Cuartero-Olivera, J., Hunter, G. and Pérez-Navarro, A. (2012). Reading and writing mathematical notation in e-learning environments. *eLC Research Paper Series, 4, 11-20.*

#02

READING AND WRITING MATHEMATICAL NOTATION IN E-LEARNING ENVIRONMENTS

J. Cuartero-Olivera[1], Gordon Hunter[2], A. Pérez-Navarro[1]

(1) jcuartero@uoc.edu
Universitat Oberta de Catalunya
(2) G.Hunter@kingston.ac.uk
Kingston University
(3) aperezn@uoc.edu
Universitat Oberta de Catalunya

Reading and writing mathematical notation in e-learning environments

ABSTRACT

How do students and teachers communicate mathematics via the internet? Why do they use these methods? Is there any better way of communicating mathematics via the internet? In addition to the time needed to understand a concept, it is also a challenge for students to write formulae in e-learning environments, since most computers and software are not designed to write formulae. Furthermore, most physics, mathematics and engineering students do most of all their initial analysis and calculations using pen and paper and then have to translate it into a computer environment. Does this extra time investment play a role in the academic results achieved?

This paper presents exploratory research into the different methods used by teachers and students to communicate mathematics via the internet and to use appropriate patterns according to the different subjects and knowledge areas. It explores the reasons that make students choose one method or another and analyses the extreme case: when students write mathematical formulae on paper and then scan this electronically.

The analysis is carried out on engineering subjects at the Universitat Oberta de Catalunya (UOC) in which mathematics plays an important role: 17,000 emails are analysed and five physics teachers are interviewed as part of a qualitative study about handwritten scanned exercises.

This paper shows that the key to explaining students' behaviour is the time factor. In order to reduce the time required to write the required mathematical formulae, the paper proposes a speech-to-text tool, such as *TalkMaths*, to help students create and edit mathematical formulae, since speech is the fastest and most natural way of communicating.

Cuartero-Olivera, J., Hunter, G. and Pérez-Navarro, A. (2012). Reading and writing mathematical notation in e-learning environments. *eLC Research Paper Series, 4, 11-20.*

KEYWORDS

e-learning, mathematical notation, mathematics communication, TalkMaths.

INTRODUCTION

In general, virtual learning environments still face challenges today, as they have to overcome time and distance barriers between students and tutors. This is true for teacher-student communication, as well as within student work groups, or when students communicate their queries to the teacher. Focusing on knowledge transfer, one of the main problems regarding scientific and engineering studies is communication between university members (both students and teachers), as a large part of this communication requires frequent and extensive use of mathematical notation.

In a virtual learning environment, the issue of learning and communicating mathematics can be compared to a disability such as visual impairment. Visually impaired students can listen to the teacher reading a given formula, although they cannot easily learn how it is expressed visually, even in an on-site learning environment (Fitzpatrick, 2007). In contrast, a virtual learning environment allows students to see a visual expression of a formula, but traditionally does not provide them with a verbal representation, which is in turn a handicap for visually impaired students. In addition, in some cases there are no auxiliary tools to help people express mathematical notation, as for example a user-friendly formula editor. When no such tool is available, the methods for expressing mathematics are still computer-aided, but they are as rudimentary as using plain text descriptions or file attachments (e.g. formulae saved as images). Issues regarding the use of mathematical notation in IT have already been described and researched (Zhao, 2008; Chen & Okada, 2001).

Wigmore, Hunter, Pflügel and Denholm-Price (2009) set out an alternative way to “write” mathematical notation within a computer-based document: by using speech. People usually acquire a very high level of competence in speaking, listening to and understanding their native language(s) at a very early age. This contrasts with the normal way in which we tend to interact with computers, mainly using keyboard-based textual communication and initiating actions and events by clicking on icons by using a mouse. This contrast increases when dealing with mathematical notation, since the keyboard is not ideally designed for writing mathematical formulae.

The following questions then arise: How do students and teachers communicate mathematics via the internet? Why do they use these methods? Is there any better way of communicating mathematics via the internet?

The objective of this work is: to present exploratory research into the different methods used by teachers and students to communicate mathematics via the internet, and to look for the reasons that make students choose one method or another.

This article is structured as follows: The Method section describes the method followed to analyse teachers’ and students’ behaviour when communicating mathematics via the internet. The findings are presented in the Results section. The Discussion section analyses the reasons why students behave in the way they do when writing mathematics and time considerations are put forward as the key factor in explaining their behaviour. A speech-to-text tool, *TalkMaths,* is then proposed as an improved method for writing mathematical formulae, and we attempt to illustrate the extra benefits of such a tool. The paper ends with our conclusions and suggestions for future work.

METHOD

This section describes the method followed to investigate how students and teachers behave when communicating mathematical formulae. We show who the participants in the study are, the material used and procedure followed.

PARTICIPANTS

The participants in this study are UOC students taking the courses listed in the subject column in Table 1. All these courses correspond to the courses designated as a whole as Computer Sciences and Telecommunications at the UOC.

MATERIAL AND PROCEDURE

The procedure followed in this paper is to observe students' behaviour within the UOC's virtual classrooms and analyse how they choose to write mathematical formulae. Since most interaction between classroom members in technological subjects at the UOC happens in the classroom forum or discussion board, only these forums are analysed in this research project.

Our method sets out to: 1) Observe and identify emails within the virtual forums for subjects with a high level of mathematical content; 2) Classify how students try to write or represent mathematical notation when there is no specialised mathematical editor available; 3) Analyse the results obtained; 4) Analyse an extreme case: those students who do not use computer tools to write mathematical formulae at all; 5) Set out a hypothesis of the key factors making students choose one option or another; and 6) Put forwards a solution for simplifying the introduction of mathematical formulae within internet-based communication.

A quantitative study was carried out covering Points 1, 2 and 3, and over more than 17,000 email messages were analysed over the course of four teaching semesters.

In addition, a qualitative study was carried for Point 4, in order to find out why some students usually deliver their solutions to exercises as scanned versions of handwritten answers. Five teachers of physics-based subjects were consulted about this point, and asked whether there is any correlation between the way students submit the exercises (handwritten or scanned) and their knowledge of the topic.

RESULTS

Analysis of the findings of our study reveal that students usually use four ways to express mathematical formulae within their electronic communications :

- **Full mathematical formulae:** The first and most common way of expressing mathematical notation is by typing out full mathematical formulae, where each one may be an equality (e.g. a=b+3), an inequality (e.g. a+2>5), or a mathematical expression containing more than one mathematical symbol or function (e.g. sin(ln(1)). (We exclude examples consisting of just a single symbol from this category.) Full mathematical formulae can also be expressed in any specific syntax compatible with programming languages or software commonly used in scientific and engineering environments. Such marked-up formulae are also considered in this category, for example specific mark-up codes which are meaningful for the LaTeX2 editor, such as "\sqrt", which would convert the encoded expression \sqrt{1-e^2} into $\sqrt{1-e^2}$

- **Individual mathematical symbols:** The "mathematical symbol" method consists of writing just one mathematical symbol at a time, whether it is in plain text (e.g. "lambda") or using the symbol itself (e.g. "λ"). Numerical expressions were only considered within this group if they are preceded (or followed) by a mathematical symbol (e.g. " >10" or "10!"). Superscript indices (e.g. "x^3", subscript indices (e.g. "a_5"), and commonly used abbreviations of mathematical functions, such as SQR, TAN, etc., are also put into this category.
- **Formula referencing:** Formula referencing (or formula citing) is used whenever a certain formula or expression is cited within the text, be it in a very colloquial, descriptive way (e.g. "the formula in the first paragraph of page 24") or by using a previously established citation system, such as numbers indexing the formulae (e.g. "the second term in equation 17").
- **File attachment:** The final method consists of attaching a file containing the formulae referenced in the email body (e.g. "the attached formula describes the relationship between the decibel rating of the system and the ratio of output signal power to input signal power"), or even writing the whole body of the message into an attached file.

Table 1 shows the findings obtained. In order to collate the results, we first considered the

Table 1: Classification and number of emails, according to subject area and type of mathematical notation used.

Subject	Total	Formulae		Symbol		Citation		Attachment	
	#	#	%	#	%	#	%	#	%
Algebra	332	235	71	77	23	29	9	24	7
Automata Theory and Formal Languages I	128	57	45	63	49	2	2	9	7
Computes Structure and Technology	287	175	61	101	35	6	2	11	4
Cryptography	224	155	69	49	22	36	16	10	4
Discrete Mathematics	244	158	65	71	29	29	12	12	5
Engineering Physics Fundamentals	228	107	47	52	23	127	56	10	4
Introduction to Mathematics for Engineering	194	147	76	22	11	10	5	23	12
Linear Systems	316	224	71	60	19	14	4	30	9
Mathematical Analysis	425	307	72	120	28	66	16	55	13
Mathematics I	352	250	71	48	14	12	3	54	15
Probability and Statistics	218	170	78	21	10	13	6	24	11
Statistics	331	228	69	83	25	45	14	21	6
Technological Fundamentals I	216	136	63	73	34	4	2	9	4
Technological Fundamentals II	328	203	62	78	24	64	20	31	9
Wiris Laboratory (Algebra)	232	127	55	49	21	4	2	67	29

total number of emails sent by students for each subject, including those containing no mathematical notation, but which were directly related to that particular subject. It can be seen that over all the subject areas, 17 % of the total emails contained some mathematical notation, but in some subjects, the proportion was found to be as high as 45 %.

If we only consider those emails which contained mathematical expressions for each subject, as in Table 1, major differences can be observed in the frequencies and proportions of the various types of mathematical notation employed (according to the previously defined classification scheme) by students of different subjects:

- Regarding the use of full mathematical formulae within the text of the email, the results ranged from 45 % in Automata Theory and Formal Languages to 78 % in Probability and Statistics.
- In the case of the use of single mathematical symbols, the usage percentage varied from 10 % in Probability and Statistics to 49 % in Automata Theory and Formal Languages.
- Regarding formula referencing, the percentage of use ranged from just 2 % in each of Automata Theory and Formal Languages, Wiris Laboratory (used in the Algebra subject) or Computer Architecture and Technology to 56 % in Engineering Physics Fundamentals.
- For the use of attachments containing mathematical notation, the percentage ranged from 4 % in each of Cryptography, Engineering Physics Fundamentals, Computer Architecture and Technology or Technological Fundamentals I to 29 % in Wiris Laboratory (Algebra).

From these results we can see that including full mathematical formulae and using symbols within the text of the message are generally the most commonly used ways of communicating formulae by UOC students in electronic messages. However, one subject area is an exception to this trend: Engineering Physics Fundamentals. In this case students mainly use the formula referencing method. Why is this ? And why do students choose one particular method rather than another?

To answer these questions, we can look at the opposite view: why do some students using the virtual classroom handwrite their exercises, then scan their solutions, instead of writing them directly using a computer, even though most modern word processors actually include an equation-editing facility?

EXERCISES WITH MANUSCRIPT FORMULAE WITHIN ONLINE ENVIRONMENTS

When students deliver their solutions to exercises to their tutors, we sometimes find that some of them write their solutions by hand, then scan the handwritten versions to produce an electronic document which they can attach to an email. To analyse this observation, we focus on three different physics-based subjects which each have about 70 students. Two facts about these courses can reveal the main reason why students choose one way or another for communicating using mathematical notation:

1. Students are penalised in terms of the grade they receive for the exercise if they submit scanned versions of their handwritten formulae. The rational behind this is that students who submit their exercises written using a computer have invested extra time and effort to do so. Despite this penalty, some students still submit scanned versions their handwritten solutions.
2. There is no correlation between students' knowledge and how they deliver their exercises: according to teachers, students only submit scanned versions of handwritten

solutions when they do not have sufficient time to write them using a computer.

Although there is no correlation between the quality of a student's solution to an exercise and the way it is delivered to the tutor, if the student submits a scanned version, the teacher can infer that that student had insufficient time to complete the exercise properly. In fact, in one of the subjects analysed, six out of the eight students who submitted some handwritten exercises subsequently left that course.

DISCUSSION

Why do students employ one particular method or another to write mathematical formulae? To answer this question, we focus on the extreme situation: those students who handwrite and then scan the formulae they use. In this case, the reason given by the students was the lack of time they had to complete the exercise, so we could therefore argue that students usually choose the path which will take the shortest time to write the formula.

Taking the time required as the key factor for writing mathematical formulae, we are able to explain why students choose one particular way or another for communicating mathematics:

- Use of full mathematical formulae within the text is the commonest method for communicating mathematical notation, but this is not really suitable for complicated structures. Why? Because this approach allows the students to write the formulae very quickly. We have to account of the fact that conventional keyboards are only designed for typing text, and not mathematical formulae, so writing a formula in a text-based way appears to be a straightforward solution.
- Use of individual symbols: many times, students were found to "write a symbol" using its name, instead of looking for it in an enriched text character set. Why? Because it is faster to write, say, "alpha" rather than looking for the corresponding symbol.
- Use of attachments: Students use file attachments for their formulae only rarely, and only when this is strictly necessary. Why do they only rarely use attachments? Because writing a document is itself quite slow, and so sending such a file attached to an email probably requires more time than writing the formulae directly into the text.
- Formula citation: to say "formula (number)" is the fastest way to "write" a formula, but this method only appears to be commonly used in the course Engineering Physics Fundamentals. Why does the use of this method occur in as many as 56 % of emails for this course? Because all the equations within the study material, as well as all the equations in the model solutions to the exercises for this subject are numbered by the tutor.

These results show that distance-learning engineering students in fact face a problem of time when dealing with mathematical notation in exercises. But these results also show that we do not currently offer a solution to the problem of writing mathematical formulae since, as has been shown, when students are short of time, they present a scanned copy of formulae handwritten on paper.

"Campus forums" do not currently provide a formula editor, and this could explain why students use the four ways described above to write formulae within the forum. However, when writing an exercise at home, they could use a word processor and its mathematical editor. Why do they not do this? Why do they (and teachers) first write out the solution on paper, and why is it that only once they have confidence in their solution does it get written using a computer? Students would probably not accept such behaviour from the teacher in class or in a real-time online interaction!

We think that the most likely explanation for this behaviour is the time factor, as shown previously. What could be an appropriate solution?

Since students write their exercises by hand, the first answer should be to use a tablet PC or an interactive pen display, like Wacom™, possibly in combination with a handwritten character recognition system which can handle mathematical symbols. But, since the main problem is the time factor, would this be the fastest way for introducing mathematical formulae into electronic typed documents on a computer?

However, there are other possibilities. Wigmore, Hunter, Pflügel, and Denholm-Price (Wigmore, 2009; Wigmore, 2010) suggest that speech is a way of introducing mathematical formulae into documents. Speech is the most natural way most people communicate. In fact, at a very early age children are able to speak with a very high level of competence. Writing is usually the next easiest for people to use: it is acquired later and it is harder to learn than speech, but is actually something that becomes quite natural - the way we write is often very close to the way we think.

However, speech and writing both contrast with the normal way in which we tend to interact with computers. For this, we usually use a keyboard and/or a mouse. The keyboard is text based and is very useful for writing text; and the mouse is useful for starting actions and events and positioning the cursor. But when dealing with mathematics, the keyboard and mouse are very far removed from the way we think: if we use a typesetting language, such as LaTeX, which is text based, we need to know the appropriate keywords to obtain the relevant symbols and formatting; and if we use an equation editor within a word processor, we need to know where the appropriate buttons are and to think not only of the formula we want to write, but also about the strategy and the path we need to follow in order to write it.

All these issues increase the complexity and consequently the time needed to write mathematical formulae using computers. As speech is the most natural way most of us have to communicate with each other, it would appear to be potentially the best and fastest way for us to communicate with computers.

But, is it possible to communicate with computers by speech? The answer is yes, since speech interfaces are now realistically feasible (McTear, 2004). And is it possible to "speak mathematics" to a computer? The answer again is yes, as illustrated by the *TalkMaths* project developed at Kingston University by Hunter, Wigmore and their partners.

THE TALKMATHS PROJECT

According to the creators of *TalkMaths*, "the ultimate aim of our research is to design, implement, test and evaluate a speech-user interface which will make performing many computer-based tasks, in the context of learning and teaching mathematics, easier through spoken commands by using and extending existing technologies" (Wigmore, 2009).

There are some other products able to translate spoken mathematics into a computer-based text format, for example *MathTalk*™ (Metroplex, 2007), and *Math Speak & Write* (Guy, 2004), but the former is not freely available and the latter is also not totally free, and also supports a rather smaller set of mathematical vocabulary than *TalkMaths* does.

TalkMaths is a speech-user interface able to transform the mathematical formulae spoken by the user into LaTeX (LaTeX, 2011) or MathML (Carlisle, 2003). Figure 1 shows the process followed by *TalkMaths*. The user dictates formulae into a microphone and the front-end automatic speech recognition system interprets the speech signal, performing a frequency-time analysis and partitioning it into phonemes. It then compares the proposed phoneme sequence with entries in a database of standard pronunciations of relevant words (the acoustic model). Then, in the language model, using grammars and statistical information about word frequencies and particular word sequences ([López-Cozar & Araki, 2005) it determines what the user is likely to be saying. Finally, the systems offers the user the best-fitting word sequence, or an ordered list of several best-fitting word sequences so the user can choose the correct option.

The question that arises is whether *TalkMaths* will save students time when writing mathematical formulae with a computer. Since saying a word is, for most people, faster than typing it or using the mouse to select it from a list, we think that the answer will be yes.

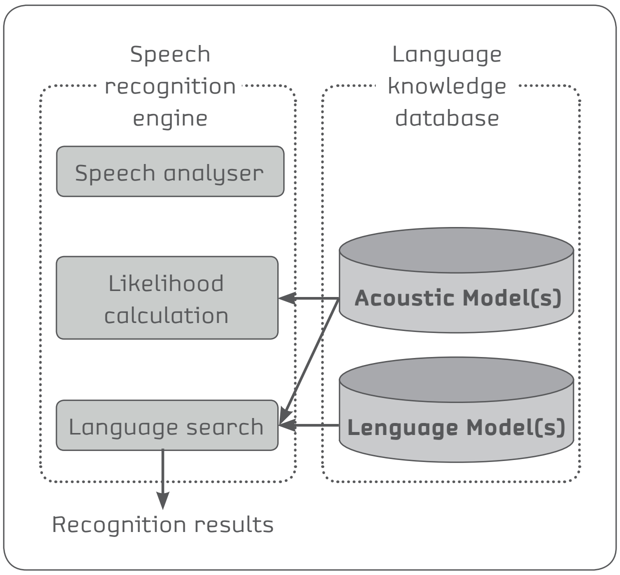
Speech as the standard modality for human-computer interaction?

In this point we suggest that speech may be the fastest way to introduce formulae into an electronic document, and also that it is technologically possible to make this a standard way of communicating when dealing with formulae on computers. However, we can go a step further, since the use of speech opens mathematical formulae up to new people and new environments, for whom and where writing a formula could be near impossible or at least require a huge amount of time and effort. Some examples are:

- **Visually or motor impaired people:** for them, speech is nearly the only way to create or edit a mathematical formula in an electronic document within a reasonable space of time.
- **Students and teachers who use their smartphone to communicate with each other:** due to the size of the device, typing with a smartphone is somewhat harder than typing with a computer, but when the text contains the complexities of a mathematical formula, typing becomes nearly impossible. Thus, students and teachers may have to wait until they are in front of a regular computer before they can write formulae into electronic documents and they therefore cannot do this whenever or wherever they want. This problem could be solved by introducing facilities for processing spoken formulae into the smartphone or similar devices.

Initial user trials of *TalkMaths* (Wigmore, 2010; Wigmore, 2011) have suggested that a spoken interface is a practical method for students with impaired use of their hands and arms to create and edit mathematical formulae in electronic documents, when compared with a more conventional interface based on use of the keyboard and mouse.

Figure 1. The speech recognition process.



Source: (Wigmore et al., 2009)

Since the main challenge for current students appears to be one of time, the solution should be a way that allows students to communicate mathematics easily and quickly using computers. Since the fastest way for people to communicate with each other is by spoken language, we believe the solution should be a feature that allows students to communicate mathematical notation using their voice.

CONCLUSION

In this paper, we have shown how students communicate mathematical formulae within a totally virtual environment at the UOC. We have seen that students usually write mathematical formulae within emails either using plain text or via citation, when the subject of interest allows citing of formulae, and only rarely use attached files. We have also seen that students only submit scanned versions of their handwritten solutions to exercises when they are short of time and need to complete the work quickly. All these indicators lead us to suggest that students choose the fastest convenient method to write mathematical formulae, i.e. the time factor is key when choosing how to communicate mathematical formulae.

As a solution to improve mathematical communication within electronic documents, speech is put forward as a more natural and potentially faster way to “write” a formula on a computer. We have shown that this is a technologically possible solution and have introduced *TalkMaths*, software which is able to translate spoken mathematics into mathematical text in standard notation and layout.

Although *TalkMaths* is still in the prototype phase, it should allow us to overcome major difficulties in the introduction of mathematical formulae to visually and motor impaired people, who can barely write or edit formulae with a keyboard or mouse; but also for teachers and students who usually work with their smartphones, where it is difficult to type even plain text.

Thus, speech is not only potentially a faster way to introduce mathematical formulae into documents; it also allows communication using mathematical notation anywhere and any time.

FUTURE WORK

Although this paper shows the possibilities of speech as a way to communicate using mathematical notation, more work should be carried out in order to test whether it would actually be the preferred option for students and whether they would find it a faster way for creating and editing formulae in electronic documents than more established methods.

An improved version of *TalkMaths*, to be deployed within a “virtual campus” and to allow the writing of mathematical formulae within emails, should be developed.

Another important development would be to complement *TalkMaths* with a text-to-speech tool (Sancho et al., 2009), so that visually handicapped people and virtual students could hear and interpret mathematical expressions which they (and other people) had written.

Cuartero-Olivera, J., Hunter, G. and Pérez-Navarro, A. (2012).
Reading and writing mathematical notation in e-learning
environments. *eLC Research Paper Series, 4, 11-20.*

References

Carlisle, D; Ion, P.; Miner, R.; Poppelier, N. (Editors) (2003). Mathematical Markup Language (MathML) Version 2.0 (Second Edition). Date of access: 03/18/2012. Available from: http://www.w3.org/TR/MathML2/.

Chen, Y. & Okada, M. (2001). Structural Analysis and Semantic Understanding for Offline Mathematical Expressions. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 15, No. 6, pp. 967-988, Sept. 2001.

Guy, C.; Jurka, M.; Stanek, S.; Fateman, R. (2004). Math Speak & Write, a Computer Program to Read and Hear Mathematical Input. Date of access: 03/18/2012. Available from: http://www.cs.berkeley.edu/~fateman/msw/AcademicPaper.pdf

López-Cozar, R. & Araki, M. (2005) Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment. *John Wiley & Sons, Ltd.*

Fitzpatrick, D. (2007). Teaching Science subjects to Blind Students. Seventh IEEE International Conference on Advanced Learning Technologies (ICALT)

LaTeX (2011). A document preparation system. Date of access: 01/05/2011. Available from: http://www.latex-project.org

McTear, M.F. (2004) Spoken Dialog Technology: Toward the conversational user interface, *Springer Verlag* (London).

Metroplex Voice Computing (2007). Date of access: 18/03/2012. Available from: http://www.axistive.com/metroplex-voice-computing.html

Sancho-Vinuesa, T.; Córcoles, C. ; Huertas, M.A.; Pérez-Navarro, A.; Marquès, D.; Eixarch, R.; Villalonga, J. (2009). Automatic Verbalization of mathematical formulae for web-based learning resources in an on-line environment, *INTED2009 Proceedings,* pp. 4312-4321

Wigmore, A. M. ; Hunter, G. J. A.; Pfluegel, E.; Denholm-Price, J. C. W. & Binelli V. (2009) *"'Let Them TalkMaths !' Developing an Intelligent System to Assist Disabled People to Learn and Use Mathematics on Computers through a Speech Interface : the TalkMaths and VoiceCalc Systems"*, 5th AAAI International Conference on Intelligent Environments (IE'09), Barcelona, Spain, July, IOS Press

Wigmore, A. M. ; Hunter, G. J. A.; Pfluegel, E.; Denholm-Price, J. C. W, Colbert, M.. (2010) "TalkMaths Better ! Evaluating and improving an intelligent interface for creating and editing mathematical text". Proceedings of 6th International Conference on Intelligent Environments, July 2010, Kuala Lumpur, Malaysia

Wigmore, A.M. (2011) "Speech-based Creation and Editing of Mathematical Content", PhD Thesis, Kingston University, U.K.

Zhao, J. (2008). Towards a User-centric Math Information Retrieval System. Bulletin of IEEE Technical Committee on Digital Libraries, Vol. 4, issue 2, fall 2008, ISSN 1937-7266