Validating the use of B-Alert Live Electroencephalography in Measuring Cognitive Load with the NASA Task Load Index

Charles Cowan[1], Justyn Girdner[1], Blaize Majdic[1], Elise Barrella[2], Robin Anderson[1], and Mary Katherine Watson[3]
[1]James Madison University, [3]The Citadel, The Military College of South Carolina, and [2]Wake Forest University

Abstract

Evaluation of cognitive functioning via physiological measures is a growing area of research in Engineering Education due to its potential implications for enhancing student performance. This paper focuses on the measurement of cognitive functioning via electroencephalography (EEG) and self-report measures, and their relationship with performance. Researchers evaluated the B-Alert X10 EEG system's reliability in measuring cognitive load, and thus indirectly evaluated its potential to measure both cognitive flexibility and cognitive efficiency in future research. Sophomore and senior undergraduate engineering students solved five engineering problems of increasing complexity while connected to the EEG. As a secondary measure, participants also completed the NASA Task Load Index, a multidimensional self-report assessment tool. The average cognitive load experienced by all participants increased as they attempted to solve problems of increasing difficulty, and sophomores experienced greater cognitive load than seniors. These findings further support electroencephalography as a valid measure of cognitive load.

Keywords

B-Alert, EEG, NASA TLX, cognitive load, problem solving

Introduction

The brain is one of the most important, and most studied, organs in the human body. It is capable of quickly processing information, making decisions to aid in survival, and controls a large portion of the biological systems required for survival. The human brain constantly evaluates an individual's surroundings, looking towards the future and generalizing about our environment. It develops strategies to enhance our opportunities and minimize dangerous encounters[1].

In addition to maintaining biological homeostasis, the brain also spends a great deal of time performing problem solving. Problem-solving engages a learner with a wide variety of cognitive components, including but not limited to concepts, rules, information networking, memory, and knowledge assessment[2]. This wide range of cognitive components can be attributed to the large variation among problem types. To show a comparison across the several types of problems, Jonassen collected hundreds of problems, analyzed their various attributes, and categorized them into 11 main groups: logical, algorithmic, story, rule-using, decision making, troubleshooting, diagnosis solution, strategic performance, case analysis, design, and dilemma[3]. These problem types vary in degrees of structuredness and complexity. Well-structured problems present all elements of the problem to the solver and have solutions that are both distinct and comprehensible, where the relationship between decision choices and problem states is known[4].

Ill-structured problems possess elements that are unknown, contain a variety of solutions (including no solution at all), and have multiple criteria for evaluating solutions[5]. Complexity on the other hand, while slightly overlapping with structuredness, has a different meaning. The complexity of a problem is defined by the number of issues, functions, or variables that are needed to solve a problem. Complexity is also impacted by the degree of connectivity among these properties, as well as their stability over time[6]. The degree to how well or ill structured a problem is, as well as it's degree of complexity, impacts how an individual solves the problem, as well as how difficult the individual perceives the problem to be.

Individuals vary in their cognitive styles and controls, which represents patterns of thinking that control the ways that individuals process and reason about information[7]. Therefore, a problem that may be difficult to solve for one individual may be considered easy to another. Nevertheless, problem solving is a very complex process, which is why it is a highly valued skill, and has a strong emphasis in the fields of mathematics, science, and engineering[8]. Ultimately, we are studying three measures of cognitive load, cognitive flexibility, and cognitive efficiency. Cognitive load is the amount of mental effort exerted by the working memory at a given time[8]. Cognitive flexibility is the ability to mentally switch between different concepts and thus to think about multiple concepts simultaneously. Cognitive efficiency is an individual's ability to use his or her mental resources to solve problems. The study in this paper focuses specifically on cognitive load.

Cognitive Load Theory and Measurement

Educational research literature is increasingly using cognitive load theory to understand how individuals learn, and to seek ways to create more effective learning styles. Cognitive load theory (CLT) is concerned with techniques for using working memory in ways that facilitate the changes in long term memory associated with schema construction and automation[9]. CLT is based on the established psychological principles of a long-term memory with a virtually unlimited capacity for storing information, and of a working memory with a limited capacity in processing information[10]. CLT research directly contributes to the design of instructional methods that effectively maximize the use of our limited cognitive processing capacity in acquiring knowledge and applying skills[10].

The load that performing a specific task imposes on the learner's cognitive system can be represented by a multidimensional construct referred to as cognitive load[12]. Cognitive load includes the concepts of mental load, mental effort, and performance[13]. Mental effort, compared to mental load and performance measures, is considered more directly related to cognitive load[11]. It is measured while participants are working on a task; and is the feature of cognitive load that refers to the cognitive capacity that is allocated to accommodate the demands imposed by the tasks being performed[10]. Mental load is task-related, and it serves as an indicator of the cognitive capacity needed to process the complexity of a task. Using an electroencephalogram (EEG), cognitive load can be measured based on the ratio of theta waves to alpha waves. A greater ratio is indicative of a higher cognitive load.

The research team chose to use the B-Alert X10 EEG for several reasons. The B-Alert X10 is a 9-sensor EEG that provides the option of recording electrocardiographic (ECG) data, or more plainly put, it can monitor participants' heartrates. The device is low-cost, portable, and wireless device. The B-Alert X10 also interacts with the B-Alert Live software suite, which utilizes thoroughly validated algorithms to measure and record cognitive load in real time[14]. This algorithm uses a linear discriminate function (LDF) to generate a value between 0 and 1, representing the likelihood that the individual is experiencing cognitive load at a specific 1-second epoch. This variable typically has a low range (usually between .6 and .8), but it usually has a low variance as well, granting it statistical power. In previous research, researchers have analyzed this LDF-based variable through visual analysis and comparison, as well as through statistical significance testing[15].

As a secondary measure of cognitive load, researchers employed the NASA-TLX (Task Load Index) self-report assessment tool. The NASA-TLX is a multidimensional assessment tool that was developed by NASA in the late 1980s for gathering information about the magnitude and sources of workload related factors[16]. The NASA-TLX carefully and specifically defines six dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. The researchers chose the NASA-TLX over other self-report measures as it was well-established due to its age, commonly cited as a valid metric, and due to the tool's growing prominence as a secondary measure during EEG studies[17]**.** This measure not only provided a secondary measure of cognitive load to compare to the team's EEG results, but also allowed the project greater confidence in the gradual rise in difficulty of its procedure's problems.

Developing the Problem Set

Five problems were selected to serve as the problem set the participants would solve. The first problem selected was an algorithmic problem, designed to be very structured and low in complexity. The second problem selected was a rule using problem, designed to be moderately structured but still low in complexity. The third problem selected was a story (word) problem, designed to be slightly less structured than the previous problem with the same level of complexity. The fourth problem selected was a more complex rule-using problem, in which the participant had to use concepts that were covered in both problem two and problem three. This problem was designed to be less structured than the others while also being moderately complex. The fifth problem selected was a troubleshooting problem, designed to be the most complicated and least structured of all the problems. Each of these problems were selected so that they would progressively increase in complexity and decrease in structuredness. The goal of this selection was to increase the cognitive load experienced while solving each problem. For more information about these five problems, please see the appendix.

The B-Alert Live System

The B-Alert X10® collects data from nine EEG sensor sites (Fz, F3, F4, Cz, C3, C4, POz, P3 and P4), two reference electrodes, and two electrocardiogram (ECG) lead sites. The device interacts with a data analysis software package from Advanced Brain Monitoring (ABM) called B-Alert Live®. This software package allows the researcher to easily analyze data via ABM's algorithms for assessing cognitive state metrics (which assesses engagement level) and cognitive workload metrics (which assess a participant's mental effort). The software package also allows the researcher to monitor distraction via an algorithm derived from the cognitive state metrics, and it monitors heart rate via ECG.

B-Alert Live systems feature automatic signal decontamination measures, including measures for electromyography (EMG), electrooculography (EOG), spikes, saturations, and excursions. The software measures engagement via a four-class quadratic DFA derived for each participant during the metric benchmarking task. This four-class model is constructed using absolute and relative power spectra variables from Fz-POz and Cz-POz. The EEG measure of mental workload was established via data from C3-C4, Cz-PO, F3-Cz, Fz-C3, and Fz-PO[18]. These measures have been previously validated in military, industrial, and educational research[19].

Participants

Participants were undergraduate engineering students at James Madison University (N = 9). All participants were either sophomores (n = 4) or seniors (n = 5), and all were male except for a single female sophomore. Participants received an $11 school dining voucher upon completion of the study as incentive.

EEG Sessions

Participants were primarily communicated to via a script/protocol, which was written to standardize the experience between participants as much as possible. Researchers gave the participants specific points during the study in which participants could ask questions about the study. After competing an informed consent document, all participants were fitted with the B-Alert X10 system and the device's connection was tested for impedance per instructions from the device's manufacturer. Participants then completed metric benchmarking tasks to create a baseline for B-Alert's algorithm to utilize. This algorithm allows B-Alert Live to compare each individual subject's baseline EEG activity to that same subject's EEG activity under load.

After participants completed the metric benchmarking tasks, a research team member provided participants a paper list of descriptions of all dimensions the NASA-TLX assessed, and data acquisition began. During data acquisition, participants completed five physics problems of increasing difficulty. Participants were offered 1-minute breaks between questions. After participants completed a problem, researchers provided participants with a NASA-TLX report sheet a scale for each of the six measured dimensions. At the end of data acquisition, participants completed a final NASA-TLX task that compared each of the six dimensions to the other five, as instructed in the NASA-TLX manual[16]. A participant's time commitment for completing the entire procedure ranged from 45 minutes to 90 minutes.

Results

When collecting the data, each participant was identified by a four-digit system. The first digit is a representation of the student's gender (0=male 1=female). The second digit is a representation of the student's grade level (0=sophomore 1=senior). The last two digits are a representation of the participant number (ex. 01). Before the data could be analyzed, it first had to be cleaned. This process began by selecting the data points that were collected while the participant was solving a problem. This was done by using the recorded start and stop times noted above. After the data associated with each problem was identified, the data was then scanned for invalid epochs. These epochs occur when more than 128 (out of a possible 256) values are deemed corrupted by the software, and thus the software labels that one second epoch as corrupted. These errors are identified as values of -99999 and are to be excluded from analysis according to B-Alert's User Manual.

Data Analysis

The researchers wanted to create a problem set that increased in difficulty from problem to problem. Table 1 shows the percentage of participants across each of the five problems that answered the problem correctly. This table shows that, generally, the problems increased in difficulty as the experiment progressed.

Table 1 Correct Responses Across Five Problems

Problem	Correct Responses	Percent Correct
1	9	100%
2	7	63%
3	7	63%
4	4	44%
5	4	44%

Researchers focused on averages of B-Alert Live's cognitive load metric, overall correctness score on the problems, and NASA-TLX self-report data. Average load can be defined as the mean intensity of load during the performance of a single task (in this experiment, a single question). This was calculated by averaging B-Alert Live's cognitive load variable across the duration of a problem. Overall correctness was scored based on a total of 15 potential points throughout all five problems. NASA-TLX score was calculated by multiplying the weight of a dimension against its raw score. This new adjusted score for each dimension was then summed and divided by 15, creating an overall workload score ranging from 1 to 10. This workload score was then averaged across all five problems for each participant, giving the researchers an overall index of load for the experiment.

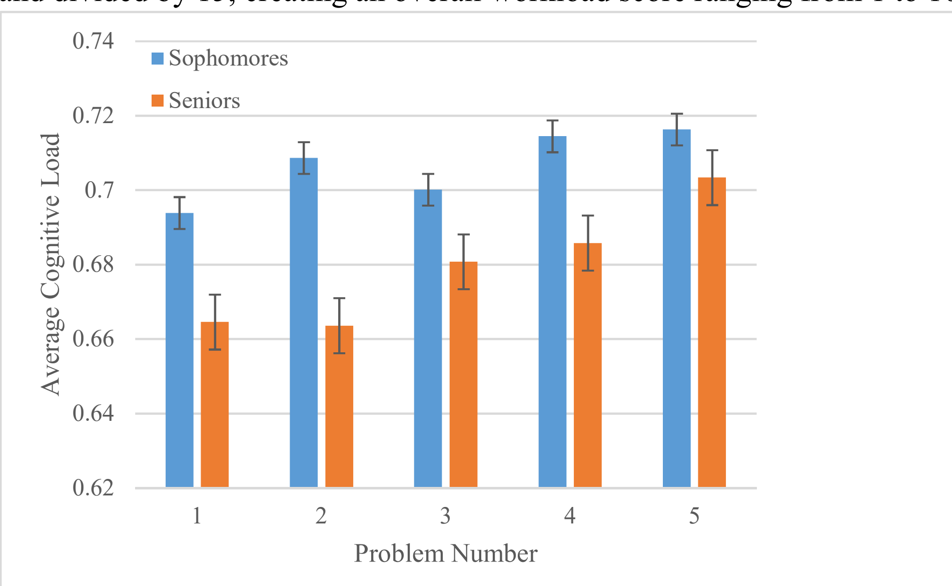


Figure 1 Average Load Compared Between Sophomores and Seniors (a higher problem number indicates a greater level of difficulty)

A consistent, gradual increase in average load throughout the five problems was observed visually, as seen in Figure 1. This figure demonstrates the average cognitive load that was experienced for each problem, and comparing sophomore averages against seniors. As hypothesized, sophomores experienced a noticeably larger amount of cognitive load than seniors did, with this difference increasing as the problems increased in difficulty.

Researchers then compared overall average load against the correctness scores of each participant. This did not reveal any visible relationship or any kind. The researchers then compared NASA-TLX overall load to each participant's score, which also showed little relationship.

Though correctness did not seem to have an impact on either an individual's cognitive load, as measured by the EEG, or via their own reporting via the NASA-TLX, there was a noticeable difference between sophomore and senior participants in self-reported overall workload. As demonstrated in Figure 2, seniors perceived that they were working harder than sophomores perceived themselves working, despite EEG evidence to the contrary.

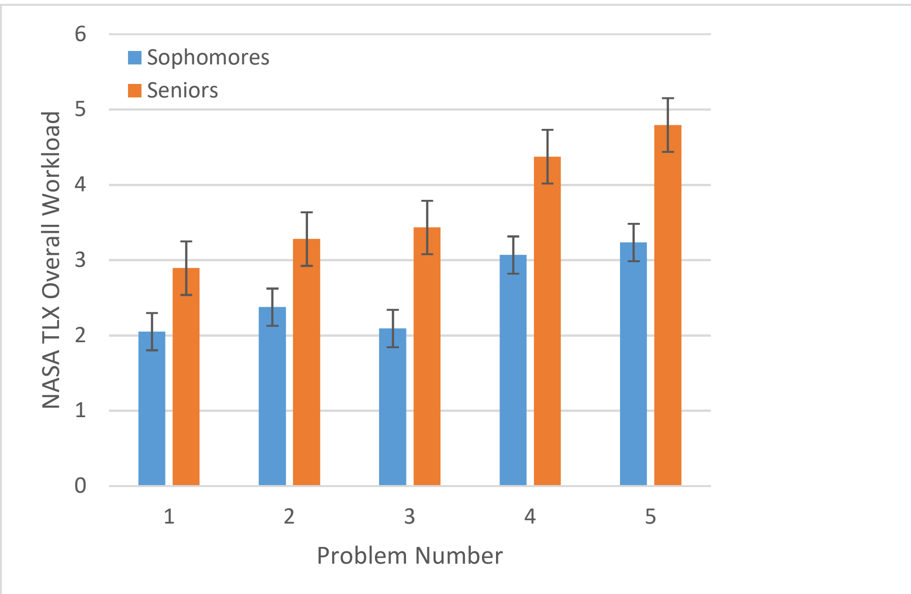


Figure 2 Overall NASA-TLX Workload Compared Between Sophomores and Seniors (a higher problem number indicates a greater level of difficulty)

With this small of a sample size ($N = 9$), we did not run inferential statistics on this data. However, inferential statistics are appropriate tools to use with both NASA-TLX results and B-Alert Live's cognitive load metric. Despite the lack of statistical analyses, there is a visible difference in self-reported workload between sophomore and senior participants.

Discussion

The researchers began this project with the goal of creating problems that scaled upward in difficulty based on Jonassen's hierarchy[2]. Based on the results demonstrated in Table 2, the researchers feel confident that the problems increased in difficulty as the experiment progressed, since correctness scores, on average, decreased as participants progressed through the procedure. Researchers predicted that sophomore students would have to work harder, and thus exhibit more cognitive load, than senior students to solve the same problems. Figure 1 supports this hypothesis as measured through EEG data.

The research team implemented the NASA-TLX as a measure of self-reported workload to have a secondary measure of problem difficulty, and to see if there was a difference between how hard individuals thought that they worked and how hard individuals worked (as measured by the EEG). This led the research team into some surprising findings as senior participants reported that they were working harder on each task than sophomores, however they were experiencing less cognitive load than sophomores. The researchers are not sure why this difference in perceived, or self-reported, workload exists. It may be attributed to senior students nearing their graduation dates, but regardless the results are interesting. The difference in perceived workload does not strongly contradict the researchers' originally hypotheses, given the other results of this study, but it does raise some interesting questions about motivation.

Future Work

The research team is currently evaluating differences in cognitive load measured through both EEG and self-report measures between linear and systems thinking. The researchers are doing this by having senior undergraduate engineering students complete two tasks related to a specific problem in the field of engineering. These two tasks are either the construction of a concept map related to the specified problem or the listing of as many related concepts to the task as possible. The listing of terms/concepts related to the problem is a linear thinking task, and concept maps involve systems thinking. The current experiment being conducted in our lab can be directly attributed to the success of the study presented in this paper. The validation of both the NASA-TLX and B-Alert's EEG quantification of cognitive load in our specific use case is essential to our current study. Our overarching goal is to find ways to create teaching methods that naturally promote the efficient management of cognitive resources, to train more effective engineers within the sustainable design paradigm. The results of the study detailed in this paper, and the experience that the research team gained during this study, provide the foundation that our current research is built upon.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant Nos. 0846468 and 1463865. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would also like to thank the student participants that took part in this study and Dr. Olga Pierrakos and Ms. Winifred Opoku who contributed to the study.

References

1. Zull, James Elwood. The Art of Changing the Brain: Enriching Teaching by Exploring the Biology of Learning. Sterling, VA: Stylus, 2002. 48-50. Print.
2. Jonassen, D. H. (1997). Instructional design models for well-structured and III-structured problem-solving learning outcomes. Educational Technology Research and Development, 45(1), 65-94.
3. Jonassen, D. H. (2000). Toward a design theory of problem solving. Educational technology research and development, 48(4), 63-85.
4. Wood, P.K. (1983). Inquiring systems and problem structures: Implications for cognitive development. Human Development, 26, 249-265.
5. Meacham, J.A., Emont, N.C. (1989). The interpersonal basis of everyday problem solving. In J.D. Sinnott (Ed.), Everyday problem solving: Theory and applications (pp. 7-23). New York: Praeger.
6. Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In RJ. Stemberg & P.A. Frensch (eds.), Complex problem solving: Principles and mechanisms (pp. 185-222). Hiilsdale, NJ: Lawrence Erlbaum Associates.
7. Jonassen, D.H., & Grabowski, B.L. (1993). Handbook of individual differences, learning and instruction. Hiilsdale, NJ: Lawrence Erlbaum Associates.
8. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285.

9. Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. Instructional Science, 32(1/2), 1-8.
10. Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist, 38(1), 63-71.
11. Myers, David G., and C. Nathan DeWall. Exploring Psychology. New York: Worth, Macmillan Learning, 2016.
12. Paas, F., & van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. Educational Psychology Review, 6, 51–71.
13. Jahns, D. W. (1973). A Concept of Operator Workload in Manual Vehicle Operations. Forschungsinstitut Anthropotechnik, Meckenheim, West Germany (Technical report No.14).
14. Berka, C., Levendowski, D. J., Westbrook, P., Davis, G., Lumicao, M. N., Olmstead, R. E., Ramsey, C. K. (2005). EEG quantification of alertness: Methods for early identification of individuals most susceptible to sleep deprivation. Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations, Orlando, FL, SPIE: The International Society for Optical Engineering.
15. Johnson, R. R., Popovic, D. P., Olmstead, R. E., Stikic, M., Levendowski, D. J., & Berka, C. (2011). Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. Biological psychology, 87(2), 241-250.
16. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183.
17. Lecoutre, L., Lini, S., Bey, C., Lebour, Q., & Favier, P. A. (2015). Evaluating EEG Measures as a Workload Assessment in an Operational Video Game Setup. In PhyCS (pp. 112-117).
18. Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, space, and environmental medicine, 78(5), B231-B244.
19. Berka & Behneman, 2011; Linking Models of Team Neurophysiologic Synchronies for Engagement and Workload with Measures of Team Communication

Charles M. Cowan

Charles Cowan is a post-graduate research assistant at James Madison University. He received an M.A in Psychological Sciences in the spring of 2017.

Justyn Girdner

Justyn Girdner is an undergraduate Engineering student at James Madison University. His expected graduation date is spring 2020.

Blaize Majdic

Blaize Majdic competed his bachelor's in Engineering at James Madison University in the spring of 2017. He is currently working on his graduate degree at Virginia Commonwealth University.

Elise M. Barrella

Dr. Elise Barrella is an assistant professor, and a founding faculty member, of Wake Forest University's Department of Engineering. Prior to joining the founding faculty team of Wake Forest University, she performed research and instructed classes at James Madison University. Her research interests include education, assessment, transportation, and sustainability.

Robin Anderson

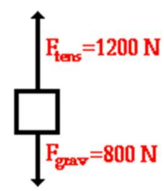
Dr. Robin Anderson is the department chair of the Department of Graduate Psychology at James Madison University. She specializes in the validation of assessment and program evaluation procedures.

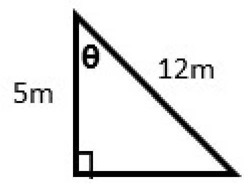
Mary Katherine Watson

Dr. Mary Katherine Watson is an assistant professor of Civil and Environmental Engineering at The Citadel. She has been with The Citadel since August of 2013.Her research interests include sustainable engineering and environmental engineering.

Appendix

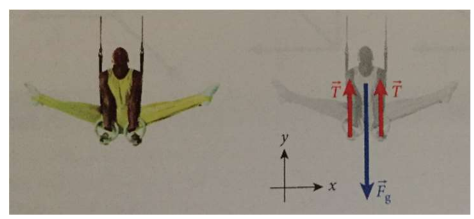
1. Examine the figure below. If there is a upward force of 1200N applied to an object, at the same time a downward force of 800N is applied to the object, determine the magnitude and direction of the net force that is applied on the object.



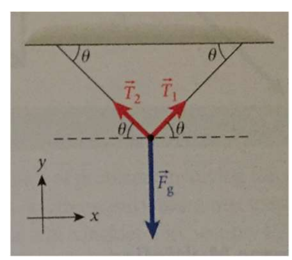


2. Using the triangle [above], solve for the angle **θ**.

3. A gymnast has a mass of 55kg and is hanging vertically from a pair of parallel rings (as shown below). If the ropes supporting the gymnast are completely vertical and attached to the ceiling above, what is the tension force in each of the ropes?



4. If the same ropes mentioned in the problem above are connected to the ceiling, where θ=45°, what is the tension force in each rope?



5. For the pulley system shown below, if the mass of block m is equal to 10kg, what force must be applied at the end of rope 1 to keep the system in static equilibrium?