The 3H and Spiral Dynamics Models
A Reconciliation

Jehanzeb Rashid Cheema

This study explores the relationship between the Spiral Dynamics and the 3H (head, heart, hands) models of human growth and development, using constructs such as empathy, moral reasoning, forgiveness, and community mindedness that have been shown to have implications for education. The specific research question is, "Can a combination of multivariate statistical techniques be utilized to find an alignment between the dimensions of these models?" We focus on practical and data-driven approaches with the primary methods including factor analysis, cluster analysis, and logistic regression. Our main finding is that the proposed methodology is robust and applicable in a variety of operational scenarios. We conclude it is feasible to empirically align and reconcile dimensions of seemingly disparate theories of educational development and human evolution with a data analysis framework based on mainstream quantitative techniques that can be easily implemented using readily available statistical software packages.

Keywords: 3H model, spiral dynamics model, cluster analysis, factor analysis, classification, cross-validation, bootstrap

Spiral Dynamics is a psycho-cultural growth model built around the evolution of human thinking and behavior. The model was pioneered by Graves (1970) and later promoted by Beck and Cowan (1996). The Spiral Dynamics model proposes a series of tiers through which individuals and societies evolve over time. Each tier or stage represents a unique set of values, perceptions, and beliefs. Theses tiers are grouped into three categories (hereafter referred to as Groups 1, 2, and 3) and are typically

---

Jehanzeb Cheema is an Instructional Associate Professor at George Mason University with a background in Education, Economics, and Health Informatics. Prior to joining Mason he taught quantitative courses at the University of Baltimore and the University of Illinois at Urbana-Champaign.

Cheema, J.R. (2024). The 3H and Spiral Dynamics Models. *Journal of Education in Muslim Societies, 6*(1), 55–76.
DOI 10.2979/jems.00016 •

color-coded: Group 1 includes individuals who are instinctive (beige), animistic (purple), or egocentric (red); Group 2 includes individuals described as absolutistic (blue), multiplistic (orange), or relativistic (green); and Group 3 includes all individuals who are systemic (yellow), holistic (turquoise), or cosmic (coral). It should be noted that within each group the sub-categories have been presented in an increasing order of magnitude. For instance, for Group 1 the three sub-categories range from instinctive to egocentric in an ascending order of magnitude. Similarly, for Group 2 the three sub-groups, as listed from lowest to highest category, are absolutistic, multiplistic, and relativistic. For a more detailed discussion and potential applications of the Spiral Dynamics model we direct the interested reader to Brown et al. (2023), Cowan and Todovoric (2000), Ivanova et al. (2021), Salters (2011), and Cacioppe and Edwards (2005).

The Spiral Dynamics model has recently been associated in the literature with efforts to align modern psychological constructs with other paradigms (Nasser, Miller-Idriss, & Alwani, 2018). One such example is the 3H model. The 3H model—where 3H stands for head, heart, and hand—was initially introduced by Orr (1992) and later promoted by Sipos et al. (2008) to highlight the comprehensive nature of evolutionary experiences in human thinking and behavior. Using the model's language, the head and heart refer to cognitive and affective links with relational understanding and engagement. In addition, the model seeks to establish the connection between critical reflection and the psychomotor and emotional domains (referred to as hand and heart respectively in the model).

There has been some theoretical work related to aligning theories of educational development and human growth and evolution. Specifically, within context of the 3H and Spiral Dynamics models Nasser et al. (2019) propose a possible alignment of the dimensions of the 3H model with those of the Spiral Dynamics model using four constructs—empathy, moral reasoning, forgiveness, and community mindedness—that span the head, hands, and heart components of the 3H model. For example, empathy involves problem solving and tends to be associated with cognitive functions (i.e. the head dimension) as well as emotional intelligence which tends to be associated with emotional functions (i.e. the heart dimension). Similarly, forgiveness involves decision-making (i.e. the cognitive or head dimension) as well as interdependence and social well-being (i.e. the behavioral or hand dimension). As these examples suggest, the four constructs could be acceptable proxies for the latent traits that constitute the 3H model. For a more focused discussion of the 3H model, especially within the context of its application in the social sciences, we refer to Singleton (2015), Abu-Nimer (2001), Spiewak and Sherrod (2011), and Sipos, Battisti, and Grimm (2008).

The current study is motivated by the work of Nasser et al. (2018) and seeks to develop an empirical approach that may be used to quantitatively align a holistic educational framework such as the 3H model with a model (such as Spiral Dynamics) that focuses on the description of evolution of human consciousness and cultural development through a series of increasingly complex value systems. We are not aware of any prior study that has attempted to specifically address alignment between these two models. We seek to keep the methodology accessible to the lay audience by employing mainstream statistical techniques that are readily implementable using standard software tools. Furthermore, our aim is a methodological template that can be applied beyond these specific models to allow evidence-based research for aligning theories that transcend the traditionally defined boundaries of the social sciences and merge seemingly disparate disciplines within that realm.

We have tested the statistical approach described using simulated data. The next three sections describe the sample and variables, statistical results, discussion of the results, and conclusions.

METHODOLOGY

Research Design

We employed a simulation-based research design with balanced groups. The study's main objective was to evaluate the feasibility and capacity of mainstream data mining methods such as factor analysis, cluster analysis, and logistic regression—among others—to align the dimensions of the 3H model with those of the Spiral Dynamics model. The specific research question is, "Can a combination of multivariate statistical techniques be utilized to find an alignment between the dimensions of the 3H model and Spiral Dynamics model?"

Data

To test the effectiveness of the approach outlined in the previous section we simulated three samples of 150, 450, and 750 observations. These three sample sizes were selected to represent the small, medium, and large samples and were subsets from an original sample size of 1,350.

The original sample size was divided into three subgroups to evaluate whether our data analysis procedures are sufficiently robust to address variation in sample size where the number of observations is as few as 150 up to samples five times as large. The three groups correspond to Groups 1, 2, and 3 based on the nine categories of the Spiral Dynamics model as discussed earlier.

Four variables simulating empathy, forgiveness, moral reasoning, and community mindedness constructs (i.e. the three dimensions of the 3H model) were generated. To mimic a real-world Likert-type scale, each variable was constrained to a 1–5 scale where the values were coded as follows: (1) strongly disagree, (2) moderately disagree, (3) undecided, (4) moderately agree, and (5) strongly agree. The values of each of the four variables were randomly selected from a discrete uniform distribution with the interval being [1, 3] for Group 1, [2, 4] for Group 2, and [3, 5] for Group 3. For example, the response outcomes for Group 1—which we earlier described as instinctive, animistic, or egocentric—ranged between strongly disagree and undecided, i.e., defined on the interval [1,3]. The choice of these intervals was dictated by (1) the underlying 1–5 Likert type scale, (2) the number of groups (k = 3), the desire to keep variation in variable values constant across groups, and (4) to obtain separation between groups while still allowing for some overlap between them. The choice of probability distribution was driven by the discrete nature of responses on the simulated 1–5 scale. It should be noted that for each sample size (low, medium, and high) by design the four variables are positively (but spuriously) correlated with each other when group membership is ignored and uncorrelated with each other within each group. Furthermore, given the overlap in variable values between groups, the distribution of each of the four variables is expected to be symmetric and unimodal within each sample (see Figure 1). Summary statistics and correlations for the simulated variables are presented in Tables 1 and 2 respectively.

Analytical Method

To measure variables of interest such as empathy, forgiveness, moral reasoning, and community mindedness, we used a single item per variable. The item wordings are as follows: Empathy, "When I see someone being taken advantage of, I feel kind of protective toward him/her;" Forgiveness, "Even if someone wrongs me, it would be wrong to seek revenge;" Moral reasoning, "Justice is the most important requirement for a society" and;

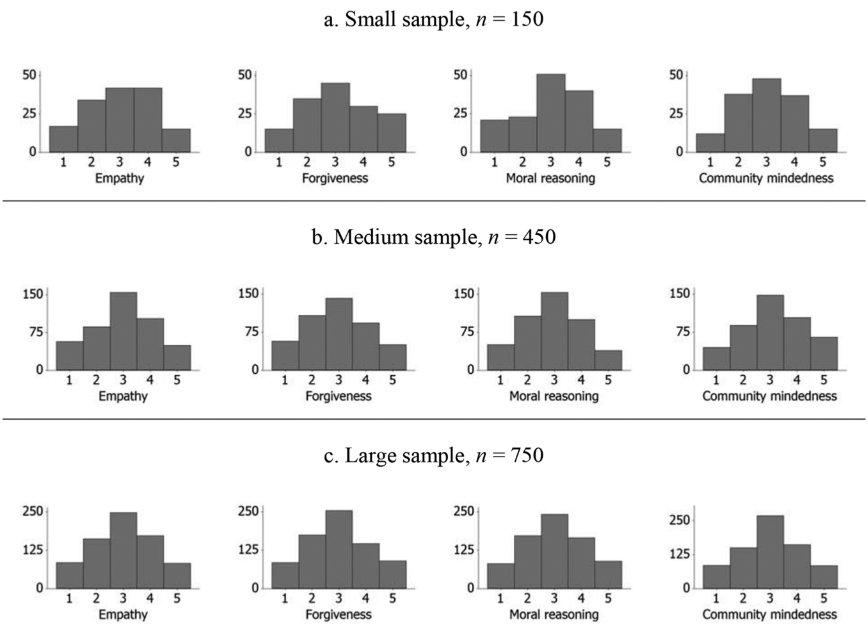

Figure 1. Histograms of simulated variables for (a) small sample, n = 150; (b) medium sample, n = 450; and (c) large sample, n = 750.

Community mindedness, "I feel confident in my ability to bring people together to address a community need." The response categories for all items ranged from (1) strongly disagree to (5) strongly agree. For detailed information on these items we direct the interested reader to Nasser et al. (2020).

To align dimensions of the 3H and Spiral Dynamics models, a number of statistical methods can be utilized. These include cluster analysis, factor analysis, and logistic regression (Everitt, Landau, Leese, & Stahl, 2011). The application of these methods depends on the assumptions made about the number of categories (or groups) proposed by a theory. For example, if one assumes there are nine categories distributed into three ordinal groups as the extended Spiral Dynamics theory suggests (Groups 1–3 as defined earlier in this paper), then cluster analysis can be performed to force each case into exactly one of these categories. For a more refined result, each case can then be further classified into each of the nine sub-groups. This reduces the problem to that of group membership where the total number of groups is already known and membership in a group is based on the observed distances between groups (Ritter,

Table 1. Summary Statistics for the Simulated Data

Sample size/Group	Empathy		Forgiveness		Moral reasoning		Community mindedness	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Small sample								
Group 1	1.96	0.81	2.02	0.80	1.92	0.88	2.14	0.78
Group 2	3.16	0.89	2.98	0.80	3.14	0.76	2.92	0.83
Group 3	3.96	0.81	4.30	0.79	4.04	0.75	4.04	0.75
Total	3.03	1.17	3.10	1.22	3.03	1.18	3.03	1.11
Medium sample								
Group 1	1.96	0.85	1.90	0.81	2.04	0.84	2.08	0.82
Group 2	3.06	0.80	2.91	0.81	2.91	0.86	3.13	0.81
Group 3	3.99	0.82	4.00	0.82	3.86	0.80	4.17	0.82
Total	3.00	1.17	2.94	1.18	2.94	1.12	3.12	1.18
Large sample								
Group 1	1.98	0.81	1.96	0.80	2.02	0.82	2.05	0.86
Group 2	3.06	0.81	2.96	0.79	2.96	0.83	3.01	0.82
Group 3	3.99	0.82	4.01	0.84	4.05	0.81	3.98	0.84
Total	3.01	1.16	2.98	1.17	3.01	1.17	3.01	1.15

Note. Sample size: small, $n = 150$; medium, $n = 450$; large, $n = 750$. The interval for all variables is [1, 3] for Group 1, [2, 4] for Group 2, and [3, 5] for Group 3. M = mean. SD = standard deviation. Paired samples t test was performed to evaluate means differences among the four variables (6 tests per group) separately within each group for a total of 18 tests. Familywise error rate was controlled using Bonferroni adjustment. None of the 18 tests was statistically significant at the adjusted alpha level of .003.

Table 2. Spearman Correlation Matrix for the Simulated Data

	Small sample			Medium sample			Large sample		
Group/Variable	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
All groups									
1. Empathy	–			–			–		
2. Forgiveness	.53	–		.52	–		.49	–	
3. Moral reasoning	.48	.55	–	.51	.53	–	.52	.51	–
4. Community mindedness	.39	.61	.48	.53	.57	.47	.47	.49	.48
Group 1									
1. Empathy	–			–			–		
2. Forgiveness	-.16	–		-.07	–		~0	–	
3. Moral reasoning	-.06	~0	–	-.01	.16*	–	.02	.02	–
4. Community mindedness	-.15	.26	-.04	.15	.10	.01	-.04	-.07	.01

(*Continued*)

Table 2. Continued

Group 2									
1. Empathy	–			–			–		
2. Forgiveness	.09	–		.03	–		-.08	–	
3. Moral reasoning	-.09	~0	–	.11	.10	–	.13*	.02	–
4. Community mindedness	-.15	.15	.02	-.01	.05	.02	-.03	.07	-.01
Group 3									
1. Empathy	–			–			–		
2. Forgiveness	.05	–		.07	–		-.01	–	
3. Moral reasoning	-.10	-.12	–	.08	.01	–	-.07	.02	–
4. Community mindedness	-.23	.08	-.25	.07	.10	-.06	.05	-.02	.04

Note: Sample size: small, n = 150; medium, n = 450; large, n = 750.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

2015). This is referred to as an unsupervised approach because there is no prior data available to guide the group membership process. This method can also be applied in an exploratory context when no prior assumption is made about the number of groups or sub-groups; instead, this information is generated from the observed sample (Roberts, 1997; Hennig, Meila, Murtagh, & Rocci, 2016).

An alternative to the unsupervised approach employed in this study is to collect group membership information directly from the study participants where each participant self-selects themselves into one of the nine categories of the Spiral Dynamics model based on their own perception of which category they belong to. This alternative (supervised) approach is commonly referred to as classification (Tabachnick & Fidell, 2013) where we can treat either the scores on 3H constructs or the categories of Spiral Dynamics model as the dependent variable in order to predict the other. This alternative approach is mentioned in passing here and, given the absence of relevant survey data, is not further explored in this paper.

The two unsupervised methods described earlier—factor analysis and cluster analysis—were developed to solve two different types of problems. The main aim in cluster analysis is to identify the natural groups that exist in a dataset. These groups are based on similarities between cases and differences among groups. Several popular algorithms for cluster analysis exist. These include *k*-means, two-step cluster analysis, and hierarchical cluster analysis, among others (Hennig et al., 2016). Each of these models allows either imposition of a pre-determined number of clusters or determination of an optimal number of clusters based on some pre-defined optimization criterion (Everitt et al., 2011).

The main objective in factor analysis is to reduce the number of dimensions (and thus reduce the complexity) of a model. Factor analysis works by combining underlying items or sub-scales (i.e. variables) into factors (or scales) based on the similarity among underlying variables. Such factor analysis can be exploratory when no prior constraint exists on the number of extracted factors or confirmatory when a predetermined number of factors is extracted from the procedure (Tabachnick & Fidell, 2013). In either case the method is considered unsupervised because of the absence of a dependent variable (Hofmann, 2001). Although the primary motivation behind factor analysis is to combine variables into factors as opposed to cluster analysis, where the focus is to combine observations into groups, factor analysis retains only that variation in extracted factors that is common across all underlying variables. In other words, any variation in underlying variables that can be attributed to random or unknown sources (i.e. all sources other than the following four constructs in the 3H model: empathy, forgiveness, moral reasoning, and community mindedness) is left out when factor scores are computed (Tabachnick & Fidell, 2013). These factor scores can then finally be used to separate observations into groups using a clustering algorithm. Theoretically, this combination of factor analysis and cluster analysis should provide a superior solution to employing cluster analysis in isolation because factor analysis helps remove noise from the underlying variables.

A combination of cluster analysis, classification, cross-validation, factor analysis, and classification performance metrics were used in this study. A brief description of each type of analysis follows.

Confirmatory Factor Analysis. Factor analysis was used to combine variables representing the 3H model—empathy, moral reasoning, forgiveness, and community mindedness—into a single index. The reason for using factor analysis was to extract common variation in the underlying variables as it directly feeds into determining group membership in the Spiral Dynamics model. This factor analysis was confirmatory based on the assumption that the four variables adequately capture the head, hands, and heart components of the 3H model. Results of cluster analysis with and without factor analysis were compared to determine whether or not factor extraction made any difference. Factor analysis was based on the correlation matrix using principal axis factoring as the extraction algorithm and the number of factors was based on an Eigenvalue exceeding 1. The Kaiser-Meyer-Olkin statistic, p value for Bartlett's test of sphericity, and Eigenvalue and percentage of total variation in the underlying variables explained by the extracted factor were also computed.

Cluster Analysis. The optimal number of clusters, k, was determined from the features of and interrelationships among the four input variables. In this part of the analysis, we did not compare cluster membership with true group membership. Given its widespread acceptance in the current literature (Hartigan & Wong, 1979; Arora, Deepali, & Varshney, 2016; Likas, Vlassis, & Verbeek, 2001; Yoo et al., 2011; Jain, 2010) the k-means algorithm was used for cluster identification, and the Davies-Bouldin index was employed as the criterion for selecting optimal k (Davies & Bouldin, 1979).

Classification Analysis. The purpose of this part of the analysis was to compare how well clusters were allocated to observations relative to true group membership. Such goodness of fit comparison is only possible due to the simulated nature of our data where we already have information on true group membership. This confirmatory analysis helped us evaluate the efficacy of our proposed methodology. For this approach, we provided the expected number of clusters (k = 3 from the Spiral Dynamics model) as an input to the classification algorithm. To keep cluster analysis results comparable with classification results, the k-means algorithm was retained for classification analysis.

Cross-Validation and Bootstrap Resampling. To ensure our fitted clustering models did not suffer from overfitting, which is a real concern whenever a model is built entirely on observed data, we employed two forms of cross-validation: k-fold cross-validation, and bootstrap validation. The k-fold cross-validation randomly divides sample data into k equal parts called folds. One of these folds is reserved for testing while those remaining are used to build the model. This approach is repeated k times giving each of the k folds the opportunity to play the role of testing set. Final results are the average of k individual fold results (Hennig et al. 2016). The number of folds selected for k-fold cross-validation in this study was 10. This number was driven by the desire to have a reasonable number of minimum observations per fold and is widely supported by existing literature (Bengio & Grandvalet, 2005; Kohavi, 1995). For bootstrap validation we used a 50–50 split for training and testing example sets. The somewhat high proportion for the testing set is justified by our smallest sample size, n = 150, which gives us just 75 observations to test the fitted model. To minimize sources of variability in results obtained from different sample sizes, the 50–50 split and number of folds (k = 10) remained constant in this study.

Bootstrap resampling was used as an additional check against overfitting in the fitted models. This method works by selecting subsamples of a given size from the original sample a specific number of times. Since the sampling is done with replacement, it is possible for the sub-sample size to be larger than the original sample size (Efron, 1979). Following our earlier justification for a 50–50 split between the training and sample sets we used one half of the original sample to generate 10,000 sub-samples in order to train the clustering algorithm. The resulting model was then used on the other half of the sample (i.e. the testing set) to evaluate its goodness of fit. The choice of 10,000 sub-samples was dictated by the approximately symmetrical distributions of our variables and the wide support for this number in the current literature (Hayes, Krippendorf, 2007; Gibson et al., 2011; Visscher, Thompson, & Haley, 1996).

Classification Performance Metrics. To test the goodness of fit of our clustering models we used several performance measures such as accuracy, precision, recall, and *F1* (Powers, 2011). Each of these measures has its pros and cons, and should be used in an appropriate combination to evaluate the performance of a clustering algorithm in a holistic manner. Calculation of these performance measures required applying the models learned from clustering analysis in a classification context, i.e. predicting true group membership from cluster membership. We also estimated a series of binary logistic regression models to evaluate the effect of study features, such as validation method, sample size, and presence or absence of factor analysis, on classification performance metrics.

RESULTS

Confirmatory Factor Analysis

To compare cluster analysis results with and without factor analysis, factor scores were computed for each sample size separately. The reason for using three separate samples is to evaluate whether the procedures explored in this study are sensitive to sample size. A satisfactory performance of these procedures at various sample sizes shows the robustness of these procedures. These confirmatory factor analysis results are summarized in Table 3. The Anderson-Rubin method produces standardized factor scores with $M = 0$ and $SD = 1$ which can thus be interpreted as z scores. The non-zero determinant values, *KMO* values generally at or above 0.8, and highly significant p values on Bartlett's test of sphericity support the adequacy of input variables for the factor analysis procedure. All factor analysis runs supported the existence of a single factor with explained variation of approximately 50%. For example, for the medium sample size category, the determinant was non-zero ($|\mathbf{R}| = 0.27$), the *KMO* statistic was 0.80, which is considered acceptable, and Bartlett's test of sphericity was statistically significant, $p < .001$. These results indicate that our data are adequate for the factor analysis procedure. Factor analysis results further suggested that for the medium sample size category all four underlying variables loaded on a single dimension as there was only one Eigenvalue that exceeded 1. The proportion of variation in these variables as extracted by this factor was 52.45%. This is a bit lower than the usual recommended textbook cut-off of 67%, but still better than combining the four values using a simple average.

Table 3. Summary of Confirmatory Factor Analysis Results

Sample size	|R|	*KMO*	Bartlett's test of sphericity χ^2	Bartlett's test of sphericity *p*	Total variance explained $Max(\lambda_i)$	Total variance explained %	$n(\lambda_i) > 1$
Small	0.28	0.77	187.42	< .001	2.52	51.36	1
Medium	0.27	0.80	579.63	< .001	2.57	52.45	1
Large	0.31	0.80	869.18	< .001	2.50	49.99	1

Note: Sample size: small, n = 150; medium, n = 450; large, n = 750. Extraction method = principal axis factoring. |**R**| = determinant of the correlation matrix. *KMO* = Kaiser-Meyer-Olkin measure of sampling adequacy. χ^2 = approximate Chi-square value. λ_i = Eigenvalue of *i*th factor.

Factor scores are summarized in Table 4. Since these are standardized, the positive estimates (excluding *SD*) reported in Table 4 indicate above-average values, while the negative estimates indicate below average values. For example, in the small sample size category, the standardized factor scores for Group 2 ranged between -0.89 and 0.80 with a mean of -0.03 (*SD* = 0.47) which was lower than the overall mean of 0 for the combined groups.

Table 4. Summary Statistics for Factor Scores by Sample Size and by Group

Sample size	Group	*Min*	*Max*	*M*	*SD*
Large	1	-2.18	0.00	-1.09	0.44
	2	-1.09	1.09	-0.01	0.46
	3	0.00	2.18	1.10	0.45
	Total	-2.18	2.18	0.00	1.00
Medium	1	-2.14	0.00	-1.08	0.48
	2	-1.07	1.07	0.00	0.47
	3	0.00	2.14	1.08	0.47
	Total	-2.14	2.14	0.00	1.00
Small	1	-1.96	-0.06	-1.08	0.45
	2	-0.89	0.80	-0.03	0.47
	3	0.11	1.67	1.11	0.40
	Total	-1.96	1.67	0.00	1.00
Total	1	-2.18	0.00	-1.08	0.46
	2	-1.09	1.09	-0.01	0.46
	3	0.00	2.18	1.09	0.45
	Total	-2.18	2.18	0.00	1.00

Note: Sample size: small, n = 150; medium, n = 450; large, n = 750. Factor scores are standardized (M = 0, SD = 1).

Cluster Analysis

Results from cluster analysis with and without factor analysis are summarized in Table 5. Figure 2 shows the average Davies-Bouldin (DB) index from *k*-means runs plotted against the number of identified clusters without factor analysis. Each plotted data point in Figure 2 represents the average of 10 runs by splitting the sample into 10 equal parts. Based on minimization of the DB index criterion, this exploratory cluster analysis suggested the optimal number of clusters to be 3, 4, and 4 in our small, medium, and large samples. It should be noted that for the medium sample a local minimum in DB index was observed at $k = 4$. This is where the arm of the DB index plot begins to form justifying 4 as the optimal number of clusters for this sample size. The number of clusters identified is very similar to the true number of groups, suggesting that the *k*-means algorithm is adequate for further analysis.

Our cluster analysis results with factor analysis are summarized in Figure 3. The number of clusters suggested by Figure 2 is $k = 2$ for all sample sizes. We note that the arms of the DB curve remain flat at $k = 3$, suggesting there is not much to lose or gain by adding a third cluster. In general, these results are very close to the true number of clusters ($k = 3$) in our simulated data, again suggesting that the *k*-means algorithm is an adequate choice for further analysis.

Table 5. Summary Statistics for Davies-Bouldin Index Values from k-Means Procedure (with and without Factor Analysis) for Small, Medium, and Large Samples

Based on factor scores?	*k*	Small sample *M*	Small sample *SD*	Medium sample *M*	Medium sample *SD*	Large sample *M*	Large sample *SD*
No	2	-0.90	0.09	-1.00	0.11	-1.02	0.07
	3	-0.98	0.12	-1.25	0.11	-1.36	0.09
	4	-0.99	0.13	-1.28	0.12	-1.33	0.09
	5	-0.82	0.11	-1.24	0.06	-1.36	0.13
	6	-0.75	0.10	-1.16	0.05	-1.28	0.05
	7	-0.70	0.09	-1.13	0.06	-1.24	0.08
	8	-0.60	0.11	-1.06	0.04	-1.20	0.09
	9	-0.53	0.07	-0.97	0.05	-1.16	0.07
	10	-0.48	0.07	-0.96	0.10	-1.13	0.05
	11	-0.46	0.21	-0.99	0.13	-1.14	0.11

(*Continued*)

Table 5. Continued

Yes	2	-0.43	0.06	-0.49	0.06	-0.50	0.06
	3	-0.44	0.12	-0.50	0.06	-0.51	0.04
	4	-0.38	0.07	-0.48	0.07	-0.51	0.04
	5	-0.35	0.08	-0.45	0.05	-0.48	0.02
	6	-0.33	0.08	-0.46	0.04	-0.47	0.02
	7	-0.26	0.08	-0.43	0.04	-0.50	0.03
	8	-0.20	0.10	-0.39	0.04	-0.47	0.05
	9	-0.16	0.08	-0.38	0.07	-0.42	0.06
	10	-0.11	0.05	-0.34	0.07	-0.38	0.05
	11	-0.13	0.16	-0.32	0.09	-0.36	0.07

Note: Sample size: small, n = 150; medium, n = 450; large, n = 750. k = number of clusters. M = mean. SD = standard deviation.

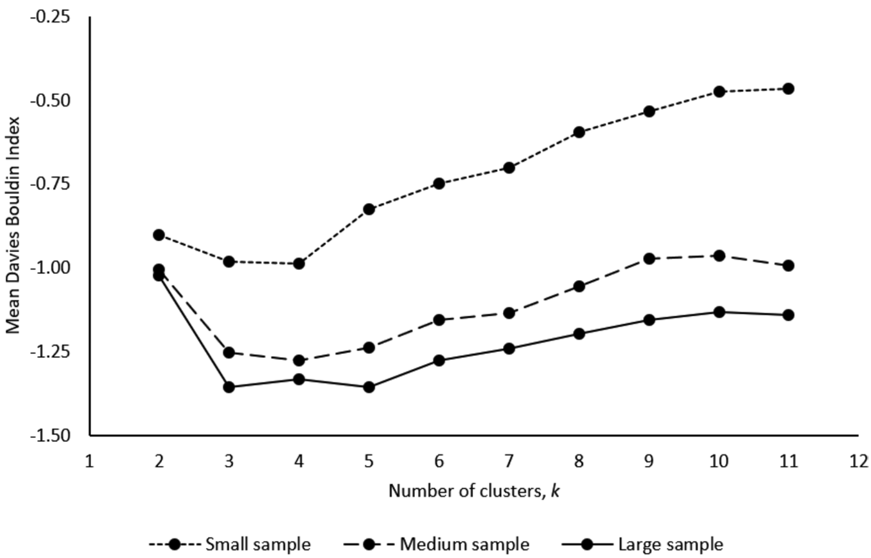

Figure 2. Visual summary of cluster analysis results (without factor analysis). Average Davies-Bouldin index plotted as a function of the number of clusters for small, n = 150; medium, n = 450; and large, n = 750 samples using k-means clustering algorithm with each data point representing 10 runs.

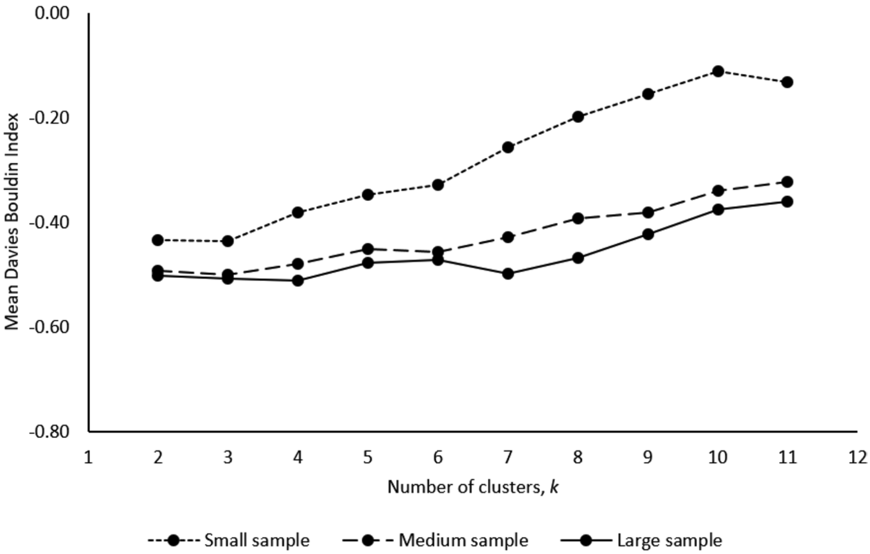

Figure 3. Visual summary of cluster analysis results (with factor analysis). Average Davies-Bouldin index plotted as a function of the number of clusters for small, n = 150; medium, n = 450; and large, n = 750 samples using k-means clustering algorithm with each data point representing 10 runs.

Classification Analysis

Classification results based on models trained by cluster analysis are summarized in Table 6. Given the generally small sample sizes simulated in this study, these results indicate an overall good prediction of true group membership. Out of the 108 values of precision, recall, and F1 reported in Table 6, we observe that 71 (65.7%) have a magnitude of 80% or more. The minimum accuracy was 79.6%, and the average margin of error in accuracy was 4.2%. Overall, the mean was 83.0% (*SD* = 8.6%) for precision, 82.4% (*SD* = 6.9%) for recall, 82.5% (*SD* = 6.4%) for F1, and 82.4% (*SD* = 2.3%) for accuracy, with the mean margin on accuracy being 4.2% (*SD* = 2.5%).

Table 6. Summary of Classification Results

Based on factor scores?	Validation method	Sample size	Group	Precision	Recall	*F1*	Accuracy in the sample %	Margin (+/-)
No	k-fold	Small	1	93.6	88.0	90.7	87.3	8.1
			2	81.6	80.0	80.8		
			3	87.0	94.0	90.4		
		Medium	1	93.0	79.3	85.6	79.6	4.5
			2	69.3	69.3	69.3		
			3	78.5	90.0	83.9		
		Large	1	84.0	92.4	88.0	83.9	4.2
			2	76.5	74.4	75.5		
			3	91.4	84.8	88.0		
	Bootstrap	Small	1	87.8	83.4	85.5	81.7	7.2
			2	75.4	68.2	71.7		
			3	81.6	94.1	87.4		
		Medium	1	91.2	79.7	85.1	80.0	2.4
			2	69.0	73.4	71.1		
			3	81.9	87.0	84.4		
		Large	1	86.6	88.1	87.3	83.3	1.4
			2	74.5	76.0	75.2		
			3	89.1	85.8	87.4		
Yes	k-fold	Small	1	97.5	78.0	86.7	84.0	8.5
			2	74.1	80.0	76.9		
			3	83.9	94.0	88.7		
		Medium	1	93.0	79.3	85.6	79.8	3.9
			2	68.3	73.3	70.7		
			3	80.7	86.7	83.6		
		Large	1	90.2	81.2	85.5	83.2	3.4
			2	70.4	85.6	77.3		
			3	93.7	82.8	87.9		
	Bootstrap	Small	1	90.5	77.9	83.7	82.0	3.0
			2	73.0	74.4	73.7		
			3	83.7	94.3	88.7		

(*Continued*)

Table 6. Continued

	Medium	1	92.2	78.8	85.0	80.4	2.5
		2	68.6	76.8	72.5		
		3	83.6	85.8	84.7		
	Large	1	88.8	85.0	86.9	83.5	1.5
		2	72.5	81.5	76.8		
		3	91.5	84.0	87.6		

Note: Sample size: small, $n = 150$; medium, $n = 450$; large, $n = 750$. Bootstrap validation is based on 10,000 samples with replacement, and a 50–50 split between training and test sets.

Logistic Regression

To evaluate the effect of validation method, sample size, and factor analysis on classification performance metrics, we also estimated a series of logistic regression models. The dependent variable in each model was the classification status with a value of 1 for cases that were correctly classified and a value of 0 for cases that were incorrectly classified. Independent variables in this model were validation method (k-fold, bootstrap), sample size (small, medium, large), and factor analysis (yes, no). All two-way interaction terms between independent variables were also included in the model which was estimated separately for each of the four individual performance metrics (precision, recall, F1, and accuracy). None of these logistic regression models returned statistically significant results, suggesting that the differences in performance metrics were statistically not significant between *k*-fold and bootstrap validation methods; between small, medium, and large samples; and between results obtained with and without factor analysis.

The main objective of the analyses described in the preceding paragraph is to provide evidence of the robustness of our analytical approach. For example, since our main results did not change when we switched the sample size or changed the validation method, it means that the data analysis technique suggested in this study can be used in a variety of scenarios. Changes in sample size, validation method, omission of factor analysis etc. are unlikely to have any undue influence on the statistical results in scenarios that are consistent with the objectives of this study. It can readily be seen from the results presented in Table 6 that shifting from non-factor analysis-based input data to that based on factor analysis, or changing the validation method from *k*-fold to bootstrap etc. had no significant result on outcome metrics such as precision, recall, F1, and accuracy. For example, in the small sample category, with all else held constant (Group = 1, Factor analysis = No), when only the validation method was changed the accuracy statistic changed from 87.3% to 81.7%. Similarly, for the same sample size category, with all else held constant (Group = 1, validation method = *k*-fold) when factor analysis was employed the accuracy statistic changed from 87.3% to 84.0%. The remaining results reported in Table 6 can be interpreted similarly. The bottom-line of this discussion is that a variation in conditions such as whether or not factor analysis was performed, change in validation method, sample size etc. did not on average have any statistically significant effect on the outcome.

DISCUSSION

The main objective of this study was to explore a practical approach to align dimensions of the 3H and the Spiral Dynamics models. Based on the specific natures of these models, we proposed a method based on a combination of cluster analysis and factor analysis which was tested with carefully simulated data designed to mimic properties of the two models. Our cross-validated and robust analytical results generated satisfactory performance metrics suggesting that the proposed alignment method will work well with real-world data.

To the best of our knowledge this is one of the first studies that has attempted to quantitatively align a holistic educational framework such as the 3H model with the Spiral Dynamics model that focuses on the description of evolution of human consciousness and cultural development through a series of increasingly complex value systems. The absence of prior empirical studies in this area makes it difficult for us to situate our study in a pre-existing research framework. In addition, we are currently not aware of any large scale effort that aims to provide empirical data for aligning the 3H and Spiral Dynamics models, apart from the Advancing Education in Muslim Societies (AEMS) initiative at the International Institute of Islamic Thought (IIIT, n.d.)—a center for educational research and Islamic thought with a focus on evidence-based research—but are confident any future research in this direction can benefit from the analytical methodology that we have developed and evaluated in this paper.

IMPLICATIONS

The main implication of this study is that the methodology used to reconcile the 3H and Spiral Dynamics models in this study is sufficiently robust to be replicated with additional models of educational and psycho-social human development. Our study should be seen as an initial attempt and a template for both academics and practitioners interested in generating evidence-based research for aligning theories that transcend traditionally defined boundaries of social sciences to merge seemingly disparate disciplines such as religion and education.

LIMITATIONS

Even though our analytical results generated encouraging performance metrics, we acknowledge that these results are based on simulated data. Notwithstanding the care exercised in this simulation, real-world results are difficult to predict and will almost invariably include additional sources of variation due to national, regional, ethnic, linguistic and other differences across sample participants.

SUGGESTIONS FOR FUTURE RESEARCH

The analytical approach described in this study can be further refined by taking into consideration additional factors that can cause variation in estimation results. This includes collecting data from more than one region within the same country, as well as from across multiple countries, including variables (such as charitable giving, virtue, nobility, religiosity etc.) in addition to the four that we used to represent the various dimensions of the 3H model, and incorporating demographic factors such as age, gender, work experience, and income level etc.

Another avenue of extension is to explore the effect of additional validation methods, sample sizes, extraction algorithms for factor analysis, and new goodness of fit metrics to demonstrate the feasibility of these techniques under a variety of scenarios.

CONCLUSION

The main conclusion of this study is that it is feasible to empirically align and reconcile dimensions of seemingly disparate theories of educational development and human evolution by developing a data analysis framework based on mainstream quantitative techniques that can be easily implemented using readily available statistical software packages. The analytical procedures used in this study have the potential for replication both within and across social science disciplines.

References

Abu-Nimer, M. (2001). Conflict resolution, culture, and religion: Toward a training model of interreligious peacebuilding. *Journal of Peace Research, 38*(6), 685–704. doi:10.1177/0022343301038006003

Arora, P., Deepali, D., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science, 78*, 507–512. doi:10.1016/j.procs.2016.02.095

Beck, D., & Cowan, C. (1996). *Spiral dynamics: Mastering values, leadership, and change.* Wiley-Blackwell.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research, 5*, 1089–1105. http://www.jmlr.org/papers/volume5/grandvalet04a/grandvalet04a.pdf

Brown, L.R., McCray, P., & Neal, J. (2023). Creating affective collaborative adult teams and groups guided by spiral dynamic theory. In: Guralnick, D., Auer, M.E., Poce, A. (eds) Innovative approaches to technology-enhanced learning for the workplace and higher education. TLIC 2022. *Lecture Notes in Networks and Systems, 581.* Springer. https://doi.org/10.1007/978-3-031-21569-8_8

Cacioppe, R., & Edwards, M. (2005). Seeking the Holy Grail of organisational development: A synthesis of integral theory, spiral dynamics, corporate transformation and action inquiry. *Leadership & Organization Development Journal, 26*(2), 86–105. doi:10.1108/01437730510582536

Cowan, C., & Todorovic, N. (2000). Spiral dynamics: The layers of human values in strategy. *Strategy & Leadership, 28*(1), 4–11. doi:10.1108/10878570010335912

Davies, D., & Bouldin, D. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2,* 224–227. doi:10.1109/TPAMI.1979.4766909

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26. https://www.jstor.org/stable/2958830

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). John Wiley & Sons.

Gibson, L., Lee, T., Koh, L., Brook, B., Gardner, T., Barlow, J., . . ., & Sodhi, N. (2011). Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature, 478*, 379–381. doi:10.1038/nature10425

Graves, C. (1970). Levels of existence: An open system theory of values. *Journal of Humanistic Psychology, 10*(2), 131–155. doi:10.1177/002216787001000205

Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 28*(1), 100–108. doi:10.2307/2346830

Hayes, A., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77–89.

Henning, C., Meila, M., Murtagh, F., & Rocci, R. (2015). *Handbook of cluster analysis.* CRC Press.

Hoffman, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning, 42*, 177–196. doi:10.1023/A:1007617005950

International Institute of Islamic Thought [IIIT] (n.d.). Mapping the terrain survey studies 2018–2019 & 2019–2020. Advancing Education in Muslim Societies (AEMS). https://iiit.org/en/aems-research/

Ivanova, S., Romin, A., Gaynutdinova, L., Kvitka, S., Paliukh, V., & Arsawan, I. W. E. (2021). *Spiral Dynamics in the management of sustainable development of society. Earth and Environmental Science, 915*. https://doi.org/10.1088/1755-1315/915/1/012039

Jain, A. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*, 651–666. doi:10.1016/j.patrec.2009.09.011

Kohavi, R. (1995). A *study of cross-validation and bootstrap for accuracy estimation and model selection.* In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (II), Montreal, Quebec (1137–1143). http://www.sciepub.com/reference/125131

Likas, A., Vlassis, N., & Verbeek, J. (2011). *The global k-means clustering algorithm* (Technical Report No. IAS-UVA-01-02). Intelligent Autonomous Systems. https://hal.inria.fr/inria-00321493/document

Nasser, I., Cheema, J., Saroughi, M., & Alwani, A. (2020). *Advancing education in Muslim societies: Mapping the terrain report 2018–2019.* The International Institute of Islamic Thought.

Nasser, I., Miller-Idriss, C., & Alwani, A. (2018). *Reconceptualizing reform of education in Muslim societies.* Paper presented at the Symposium on Muslim Philanthropy and Civil Society, Indianapolis, IN.

Orr, D. (1992). *Ecological literacy: Education for a post modern world.* State University of New York.

Powers, D. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63. https://bioinfopublication.org/files/articles/2_1_1_JMLT.pdf

Ritter, G. (2015). *Robust cluster analysis and variable selection.* CRC Press.

Roberts, S. (1997). Parametric and non-parametric unsupervised cluster analysis. *Pattern Recognition, 30*(2), 261–272. doi:10.1016/S0031-3203(96)00079-9

Salters, D. (2011). Transactional analysis and spiral dynamics. *Transactional Analysis Journal, 41*(3), 265–276. doi:10.1177/036215371104100308

Singleton, J. (2015). Head, heart, and hands model for transformative learning: Place as context for changing sustainability values. *Journal of Sustainability Education, 9*. http://www.jsedimensions.org/wordpress/wp-content/uploads/2015/03/PDF-Singleton-JSE-March-2015-Love-Issue.pdf

Sipos, Y., Battisti, B., & Grimm, K. (2008). Achieving transformative sustainability learning: Engaging head, hands and heart. *International Journal of Sustainability in Higher Education, 9*(1), 68–86. doi:10.1108/14676370810842193

Spiewak, G., & Sherrod, L. (2011). *The shared pathways of religious/spiritual engagement and positive youth development.* In A. Warren, R. Lerner, & E. Phelps (Eds.), Thriving and spirituality among youth: Research perspectives and future possibilities (pp. 167–182). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118092699

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Visscher, P., Thompson, R., & Haley, C. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics Society of America, 143*, 1013–1020. http://www.genetics.org/content/143/2/1013

Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J., & Hua, L. (2012). Data mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems, 36*(4), 2431–2448. doi:10.1007/s10916-011-9710-5