Adaptive Assessment Improving Interaction in an Educational Hypermedia System

Evangelia Gouli, Harry Kornilakis, Kyparissia Papanikolaou, Maria Grigoriadou

Department of Informatics and Telecommunications, University of Athens,
Panepistimioupolis,
GR-15784 Athens, Greece
{ lilag, harryk, spap, gregor }@di.uoa.gr

Tel. 7275230, 7275205
Fax. 727521

SUMMARY

In this paper, we present an adaptive assessment framework as part of the diagnostic module in the Adaptive Educational Hypermedia System INSPIRE-INtelligent System for Personalized Instruction in a Remote Environment. The main goal of the assessment procedure is to provide assessment of learner's knowledge level and to give the learner a chance to keep track of his/her learning progress. Adaptive assessment aims to assess a learner's competency by posing a minimum number of questions. The proposed adaptive assessment framework facilitates the interaction between the learner and the system as it provides a more accurate estimation of the learner's proficiency in an efficient way.

KEYWORDS : Adaptive Assessment, Adaptive Educational Hypermedia Systems, Item Response Theory

Name and version of software used to generate the submitted document: Word 2000

Type of submission: Full Paper

Descriptive terms to describe the submission, from the list of options: User models and models of interaction

Adaptive Assessment Improving Interaction in an Educational Hypermedia System

Evangelia Gouli, Harry Kornilakis, Kyparissia Papanikolaou, Maria Grigoriadou

Department of Informatics and Telecommunications, University of Athens,
Panepistimioupolis,
GR-15784 Athens, Greece
{ lilag, harryk, spap, gregor }@di.uoa.gr

SUMMARY

In this paper, we present an adaptive assessment framework as part of the diagnostic module in the Adaptive Educational Hypermedia System INSPIRE-INtelligent System for Personalized Instruction in a Remote Environment. The main goal of the assessment procedure is to assess the learner's knowledge and to give the learner a chance to keep track of his/her learning progress. Adaptive assessment aims to assess a learner's competency by posing a minimum number of questions. The proposed adaptive assessment framework facilitates the interaction between the learner and the system as it provides a more accurate estimation of the learner's proficiency in an efficient way.

KEYWORDS: Adaptive Assessment, Adaptive Educational Hypermedia Systems, Item Response Theory

INTRODUCTION

Adaptive Educational Hypermedia Systems (AEHS) [3][4] is a novel area of research aiming to offer personalized support according to the personal needs and abilities of each learner. The presence of an effective assessment mechanism in such a system is essential, since the assessment of learning is a crucial part of the instructional design process and therefore of an educational system [12]. Assessment is part of the developmental process of learning [9] and it is related to the accomplishment of the learning outcomes. Through the assessment, the learner is able to identify what he/she has already learned and which are his/her strengths and weaknesses, to observe his/her personal learning progress and to decide how to further direct his/her learning process. Furthermore, in the context of an AEHS the assessment is exploited by the system itself as part of the diagnostic process [18]. The estimated learner's proficiency can then be used to guide the adaptation of the system.

Adaptive assessment [20] is a type of assessment in which, in contrast to the static nature of tests with fixed questions, the assessment process is dynamic. In an adaptive test, the selection of each question and the decision to stop the test are dynamically adapted to the learner's performance in the test [8]. Several research papers [19][15] have shown that adaptive assessment is usually more efficient than non-adaptive methods. Less time is needed to administer adaptive tests since fewer questions are required to achieve acceptable accuracy in the evaluation of the learner. Further, adaptive assessment can provide a more accurate estimation of the learner's proficiency. Such a property is very useful for the assessment mechanism of an AEHS, since in case the system uses the learner's proficiency as a source of adaptation, its accurate estimation is critical to the system's efficiency. Also, adaptive assessment facilitates the interaction between the learner and the system. Adaptive assessment is less tedious for the learners, as the generated tests are tailored to their proficiency and thus they don't have to answer numerous questions that are either too easy or too difficult for them.

This paper describes an adaptive assessment framework designed for the Adaptive Educational Hypermedia System-INSPIRE [7]. The assessment mechanism assesses the learning progress and offers feedback to the learner based on the answers given by him/her. In the next sections, an outline of the INSPIRE is briefly presented and a general description of the adaptive assessment is given. Afterwards, the design of the adaptive assessment in INSPIRE and the adaptive assessment algorithm are presented. The paper ends with conclusions and further research.

AN OUTLINE OF INSPIRE

INSPIRE is an AEHS that aims to facilitate distance learners during their study, by restricting the domain knowledge at the beginning of the interaction, a strategy more appropriate for novices [2], and enriching it, progressively, following their performance. Based on the learning goals that the learner selects, INSPIRE generates lessons that correspond to specific learning outcomes, accommodating learner's proficiency and learning style. For a detailed discussion of the system, see [7,14].

The domain knowledge in INSPIRE is represented in three hierarchical levels of knowledge abstraction: *learn-*

ing goals, *concepts* and *educational material* [14]. The learner selects a learning goal that corresponds to a topic of the domain knowledge. A subset of concepts of the domain knowledge is associated with each goal. The *outcome concepts*, i.e. the most important of them, are presented through various modules of educational material, the *knowledge modules*. Also, a number of prerequisite and related concepts are associated with each outcome concept. Furthermore, the presentation of each outcome concept is associated with the *three different levels of performance*, Remember, Use and Find [11].

The design of the assessment mechanism in INSPIRE, is based on the three following principles [9]: i) assessment should reflect what is most important for the learner to learn ii) the assessment process should enhance learning and support instructional practice iii) every learner should have the opportunity to learn through assessment.

Following the guidelines proposed in [10] the assessment is oriented to:

- assess the learner's knowledge before or through a lesson.
- stimulate the learner to contemplate on the material he/she has studied.
- support the learner in revising the concepts that he/has already studied.
- indicate whether concepts of the course have been clearly understood by the learner.

ADAPTIVE ASSESSMENT

An adaptive assessment is tailored to the learner's proficiency and provides the learner with questions, which depend on his/her previous answers. The estimation of the learner's performance isn't based only on the percentage of correct responses, but also on the difficulty level of the questions the learner was able to answer correctly [6].

In general, the adaptive assessment procedure works as follows. Initially, a question of moderate difficulty is presented, because an estimation of the initial learner's proficiency is not available and therefore a moderate knowledge level is assumed. Questions are selected in a way such that their difficulty matches the learner's estimated proficiency. If the answer to the question is correct, the learner's estimated proficiency is updated and is set to be higher than before; otherwise it is estimated as lower. Then, the next question is selected and presented based on the learner's estimated proficiency. After the answer, the learner's proficiency is re-estimated and so on. As this process progresses, the distance between the estimated proficiency and the true proficiency of the learner is gradually becoming smaller. After a number of questions, the assessment will hopefully reach an accurate estimation of the learner's actual proficiency. The assessment procedure is terminated when specified termination criteria, related to the length of the test or the desired accuracy, are carried out. It is obvious that the estimation not only depends on the number of questions answered correctly but also the level of difficulty of the answered questions is taken into account. Following this procedure, a student answering questions correctly will be administered progressively more difficult questions, while a student answering incorrectly will be administered progressively easier questions.

Several systems use adaptive assessment methods based on different approaches. Collins, Greer and Huang in [5] use granularity hierarchies and Bayesian nets to provide adaptive assessment of multiple traits in a single test. Huang in [8] described an adaptive testing algorithm, CBAT-2 that generates content-balanced questions based on the portion of the course curriculum that meets the goals of a test and uses a simple machine learning procedure to determine the question parameter values. Finally, the SIETTE system [16] is an Internet based evaluation system and has a complete set of tools to support teachers in test development and students to assess themselves. Adaptive assessment capabilities are provided to the system based on the Item Response Theory (IRT) model [17].

ADAPTIVE ASSESSMENT IN INSPIRE

In Inspire, for the estimation of the learner's proficiency, a number of questions has to be posed. The tutor specifies these questions and associates them with the knowledge modules of the outcome and prerequisite concepts. It is possible to have questions associated with more than one knowledge module. A number of parameters is associated with each question such as the level of difficulty, the level of performance, number of times that the question has been answered correctly or incorrectly by any learner etc. These parameters allow the system to determine the appropriate question that corresponds to each learner according to his/her proficiency level and his/her navigation behavior in the knowledge modules. The tutor also defines a number of parameters in the test specification i.e. the minimum number of questions that will be posed for different levels of performance, the maximum number of questions that will be posed in the test, the desired level of estimation accuracy etc.

Success or failure in an assessment often motivates learners. A difficult assessment may encourage learner to put more effort. This approach increases motivation by increasing learner's stress, which can eventually result in a negative impact on the learner's performance. If learners are successful in an assessment, they receive satisfaction and reassurance that they are reaching the required standard [10]. In our approach, the assessment mechanism poses questions that improve learner's motivation and reduce stress by taking into consideration the difficulty level of each question and the current learner's proficiency. Tediousness from answering many easy questions and frustration from answering too many hard questions are avoided [6]. Furthermore, the assessment mechanism provides meaningful feedback aiming to improve learner's performance. The correct answer is given for each question, along with hints if the answer isn't correct and links to locations where the learner could find supplementary knowledge.

The learner is offered two options regarding his/her assessment. The first is to take a self-assessment test on the knowledge modules that he/she has already studied, so that he/she can review his/her learning progress. The second option is to take a summative assessment test on any outcome concept or the entire learning goal.

The self-assessment test is constructed based on the learner's proficiency and the navigation that he/she has already performed. Thus, if the learner has studied the educational material concerning the remember level of performance, and he had a limited navigation in the educational material concerning the use level of performance, then the self-assessment test will present questions associated mainly with the remember level of performance.

The summative assessment test assesses the learning outcomes of the concept independently from the navigation behavior of the learner. Also questions associated with the prerequisite concepts are posed.

The main components of the adaptive assessment framework are:

- Question and Feedback Knowledge Base (QFKB): The collection of the questions that are posed in a test, and the text that will be used as feedback to the learner for each of the possible answers that might be given to a question.
- Test Adaptive Generation module (TAG): The part of the system that is responsible for selecting the questions, which are posed to the learner. It takes into account the test specifications, the parameters of the questions, such as difficulty, and the current estimation of the learner's proficiency.
- Learner Assessment module (LA): The part of the system that is responsible for assessing the learner's proficiency and for informing the learner about his/her learning progress.
- Presentation Module (PM): The part of the system that is responsible for the presentation of the question that is selected by TAG and for the presentation of the feedback, which depends on the learner's answer.
- Question and Test Editor (QTE): A tool that allows tutors to give test specifications and insert questions and feedback text in the QFKB. Such a tool allows the tutor to concentrate on the authoring of questions and their quality and relieves him of the strenuous technological details.

ADAPTIVE ASSESSMENT ALGORITHM

A high level description of the adaptive assessment algorithm is:

Make an initial estimation of the learner's proficiency
Repeat
> *Select a question based on the current estimation and pose it to the learner.*
> *Depending on the learner's answer, update the estimation.*

until one of the termination criteria is met

In the above description of the algorithm three important points have to be made more specific:

- how is each question selected?
- how is the learner's proficiency estimated and updated?
- what are the termination criteria?

Each of these questions is covered in detail in the following paragraphs.

Question Selection

For the selection of the questions in the adaptive assessment algorithm, we follow the framework of the Item Response Theory (IRT) [17]. IRT is a statistical model that has been widely used in the implementation of computerized adaptive assessments. Its aim is to provide information about the functional relation between the estimate of the learner's proficiency on a concept and the likelihood that the learner will give the correct answer to a specific question (the probability of the correct response).

According to IRT, in order to select a question, we have to calculate the Item Characteristic Curve (ICC) and the Item Information Function (IIF) of each question. The most appropriate question is the one that provides the most information about the learner's proficiency, i.e. the one with the highest value of IIF for the learner's proficiency. In our approach, in addition to the value of the question's IIF, the algorithm takes into consideration an additional factor, which is the weight of the knowledge module that the question refers to. The tutor has the opportunity to assign different weights to the various modules, declaring how important each of them is. Therefore questions that refer to more important modules have a greater chance of being chosen.

Each question in the QFKB has its own Item Characteristic Curve (ICC), which represents the probability that the learner with a certain proficiency will be able to provide a correct answer. This probability depends on two parameters that are specific to each question. These parameters are the difficulty of the question and its guessing factor. In combining these parameters we make use of the logistic ICC function used by Huang [8], which is a variation of Birnbaum's [1] logistic ICC adjusted for two parameters, instead of three[1]. Therefore the probability that a learner, whose proficiency is θ, will answer a question of difficulty b and guessing factor c is:

$$P(\theta) = c + \frac{1-c}{1+e^{-2(\theta-b)}}$$

In figure 1 appears the Item Characteristic Curve of a question that has difficulty level b=1 and guessing factor c=0.25. The proficiency level θ, which appears in the x-axis, may assume values between -3 (novice) and 3 (expert). In the figure we notice that the probability to answer the question increases with higher values of θ. Furthermore everyone has a probability of answering at least equal to the guessing factor c.

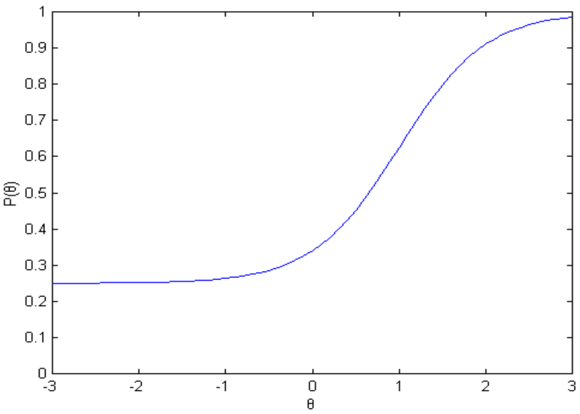

Figure 1: Item Characteristic Curve for a question with b=1 and c=0.25

The difficulty level of each question (parameter b) is initially assigned by the tutor and as the question is used in the assessment tests, it is re-estimated according to the formula described by Huang [8]. This is a method that is able to adjust the difficulty of a question depending on the number of times that it has been answered correctly or incorrectly. The guessing factor (parameter c) of a question is equal to the ratio of the number of possible correct answers to the total number of possible answers to the question. For example, it is 0.5 for a true-false question, 0 for a free-text question etc. [8]

[1] Birnbaum makes use of an additional parameter called discriminatory power. We have chosen to omit this parameter since it is very hard for a tutor to specify its value. [8]

For the selection of a question the Item Information Function (IIF) is calculated for each question. This function is a representation of the amount of "information" provided by each question. For example, the IFF of a difficult question assumes a value close to 0 for small values of θ, meaning that asking a novice a difficult question provides little information about his/her proficiency. The chosen question is the one with the highest value of IFF for the current estimation of the learner's proficiency. This is the question that, when asked, will provide the most information. Usually such questions are those with difficulty similar to the learner's proficiency and low guessing factor. According to IRT, a question's information function $I(\theta)$ can be quantified as the standardized slope of its ICC $P(\theta)$.

$$I(\theta) = \frac{(P'(\theta))^2}{P(\theta)(1-P(\theta))}$$

Figure 2 shows the Item Information Function for the question whose ICC appears in figure 1. The function assumes it's maximum value when θ is close to the difficulty of the question, while it is almost equal to 0 when θ is very low or very high. The interpretation of this is that most information can be provided when the proficiency of the learner matches the difficulty of the question.

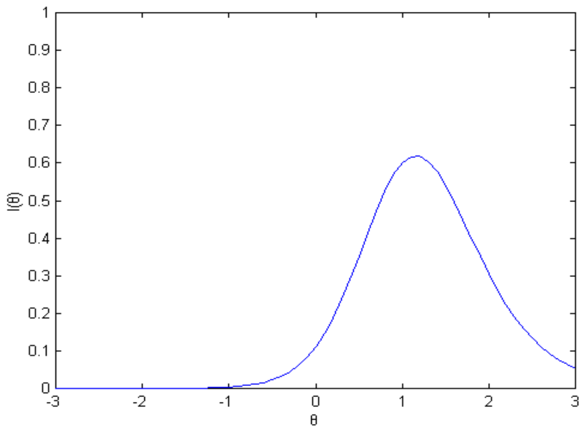

Figure 2: Item Information Function for a question with b=1 and c=0.25

Proficiency Estimation and Termination Criteria

In order to estimate the proficiency of the learner we use an iterative approach outlined by Lord [13], which is a modification of the Newton-Raphson iterative method for solving equations. According to this approach, initially the learner's proficiency is estimated as moderate. After each answer, the system's estimation of the learner's proficiency is adjusted by a quantity, which depends on the current estimation of his/her proficiency and on all the previous answers given by the learner. The update of the estimated proficiency θ after the n-th question is made according to the following formula.

$$\theta_{n+1} = \theta_n + \frac{\sum_{i=1}^{n} S_i(\theta_n)}{\sum_{i=1}^{n} I_i(\theta_n)}$$

with

$$S_i(\theta) = [u_i - P_i(\theta)] \frac{P_i'(\theta)}{P_i(\theta)[1 - P_i(\theta)]}$$

In the above formula θ_n is the proficiency estimation after n questions, P_i and I_i are the ICC and IFF of the i-th question, respectively and u_i is the answer given to the i-th question (u=1: correct; u=0: incorrect).

Figure 3, presents an example of the way that the estimation of the learner's proficiency changes as more questions are posed to the learner. The number of questions appears in the x-axis and the estimated proficiency appears in the y-axis. The estimation is initially equal to 0. By examining the figure, we notice that after only 13 questions the algorithm has converged and the changes to the learner's estimated proficiency are very small.

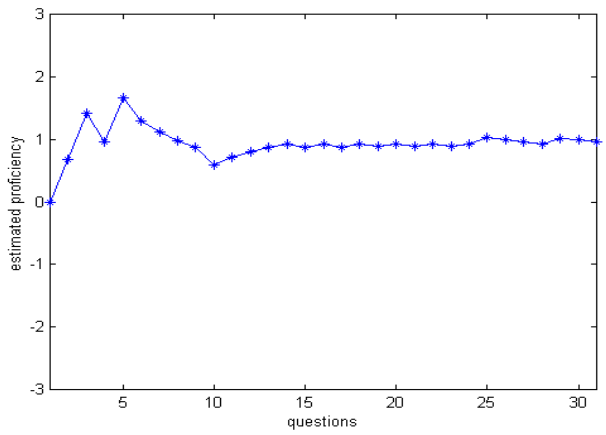

Figure 3: Estimation of the learner's proficiency plotted against the number of questions he/she has answered

The assessment procedure is terminated once any of the termination criteria has been met. These criteria have been specified by the tutor, with the aid of the QTE, as part of the test specifications. Two possible termination criteria are: if the number of questions posed exceeds the maximum number of questions allowed and if the accuracy of the estimation of the learner's proficiency reaches a desired value.

CONCLUSIONS AND FURTHER RESEARCH

In this paper, we have presented how adaptive assessment can be used to support learners during studying and to facilitate their interaction with the adaptive educational hypermedia system INSPIRE. The general framework for the assessment has been provided and the algorithm used to implement the adaptive assessment has been covered in greater detail.

The main design considerations of our research work that were discussed in the paper are:

- Adaptive assessment can provide accurate estimation of the learner's proficiency in an efficient way without forcing him/her to answer questions that are either too easy or too difficult for him/her.
- The accurate estimation is essential for an adaptive educational hypermedia system as it is used to guide the adaptation of the system.
- Accurate estimation takes into consideration the importance of the different knowledge modules providing tutors with the option to define weights on each knowledge module.
- Providing opportunities for self-assessment and summative assessment motivates learners to reflect on their learning and to control their studying effectively.
- Offering extensive feedback enhances the learning process.

Future plans include the implementation of the entire adaptive assessment mechanism and its integration as part of the diagnostic module in INSPIRE. The effectiveness of the adaptive approach will be evaluated and compared to fixed-question tests. To this end, experiments with real users will be performed in order to test whether adaptive assessment leads to improved interaction between the learner and the system. Also, our future work includes research on the operation of the Presentation Module. The goal of this research will be to find the optimal method of presenting the various types of questions and their feedback, so as to assist the user in the assessment and facilitate the interaction between the system and the user. Finally, a remaining question under investigation is how easily the tutors could define the required questions and the corresponding parameters, which are needed for the assessment and whether the use of the Question and Test Editor component could make this task easier.

BIBLIOGRAPHY

1. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M., Novick, M.R. (eds.) *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA, 1968.

2. Bransford, J.D., Brown, A.L., Cocking, R.R. (eds.) *How People Learn. Brain, Mind, Experience, and School.* National Academy Press, Washington, 1999.

3. Brusilovsky, P. *Methods and Techniques of Adaptive Hypermedia.* User Modeling and User-Adapted Interaction, Vol. 6, No. 2-3, 1996, pp. 87-129.

4. Brusilovsky, P. *Adaptive and Intelligent Technologies for Web-based Education*. In: Rollinger, C., Peylo, C. (eds.) Kunstliche Intelligenz, Special Issue on Intelligent Systems and Teleteaching, Vol. 4, 1999, pp. 19-25.

5. Collins, J., Greer. J., Huang, S. Adaptive Assessment Using Granularity Hierarchies and Bayesian Nets. In: Frasson, C., Gauthier, G., Lesgold, A. (eds.) *Intelligent Tutoring Systems, Third International Conference, ITS'96* (June 12-14, Montreal, Canada), Springer, 1996, pp. 569-577.

6. Foster, D. F. *Computerized Adaptive Testing.* Certification Magazine, December 2000. Available from http://www.galton.com/documents.

7. Grigoriadou, M., Papanikolaou, K., Kornilakis, H., Magoulas, G. INSPIRE: An Intelligent System for Personalized Instruction in a Remote Environment, *8th International Conference on User Modeling (UM2001), Third Workshop on Adaptive Hypertext and Hypermedia,* Germany, 2001.

8. Huang, S. A Content-Balanced Adaptive Testing Algorithm for Computer-Based Training Systems. In: Frasson, C., Gauthier, G., Lesgold, A. (eds.) *Intelligent Tutoring Systems, Third International Conference, ITS'96* (June 12-14, Montreal, Canada), Springer, 1996, pp. 306-314.

9. Kommers, P., Grabinger, S., Dunlap, J. *Hypermedia Learning Environments. Instructional Design and Integration.* Lawrence Erlbaum Associates, 1996.

10. McCormack, C., Jones, D. *Building a Web-based Education System.* Wiley Computer Publishing, 1997.

11. Merril, D. Component Display Theory. In: Reigeluth, C.M. (eds.) *Instructional design theories and models: An overview of their current status*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

12. Mislevy, R. J. *Some recent developments in assessing student learning*. Educational Testing Service, Princeton, NJ, 1996.

13. Lord, F.M. *Application of Item Response Theory to Practical Testing Problems.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1980.

14. Papanikolaou, K., Magoulas,G., Grigoariadou, M. A Connectionist Approach for Supporting Personalized Learning in a Web-based learning Environment. In Brusilovsky, P., Stock, O., Strapparava, C., (eds.), *Adaptive-Hypermedia and Adaptive Web-based Systems*. Lecture Notes in Computer Science, Vol. 1892. Springer-Verlag, Berlin, 2000, pp. 189-201.

15. Ponsoda, V., Wise, S. L., Olea, J., Revuelta, J. *An investigation of self-adapted testing in a Spanish high school population*. Educational and Psychological Measurement, Vol. 57, No. 2, 1997, pp. 211-221.

16. Rios, A., Millan, E., Trella, M., Perez-de-la-Cruz, J., Conejo, R. Internet Based Evaluation System. In: Lajoie S., and Vivet M., (eds.) *Artificial Intelligence in Education, Open Learning Environments, New Computational Technologies to Support learning Exploration and Collaboration, Frontiers in Artificial Intelligence and Applications*, IOS Press, 1999, pp. 387-394.

17. van der Linden, W. J., Hambleton, R.K. (eds.) *Handbook of Modern Item Response Theory*. Springer-Verlag, London, 1997.

18. VanLehn, K. Student Modeling. In: Polson, M. C., Richardson, J.J. (eds.) *Foundations of Intelligent Tutoring Systems*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988, pp. 55-78

19. Vispoel, W. P., Rocklin, T. R., Wang, T. *Individual differences and test administration procedures: a comparison of fixed-item, computerized-adaptive, and self-adapted testing.* Applied Measurement in Education, Vol. 7, No. 1, 1994, pp. 53-79.

20. Weiss, D. J. *Improving measurement quality and efficiency with adaptive testing.* Applied Psychological Measurement, Vol. 6, 1982, pp. 473-492.