

Collaborative E- Learning Algorithm for Domain Knowledge Acquisition using Unbiased Matching

Monday Eze*

*Department of Computer Science/Mathematics/Statistics/Informatics, Federal University, Ndufu-Alike, Ikwo (FUNAI), Ebonyi State, Nigeria.

Abstract- The need to acquire domain knowledge at a fast pace cannot be overemphasized. Domain knowledge acquisition is not limited to any particular field, but rather, is a requisite path to giant strides in the sciences, technology and academics. Apart from being a major goal at all levels of academics, knowledge acquisition is inevitable in some fields of computing such as Artificial Intelligence and Expert Systems, where system construction is hinged on the ability of the system developers to elicit knowledge from the domain experts. Collaborative e-learning is a positive deviation from the traditional learning paradigm, in that knowledge is shared between the participants, through information technology. However, one of the challenges of collaborative learning is how to organize the team of learners, so as to maximize the learning impacts. Another important issue is, how to devise empirical strategies, for measuring the impact of e-learning methodologies. This work proposes a collaborative e-learning algorithm called the Community Bi-Partition Learning Model (CBLM). The algorithm applies a computational technique known as unbiased matching, to partition a learning community in order to speed up the process of knowledge acquisition. A validation approach for measuring the impact of the e-learning algorithm is also presented.

Index Terms- Collaborative e-Learning, Community Bi-Partition, Domain Knowledge, Knowledge Acquisition.

I. INTRODUCTION

According to [1], the illiterates of the 21st century are not those who cannot read and write, but those who cannot learn, unlearn, and relearn. Advancements in information and communication technology have caused a paradigm shift in the way knowledge is acquired or shared. A lot of research and industrial interests are increasingly focusing on such knowledge concepts as collective intelligence, collaborative learning, and so on. No wonder, Wikipedia could create the world's largest knowledge reservoirs from thousands of collective contributors worldwide [2]. The work being presented here is a collaborative e-learning algorithm for fast tracking domain knowledge acquisition. Though it was experimented on Project Management as a knowledge domain with a learning community of minimal size, this model is expected to work for a larger community, in diverse fields of learning.

E-learning is defined by [3], as the use of information and computer technologies to create learning experiences. Some of the common e-learning strategies enumerated by [3] are standalone courses, virtual-classroom courses, learning games and simulations, embedded e-learning, blended learning, mobile learning, and knowledge management. A number of researchers have identified some problems related to e-learning. The necessity for improved peer-to-peer learner-matching in a collaborative learning environment was emphasized by [4]. Another issue of importance raised by [5] is how to measure the value of e-learning experience using information technology. This paper tries to tackle these issues using community bi-partition, based on unbiased matching.

II. THE PROPOSED ALGORITHM

The Community Bi-Partition Learning Model (CBLM) is designed with the view of enhancing the process of domain knowledge acquisition through collaborative e-learning. The algorithm aims to increase the overall learning pace of a knowledge seeking community. The impact of a learning system is measured in terms of its ability to spread domain knowledge to an increased population of a learning community, at the shortest possible time. Like other collaborative learning systems, this model encourages knowledge sharing, which is an inevitable strategy for passing knowledge from the 'more knowledgeable' to the 'less knowledgeable' members of a learning community. For the sake of clarity, some relevant definitions will be presented here.

A *community* is a group of people interacting, in order to achieve a common goal [7]. A *learning community* is therefore, a group of people who have come together to learn, so as to acquire a particular domain knowledge. It is necessary to mention that a community could interact physically, or online. For instance, a group of selected employees of an organization participating in management training, or a group of undergraduates on excursion training, are both examples of a learning community.

The term is also used interchangeably with 'knowledge seeking community'. A *learner* is any member of a learning community. The term *graduation* depicts the completion of the learning process. In other words, graduation is the point where the participants are believed to have acquired the domain knowledge. The essence of learning as summarized in the block diagram in Fig 1, is therefore to transform a knowledge seeking community into a knowledgeable (graduate) community.

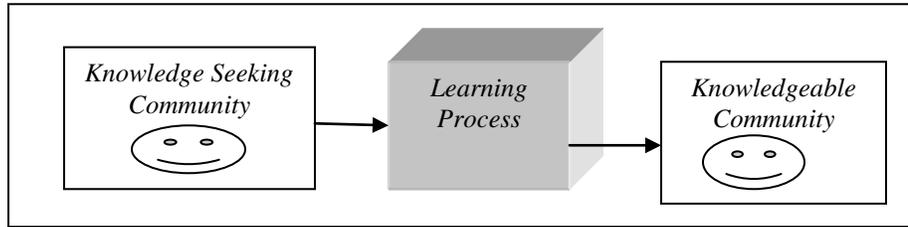


Figure 1: Block Diagram for the Essence of Learning

The *population* of a learning community is the total number of learners enrolled into the learning process. *Domain knowledge* [8] is the knowledge related to a particular niche area. For instance, a learning community may delve into the study of Aircraft Engineering, Tropical Infectious Diseases, Banking Operations, Linear Algebra, and so on. The term *homogenous* refers to a community that is in its natural form. On the contrary, a *partitioned community* is one that has been partitioned [9], using a requisite *bi-partition algorithm*. Such an algorithm breaks the learning community into two partitions - the more knowledgeable (*MK*) set and the less knowledgeable (*LK*) set respectively. After bi-partition, a computational technique called *unbiased matching* is applied to pair up the members of the *MK* set with an appropriate member of the *LK* set, in order to maximize the e-learning impact in the community. The term *unbiased matching* stems from the fact that the matching operation is fairly *autonomous*. In other words, the process of drawing the *MK* and *LK* pairs is neither based on an external human intervention, nor any form of random computer operation. On the contrary, it is purely based on the individual performances of the learners in a preliminary assessment test. The detailed system designs and model runs of the overall algorithm will be presented at this point.

III. THE SYSTEM DESIGN AND MODEL RUNS

A. Workflow Design

The workflow design for the proposed system is shown in Fig. 2. The diagram is made up of four main symbolic components. These are, the rectangular boxes (HL-Com, GL-Com, and PL-Com) representing the three community variations, the cuboids (C-Room, P-Exam, B-PartOP and PMC-Room) representing a variety of learning operations, the cylindrical figure (X-Result) which is an output data representing the pre-partition examination results, and eight arrows, labeled R, X, 1, ... 6, which indicate the model flow. The term pre-partition examination depicts an assessment, whose result forms the basis for community bi-partitioning.

As shown in the workflow, the three communities HL-Com, GL-Com, and PL-Com represent the homogeneous learning, the graduate learning, and the partitioned learning communities respectively. The homogeneous learning community is made up of the people enrolled into the collaborative e-learning from onset, while the graduate learning community comprises of the knowledgeable community; that is those who have graduated. On the other hands, the partitioned learning community is the structured community, arising from the bi-partition operation. The unbiased matching operation is usually performed on the partitioned community. It is important to note that the community population P_C does not change, unless someone drops out of the learning community prior to graduation.

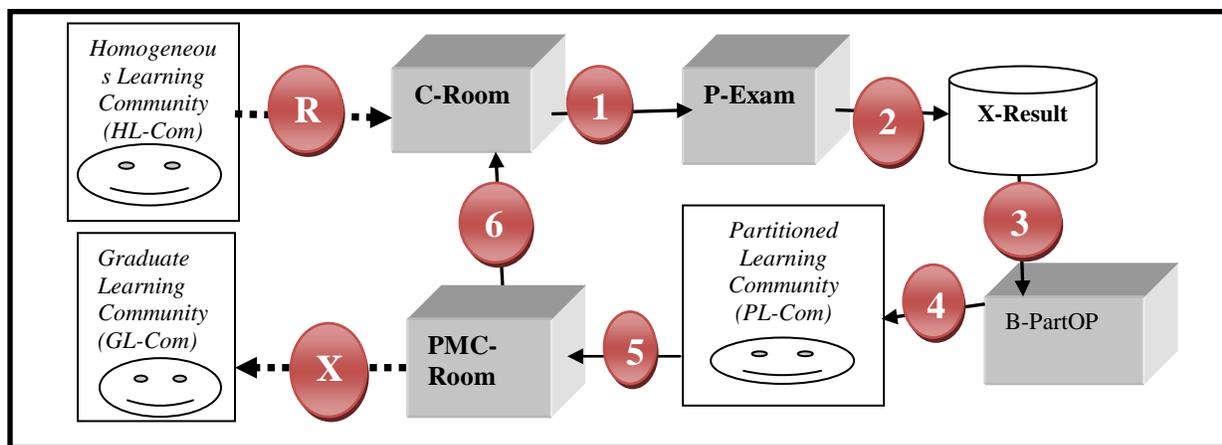


Figure 2: The Community Bi-Partition Learning Model Workflow

As shown in the workflow diagram, the HL-Com is enrolled into the system through the entry path R, while the graduates GL-Com leave the system through the exit path X. On entry into the e-learning system, HL-Com passes through the C-Room operation, which represents the class room. This is an active e-learning session, after which there is a preliminary examination, P-Exam. The assessment result, X-Result is used as an input into the bi-partition operation, B-PartOP, which transforms the original homogenous community into the partitioned learning community, PL-Com. After the unbiased matching, the learners undergo a post match learning operation called PMC-Room (post match class room). This is a one-on-one revision of the previous e-learning contents by the unbiased partners. The essence of the PMC-Room is, to ensure that domain knowledge is passed from the more knowledgeable (MK) to the less knowledgeable (LK) members of the community.

It is important to state that the arrow links from 1 to 6 form what is termed in this research as a *single session* of e-learning. The overall e-learning duration could comprise of a number of learning sessions. The current research is based on three e-learning sessions.

B. Experimental Design and Model Runs

The experimental design diagram is shown in Fig. 3. A total of 30 university undergraduates, within the age brackets of 18 to 25 years were randomly selected, and enrolled for the collaborative e-learning experiment. This population forms the homogenous learning community.

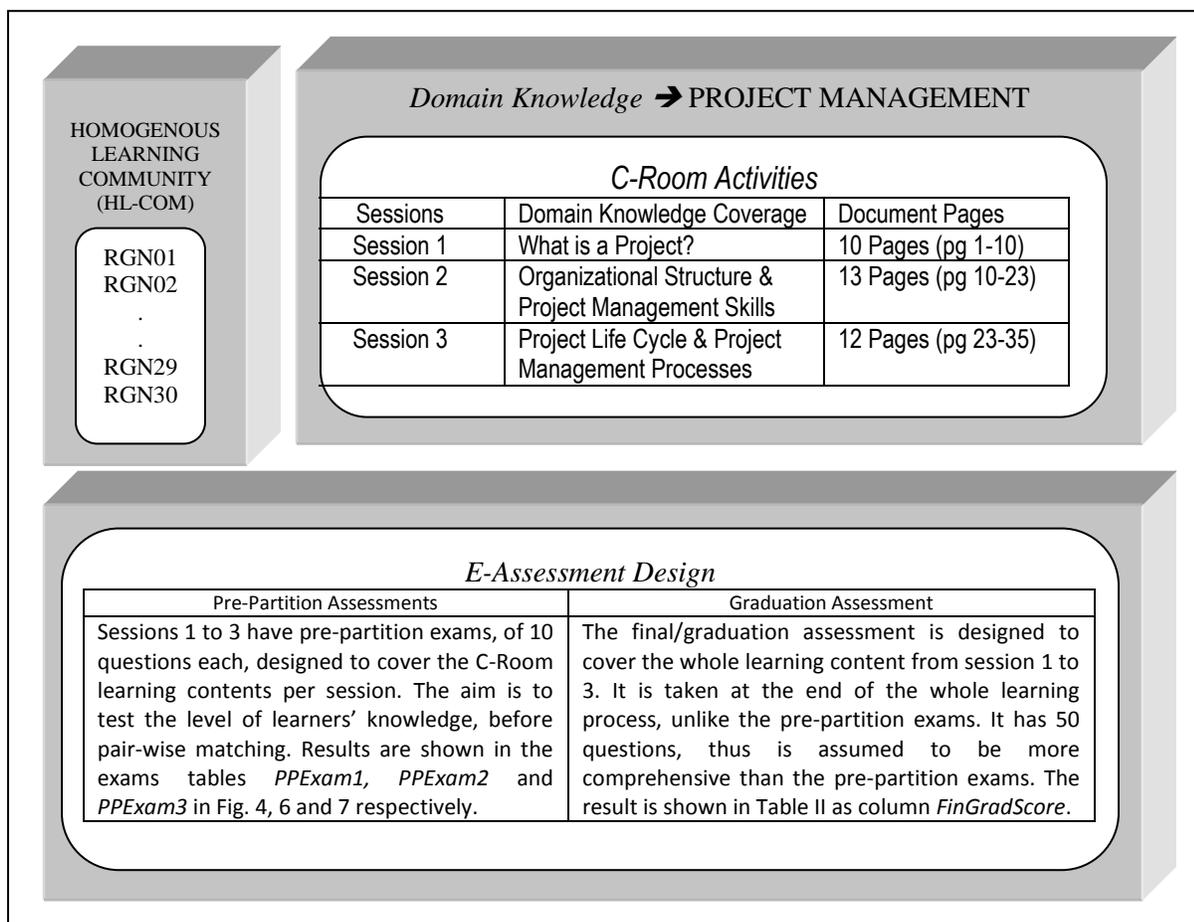


Figure 3: Collaborative e-Learning Experimental Design

As indicated in the diagram, the learners are assigned unique registration numbers RGN01, RGN02, ... RGN30, which remain permanent, throughout the learning duration. The chosen knowledge domain for the experiment is Project Management. The learning content for this experiment is drawn from a Project Management Professional (PMP) Certification book by [6]. The soft copy of this document was downloaded in PDF format. This online text book has 614 pages; though, only Chapter 1 which is made up of 35 pages was used for the learning experiment. The collaborative e-learning duration was divided into three sessions, with each session covering specific areas of the knowledge domain, as already indicated in Fig. 3.

The results of the first e-learning session are shown in Fig. 4. The examination table PPEXam1 consists of two columns - RegNum for the learners' registration numbers, and ESco1 for the corresponding pre-partition exam scores.

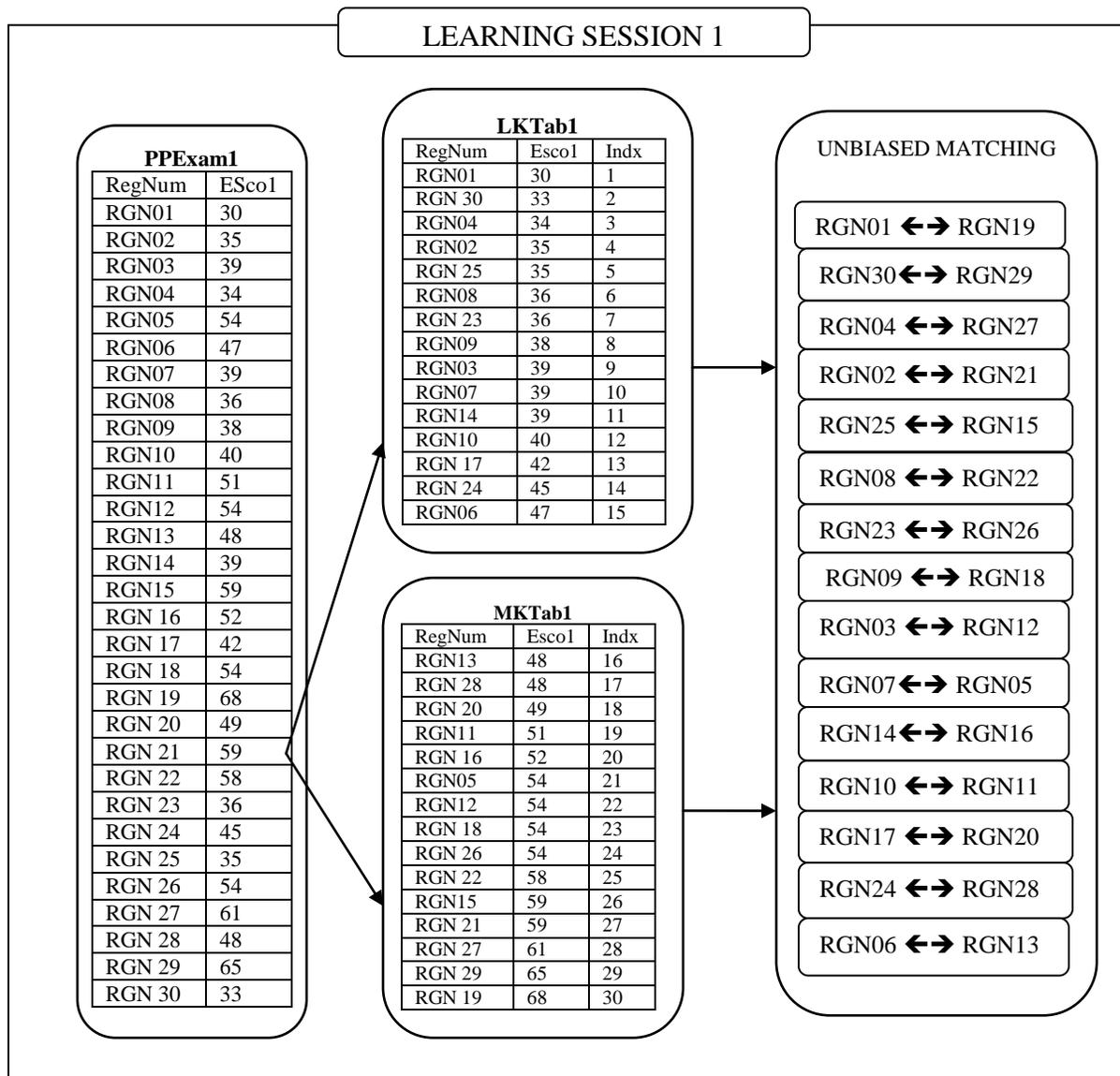


Figure 4: Experimental Result of Learning Session 1

The system takes PPEXam1 as an input, and then creates an intermediary version of the table, having three columns. The third column which is designated as *Indx*, sorts the pre-partition scores in an ascending order of magnitude. The resulting intermediary table is then partitioned into two separate tables, using the two bi-partition rules. The first bi-partition rule, which is shown in equation (1), determines the number of teams that could be formed. Here the key parameter is, whether the population size of the learning community P_C is numerically odd or even. The second bi-partition rule, which is shown in Fig. 5 as a pseudo code, partitions the homogenous learning community (HL-Com) into two groups - the *LK* and *MK* sets.

$$NLT = \begin{cases} \frac{P_C}{2}, & \text{If } P_C \text{ is EVEN} \\ \frac{(P_C - 1)}{2}, & \text{If } P_C \text{ is ODD} \end{cases} \quad (1)$$

where P_C is the population size and NLT is the number of learning teams.

As already mentioned, the LK set consists of the lowest knowledgeable persons, while the MK consists of the most knowledgeable persons in the class. The classification of learners into these sets is purely on the basis the pre-partition assessment result. Thus, a learner could be in one set in a particular session, and then drop into the opposite set in another session, depending on the prevailing performance. The two sets for the e-learning session 1 are tables LKTab1 and MKTab1 respectively, as shown in Fig. 4.

```

LKPartition Set Formation
  If PC is EVEN,
    LKPartition = {All the Learners with indices 1 to N/2}
  ElseIf PC is ODD then
    LKPartition = {All the Learners with indices 1 to (N-1)/2 }
    PLUS the Triplet Partner having indices N
  EndIf

MK Partition Set Formation:
  If PC is EVEN,
    MKPartition = {All the Learners with indices [N/2]+1 to N }
  ElseIf PC is ODD then
    MKPartition = {All the Learners with indices (N+1)/2 to (N-1)}
  EndIf
    
```

Figure 5: Pseudo Code for the Bi-Partition Rule 2

The strength of the overall model is that the system always takes cognizance of the most knowledgeable, and the least knowledgeable set of learners, thus ensuring that knowledge is shared through the process of unbiased matching. The formation of the matching pairs is based on the *unbiased matching equation* given by equation (2).

$$Indx_{LK} + Indx_{MK} = P_C + 1 \quad (2),$$

where $Indx_{LK}$ is the index of the partner from the LK set, $Indx_{MK}$ is the index of the partner from the MK set, and P_C is the numerically even population size.

The *verification table*, satisfying the unbiased matching equation for the session 1 experiment is Table I. Given the population size of the learning community $P_C = 30$, the pair-wise indices for the LK and MK partitions always adds up to 31, which is the same as $P_C + 1$.

Table I: Matching Verification Table

$Indx_{LK}$	1	2	3	4	5	6	15
$Indx_{MK}$	30	29	28	27	26	25	16
$Indx_{LK} + Indx_{MK}$	31	31	31	31	31	31	31
$P_C + 1$	31	31	31	31	31	31	31

The matching equation (2) takes a modified format, when the population size is odd. This second matching equation, otherwise called the *unbiased triplet rule* is given by equation (3). Thus, apart from forming the pair-wise partners as earlier discussed, one of the teams is promoted to a triplet, having three partners.

$$Low(Indx_{LK}) + SecHigh(Indx_{MK}) + High(Indx_{MK}) = 2 * P_C \quad (3)$$

where the symbol *Low* stands for ‘the lowest’, *SecHigh* stands for ‘the second highest’, and *High* stands for ‘the highest’ indices respectively.

The interpretation of equation (3) is that the learners with the lowest index from LK set, the second highest index from MK set, and the highest index from MK set form the unbiased triplets. Furthermore, the sum of their corresponding indices is always equal to double the community population size P_C . For instance, if the community population size is $P_C = 89$, which is an odd number, then the unbiased triplets consists of the learners, having the indices $Indx=1$, $Indx=88$ and $Indx=89$ respectively.

The computational implementation of unbiased matching for a numerically odd population size involves the following steps. After the indexing operation, the system excludes the highest indexed learner from the community, thus giving rise to an intermediary population size ($P_C - 1$), which is even. The bi-partition rules in equations (1) and (2) are used to form the pair-wise partners. Thereafter, the learner with highest index, who was originally excluded, is then promoted to join the first pair-wise partners, thus giving rise to a single triplet partner, in line with equation (3).

The formation of the unbiased partners is carefully designed such that knowledge transfer is highly encouraged, both in pair-wise and triplet partnership cases. For instance, the third candidate promoted to form the triplet partnership is usually the most knowledgeable candidate in the community, based on the pre-partition assessment test. Also one of the triplet partners is the least knowledgeable in the community. Thus, the collaborative algorithm does a purposeful matching of these candidates, so as to enhance knowledge flow from the most knowledgeable, to the least knowledgeable partners in the learning community. The details of the outcomes of the e-learning sessions 2 and 3 are shown in Fig. 6 and 7 respectively.

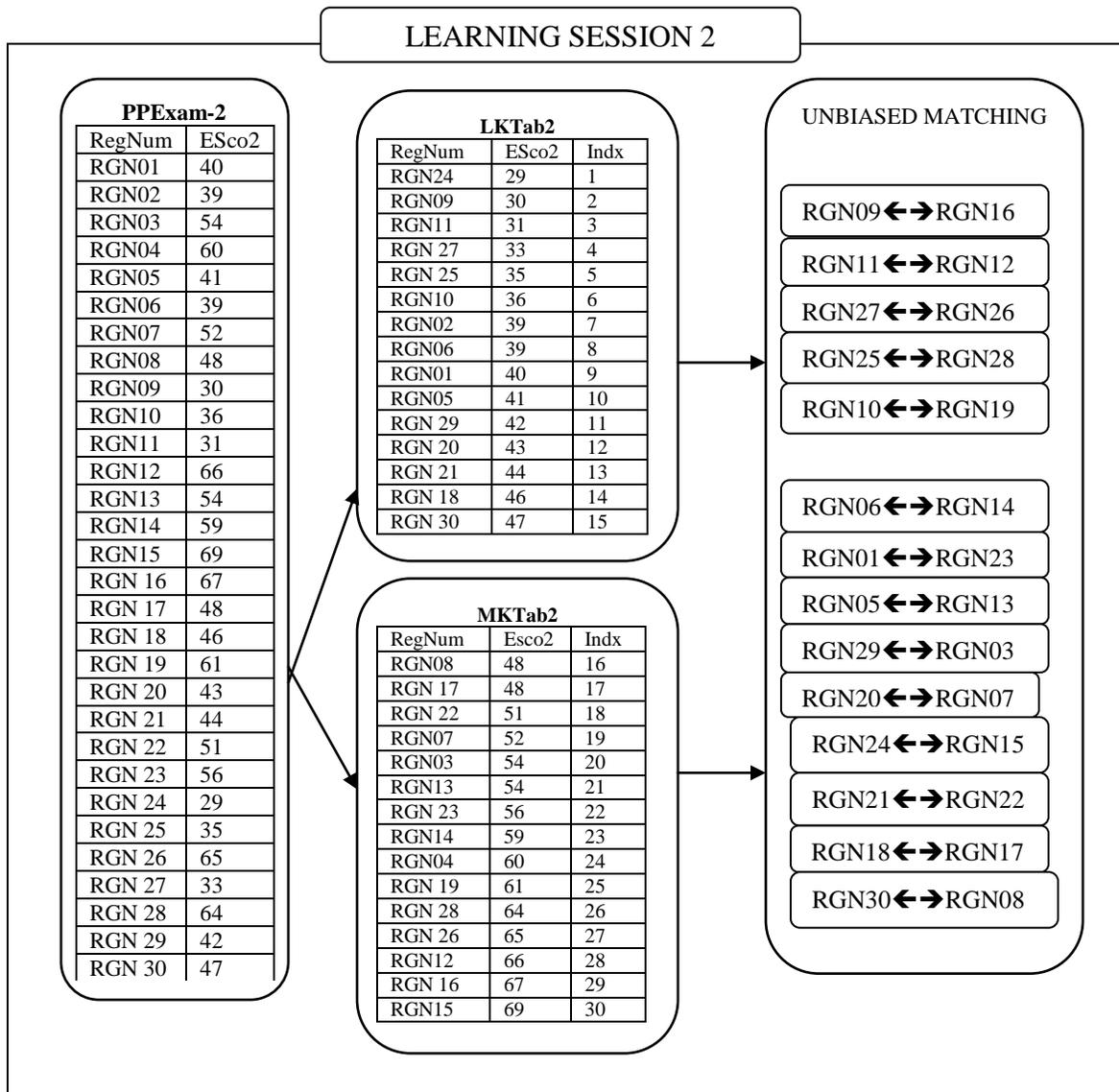


Figure 6: Experimental Result of Learning Session 2

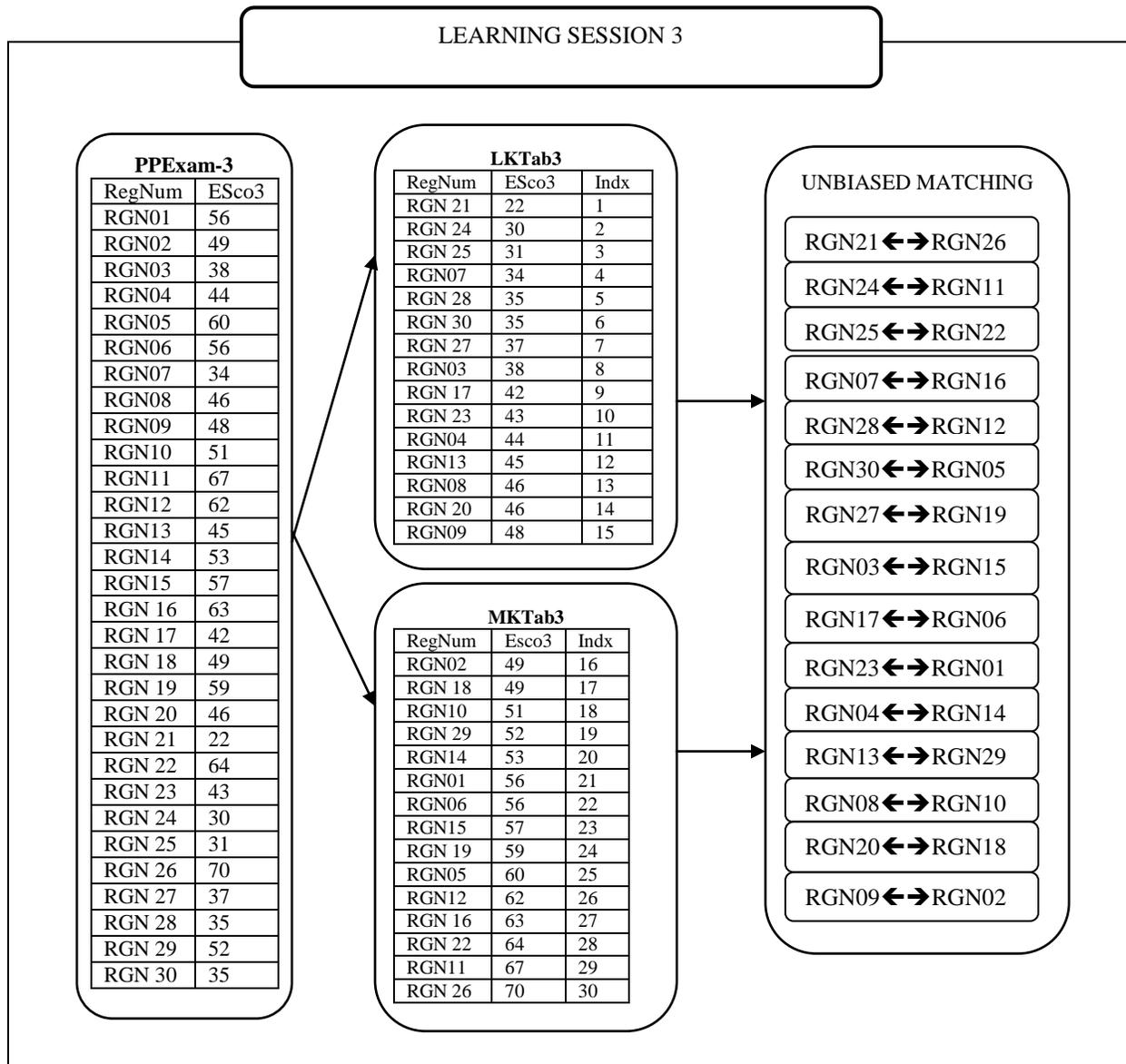


Figure 7: Experimental Result of Learning Session 3

IV. SYSTEM ANALYSIS AND VALIDATION OUTPUTS

The workflow for validating the system performance is shown in Fig. 8. The strategy adopted is to first, derive the average pre-partition score using equation (4).

$$AvgPreGradScore = \frac{ESco_1 + ESco_2 + \dots + ESco_s}{S_N} \quad (4),$$

where $ESco_1, ESco_2, \dots, ESco_s$ are the pre-partition scores per session, which have been shown in Fig. 4, 6 and 7 respectively; S_N is the number of e-learning sessions, which is 3 in the current study.

Next, the graduate learning community (GL-Com) is passed through a final graduate assessment test, designated as G-Exam in the workflow. The result of this is designated as *FinGradScore*.

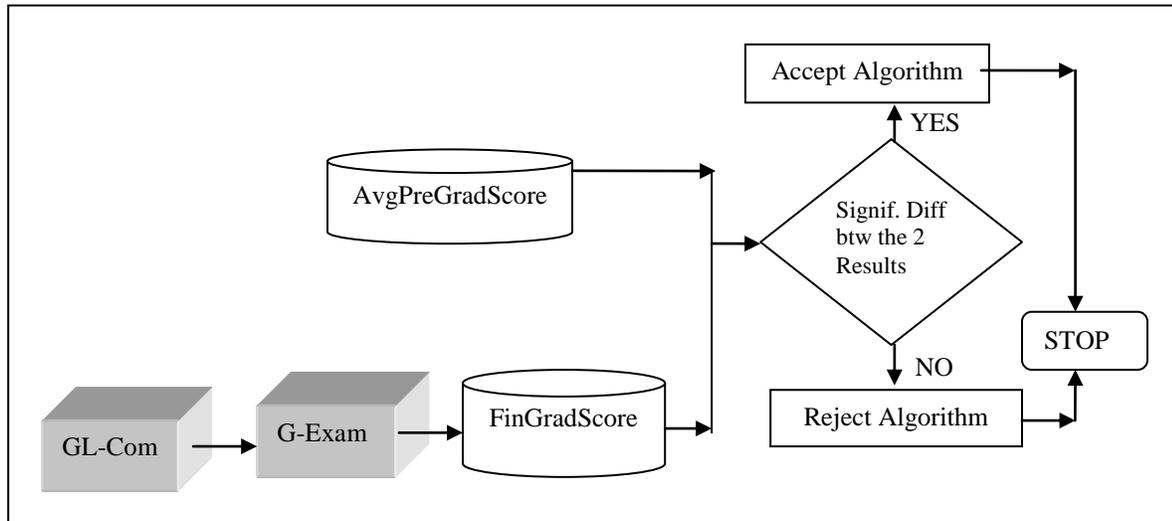


Figure 8: Model Validation Workflow

The basis of system validation is therefore, to compare the *AvgPreGradScore* with the *FinGradScore*. Since the *AvgPreGradScore* measures the level of competence before graduation, while *FinGradScore* represents the learner’s competence after graduation, it is expected that there should be a significant difference between the two scores. The Global Assessment Table (GAT) shown as Table II is used for a number of performance analysis, as will be discussed.

A. Least IQ Improvement Analysis

The term ‘*Least IQs*’ in this research refers to each of the learners who had the lowest domain knowledge, for each of the learning sessions. These are the learners who had the lowest indices in each of the LK sets, based on the pre-partition exams. From the tables LKTab1, LKTab2 and LKTab3 in Fig. 4, 6 and 7, the learners in this category are, RGN01 for session 1, RGN24 for session 2 and RGN21 for session 3 respectively. The Least IQ Improvement Analysis is used to investigate whether there was significant value added to the learners, after completion of the knowledge acquisition training. To measure the impact of collaborative e-learning, the *AvgPreGradScore* is compared with the *FinGradScore* for all the Least IQs. The assessment scores for the learners in the *Least IQs* detailed in the Global Assessment Table (GAT), are used to plot the improvement analysis chart shown in Fig. 9. From the chart, each of the learners in the *Least IQ* category improved their domain knowledge by at least 20 units as a result of the learning process.

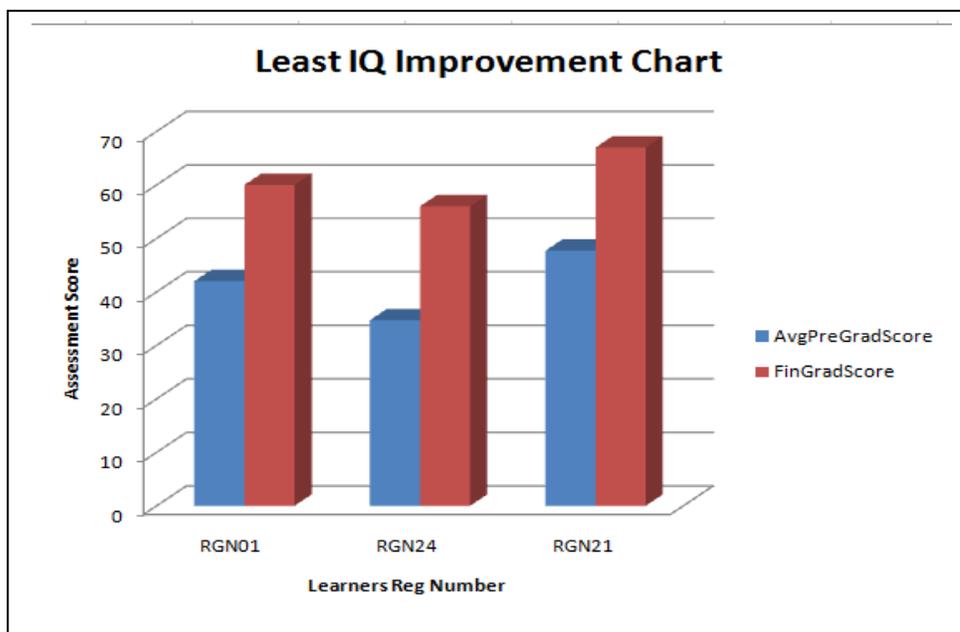


Figure 9: Least IQ Improvement Analysis Chart

B. Low Consistency Improvement Analysis

The term ‘Low Consistency’ in this research refers to the set of learners who, may not necessarily be among the *Least IQs*, but were however consistently in the LK set, throughout the duration of the e-learning. From the experimental data, the three learners RGN09, RGN25 and RGN30 remained in the LK tables LKTab1, LKTab2 and LKTab3, throughout the learning period. Thus, they are in this category. Fig. 10 is the analysis chart depicting the level of improvement for the learners in the ‘Low Consistency’ category.

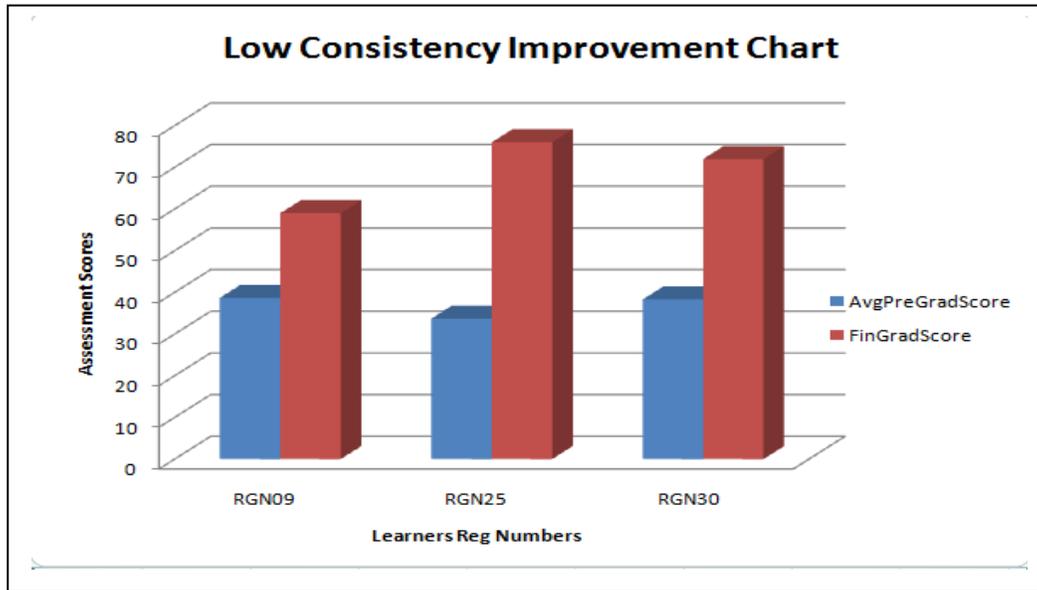


Figure 10: Low Consistency Improvement Chart

This improvement analysis chart is also derived from the Global Assessment Table (GAT), by comparing the *AvgPreGradScore* with the *FinGradScore* for learners concerned.

C. Global Gap analysis

A gap analysis was done, by comparing the *AvgPreGradScore* with the *FinGradScore* for the whole community. The result of this is shown in Fig. 11, where the x-axis is calibrated as the registration numbers, while the y-axis represents the assessment results. As indicated in the figure, the blue coloured graph represents the *AvgPreGradScore*, while the red coloured one represents the *FinGradScore* plot. The fact that there is a clear demarcation or gap between the two graphs for the whole community implies that, there is a global improvement after the learning process.

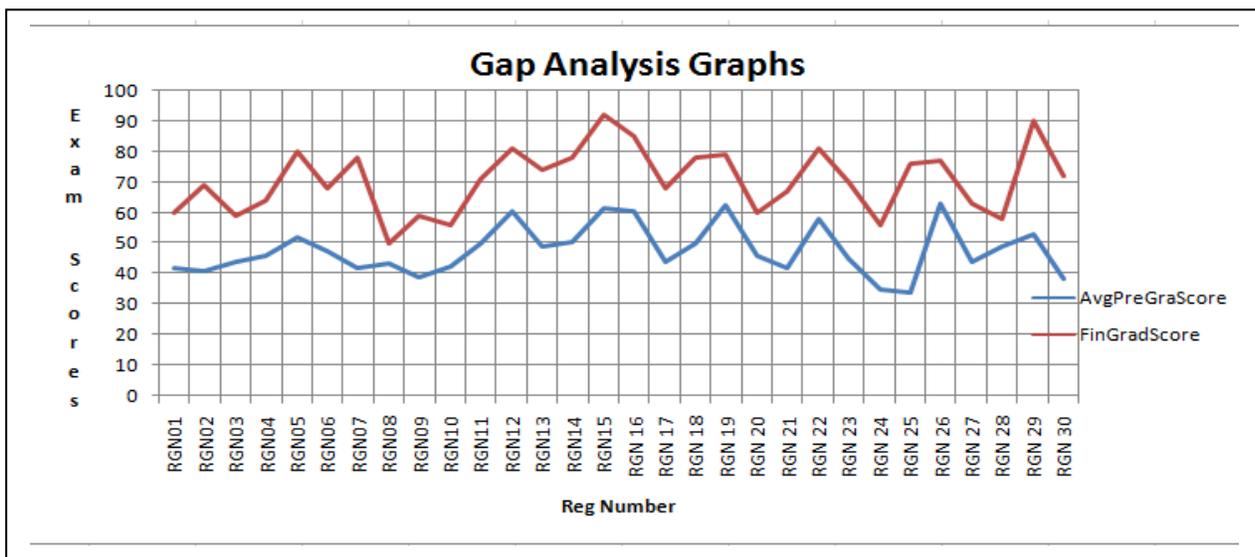


Figure 11: Gap Analysis Result

V. DISCUSSION AND CONCLUSION

Apart from encouraging collaborative learning and knowledge sharing through the implementation of a matching algorithm, this work also presents practical strategies for validating e-learning models. Fig. 12 is a final output which summarizes the overall benefit of this work to a learning community. The figure which is termed as *Collaborative e-Learning Benefit (Celeb) Graph* is derived by plotting the *Diff* column of the Global Assessment Table (GAT) against the corresponding learners’ registration numbers. The *Diff* column calculates the overall difference between the *AvgPreGradScore* and the *FinGradScore* for the whole e-learning duration.

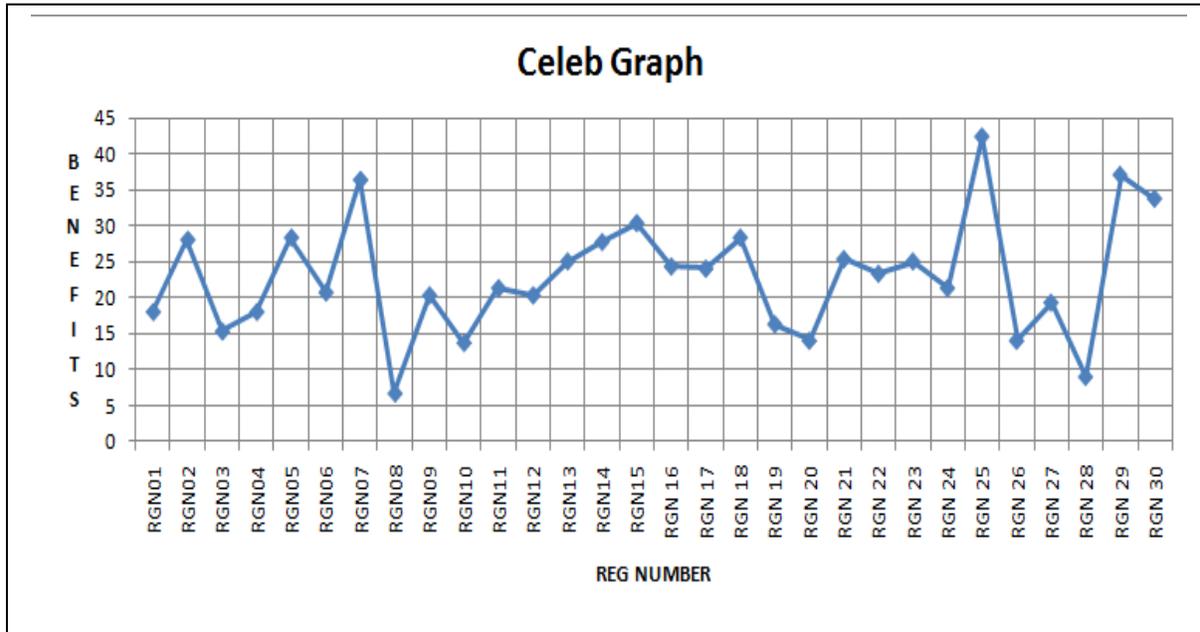


Figure 12: Benefit Analysis Result

Based on the Celeb graph, one can easily pinpoint that RGN25 has the highest benefit from the domain knowledge acquisition, followed by RGN29 and RGN07 in that order. In conclusion, this algorithm is has been fully presented in such a way that it can be easily implemented in a number of computing environments. It is also possible to adapt it in diverse learning environments, including academic institutions, business training organizations, among others.

Table II: Global Assessment Table (GAT)

Reg	E1	E2	E3	AvSc	FgS	Diff
RGN01	30	40	56	42.0	60	18.0
RGN02	35	39	49	41.0	69	28.0
RGN03	39	54	38	43.7	59	15.3
RGN04	34	60	44	46.0	64	18.0
RGN05	54	41	60	51.7	80	28.3
RGN06	47	39	56	47.3	68	20.7
RGN07	39	52	34	41.7	78	36.3
RGN08	36	48	46	43.3	50	6.7
RGN09	38	30	48	38.7	59	20.3
RGN10	40	36	51	42.3	56	13.7
RGN11	51	31	67	49.7	71	21.3
RGN12	54	66	62	60.7	81	20.3
RGN13	48	54	45	49.0	74	25.0
RGN14	39	59	53	50.3	78	27.7
RGN15	59	69	57	61.7	92	30.3
RGN16	52	67	63	60.7	85	24.3
RGN17	42	48	42	44.0	68	24.0
RGN18	54	46	49	49.7	78	28.3
RGN19	68	61	59	62.7	79	16.3
RGN20	49	43	46	46.0	60	14.0
RGN21	59	44	22	41.7	67	25.3
RGN22	58	51	64	57.7	81	23.3
RGN23	36	56	43	45.0	70	25.0
RGN24	45	29	30	34.7	56	21.3
RGN25	35	35	31	33.7	76	42.3
RGN26	54	65	70	63.0	77	14.0
RGN27	61	33	37	43.7	63	19.3
RGN28	48	64	35	49.0	58	9.0
RGN29	65	42	52	53.0	90	37.0
RGN30	33	47	35	38.3	72	33.7

Key to GAT Columns: Reg=RegNum, E1=ESco1, E2=ESco2, E3=ESco3, AvSc=AvgPreGradScore, FgS=FinGradScore

REFERENCES

- [1] M. Rosenberg, "E-Learning Strategies for Delivering Knowledge in the Digital Age", McGraw-Hill, New York, 2001
- [2] T.W. Malone, R. Laubacher, & C. Dellarocas, "The Collective Intelligence Genome", Vol. 51, No.3, 2010.

- [3] W. Horton, "E-Learning by Design", Pfeiffer Publishers, San Francisco, 2006.
- [4] R. Nagarajan, "Introduction to E-Learning: V Semester Course Material", Centre for Distance Education, Anna University Chennai, 2008
- [5] S. Naidu, "E-Learning: A Guidebook of Principles, Procedures and Practices", Commonwealth Edu. Media Center for Asia (CEMCA), New Delhi, 2006
- [6] K. Heldman, "Project Management Professional (PMP) Examination Study Guide", 5th ed, Wiley Publishing Inc, Indiana, 2009
- [7] J. Bacon, "The Art of Community: Building the New Age Participation", O'Reilly Media Inc, Sebastopol, 2009
- [8] D. Andrzejewski, Z. Xiaojin, & M. Craven, "Incorporating Domain Knowledge into Topic Modeling via Dirichlet Forest Priors". In the Proc. of the 26th Int. Conf. on Machine Learning, Montreal Canada, 2009.
- [9] B. Siddhartha, R.G. Bibek, C. Rupash, B. Arnab, R. Suvadip, & B. Amit, B. "Circuit Bipartition Using Genetic Algorithm". Int. Journal of Adv. Eng. & Global Tech., Vol 1, Issue 3, Oct. 2013
- [10] PeopleSoft, "PeopleTools 8.12 PeopleSoft Workflow PeopleBook", 2001.

AUTHORS

Author – Eze, Monday Okpoto (MBA , MSc , PhD) is currently a Computer Science Lecturer / Academic Adviser in the Department of Computer Science/Mathematics/Statistics/Informatics, Federal University, Ndufu-Alike, Ikwo (FUNAI), Ebonyi State, Nigeria. E-mail: eze_monday@yahoo.com, Mobile : +234 8028669172.