

Factor Structure of the Women's Heptathlon: Applications of Traditional Factor Analysis and Structural Equation Modelling

Ian Tim Heazlewood

The women's junior and senior Heptathlon in the sport of Athletics is an event that is conducted over two consecutive days using seven event performances in the following order; day 1 consists of the 100m hurdles, high jump, shot put and 200m; and day 2 consists of the long jump, javelin throw and 800m. Training recommendations are based on conceptual models and suggested motor fitness constructs between the different events within the Heptathlon. In this model, sprint events such as the 100m hurdles and 200m; jump events such as long jump and high jump; throws such as shot put and javelin and the 800m should represent four underpinning factors that represent the seven the events. The predominant research issue was do two statistical multivariate analytical approaches, one based on traditional factor analysis and the second based on structural equation modelling (SEM) support the four factor conceptual model? The statistical method known as confirmatory factor analysis was applied to a set of IAAF top 173 heptathlon scores in 2010 to uncover the latent structure or factors of the seven events. The traditional factor analysis, principal component analysis, generated the simplest and interpretable factor structure that supported the hypothesised four factor model. The SEM approach produced more paradoxical results. The results suggest training modules can be developed based on the factor structure of the Heptathlon.

Introduction

The women's junior and senior Heptathlon in the sport of Athletics is an event that is conducted over two consecutive days using seven event performances in the following order; day 1, consists of 100m hurdles, high jump, shot put and 200m; and day 2, consists of long jump, javelin throw and 800m. The individual event performances are converted to point scores using International Association of Athletic Federations (IAAF) scoring tables, which are then summed to assess rank performance (IAAF, 2009). Literature on training for the Heptathlon in terms of evidence based training is minimal and limited recommendations that exist for adult athletes are based on case study research (Daly, 1984), extrapolating from how the high performance athlete trains (Daly, 1984), training recommendations that are based on conceptual models (Hancock, 1987; Mackenzie, 2007; Marra, 1985; Sarponov, 1982; Telfer, 1988) and

suggested exercise physiological links between the different events within the Heptathlon (Mackenzie, 2007), based on motor fitness constructs of endurance, strength, speed, skill mobility and so on. In the conceptual-exercise physiological model, sprint-power events, such as the 100m hurdles, 200m, long jump and high jump should be correlated and some training transfer should occur between these events. That is, sprint training should assist long jump performance. In terms of an evidenced based approach Heazlewood (2008) evaluated the factor structure of the Heptathlon based on the top 100 heptathletes from the 2006 IAAF athletic year and derived a three factor model to partially explain the interrelationships or factor structure between the seven events. In this model factor 1 displayed significant factor loadings with 110m hurdles, 200m, high jump and long jump; factor 2 with shot put and javelin and factor 3 with 800m. Some factor complexity was associated with shot put, 200m and 800m, which loaded across more than one factor. However, whether or not this factor structure is invariant across different

Ian Tim Heazlewood the Department of Exercise and Sports Science, Charles Darwin University, Australia.

competitive seasons has not been established. The conceptual-exercise physiological model of Mackenzie (2007) is represented in table 1, represents a more complex explanation on which motor fitness factors are associated with

each event in terms of high, medium or low dependence, although no explanation is provided as to how the weightings of high, medium and low were derived or what exact quantities they represent.

Table 1. Hypothesised Factor Structure Based on Motor Fitness Factors Contributing to Each Event (Mackenzie, 2007).

Event	Aerobic Endurance	Gross Strength	Skill	Relative Strength	Running Speed	Mobility	Explosive Strength	Speed Endurance	Strength Endurance
100m hurdles	-	Med	High	High	High	High	High	Med	-
High jump	-	Low	High	High	High	High	High	-	-
Shot put	-	High	High	Med	Low	Med	High	-	-
200m	Low	Med	Med	High	High	High	High	High	High
Long jump	-	Low	High	High	High	High	High	-	-
Javelin	-	Med	High	High	Low	High	High	-	-
800m	High	-	Low	Low	Med	Low	-	-	High

Explained in greater detail based on motor fitness constructs or factors, the 100m hurdles is conceptualised to depend highly on skill, relative strength, running speed, mobility, explosive strength, and to a medium degree on gross strength and speed endurance; the high jump to depend highly on skill, relative strength, running speed, mobility, explosive strength, and to a low degree gross strength; whereas the 200m sprint will depend highly on relative strength, running speed, mobility, explosive strength, speed endurance and strength endurance to a medium degree on gross strength and skill and to a low degree on gross strength. The other four events in table 1, that is, shot put, long jump, javelin and 800m are also described in terms of factors contributing to successful performance in these events. This model has some merit; however relative contribution to the underlying factors to the events has not actually been quantified, so the actual statistical weights for the factors have not been identified.

The statistical methods to indentify the interrelationships between variables can be solved at the bivariate level such as bivariate correlation or at a more complex explanatory level using traditional factor analysis methods or “moving beyond the basic techniques” by applying the multivariate technique of structural equation modelling (abbreviated as SEM) (Hair et al., 2006, p. 705). Both multivariate

approaches can be applied to evaluating the interrelationships between the seven heptathlon events and then assessed for their ability to provide factor simplicity and interpretability. In the context of this research confirmatory factor analysis will be applied using both methods. Confirmatory factor analysis attempts to confirm an underlying theoretical structure of a relatively large set of possible explanatory variables. The researcher's a priori assumption is that explanatory variables may be associated with specific factors. This is the most common form of factor analysis. There is a prior theory and the researcher uses factor loadings to explore and identify the factor structure of the data (Kim & Mueller; 1978; Norusis, 1985, Hair et al.; 1998; Hair et al.; 2006; Costello & Osborne, 2005). The advantage of confirmatory factor analysis is that it allows for testing hypotheses about a particular factor structure. Confirmatory factor analysis is theory based unlike exploratory factor analysis and confirmatory factor analysis is a complex procedure with some absolute guidelines and many options (Costello & Osborne, 2005).

The more specific objectives of factor analysis (Kim & Mueller, 1978; Norusis, 1985; Hair et al., 2006) are to.

1. To identify underlying constructs or factors which explain the correlations among a set of variables?
2. To test hypotheses about the structure of the variables.
3. To summarise a large number of variables with a smaller number of derived latent variables or factors.
4. To determine the number of dimensions to represent a number of variables and the correlation matrix.
5. To achieve the simplest, most interpretable and pragmatically more meaningful factor solution.

There are a number of methods for factor analysis and these are principal component analysis, maximum likelihood, image, principal axis, generalised least squares, unweighted least squares and alpha factoring . The statistical approach of each method is slightly different (Kim & Mueller; 1978; Norusis, 1985; Hair et al., 2006; Garson, 2011a). For example, principal component analysis transforms a set of observed variables into another reduced set of latent variables or factors; maximum likelihood is to solve the factor solution that best fits the observed correlations; image factoring distinguishes between the common variance part of a variable and the unique variance part; principal axis method attempts to reproduce the correlations of the variables and accounts for the covariation among the variables. ; generalised least squares is to minimise the residual correlations after extracting a specific number of factors and to test the fit between the reproduced correlations from the model to observed correlations; and in alpha factoring the variables included in the “analysis are considered a sample from a universe of variables” (Kim & Mueller, 1978. P. 26).

Confirmatory factor analysis is a special case of the structural equation model (SEM), also known as the covariance structure and with confirmatory factor analysis it is possible to place substantively meaningful constraints using SEM on the factor model, such as setting the effect of one latent variable to equal zero on a subset of the observed variables. SEM consists of two components; a measurement model linking a set of observed variables to a usually

smaller set of latent variables and a structural model linking the latent variables through a series of recursive and nonrecursive relationships. Confirmatory factor analysis corresponds to the measurement model of SEM and as such is estimated using SEM software, such as Amos (Albright, 2006; Arbuckle, 2006, 2009).

Research Questions

The predominant research question is do different statistical analytical approaches support the conceptual model where four underlying factors represent the interrelationships between the seven events based on factors of sprint, jumps, throws and endurance abilities or factors? In the context of training, will heptathletes expect some training and performance transfer between events or do the seven events represent unique constructs that have to be trained individually? That is, do a number of events really represent a smaller number of factors that may be utilised to develop training approaches based on events that load on specific factors or are events unique and must be trained in isolation? If significant correlations exist among the seven individual events, factor analysis and structural equation modelling can be utilised to reduce the larger variable set, the seven events, into a smaller number of latent non-measured variables or factors. In terms of a conceptual model a four factor model can be hypothesised where 100m hurdles and 200m should be dependent on sprint speed, jump events dependent on “explosive” power, throwing events dependent on strength-power and the 800m dependent on endurance.

Hypotheses

The two predominant hypotheses are:

1. If the interrelationships of the seven events in the heptathlon are invariant across different athletic seasons then the three factor model of Heazlewood (2008) should be reproduced.
2. If the seven events load with the constructs of speed, explosive-power, strength-power and endurance then a four factor model should be derived. This model is presented in

figure 1. The path from the endurance factor to 200m as well as to 800m is based on some previous research (Heazlewood, 2008) that correlations exist between 200m and 800m for the 2006 Heptathletes. This argument also applies to the path from the sprint factor to 800m.

3. If the factor structure represents true interrelationships between the seven individual events then the different factor analysis methods should provide reasonably consistent factor solutions that are simple and interpretable. The

theoretical model is illustrated in figure one using a structural equation model where the four underpinning factors are endurance, sprint, jumps and throws ability. In the hypothesised model the observed or endogenous variables, seven events, are in rectangles and the unobserved, latent, factors or exogenous variables of endurance, sprint, jumps and throws are within ellipses. The double headed arrow is a method to examine the correlations between the different factors and to assess statistical independence (orthogonal or oblique factors).

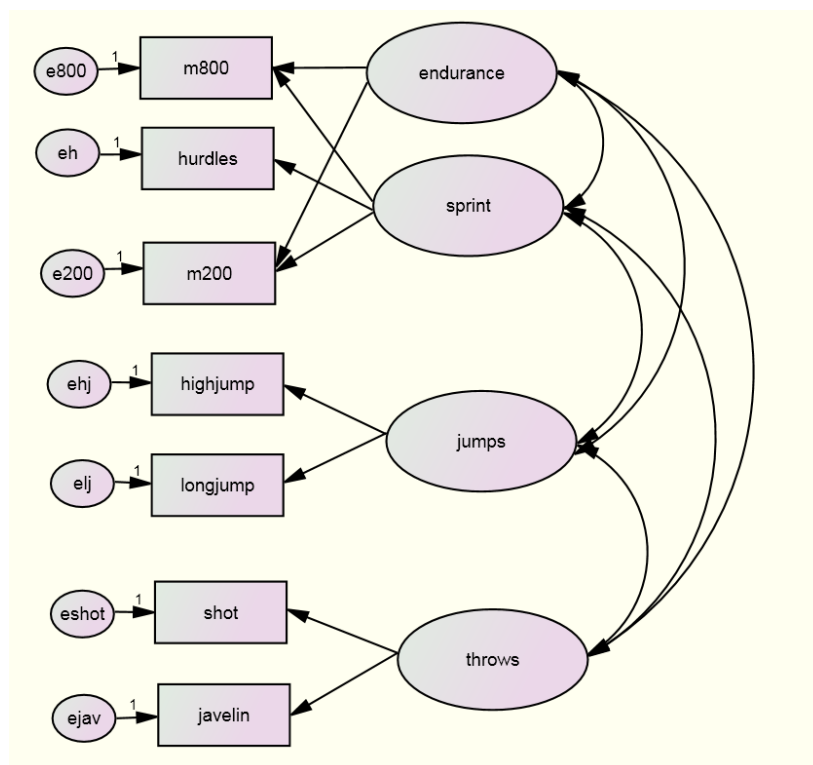


Figure 1. Hypothesised model of the interrelationships of the seven events as they related or are loaded with underpinning conceptual factors of sprint, jump, throw and endurance ability. Sprint to 800m and endurance to 200m path coefficients are predicted to be causally weak.

4. The factor structure if simple, interpretable and parsimonious can provide evidence based approaches to designing overall training programs and individual training sessions to optimise heptathlon training as

Methods

The statistical method known as confirmatory factor analysis was applied to a set of IAAF data (IAAF, 2010) based on world athletic rankings for top 173 women’s Heptathlon scores in the 2010 athletic season. The research aim was to confirm the structure or three factors of the seven events that were established by

the underpinning factors will represent specific event competition clusters and clusters on which training units can be developed.

Heazlewood (2008) or the four factor model hypothesised in this research.

The original scores for the seven events were input into an SPSS data file (PASW Statistics 18.0, 2007) using distances and heights for field events to the nearest 0.01 metre and time to the nearest 0.01 second for track events. The performances determined by the IAAF with the

2010 performances to be significantly wind-assisted were excluded from the analysis. A number of traditional statistical factor solution methods were applied and these were principal component analysis, maximum likelihood and principal axis factoring. Two factor solutions were derived, which were to select significant factors that displayed eigenvalues greater than 1 and to test a researcher defined four factor model based on the hypothesised conceptual factor structure of sprint speed, jumps explosive-power, throws strength-power and endurance.

As well, structural equation modelling an alternative approach to evaluating factor structure was applied using Amos 18 (SPSS Inc, 2007) software. Amos is short for Analysis of MOment Structures. It implements the general approach to data analysis known as structural equation modelling (SEM), also known as analysis of covariance structures, or causal modelling. Structural equation modelling represents a combination of factor analysis and multiple linear regression analysis was applied, which is based on reproducing the covariance matrix based on the seven events (Arbuckle, 2009), whereas the input matrix for the principal component analysis is the correlation matrix. The factor analytic method that provided the simplest, most interpretable and pragmatically meaningful factor solution was principal component analysis. Principal components analysis is a method of factoring a correlation matrix directly, without estimating communalities. Linear combinations of variables are estimated, which explain the maximum amount of variance in the variables. The components from principal components analysis reflect both common and unique variance of the variables and may be seen as a variance focused approach seeking to reproduce both the total variable variance with all components and to reproduce the correlations. Principal components analysis is far more common than principal factor analysis, however, and it is common to use "factors" interchangeably with "components."

In principal factor analysis the first component accounts for the most variance in the variables,

then the second component accounts for the most variance in the variables residualized for the first component, and so on. The method transforms a collection of measured variables into a set of orthogonal maximum variance linear combinations. Orthogonal in this context means the principal components or factors are not correlated with one another. Other factor solutions were considered and attempted, such as the maximum likelihood, alpha factoring, generalised least squares and principal axis factoring. A number of initial tests were applied to assess the structure and quality of the correlation matrix to be used in the factor analysis. Kaiser-Meyer-Olkin measure of sampling adequacy was applied and indicated that factor analysis was appropriate (value=0.587) and Bartlett's Test of Sphericity (approximate chi square = 93.5, df = 21, $p < .001$) was applied to test the hypothesis that the correlation matrix was not an identity matrix. The factor analysis derived the descriptive statistics, the correlation matrix, communalities, total variance explained, component matrix, a scree plot, rotated component matrix, component transformation matrix, and component plot in rotated space.

There are different but common assessment criteria for determining the number of significant factors that have been identified. Specifically, (Kim & Mueller; 1978; Norusis, 1985; Hair et al., 2006):

1. The Kaiser rule to exclude all components or factors with eigenvalues less than one.
2. The Cattell scree test, which plots the components as the X axis and the corresponding eigenvalues as the Y axis, is another test. As graph moves to the right, toward the less significant components/factors, the eigenvalues decrease. When the decrease ceases the curve makes a deflection or scree toward a less steep decline and Cattell's scree test excludes all further components after the one starting the deflection.
3. Variance explained criteria where you use the rule of keeping enough factors to account for

90% (sometimes 70% plus) of the variation. Where the research goal emphasizes parsimony (explaining variance with as few factors as possible) but notes the criterion can be as low as 50%.

4. The Joliffe criterion is a more liberal, which may result in twice as many factors as the Kaiser criterion as it excludes all components with eigenvalues under 0.7.

5. With a sample size of 150 plus significant factor loadings should be .45 or higher (Hair et al., 2006).

6. Although not a strictly mathematical criterion, construct comprehensibility is an important criterion, for limiting the number of factors to those whose dimension of meaning is readily comprehensible or interpretable. This

approach often identifies the first two or three components/factors as theoretically important.

The path diagram that is illustrated in figure 1 follows the normal convention for constructing the path diagram by which the SEM model solves or attempts to fit the model. This is an application of confirmatory factor analysis based on SEM principals in which there is unmeasured covariance between each possible pair of latent variables. In reference to figure 1, the straight arrows are from the latent variables to their respective indicators or actual events, as well as straight arrows from the error and disturbance terms to their respective variables or events. To re-emphasise, the double headed arrows connecting the four factors enables the computation of the bivariate correlations between these factors and indicates whether these are orthogonal or oblique (related) factors.

Results

The descriptive statistics for the sample based on performances in the total points and for the seven individual events. The range, minimum

score, maximum score, mean, and standard deviation are displayed in table 2.

Table 2. Range, Minimum and Maximum Scores, Means and Standard Deviations of

Total Points and for each Event.

Event	Range	Minimum	Maximum	Mean	Std. Deviation
Total Points	1312	5511	6823	5903	262.91
100m Hurdles(s)	1.98	12.89	14.87	14.00	.37
High jump(m)	.37	1.54	1.91	1.74	.07
Shot Put(m)	6.88	9.00	15.88	12.68	1.15
200m(s)	12.69	23.21	35.90	24.98	1.04
Long jump(m)	1.57	5.22	6.79	6.03	.25
Javelin(m)	27.30	25.86	53.16	41.61	5.37
800m(s)	30.01	126.43	156.44	137.73	5.13

The correlation matrix, which is illustrated in table 3, a conceptual and statistical starting point in factor analysis, indicated very significant absolute correlations between events.

Table 3. Correlation Matrix Indicating Pearson Correlations and Level of Significance for Each Event.

Event	100m Hurdles	High Jump	Shot Put	200m	Long jump	Javelin	800m
100m hurdles	1	-.080	-.127	.386**	-.223**	-.068	.159*
High jump	-.080	1	.109	-.097	.332**	-.023	-.046
Shot put	-.127	.109	1	.005	.101	.267**	-.102
200m	.386**	-.097	.005	1	-.255**	.099	.194*
Long jump	-.223**	.332**	.101	-.255**	1	.038	-.061
Javelin	-.068	-.023	.267**	.099	.038	1	.029
800m	.159*	-.046	-.102	.194*	-.061	.029	1

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Specifically, the 100m hurdles and 200m ($r = .386$; $p < .01$), 100m hurdles and long jump ($r = -.223$; $p < .01$), 100m hurdles and 800m ($r = .159$; $p < .05$), high jump and long jump ($r = .332$; $p < .01$), shot put and javelin ($r = .267$; $p < .01$), 200m and 800m ($r = .194$; $p < .05$) and 200m and long jump ($r = -.255$; $p < .01$), and clustering of events suggesting the appropriateness of applying factor analysis.

The resulting initial factor solution utilising principal component analysis based on a

researcher conceptually defined four factor model, explained 73.064% of the total variance explained. It is important to note that the fourth derived factor has an eigenvalues slightly less than one, the usual cut-off point in determining significant factors. A number of orthogonal factor rotations were applied to improve the factor loadings and the most interpretable factor solution was a varimax rotation with Kaiser normalisation. Both the initial and rotated factor solutions with total variance explained are presented in table 4.

Table 4. Total Variance Explained by Four Significant Factors Model. Eigenvalues

Greater than One were Selected as Significant.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.818	25.968	25.968	1.818	25.968	25.968	1.455	20.789	20.789
2	1.287	18.389	44.357	1.287	18.389	44.357	1.328	18.968	39.757
3	1.102	15.740	60.097	1.102	15.740	60.097	1.301	18.583	58.340
4	.908	12.967	73.064	.908	12.967	73.064	1.031	14.724	73.064
5	.698	9.968	83.032						
6	.621	8.874	91.906						
7	.567	8.094	100.000						

a. Extraction Method: Principal Component Analysis.

b. Four factors or components derived.

The initial component matrix is displayed in table 5 and the more important and interpretable varimax rotated component matrix in table 6.

Factor loadings of .45 or greater are identified as significant based on a sample size of 150 plus cases (Hair et al., 2006; Norusis, 1985) and

factor loadings in to .90s are considered marvellous, .80s meritorious and .70s moderate according to (Kim & Muller, 1978). The focus of the results will be on table 6, the varimax rotated component matrix. The varimax method attempts to minimise the number of variables that have high loadings on a factor and simplifying the columns in the factor matrix and is generally considered the best method in a achieving a simplified factor structure. This four factor solution loaded 110m hurdles (.818), 200m (.793) with factor 1; high jump (.876) and long jump (.711) on factor 2; the shot put (.750)

and javelin (.818) on factor 3; and the 800m (.940) on factor 4. The positive and negative loadings on indicate directional concepts in magnitude change of the variables with each factor. What is important to emphasise is essentially no factor complexity occurred, making the final rotated solution simple, interpretable and parsimonious and suggesting strongly that the interrelationships between the seven events can be accurately represented by a four factor model. The factors in this model are orthogonal, which means the factors are not correlated ($r=0$) with each other.

Table 5. The initial component matrix based on four factor model.

Event	Component			
	1	2	3	4
100m hurdles	.673	.051	.341	.364
High jump	-.474	.061	.686	.284
Shot put	-.322	.707	-.116	.217
200m	.662	.388	.234	.228
Long jump	-.657	.017	.452	-.114
Javelin	-.084	.787	-.159	-.252
800m	.400	.107	.466	-.720

- a. Extraction Method: Principal Component Analysis.
- b. Four components or factors extracted.

The varimax rotation approach attempts to minimise the number of variables that have a high loading on a factor, which should enhance the interpretability of the factor solution. Table

6 indicates this approach has been very successful as the variable loadings across the four orthogonal factors display distinct factor simplicity.

Table 6. The final varimax rotated component matrix as four factor model.

Event	Component			
	1	2	3	4
100m hurdles	.818	-.037	-.180	.047
High jump	.072	.876	.003	-.085
Shot put	.015	.173	.750	-.266
200m	.793	-.125	.175	.144
Long jump	-.359	.711	.071	.092
Javelin	-.034	-.088	.818	.192
800m	.151	.001	-.021	.940

- a. Extraction Method: Principal Component Analysis.
- b. Rotation Method: Varimax with Kaiser Normalization.
- c. Rotation converged in 5 iterations.

The structural equation models (SEM) are more complex in the analysis the model that is presented in figure cannot be solved due to unidentified of a model. Identification is essentially whether or not enough information exists to generate a solution in SEM for a set of structural equations. The identification problem

results in an inability of the hypothesised factor structure or model to generate unique estimates, which can interfere with the SEM program producing solutions (Hair et al., 2006). A model is under identified if there are more parameters to be estimated than there are elements in the covariance matrix. The

mathematical properties of under identified models prevent a unique solution to the parameter estimates and prevent goodness of fit tests on the model (Garson, 2011b). As a consequence the path from the latent variable or factor endurance to the 200m and the path from sprint to 800m was included to “solve” the SEM, resulting in a model in figure 2. The standardised regression weight loadings in figure 2 are not as substantive as the factor loadings in table 6 nor do the loadings enable a simple and interpretable factor structure.

The model fit based on goodness of fit indices or GOF (Hair et al., 2006; Garson, 2011b) indicated minimum was achieved, Chi-square = 4.96, degrees of freedom = 8 and probability

level = .76 (number of distinct sample moments 28; number of distinct parameters to be estimated 20 and degrees of freedom $28 - 20 = 8$). This indicates a reasonable fit of the model to the data or the structural model with the measurement model as they are statistically different (null hypothesis supported), however it is not the only possible fit as the path diagram in figure 2 is an outcome of conceptual modelling based on arguments presented in the introduction. Other fit indices are presented in table 7. The goodness-of-fit-index (GFI) is a fit statistic is less sensitive to sample size issues such as Chi-square and values of .90 to .95 are considered to indicate good fit and it should be noted in the default model the GFI is .992, once again indicating good model fit.

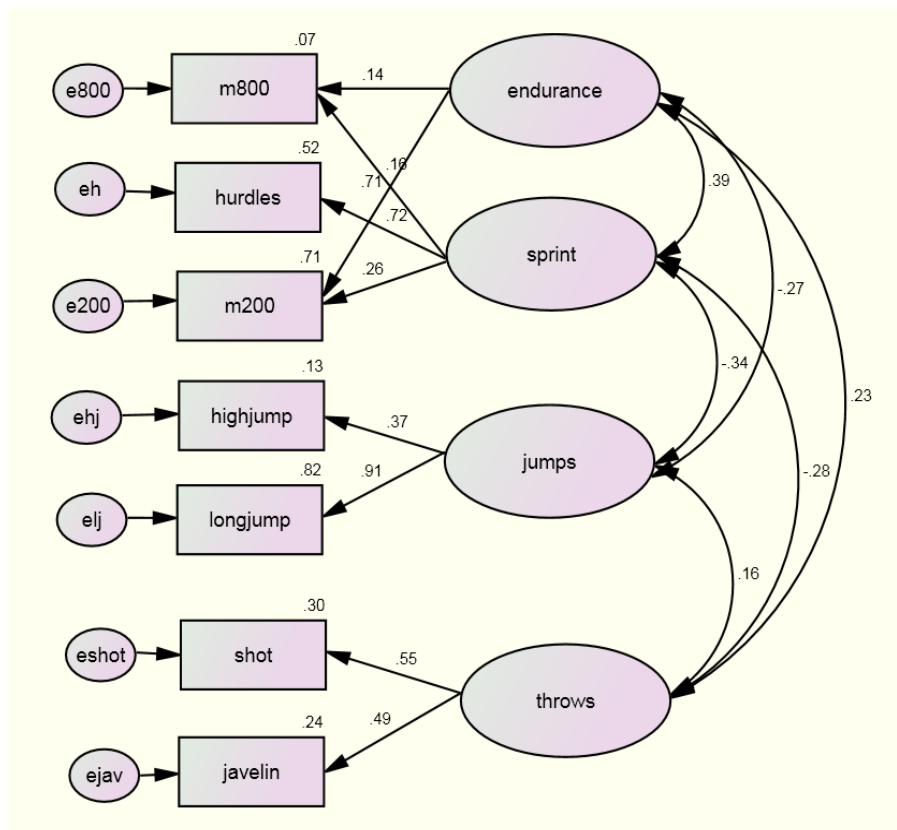


Figure 2. Evidenced based model of the interrelationships of the seven events as they related or are loaded with underpinning conceptual factors of sprint, jump, throw and endurance ability.

The root mean square (RMR) of .223, is higher than the usual cut-off value of .08 and to a degree supports the model fitting the data. Table 8 contains the normed fit index (NFI), Tucker

Lewis index (TLI) comparative fit index (CFI) and values of .90 or above are considered to indicate good fit.

Table 7. Root Mean Square (RMR) and

Goodness-of-Fit-Index (GFI) Values.

Model	RMR	GFI	AGFI	PGFI
Default model	.223	.992	.972	.283
Saturated model	.000	1.000		
Independence model	.435	.854	.806	.641

Table 8. Baseline Comparisons NFI, TLI and CFI.

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.948	.863	1.035	1.107	1.000
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Table 9 indicates the root mean square error of approximation (RMSEA), which involves a correction for large sample sizes and values less than .10 are thought to represent good model fit.

Table 9. Root Mean Square Error of Approximation (RMSEA).

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.062	.911
Independence model	.143	.115	.173	.000

Collectively, all these indices of goodness-of-fit measures indicate good and acceptable model fit with the 2010 IAAF heptathlon data for SEM; however the regression weights for the seven events with the four factors do not display the factor simplicity that was evident when applying “more traditional” confirmatory factor analysis utilising principal component analysis. For example the regression weights are throws

to shot is .56, throws to javelin .49; jumps to high jump .37 and jumps to long jump .91; sprints to 100m hurdles .72, sprints to 200m .26 and sprints to 800m as expected a low; and endurance to 800m .14 and an unexpected value for endurance to 200m .71. The maximum likelihood solution applied in SEM did not simplify or enhance interpretability when compared to principal component analysis.

Discussion

Comparing the two statistical approaches resulted in the more traditional multivariate factor analysis providing a simpler, more interpretable and parsimonious factor solution that confirmed the hypothesised four factor theoretical model that enables more practical applications to designing training modules based on clustering events based on the clear factor loadings using principal component analysis. That is 100m hurdles with 200m, high jump with long jump, shot put with javelin and 800m or endurance as a stand alone training module. However, the factor structure based on the IAAF 2010 data was significantly different from the factor structure based on the IAAF

2006 findings, which suggest the factor structure invariance over time may not exist.

The SEM factor approach did not display the simple factor structure in terms of regression weights as did principal component analysis and paradoxical effects were especially noted where the endurance to 800m was causally low and the endurance to 200m value much larger expectantly. The correlations among the factors were assessed to be non-significant suggesting that theoretically they represent more discrete or unique factors and real underpinning abilities on which Heptathlon performance is dependent. The next step is to redraw a number of possible paths, however one statistical issue is the factor of endurance requires more than one measured variable path as to the 800m as the simultaneous

equation solutions of SEM have to be mathematically solvable.

These findings suggest that training can be clustered into similar conceptual events to enhance transfer based on different energy systems and strength-power-speed requirements that underpin each event in the Heptathlon. In this context, the training can be clustered into four predominant training modules based on the sprints-hurdles (speed-power training), jumps (power-plyometric) training, shot put and javelin throws (strength-power training) and finally an isolated training session for the only endurance event the 800m. It must be emphasised the initial principal component solution, which derived the simplest and most interpretable theoretical model had four derived factors as having zero correlations. This implies that the training models are representing unique or stand alone constructs, which must be trained as discrete and unique factors. As the 800m or endurance event appears as essentially a stand alone event based on the exploratory factor analysis and represents only 1/7 of the Heptathlon. The consequence of this finding suggests training for this event probably only requires only one seventh of the total allocated training time be donated to improving 800m scores as its impact is not as significant in point achieved in the Heptathlon.

Conclusion

The traditional factor analysis in this research provided a more substantive theoretical model of the seven events in the Heptathlon and confirmed the hypothesised factor structure based on motor fitness constructs of speed, power, strength and endurance underpinning each event, when compared to structural equation models of the Heptathlon factor structure using the IAAF 2010 rankings. These findings suggest methods of Heptathlon training into discrete training modules based on the four factor structure. Finally, the factor structures displayed in 2006 and 2010 indicate the correlations or interrelationships between the seven events are changing over time.

Recommendations

The IAAF data 2006 to 2010 inclusive, should be pooled and the identical approach applied to assess the factor structure from a larger data set as, well as to compare the factor structure for 2006, 2007, 2008, 2009 and 2010 years using the SEM statistical multigroup approach to compare the structure across different competitive seasons.

References

- Albright, J. J. (2006). *Confirmatory factor analysis using Amos, Lisrel and Mplus*. The Trustees of Indiana University: Indiana University. Retrieved 9/1/2006 from <http://www.indiana.edu/~statmath>
- Arbuckle, J. L. (2006). *Amos 7.0: User's guide*. Chicago: SPSS Inc.
- Arbuckle, J. L. (2009). *Amos 18: User's guide*. Chicago: SPSS Inc.
- Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. [Electronic version]. *Practical Assessment, Research & Evaluation*, 10 (7): 1-9.
- Daley, J. (1984). How Glynis Nunn trains. *Modern Athlete and Coach*, 22, 29-30.
- Garson, G. D. (2011a). **Factor analysis. Statnotes: Topics in Multivariate Analysis.** Retrieved 5/2/2011 from <http://faculty.chass.ncsu.edu/garson/PA765/factor.htm>
- Garson, G. D. (2011b). **Structural equation modelling. Statnotes: Topics in Multivariate Analysis.** Retrieved 5/2/2011 from <http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>.
- Hair, J. E., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. (5th Ed.) Upper Saddle River: Prentice Hall.
- Hair, J. E., Block, W., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate*

- data analysis*. (6th Ed.) Upper Saddle River: Pearson - Prentice Hall.
- Hancock, T. (1987). Problems confronting coaches in multiple events. *Modern Athlete and Coach*, 25, 36-38.
- Heazlewood, T. (2008). Factor structure of the women's Heptathlon: Implications for training. *Proceedings of First Joint International Pre-Olympic Conference on Sports Sciences and Sports Engineering. Volume I: Computer Science in Sport*. Ed. Y. Jiang, A. Baca & H Zhang. Nanjing, China, August 4-7, 2008, 283-288.
- International Association of Athletics Federations. (2009). *Competition rules 2009*. International Association of Athletics Federations (IAAF): Monaco.
- International Association of Athletics Federations. (2006). *Heptathlon 2006 outdoor lists*. Retrieved from IAAF website on December 1, 2007 <http://www.iaaf.org/statistics/toplists/inout=O/ageGroup=N/season=2006/gender=W/discipline=HEP/legal=A/index.html>.
- International Association of Athletics Federations. (2010). *Heptathlon 2010 outdoor lists*. Retrieved from IAAF website on January 5, 2011 <http://www.iaaf.org/statistics/toplists/inout=O/age=n/season=2010/sex=W/all=n/legal=A/disc=HEP/detail.html>
- Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. London: Sage.
- Mackenzie, B. (2007). Heptathlon. *Sports Coach*. Retrieved 5/1/2006 from <http://www.brianmac.co.uk/hepth/>
- Marra, H. (1985). Decathlon hints. *Modern Athlete and Coach*, 23, 15-18.
- Norusis, M. J. (1985). *SPSSX*. Sydney: McGraw-Hill Book Company.
- Sarponov, E. (1982). Looking at the future of Heptathlon. *Modern Athlete and Coach*, 20, 39-40.
- SPSS Inc. (2007). *PASW statistics 18.0: User's guide*. Chicago: SPSS Inc.
- Telfer, P. (1988). The javelin in the Heptathlon. *Modern Athlete and Coach*, 26, 32-34.