

The Use of Mathematical Formulae in an E-Learning Environment

Josep Cuartero-Olivera, Antoni Pérez-Navarro
and Teresa Sancho-Vinuesa
Universitat Oberta de Catalunya
Spain

1. Introduction

The use of mathematical formulae in engineering studies is as important as the subject content itself, especially in online distance education. In engineering studies, which have a strong technical component, both students and teachers must use formulae to express and solve their doubts, prove their knowledge or even quote any given piece of support material. In addition, in online distance education, communicating mathematics is not as easy as writing on a piece of paper or on a blackboard. As well as mastering the language of mathematics to express them properly, which is a problem that also exists in on-site environments, there is the problem of *writing* mathematics mainly through text e-mails, which is the main way to communicate within an on-line environment. The data gathered for this research at the *Universitat Oberta de Catalunya* (UOC) involves 4 terms, 15 engineering-related subjects and more than 17,000 e-mails. Among this large volume of e-mails, the use of mathematical notation is present in over 4,000 of them, representing an average of 23% of the total. As this preliminary result is quite significant, the aim of this chapter is to analyse the use of all the different strategies for communicating ideas with a mathematical content through the Internet and studying the impact for each one of them in order to find usage patterns.

Regarding virtual learning environments, as it is not possible to find previous studies about the use of mathematical notation within them, this work presents research of the different methods used by teachers and students to communicate mathematics through the Internet, and the use patterns regarding different subjects and knowledge areas. In order to do so, the core of this chapter consists of exploratory research as to which are the mentioned use patterns and tries to find relationships between them.

This chapter is structured as follows: as a first step, Section 2 explains in detail the problem addressed by this research. In Section 3, different methods for expressing mathematical notation in the particular case of a virtual learning environment like UOC are described. Next, Section 4 will introduce the scenario in which this research has been conducted. After that, Section 5 will focus on the study of how every one of these notation methods is used. Some statistical measurements are presented in order to try to find behaviour patterns through the different subjects and/or knowledge areas. The chapter ends with the conclusions and future lines of work.

2. Engineering studies in a virtual learning environment

Virtual learning environments are, in general, still challenging nowadays, as they have to deal with the barriers of time and distance. This is true not only for communication from teacher to students, but also among student workgroups or for students to communicate their queries to the teacher. The challenges become greater in engineering studies, where there are additional obstacles to overcome; for example the use of laboratories (having to become virtual laboratories) or granting access to high-profile computational tools. Focusing on the issue of knowledge transfer, one of the main problems regarding engineering studies is communication among university members, as a great part of this communication implies the use of a large amount of mathematical notation.

In the case of 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 reading of a given formula by the teacher, while they cannot easily learn how it is expressed visually, even in an on-site learning environment (Fitzpatrick, 2007). In reverse, a virtual learning environment allows students to see visual expression of formulae, but traditionally does not provide them with a verbal representation, which is also a handicap for visually impaired students. In addition, in some cases there is no auxiliary tool to help express mathematical notation, for example a formulae editor. When there is no such tool available, the methods for expressing mathematics are still computer aided, but they are as rudimentary as plain text or file attachments.

The issues of the use of mathematical notation regarding information systems has previously been stated and researched. For instance, there are differences in handling math expressions in one way or another for the indexing and retrieval of mathematics educational material in a search engine (Zhao, 2008). For that purpose, the author proposes the use of links between math expressions and text keywords. In this way, a semantic expression like “area of the function \cos ” can be linked to mathematics content about the resolution of integral functions for the particular \cos function. From a different point of view, other authors propose a five step process for the recognition and semantic understanding of mathematical formulae, basically consisting of (Chen & Okada, 2001):

- Pre-processing: In this first step, the mathematical expression is scanned to obtain an image. This image is processed in order to remove any noise and then split into mathematical symbols, digits and letters.
- Character recognition: The individual symbols, digits and letters are processed through a character recognition system and then classified into dyadic operators, monadic operators or atom characters.
- Rule base: Any ambiguity in the mathematical expression is eliminated by using a rules system. This system consists of mathematical rules, a sense-based dictionary for handling layout-dependent ambiguity and an experience-based dictionary for handling layout-independent uncertainty.
- Expression understanding: According to the layout of the mathematical expression, a layout tree is generated and parsed. This layout tree contains type and position of symbols, their sizes and centrelines and their parent-child relationships according to the expression layout. The result of this step is a semantic tree based on the mathematical rules used in step 3. This new semantic tree contains the types of symbols, their parent-

child relationships according to the expression semantics, the mathematical meaning of the combination and information about the expression constants and variables.

- Translating: In the last step, a recognized expression is produced and translated into a script or a source code suitable for being used in third-party software like TeX or Mathematica.

As we stated in Section 1, currently there is not much research into the particular context of a virtual learning environment, regarding the use of the mathematics language. The main communication method in this kind of environment is the use of e-mail, so students have to adapt their communications, including the mathematical notation within them, to this particular communication tool. Next, Section 3 shows how this is carried out by students and teachers and also explains the different available methods.

3. Writing mathematical notation within e-mails

The teaching and learning activity in a virtual campus, like in the case of UOC, is developed mainly within the virtual classroom. The virtual classroom is a space where teachers and students can communicate in a few different ways:

- Classroom forums
- Discussion boards
- Delivery board
- Private e-mail

Most of the interaction among the classroom members happens in the classroom forum or discussion board, where any of the members can send text messages and attach documents, while all the rest of members can read any of the messages and reply to them. Besides this communication space, students and teachers can also communicate by using their private e-mail address or the delivery board, where students send their due work, like continuous assessment tests. In the case of the private e-mail or the delivery board those messages are private to the sender and the recipient. However, the message format within all of these communication spaces is the same one: an e-mail written with the virtual campus e-mail editor.

As previously stated, there are several tools to improve mathematics communication in virtual learning environments, mostly formulae editors. Some of these formulae editors work as a standalone program, like the Wiris Editor (Wiris, 2011) or Microsoft Equation Editor (Microsoft, 2011), which can also be embedded into other software like Moodle. Some others are available directly on a website, like the LaTeX Equation Editor (The Number Empire, 2011) or sMArTh (sMArTH, 2011). Most of them use very common mathematical notation languages, like LaTeX (LaTeX, 2011) or the commonly known as MathML, Mathematical Markup Language (World Wide Web Consortium, 2011), which make them very feasible as plug-ins for other environments.

At UOC, such a mathematical notation tool has been available only during the last few terms, and it is still considered to be in its early stages. In fact, several tools are being tested and it is still to be decided which one is the best option. Before this tool was available, students and teachers needed to use other communication methods. These methods can range from the virtual campus e-mail editor itself -either in its plain text or Rich Text

Format versions- to more evolved tools for attaching documents like Microsoft Word or OpenOffice.org Writer –where these tools can be considered as the transfer method of the formulae itself, instead of the e-mail body.

Two main methods for expressing mathematics in web-based environments have been covered: through pictures and through coding with MathML (Yue-sheng & Jia-yi, 2008). These methods are focused on the way a formula is usually represented in a web page, from a technological point of view. Since this research though is focused on how students write formulae, it has to respond to a different classification, based on the technique used to write them. For instance, a symbol can be written with any rich text editor (for example the Σ symbol), or even with a plain text editor depending on the symbol (for example the + symbol). However, the construction of a given formula is frequently not possible using only text and needs other tools to help with its visual representation, for instance as shown in Figure 1:

$$\lim_{\Delta x \rightarrow 0} \sum_{i=1}^{\infty} f(x_i) \cdot \Delta x$$

Fig. 1. Sample formula

It is also possible to cite a formula by using any simple text editor, while in the case of a formula attached to an e-mail body an external tool is needed to generate the attachment itself. Therefore, a wider classification has been used, obtained from the observation of the behaviour of students and teachers: full mathematical formulae, mathematical symbols, formulae referencing and attachment. In order to better understand the results, all these different ways of communicating are explained and delimited through the next sections.

3.1 Full mathematical formulae

The first and most common method for expressing mathematical notation is through full mathematical formulae, understanding it as an equality (i.e. $a=b+3$), an inequality (i.e. $a+2>5$), or a mathematical expression consisting of a combination of more than one mathematical symbol (i.e. $\sin(\ln(1))$). Full mathematical formulae can also be expressed in any specific syntax delivered by programming languages or software commonly used in engineering environments. These variations are also considered in this group, for example specific mark-up codes like “\sqrt”, which are meaningful for the LaTeX2 editor, converting the expression $\sqrt{1-e^2}$ into $\sqrt{1-e^2}$.

3.2 Mathematical symbols

The mathematical symbol method consists of writing just one mathematical symbol at a time, whether it is in plain text (i.e. λ) or by using the symbol itself (i.e. λ), and exclusively when the symbol is not part of a whole mathematical formula. Numeric expressions have been considered into this group only if they are preceded (or followed) by a mathematical symbol (i.e. >10 or $10!$). Hyper-index, sub-index, or commonly used abbreviations of mathematical expressions like SQR, TAN, etc. also fall into this group.

3.3 Formulae referencing

Formulae referencing is used whenever a certain formula or expression is cited within the text, whether it is in its most common way (i.e. the formula on the first paragraph of page 24) or by using a previously established citation system (i.e. formula 17).

3.4 Attachment

The last method consists of attaching a file containing the formulae referenced in the e-mail body (i.e. the attached formula is wrong), or even writing the whole body of the message in an attached file. The attachment might be an image, some kind of text document (RTF, Microsoft Word, OpenOffice.org Writer, etc.), or even a scan of handwritten formulae.

According to Zhao classification of mathematical educational resources and some math information indexing and retrieval systems analysed through his research (Zhao, 2008), mathematical expressions can be considered as syntactically math-aware whenever the retrieval system reads the syntactical structure of the math expression to be searched for. On the other hand, if the system is capable of also capturing the semantics of the expression, then it is considered as semantically math-aware. Other systems, incapable of recovering neither the math-related syntactic or semantic meaning, are considered as math-unaware. Following this classification, full mathematical formulae and mathematical symbols could be considered as syntactically math-aware, and formulae referencing as semantically math-aware. However, as we have seen, there are not many different ways of expressing mathematical notation, although as explained in this section they are quite different from one another. Therefore, the next sections address the scenario for this research and the analysis of how each method is used by students and teachers, what patterns can be found among different subjects and knowledge areas and what the possible reasons are for a particular behaviour.

4. Scenario under study

As described in section 3, there are several ways of including mathematical expressions in a digital text subject to be sent by email. As most of the communication between teachers and students takes place in the classroom forum, for this research we have only considered the e-mails sent to that communication space. Therefore, in this section an exhaustive analysis of over 17,000 e-mail messages is made in order to classify them according to the type of mathematical expression method used. These e-mail messages have been gathered from 15 different subjects, all of them related to engineering degrees.

The interaction among students at the UOC is mainly developed through the virtual campus. Therefore, it is very common to find e-mails not directly related to the subject, for example introductory messages, technical problems or Christmas greetings. In order to avoid any kind of noise in the results, all these e-mails have been carefully discarded. This cleansing leaves still more than 15,000 e-mails. In a first search through this data, it has been detected that the use of mathematical notation is present in over 4,000 of the 15,000 e-mails, representing an average of 27% containing some kind of mathematical expression. These 4,000 e-mails are the ones we are going to take into consideration for the rest of this work.

Having prepared the e-mails which are going to be processed, this work focuses first on statistically analysing the types and frequency of mathematical notation used in them. With this information, a careful search through the data will make it possible to detect any particular usage pattern or specific student behaviour depending on variables such as the subject, the knowledge area, the type of studying material and other subject-related variables. Therefore, the next section deals with processing and analysing all this data in order to explain different behaviours in different classrooms.

5. Analysis of the results

Having agreed the motivation for this study, the characteristics of the research subject and the scenario in which it is developed, this section addresses the core of the research. We will first present a basic study of the way students and teachers communicate using mathematical expressions, consisting of the use frequency for every different expression type. Afterwards, the analysis will focus on determining some similarities and differences in the usage pattern for different subjects and/or knowledge areas. Finally, and in order to be able to better explain the reasons for those patterns, some of the most significant descriptive statistics will be developed.

5.1 A basic classification

The first question arising out of this study is to what degree mathematical expressions are used, regarding the different types of notation. We must bear in mind that students are not required to use any specific notation method, so they are free to use whichever method they think is most convenient for their communication needs. As previously noted, and unless it is stated differently, this research and its calculations will consider only the 4,000 e-mails containing some kind of mathematical notation. Therefore, Table 1 gathers the use percentage for every mathematical notation type, showing that around 66% of the e-mails include full mathematical formulae. The rest of the e-mails include, in order of use frequency, single mathematical symbols (24%), formulae citation (11%) and attachments with some mathematical notation (10%). It should be noted that e-mails may fall under more than one category if they use more than one of the different methods. Full results are shown in Table 1.

Total	Formulae		Symbol		Citation		Attachment	
# e-mails	# e-mails	Total %	# e-mails	Total %	# e-mails	Total %	# e-mails	Total %
4055	2679	66	967	24	461	11	390	10

Table 1. Classification of e-mails according to the mathematical notation type used

Furthermore, we want to analyse if this same frequencies apply to individual subjects. The aim is to find out if those frequencies exist regardless of the knowledge area of a particular subject, its content or its methodology. As a first step we will analyse the data in Table 1 but this time grouped by subject. Table 2 shows these results.

As we can see in Table 2, not all the individual subjects have the same average percentages regarding the use of one or other type of mathematics expression. The overall pattern is the same, but there are subjects where full mathematical formulae is used less in favour of citation, or the use of many more mathematical symbols. This fact can be due to differences

between the methodology and the study materials of every subject, or also due to different student profiles. In both cases, the results now require us to take into consideration each individual subject. Therefore, as a further step, it is necessary to find out what the patterns are for every subject, which are in turn classified under different knowledge areas. The next section takes care of this matter.

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
Computers 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 on 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

Table 2. Classification of e-mails according to the subject and the mathematical notation type used

5.2 Different subjects, different behaviour

After these first results, and as we have observed thanks to the statistics in Table 2, the next questions deal with the behaviour of students and teachers according to a particular subject. The main goal is to find out if the same average behaviour can be applied to all of the studied subjects or, if not, what are the possible reasons why there is a different behaviour, by finding subject-related variables affecting that overall pattern.

In order to calculate the next measurements, firstly we will consider the total number of e-mails for every subject, including the ones with no mathematical notation but directly related to that particular subject. We can see there is quite a significant difference between subjects regarding the type of mathematical notation used: depending on the subject, we find 17% of the total e-mails contain mathematical notation, while it can rise to 45% for other subjects.

If we consider only e-mails containing mathematical expressions for every subject, as we previously did in Table 2, we observe differences in the use of one or other type of

expression. Table 3 shows the full results of this data, but as a main result it is possible to observe that the percentages are quite different from one subject to another:

- Regarding full mathematical formulae, the results range from 45% in Automata Theory and Formal Languages to 78% in Probability and Statistics.
- In the case of single mathematical symbols, the percentage varies from 10% in Probability and Statistics to 49% in Automata Theory and Formal Languages.
- Regarding formulae citations, the percentage ranges from 2% in Automata Theory and Formal Languages, Wiris Laboratory (Algebra) or Computers Structure and Technology to 56% in Engineering Physics Fundamentals.
- For attachments containing mathematical notation, the percentages range from 4% in Cryptography, Engineering Physics Fundamentals, Computers Structure and Technology or Technological Fundamentals I to 29% in Wiris Laboratory (Algebra).

As it can be seen, there are significant differences regarding both the percentage of e-mails containing mathematical notation and the use of one or another expression method. Considering these results, the next question that arises is about the relationship between similar behaviours. As subjects can be classified within different knowledge areas, Table 3 also contains the same statistics, this time calculated for each one of those areas.

The most significant fact regarding the differences between knowledge areas is about Physics. In that knowledge area there is, compared with the other areas, quite a significant increase in overall mathematical notation use: while for Mathematics this percentage is around 28% and for Technology it is around 21%, for the Physics area it increases to 39%. Analysing this fact in detail and if we have a closer look at the different notation methods, the increase is mostly related to citations: 56% against 9% in the other areas. The reasons for this behaviour pattern are two-fold:

1. There is a well defined citation method in the subject falling under this knowledge area (Engineering Physics Fundamentals), which is responsible for this increase in the use of mathematical notation. This citation method consists of uniquely identifying with a number every single formula used within the subject. In every work document during the term, as well as within communications between students and teachers, formulae are referenced by using those unique numbers. Therefore, it can be easier and faster for both students and teachers referencing any of the formulae and thus the mathematical notation percentage increases. This same fact causes the rest of the notation types to be less used for the Physics area. Technological Fundamentals II (Circuit Theory) uses the same citation method and as we can see it is the second subject where the citation method is used more, with a percentage of 20%.
2. The subject itself: in Physics there are many formulae that students have to learn and understand. This explains the difference with Technological Fundamentals II, where the number of formulae is very much smaller.

Again, the usage figures for full mathematical formulae have an expected pattern, according to the results previously shown in Table 2. While for the Mathematics area it increases to 69%, very similar to the 62% for the Technology area, it drops to 47% for the Physics area. This behaviour is because of two main reasons: the first one is that the Technology area has a lower amount of formulae use within the subjects than the Mathematics and Physics areas.

Area	Subject	Total e-mails		Math notation		Formulae		Symbol		Citation		Attachment	
		#	#	%	#	%	#	%	#	%	#	%	
Mathematics	Algebra	1170	332	28	235	71	77	23	29	9	24	7	
	Automata Theory and Formal Languages I	610	128	21	57	45	63	49	2	2	9	7	
	Cryptography	1122	224	20	155	69	49	22	36	16	10	4	
	Discrete Mathematics	1020	244	24	158	65	71	29	29	12	12	5	
	Introduction on Mathematics for Engineering	726	194	27	147	76	22	11	10	5	23	12	
	Linear Systems	846	316	37	224	71	60	19	14	4	30	9	
	Mathematical Analysis	953	425	45	307	72	120	28	66	16	55	13	
	Mathematics I	1198	352	29	250	71	48	14	12	3	54	15	
	Probability and Statistics	610	218	36	170	78	21	10	13	6	24	11	
	Statistics	1413	331	23	228	69	83	25	45	14	21	6	
	Wiris Laboratory (Algebra)	1120	232	21	127	55	49	21	4	2	67	29	
	AREA TOTALS		10788	2996	28	2058	69	663	22	260	9	329	11
Physics	Engineering Physics Fundamentals	581	228	39	107	47	52	23	127	56	10	4	
	AREA TOTALS	581	228	39	107	47	52	23	127	56	10	4	
Technology	Computers Structure and Technology	1665	287	17	175	61	101	35	6	2	11	4	
	Technological Fundamentals I	1078	216	20	136	63	73	34	4	2	9	4	
	Technological Fundamentals II	1230	328	27	203	62	78	24	64	20	31	9	
	AREA TOTALS	3973	831	21	514	62	252	30	74	9	51	6	

Table 3. Usage of mathematical notation types by subject and aggregations by knowledge area

The second reason is that in the Physics area the use of a citation method is favoured as we previously explained.

Still looking at full mathematical formulae and focusing on the Mathematics area, this same irregular behaviour can be verified. The average use percentage of full mathematical formulae for this knowledge area is around 69%, which is quite high, but the behaviour is

not the same for all of the subjects in this area. While most of them fall into the range 65% to 75%, there are two subjects where the percentage drops dramatically to 45% and 55%. These subjects are, respectively, Automata Theory and Formal Languages I and Wiris Laboratory (Algebra). Similarly, as previously explained, these subjects do not contain as much mathematical formulae as the rest of the subjects and therefore the use percentage decreases. On the other hand, the use percentage of mathematical symbols in the subject Automata Theory and Formal Languages I is quite high, since this subject contains a high amount of single mathematical symbols instead of full mathematical formulae.

Besides these facts, as for the rest of the notation types there is no significant difference. Again, the conclusion we come to is that in some cases the use of one or other type of notation is highly dependant on the subject, depending on the content itself and on a previous agreement between teachers and students for using some specific notation method as we could see in the Physics subject. In this way, it seems that it is easier and more feasible for students and teachers to express mathematics by the use of a previously established citation system. But again, for the rest of the subjects, apparently the use of one or other method is more likely to be linked to the students' particular preferences.

In the next section we will develop more statistics in order to find yet more specific relationships between subjects and knowledge areas.

5.3 A global statistical analysis

The previous section has shown that there are significant differences in the use of notation between different subjects or knowledge areas. At this point it is important to develop some global descriptive statistics in order to better understand the links between different expression methods.

Table 4 shows the main statistical measurements, calculated for every notation type and knowledge area. For each of these groups, it shows the mean, the minimum and maximum, the standard deviation and the mode, all values represented in percentages. Regarding the special case of Physics, as there is just one subject under this knowledge area, we will not consider its standard deviation.

As it can be seen in Table 4, there are two cases in which the mean does not fall into the mode range:

- Mathematics area, formula within the e-mail body
- Mathematics area, attachment with formula

In none of these cases, though, the difference between the mean and the mode is very significant. This might only be a symptom of an abnormal distribution, and it is not surprising because as we described in previous sections there is a very different pattern in a few subjects for using one or other mathematical notation type depending on the subject and area.

The Mathematics knowledge area is the one showing a larger difference overall between the minimum and maximum percentages for every notation type. This was already explained in a previous section, the reason being there are two subjects in this area (Automata Theory and Formal Languages I and Wiris Laboratory) which do not follow the regular pattern of

the other subjects because of their content type. That is also confirmed by this area having the overall highest standard deviations, especially concerning the most used notation types: full mathematical formulae (with a standard deviation of 9) and mathematical symbol (with a standard deviation of 10).

	Knowledge area	Mean	Min	Max	Standard deviation	Mode
Formula within the e-mail body	Mathematics	67	45	78	9	>= 70% - < 80%
	Physics	47	47	57	-	>= 40% - < 50%
	Technology	62	61	63	1	>= 60% - < 70%
Mathematical symbol	Mathematics	23	10	49	10	>= 20% - < 30%
	Physics	23	23	23	-	>= 20% - < 30%
	Technology	31	24	35	5	>= 30% - < 40%
Citation	Mathematics	8	2	16	5	>= 0% - < 10%
	Physics	56	56	56	-	>= 50% - < 60%
	Technology	8	2	20	8	>= 0% - < 10%
Attachment with formula	Mathematics	11	4	29	7	>= 0% - < 10%
	Physics	4	4	4	-	>= 0% - < 10%
	Technology	6	4	9	3	>= 0% - < 10%
Attachment without formula	Mathematics	5	1	14	4	>= 0% - < 10%
	Physics	1	1	1	-	>= 0% - < 10%
	Technology	6	4	9	2	>= 0% - < 10%
Attachment with graphics	Mathematics	2	0	5	2	>= 0% - < 10%
	Physics	2	2	2	-	>= 0% - < 10%
	Technology	5	1	10	4	>= 0% - < 10%
Any kind of mathematical notation	Mathematics	28	20	45	8	>= 20% - < 30%
	Physics	39	39	39	-	>= 30% - < 40%
	Technology	21	17	27	4	>= 20% - < 30%

Table 4. Statistical analysis grouped by notation type and knowledge area

Table 5 shows the same statistical measurements groups as in Table 4, but this time regardless of the knowledge area. As it can be seen in the results, the percentages are more

dispersed, showing a high standard deviation on all three most commonly used notation types: formula within the e-mail body, mathematical symbol and citation.

As it can be observed, the only group mismatching the mean into the mode range with a significant percentage is Formula within the e-mail body. But analysing the data in Table 3, we can see that it is only due to a very irregular use of mathematical formulae: while the mode stays at the range 70%-80%, the rest of the subjects not falling into this range belong to a few different ranges. Therefore, we can discard the statistics in Table 5 as they are not explanatory for this study.

	Mean	Min	Max	Standard deviation	Mode
Formula within the e-mail body	65	45	78	10	$\geq 70\% - < 80\%$
Mathematical symbol	24	10	49	10	$\geq 20\% - < 30\%$
Citation	11	2	56	13	$\geq 0\% - < 10\%$
Attachment with formula	9	4	29	6	$\geq 0\% - < 10\%$
Attachment without formula	5	1	14	4	$\geq 0\% - < 10\%$
Attachment with graphics	3	0	10	2	$\geq 0\% - < 10\%$
Any kind of mathematical notation	28	17	45	8	$\geq 20\% - < 30\%$

Table 5. Statistical analysis grouped by notation type

Finally, Table 6 shows, according to the mode, the most popular notation types within each knowledge area. This rank also states that one or other notation type use highly depends on the subject and area, Physics being a good example of that: Mathematics and Technology areas both have formula within the e-mail body as the most commonly used notation type,

	Most commonly used	Second commonly used
Mathematics	Formula within the e-mail body	Mathematical symbol
Physics	Citation	Formula within the e-mail body
Technology	Formula within the e-mail body	Mathematical symbol

Table 6. Most commonly used citation methods by knowledge area

while Physics has citations as its preferred type. Furthermore, Physics does not have mathematical symbols in second place as Mathematics and Technology do, but formula within the e-mail body instead. This means that for Physics, when citation is not being used, the pattern reflects the one in Mathematics and Technology.

As we have seen, these main statistical measurements neither completely explain the overall behaviour of students choosing a particular mathematics expression method. More information is needed, basically in the way of a much larger e-mail sample, so it is possible to understand why a student expresses mathematics in a particular way. Therefore, this research leads us to conclude that a deeper study is needed in order to analyse different patterns linked to particular students.

6. Conclusion

In this chapter it has been shown: 1) which strategies and methods students use to communicate mathematical formulae in a web based e-learning environment, by means of the analysis of 17,000 messages; and 2) how important each method is depending on the subject and on the knowledge area.

This study has been developed exclusively using an e-mail web application that lacks a formulae editor, in order to explain the way students communicate using mathematical notation. In the course of this research, it has been seen that the use of mathematical formulae in virtual learning environments has to be carefully studied in order to provide students with better communication, as well as a better understanding of mathematics in engineering degrees.

From the study, it can be concluded that:

- Mathematical formulae appear in 30% of the e-mails for the analysed subjects. This shows that in the area of e-learning for technical and scientific degrees formulae play a key role in communication. When a technological solution is not available, which is the case, students manage to find a way to communicate mathematics. However, it has to be taken into account that this is an extra handicap for students in subjects that they traditionally find difficult. The challenge of communication, besides the inherent difficulty of the subjects, can cause some students not to ask questions.
- Mathematical expressions appear in different ways: as a symbol, as a formula written in pseudocode (LaTeX style), as a cited formula and as an attachment.
- For some subjects, the method used to communicate mathematical formulae depends on two factors:
 - The subject itself: some subjects have more formulae (like Physics) and others have more symbols (like Automata Theory and Formal Languages I). The complexity of the formulae and the role they play in the subjects will determine how much mathematical formulae will appear. Then, the overall amount of mathematical notation used by teachers and students seems to relate to the amount of mathematical notation content within the subject itself better than to some other external factors.
 - The features of the study materials: some subjects, like Physics or Technological Fundamentals II, have a very good citation method since every formula is numbered. This makes it easier for students and teachers to cite formulae by their

number and therefore causes a significant boost to the use of formulae thanks to the simplicity of the citation method. Assuming that students and teachers use (or should use) mathematical notation whenever they need to, and regarding the increase of mathematical notation use in Physics, it can be concluded that the lack of such an easy pre-established notation method causes difficulties in communication among the members of a virtual classroom community.

- There is no pattern regarding the use of mathematical formulae which is valid for all the subjects and knowledge areas. When a concrete type of notation is considered, the results show there is an overall common pattern among all the subjects, full mathematical formula, symbol and citation being the most commonly used. The exception though occurs when a certain notation method is established beforehand, in which case it seems easier and more likely to be used by students and teachers according to the increase of use observed in the particular case of the Physics subject.
- Therefore, there are signs leading to the existence of student patterns and profiles, more than an overall pattern for every subject. In some cases, when a subject offers an easy and feasible method for expressing mathematics, such is the case for citation, students and teachers tend to adopt it and in that way increase the use of mathematics content within e-mail. In the rest of cases, the student preference seems to be the main reason for the selection. In that case, we need to analyse what leads a student to choose one or other method and if that choice can be linked to a better understanding of the subject, thus a better academic performance. Or furthermore, from a different point of view, if students that have a better academic performance are linked to one particular type of mathematical expressions.

All these conclusions show the importance of mathematical notation for students of technological subjects. For some subjects, this study has detected several key points as indicators for the use of a specific mathematics expression method. For example, in certain subjects, a well-established citation system makes it easier and faster for students to use citation instead of any other method. In the same way, other features like the structure of the study materials or even its content, can also affect the behaviour of students. As for other subjects, further work has to be developed in order to find proper key indicators, which apparently can be related to particular student profile or preferences.

In spite of the large volume of e-mails processed in this research, more than 17,000, the information gathered from them is not enough to identify these student profiles. Currently, the information related to one particular student through different subjects and terms is not significant enough, statistically speaking, to be able to determine if they are following a particular pattern. Therefore, future research must bear this in mind and target particular students behaviour instead of overall subject behaviour. Once this information is available, future studies can also try to find links between the academic performance of students and mathematical expressions use patterns. For example, it is possible to find out if a specific behaviour pattern, varying from the classroom average, leads to a different academic performance, either if that performance is reflected in the students' final marks or on a higher rate of students following continuous assessment during the term. Furthermore, not only the use of the communication method chosen by the student, but the variation in the use of different methods, the usage amount of each

one of them and even the content of the e-mails itself can lead us to detect different student profiles from which we could have another very interesting point of view. For example, the use of a richer language or the development and discussion of a given formula through a thread of e-mails can help teachers identify the expected performance for a particular student and therefore help them focus on the students who are not following this pattern.

According to the results of this research, the contents and structure of a subject can lead students to communicate mathematics in a particular way, sometimes more frequently than the average. However, this does not necessarily mean a better overall performance in a subject, as students would perhaps perform better if the subject was, conversely, designed according to the preferences of the students, providing them with the necessary tools for this purpose.

Finally, the use of mathematical language within a virtual classroom is a handicap for e-learning since students and teachers are only able to express themselves by the use of e-mail but, furthermore, we must take into account that this problem can be much worse for disadvantaged student groups – as for example students with visual impairments – especially when we consider the similar difficulties that both students in a virtual environment and students with visual impairments face on a daily basis (as was explained in Section 2). Therefore, these are the main aspects that will be explored in future work.

7. References

- 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.
- 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>
- Microsoft (2011). Microsoft Equation Editor. Date of access: 01/05/2011. Available from: <http://www.microsoft.com/education/en-us/teachers/how-to/Pages/mathematical-equations.aspx>
- sMArTH (2011). An online equation editor for MathML and LaTeX. Date of access: 01/05/2011. Available from: <http://smarth.sourceforge.net>
- The Number Empire (2011). LaTeX Equation Editor. Date of access: 01/05/2011. Available from: <http://www.numberempire.com/texequationeditor/equationeditor.php>
- Wiris (2011). The global solution for maths education. Date of access: 01/05/2011. Available from: <http://www.wiris.com/en/editor>
- World Wide Web Consortium (2011). MathML, W3C Math Home. Date of access: 01/05/2011. Available from: <http://www.w3.org/Math>
- Yue-sheng, G. & Jia-yi, Z (2008). Uploading Strategy of the Formula in the Web-based Mathematics Testing System. International Conference on Computer Science and Software Engineering.

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