Disrupting the Rote Learning Loop: CS Majors Iterating Over Learning Modules with an Adaptive Educational Hypermedia

Muhammad Mustafa Hassan[1,2(✉)] [ORCID] and Adnan N. Qureshi[1]

[1] University of Central Punjab, Lahore 54000, Pakistan
{Mustafa.hassan,dr.qureshi}@ucp.edu.pk
[2] University of Eastern Finland, 80101 Joensuu, Finland

Abstract. The rote learning problem has plagued the education systems of developing world since long. To name a few, improperly designed assessments, teachers' authority, rewarding verbatim answers, sheer class sizes, and individual learner differences are amongst the most notable mediators. The authors report on the design and development of an adaptive educational hypermedia, which disrupts the rote learning loop by hitting a few of the aforementioned reasons. The reported system provides a personalized learning experience to each learner, adapting on the basis of cognitive and learning styles. Further, the assessments are designed in a way that they loop each failed learning via variated paths, hence eliminating chances of rote learning. Moreover, the failed perturbations are traced back to the problematic domain segment for further knowledge acquisition. In-situ evaluations of the system with end-users (real students of Bachelor of Science in Computer Science) reveal a difference between control and experimental groups. The effect size is however moderate.

Keywords: Rote learning · Adaptive educational hypermedia
AEH · Contextualization · Adaptor-innovator model · Cognitive styles
Learning styles · ESL

1 Introduction

The problem of rote learning has since long plagued educational communities around the globe. The learning approaches adopted by students, in general, have direct relationship with understanding of the concepts and subsequent performance [1], and the rote learning style, in particular, degrades the quality of learned knowledge [2]. Nonetheless, rote learning—bare memorization—in itself is not a problem. It is an efficient learning technique needed in some learning contexts which need memorization [3], for example remembering alphabet of a language. However, it becomes a problem when a student adopts it as her primary learning strategy, applied to most—if not all—of her learning contexts [3]. The said student is mostly compelled to choose rote learning because it reduces cognitive load otherwise endured in understanding complex concepts, a situation commonly aroused in science subjects. However, in the long term, the knowledge learned through rote learning has less retention span when compared to other learning methods, like that of meaningful learning [4].

The recent research in educational psychology suggests that more abstract concepts are more negatively affected by rote learning approach [5]. Similarly, disciplines relating to problem solving are heavily affected by this learning technique [3]. The problem—though persistent throughout science spectrum—exhibits itself especially in mathematics and computer science, being the fields of higher abstraction applied to problem solving [6]. Further, this very mechanism of learning can even contribute to development of non-interest in science subjects [7].

Contrary to the popular belief, rote learning is not abundant in only developing countries [5]. Even after 30 years of education reforms in developed world, rote learning problem still prevails in many countries, for example in USA and Sweden [3]. The recent research tells that the students still choose rote learning when working with complex concepts [6]. However, the dynamics are different from the developing world where the pupils are trained to become rote learners from the very early age [8]. The researchers note that the educational system in developing countries, from teaching to assessment supports rote learning [9]. The students are rewarded for the reproduction and imitation of the given concepts verbatim [10]. Especially in Pakistan—where this research is carried out—it is frequently reported that learners are trained to reproduce what has been articulated to them [10]. They seldom use their own intuition in problem-solving of any kind, merely applying imitations of solutions earlier learned [8].

In this work, the authors report on the design, development, and testing of an Adaptive Educational Hypermedia (AEH) system to disrupt the rote learning loop of verbatim repetition and reproduction of concepts. The system is envisioned and developed for CS majors in Pakistan and is situated in their very own context. The paper also report in-situ evaluations of first prototype of the proposed system in real environment with end-user learner. The rest of the report is structured as following. The next section reviews the related work, followed by a discussion on the proposed innovation's architecture. Section 4 details the research methodology used to experiment with the prototype of AEH. The results of the analysis are presented in Sect. 5, with discussion in Sect. 6. Finally, the authors conclude the paper in Sect. 7.

2   Related Work

Liu and Hmelo-Silver [11] report on the design and development of educational hypermedia to promote meaningful learning of complex systems in science students. They argued that conventional learning methodologies for complex systems did nothing more than piling up the information in learners' head. To resolve the issue, two different hypermedia versions were tested with 7th graders and pre-service teachers. The authors found that both type of hypermedia support avoidance of rote learning.

Jacobson and Archodidou developed Knowledge Mediator Framework (KMF) [12]. With a proof-of-concept applied to high school students in learning neo-Darwinian evolutionary biology, they tested the efficacy. Their results showed a significant improvement in students' progress, as well as in learning patterns. They noted that the students started developing expert-like models in their solutions.

Though not in a strict hypermedia sense, Zydney and Grincewicz experimented with a multimedia learning environment with videos to enhance students' meaningful learning abilities [13]. Their study found that the amount of time students spent with the system was a predictor of their performance.

Rum and Ismail [14] used metacognitive tools to assist students in learning programming in meaningful ways. They devised six different strategies, however all metacognitive. The tools were implemented with the help of an educational hypermedia. They enrolled 30 participants in experimental group and 36 in control group. The experimental group exhibited performance improvement over the control group.

3   The Proposed Innovation

The proposed AEH is composed of 5 modules, namely student model, assessment engine, adaptation controller, content store, and the interface. The students communicate with the interface, which presents learning/assessment activities to the learner, formatted and selected by the adaptation controller. The adaptation controller works with the information stored in the student model to appropriately select and format the content which is stored in the content store. The selection of learning activities is based on three criteria items relating to a student, namely cognitive style, learning style, and background knowledge. Moreover, the adaptation controller selects/formats assessment activities based on the input provided by the assessment engine. The schematic of the system is depicted by Fig. 1.
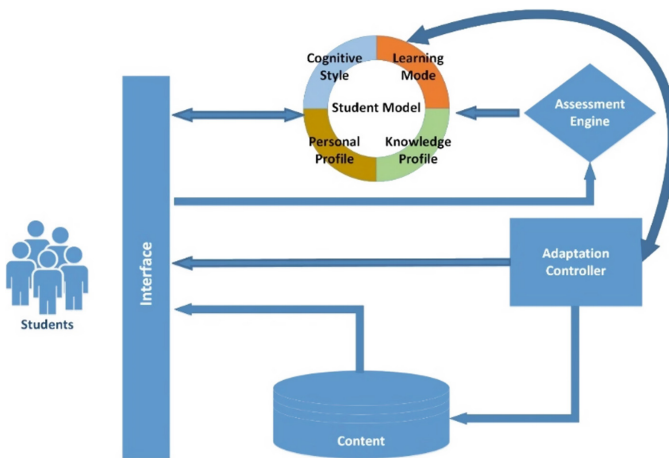

Fig. 1. Schematic of the proposed innovation

Nonetheless, situating every factor in the context of the learner is important for an effective learner model, but some attributes weigh more than others. An important aspect to consider in the case of developing nations is the difference of cognitive style

of learners as compared to the ones from developed nations. The authors thus chose Kirton's innovator-adaptor model [15] which is more closely related to rote/meaningful learning than other cognitive models. The authors' stance also finds support in an experiment conducted at the University of Central Punjab (UCP), Pakistan. However the result of that experiment is the topic of another paper [16].

The learning styles model—not to be confused with cognitive styles catering rote/meaningful learning—used in the system is based on VARK (Visual, Auditory, Read-Write, Kinesthetic) model by Fleming and mills [17]. The VARK model is repeatedly reported to be found in learners from developing nations [18]. The authors, however, could not implement kinesthetic style due to software limitations.

Finally, the knowledge profile stores information about the current progress of the learner, upon which a new activity—learning or assessment—is selected for presentation. For further details of the system, the interested reader is referred to [16].

3.1 Disrupting the Rote Learning Loop

The system deploys a novel mechanism to disrupt the rote learning loop. The schematic of the proposed design is given in Fig. 2. Upon starting a learning session, the system —based on learner's model—selects and presents a particular lesson—say Γ—to learner. A learning lesson typically comprises 9 learning activities—3 visual, 3 auditory, and 3 reading, denoted Γv, Γa, Γr, respectively—all focused on the learning theme of that particular lesson. The arrangement of the presentation is decided according to learner's attributes. One learner may get visual first, then auditory, and then read, while another may receive auditory-visual-read, or any other combination.
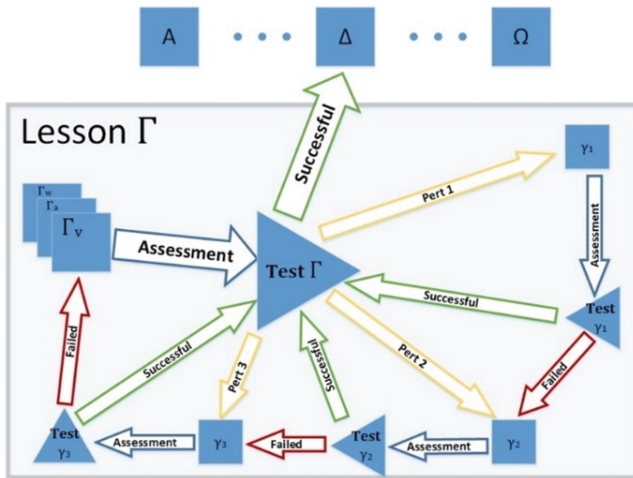

Fig. 2. The learning loop disrupting the rote memorization

After completing the lesson Γ, the learner takes assessment "Test Γ". The tests are designed in a way that they track down the perturbations back to the segment of

knowledge where the misconception is stemming from. For example, consider a question on loops in C++. The question may have 4 answers, amongst which one is correct. The rest three are distractors, designed in a way that they point into the direction where the learner's knowledge is erroneous. For example, answer *b* tells that learner has a problem with understanding conditional statements and relational operator, and option *c* may reveal that the learner is not good with pre/post increment concept. Designing the assessment in such fashion allows tracking the source of perturbations.

If a learner is successful in assessment, she moves to next segment of knowledge taking the shortest possible path within knowledge domain. However, if the assessment is incorrect, or partially correct, she may take one of several possible paths. A major error in learning of lesson Γ takes her to sub-activity $\gamma1$, a lesser problem to $\gamma2$, and a still lesser problem to $\gamma3$—increasing subscripts denote reducing magnitude of error. If the learner is taken to $\gamma1$—a sub-activity of Γ presented in a different way to avoid rote learning—she has to take sub-activity assessment "Test $\gamma1$" as well. Completing $\gamma1$ successfully moves her back to main Γ. However, if she is not able to successfully complete $\gamma1$, she is taken one level further down to $\gamma2$. A successful completion of "Test $\gamma2$", takes to the main assessment, and failure takes one step further down to $\gamma3$. If the learner is not able to complete the simplest level $\gamma3$, she is then taken back to do the entire learning of lesson Γ again.

3.2   Assessment Model

An important aspect in learning is assessment. If learners are expected—or trained—to produce verbatim (principles, rules, formulas, definitions) answers in assessments, they incline towards bare memorization of facts, i.e. rote learning [19]. To incline them towards more meaningful learning experiences, the assessments shall be designed and implemented with different expectations—no verbatim answers expected.

The test to be conducted on the students was divided into two parts: standardized and adaptive. The standardized test had same questions and rubric for all the participants [19] and the scoring was done as (1), where $Q_{\gamma s}$ represents the quiz from lesson γ with standardized questions, r_i represents the response to the ith item in the respective quiz. The response is calculated as (2).

$$Q_{\gamma s} = \sum_{i=0}^{n} r_i \qquad (1)$$

$$r = \begin{cases} 1, & correct\ answer \\ 0, & incorrect\ answer \end{cases} \qquad (2)$$

Therefore, this part of the assessment followed "criterion-referenced score interpretations" scheme [20], which only considers if the students' answer is correct or not. As an outcome of this approach, a student may simply be declared as 'fail' or 'pass' for the respective test if the cumulative score is $\geq 50\%$.

The second part of the assessment consisted of an adaptive approach to rate the learner amongst peers. The question bank consisted of calibrated (criterion: difficulty

level) items which had been meticulously designed by pedagogues. The starting point of this test was based on the score obtained in the standardized test. The score of standardized test was stratified into 50–69%, 70–85% and >85%. This allowed the appropriate entry point for the candidate into the adaptive quiz. The scoring of the adaptive part was based on the formula in (3):

$$Q_{\gamma a} = \sum_{i=0}^{n} r_i \times \frac{1}{a_i} \times \frac{1}{w_i} \times \frac{1}{t_i} \qquad (3)$$

The penalty terms in (3) are a (number of times the student changed answer options before submitting), w (number of times the student attempted the same question), and t (the time taken by the student to answer, 1 in case of predefined time limit (90 s) and 2 in case of more time), defined respectively in (4), (5), and (6).

$$a = \begin{cases} 1, & \textit{first click} \\ k, & k^{th} \textit{ change of option} \end{cases} \qquad (4)$$

$$w = \begin{cases} 1, & \textit{first attempt} \\ k, & k^{th} \textit{ attempt of the item} \end{cases} \qquad (5)$$

$$t = \begin{cases} 1, & \textit{within } 90 \textit{ seconds} \\ 1 + \left(\frac{1}{p} \times k\right), & k^{th} \, 30 \textit{ second interval} \end{cases} \qquad (6)$$

The value of p can be empirically estimated. We used $p = 10$ for the experiments. For the adaptive part of the quiz, a student can get the right answer in the first click and attempt, in which case the r_i will be the score of the respective i^{th} quiz item. For all other cases, the score $Q_{\gamma a}$ will depend upon the contribution from the penalty terms and, thus, can never be 100%.

4   Research Methodology

The efficacy of first prototype was tested with CS1 students at the UCP. All participants were enrolled into same 5 courses, including CS1, Basic Electronics, English 1, Social Studies, and Logical Thinking. The students were not given the choice to select subjects themselves—UCP freshmen are offered a pre-designed track in 1st semester.

The complete enrollment of 4 semesters (S15, F15, S16, and F16) was inducted into the experiment. In S15, the students were taught with conventional methods. The educational process was watched closely and the results were recorded. Meantime, the content developers created English language content, and the system developers prepared the first prototype. As soon as F15, the system was ready to go under first efficacy testing. The content/system refinement continued in parallel with QA and error correction, resulting in an updated prototype for S16, and a further improved form in F16. The results of all these semesters were subjected to statistical analysis.

The course chosen for the analysis was English 1, since the second language learning is an area which is especially affected with rote learning mechanism. Nonetheless, building vocabulary may be argued to base on a bare-memorization technique, but comprehending information from a passage needs some creative thinking.

4.1   Participants

A total of 1161 students participated in the experiment, of which 82 withdrawn from the course of their studies in English 1, or dropped from the program altogether. Of the rest 1079, 108 belonged to the control group. The remaining 971, were subjected to different level of treatment. The students in F15 had the first version of the system, which was improved for S16, and in a still improved form for F16. Hence, the level of treatment for the subsequent semesters was increasing. Table 1 enlists the total number of students in each group. The students under the head grading are those whose data were included into the analysis. The enrollment of both S15 and S16 was divided into 3 sections, while F15 and F16 had 10 and 11 sections respectively.

Table 1. Number of students enrolled, withdrawn, and continued in each semester

	S15	F15	S16	F16
Sections	3	10	3	11
Graded	108	352	133	486
Withdrawn	30	6	5	41
Total	138	358	138	527

Table 2. Number of sections in each semester with respective enrolment

Sections	S15	F15	S16	F16
A	42	27	46	47
B	38	29	49	46
C	28	36	38	–
D	–	27	–	41
E	–	35	–	50
F	–	40	–	51
G	–	40	–	37
H	–	39	–	35
I	–	40	–	48
J	–	39	–	51
K	–	–	–	46
L	–	–	–	34
Total	108	352	133	486

Since, all students belonged to the same semester of same program, hence they were assumed to have similar profile, including prior knowledge, the skills learned, the courses taken, and the level of studies already achieved. Moreover, they were enrolled in UCP via the same admission process/criteria, passing the same admission test, and fulfilling the same entry requirements, hence ensuring a similar knowledge profile across entire population.

The same team taught all 4 semesters. However, the number of teachers engaged differed for each semester. F15 and F16 being more populous had more teachers engaged into teaching than spring semester, as detailed in Table 2.

4.2 Procedure

Since, the fall semesters normally gets more intake, F15 and F16 had more sections as compared to the spring semester. The section assignment was on first come first serve basis. At any given time, only one section was open. As soon as a student was admitted to the program, she was assigned to the open section. Once, the opened section had received enough enrollments, it was closed, and another section in line was open. For example, the first student was admitted to section A, and all the forthcoming 49 other students were assigned to that very section. Once, the section A had 50 enrollments, it was marked close and section B was opened, and so on. Moreover, the teacher assignment for the sections was not known at the time of student enrollment, hence, no teacher preference bias was induced.

Contrary to the conventional learning procedure in S15, the subsequent semester were mostly automated. The lessons were delivered mostly with the help of AEH, though the teachers taught some portions manually as well. The quizzes and assignments were mostly delivered and assessed and recoded through the AEH interface.

Each instrument had a specific weight in the final grade of the students. For example, the quizzes comprised 15% of the total weight, and so do the assignments. The class participation comprised of 5%, and the presentation was weighed 10%. The mid-term and the final-term was 25% and 30% of the total grade, respectively.

All the instruments were designed in a way that they minimized the chances of producing verbatim answers, even in S15. However, S15 procedure does not have perturbation tracking mechanism providing learning iterations over learning modules. The instruments in each semester were analyzed and improved for further administration into upcoming semesters. The major improvement was introducing answers which were more innovative and creative.

4.3 Tools and Materials

Both manual and automated system included several teaching interventions and assessment instruments designed on similar pedagogical pattern, though differing with respect to technology. Nonetheless, the exact number might have differed in a few cases, but all students of all sections of all semesters received 45 contact hours, either with AEH or without it. On the assessment side, 14 short quizzes were administered on average, on weekly basis—one quiz a week—to track the perturbations in students' current state of knowledge. Similarly, 4 assignments, a class participation activity, a presentation, a mid-term and a final-term exam were administered, either manually or via AEH.

In S15, the students used paper based instruments including all quizzes, assignments, and exams. Nonetheless, the presentation and the class participation activities— and even assignments in some cases—involved the use of multimedia and word processing. Contrarily, almost all tools and materials in F15, S16, and F16 were computer

based. All quizzes were administered electronically through the use of AEH, as well as assignments which were delivered and collected through the same platform. However, the exams—mid and final—still remained paper based. One standard outline was followed throughout four semesters.

5   Statistical Analysis and Results

The dependent variable (the numerical grade of students) produced 4 distributions, namely S15 [$N = 108$, $M = 66.93$, $SD = 9.49$], F15 [$N = 352$, $M = 69.71$, $SD = 10.77$], S16 [$N = 133$, $M = 73.26$, $SD = 10.53$], and F16 [$N = 486$, $M = 80.96$, $SD = 7.40$], chronologically representing each semester included into the study. The sample size in each semester varied—the authors had no control over enrollment. To the authors' surprise, no assumptions of parametric analysis was tenable in any distribution. The Shapiro-Wilk test—used to test the normality of data in samples—did not accept the null hypothesis of normality for any distribution [S15: ($p = .004$), F15: ($p < .001$), S16: ($p = .007$), F16: ($p < .001$)]. The same was confirmed through the visual inspection of Q-Q plots as depicted by Fig. 3. Adding to non-normality, large number of outliers appeared in some distributions, as depicted by the boxplot in Fig. 4. The assumption of homoscedasticity was not tenable as well. The Leven's test of homogeneity of variance failed to accept the null hypothesis of equivalent variance in all 4 distributions [$F(3, 1075) = 16.68$, $p < .001$].
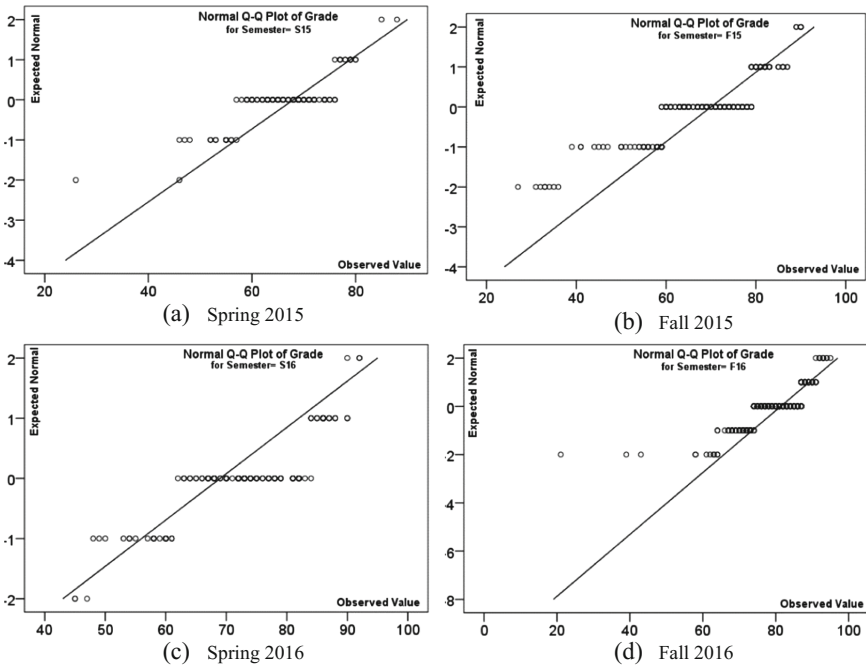
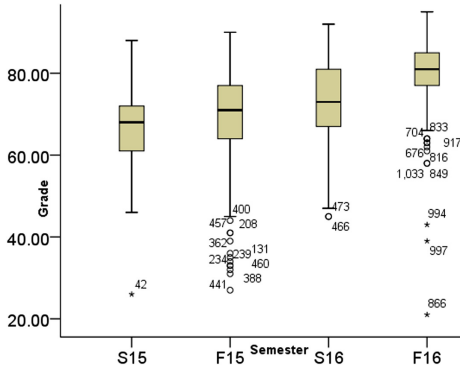

Fig. 3. Q-Q plots of all four distributions
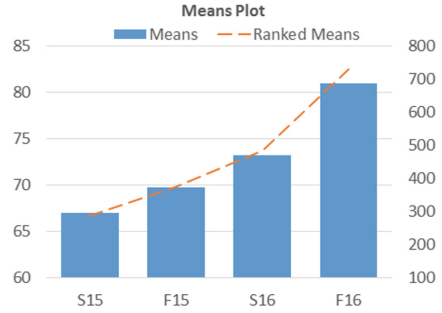
Fig. 4. Boxplot of grade distributions

Fig. 5. Means plot comparing arithmetic means of the distributions with ranked means of transformations

Since, the assumptions of parametric analysis were not tenable, the authors opted for Kruskal-Wallis test for Analysis of Variance, reported to be the most favored nonparametric test [21]. The Kruskal-Wallis H statistic showed that there was a statistically significant difference between distributions [$\chi^2(3) = 352.02$, $p < .001$, $R_{S15} = 287.56$, $R_{F15} = 376.78$, $R_{S16} = 483.55$, $R_{F16} = 729.76$]. Further analysis with non-parametric Jonckheere-Terpstra test for ordered alternatives showed that there was a statistically significant trend of increasing median amongst distributions [$J_{JT} = 296705.00$, $z = 18.67$, $p < .001$], remarkably in a chronological order. The same was confirmed by the visual analysis of the means plot presented by Fig. 5. Both means—the means of the original distribution and the means of transformations—had an increasing trend in the chronological order of semesters, i.e. progressing from S15 to F15, and then from S16 to F16, with F16 having the highest mean and median.

6   Discussion and Implications

The statistical analysis of data generated in 4 semesters revealed interesting facts. First, the H-statistic indicated differences in grade distributions of 4 semesters. Then, the means of the raw scores turned out to be rising in chronological order, indicating a positive change in learning, assessed through quizzes, assignments and exams. The positive change was however not attributed to the chance alone, since the data was subjected to chance-corrected statistical methods, like that of Kruskal-Wallis test. The H-test also indicated a difference in the means of transformations created from the original grade distributions. Moreover, the ranking test (Jonckheere-Terpstra) indicated the same chronological ordering, as was observed through visual analysis of means of raw distributions.

The authors draw the most important implication that the rote learning hinders the meaningful learning, and hence performance and creativity of the students in respective area. The argument is backed-up in the literature, as well as finds supports in the experiment reported here. Albeit being a less-respected learning technique, the students

are somehow compelled to choose rote learning due to its ability to strip the complexity off the topic. Contrarily, the educators' community wants pupils to learn things in more meaningful ways. Nonetheless, achieving this goal is difficult with conventional pedagogies. The teachers can deliver lectures in novel ways, engage students in meaningful activities, design creative assessments, reward for novelty and innovation, but administering all this with a large number of enrollment seems difficult, if not impossible at all.

One promising solution is the use of AEH systems, which were previously targeted over the customized learning experience. The authors suggest that the AEH—and the learning and assessment modules—shall be designed in a way that they enforce students to refrain from rote learning. In support of their argument, the authors have demonstrated how an AEH system can be designed and implemented to disrupt the rote learning loop.

Nonetheless, controlling all variables in a social science educational setting was not possible. However, the researchers tried to keep the execution all the same across all these years. The only difference induced was in form of educational technology used for learning and assessment. Nonetheless, the students also were changed during each term, however, the induction process for new students remained the same and the students—though changing personally—belonged to the same population.

7   Conclusions and Future Work

AEH can help in changing the preferred learning strategy of the student. One possible course of action is to design AEH with situated learner attributes considering factors which are compelling student body to learn via rote methods at large. Additional to the learning and cognitive attributes, the learning management can, as well, push the students towards rote learning. Learning management inculcating rote learning involves (1) designing and implementing such learning activities which loop on the same thing several times, (2) designing and implementing assessments which are answered with remembered concepts, and (3) rewarding for the verbatim answers. The design of an AEH shall also consider these factors as well. The authors have built such an AEH with a proposed novel assessment system, and the experiments show positive effects on learning.

In the future, the authors want to run further real-time experiments with the system to gather more data on the efficacy. The authors also plan on running two separate subjects, one with the AEH, and the other in a conventional manner to compare the results.

References

1. Chin, C., Brown, D.E.: Learning in science: a comparison of deep and surface approaches. J. Res. Sci. Teach. **37**(2), 109–138 (2000)
2. She, H.-C.: Promoting students' learning of air pressure concepts: the interrelationship of teacher approaches and student learning characteristics. J. Exp. Educ. **74**(1), 29–51 (2005)
3. Lithner, J.: Learning mathematics by creative or imitative reasoning. In: Cho, S. (ed.) Selected Regular Lectures from the 12th International Congress on Mathematical Education, pp. 487–506. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-17187-6_28
4. Mokhtar, A.A., Rawian, R.M., Yahaya, M.F., Abdullah, A., Mohamed, A.R.: Vocabulary learning strategies of adult ESL learners. Engl. Teach. **36**, 133–145 (2017)
5. Kılıç, D., Sağlam, N.: Students' understanding of genetics concepts: the effect of reasoning ability and learning approaches. J. Biol. Educ. **48**(2), 63–70 (2014)
6. Lithner, J.: A research framework for creative and imitative reasoning. Educ. Stud. Math. **67**(3), 255–276 (2008)
7. Hacieminoglu, E.: Elementary school students' attitude toward science and related variables. Int. J. Environ. Sci. Educ. **11**(2), 35–52 (2016)
8. Pell, A.W., Iqbal, H.M., Sohail, S.: Introducing science experiments to rote-learning classes in Pakistani middle schools. Eval. Res. Educ. **23**(3), 191–212 (2010)
9. Sonawat, R., Kothari, M.: Rote learning and meaningful learning in mathematics: perspectives of primary school teachers and learning in children. Indian J. Posit. Psychol. **4**(1), 49–54 (2013)
10. Safdar, M.: Meaningful learning and rote learning in physics: a comparative study in city Jhelum (Pakistan). Middle Eastern Afr. J. Educ. Res. **6**, 60–77 (2013)
11. Liu, L., Hmelo-Silver, C.E.: Promoting complex systems learning through the use of conceptual representations in hypermedia. J. Res. Sci. Teach. **46**(9), 1023–1040 (2009)
12. Jacobson, M.J., Archodidou, A.: The design of hypermedia tools for learning: fostering conceptual change and transfer of complex scientific knowledge. J. Learn. Sci. **9**(2), 145–199 (2000)
13. Zydney, J.M., Grincewicz, A.: The use of video cases in a multimedia learning environment for facilitating high school students' inquiry into a problem from varying perspectives. J. Sci. Educ. Technol. **20**(6), 715–728 (2011)
14. Rum, S.N.M., Ismail, M.A.: Metacognitive support accelerates computer assisted learning for novice programmers. Educ. Technol. Soc. **20**(3), 170–181 (2017)
15. Kirton, M.: Adaptors and innovators: a description and measure. J. Appl. Psychol. **61**(5), 622–629 (1976)
16. Hassan, M.M., Qureshi, A.N.: Situating adaptive educational hypermedia into the local context of developing nations. In: Proceedings of the 2017 2nd International Conference on Communication and Information Systems (ICCIS 2017), Wuhan, China (2017)
17. Fleming, N.D., Mills, C.: Not another inventory, rather a catalyst for reflection. To Improve the Academy, Paper 246. DigitalCommons@University of Nebraska, Lincoln (1992)
18. Saga, Z., Qamar, K., Trali, G.: Learning styles-understanding for learning strategies. Pak. Armed Forces Med. J. **65**(5), 706–709 (2015)
19. Ye, L., Lewis, S.E.: Looking for links: examining student responses in creative exercises for evidence of linking chemistry concepts. Chem. Educ. Res. Pract. **15**, 576–586 (2014)
20. Popham, W.J.: Why standardized tests don't measure educational quality. Educ. Leadersh. **56**(6), 8–15 (1999)
21. Lix, L.M., Keselman, J.C., Keselman, H.J.: Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance "F" Test. Rev. Educ. Res. **66**(4), 579–619 (1996)