The importance of using and not using symbols in school mathematics

Gerrit Stols
University of Pretoria
gerrit.stols@up.ac.za

Abstract
Our high school mathematics bristles with symbols although school algebra historically developed without the use of modern symbols. Students often experience problems because of the use of these symbols. Mathematical symbolic notation, as it is used today, is a relatively new phenomenon in algebra. Although history indicates that algebra showed enormous growth once symbolic notation came into use, it also indicates that not all the symbols play an equally important part in the development of mathematics. The true strength of algebra lies in the use of symbols for the unknowns and variables. It is therefore important that teaching focus on the use of these symbols and not on the use of all the other symbols.

Keywords: symbols, mathematics history, notation, algebra

1. Introduction
Mathematical symbolic notation as we know it today is a concise and accurate form of notation which is internationally accepted. Because symbolic notation is so concise and well defined, mathematicians and mathematics teachers tend to use symbolic notation. Symbols are so important in mathematics, and especially in algebra and in the history of algebra, that Kieran (1992, p. 391) defines algebra as “the branch of mathematics that deals with symbolizing general numerical relationships and mathematical structures and with operating on those structures”. School mathematics without symbols would be unthinkable today. High school mathematics uses a plethora of plus signs, minus signs, division signs, letters, brackets, exponents, root signs, equal signs, logarithmic notation and many other symbols. Students often experience problems with the use of symbols, however (Tall et al., 2001). As a result some students regard algebra as a mysterious field of endeavour (Cunningham 1986). Whitehead (1911) describes it as follows:

Mathematics is often considered a difficult and mysterious science, because of the numerous symbols which it employs. Of course, nothing is more incomprehensible than a symbolism which we do not understand. Also a symbolism, which we only partially understand and are unaccustomed to use, is difficult to follow. p. 59

Higginson expresses the following concern: “Much of the power of mathematics stems from the potency of its symbols. There is, however, a price to be paid for this potency. The symbols which serve as highly effective tools for some are most formidable barriers for others”. Back in 1842 De Morgan (Cajori, 1928b) uttered a warning against the excessive use of symbols: “Too much abbreviation may create confusion and doubt as to the meaning...” (p.327). But on the other hand Blackhouse et al. (1992) contend that the strength of
The question is how many symbols, if any, should be used in the teaching of mathematics at school level. Are symbols necessary, or are they merely convenient? The purpose of this study is to determine from a historical perspective which symbolic notation is essential for use in school mathematics.

2. The historical development of symbols

The development of algebraic symbols went through three phases: a rhetorical phase, a syncopating phase and a symbolic phase (Groza, 1968; Kieran, 1992). Nesselmann identified these three phases of the development of symbolic algebra as long ago as 1842 (Eves, 1980; Smith, 1953). These phases extend from 1700 BC to about 1700 AD and are characterised by the systematic development of symbols and the solution of equations by various methods (Baumgart 1989). Smith (1953) points out that these phases overlapped historically and can therefore not be separated exactly, although we can distinguish between them.

2.1 The development of the “variable” in algebra

The rhetorical phase is characterised by the absence of algebraic symbols. The solutions to problems were written out without the use of abbreviations or symbols and as a result the solutions were very long and were difficult to follow cognitively. The following problem from Baumgart (1989) is an example of a rhetorical problem derived from Babylonian algebra: “Length, breadth. I multiplied length by breadth to get an area of 252. I added the length and the breadth: 32. Required: length and breadth” (p. 235). This is typical of the problems found on cuneiform clay tablets dating from the reign of King Hammurabi, 1700 BC. These algebraic problems were expressed and solved in words (Kline, 1972). Words such as us (length), sag (breadth), and aša (area) were often used as the unknowns because most algebraic problems had their origin in geometrical contexts (Kline, 1972, p. 9). Although the Babylonian mathematicians were able to solve a wide variety of equations, the lack of symbols prevented them from formulating general solutions (McQualter, 1983). According to the Rhind and Moscow papyri (which date respectively from about 1850 BC and 1650 BC), the solving methods of the Egyptians are not as sophisticated as those of the Babylonians (Baumgart, 1989). They solved the problems by estimation. This method is known as the “Rule of False Position” and was still used as late as the sixteenth century (Hooper, 1948, p. 78). The limited Egyptian algebra contained almost no symbols, except the feet of a man walking forwards and backwards, used as symbols for addition and subtraction respectively (Kline, 1972). The fact that they use direction in doing algebra shows the strong link between geometry and algebra at that time.

Syncopating algebra was characterised by the use of stenographic abbreviations for quantities, ratios and calculations that occurred repeatedly. Whereas the Egyptians and Babylonians wrote equations and solutions in words, Diophantus (250 AD) introduced
symbolic abbreviations for the different terms of an equation. He broke with the traditional Greek use of symbols by working with powers greater than three (Katz, 1993). This means that the equations can no longer be represented geometrically (in three dimensions) by figures. He gained fame by publishing his book *Arithmetica*, in which he covered indefinite equations, usually two or more equations with different variables for which there are an endless number of rational solutions (Baumgart, 1989). In the *Arithmetica* Diophantus used abbreviations to represent the unknown, powers of the unknown, subtraction, equality and reciprocals (Eves, 1980). Diophantus was the first person to use letters for unknown quantities. By so doing he made the transition from arithmetic to algebra (Schoenfeld & Arcavi, 1988). The unknown number in algebra was defined by Diophantus as an undefined number of units and was represented by the Greek letter \( \varsigma \) (Kline, 1972, p. 139). Katz (1993, p. 163) showed that Diophantus’s symbol for the unknown is derived from the Greek letters, \( \alpha \) and \( \rho \), of the word “arithmos” which means number. \( \rho \) was used for \( \mu \nu \alpha \varsigma \) (monas or unit). The manuscript therefore contains expressions like \( \Delta^{\nu} \varsigma \alpha \beta \mathrm{M} \theta \), which stands for 3 squares, 12 numbers, and 9 units, or, as we would write this, \( 3x^2 + 12x + 9 \) (Katz, 1993, p. 163). Diophantus also used the above symbols together with the sign \( X \) to represent reciprocals, for example \( \Delta^{\nu} x \) was used to represent \( 1/x^2 \) (Katz, 1993, p. 163). However, there was no connection between the symbol \( \varsigma \), which stands for the unknown, and its square \( \Delta^{\nu} \varsigma \) as there is between \( x \) and \( x^2 \) today. This was the biggest shortcoming of Diophantus’s notation. Diophantus was able to solve quadratic equations and in the introduction to the *Arithmetica* he described a method by means of which equations could be simplified (Katz, 1993). Nevertheless the method of “False position”, a technique that was already in use among the Egyptians, was given a central place in Diophantus’s Book IV. Diophantus’s biggest contribution to the development of algebra was the use of a letter to represent unknown quantities. Cajori (1928a) observes, however, that Diophantus really made very little use of symbols. He often described calculations in words, using only one symbol for the unknown. According to Kline (1972), Diophantus wrote his solutions like a piece of continuous prose.

For algebra the time was ripe to enter the symbolic phase. The first authoritative algebraic work to be printed, according to Smith (1953), is the *Ars Magna* by Cardan in 1545. The primary focus of this book is the solving of algebraic equations; it dealt with the solving of cubic and biquadratic equations. Even complex numbers were used in this work, which was the first step in the direction of modern algebra. The following major work that appeared was the *General Trattato* of Tartaglia in 1556. In 1591 the French jurist, François Viète, made such improvements to symbolic algebra, according to Groza (1968, p. 251), that he has often been called the “father of algebra”. Groza (1968) explains that he used capital letter vowels to indicate unknown quantities and Eves (1980) add that he used consonants to indicate known quantities. Although some of Viète’s predecessors also used capitals to represent general quantities, Viète was the first to do so systematically. In the introduction to his *In Artem Analyticem Isagoge* (Introduction to the Art of Analysis) of 1591, Viète made the following observation: “Numerical logistic is that which employs numbers; symbolic logistic that which uses symbols, as, say, the letters of the alphabet”. Viète (as quoted by Katz, 1993, p. 340) manipulated both letters and numbers:
Given terms are distinguished from unknown by constant, general, and easily recognized symbols, as by designating unknown magnitudes by the letter A and the other vowels E, I, O, U, and Y and the given terms by the letters B, G, D and other consonants.

In an example from Viète’s *Opera Mathematica* (1646), we can clearly see how he uses unknowns: “I QC - 15 QQ + 85 C - 225 Q + 274 N aequatur 120”. In modern notation this would be $x^6 - 15x^4 + 85x^3 - 225x^2 + 274x = 120$. This notation enabled Viète to solve quadratic equations with the aid of analytical instead of geometric methods. Viète made only partial use of modern notation, but took the important step of allowing letters to represent numerical constants (variables), which enabled him to break away from the generally verbal style of his predecessors. This made it possible to solve problems with the aid of fixed rules. Viète was able to concentrate his attention on solving procedures rather than on the solution itself, was able to give general rather than specific examples and was also able to provide formulas rather than rules. Gellert et al (*The VNR Concise Encyclopedia of Mathematics, 1975*) formulates this as follows: “Variables are useful in two ways: they make it easy to state laws, and the solution of a problem expressed in terms of variables yields the result for arbitrarily many individual cases without new calculations, by mere substitution”. There was no single discovery which changed algebra as much as Viète’s use of letters for both known and unknown quantities (Phillip 1992). Harper (1987, p. 79) expresses this as follows: “This deceptively simple, yet brilliant conceptual advance was to change the face of algebra”. The use of letters to represent a “given” quantity established a new numerical concept in mathematics – the algebraic number concept. Numbers that are represented by these letters have no quantity or specific order. As Harper (1987, p. 78) puts it: “Each letter at once represents each, every and all numerals within a given set, and order has to be defined when necessary”. These generalisations were used to make general statements (e.g. $a + b = b + a$), to investigate generalities in mathematics and to handle a finite or infinite number of cases simultaneously. René Descartes was one of the first people to use Viète’s letters. He used the last letters of the alphabet, such as x, y and z, to represent unknown quantities and the first letters of the alphabet to represent known ones (Eves 1980). In his book *La Géométrie* (1637) we come across the expression $x^3 - \sqrt{3}xx + \frac{26}{27}x - \frac{8}{27\sqrt{3}} = 0$, which closely resembles our modern notation (Resnikoff & Wells, 1973). Descartes used $\infty$ to represent $\approx$ and preferred $xx$ to $x^2$. He even frequently used $xxx$ instead of $x^3$ (Groza, 1968). Descartes’s notation was generally accepted and used, with the exception of the equals sign. Around 1800 $x^2$ and $x^3$ were also generally used (Groza 1968). During the symbolic stage, algebraic notation was considerably modified and refined. It was only in the time of Leibniz (1646-1716) that algebraic notation was largely standardised, but even today certain exceptions exist.

2.2 The development of symbolic notation for addition and subtraction
The first symbols for these operations that were discovered were the Egyptian symbols, the hieroglyphs for addition and subtraction that were found in the Ahmes Papyrus discovered in 1550 BC (Smith 1953). Although Diophantus (275 BC) used the symbol $\uparrow$ for subtraction, he did not use any symbol for addition or multiplication (Katz, 1993). The symbol $\uparrow$ comes from a combination of the letters $\Lambda$ and $I$ in the Greek word *leipis* ($\Lambda E I \Pi \Sigma$), which means
“lacking” (Eves, 1980, p. 129). Diophantus represented addition by means of contiguity, as in $K \sum_{x=1}^{10} m_{27} - 25$ (in modern notation $\frac{2x^2 + 100 - 20x}{10x - x^2} = 25$) (Resnikoff & Wells, 1973, p. 205). Addition was represented by $et$ and subtraction by $m$. The first printed + and − signs were used in 1489 by Johann Widman (Eves, 1980). These signs were not used as operation symbols. The plus sign is an abbreviation for the Latin word $et$ (Eves, 1980; Smith, 1953). According to Eves (1980) and Smith (1953), the + and − signs were used to indicate algebraic operations for the first time in 1514 by the Dutch mathematician, Van der Hoecke. For example, he wrote $4x^2 - 51x - 30 = 45 \frac{3}{5}$ as “4 Se. - 51 Pri. - 30 N. Dit is ghelijc $45 \frac{3}{5}$”. The mathematician who is credited with the general use of these symbols is Stifel, who wrote $3x + 2$ as “3 sum: +2” (Smith, 1953, p. 400). According to Resnikoff and Wells (1973, p. 205), he also, for example, wrote $\frac{9x + 8y}{6t} \text{ perit} \frac{3p}{2} \text{ facit} \frac{27y + 24t}{12t}$ instead of $\frac{9x^2 + 8x^2}{6x^2} \frac{3x}{2} \text{ as } 27x^5 + 8x^2 \text{ instead of } \frac{12x^8}{12x^8}$, as it would be in our modern notation. Since that time the symbols for addition and subtraction were widely used by German and Dutch writers but the exact form of the symbols was only standardised in the eighteenth century. In 1752, in the edition by Bartjens, for example, $x^2 = -2375x + 1785000$ was written as $xx = \div -2375 x \times 1785000$ (Smith, 1953, p. 400).

2.3 The development and use of the “equals sign”

As early as 275 BC Diophantus used the abbreviation $\sum$ for $\sum_{x=1}^{10} m_{27}$ (equals) in his Arithmetica to indicate equality, although the classical writers preferred to write out words like æqualis, æquantur, esgale, faciunt, ghelijk, orgleich (Smith, 1953, p. 410). Cajori (1928a) points out cases where æquales was abbreviated to aeq. The symbol for equality, $\equiv$, was used in 1630 by Descartes. According to Cantor (1913) the symbol is derived from the ligature of the two letters $ae$ in the word æquales (Cajori, 1928a). In 1557 Robert Recorde (1510) in his book, Whetstone of Witte, used the modern equal sign ($=$) for the first time (Eves, 1980; Smith, 1953). He gave the following reason for using two parallel line segments for the equal sign: “bicausenoe 2 thynges can be more equalle” (Smith, 1953, p. 412). But a hundred years after Recorde, the most prestigious mathematicians of the time were still using no symbol to denote equality (Cajori, 1928a). Not before the late eighteenth century was the sign accepted by all mathematicians.

2.4 The development and use of the “root sign”

In the early Renaissance period the word radice was used in Europe by Regiomontanus to represent the square root. He wrote the algebraic expression $5 - \sqrt{21 \frac{8}{27}} = x$ as “5 m Radice de $21 \frac{8}{27}$ ecce valorrei” (Resnikoff& Wells, 1973, p. 205). The symbol most generally used by the Latin writers to represent a root is $R$, an abbreviation of radix (Smith, 1953, p. 407). This sign, with numerous variations, was used in printed books for almost a century. Pacioli wrote $\sqrt{7 + \sqrt{14}}$ as RV7 p R 14, where RV stands for the radix universalis (which indicates that the root of everything that follows should be taken), while Bombelli wrote it
as $R_{7 \ p \ R14}$. Smith (1953, p. 408) found that the same sign was used for response and for res, for the unknown quantity. The root sign ($\sqrt{}$) was used for the first time in 1525 by ChristoffRudolff in his book Die Coss (Eves, 1980; Smith, 1953). According to Eves (1980), the origin of this sign was the small r, which stands for radix, but Smith (1953) does not believe there is any evidence corroborating this. In fourteenth century manuscripts forms such as $\gamma$, $r$, and $\sqrt{}$ were used for the letter $r$. Therefore in all probability the root sign is derived from the small letter $r$, which stands for radix. The use of symbolic notation to represent any root (cube root, biquadratic root, etc) was not firmly established for a long time. Vlacq used $\sqrt{}$ for a square root, $\sqrt[3]{}$ for a cube root and $\sqrt[n]{}$ for a biquadratic root (Smith, 1953, p. 408). Subsequently Rahn, Ramus-Schoner, Gosselen, Stevin and Biodini still used different variations for the representation of any root. In the seventeenth century this notation largely became standardised, but it was not until 1928 that the alternative exponent form was officially proposed in a report by the National Committee on Mathematical Requirements:

With respect to the root sign, $\sqrt{}$, the committee recognizes that convenience of writing assures its continued use in many cases instead of the fractional exponent. It is recommended, however, that in algebraic work involving complicated cases the fractional exponent be preferred. Attention is called to the fact that the symbol $\sqrt[n]{a}$ means only the positive square root and that the symbol $\sqrt[1/n]{a}$ means only the principal $n$th root, and similarly for $\alpha^{1/2}, \alpha^{1/n}$ (as quoted by Cajori, p. 379).

3. Discussion

The reasons for the development of algebraic notation as well as the necessity for the use of symbols in algebra are evident if we study the history. The history of the development of algebraic symbols indicates that algebra showed enormous growth once symbolic notation came into use. Scott (1960, p. 41) remarked on this as follows: “Euclid, Archimedes, Apollonius, had brought mathematics to a state beyond which further advance was hardly possible until new methods, and a very much improved system of notation, had been devised”. All forms of symbolic notation did not play an equally important part in the development of mathematics, however. The true strength of algebra lies in the use of symbols for the unknowns and variables. Diophantus’s most important contribution to the development of algebra was the use of a letter to represent unknown quantities. Fixed rules could then be applied to find the unknown quantity through manipulation. But Viète took the important step, namely allowing letters to represent both known and unknown quantities. This enabled Viète to give general examples rather than specific ones and work out formulas instead of rules. This major discovery of Viète’s was to transform algebra. As Schoenfeld and Arcavi (1988, p. 423) put it: “...variables are a formal tool in the service of generalization”. As soon as a problem is expressed in a symbolic form it is a simple task to solve it. Hooper (1948: 65) found that this modern method of solving takes about one-tenth of the time that it would take simply to read the solution given in the Ahmes papyrus. Whereas Diophantus’s solutions were not general, Viète’s solutions could be applied to any problem. This means that a large number of problems can be represented by a single equation. Once this general equation has been solved, all problems represented by the general equation have been solved as well. An example of this is the equation $ax^2 +$
\[ bx + c = 0, \text{ for which the general solution is } x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \] The advantage of algebraic problems is therefore that they can usually be solved by means of a standard algorithm, which is why Resnikoff and Wells (1973, p. 202) commented as follows: “In this sense the urge to algebraize is one of economy of effort”.

In the history of mathematics it is clear that the benefits of using symbols exceed the disadvantages and that the use of particular symbols is necessary. There are symbols without which progress would be impossible and which have played a more prominent role in the development of algebra than have other symbols. These symbols include the letters used in the place of unknown, and later in the place of known quantities. It is clear from the history of the development of algebraic symbols that, apart from the use of unknowns and variables, there was no urgency about the introduction of symbols. The development of elementary algebra is not dependent on the use of symbolic notations such as =, −, +, \sqrt{ } and exponents. If we examine almost any algebra textbook from the 17th and 18th centuries, we can see that virtually no symbols were used. Hodder, for example, used virtually no symbols before page 201; he remarks as follows: “Note that a + thus, doth signify Addition, and two lines = Equality, or Equation, but a \times thus Multiplication” (Smith, 1953, p. 395). No other symbols were used. It is surprising that linear algebra had developed almost fully before the symbolic stage was reached. Maestro Dardi of Pisa was able to solve 198 different kinds of equations in the fourteenth century without the aid of algebraic notation. Some of these equations were complex and contained powers of twelve. In Western Europe algebra remained rhetorical up to the fifteenth century, with syncopation making its appearance only in a few isolated places. Symbolic algebra did not appear in Western Europe until the sixteenth century, but developed so slowly that it was not until the seventeenth century that it was generally used. Historically, therefore, the symbolisation of algebra took place very slowly, and our modern algebraic notation is very young in relation to the whole history of the discipline. It was not until the twentieth century that firm directions were given about the use of root notation and alternative exponential notation. Mathematical symbols developed over a long period and this was a natural process that took centuries to standardise. It is notable that elementary algebra had virtually been perfected before a proper standardised symbolic notation had been developed.

4. Conclusion

Although the occasional use of symbols makes matters easier, many of the symbols used in school mathematics are unnecessary. Alexandre Savérien expressed himself strongly (back in the eighteenth century) against the excessive and unnecessary use of symbols:

... is \sqrt{aa + 2ab + bb} not better than w which some wish to substitute for it?
... To invent new characters which signify nothing else than those which have been already accepted, is a wanton embroilment of things ... I believe that the less one uses characters, the more one learns of mathematics (as quoted by Cajori 1928b, p. 330).
The advantage of mathematical notations, however, is their precision and accuracy. Natural language can often be differently interpreted, but mathematical symbols can be used with great accuracy. The accuracy and precision of algebraic symbols can best be illustrated by means of practical examples from the work of a mathematician. In Enderton’s book *Elements of set theory* the extension principle is formulated in words: “If two sets contain exactly the same elements, they are equal”. But then Enderton (1977, p. 2) makes the following observation: “... we can state things more concisely and less ambiguously by utilizing a modest amount of symbolic notation”. The same extension principle then becomes: ‘If A and B are sets, so that for each object t, t ∈ A if and only if t ∈ B, then A = B’. Once Enderton introduced symbols of mathematical logic as well, the principle was written as ‘∀A∀B[(A and B have the same elements) ⇒ A = B]’ and when more symbols were added the extension principle was written as ‘∀A∀B[∀x(x ∈ A ⇔ x ∈ B) ⇒ A = B]’. According to Enderton (1977, p. 13), the reason for this is: “The symbolic language, when used in judicious amounts to replace the English language, has the advantages of both conciseness (so that the expressions are shorter) and preciseness (so that expressions are less ambiguous)”. The question is: what degree of precision is really necessary for mathematics at secondary level? Is it really necessary to symbolise the extension principle fully? It is not really necessary for us to state the axioms in symbolic form. In advanced mathematics symbolic notation has more advantages. To name an example: In the process of abstracting the admission requirements of a set, certain paradoxes arise in the set theory. In 1906, for example, Berry pointed out that the set

{\{x / x is all positive whole numbers that could occur in one typed line\}} gives rise to a paradox. This paradox can be avoided by requiring that the admission requirements of a set be written in a completely unambiguous form. Enderton (1977, p. 22) expressed the following view: “We are saved from this disaster by our logical symbols. By insisting that the formula be expressible in the formal language...” and then observes: “Those symbols are your friends!”

5. Recommendation

It is clear from the historical development of symbols that the development and use of unknowns and variables are essential for school students’ own mathematical development. This development has had a long history. Students should therefore be given sufficient time to progress through the same process. Instructional activities should be specifically developed to speed up the students’ own use of variables and unknowns.

The use of algebraic symbols has its advantages, but it also has some decided drawbacks. Mathematical notations and symbols are a powerful aid in the service of mathematics. The introduction of symbols, apart from the use of the variable and unknowns, can be left to students themselves under the guidance of the teacher. It is, at the end of the high school, important that the students use the standardised symbolic notation, but according to Rubenstein and Thomson (2001) meaning must precede symbolisation. Symbols were originally tools used by mathematicians, just like abbreviations for words, to make life easier for them. At school level symbols appear to be regarded as masters to be served, rather than as tools to be used by students. Consider, for example, the confusion among students.
as a result of the early use of exponents. This is the reason why some students believe that \(x^2 + x^3 = x^5\).

The findings can best be summed up by quoting the words of Cajori (1928a, p. 431):

The experience of the past certainly points to conservatism in the use of symbols in elementary instruction . . . . The conclusion reached here may be stated in terms of two schoolboy definitions for salt. One definition is, “Salt is what, if you spill a cupful into the soup, spoils the soup.” The other definition is, “Salt is what spoils your soup when you don’t have any in it.”

References


