their process- and product-oriented motor performance when kicking a stationary beep kickball (The Beep Kickball Association, Atlanta, GA, US) with the preferred-limb. Kicking a stationary ball was chosen as (a) it directly translated to the game of football (e.g., striking a ball) and (b) the measurement properties of the skill have been shown to be robust in adolescents with VI (Brian et al., 2018). When collected in tandem, both product and process outcomes can provide complementary yet independent information concerning kicking performance. As a result, it is reasonable to suggest that higher-skilled players should perform with improved qualitative and quantitative scores.

TGMD-3 kick. The Test of Gross Motor Development-third edition (TGMD-3; Ulrich, 2019) kick is a process-oriented item with four components. The four components include (1) stepping alongside or near the ball, (2) contacting the ball with the instep, (3) a rapid, continuous approach, and (4) and an elongated stride prior to contact (Ulrich, 2019). Kicking criteria from two scored trials with four criteria per trial are coded with either a one (met criteria) or zero (did not meet the criteria) with total scores ranging from 0-8 points. TGMD-3 scores have strong measurement properties in adolescents with VI up to 18 years of age including modifications specifically for individuals with VI such as utilizing a beep kickball to perform the kick (Brian et al., 2018). Per the standard TGMD-3 guidelines, only the preferred limb is assessed.

Peak ball velocity using radar. Product-oriented performance was quantified as peak beep kickball velocity in miles per hour using a Stalker Pro II radar gun (Stalker Radar, Plano, TX, US; Stodden et al., 2014). Peak ball velocity can occur concurrently with another assessment, such as the TGMD-3 trials.

Procedures

The Institutional Review Board of the University of South Carolina (ID: Pro00085394) and the summer camp sites where data were collected approved all procedures prior to participant recruitment and data collection. Throughout the study, the research team adhered to all Institutional Review Board rules, regulations, and training requirements. Assessments for adolescents with VI were administered in the months of June, July, and August in 2019 at three separate summer camps. Upon arrival to camp, parents and participants completed parental consent and child assent before being enrolled in the study.

Each task was administered using stations run by the researchers and research assistants. Prior to data collection, all researchers and assistants were trained on their specific assessment and on the verbal or directional cues that could be provided to the participants (e.g., "Pass towards my voice."). Researchers followed a least-to-most prompting strategy for auditory cueing. This was to ensure that the participants were being evaluated on their skill, and not their comprehension of the assessment. All assessment items were pre-taught and described (Lieberman et al., 2013). For example, dribbling pre-teaching included a description of the task being assessed and having the athlete walk the triangle assessment space prior to completing the official trials.

At each of the three camps, process (TGMD-3) and product (peak ball velocity) measures were also collected by the researchers and research assistants. Kicking a stationary beep kickball with the preferred limb was digitally recorded and retroactively coded pursuant to TGMD-3 protocols and modifications for adolescents with VI (Brian et al., 2018; Ulrich, 2019). Further, maximum kick velocities, measured in miles per hour during each of the scored TGMD-3 kick trials were assessed and recorded live using a radar gun.

Data Analysis

Descriptive statistics were calculated followed by numerous inferential analyses. Where applicable, statistical significance was determined when $p < .05$. All analyses were conducted using R statistical software version 4.0.3.

Missingness. Missingness was assessed using the *VIM* package (Templ et al., 2021) prior to statistical analyses. A cumulative total of 9.1% of the data was missing due to inclement weather. As such, much of the missing data was contained to specific individuals at a single location for a specific skill or trial. While elevated levels of missingness of values greater than 10% can lead to increased bias (Bennett, 2001), it has been suggested that the pattern and mechanism of the missing data can be of greater importance (Tabachnick & Fidell, 2012). Based on the circumstances of the data collection process, it was concluded that the missing data were missing completely at random (Kang, 2013; Rubin, 1976). Likewise, there was a univariate pattern of missingness such that the general pattern of missingness was not a random, arbitrary pattern (Dong & Peng, 2013). Both the missing completely at random mechanism and the univariate pattern are beneficial traits for imputation (Dong & Peng, 2013).

Typical practices for missing data include listwise and pairwise deletion; however, these methods have severe shortcomings that imputation can potentially remediate (Dong & Peng, 2013). Thus, all missing cells were imputed using the missForest package which is a nonparametric iterative imputation method based on a random forest (Breiman, 2001). The missForest package can handle different variable types and high-dimensional data. Further, the computation speed is expeditious, and the package does not require the use of a tuning parameter (Stekhoven & Bühlmann, 2012).

An additional benefit of the missForest package is that it returns two out-of-bag errors: normalized root mean squared error for continuous data and proportion of falsely classified entries for imputed categorical data. In both cases, good out-of-bag error performance occurs when a value is close to zero while poor performance occurs when a value is closer to one (Oba et al., 2003; Stekhoven & Bühlmann, 2012). Only the normalized root mean squared error was calculated as the raw data were continuous in nature. The normalized root mean squared error for the current imputation was .21 highlighting a satisfactory imputation error estimation.

Quartile scoring. Due to contrasting levels of variance within the raw scores of each skill and the non-uniform units between items (e.g., seconds, points), all raw, continuous-scaled item scores were converted to an ordinal four-point scale using within-sample quartile cut-off points. This process is known as discretization or binning. Quartiles were selected as the appropriate binning procedure based on the observations of Altman and Bland (1994). Discretization is commonly viewed as a downgrading of measurement where the loss of information is difficult to compare across studies (Bennette, & Vickers, 2012). However, such a process can have practical advantages (Kim & Frisby, 2019). Based on these conversions, the assessment included five items each worth four maximum points and became worth 20 total points. Whereas limitations exist, it is suggested that these quartile-based cut-offs and scores may be used for initial scoring and interpretation by professionals, where higher total scores = better overall performance, in amateur adolescents with VI.

Analytic Approach

Pre-screening of the Dribble item. During the original data collection, participants completed the dribble task to the left and right sides, two times each side. However, doing so was a lengthy process. Therefore, we investigated for statistical differences between the quickest right and left side directions for the dribble item using an Exact Wilcoxon signed-rank test with the Pratt modification. If the two quickest times were not statistically different, it would be reasonable to suggest that participants only need to travel in one direction for two trials which would significantly cut down on administration time.

Item associations. Zero-order Spearman correlations (ρ) were used to assess the direction and strength of the monotonic relationship between the individual items. Absolute value two-tailed bivariate coefficients were classified as very strong (≥ .90), strong ($.70 \geq \rho \leq .89$), moderate ($.50 \geq \rho \leq .69$), low ($.30 \geq \rho \leq .49$), or negligible ($.00 \geq \rho \leq .29$) (Hinkle et al., 2003). Overall, it has been suggested that correlations between individual items should range between .30 and .70 (Ferketich, 1991). Evidence of association was concluded if statistical significance was confirmed (i.e., $p < .05$).

Internal consistency. Internal consistency is the interrelatedness of components within an assessment (Tavakol & Dennick, 2011). To assess internal consistency of the scores, Cronbach's alpha, McDonald's (ordinal) omega (Crutzen & Peters, 2015), the average bias-corrected inter-item correlation (Fisher, 1958), and item-total correlations (without item in total) were investigated. An internal consistency coefficient of ≥ .70 but ≤ .90 for Cronbach's alpha and McDonald's omega was deemed as strong yet non-redundant (Tavakol & Dennick, 2011). The average bias-corrected inter-item correlation should fall between .15 and .50. Specifically, the average should be between .40 and .50 for a narrow construct (Clark & Watson, 1995). Item-total correlations, without item in total, required a minimum correlation of .30 (Cristobal et al., 2007). Changes in Cronbach's alpha with item removal were also calculated and assessed.

Item difficulty. Item difficulty is the proportion of individuals who select the "correct" or received the maximal score on an item. That is, items should not have an obvious floor with values closer to 0 (too hard) or ceiling effects closer to 1 (too easy). Difficulty values for items should range between .20 and .80 whereas an average item difficulty of .50 is typically sought by developers (McIntire & Miller, 2007). For a four-point quartile-based scale, the optimal ideal difficulty was determined to be .50. The ideal difficulty = $p + (1 - p) / 2$ where p = chance of scoring a one, two, three, or four = .25.

Convergent validity. "Convergent validity reflects the extent to which two measures capture a common construct" (Carlson & Herdman, 2012, p. 18). To investigate convergent validity, zero-order Spearman correlations (ρ) were used to assess the direction and strength of the monotonic relationship between the total quartile scores and a kicking process score, as measured by the total score for the TGMD-3 kick, and a kicking product outcome, as represented by maximal ball velocity. Correlation magnitude was determined using the aforementioned cut-offs provided by Hinkle and colleagues (2003). Evidence of convergent validity was concluded if statistical significance was confirmed.

Construct validity—known groups method. The known groups method, such that evidence that assessment scores can discriminate between specific groups (Portney & Watkins, 2015) can be used to provide general evidence of construct validity. Construct validity can be defined as "representing the correspondence between a construct (conceptual definition of a variable) and the operational procedure to measure or manipulate that construct" (Schwab, 1980, p. 5). It was logical to hypothesize that, on average, adolescents with lower degrees of vision would have lower total assessment scores (Brian et al., 2018; Skaggs & Hopper, 1996; Haibach et al., 2014). Typically, an analysis of variance would be used to examine differences between groups using the four USABA degrees of visual classification (i.e., B1, B2, B3. B4). However, given that there were four unequal sized groups, a linear regression analysis was performed to determine if degree of vision was predictive of total BFST-AVI scores. Linear regression was preferred as it is viewed as a more flexible alternative to the analysis of variance (Plonsky & Oswald, 2017). Specifically, the B1, B2, B3, and B4 visual classifications were dummy coded as 1, 2, 3, and 4, respectively. This analysis was performed after all assumptions were met as determined by visual inspection of the residual versus fitted, normal Q-Q, and scale-location plots. Effect size was measured via Cohen's (1988) f^2 using small (.02), medium (.15), and large (.35) cut-offs.

Results

This convenience sample was comprised of adolescents with VI aged 9 to 18 years old (N = 57, boys = 54%, girls = 46%). Descriptive information for the sample was as follows: M_{age} = 13.55 years (SD = 2.15), M_{height} = 1.56 m (SD = 0.13), M_{weight} = 54.35 kg (SD = 20.43), M_{BMI} = 21.85 kg/m^2 (SD = 5.98), and $M_{BMI\%}$ = 60.91 (SD = 32.14). Multimorbidity (i.e., coexistence of at least two chronic conditions [Espinoza et al., 2018]) was self-reported by 39% of the sample. Disabilities occurring in tandem with VI were highly variable (e.g., developmental disabilities, orthopedic impairments, etc.). However, all conditions were recognized disabilities pursuant to federal law within the US (Individuals with Disabilities Education Act, 2004; 34 CFR § 300.8). Regarding race, 65%, 16%, 14%, and 5% were White, Black, Asian, and Other, respectively. Concerning visual classification, 17.5% were B4, 38.5% were B3, 21% were B2, and 23% were B1 per the IBSA/USABA visual classification scales. Further, 81% of the sample stated that their preferred foot was their right foot. This was ascertained by asking each participant "Which foot do you prefer to kick a ball with?".

It was determined that participants could self-select the direction that they would like to travel for the Dribble trials as pre-screening of the data showed that there was not a statistically significant difference between traveling to the right or left directions (Exact Wilcoxon signed-rank test with Pratt modification: p = .09). For the current sample, the fastest Dribble time of all trials was used for anlaysis.

The within-sample raw score statistics for the BFST-AVI and the within-sample raw-to-quartile point conversions for each item can be found in Table 1 and Table 2, respectively. Henceforth, only the quartile point scale will be examined (see Table 3).

Table 1. BFST-AVI raw score descriptive statistics.

Component	Mean	SD	Median	Skewness	Kurtosis
Dribble	20.89	10.8	17.7	1.41	1.97
Accuracy Pass	3.86	2.68	4	0.68	0.02
Shooting	4.91	3.15	5	0.39	-0.37
Continuous Pass	16.74	9.17	14	0.78	-0.29
Goalkeeping	13.88	9.51	13	0.93	1.58

Table 2. BFST-AVI raw score to quartile score conversion table.

Component	1 (≤25th)	2 (≤50th)	3 (≤75th)	4 (>75th)
Dribble	>25	≤25	≤18	≤14
Accuracy Pass	≤2	≤4	≤5	>5
Shooting	≤3	≤5	≤6	>6
Continuous Pass	≤10	≤14	≤22	>22
Goalkeeping	≤8	≤13	≤18	>18

Note. Working example using the dribble task: an individual's best (raw) dribble time is 17.99 seconds; they would subsequently receive a quartile score of 3 for the dribble task.

Table 3. BFST-AVI quartile score statistics.

Component	Mean	SD	Median	Skewness	Kurtosis	Difficulty	α (if dropped)	Item-Total Corr.*
Dribble	2.56	1.09	3	-0.07	-1.32	.64	.71	.79
Accuracy Pass	2.25	1.15	2	0.48	-1.25	.56	.75	.65
Shooting	2.32	1.20	2	0.25	-1.51	.58	.80	.48
Continuous Pass	2.40	1.16	2	0.13	-1.48	.60	.80	.48
Goalkeeping	2.44	1.12	2	0.08	-1.39	.61	.77	.59

* = correlation without item in total

The median and average total quartile score for the sample were 11.00 and 11.96 (*SD* = 4.29) points, respectively. Total quartile scores were not found to be normally distributed according to a Shapiro-Wilk analysis (*p* = .010; skewness = .28; kurtosis = -1.01). No total quartiles scores were identified as outliers using the 'fence' method (±1.5 * interquartile range; interquartile range = type 7).

Zero-order Spearman correlations for the individual items ranged from .23 to .65. One correlation was below .30 whereas no correlations were above .70. Nine of the 10 correlations were statistically significant (see Table 4). The only exception was the correlation between the Shooting and Continuous Pass skills. Overall, these results suggested that the individual item scores were monotonically related in adolescents with VI.

Table 4. Quartile score spearman correlation matrix.

Component	Dribble		Accuracy Pass		Shooting		Continuous Pass	
	rho	p	rho	p	rho	p	rho	p
Accuracy Pass	.65	<.001						
Shooting	.46	<.001	.49	<.001				
Continuous Pass	.54	<.001	.30	.02	.23	.09		
Goalkeeping	.64	<.001	.45	<.001	.30	.02	.40	.002

The Cronbach's alpha and McDonald's (ordinal) omega for all items were .81 (95% CI = .73, .89) and .85 (95% CI = .79, .91), respectively. The average bias-corrected inter-item correlation was .46 suggesting that the items were measuring a narrow construct. All item-total correlations, without item in total, were above the proposed .30 threshold with a spread = .48 to .79 (see Table 3). These outcomes suggest strong internal consistency for the assessment scores in adolescents with VI.

Item difficulty values for each item can be found in Table 3. All values ranged between .56 and .64 and therefore (a) fell within the .20 to .80 cut points and (b) were close to the previously calculated optimal ideal difficulty of .50. The average item difficulty was .60

which was relatively close to the .50 average typically sought by assessment developers. Therefore, all items were shown to have an appropriate level of difficulty in adolescents with VI.

Concerning the convergent validity measures, the average peak beep kickball velocity was 23.07 miles per hour (*SD* = 8.01; *Mdn* = 23; skewness = .06; kurtosis = .07). In metric, this equates to 37.13 kilometers per hour (*SD* = 12.89). The average TGMD-3 kick score was 4.28 points (*SD* = 2.22; *Mdn* = 4; skewness = .04; kurtosis = -1.03). Total scores were found to statistically, positively, and moderately associate with the kicking process (ρ = .55; 95% CI = .34, .71; $p < .001$) and product assessments (ρ = .69; 95% CI = .52, .80; $p < .001$). These results provide evidence of convergent validity for the total quartile scores in adolescents with VI.

Based on the known groups' analysis, degree of vision was shown to be predictive of total BFST-AVI scores in a positive and linear fashion. The following regression equation (p = .002) was;

$$y = 7.88 + 1.63x_1$$

Specifically, it can be stated that for every one unit increase in the dummy coded USABA classification scale (i.e., B1-B4 transposed as 1-4), predicted total BFST-AVI scores increased by 1.63 points when all other variables in the model are held constant. The adjusted R^2 was .14 suggesting that 14% of the variability for BFST-AVI total scores was explained by the linear relationship between the USABA visual classification level and BFST-AVI total scores in adolescents with VI. Further, using magnitude-based inference, the Cohen's f^2 was .16 suggesting that USABA degree of vision had a medium-sized effect on total BFST-AVI scores (Cohen, 1988). These outcomes indicate that, on average, increased BFST-AVI total score performance appears to be linearly associated with increased vison in adolescents with VI. As such, these results provide evidence of construct validity for the total quartile scores in adolescents with VI.

Discussion

The purpose of this study was to investigate the immediate measurement properties of scores from the BFST-AVI, a novel football assessment for amateur-skilled adolescents with VI in the US. It was hypothesized that scores would, in general, have sound measurement properties in adolescents with VI. This discussion will initially focus on the measurement results and then delineate the practical implications of this study.

Measurement Properties of the BFST-AVI

Based on, and limited to, the statistical analyses that were used within the current study, individual and total quartile scores from the BFST-AVI were deemed to have acceptable forms and levels of reliability and validity in the current population. However, one result, in particular, was not ideal. The Shooting and Continuous Passing tasks did not statistically correlate (ρ = .23, p = .09). This may have occurred because shooting and continuous passing may be independent football skills in amateur skilled adolescents with VI when being scored using the BFST-AVI. This conclusion would align with Henry's (1958) specificity hypothesis which would suggest that the BFST-AVI skills have individualized task requirements and developmental trajectories. As such, performance on one skill may not be highly transferable or directly associate with performance of another skill. However, it is reasonable to suggest that a statistically significant relationship of some magnitude should exist between these two skills. It is possible that (a) the BFST-AVI assessment did not optimally capture a relationship between these tasks or (b) that there may not be a monotonic relationship between these two skills in amateur-skilled adolescents with VI.

Practical Implications

Given the totality of the results, the BFST-AVI can be adopted by practitioners to assess football skill in amateur-level American adolescents with VI for talent identification, skill assessment, and related purposes. This outcome is significant as the BFST-AVI is the first vetted football assessment of its kind for adolescents with VI.

As such, there are multiple direct and indirect practical implications which could occur from this study. The most tangible outcome of this study is that the BFST-AVI can be used to evaluate football talent and skill development in adolescents with VI. Specifically, given that scores from the BFST-AVI have been found to have multiple forms and acceptable levels of reliability and validity in adolescents with VI, general and adapted physical educators may be more inclined to include football, blind or adapted, as a unit within their curriculum. As such, these practitioners could use the BFST-AVI for the measurement of psychomotor learning outcomes, for ability grouping during a football-based unit, or may want to include the BFST-AVI as an assessment within a student's individualized education plan. The BFST-AVI could also be used by coaches, researchers, and at sport camp-type settings which heavily rely on motor skill assessments of adolescents with VI (e.g., tracking performance before, during, or after a football intervention).

A domain in which the BFST-AVI may be increasingly relevant is talent identification. Currently, when considering blind football, there is a clear desire to (a) develop American national team(s), (b) to raise the participation and accessibility of blind football for individuals with VI, and (c) to strategically increase the profile of blind football in the US from a commercial standpoint (Oh et al., 2021). To achieve these goals, immediate talent identification is of utmost importance. Thus, adoption and use of the BFST-AVI could support these goals by enabling early-stage talent identification at the grassroot level in adolescents with VI which could promote a larger, more systemic, hierarchical talent stream. This point is further underscored by the current push to develop and field an US blind football team for the 2028 Paralympics. Specifically, it has been suggested that such a US blind football team should be established by 2023 and that national governing bodies should be the motive force behind these processes (Oh et al., 2021).

However, the foundation for such progress cannot exist without a sound, early-stage assessment like the BFST-AVI. Therefore, identification of emergent football talent at the grassroots levels in adolescents with VI (i.e., B1) is particularly timely. Yet, it is important to note that various obstacles remain in relation to coaching and the scouting of talent within blind football (Mycock & Molnár, 2020; Oh et al., 2021).

There are numerous indirect implications that could extend to adapted physical education and Paralympic sport communities from this study. For example, adoption of the BFST-AVI by key stakeholders could increase the awareness of blind football within adolescents with VI as well as the profile of adapted forms of football on a national scale. Given that adolescents with VI who participate in blind soccer may incur biopsychosocial (Gamonales et al., 2018; Lankhorst et al., 2019) and performance-related (Bergeron et al., 2015) benefits and that there is a considerable desire to raise the profile of blind football in the US (Oh et al., 2021), accessible, unstructured experiences with blind football may be an important strategy for initiating and maintaining participation rates within the sport in the future (Haegele, 2019). Likewise, if the societal profile of blind football is increased, it is probable that sports such as blind or partially sighted football could be used as a vessel to promote disability awareness in those without VI within the general physical education curriculum (Tindall, 2013; Tindall et al, 2013). Further, this study could inspire others to develop, vet, and disseminate other sport-specific assessments surrounding specific adapted and Paralympic sports or populations.

Another implication of the current study is prospective in nature and relates to the dissemination of the BFST-AVI. Currently, it is planned that the BFST-AVI will reach practitioners and professionals via the USABA website, social media platforms of related organizations and collaborators, listservs, conference and professional development presentations, and future publications concerning the BFST-AVI. Further, it is planned that a collection of training videos and related resources, which will be hosted and made publicly available online, will be developed to assist with the training and standardization of the BFST-AVI processes.

Limitations and Future Research

This study is not without limitations. A convenience sample was used due to the low incidence of children and adolescents with VI. This could lead to under-representation or over-representation of the groups within the sample. It should also be noted that while the sample size was smaller in absolute terms, in relative terms, the current sample was sizeable considering the population. Our sample was heterogenous as it included individuals with a variety of ages, sexes, and vision levels. However, heterogeneity does increase the external validity and generalizability of the results. Further, these scores are based upon amateur adolescents with VI ranging from B1 to B4 from the US. Therefore, the robustness of these quartile scores cannot be extrapolated to other populations.

It is also important to note that the tasks which comprised the BFST-AVI were inspired by blind football. The BFST-AVI is not representative of all potential or general football-related skills (e.g., heading, crossing, tackling, free kicks, throw-ins). Therefore, it is important to acknowledge that the BFST-AVI was developed with blind forms of football in mind, the skill level of the participants (i.e., amateur), and data collection factors such as the training of evaluators, assessment duration, and environmental considerations such as space and equipment.

Further, as the Continuous Pass and Goalkeeping tasks required an administrative ‘partner,’ inter-individual variability such as inconsistent ball passes or tosses may influence BFST-AVI performance. Future research should consider using processes and equipment that would standardize the whole assessment, such as a ball launcher for the Goalkeeping task. Viewing the BFST-AVI as a grassroots assessment, researchers may also consider developing cut-offs and normative values in amateur-skilled adolescents with VI in the future to aid in talent identification and assessment purposes. Likewise, investigations focusing on the predictive validity of BFST-AVI scores in relation to future Paralympic or elite-level team selections may prove useful.

Concerning the analytic approach used within the current study, not all forms of reliability or validity were investigated using BFST-AVI scores. Last, it is important to stress that measurement results are temporal and could change over time. Based on these limitations, future research should utilize larger sample sizes which could lead to more accurate parameter estimates. Measurement investigations using BFST-AVI scores could also be completed using other populations such as, elite athletes, younger or older individuals with VI, or individuals from other countries. Further, it is likely that a reproducibility study will be needed in the future as measurement properties should not be assumed to be constant. Likewise, investigations which explore additional forms of validity and reliability, nuances within each task such as preferred and non-preferred limb performance, the investigation of additional skills such as crossing, and studies exploring factors such as a biological sex are warranted.

Perspectives

The BFST-AVI, an assessment composed of five football-related tasks, and is the first football-based assessment created specifically for adolescents with VI. Individual and total scores from the BFST-AVI had multiple forms and acceptable levels of reliability and validity in American adolescents with VI. This is significant for practice as the BFST-AVI can immediately be adopted by educators, coaches, and talent scouts to assess football skill performance for talent identification or gross motor skill assessment in American adolescents with VI at the amateur level. Likewise, the BFST-AVI could have numerous indirect benefits such as helping increase the awareness, participation, or accessibility of blind football for adolescents with VI and increase the commercial profile of blind football. For additional details concerning the BFST-AVI protocols, the BFST-AVI manual can be found online manual (Gilbert & Pennell, 2019) or by contacting the corresponding author.

Author affiliations:

1 Natural Science Division, Pepperdine University, Malibu, CA; adam.pennell@pepperdine.edu

2 Department of Physical Education, State University of New York at Cortland, Cortland, NY; emily.gilbert@cortland.edu

3 Department of Physical Education, University of South Carolina, Columbia, SC, USA; abrian@sc.edu

4 Department of Kinesiology, Sports Studies, and Physical Education, State University of New York's at Brockport, Brockport, NY; pbeach@brockport.edu; llieberman@brockport.edu

* Correspondence: adam.pennell@pepperdine.edu; Tel.: +01-310-506-4162

Author Contributions: Conceptualization, A.P. and E.G.; Methodology, A.P. and E.G.; Formal Analysis, A.P.; Writing-Original Draft Preparation, A.P.; Writing-Review & Editing, A.P., E.G., A.B., P.B., and L.L.

Funding: This research was funded by the Gibney Family Foundation.

Acknowledgments: None.

Conflicts of Interest: The authors declare no conflict of interest.

References

Altman, D., & Bland, J. (1994). Statistics notes: Quartiles, quintiles, centiles, and other quantiles. *British Medical Journal*, *309*, 996. http://doi.org/10.1136/bmj.309.6960.996

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (2014*). Standards for educational and psychological testing*. American Educational Research Association.

Bennett, D. (2001). How can I deal with missing data in my study?. *Australian and New Zealand Journal of Public Health*, *25*(5), 464-469. http://doi.org/10.1111/j.1467-842X.2001.tb00294.x

Bennette, C., & Vickers, A. (2012). Against quantiles: Categorization of continuous variables in epidemiologic research, and its discontents. *BMC Medical Research Methodology*, *12*(1), 21. http://doi.org/10.1186/1471-2288-12-21

Bergeron, M., Mountjoy, M., Armstrong, N., Chia, M., Côté, J., Emery, C., Faigenbaum, A., Hall Jr., G. Kriemler, S., Léglise, M., Malina, R., Pensgaard, A., Sanchez, A., Soligard, T., Sundgot-Borgen, J., van Mechelen, W., Weissensteiner, J., & Engebretsen, L. (2015). International Olympic Committee consensus statement on youth athletic development. *British Journal of Sports Medicine*, *49*(13), 843-851. http://doi.org/10.1136/bjsports-2015-094962

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32. http://doi.org/10.1023/A:1010933404324

Brian, A., Taunton, S., Lieberman, L., Haibach-Beach, P., Foley, J., & Santarossa, S. (2018). Psychometric properties of the Test of Gross Motor Development-3 for children with visual impairments. *Adapted Physical Activity Quarterly*, *35*(2), 145-158. http://doi.org/10.1123/apaq.2017-0061

Carlson, K., & Herdman, A. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17-32. http://doi.org/10.1177/1094428110392383

Clark, L., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309-319. http://doi.org/10.1037/1040-3590.7.3.309
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
Cristóbal, E., Flavián , C., & Guinalíu, M. (2007). Perceived e-service quality (PeSQ): Measurement validation and effects on consumer satisfaction and web site loyalty. *Managing Service Quality: An International Journal*, *17*(3), 317-340. http://doi.org/10.1108/09604520710744326
Crutzen, R., & Peters, G. (2015). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, *11*(3), 242-247. http://doi.org/10.1080/17437199.2015.1124240
Dong, Y., & Peng, C. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222 http://doi.org/10.1186/2193-1801-2-222
Espinoza, S., Quiben, M., & Hazuda, H. (2018). Distinguishing comorbidity, disability, and frailty. *Current Geriatrics Reports*, *7*(4), 201-209. https://doi.org/10.1007/s13670-018-0254-0
Fédération Internationale de Football Association. (2007). *Big Count 2006: Statistical summary report by association*. https://resources.fifa.com/image/upload/big-count-summary-report-association-520044.pdf?cloudid=vrnjcgakvf7nds6sl5rx
Fédération Internationale de Football Association. (2018*). 2018 FIFA World Cup Russia: Global broadcast and audience summary*. https://resources.fifa.com/image/upload/njqsntrvdvqv8ho1dag5.pdf
Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. *Research in Nursing & Health*, *14*(2), 165-168. http://doi.org/10.1002/nur.4770140211
Fisher, R. (1958). *Statistical methods for research workers* (13th ed.).Oliver & Boyd.
Gamonales, J., Muñoz-Jiménez, J., León-Guzmán, K., & Ibáñez, S. (2018). 5-a-side football for individuals with visual impairments: A review of the literature. *European Journal of Adapted Physical Activity*, *11*(1). http://doi.org/10.5507/euj.2018.004
Gilbert, E., & Pennell, A. (2019). 5-a-side soccer manual for the blind and visually impaired. https://www.usaba.org/wp-content/uploads/5-a-side-soccer-Manual-2019.pdf
Haegele, J. A. (2018). Youth leisure-time physical activity from the perspectives of young adults with visual impairments. *European Journal of Adapted Physical Activity*, *11*(2). http://doi.org/10.5507/euj.2018.010
Haibach, P., Wagner, M., & Lieberman, L. (2014). Determinants of gross motor skill performance in children with visual impairments. *Research in Developmental Disabilities*, *35*(10), 2577-2584. http://doi.org/10.1016/j.ridd.2014.05.030
Henry, F. (1958). Specificity versus generality in learning motor skill. *Annual Proceedings, College Physical Education Association*, *61*, 126-128.
Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Houghton Mifflin.
Individuals with Disabilities Education Act, 20 U.S.C. § 1400. (2004).
International Blind Sports Association. (2017a). *Football five-a-side laws 2017-2021: B1 category*. https://blindfootball.sport/wp-content/uploads/2021/02/960-Rules-IBSA-Blind-Football-B1-Rulebook-2017-2021.pdf
International Blind Sports Association. (2017b). *Football five-a-side laws 2017-2021: B2-B3 category*. https://blindfootball.sport/wp-content/uploads/2021/02/622-Rules-IBSA-Partially-Sighted-Football-B2-B3-Rulebook-2017-2021.pdf
International Blind Sports Association. (2021, April 15).Overview. https://blindfootball.sport/about-football/overview/
Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402-406. http://doi.org/10.4097/kjae.2013.64.5.402
Kim, S., & Frisby, C. (2019). Gaining from discretization of continuous data: The correspondence analysis biplot approach. *Behavior Research Methods*, *51*(2), 589-601. http://doi.org/10.3758/s13428-018-1161-1

Lankhorst, K., Takken, T., Zwinkels, M., van Gaalen, L., Velde, S., Backx, F., Verschuren, O., Wittink, H., & de Groot, J. (2019). Sports participation, physical activity, and health-related fitness in youth with chronic diseases or physical disabilities: The health in adapted youth sports study. *Journal of Strength and Conditioning Research*, *35*(8), 2327-2337. http://doi.org/10.1519/JSC.0000000000003098
Lieberman, L., Ponchillia, P., & Ponchillia, S. (2013). Physical education and sports for people with visual impairments and deafblindness: Foundations of instruction. AFB Press.
McIntire, S., & Miller, L. (2007). *Foundations of psychological testing: A practical approach* (2nd ed.). Sage Publications.
Mycock, D., & Molnár, G. (2020). 'The blind leading the blind'- A reflection on coaching blind football. *European Journal of Adapted Physical Activity*, *14*(3). http://doi.org/10.5507/euj.2020.011
Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., & Ishii, S. (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, *19*(16), 2088-2096. http://doi.org/10.1093/bioinformatics/btg287
Oh, Y., Arthur-Banning, S., & Domka, M. (2021). SWOT analysis on the potential growth of Football 5-a-side programme across the United States: An exploratory case study approach for athletes with visual impairment. *Sport in Society*, 24(9), 1683-1697. http://doi.org/10.1080/17430437.2020.1768242
Plonsky, L., & Oswald, F. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, *39*(3), 579-592. http://doi.org/https://doi.org/10.1017/S0272263116000231
Portney, L.(2020). *Foundations of clinical research: Applications to evidence-based practice* (4th ed.). F.A. Davis Co.
Rubin, D. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592. http://doi.org/10.1093/biomet/63.3.581
Schwab, D. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior*, *2*, 3-43.
Skaggs, S., & Hopper, C. (1996). Individuals with visual impairments: A review of psychomotor behavior. *Adapted Physical Activity Quarterly*, *13*(1), 16-26.
Stekhoven, D., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112-118. http://doi.org/10.1093/bioinformatics/btr597
Stodden, D., Gao, Z., Goodway, J., & Langendorfer, S. (2014). Dynamic relationships between motor skill competence and health-related fitness in youth. *Pediatric Exercise Science*, 26(3), 231-241. https://doi.org/10.1123/pes.2013-0027
Tabachnick B., & Fidell, L. (2012) *Using multivariate statistics* (6th ed.). Pearson.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55. http://doi.org/10.5116/ijme.4dfb.8dfd
Templ, M., Kowarik, A., Alfons, A., de Cillia, G., Prantner, B., & Rannetbauer, W. (2021). *VIM: Visualization and imputation of missing values*. https://cran.r-project.org/web/packages/VIM/index.html
Tindall, D. (2013). Creating disability awareness through sport: Exploring the participation, attitudes and perceptions of post-primary female students in Ireland. *Irish Educational Studies*, *32*(4), 457-475. https://doi.org/10.1080/03323315.2013.859339
Tindall, D., McMahon, K., Willson, J., & Foley, J. (2013). Disability awareness through futsal. *Physical Education Matters*, *8*(2), 26-29.
Ulrich, D. (2019). *The Test of Gross Motor Development* (3rd ed.). PRO-ED, Inc.
United States Association of Blind Athletes. (2017, December 20). *Visual classifications*. https://www.usaba.org/membership/visual-classifications/
United States Association of Blind Athletes. (2021, September 28). *USA blind soccer*. https://www.usaba.org/sports/blind-soccer/