Detecting Learners' Profiles based on the Index of Learning Styles Data*

Silvia Rita Viola[a], Sabine Graf[b], Kinshuk[c]

[a]Dipartimento di Ingegneria Informatica, Gestionale e dell' Automazione, Universita' Politecnica delle Marche, Ancona, Italy

[b]Women's Postgraduate College for Internet Technologies, Vienna University of Technology, Vienna, Austria

[c]School of Computing and Information Sciences, Athabasca University, Athabasca, Canada

{sr.viola, sabine.graf, kinshuk}@ieee.org

ABSTRACT

Adaptivity based on learning styles is considered by several adaptive systems, aiming at providing content that matches with the learning styles of students in order to make learning easier for them. However, for providing proper adaptivity, the learning styles of students have to be identified first. In most systems, learning style questionnaires are used for this issue, but such questionnaires have to deal with several problems, for example, some lack in validity and reliability. In this paper, we describe a model for profiling learners that is based on the answers to the Index of Learning Styles questionnaire, a commonly used instrument for detecting learning styles in traditional and online learning. The introduced model aims at overcomes the limitations of the questionnaire with respect to validity and reliability. . It uses Multiple Correspondence Analysis together with proximity measure to detect the most likely style for a learner. The results show that the model can be considered sufficiently reliable for detecting profiles, while less reliable for the active and reflective styles, and that the sensitivity of the proximity measure is an important issue and needs to be addressed. Additionally, an analysis of the profiles shows that some of them overlap because of their reciprocal influences. It can be concluded for the effectiveness of the approach for finding authentic profiles, even when unexpected relationships are found. Therefore, the model provides us with more accurate information about the learner by identifying the most influent learning style of a learner and its main characteristics.

Keywords

Learning styles, data mining, learners' models, learners' profiles, Index of Learning Styles

INTRODUCTION

Incorporating learning style in technology enhanced learning has potential to help student in learning and make learning easier for them. Felder [6] for example pointed out that learners with a strong preference for a certain learning style might have difficulties in learning if their learning style is not supported by the teaching environment. Several adaptive systems were developed which address this issue and aim at providing courses that fit to the individual learning style of students. Some examples of such systems include CS383 [2], IDEAL [14], MAS-PLANG [13], and INSPIRE [12].

Several learning style models are presented in literature such as the model by Kolb [10], Honey and Mumford [8], and Felder and Silverman [6]. In this paper, we focus on the Felder-Silverman learning style model (FSLSM) which is often used for providing adaptivity regarding learning styles in technology enhanced learning environments. FSLSM characterizes each learner according to four dimensions, each of them including 2 poles: active (A) learners learn by trying things out and working with others whereas reflective (R) learners learn by thinking things through and working alone. Sensing (Sen) learners like to learn concrete material and tend to be practical whereas intuitive (I) learners prefer to learn abstract material such as theories and their meanings and tend to be more innovative than sensing learners. Visual (Vis) learners remember best what they have seen whereas verbal (Ver) learners get more out of words, regardless whether they are spoken or written. Sequential (Seq) learners learn in linear steps and prefer to follow linear stepwise paths whereas global (G) learners learn in large leaps and are characterized as holistic.

For providing adaptivity based on learning styles, the learning styles have to be detected first. In most adaptive systems, this is done by asking students to fill out a questionnaire. Based on the students' answers, their learning styles can then be calculated. For what concerns the FS LSM, in order to detect both the preference itself and the degree of preference of learners for each dimension, the Index of Learning Styles (ILS) questionnaire has been developed by Felder and Soloman [5]. The ILS is a 44-item questionnaire which aims at identifying learning styles according to FSLSM.

Since the correct identification of learning styles is a crucial part in order to provide proper adaptivity, the reliability and validity of the learning style questionnaires are important issues. For being defined

reliable and valid, a questionnaire has to show internal consistency (that is, each set of items of the questionnaire has to measure a single latent variable) and validity (that is, the tool has to actually measure the latent variable that it is devoted at measuring).

However, a comprehensive study about learning styles conduced by Coffield et al. [4] showed that learning style questionnaires often lack this requirement. For what concerns the ILS questionnaire, a previous study [15] showed that ILS does not fully satisfy these requirements as well: in fact, it has been shown there that the poles of ILS can not be considered linearly independent, and the correlations between each others are in some cases very strong.

In this paper, we introduce an approach for learning style detection aimed at overcoming the limitations of the ILS questionnaire discussed before. The derived model is based on Multiple Correspondence Analysis (MCA) and incorporates the dependency of poles, latent dimensions and correlations between dimensions, which were found in our previous analysis.

Our proposed approach aims at improving authenticity of learner profiling by a fully data-driven approach [3]. The learner profile is here meant both as the detection of the most likely style for the learner, and the detection of the main characteristics of such profiles as they emerge from the answers given to ILS.

The paper is structured as follows: in the next section, the materials and methods are described; subsequently, we discuss the results of the approach and section 4 concludes our paper.

1. MATERIALS AND METHODS

1.1 Materials
The ILS questionnaire proposes a list of questions used to identify the learning style of each learner. The resulting index of preference for each dimension is expressed by an odd integer ranging [-11, +11], since 11 questions are posed for each of the four dimensions DIM={A/R, SEN/I, VIS/VER, SEQ/G}, with steps +/- 2. For each question q_i there are two possible values: +1 if the answer expresses the preference for the first style, -1 if the answer expresses the preference for the second style. The degree of preference for each dimension is just the algebraic sum of all values of the answers to the eleven questions.

The vector of indexes $I=\{i_{A/R}, i_{SEN/I}, i_{VIS/VER}, i_{SEQ/G}\}$ is a four feature vector containing the index of learning

styles for each dimension. is the set of questions belonging to each dimension.

For the analysis a dataset containing the ILS answers of 469 individuals collected at Vienna University of Technology and at Massey University in New Zealand was used.

To characterize this dataset, we have calculated the

frequency of occurrence of each pole of the ILS styles (table 1) and of the values for each answer (table 2). Table 1 considers all the learners showing an ILS score greater/less than zero, irrespective for the degree of preference. It can be noticed that:

a) the occurrence of each pole within the set of styles is not equally frequent with respect to all others;

b) the occurrence of each value for each answer is not equally frequent with respect to all others.

Table 1 – Absolute and relative frequencies for each FSLSM pole.

	A	R	Sen	I	Vis	Ver	Seq	G
F	260	209	286	183	400	69	220	249
%	.55	.45	.61	.39	.85	.15	.47	.53

Table 2 – Relative frequencies of occurrence of the value +1 for each ILS answer.

ACT/REF		SEN/INT		VIS/VER		SEQ/GLO	
q29	.76	q42	.55	q35	.52	q4	.29
q1	.77	q22	.58	q3	.84	q28	.27
q17	.38	q30	.58	q7	.77	q8	.39
q25	.49	q2	.66	q11	.76	q12	.71
q5	.51	q26	.43	q19	.83	q16	.62
q9	.57	q6	.68	q23	.83	q40	.47
q21	.39	q10	.37	q27	.73	q24	.40
q33	.52	q18	.75	q31	.77	q32	.57
q41	.41	q38	.66	q39	.66	q20	.57
q37	.58	q14	.51	q15	.41	q44	.64
q13	.39	q34	.36	q43	.80	q36	.51

These results are known by literature [7], nevertheless it can be hypothesized that are likely to impact on the definitions of learning styles and of their characteristics, and therefore should be taken into account when handling ILS data.

1.2 Methods
To realize a model that incorporates the different occurrence of each pole as well as each answer, we used Multiple Correspondence Analysis (MCA) technique. MCA [9, 11] is a well known multivariate technique for data dimensionality reduction and graphical exploration, especially for categorical data [1]. MCA works performing the optimal projection of a matrix in rows and columns space simultaneously on few dimensions (usually two), looking at independence of each value according to chi square metric. Chi-square metric is aimed at looking at the independence of both individuals and variables by comparing the differences between expected values

and observed ones.

Let M be the 469x44 matrix containing one row per learner and one column per ILS answers. The matrix M contains then positive and negative values expressed according to a binary scale. In order to make data consistent with the application of MCA, we transformed data on absolute frequencies as follows: for every question qi, Q=44, 2 numerical variables, namely the two answers to each question,

$$a_i^{(1)} = 1 \text{ if } q_i = 1 , 0 \text{ otherwise}$$

$$a_i^{(2)} = 1 \text{ if } q_i = -1, 0 \text{ otherwise}$$

were obtained. Let A be the 469 x 88 matrix containing in rows individuals and in columns the a_i, i=1,...,88.

Learners showing a preference for each of the eight styles, irrespective with the strength of preference, were selected and grouped (table 1). More formally, all the students showing $i_{DIM} > 0$ or $i_{DIM} < 0$, for each DIM respectively were selected and grouped according to their preferences. Then the frequencies of answers to each of all 88 answers were counted and divided by the cardinality of each group of students, grouped according to their preferences.

Let S be the matrix having in rows the frequencies of answers for each of the 88 answers and in columns the eight possible Felder-Silverman learning styles, i.e. the relative frequency of occurrence of a certain answer inside the set of learners classified as belonging to a certain style. Such a matrix is inversely weighted in rows by the frequency of each style, and in columns by the frequency of each answer.

Then S was decomposed according to Multiple Correspondence Analysis (MCA) algorithm, and the first two non trivial axes, i.e. the second and the third, were selected for the low dimensional representation. The first 2 non trivial axes, that is the axes 2 and 3, have been selected, according to the order of magnitude of the related singular values, that is 10-1, and achieving a cumulative variance – calculated over the non-trivial singular values - of about 64%. In this study, the MCA algorithm proposed by Greenacre [9] has been chosen. It consists on a Singular Value Decomposition (SVD) of the matrix S, weighted by positive definite diagonal matrices containing the inverse of the square roots of marginal frequencies. Once obtained the model, given by the selected axes, we projected A into the space. The vectors are treated as supplimentary variables, and therefore are inversely weighted in row before the projection [11]. The projected points represent the coordinates of each individual belonging to A on the MCA plane.

The characterization of each individual with respect to the styles can be addressed in terms of proximity between points on the plane. The closer the individual to each style, the stronger the impact of that style on the individual. To detect these influences, a suitable proximity measure is needed. The choice for a suitable proximity measure is an important issue for the effectiveness of the model. To select an appropriate and efficient proximity measure, we have tested different proximity measures, including euclidean distance, infinity norm distance, weighted euclidean distances, and cosines. Looking at the comparison of these proximity measures, we choose the cosine between the points representing styles and the points representing learners in the MCA plane, because of the stability properties it exhibits. For every point representing learners, the cosine with every point representing styles is calculated. Cosine is an angle-based normalized proximity measure that measures the association between two vectors in terms of linear dependence between them. The sign of the cosine indicates a positive association when positive, and a negative association when negative; the strenght of association is given by the absolute value of the cosine: the greater the absolute value, the stronger the association, both when positive and when negative.

2. RESULTS

2.1 Characteristics of the model

The results of MCA show that the matrix S has rank 5; except for the first trivial singular value that is equal to 1, two singular values have of order of magnitude 10^{-1}, and two singular values have of order of magnitude 10^{-2}.

Looking at the most represented answers on each axis of the model, the latent dimension underlying the first non trivial axis clearly refers to the visual or verbal preference: the positive semiaxis encodes the verbal preference, while the negative one encodes the visual preference. Nevertheless, other contributions in the negative semiaxis refer to the preference for a reversed social behaviour and for the need for having a whole vision of a subject; other contributions in the positive semiaxis refer to the preference for a regular learning process and to the attention for details.

The latent dimension underlying the second non trivial axis refers to the preference for: a) what students prefer to learn: concrete facts or abstract theories and concepts; b) the way in which students prefer to learn: bottom-up, inferential, regular, and careful with details or unregular, intuitive, top-down, and not careful with details.

The positive semiaxis encodes the preference for theories and unregular learning paths; the negative semiaxis encodes the preference for concrete material and regular learning paths.

Figure 1 shows graphically the MCA plane made by the second and third axes, and the position of the points representing styles on that plane.
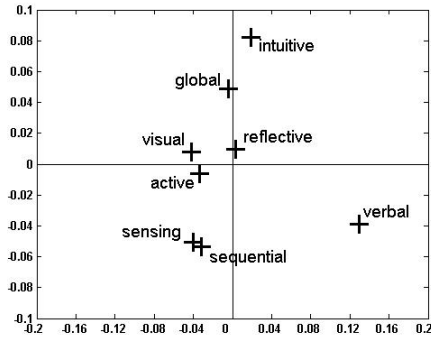
Figure 1: The eight styles on the MCA plane.

In Figure 1, the styles are arranged according to chi-square metrics, that measures the differences between expected values and observed ones. The values nearest to zero are the ones nearest to the expected ones. On the contrary, the values far from zero show greater differences. The more two point are close, the stronger is the association between them.

Therefore, as it can be seen from Figure 1, dependencies between styles exist. Looking at correlation coefficients for the columns of S, we find that the active and visual poles and the sensing and sequential poles are highly correlated (Pearson's coefficients .918 and .92 respectively for active-visual and sensing-sequential poles, p values both less than 10-4). Additionally, another strong correlation can be found between visual and sensing poles (Pearson's coefficients .918, p values less than 10-4) although it is not evident in Figure 1 (it is represented on the fourth axis). The dependencies found in this dataset are coherent with the ones described in [15], and should be taken into account when dealing with ILS-based learner modelling.

These dependencies between styles affect the reliability of detection of the preference of learners. In fact, the more two poles show associations – in terms of shared answers - between each others, the more difficult is to distinguish a clear preference of a learner for each of them. These dependencies are also reflected looking at the coordinates of the x-axis and of the y-axis. In fact, some styles are very close each others on one or both axes. This happens strongly between sensing and sequential, and between reflective and global on the x-axis; it happens less strongly on the x-axis between visual and active, and between intuitive and both reflective and global. On the y-axis, this happens strongly between sensing and sequential, and visual and reflective; it happens less strongly on the y-axis between verbal and both sensing and sequential, and between visual and both active and reflective. This closeness indicates shared characteristics of the styles, given by shared answers of learners belonging to each pole. As a consequence, the closer the styles are to each other, the more difficult to characterize clearly them with respect to all the other styles looking at the ILS answers.

2.2 Learners' profiles

Table 3 shows, for each pole of the ILS model, some results drawn by looking at the cosines between the points representing styles and the learners in the dataset. All the learners showing a cosine greater than zero, greater that 0.6, and greater than 0.8 have been selected for each pole. Then, for every set of learners grouped according to the cosines, the results have been compared with the number of questions answered according to a certain style, to investigate the behaviour of the model. Within these sets, the number of learners showing a number of answers greater than 5 for a certain style has been compared with the results (ILS>5). Moreover, the learners showing a number of answers equal to five according to the certain styles, are indicated (ILS =5). The percentage of agreement between the model presented here and the number of answers for every style has been used as a criterion for the reliability of the model. It should be noticed that while this criterion considers the number of answers, it does not takes into account at all which answers are given by the learners, and therefore this should be considered as a criterion on the whole.

Table 3 – The results obtained according with the values of cosines greater than zero (rows 2-6), greater that 0.6 (rows 7-11) and greater than 0.8 (rows 12-16) compared with values from ILS.

St	A	R	Sen	I	Vis	Ver	Seq	G
c>0	346	204	299	206	365	104	286	231
I>5	212	108	261	163	364	69	180	179
%	61.2	52.9	87.2	79.1	99.7	66.3	62.9	77.4
I=5	57	29	25	25	1	24	52	27
%	16.4	14.2	8.3	12.1	0.3	23	18.1	11.7
c>.6	266	128	225	134	269	67	212	157
I>5	171	71	210	121	269	59	155	131
%	64.3	55.4	93.3	90.3	100	88	73.1	83.4
I=5	42	17	12	9	0	7	37	16
%	15.7	13.2	5.3	6.7	0	10.4	17.4	10.2
c>.8	184	71	157	82	166	40	129	104
I>5	123	43	151	77	166	36	103	88
%	66.8	60.5	96.1	93.9	100	90	79.8	84.6
I=5	30	8	6	4	0	4	18	9
%	16.3	11.2	3.8	4.8	0	10	13.9	8.6

It can be seen from table 3 that the results given by the model, with respect to the aforementioned criterion, indicate an agreement:

a) in more that 70% of the cases when dealing with sensing, intuitive, visual and global styles;

b) in a percentage between 50% to 70% of the cases when dealing with verbal, sequential, active and reflective, that is the worst case.

4

Furthermore, while increasing the threshold given by the absolute values of the cosines, the performances increase considerably for verbal and sequential, while they do not increase considerably in all other cases, and in particular for active and reflective styles.

Eventually, it can be seen that for the styles listed in a) the percentage of strong disagreement between the model and the number of questions answered according to a certain style (ILS <5) is less frequent than for the styles listed in b), whatever the threshold of the cosine would be. The reflective style is hereby the worst case.

This behaviour can be motivated looking at the dependencies of styles and at the closeness of styles between each others on both axes. Reflective is in fact the style closest to zero; it is very close to visual on the x-axis and to global on the y-axis. For what concerns active, it is very close to sensing, and close to visual and sequential, on the x-axis; on the y-axis, it is close to visual, which is also the most variant group in the dataset, and therefore has a strong influence.

As a conclusion, it can be concluded that, on the whole, the model can be considered reliable for detecting all the styles, while less reliable for active and reflective poles; furthermore, the threshold of the cosine according to which the learners' profiles are selected can be considered a critical parameter, and need to be carefully checked.

2.3 Characteristics of the profiles

Table 4 – The most frequent answers given by the 25 learners closest to each style

	Act	Ref	Sen	Int	Vis	Ver	Seq	Glo
1	7a	3a	6a	3a	11a	15b	19a	23a
2	43a	34b	36a	34b	31a	31b	20a	7a
3	38a	10b	44a	10b	3a	41b	12a	3a
4	29a	26b	20a	6b	7a	4b	38a	8b
5	31a	28b	12a	26b	18a	35b	6a	28b
6	19a	23a	43a	28b	19a	14b	15b	4b
7	11a	4b	19a	4b	6a	26b	18a	43a
8	23a	31a	31a	23a	1a	33b	30a	19a
9	3a	35b	38a	35b	28b	20a	36a	10b
10	6a	6b	18a	8b	4b	34b	44a	6b
11	1a	13b	2a	43b	43a	10b	16a	26b

Table 4 shows the most frequent ILS answers given by the 25 learners that are closest to each style, according to the model. We look at the first 11 questions, since for each style 11 questions exist that are developed to indicate a preference for the respective style.

The results of table 4 confirm the dependencies within styles. Looking at the representative answers for the

profiles grouped as active (7a, 43a, 31a, 19a, 11a, 23a, 3a from visual style according to ILS; 38a and 6a from sensing style; 29a and 1a from active style), we can see that they indicate a profile with strong preference for visual material, but also with a preference for concrete materials and concrete learning experiences.

For the profiles grouped as reflective, answers from many different styles are present (3a, 23a, 31a belonging to visual style; 35b belonging to verbal style; 34b, 10b, 26b, 6b belonging to intuitive style; 28b, 4b belonging to global style; 13b belonging to reflective style) whereby only one answer belongs to the reflective style according to ILS. These answers indicate from one side the preference for visual material; from the other, the preference for abstract materials and for a reserved social behaviour, and the need for having a whole vision of a subject.

Looking at the profiles grouped as sensing (6a, 38a, 18a, 2a belonging to sensing style; 36a, 44a, 20a, 12a belonging to sequential style; 43a, 19a, 31a belonging to visual style), we can see that they indicate a profile with strong preference for visual materials, concrete materials and learning experiences, and for regular learning paths; moreover, they show a preference for a more sociable behaviour and for realism.

Looking at the profiles grouped as intuitive (3a, 23a belonging to visual style; 34b, 10b, 6b, 26b belonging to intuitive style; 28b, 4b, 8b belonging to global style; 35b, 43b belonging to verbal style), we can see that they indicate a profile with strong preferences for abstract materials and theories, and the need for having a whole vision of a subject. A preference for visual materials is also expressed, although it is not univocal.

For the profiles grouped as visual (11a, 31a, 3a, 7a, 19a, 43a belonging to visual style; 18a, 6a belonging to sensing style; 4a belonging to sequential style; 28b belonging to global style; 1a belonging to active style) the most representative answers indicate a preference for visual materials, together with a preference for concrete materials and learning experiences, and the need for having a whole picture of a subject.

Looking at the profiles grouped as verbal, answers (15b, 31b belonging to verbal style; 41b, 33b belonging to reflective style; 4b belonging to global style; 14b, 26b, 34b, 10b belonging to intuitive style; 20a belonging to sequential style) indicate the preference for verbal presentations of materials; moreover, it can be seen the preference for a reversed social behaviour and for learning abstract materials. The need for having a whole picture of a subject is also expressed, although it is not univocal.

For the profiles grouped as sequential, answers (19a belonging to visual style; 15b belonging to verbal style; 20a, 12a, 36a, 44a, 16a belonging to sequential style; 38a, 6a, 18a, 30a belonging to sensing style) indicate the preference for regular learning paths and for concrete materials; moreover, answers indicate the

need for having a detailed vision of a subject.

Looking at the profiles grouped as global, answers (23a 7a 3a 43a 19a belonging to visual style; 8b 28b 4b belonging to global style; 10b, 6b, 26b belonging to intuitive style) indicate a preference for visual materials; moreover, it can be seen a preference for abstract materials, and the need for having a whole vision of a subject.

These profiles are coherent with the latent dimensions underlying the first and second axis of the model, according to figure 1, and are all quite different from the ones underlying ILS.

As a conclusion, it can be stated that because of the reciprocal influences between styles, the profiles overlap each other; therefore they are not equally easy to distinguish in themselves, which is reflected by the model.

3. CONCLUSIONS AND FUTURE WORK
In this paper, an approach for profiling learners based on data from the ILS questionnaire is presented. The approach is data driven and is based on the analysis of the answers given to the ILS questionnaire.

The results show that the approach can be considered sufficiently reliable for detecting the preference of learners for a learning style when compared with the number of ILS answers given by the learner according to that style.

Additionally, this approach allows the exploration of the characteristics of the profiles in terms of similar answers given. The results show that, because of the dependencies between styles, some styles overlap each other. As a consequence, the characteristics of each profile as given by this model are quite different from the ones underlying ILS.

We plan to incorporate the learning styles characterization achieved here within Learning Environments, in order to improve them by providing proper adaptivity and personalization with respect to the profile of every learner. Therefore, future work will deal with improving adaptivity and personalization of learning environments by using the results achieved here.

REFERENCES
[1] Agresti, A.(2002) Categorical Data Analysis. Wiley Interscience, NY, 2nd Edition.

[2] Carver, C. A., Howard, R. A., and Lane, W. D. (1999). Addressing different learning styles through course hypermedia. IEEE Transactions on Education, 42, 1, 33-38.

[3] Cherkassky, V. and Mulier, P. (1998) Learning from data. Wiley Interscience, New York.

[4] Coffield, F., Moseley, D., Hall, E., and Ecclestone, K. (2004). Should we be using learning styles? What research has to say to practice. London: Learning and Skills Research Centre / University of Newcastle upon Tyne.

[5] Felder, R. M., and Soloman, B. A. (1997). Index of Learning Styles questionnaire. Retrieved 30 April, 2006, from http://www.engr.ncsu.edu/learningstyles/ilsweb.html.

[6] Felder, R. M., and Silverman, L. K. (1988). Learning and teaching styles in engineering education. Engineering Education, 78, 7, 674–681.

[7] Felder, R. M. and Spurlin, J. (2005) Applications, Reliability and Validity of the Index of Learning Styles. Int. J. on Engineering Education 21(1), pp. 103-112.

[8] Honey, P., and Mumford, A. (1982). The manual of learning styles. Maidenhead: Peter Honey.

[9] Greenacre, M. (1984) Theory and Applications of Correspondence Analysis. Academic Press, London.

[10] Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Englewood Cliffs, New Jersey: Prentice-Hall.

[11] Lebart, L., Morineau, M. and Warwick, K. M. (1984) Multivariate Descriptive Statistical Analysis. Wiley Interscience, New York.

[12] Papanikolaou, K. A., and Grigoriadou, M. (2003). An instructional framework supporting personalized learning on the web. Proceedings of the International Conference on Advanced Learning Technologies (Los Alamitos, CA), IEEE Computer Society, 120–124.

[13] Peña, C.-I., Marzo, J.-L., and de la Rosa, J.-L. (2002). Intelligent agents in a teaching and learning environment on the web. V. Petrushin, P. Kommers, Kinshuk & I. Galeev (Eds.), Proceedings of the International Conference on Advanced Learning Technologies (Palmerston North, NZ), IEEE Learning Technology Task Force, 21-27.

[14] Shang, Y., Shi, H., and Chen, S.-S. (2001). An intelligent distributed environment for active learning. ACM Journal of Educational Resources in Computing, 1, 2, 1-17.

[15] Viola, S. R., Graf, S., Kinshuk, and Leo, T. (2007). Investigating relationships within the Index of Learning Styles: A data-driven approach. International Journal of Interactive Technology and Smart Education, 4, 1, 7-18.