Desirable features of educational theorem provers - a *Cognitive Dimensions* viewpoint

Gada Kadoda — Bournemouth University

Roger Stone — Loughborough University

Dan Diaper — Bournemouth University

Keywords: Theorem Provers, Proof Systems, Usability Analysis, Cognitive Dimensions.

Extended Abstract

Recently, the necessity of bridging the gap between the established areas of theorem provers and their user interface has been given more attention. The relative complexity of the theories underlying theorem provers makes them inaccessible to a wide range of software engineers who are not experienced mathematicians. Improving these systems' user interfaces may play an important role in stimulating their use by more software practitioners. This paper concentrates on the use of theorem provers for educational applications, i.e. where their purpose is as tools to teach theory proof.

This paper reports on the results of a questionnaire-based study that utilised Green's Cognitive Dimensions' framework [Gre89,96] to analyse the usability of theorem provers. The results of the study consist of four parts. First, it provides a survey of features supported by state-of-the-art theorem provers in the context of Green's dimensions. Second, a defined relationship between the dimensions in the context of theorem provers may pave the way for establishing the framework as an evaluative technique for such systems. Third, the analysis of the results uses the orthogonal nature of the dimensions to examine how changes to one dimension affect another, or what Green [Gre96] referred to as the 'standard remedies and their trade-offs'. A desirable feature in theorem provers, for instance, is to allow reuse of previously

proven theorems (abstraction). This can create problems of consistency with the system handling built-in rules in a different way to user-introduced rules. Another way to look at the interaction between the dimensions is by considering the intended use of a theorem prover and the application domain. For example, design decisions for an educational system are bound to require a different view of the dimensions 'trade-off' adjustment compared to a commercial system.

Fourth, the paper will use the questionnaire study data to evaluate the OPEn proof editor [Sto94]. In particular, the 'trade-off' adjustment mentioned above can provide a basis for comparing relatively more powerful proof systems to identify the extent to which the Loughborough built proof editor OPEn supports the appropriate concepts in its intended educational context.

A list of user interface issues for theorem provers was assembled through an extensive literature review on the subject. For instance, structuring of proofs, management of long listings of proofs, amount of feedback given to users [Lin88], user vs. system control of the proof, and the system's ability to project different views of formal objects at different levels of detail [Jon91]. In addition, Norman's stages of user activities model [Nor86] was used to examine the interface of a typical theorem prover, identifying even more user interface issues, and to present the complete set of these in an organised manner [Kad97a].

This set was used to construct a questionnaire that is split into two main sections. The first section is used to acquire miscellaneous information about the interface of the theorem provers. For example, classification in terms of the software development life-cycle support, interaction style, application domain and purpose. The second section is used to examine the interface of the theorem provers in terms of Green's cognitive dimensions.

The survey questionnaire was distributed to the developers and users of 27 proof systems. The response rate was approximately 63% in total covering 17 different proof systems. Descriptive statistics was initially used to give an overview of the sample, the Pearson Correlation Coefficient was then used to examine the association between the variables. An ANOVA test was used to further investigate the inter-relationships of the variables. An overall view of the results is illustrated in Figure 1 which shows the features that can be supported by a theorem prover for it to have a high performance on a particular dimension. Full listings of the results can be found in [Kad97b].

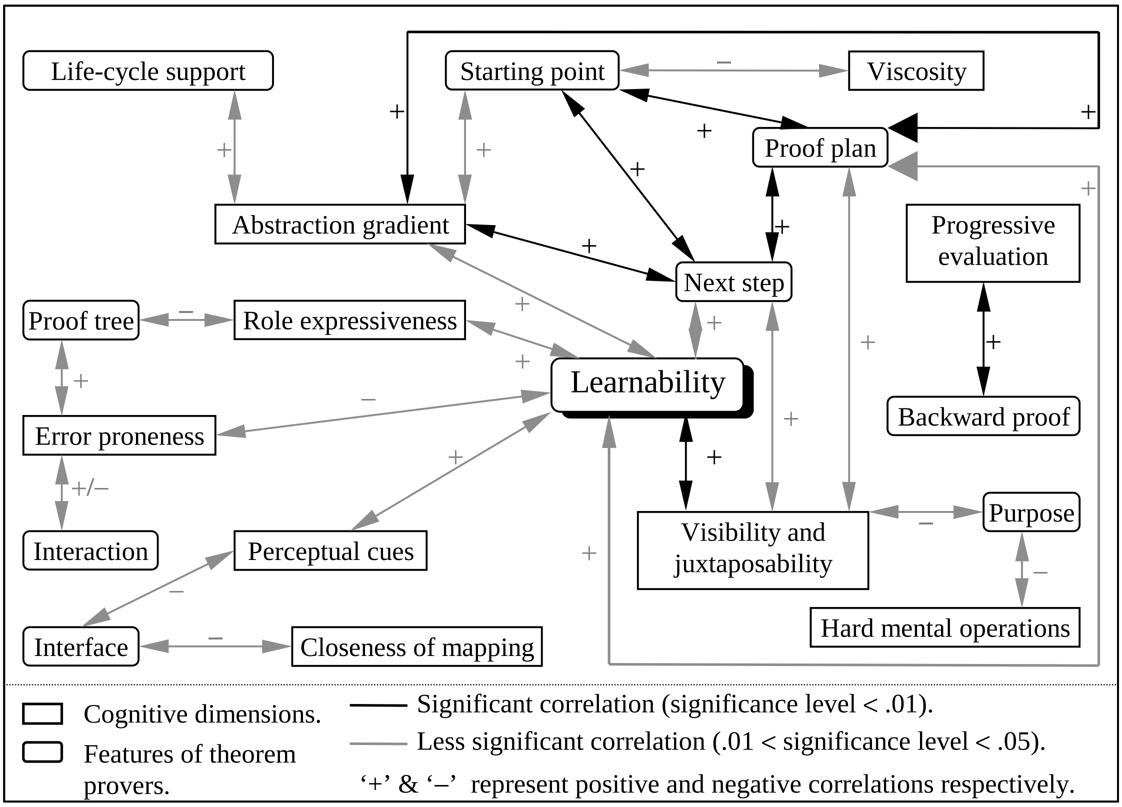

Figure 1 *[Relationships among theorem provers features and Green's Cognitive Dimensions]*

The results of the study highlighted a list of desirable features for a proof system to be successful as a teaching tool. These are shown in a form of a checklist (Table 1) that displays the type of support to be provided by a system and a set of 'trade-off' relationships.

high ☑	*low* ☒	*no effect*
• assistance(proof plan, next step) • abstraction gradient • visibility and juxtaposability • visibility and juxtaposability • perceptual cues • meaningful error messages (error proneness) • role expressiveness (obvious proofs and clear substructure of lemmas) • consistency (handle all classes of rules in the same way)	• viscosity • diffusiveness (large proof scripts)	• domain • automation • life-cycle support • building a proof tree • support of backward proof • development record keeping • closeness of mapping • secondary notations • hard mental operations

Table 1 *[Checklist for a proof system that has a high learnability rating]*

The evaluation of the OPEn proof editor with regard to other proof systems in the sample is carried out in two stages. Firstly, the evaluation is approached from the viewpoint that if OPEn responses lie within the majority's responses (40 - 60 %), then the editor supports the right concepts. Secondly, by examining OPEn against the checklist in Table 1.

For 3/4 of the questions the response for OPEn lies within the majority of responses for the particular question. The system however was found lacking in:

- ♦ generalisation of proof (abstraction gradient);
- ♦ proof visualisation (visibility and juxtaposability);
- ♦ display of error messages.

The critique of the study is mainly based on the comments provided by developers and users of the theorem provers who responded to the questionnaire. Firstly, the subjective nature of some of the questions was identified as the principal issue in the comments. Secondly, the inapplicability of the some questions to all theorem provers. Lastly, being unfamiliar (in the practical sense) with many of the systems made it difficult to evaluate the responses regarding those systems. The answers being provided mostly by members of the development teams of theorem provers raises the issue of the differences of what the designers think is not necessarily what the users find.

According to Shackel, there are three types of measurements of usability as there are three procedures of its evaluation [Sha91]. The cognitive dimensions framework can be categorised as an attitude criteria that assesses the users' view of the system and relies on their judgements, and the study conducted used expert reviewers. However, neither the type of measurement used in this study nor its procedure are sufficient on their own for evaluating the usability of theorem provers. Yet, the different concepts behind the cognitive dimensions framework as a usability evaluation technique and its dimensions' 'trade-off' feature has illustrated the potential of the framework to serve as a discussion tool, which could be considered as a merit of the study.

Future work could investigate trade-offs between dimensions through a more controlled study with some focus on the purpose and subject group factors. This could yield different sets of trade-off relationships between dimensions depending on either factor which would shed light on: what are the issues to be considered when building a educational system or commercial system; identifying where user and designer views do conflict so as to help the latter in building systems to match the former.

References :

[Ait95] Aitken J. S. Gray P. Melham T. and Thomas M. Interactive Theorem Proving : An Empirical Study of User Activity, http://www.dcs.gla.ac.uk/~stuart/ITP/ITP.html, To Appear in *Journal of Symbolic Computation*, 1995.

[Gre89] Green T. R. G. *Cognitive Dimensions of Notations*, *People and Computers V* : 443-460, Edited by Sutcliffe A. and Macaulay L., 1989.

[Gre96] Green T. R. G. Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework, *Journal of Visual Languages and Computing*, 7 : 131-174, 1996.

[Jon91] Jones C. B. Jones K. D. Lindsay P. A. and Moore R. *Mural: A Formal Development Support System*, London : Springer-Verlag, 1991.

[Kad97a] Kadoda G. and Stone R.G., Cognitive Dimensions Analysis of Theorem Provers, Proceedings of the 3rd International Workshop on User Interface Design for Theorem Proving Systems: France, 1997.

[Kad97b] Kadoda G. Formal Software Development Tools (An investigation into usability). PhD Thesis, Loughborough University, 1997.

[Lin88] Lindsay P. A. A survey of mechanical support for formal reasoning, *Software Engineering Journal* 3(1) : 3-27, 1988.

[Mer96] Merriam N. A. *Two Modelling Approaches Applied to User Interfaces to theorem Proving assistants,* Proceedings of the 2nd International Workshop on User Interface Design for Theorem Proving Systems: University of York, UK, 1996.

[Nor86] Norman D. A. *Cognitive Engineering,* In: *User Centred System Design - New Perspectives on Human-Computer Interaction* : 31-61, Edited by Norman D.A. and Draper S. W., 1986.

[Sto94] Stone R. G. *OPEn - an Outline Proof Editor*. Loughborough : Department of Computer Studies, Loughborough University, UK, 1994.