On the learning of algebra

H. Wu

Department of Mathematics #3840
University of California, Berkeley
Berkeley, CA 94720-3840
USA
http://www.math.berkeley.edu/∼wu/

wu@math.berkeley.edu

One of the main goals of California's Mathematics Content Standards is that algebra be taught in the eighth grade so that all high school graduates know algebra. Achieving this goal of "algebra for all students" would be easier if we have a quantitative understanding of how the learning of algebra takes place, but we don't. Nevertheless, there is at least a general agreement that the learning of algebra would be extremely difficult if it is not accompanied by computational fluency in the arithmetic operations of integers and fractions *together with a good understanding thereof*. Let us try to understand why this is so.

School algebra is sometimes referred to as "generalized arithmetic". If one looks at the algebra taught in school, superficially there seems to be not much difference between arithmetic and algebra because, in both cases, the operations involved are the usual standard ones: $+$, $-$, $\times$, $\div$, with a few $\sqrt{\ }$ and $\sqrt[3]{\ }$ thrown in here and there. But the key difference, not visible to the naked eye, lies in the *generality* of the reasoning in algebra. For example,

[0] May 25, 2000.
This article was written in January of 2000 at the request of Janet Nicholas, then a member of the California State Board of Education. It was meant for the consumption of parents and school administrators.
Revised January 17, 2001. A suggestion of Jeff Rabin was the reason for the revision as well as the improvement. I wish to thank him warmly.

a sixth grader can easily compute the square of a sum of two numbers: $(5+1.6)^2 = 6.6^2 = 43.56$, $(1.3+0.1)^2 = 1.4^2 = 1.96$, etc. This is no more than simple arithmetic: add two numbers and then multiply the result by itself. Sometimes, however, the straightforward method of computation is inadequate, and we want to know if there a different way to express the square of a sum, for instance, as a sum of products in terms of the given numbers. There are many reasons for asking this question, one of which may be explained as follows. Using $(1.3+0.1)^2$ for the purpose of discussion, suppose we regard 1.3 as fixed and, instead of just asking for its square, also want to know how much bigger or smaller the square of a number near 1.3, such as $1.3+0.1$, would be. This calls for a computation of the difference $(1.3+0.1)^2 - 1.3^2$. Of course, the real question is not about one number $1.3+0.1$ near 1.3, but *any* number near 1.3: e.g., $1.3+(-0.1)$, $1.3+0.01$, $1.3+(-0.004)$, etc. So we may as well look at the difference $(1.3+h)^2 - 1.3^2$ for any small number h.

Now $(1.3+h)^2 - 1.3^2$ is a mysterious expression! It is impossible to get an intuitive feeling of how the magnitude of h affects the difference $(1.3+h)^2 - 1.3^2$ just by looking at it. If we are given $5h^2 + 13$, for example, we would feel more comfortable because, ignoring the term 13, we have a general idea of how h^2 behaves: if $h = 0.1$, then $5h^2 = 5 \times 0.01 = 0.05$, and if $h = 0.004$, then $5h^2 = 5 \times 0.000016 = 0.00008$. Or if we are given $16h+3$, we would feel even better because, except for the constant term 3, this expression is proportional to h. What we are driving at is that, with expressions such as $5h^2+13$ or $16h+3$, h is displayed *explicitly*, whereas in $(1.3+h)^2 - 1.3^2$ the role played by h is opaque: h is "hidden" inside $(1.3+h)^2$. Is there perhaps some way to "liberate" h from inside the square *regardless of what h may be*? When we start to think this way, we have already left arithmetic behind and are engaged in what is known as *algebraic thinking*. But first let us see how to liberate h.

There is clearly no point in looking at a specific number such as 1.3, so we do it in general. Let x and y be two *arbitrary* numbers, and we ask for the value of $(x+y)^2$. By definition, $(x+y)^2 = (x+y)(x+y)$, so if we use A to stand for the first factor $(x+y)$, then the *distributive law* —- which says that no matter what real numbers a, b and c may be, it is always true that $a(b+c) = ab+ac$ and $(b+c)a = ba+ca$ —- yields:

$$(x+y)^2 = A(x+y) = Ax + Ay.$$

But $Ax = (x+y)x$ and to the right side, we can again apply the distributive law to get $Ax = x^2 + yx$. Similarly, $Ay = xy + y^2$, so that[1]

$$(x + y)^2 = x^2 + yx + xy + y^2.$$

Remembering that the multiplication of numbers is commutative (in the sense that $yx = xy$ no matter what the numbers x and y may be), we finally get:

$$(x + y)^2 = x^2 + 2xy + y^2 \tag{1}$$

Putting everything together for easy reference, we can summarize the derivation of identity (1) in the following sequence of steps:

$$\left.\begin{array}{rcl} (x+y)^2 & = & (x+y)(x+y) \\ & = & (x+y)x + (x+y)y \\ & = & (x^2+yx) + (xy+y^2) \\ & = & x^2 + 2xy + y^2. \end{array}\right\} \tag{2}$$

We have thus shown that $(x + y)^2$ can be evaluated as the sum of three products xx, $2xy$ and yy, and this is valid *regardless of what x and y may be.* This feature of *generality*, together with the extensive use of symbolic notation to express this generality, are what distinguishes algebra from ordinary arithmetic.

To go back to our example, we can use identity (1) —- with 1.3 replacing x and h replacing y — to obtain $(1.3 + h)^2 = 1.3^2 + 2.6\,h + h^2$. It follows that

$$(1.3 + h)^2 - 1.3^2 = 2.6\,h + h^2\,, \tag{3}$$

which then completely solves our original problem of making the role of h explicit in the expression $(1.3 + 0.1)^2 - 1.3^2$ *without* knowing the exact value of h. In the process, we come to appreciate the power of generality: we lose nothing in the way of precision but gain enormously in the sweeping nature of our conclusion. The identity (3) is valid *for all values of h.* Thus, if we let h be 0.1, then (3) gives $(1.3 + h)^2 - 1.3^2 = 2.6\,h + h^2 = 2.6 \times 0.1 + 0.1^2 = 0.26 + 0.01 = 0.27$, which is of course exactly $1.4^2 - 1.3^2$ $(=1.96 - 1.69)$.

Let us examine critically how the the sequence of steps (2) leading up to identity (1) differ from the usual computations in arithmetic. Because the exact values of x and y are not given, doing any of the steps in (2) by following a mechanical procedure of adding and multiplying is out of the

[1] We have chosen to soft-pedal the associative law of addition here.

question. We must go beyond *procedures* and rely on a *general understanding* of the operations of addition and multiplication. Whereas each computation of the type $(1.3 + 0.1)^2$ focuses on the individual numbers in question, the steps in (2) can be performed only when a certain abstract understanding of all numbers as a whole has been attained. It requires more than looking at individual numbers. Instead, one must focus on the abstract nature of the operations on real numbers. In saying this, we are aware of the possibility of performing the computations in (2) by rote, but solid evidence is available to show that the rote approach leads to errors in the long run. There is no known rote method of mastering abstract mathematics.

It may be profitable to revisit the computation in (2) from the opposite point of view. A natural question is whether someone who cannot perform computations of the type $(1.3 + 0.1)^2 = 1.96$ with ease is nevertheless able to do abstract computations such as (2). By and large, the answer is no, because a purely formal understanding of general properties of numbers such as the commutative law of multiplication or the distributive law, when not backed up by vast concrete experiences with specific numbers, tends to be fragile and incapable of supporting the kind of structural thinking implicit in the computations of (2). Working mathematicians know only too well that abstractions are rooted in concrete special cases. It may be somewhat tenuous, but an analogy can be made to the effect that someone who has limited knowledge of people as individuals would be equally limited in his overall understanding of human beings. In this sense, proficiency in arithmetic provides the foundation that makes possible the understanding of algebra.

An additional comment can be made about the kind of *symbolic computation* exhibited in the steps of (2). In the mathematics leading up to algebra, there is one subject which develops the skill of symbolic computation, and it is *fractions*. When fractions are taught correctly —- and it must be admitted that often it is not done correctly —- students get to manipulate integers almost on a symbolic level, e.g.

$$\frac{5}{7} + \frac{2}{3} = \frac{(5 \times 3) + (2 \times 7)}{7 \times 3}, \qquad \frac{2/9}{6/5} = \frac{2 \times 5}{9 \times 6}, \quad \text{etc.}$$

Of course, when we try to express these facts in complete generality, we have no choice but to use symbolic notation: for all integers a, b, c, and d, with $b \neq 0$ and $d \neq 0$,

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd},$$

and if also $c \neq 0$,

$$\frac{a/b}{c/d} = \frac{ad}{bc}\,.$$

In fact, it is generally conceded that the understanding of fractions holds the key to understanding algebra. For, beyond the matter of symbolic computations, a proper understanding of fractions —- again, when it is properly taught —- requires an understanding of the structure of the integers. In discussing whole number algorithms or fractions, it is the ability to reason about the operations on numbers that will safely lead students to algebraic competency. For example, in order to understand the meaning of the multiplication of two fractions, one is forced to first achieve an abstract understanding of the multiplication of two whole numbers. Similarly, understanding why "invert and multiply" is the right way to divide fractions requires an abstract understanding of the division of whole numbers. For this reason, the subject of fractions occupies an intermediate position in school mathematics between the arithmetic of whole numbers and algebra. One way to improve students' success rate in algebra is naturally to improve the teaching of fractions.

In trying to achieve success in teaching algebra, we are confronted with a basic fact in mathematics, namely, that mathematics is cumulative and hierarchical: learning any one topic requires a knowledge of most, if not all, of the topics preceding it.[2] The discussion above is a testimony to this fact. For this reason, if we hope to achieve the goal of teaching algebra to all eighth graders, a solid and substantive mathematics curriculum in grades K through 7 is absolutely essential. In this light, the rigorous demands of the California Mathematics Content Standards in grades K through 7 now begin to make sense.

[2] We have to put this statement in the context of school mathematics. In higher mathematics, this same hierarchical structure still holds, but in a modified form.