Consistency of Students' Pace in Online Learning

Arnon Hershkovitz and Rafi Nachmias
{arnonher, nachmias}@post.tau.ac.il
Knowledge Technology Lab, School of Education, Tel Aviv University, Israel

Abstract. The purpose of this study is to investigate the consistency of students' behavior regarding their pace of actions over sessions within an online course. Pace in a session is defined as the number of logged actions divided by session length (in minutes). Log files of 6,112 students were collected, and datasets were constructed for examining pace rank consistency in three main situations: day/night sessions, beginning/end (for both situations, sessions of the same learning mode were taken), and a comparison between sessions from different learning modes. For each dataset, students were ranked twice, according to their pace in the two sub-groups, and these ranks were correlated. Results obtained with this study's data suggest that pace is sometimes not consistent, hence might not be considered as a characterizing measure for the whole learning period. A discussion of this study and further research is provided.

1 Introduction

Log files are the essential basis for many Data Mining research, however raw data from these files are usually being transformed into variables on which algorithms and statistical tests might be applied. In EDM research, all levels of aggregation into variables should be considered: keystroke level, answer level, session level, student level, classroom level, and school level [3]. While discussing individual differences between users (i.e., aggregating or estimating in student level), a question might arise: Do variables taken into consideration indeed characterize the learner (even regarding the limited context of domain and environment)? Not only that such a variable (e.g., session length, response time, intense of activity, preferred tasks) might introduce a large variance when repeatedly measured for the same student, there is also a possibility that this inconsistency represents a non-trait measure, hence this variable does not and should not represent a student.

In this study, we chose to examine the pace of actions within a Web-based learning environment. It is a time-related variable occasionally being calculated in the student level. However, in configurations where students have the freedom to choose when, where and what/how to learn, and while their sessions might extend over a long period (days or weeks) – it is not clear that a student has a "characterizing pace", and that we can try to compare students by their pace.

Moreover, pace measuring is just one example from a large set of variables often being used in student models, and an important purpose of this study is to shed light on some obstacles for using such variables.

2 Background

Logged data for calculating pace of activity in a learning environment, was studied – probably for the first time – almost twenty years ago in a Computer-based Instruction

(CBI) configuration [7]. The results suggested that "students exhibit a characteristic rate of responding or way of approaching CBI activities". Although this conclusion treats pace as measuring response or approach to activities, it seems that the basic definition of pace, as the researcher had defined it - number of activities completed, divided by total time on task – tend to be more cognitive than behavioral.

In fact, pace (also referred as *speed*, *rate*) is somehow a slippery term in EDM research, as it might relate to two different phenomena: a) Pace of learning – measured by completion rate per time-unit [7] or by time taken to complete a task – e.g., in [10, 16] (notice the difference in units between these two measures); b) Pace of action – measured by number of actions per time-unit [13, 14]. These two measurement are, of course, not independent, as pace of action might affect pace of learning, and vice versa: If we take, for example, two students with the very same cognitive skills needed for a given task but with different values of pace of action, the student which is more speedy has an advantage in completing the task quicker; on the other hand, student's pace of action might be affected by learning occurred or knowledge application needed between consecutive actions.

Although pace (in either interpretation) might change noticeably between tasks, it is sometimes being treated as characterizing the student for the whole learning period. Therefore, parameters measuring pace are being averaged over multiple sessions (as was previously done by the authors in [13]) or being calculated on the whole learning period level in the first place [8].

Considering pace as representing students might lead to a calculation of relative pace. For example, Beck's disengagement model [4] has a student-specific parameter of *reading speed*, for accounting inter-students variability; this parameter fine-tunes the model by considering the student's speed relative to the class' average, and is calculated and applied across all question types. Another relative calculation of time-related measuring is presented in [18], where student's *working time* was calculated as the ratio between the student's completion time for a given task divided by the class' average completion time. Both these studies rely on the hidden assumption that student's rank, regarding her or his activity's speed or time, is consistent over tasks and/or over time. The examination of that hidden assumption is the core of this research.

3 Methodology

To determine whether pace of action does characterize learners, we examined consistency of pace ranking, i.e., of students' ranking by their pace. If pace does characterize students, pace ranking is expected to be consistent (to a certain measure) over different situations. The following three situations were examined:

a) Day/night – median pace for each student is considered for calculating her or his rank in day/night sessions within the same learning mode
b) Over time – pace ranks are based on pace measures for beginning and last sessions within each learning mode. Second session was chosen to represent the beginning, since pace in first session might be greatly biased

c) Across learning modes – median pace in each mode serves as the basis for pace ranks.

In addition, we examined another situation, which is quite more technical: Pace ranks are based on median pace in two randomly-divided groups of sessions for each student (first, in general, and then within each learning mode).

Different datasets were constructed for each of the above situations, as will be described in section 3.4. Following is a description of the learning environment, the log file, the data collection and preprocessing, and the datasets construction.

3.1 The Learning Environment

A simple yet very intensive online learning unit was chosen as the research field. This fully-online environment focuses on Hebrew vocabulary and is accessible for students who take a face-to-face preparatory course for the Psychometric Entrance Exam (for Israeli universities). The online system is available for the participants from the beginning of the course and until the exam date (between 3 weeks and 3 months in total).

The system includes a database of around 5,000 words/phrases in Hebrew and, offers the students with a few learning modes: a) Memorizing, in which the student browses a table of the words/phrases along with their meanings; b) Practicing, in which the student browses the table of the words/phrases without their meaning. The student may ask for a hint or for the explanation for each word/phrase; c) Gaming; d) Self-testing, in the same format of the exam the students will finally take; and e) Searching for specific word/phrase. The first two modes (Memorizing, Practicing) have a very similar interface of a multi-page table in each row of which there is a word/phrase; while in the Memorizing mode, the meaning of that word/phrase is shown, in the Practicing mode it is hidden and will be revealed only upon the student's request.

3.2 Log File Description

The researched system logs the students' activity, thus each student is identified by a serial number. Each row in the log file documents a session, initiated by entering the system and ended with closing the application window. For each session, the following attributes are kept: starting date, starting/ending time, ordered list of actions and their timestamps; actions documented are every html/asp page in the system, not including actions within Java/Flash applets.

3.3 Data Collection and Preprocessing

For examining the research hypothesizes, we used logged data from April 2006 – May 2007. The original data included 181,111 sessions of 11,068 students. Cleaning was done for keeping only the following: a) active sessions – session that lasted at least one minute and less than one hour, and that had at least five documented actions; b) active students – students who had at least three active sessions. The cleaned log had 64,700 (active) sessions of 6,112 (active) students. Pace for each session was calculated as the number of actions in the session, divided by the session length (in minutes).

Next, we mapped and coded the actions within each session to one of the four learning modes: Memorizing, Practicing, Self-testing, Searching; gaming was not coded because most of the gaming-related pages are implemented in Java, and therefore they were not documented. Then, each session was coded into one of the four modes if at least 60% of its actions were of that same mode. It turned out that about 30% of the sessions were coded as "Memorizing", 20% were coded as "Practicing", only about 1% of the sessions were "Searching", and only a few sessions were "Self-testing"; the rest were not categorized to any of the modes (i.e., they were mixed sessions). Therefore, our study is focused only in the two eminent modes.

3.4 Constructing the Datasets for Testing the Hypotheses

Eight different datasets were constructed, in order to investigate the consistency of pace rank between day/night sessions, between beginning/end sessions, across learning modes, and among random divisions of the sessions. A detailed description is given in Table 1.

Table 1. Description of the datasets for investigating pace rank consistency

Dataset	Learning Mode(s)	Sessions Were Included for Students With…	Total Students	Total Sessions	Pace calculation for student-group
Dataset1$_M$ Day/night	Memorizing	at least 3 sessions in each group of day/night sessions	331	3,823	Median
Dataset1$_P$ Day/night	Practicing	at least 3 sessions in each group of day/night sessions	285	4,389	Median
Dataset2$_M$ Beginning/end	Memorizing	at least 3 Memorizing sessions	2,650	16,724	One sample
Dataset2$_P$ Beginning/end	Practicing	at least 3 Practicing sessions	1,358	11,409	One sample
Dataset3 Across modes	Memorizing + Practicing	at least 3 sessions of each mode (Memorizing, Practicing)	768	12,593	Median
Dataset4$_A$ Random division	All	no limitations	6,112	64,700	Median
Dataset4$_M$ Random division	Memorizing	at least 3 sessions in each of two randomly divided sub-groups of the sessions	758	8,445	Median
Dataset4$_P$ Random division	Practicing	at least 3 sessions in each of two randomly divided sub-groups of the sessions	526	7,739	Median

For each dataset, we sorted the students twice, according to their pace in the relevant sub-groups (the student with the highest pace was ranked as "1", the student with the second-highest pace was ranked as "2", and so on). These two ranks were correlated using Spearman's rho (ρ) and Kendall's tau (τ), two common alternatives for non-parametric correlation coefficients ([-1,1]) which are often being compared, however without a sharp recommendation towards neither of them [9, 12, 17]; it is known that the Kendall's coefficient is usually lower than the Spearman's.

4 Results

Day/Night Consistency

Results for $Dataset1_M$ and $Dataset1_P$, in which day/night situation was examined in the two learning modes, are given in Table 2. It might be concluded from the results that there is a significant relatively high correlation between pace ranks between day and night in both modes. It was also found that there is a significant difference when comparing means of pace values between day and night groups: Mean pace over night sessions was higher than the mean pace over day sessions; t values were 2.11^{*} ($df=330$) for $Dataset1_M$, and 2.33^{*} ($df=284$) for $Dataset1_P$.

Table 2. Day/night consistency of pace rank

Dataset	N (Students)	Mode	Group 1	Group 2	ρ	τ
$Dataset1_M$	331	Memorizing	Day	Night	0.59^{**}	0.43^{**}
$Dataset1_P$	285	Practicing	Day	Night	0.53^{**}	0.39^{**}

* $p<0.05$, ** $p<0.01$

Beginning/end Consistency

Results for $Dataset2_M$ and $Dataset2_P$, examining consistency of pace ranks over time, are given in Table 3. As might be seen, correlation coefficients are pretty low. On average, beginning and last sessions are differed by pace of action within them: Students tend to work faster at the end, as shown by t values of 3.33^{**} ($df=2,649$) for $Dataset2_M$, and 3.64^{**} ($df=1,357$) for $Dataset2_P$.

Table 3. Over time consistency of pace rank

Dataset	N (Students)	Mode	Sample 1	Sample 2	ρ	τ
$Dataset2_M$	2,650	Memorizing	2nd session	Last session	0.26^{**}	0.18^{**}
$Dataset2_P$	1,358	Practicing	2nd session	Last session	0.20^{**}	0.14^{**}

** $p<0.01$

Another way of looking at these results is to scatter plot a two-dimension representation of the students according to their ranks in both groups, and to look at the four quadrants formed by the median lines. If pace rank is consistent, it is anticipated that the faster students will be faster in both dimensions, and same for the slower students, hence quadrants I (top-right) and III (bottom-left) should be occupied with most of the dots (students).

For example, let's take a look at such a scatter plot for *Dataset2$_P$*, which relates to the beginning/end situation for the Practicing learning mode. The examination of pace rank consistency for this dataset showed a low yet significant correlation (ρ=0.20**). The scatter plot for this example is presented in Figure 1. According to our calculations, the first and the third quadrants each holds 30% of the dots, which means that the second and fourth quadrants hold together 40% of the students.

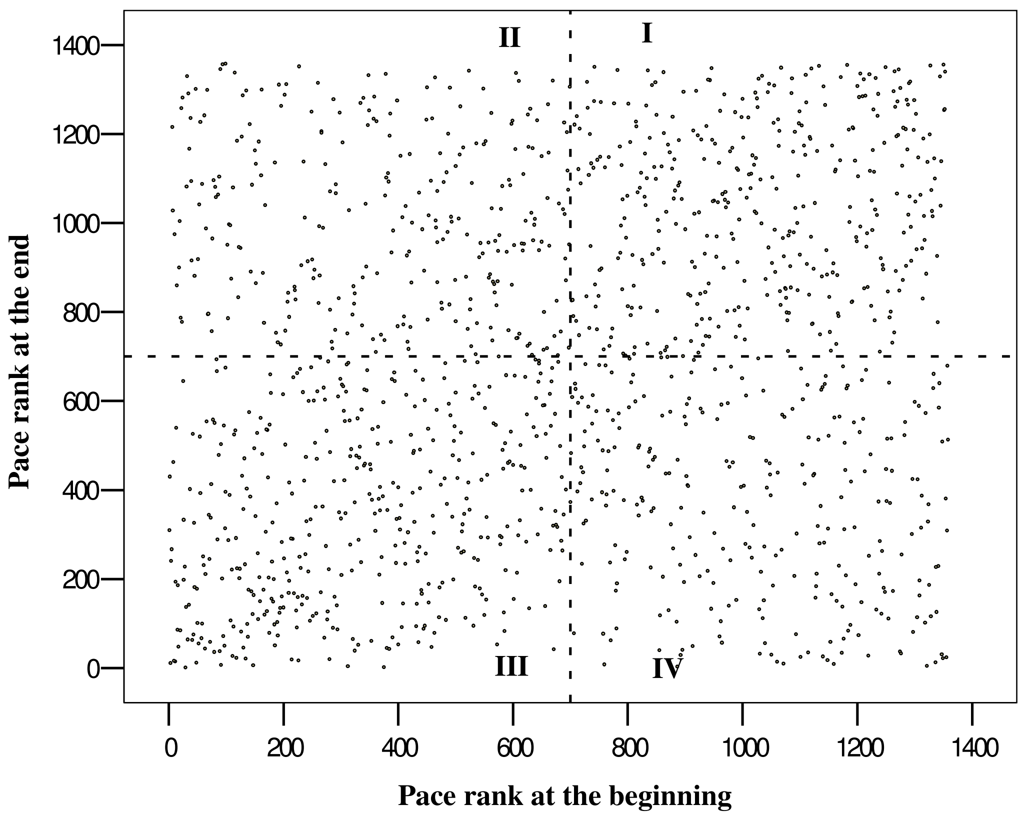

Figure 1. Scatter plot of pace ranks at the beginning (x) and the end (y) for *Dataset2$_P$* (Practicing learning mode), N=1,358

Across Modes Consistency

Results for *Dataset3* are given in Table 4, representing the examination of pace rank consistency across learning modes. Correlation coefficients are relatively low for this situation. Furthermore, there is a significant difference between the means of the two groups: On average, Memorizing sessions were faster than Practicing sessions with $t(767)=7.99^{**}$.

It is a good point to recall the similarities and differences between the two learning modes being discussed here. While Memorizing and Practicing modes share a very similar GUI, and work according to the same principle (browsing over pages each consisting of a 10-row table of words/phrases), the main difference is that the Memorizing tables show the meaning of the term, while the Practicing tables hide it. As suggested by the results, students spend more time on Memorizing pages than on Practicing pages, and pace ranks across modes have a low correlation. This might imply that pace of action is affected by a set of skills needed for progressing in either of the modes.

Table 4. Across modes consistency of pace rank

Dataset	**N (Students)**	**Group 1**	**Group 2**	**ρ**	**τ**
Dataset3	768	Memorizing	Practicing	0.34^{**}	0.23^{**}

** $p<0.01$

Random Division Consistency

Results for *Dataset4*$_A$, *Dataset4*$_M$ and *Dataset4*$_P$ are given in Table 5. These three datasets relate to a more technical situation than the previous ones: random division of each student's sessions to two groups, and examination of pace rank consistency between these two groups. While *Dataset4*$_A$ takes into consideration all the sessions from the log file, *Dataset4*$_M$ and *Dataset4*$_P$ relate only to Memorizing and Practicing sessions, accordingly.

Table 5. Random division consistency of pace rank

Dataset	**N (Students)**	**Mode**	**Group 1**	**Group 2**	**ρ**	**τ**
Dataset4$_A$	6,112	All	Random	Random	0.36^{**}	0.25^{**}
Dataset4$_M$	758	Memorizing	Random	Random	0.62^{**}	0.45^{**}
Dataset4$_P$	526	Practicing	Random	Random	0.56^{**}	0.41^{**}

** $p<0.01$

It might be seen that for the general case – correlation is relatively low, however when examining pace ranks within the same learning mode, correlation is resulted with relatively high values of coefficients. Also, no significant difference was observed in the means between the two groups within each of the datasets.

To conclude the results of this study, there were only two situations in which pace rank was found to be consistent with relatively high values of correlation coefficients: a) Day/night division within the same learning mode; and b) Random division of each student's sessions within the same learning mode. In all the other situations - namely: over time, across modes, and all-inclusive random division - pace rank consistency was found to be relatively low, with correlation coefficients (ρ) between 0.20^{**} and 0.36^{**}.

5 Discussion

Many EDM studies often handle fine-grained data in the action/session level, like pace measures. However, when examining the student level, mainly since vector variables are not easy to cope with while applying data mining algorithms, scalar measures of these variables are often being used (e.g., average or median pace over different sessions). Time-related variables (usually describing the time taken for answering a question or for completing a task) are quite common in EDM research [1, 8, 11], but others are also often being averaged, for example: attempts for answering a question [1, 11], hint/help usage (usually per question) [1], and intense of activity (usually in terms of number of actions per session or frequency of certain activities) [6, 15]. While doing this, a hidden assumption – regarding the variable in question being a trait – is lying behind the calculations. It is our obligation to deeply investigate the consistency of each variable before projecting it on a 1-dimensional measuring scale and assuming it is of a trait type, as was clearly presented by Baker [2].

This is why we choose a rather primitive variable, namely pace of actions, in order to study its consistency. As the results obtained with our data suggest, correlation between pace ranks in different situations was sometimes very low. The minimal correlation coefficient (for *Dataset2$_P$*) was 0.20^{**}, which is almost a zero correlation. The maximal correlation coefficient (for *Dataset4$_M$*) was 0.62^{**}, which is relatively high but still quite far from a perfect correlation.

To be honest, these results was, at first, very surprising, as we expected to see much higher correlation values. The fact that for one situation (beginning/end consistency, Practicing mode) 40% of the students were located at the second and fourth quadrants of the pace ranks scatter plot (Figure 1) – indicating they were above the median rank in the beginning and below it in the end, or vice versa – is thought-provoking, and explicitly shedding light on the questionability of the assumption of pace rank consistency.

Furthermore, the surprisingly low correlations might imply that our choice of pace was not at all of a simple variable as we first thought, as pace of actions depicts different kinds of processes in which the online student is involved while learning, e.g., reading, memorizing, recalling previous knowledge, thinking, processing, typing, and navigating. Besides the clear effect of different learning components on learning time/pace, individual components also heavily affect it, such as ability to understand instruction or quality of instruction events, as was seminally proposed by Carroll [5]. Considering that pace measurement embodies different task-related and/or student-related components (and potentially others), it is clear that replicating this study with different learning systems and/or with different pace metrics is necessary before generalizing any conclusion regarding the consistency phenomenon.

In general, many educational studies investigate all kinds of students' attributes; however, EDM researches often analyze data drawn from relatively long periods of time, therefore our hand on the reduction trigger is likely to be more itchy. Further research and a deeper investigation is needed in order to better understand which behavioral attributes in online learning are indeed students' traits and which are heavily situation dependent.

References

[1] Arroyo, I. and Woolf, B.P. (2005). Inferring learning and attitudes from a Bayesian Network of log file data. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*. Amsterdam, The Netherlands.

[2] Baker, R. (2007). Is gaming the system state-or-trait? Educational Data Mining through the multi-contextual application of a validated behavioral model. *Proceedings of Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling*. Corfu, Greece.

[3] Baker, R.S.J.d., *Data Mining for Education*, in *International Encyclopedia of Education (3rd edition)*, B. McGaw, P. Peterson, and E. Baker, Editors. In Press, Elsevier: Oxford, UK.

[4] Beck, J.E. (2004). Using response times to model student disengagement. *Proceedings of ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*. Maceio, Brazil.

[5] Carroll, J.B. (1963). A model of school learning. *Teachers College Record, 64*(8), 723-733.

[6] Chen, G.-D., Liu, C.-C., Ou, K.-L., and Liu, B.-J. (2000). Discovering decision knowledge from Web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research, 23*(3), 305-332.

[7] Clariana, R.B. (1990). Rate of activity completion by achievement, sex and report in computer-based instruction. *Journal of Computing in Childhood Education, 1*(3), 81-90.

[8] Cocea, M. and Weibelzhal, S. (2006). Can log files analysis estimate learners' level of motivation? *Proceedings of 14th Workshop on Adaptivity and User Modeling in Interactive Systems*. Hildesheim, Germany.

[9] Daniel, W.W., *Applied Nonparametric Statistics*. 1978, Boston, MA: Houghton Mifflin.

[10] de Vicente, A. and Pain, H. (2002). Informing the detection of the students' motivational state: An empirical study *Proceedings of The Sixth International Conference on Intelligent Tutoring Systems (ITS 2002)*. Biarritz, France.

[11] Feng, M., Beck, J., Heffernan, N., and Koedinger, K. (2008). Can an Intelligent Tutoring System predict math proficiency as well as a standardized test? *Proceedings of First International Conference on Educational Data Mining*. Montreal, Canada.

[12] Gibbons, J.D., *Nonparametric Measures of Association*. Quantitative Applications in the Social Sciences. 1993, Newbury Park, CA: Sage Publications.

[13] Hershkovitz, A. and Nachmias, R. (2008). Developing a log-based motivation measuring tool. *Proceedings of First International Conference on Educational Data Mining*. Montreal, Canada.

[14] Kornbrot, D. and Macleod, M. (1990). Monitoring and Analysis of Hypermedia Navigation. *Proceedings of INTERACT '90, The 3rd IFIP Conference on HCI*. Cambridge, UK.

[15] Perera, D., Kay, J., Yacef, K., and Koprinska, I. (2007). Mining learners' traces from an online collaboration tool. *Proceedings of Educational Data Mining Workshop at the 13th International Conference on Artificial Intelligence in Education*. Marina del Ray, CA.

[16] Qu, L. and Johnson, W.L. (2005). Detecting the learner's motivational states in an interactive learning environment. *Proceedings of Artificial Intelligence in Education*. Amsterdam, The Netherlands.

[17] Sprent, P., *Applied Nonparametric Statistical Methods*. 2000, London, UK: CRC Press.

[18] Zhang, G., Cheng, Z., He, A., and Huang, T. (2003). A WWW-based learner's learning motivation detecting system. *Proceedings of International Workshop on Research Directions and Challenge Problems in Advanced Information Systems Engineering*. Honjo City, Japan.