Learning algebra

Eugene, OR
October 17, 2009

H. Wu

*I am grateful to David Collins and Larry Francis for many corrections and suggestions for improvement.

This is a presentation whose target audience is primarily mathematics teachers of grades 5–8. The main **objectives** are to:

1. Explain the inherent conceptual difficulties in the learning of algebra.

2. Explain the artificial difficulties created by human errors.

3. Give two examples to illustrate what can be done to smooth students' entry into algebra.

1. Inherent conceptual difficulties

Arithmetic is about the computation of specific numbers. E.g.,

$$126 \times \frac{3}{8} = ?$$

Algebra is about what is true in general for **all** numbers, **all** whole numbers, **all** integers, etc. E.g.,

$$a^2 + 2ab + b^2 = (a+b)^2 \quad \text{for all numbers } a \text{ and } b$$

Going from the specific to the general is a giant conceptual leap. It took mankind roughly 33 centuries to come to terms with it.

1a. *Routine use of symbols*

Algebra requires the use of symbols at every turn. For example, we write a general quadratic equation without a moment of thought:

> *Find a number x so that*
>
> $$ax^2 + bx + c = 0$$
>
> *where a, b, c are fixed numbers*.

However, the ability to do this was the result of the aforementioned 33 centuries of conceptual development, from the Babylonians (17th century B.C.) to R. Descartes (1596-1640).

What happens when you don't have symbolic notation?

From al-Khwarizmi (circa 780-850): What must be the square which, when increased by 10 of its own roots, amounts to thirty-nine? The solution is this: You halve the number of roots, which in the present instance yields five. This you multiply by itself: the product is twenty-five. Add this to thirty-nine; the sum is sixty-four. Now take the root of this, which is eight, and subtract from it half the number of the roots, which is five; the remainder is three. This is the root of the square which you sought for.

An annotation:

What must be the square $[x^2]$ which, when increased by 10 of its own roots $[+10x]$, amounts to thirty-nine $[= 39]$? The solution is this: You halve the number of roots $[\frac{10}{2}]$, which in the present instance yields five. This you multiply by itself: the product is twenty-five. Add this to thirty-nine; the sum is sixty-four. Now take the (square) root of this, which is eight, and subtract from it half the number of the roots, which is five; the remainder is three. This is the root of the square which you sought for.

Solve $x^2 + 10x - 39 = 0$:

$$\frac{-10 + \sqrt{10^2 + 4 \times 39}}{2} = -\left(\frac{10}{2}\right) + \sqrt{\left(\frac{10}{2}\right)^2 + 39} = 3$$

Therefore, do not coddle your students in grades 3-8 by minimizing the use of symbols. *Celebrate the use of symbols instead.*

Teachers of primary grades: please use an n or an x, at least from time to time, whenever a $\square$ appears in a problem promoting "algebraic thinking", e.g.,

$$5 + \square = 13$$

There is *no* "developmental appropriateness" issue here. (See the Learning-Processes Task Group report of the National Math Panel, or the many articles of Daniel Willingham in *American Educator*.)

1b. *Concept of generality*

Generality and symbolic notation go hand-in-hand. How can we do mathematics if we don't have symbols to express, for example, the following *general* fact about a positive integer n?

> *The difference of any two nth powers is equal to the product of the difference of the two numbers and the sum of products consisting of the $(n-1)$th power of the first number, then the product of the $(n-2)$th power of the first and the first power of the second, then the product of the $(n-3)$th power of the first and the second power of the second, and so on, until the $(n-1)$th power of the second number.*

In symbols, this is succinctly expressed as the identity:

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + ab^{n-2} + b^{n-1})$$

for all numbers a and b

As an example of the **power of generality**, this identity implies (i) $17^7 - 6^7$ is not a prime number, nor is $815^{73} - 674^{73}$, etc., and (ii) one can sum any geometric series, e.g., letting $a = 1$ and $b = \pi$,

$$1 - \pi^n = (1-\pi)(1 + \pi + \pi^2 + \pi^3 + \cdots + \pi^n)$$

implies

$$1 + \pi + \pi^2 + \pi^3 + \cdots + \pi^n = \frac{\pi^n - 1}{\pi - 1}$$

The **need** for generality manifests itself in another context. Essentially all of higher mathematics and science and technology depends on the ability to represent geometric data algebraically or *analytically* (i.e., using tools from calculus). Thus something as simple as the algebraic representation of a line requires the language of generality. E.g.:

Consider all pairs of numbers (x, y) *that satisfy* $ax + by = c$, *where* a *and* b *are fixed numbers. Such a collection is a* **line** *in the coordinate plane.*

1c. *Abstract nature of algebra*

The main object of study of arithmetic is numbers: whole numbers, fractions, and negative numbers.

Numbers are tangible objects when compared with the main objects of study of algebra:

> *equations, identities, functions and their graphs, formal polynomial expressions (polynomial forms), and rational exponents of numbers (e.g.,* $2.4^{-6.95}$*).*

Among these, the concept of a function may be the most fundamental. Functions are to algebra what numbers are to arithmetic. For example, consider the function

$$f(x) \;=\; e^x\,(3x^4 - 7x + 11)$$

One cannot picture this function as a *finite collection of numbers*, say

$$e^5\,(3\cdot 5^4 - 7\cdot 5 + 11), \text{ and}$$

$$e^{21}\,(3\cdot 21^4 - 7\cdot 21 + 11),$$

but must take into account all the numbers $\{f(x)\}$ all at once. This is a difficult step to make for all beginning students, and no one approach can eliminate this difficulty. Graphing f provides partial help to visualizing all the $f(x)$'s, but the concept of a graph is itself an abstraction.

1d. *Precision*

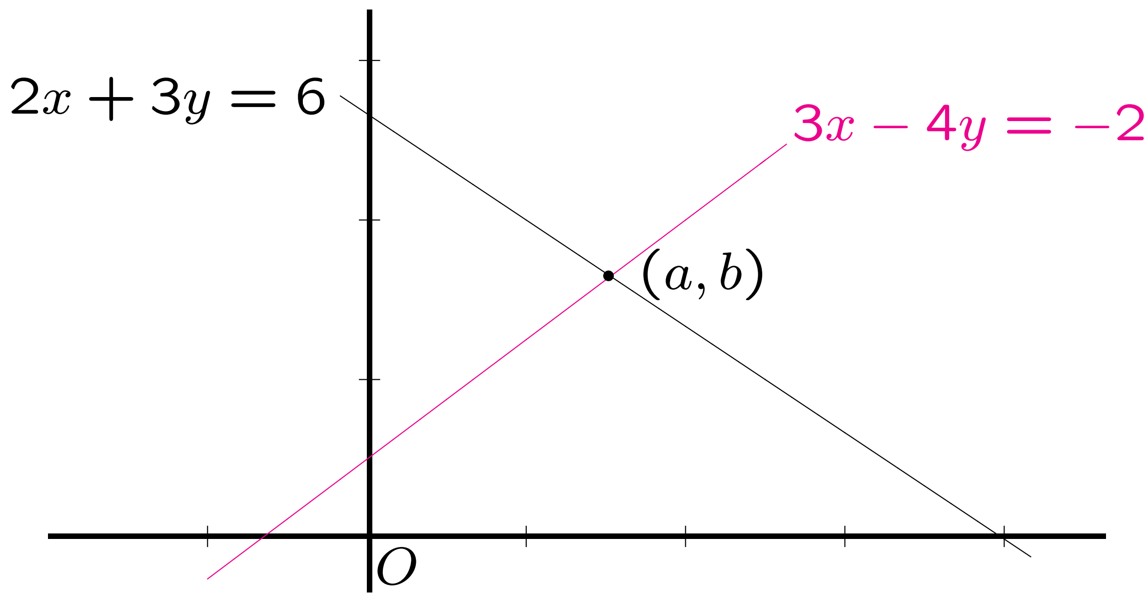
It is in the nature of an abstract concept that, because it is inaccessible to everyday experience, our only hope of getting to know it is by getting a precise description of what it is. This is why the more advanced the mathematics, the more abstract it becomes, and the more we are dependent on precision for its mastery.

While all of mathematics demands precision, the need for precision is far greater in algebra than in arithmetic.

Here is an example of the kind of precision necessary in algebra. We are told that to solve a system of equations,

$$\begin{aligned} 2x + 3y &= 6 \\ 3x - 4y &= -2 \end{aligned}$$

we just graph the two lines $2x + 3y = 6$ and $3x - 4y = -2$ to get the point of intersection (a, b), and (a, b) is *the* solution of the system.



But why is (a, b) **the** solution?

Because, by definition, the graph of $2x + 3y = 6$ consists of all the points (x, y) so that $2x + 3y = 6$. Since (a, b) is on the intersection of the graphs, in particular (a, b) lies on the graph of $2x + 3y = 6$ and therefore $2a + 3b = 6$.

For the same reason, we also have $3a - 4b = -2$.

So (a, b) is a solution of the system

$$\begin{aligned} 2x + 3y &= 6 \\ 3x - 4y &= -2 \end{aligned}$$

Now suppose the system has another solution (A, B). Thus $2A+3B = 6$ and $3A - 4B = -2$, by definition of a solution. Since the graph of $2x + 3y = 6$ consists of all the points (x, y) so that $2x+3y = 6$, the fact that $2A+3B = 6$ means the point (A, B) is on the graph of $2x+3y = 6$. Similarly, the point (A, B) is on the graph of $3x - 4y = -2$. Therefore (A, B) is on the intersection of the two graphs (which are lines).

Since two non-parallel lines meet at exactly one point, we must have $(A, B) = (a, b)$. So (a, b) is **the (only)** solution of the system.

If we do not emphasize from the beginning the *precise definition of the graph of an equation*, we cannot explain this fact.

Consider another example of the need for precision in algebra: the *laws of exponents*. These are:

For all positive numbers x, y, and for all **rational numbers** r and s,

$$\begin{aligned} x^r x^s &= x^{r+s} \\ (x^r)^s &= x^{rs} \\ (xy)^r &= x^r y^r \end{aligned}$$

For example,

$$2.4^{-3/5}\, 2.4^{2/7} = 2.4^{-3/5+2/7}$$

These laws are difficult to prove.

On the other hand, the starting point of these laws is the set of "primitive" laws of exponents which state:

For all positive numbers x, y, and for all **positive integers** m and n,

$$\begin{aligned} x^m x^n &= x^{m+n} \\ (x^m)^n &= x^{mn} \\ (xy)^m &= x^m y^m \end{aligned}$$

These are trivial to prove: just count the number of x's and y's on both sides, e.g., $x^8 x^5 = x^{8+5}$, or

$$xy \cdot xy \cdot xy \cdot xy \;=\; x^4 y^4$$

The two sets of laws look entirely similar, but the substantial difference between the two comes from the different *quantifications* of the symbols s, t, and m, n. The former are rational numbers and the latter are positive integers.

In algebra, it is therefore not sufficient to look at formulas *formally*. We must also pay attention to *exactly what each symbol represents* (i.e., its quantification). This is precision.

We will have more to say about the laws of exponents later.

The preceding difficulties in students' learning of algebra are real. *They cannot be eliminated*, any more than teenagers' growing pains can be eliminated. However, they *can* be minimized provided we have

a good curriculum,
good textbooks, and
good support from the educational literature.

Unfortunately, all three have let the students down.

My next goal is to describe some examples of this letdown, and also discuss possible remedies.

2. Difficulties due to human errors

2a. *Variables*

We all know that algebra is synonymous with "variables". What do we tell students a "variable" is? Here are two examples from standard textbooks.

> *A* **variable** *is a quantity that changes or varies. You record your data for the variables in a table. Another way to display your data is in a coordinate graph. A* **co-ordinate graph** *is a way to show the relationship between two variables.*

Another view:

Variable *is a letter or other symbol that can be replaced by any number (or other object) from some set. A* **sentence** *in algebra is a grammatically correct set of numbers, variables, or operations that contains a verb. Any sentence using the verb "$=$" (is equal to) is called an* **equation**.

A sentence with a variable is called an **open sentence**. *The sentence* $m = \frac{x}{5}$ *is an open sentence with two variables.*

An expression, such as $4 + 3x$, *that includes one or more variables is called an* **algebraic expression**. *Expressions are not sentences because they do not contain verbs, such as equal or inequality signs.*

In both cases, the author(s) seemed less interested in giving a detailed explanation of what a "variable" is than in introducing other concepts, e.g., "coordinate graph", "sentence", "equation", "expression", etc., to divert attention from "variable" itself, which is supposed to be *central*.

However there is no mistaking the message: a "variable" is something that **varies**, and that the minute you put down a symbol on paper, it becomes a "variable".

From your own teaching experience, can students make sense of "something that varies"? And do they know what they are doing when they put a symbol on paper?

2b. *Expressions*

The recent set of **Common Core Standards (CCS)**, released on September 17, 2009, has this to say about "expressions":

> *Expressions are constructions built up from numbers, variables, and operations, which have a numerical value when each variable is replaced with a number.*
>
> *Expressions use numbers, variables and operations to describe computations.*
>
> *The rules of arithmetic can be applied to transform an expression without changing its value.*

CCS has wisely chosen to bypass defining a "variable" and get to "expression" directly. Since CCS has aspirations to be the *de facto* national standards, its pronouncement on what an "expression" is must be taken seriously.

So what is an "expression" according to CCS? It is a construction. But what is a construction? Is it **any** assemblage of symbols and numbers? And what are the rules of the "operations" for the assemblage? **WHY** can the "rules of arithmetic be applied to transform an expression without changing its value"?

Do YOU think this tells you what an "expression" is? More importantly, can you use this to teach your eighth grader what an expression is?

2c. *Equations*

CCS says:

> *An equation is a statement that two expressions are equal.*

Without knowing what an "expression" is, we must now confront an "equation".

According to CCS,

$$xyzrstuvw \;=\; 2a + 3b + 4\,\frac{cdefghijklmn}{opq}$$

is an equation. What does it mean? Is this what you want to tell your students, and if so, do they know what you are talking about?

Let us approach "variable", "expression", and "equation" in a way that is consistent with how mathematics is done in mainstream mathematics.

The fundamental issue in algebra and advanced mathematics is **the correct way to use symbols.** Once we know that, then basically everything in school algebra falls back on arithmetic.

There will be no guesswork, and *no hot air*.

The cardinal rule in the use of symbols is to

specify explicitly what each symbol stands for.

This is called the **quantification** of the symbols.

Make sure your students understand that.

There is a good reason for this: symbols are the pronouns of mathematics. In the same way that we do not ask “Is he six feet tall?” without saying who “he” is, we do not write down

$$xyzrstuvw = a + 2b + 3cdefghijklmn$$

without first specifying what a, b, . . . , z stand for either (this is an equality between what? Two random collections of symbols?? What does it mean??).

Let x and y be two (real) numbers. Then *the number* obtained from x and y and a fixed collection of numbers by the use of the four operations $+$, $-$, $\times$, $\div$, together with $\sqrt[n]{\ }$ (for any positive integer n) and the usual rules of arithmetic, is called **an expression in x and y**. E.g., $\dfrac{85xy^2}{\sqrt{7}+xy} - \sqrt[3]{x^5 - \pi y}$.

An expression in other symbols a, b, . . . , z is defined similarly.

It is only when we make explicit the fact that, *in school algebra* (with minor exceptions), **each expression involves only numbers** that we can finally make sense of CCS's claim that "The rules of arithmetic can be applied to transform an expression without changing its value."

Still with x and y as (real) numbers, we may wish to find out all such x and y for which two given expressions in x and y are equal. For example, are there x and y so that

$$x^2 + y^2 + 3 \ = \ 0 \ ?$$

(No.) Note that 0 is the following expression in x and y: $0 + 0 \cdot x + 0 \cdot y$. Another example: Are there x and y so that

$$3x - 7y \ = \ 4 \ ?$$

(Yes, infinitely many.)

In each case, the equality of the given expressions in x, y is called **an equation in x and y**. To determine all the x and y that make the expressions equal is called **solving the equation**.

There are some subtle aspects to the quantification of symbols. For example, the meaning of a "quadratic equation $ax^2+bx+c=0$", when completely spelled out, is this:

> *Let a, b, c be fixed numbers. What numbers x satisfy $ax^2+bx+c=0$?*

Here a, b, c and x are all symbols, yet they play different roles. Because each of a, b, c stands for a *fixed* number in this equation, it is called a **constant**. *A priori*, there can be an infinite number of x, yet to be determined, that satisfy this equality. For this reason, this x is traditionally called a **variable**, or an **unknown**.

It is important to realize that the precise quantification of x in the meaning of a quadratic equation (or *any* equation) renders the terminology of a "variable" superfluous. There is no need for the term "variable".

Out of respect for tradition, the word "variable" *is* used in higher mathematics and in the sciences, *not* as a well-defined mathematical concept, but as an informal and convenient shorthand. For example, "a function of three variables $f(x, y, z)$" expresses the fact that the domain of definition of f is some region in 3-space. However, there is no need to teach an informal piece of terminology as a fundamental concept in school algebra.

2d. *Solving equations*

Textbooks tell you how to "solve" $x^2 - x - 1 = 0$ by completing the square:

$$\begin{aligned} x^2 - x &= 1 \\ (x^2 - x + \frac{1}{4}) &= 1 + \frac{1}{4} \\ (x - \frac{1}{2})^2 &= \frac{5}{4} \\ x - \frac{1}{2} &= \pm\frac{1}{2}\sqrt{5} \\ x &= \frac{1}{2}(1 \pm \sqrt{5}) \end{aligned}$$

The whole process is highly unsavory: It is a case of wanton manipulations of symbols with no reasoning whatsoever.

First, x is just a symbol, but the equality $x^2 - x = 1$ means the two *symbols* x^2 and $-x$ combined is equal to the *number* 1. *How can a number be equal to a bunch of symbols?*

The passage from $x^2 - x = 1$ to $(x^2 - x + \frac{1}{4}) = 1 + \frac{1}{4}$ is usually justified by "equals added to equals are equal", which is in turn justified by some *metaphors* such as adding $\frac{1}{4}$ to two sides of a balance, with $x^2 - x$ on one side and 1 on the other.

Finally, even ignoring all the questionable steps, how do we know $\frac{1}{2}(1 \pm \sqrt{5})$ *are* solutions of $x^2 - x - 1$? In other words, have we proved that the following is true?

$$(\frac{1}{2}(1 \pm \sqrt{5}))^2 - \frac{1}{2}(1 \pm \sqrt{5}) - 1 \ = \ 0$$

Let us now solve the equation correctly. Assume for the moment that there is a number $\mathbf{x}$ so that $\mathbf{x}^2 - \mathbf{x} - 1 = 0$. Then both sides of the equal sign are *numbers* and we are free to compute with numbers to obtain:

$$\begin{aligned}
\mathbf{x}^2 - \mathbf{x} &= 1 \\
(\mathbf{x}^2 - \mathbf{x} + \frac{1}{4}) &= 1 + \frac{1}{4} \\
(\mathbf{x} - \frac{1}{2})^2 &= \frac{5}{4} \\
\mathbf{x} - \frac{1}{2} &= \pm\frac{1}{2}\sqrt{5} \\
\mathbf{x} &= \frac{1}{2}(1 \pm \sqrt{5})
\end{aligned}$$

Notice that the whole process looks the same as before, but it now makes perfect sense as a computation in numbers.

We have now proved that if $\mathbf{x}$ is a number so that $\mathbf{x}^2 - \mathbf{x} - 1 = 0$, then necessarily $\mathbf{x} = \frac{1}{2}(1 \pm \sqrt{5})$. But with this as a hint, we can now directly check that $\frac{1}{2}(1 \pm \sqrt{5})$ are solutions of $x^2 - x - 1 = 0$, as follows. Let $\mathbf{x} = \frac{1}{2}(1 \pm \sqrt{5})$. Then

$$
\begin{aligned}
\mathbf{x}^2 - \mathbf{x} - 1 &= \mathbf{x}\,(\mathbf{x} - 1) - 1 \\
&= \left(\frac{1}{2}(1 \pm \sqrt{5})\right)\left(\frac{1}{2}(1 \pm \sqrt{5}) - 1\right) - 1 \\
&= \left(\frac{1}{2}(1 \pm \sqrt{5})\right)\ \left(\frac{1}{2}(-1 \pm \sqrt{5})\right) \quad - 1 \\
&= \frac{1}{4}(-1^2 + (\sqrt{5})^2) \ - 1 \ = \ 0
\end{aligned}
$$

Conclusion: the two numbers $\frac{1}{2}(1 \pm \sqrt{5})$ are solutions of $x^2 - x - 1 = 0$, and they are the *only* solutions.

Why should students learn how to solve equations correctly?

(i) They come to the full realization that achieving algebra depends on a robust knowledge of (rational) numbers, and not on the unknowable concept of a "variable". Basically, **everything in school algebra falls back on arithmetic.**

(ii) They realize that solving equations is not a sequence of mindless manipulations of symbols but a progression of well-understood procedures with numbers, based on **reasoning**. Mathematics becomes **knowable**.

(iii) For the case of quadratic equations, they learn the *how* and the *why* of the **quadratic formula** and gain the confidence that they can derive it at will. This knowledge diminishes the need for memorization.

2e. *Precision*

Of the endless examples of how school textbooks ignore precision, we will discuss two. The first is about the definition of rational exponents of a (positive) number. The starting point is always the following laws of exponents for positive integers:

For all numbers x, y, and for all **positive integers** m and n,

$$\begin{aligned} x^m x^n &= x^{m+n} \\ (x^m)^n &= x^{mn} \\ (xy)^m &= x^m y^m \end{aligned}$$

Most textbooks now present the "proof" that, e.g., $5^0 = 1$:

> *Because* $x^m x^n = x^{m+n}$, $5^2 \cdot 5^0 = 5^{2+0} = 5^2$. *Dividing both sides by* 5^2, *we get* $5^0 = 1$.

Comments This is not a proof, because the formula used to justify $5^2 \cdot 5^0 = 5^{2+0}$ is $x^m x^n = x^{m+n}$, which is only valid for $m > 0$ and $n > 0$. We must use each fact *precisely* without unwarranted extrapolation.

In addition, because we are still trying to find out what 5^0 means, we cannot use it in an equation *in order to compute its value*. This is called **circular reasoning**.

Another "proof" that $5^0 = 1$ is to appeal to patterns:

$$\cdots 5^3 = 125, \quad 5^2 = 25, \quad 5^1 = 5, \quad 5^0 = ?$$

As we go to the right, each number is obtained from the preceding one by dividing by 5*. Thus going from* $5^1 = 5$ *to* 5^0*, we must divide 5 by 5, yielding* $5^0 = 1$*.*

Comments How do we know that the pattern would persist all the way down to 5^0?

These are not proofs. The fact that $5^0 = 1$ is a matter of **definition**.

We now explain why we make the following **definitions**: for any $x > 0$, any positive integer m, n,

$$\boldsymbol{x^0 = 1, \quad x^{m/n} = (\sqrt[n]{x})^m, \quad x^{-m/n} = \frac{1}{x^{m/n}}}$$

Again the starting point is the law of exponents for all **positive integers** m and n, and for any $x > 0$,

$$\begin{aligned} x^m x^n &= x^{m+n} \\ (x^m)^n &= x^{mn} \\ (xy)^m &= x^m y^m \end{aligned}$$

These laws are so "nice" that people decide that they should remain true even when m and n are rational numbers. The definitions of x^0, x^{-n}, etc., are made with this goal in mind.

If $x^r x^s = x^{r+s}$ is to be true for all rational numbers r and s, then for any positive integer n, we must have

$$x^n x^0 \; = \; x^{n+0} \; = \; x^n$$

so that, dividing both sides by x^n, we obtain $x^0 = 1$.

Thus, if $x^r x^s = x^{r+s}$ has any hope of being true for all rational numbers r and s, we must have $x^0 = 1$. This is why we adopt it as the definition.

Similarly, **if** $(x^r)^s = x^{rs}$ is to be true for all rational numbers r and s, then for any positive integer n, we must have

$$(x^{1/n})^n = x^{(1/n)n} = x^1 = x$$

This means $x^{1/n}$ is a number whose nth power is x itself. You recognize this number $x^{1/n}$ to be the **positive $\boldsymbol{n}$th root $\boldsymbol{\sqrt[n]{x}}$ of $\boldsymbol{x}$**.

Again, **if** $(x^r)^s = x^{rs}$ is to be true for all rational numbers r and s, we must have $x^{m/n} = (x^{1/n})^m = (\sqrt[n]{x})^m$. This is why we adopt the **definition:** for all positive integers m and n,

$$\boldsymbol{x^{m/n} = (\sqrt[n]{x})^m}$$

The same reasoning yields the "correct" definition that

$$\boldsymbol{x^{-m/n} = \frac{1}{x^{m/n}}}$$

With these definitions in place, we are now at least in a position to ask whether the following is true:

For all positive numbers x, y, and for all **rational numbers** r and s,

$$\begin{aligned} x^r x^s &= x^{r+s} \\ (x^r)^s &= x^{rs} \\ (xy)^r &= x^r y^r \end{aligned}$$

The proof that these laws of exponents for rational numbers r and s are true is tedious and boring, and is not our concern here. We are interested, rather, in students' ability to think clearly: whether they know the distinction between a valid proof and a seductive but incorrect argument, and whether they know the distinction between a definition and a theorem.

It takes precise reasoning to draw such distinctions.

Such precise reasoning is what it takes to learn advanced mathematics and science.

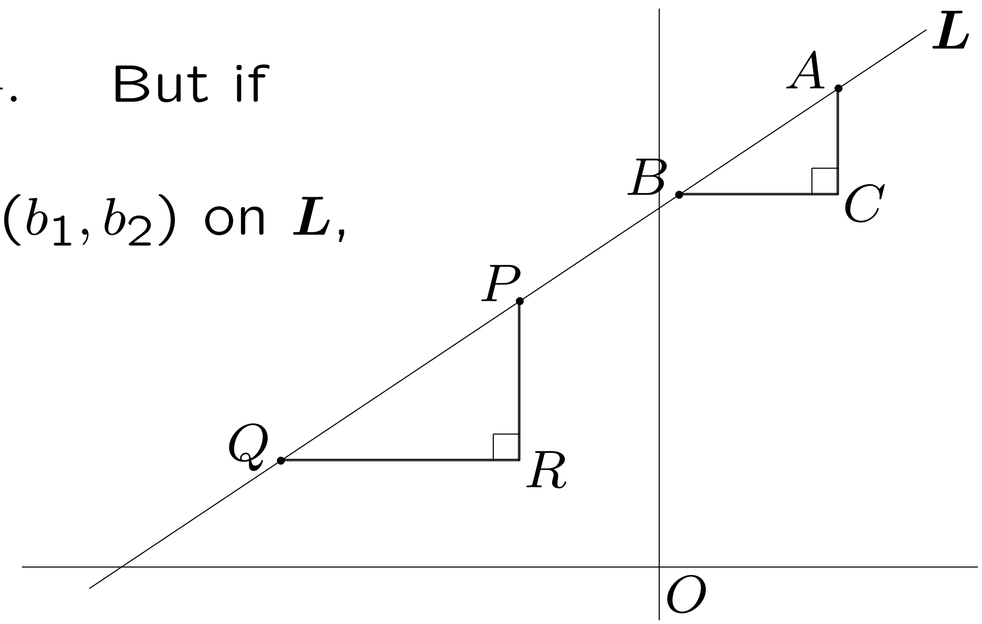
As a second example of the need for precision, consider the definition of the slope of a line $\boldsymbol{L}$. Let $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ be distinct points on $\boldsymbol{L}$. Then the usual definition is:

Slope of $\boldsymbol{L}$ is $\dfrac{p_2 - q_2}{p_1 - q_1}$. But if

$A = (a_1, a_2)$ and $B = (b_1, b_2)$ on $\boldsymbol{L}$,

we **must prove:**

$$\frac{p_2 - q_2}{p_1 - q_1} = \frac{a_2 - b_2}{a_1 - b_1}$$



This proof requires the concept of similar triangles: $\triangle ABC \sim \triangle PQR$.

What is alarming is the fact that textbooks want students to conflate **the slope of a line L** with **the slope of two chosen points P and Q on L.**

If students get comfortable with blurring the distinction between two concepts as different as these, they may never be able to learn any mathematics of value again. Without precision, there is no mathematics.

In the short term, students will not understand why the graph of $ax+by=c$ is a line, and consequently will have trouble *struggling to memorize* the four forms of the equation of a line.

2f. *Hidden agenda*

Too often we teach students mathematics that *suppresses a needed assumption*.

For example, textbooks and standardized tests promote pattern problems to get students into "algebraic thinking". Here is a typical test item:

> *What is the next number in the sequence 2, 3, 5, 9, 17 where each number is obtained from the preceding number by the same fixed rule?*

$$2, 3, 5, 9, 17$$

If your answer is 33, you will make many people very happy because you believe exactly what they want you to believe, namely, that the rule is: "If the number is k, the next number is $\mathbf{2}\boldsymbol{k}-\mathbf{1}$".

What if I tell you the rule is actually

$$\boldsymbol{(2k-1)+(k-2)(k-3)(k-5)(k-9)}$$

Make sure that this also works.

Then the next number would not be 33, but

$$33 + 15 \cdot 14 \cdot 12 \cdot 8 = 20193.$$

The hidden assumption of this pattern problem is that

the rule should be a *linear* function of k.

Here is another example of a hidden agenda. Consider the following standard problem:

> *Janice walked 3.6 miles. It took her 35 minutes to walk 2.1 miles. How long did it take her to walk the whole 3.6 miles?*

The usual way to do it this to *set up a proportion*: if it took her x minutes to walk 3.6 miles, then

$$\frac{2.1}{35} = \frac{3.6}{x}$$

Cross-multiply, and we get the answer: $x = 60$ minutes.

How to explain "setting up a proportion"?

Janice's walking routine was actually the following:

> Walk briskly for 2.1 miles in 35 minutes, rest 40 minutes, and walk another 1.5 miles in 45 minutes.

In this case, it actually took her $35 + 40 + 45 = 120$ minutes to walk the 3.6 miles, so that *this*, and not 60 minutes, should be the correct answer to the problem.

Maybe "setting up a proportion" is not such a good idea after all.

If we insist on "setting up a proportion", then we must (i) revise the problem by making explicit the hidden assumption(s), and (ii) explain why it is valid to "set up a proportion".

Revised problem:

> *Janice walked 3.6 miles at a constant speed. It took her 35 minutes to walk 2.1 miles. How long did it take her to walk the whole 3.6 miles?*

The hidden assumption is that she walked at a *constant speed*.

Now we explain *constant speed*, which is usually suppressed in textbooks. First, assume that in a fixed time interval from time s to time t (in minutes), she walks d miles. Then we say that her **average speed in the time interval** $[\boldsymbol{s}, \boldsymbol{t}]$ is $\frac{d}{t-s}$ mi/min.

Suppose Janice walked as described:

> Walk briskly for 2.1 miles in 35 minutes, rest 40 minutes, and walk another 1.5 miles in 45 minutes.

Then her average speed in the time interval $[0, 35]$ is $\frac{2.1}{35} = 0.06$ mi/min. Her average speed in the time interval $[0, 40]$ is $\frac{2.1}{40} = 0.0525$ mi/min. Her average speed in the time interval $[0, 75]$ is $\frac{2.1}{75} = 0.028$ mi/min. Her average speed varies with the choice of the time interval.

We say she walks **at a constant speed** if her average speed in **any** time interval is equal to a fixed constant v (mi/min). We then call v her **speed**.

We go back to our problem:

> *Janice walked 3.6 miles at a constant speed. It took her 35 minutes to walk 2.1 miles. How long did it take her to walk the whole 3.6 miles?*

As before, let x be the number of minutes it took her to walk 3.6 miles. Her average speed in the time interval $[0, 35]$ is $\frac{2.1}{35}$ mi/min. Also, her average speed in the time interval $[0, x]$ is $\frac{3.6}{x}$ mi/min. Since her speed is constant,

$$\frac{2.1}{35} = \frac{3.6}{x}$$

This was the "proportion" we set up by rote last time. Now cross-multiply, and we get the same answer: $x = 60$ minutes.

If we use algebra, the explanation becomes more transparent.

Let $f(t)$ be the number of miles Janice walked from 0 minute to t minutes. Since she walked at a constant speed (say v mi/min), her average speed in the time interval $[0, t]$ for *any* t is

$$\frac{f(t)}{t-0} = v, \quad \text{which means} \quad f(t) = vt$$

The function $f(t)$ is a linear function without constant term.

Given $f(35) = 2.1$. We want the time t_0 so that $f(t_0) = 3.6$. Now $f(35) = v(35)$, so $v = \frac{2.1}{35}$ mi/min. Since $f(t_0) = vt_0 = \frac{2.1}{35}\, t_0$, we have

$$\frac{2.1}{35}\, t_0 = 3.6, \quad \text{and} \quad t_0 = 60 \text{ min}$$

Moral: Mathematics is WYSIWYG. Every assumption needed for the solution of a problem must be on the table. This openness makes mathematics learnable, and the same openness gives every student an equal opportunity to learn.

The teaching of mathematics must respect this WYSIWYG characteristic.

2g. *Coherence*

Mathematics is more learnable if there is continuity from topic to topic, and from grade to grade.

Consider the addition of rational expressions, e.g.,

$$\frac{x}{x+1} + \frac{1}{x-1} = \frac{x(x-1)+(x+1)}{(x+1)(x-1)} = \frac{x^2+1}{x^2-1}$$

If x is a "variable", this addition would be coming out of nowhere and would make sense only if we *arbitrarily decree*, as CCS does, that variables obey the usual rules of arithmetic. An arbitrary decree is not the kind of continuity we seek.

On the other hand, if we regard all expressions in a number x **as numbers**, then letting A, B, C, D be the numbers

$$A = x, \quad B = x + 1, \quad C = 1, \quad D = x - 1,$$

we see that

$$\frac{x}{x+1} + \frac{1}{x-1} = \frac{x(x-1) + (x+1)}{(x+1)(x-1)} = \frac{x^2+1}{x^2-1}$$

is just

$$\frac{A}{B} + \frac{C}{D} = \frac{AD + BC}{BD}$$

On the most basic level then, we make the connection that the addition of rational expressions is just the ordinary addition of fractions provided the latter is taught correctly without reference to Least Common Denominator.

Let us look more closely at

$$\frac{x}{x+1} + \frac{1}{x-1} \;=\; \frac{x(x-1)+(x+1)}{(x+1)(x-1)} \;=\; \frac{x^2+1}{x^2-1}$$

If this is supposed to be true for **all** numbers, then it is true for (say) $x = \frac{4}{3}$. What we have is therefore the following addition:

$$\frac{\frac{4}{3}}{\frac{7}{3}} + \frac{1}{\frac{1}{3}} \;=\; \frac{\frac{4}{3}\cdot\frac{1}{3} + \frac{7}{3}}{\frac{7}{3}\cdot\frac{1}{3}}$$

Now we realize we are not just adding fractions, but are adding **complex fractions**, which are division of fractions by fractions. Have you seen textbooks discussing *the addition of complex fractions*? If not, how to transition from fractions to algebra?

There is more. Suppose we let $x = \pi$, then

$$\frac{x}{x+1} + \frac{1}{x-1} \;=\; \frac{x(x-1)+(x+1)}{(x+1)(x-1)} \;=\; \frac{x^2+1}{x^2-1}$$

implies

$$\frac{\pi}{\pi+1} + \frac{1}{\pi-1} \;=\; \frac{\pi(\pi-1)+(\pi+1)}{(\pi+1)(\pi-1)} \;=\; \frac{\pi^2+1}{x^2-1}$$

Now we are talking about the addition of "fractions" whose numerators and denominators are *irrational numbers*. Have you seen textbooks discussing *the addition of such "fractions"*? Again, how to transition from fractions to algebra?

We have just touched the tip of the iceberg of the phenomenon of our incoherent curriculum. It is time we leave behind the discussion of the difficulties of learning algebra caused by human errors and take a closer look at curriculum.

We will give a brief discussion of our K-8 curriculum and then use it as a springboard to launch two illustrative examples of how things can be done better in the pre-algebra curriculum.

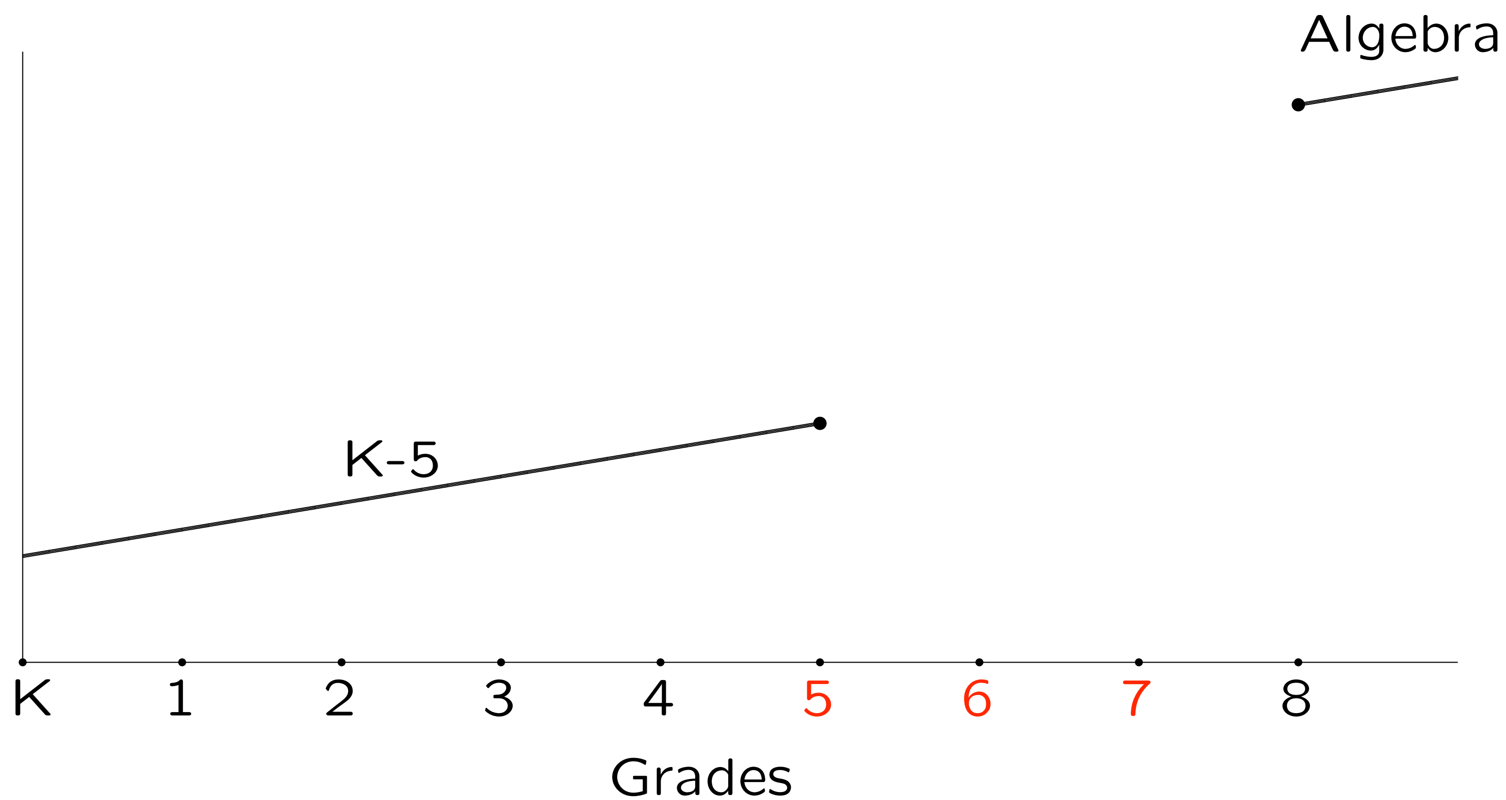
Allowing for some minor variations, the essence of the K-8 curriculum can be laid bare in the schematic diagram:

Whole numbers $\longrightarrow$ Fractions $\longrightarrow$ Rational numbers $\longrightarrow$ Algebra

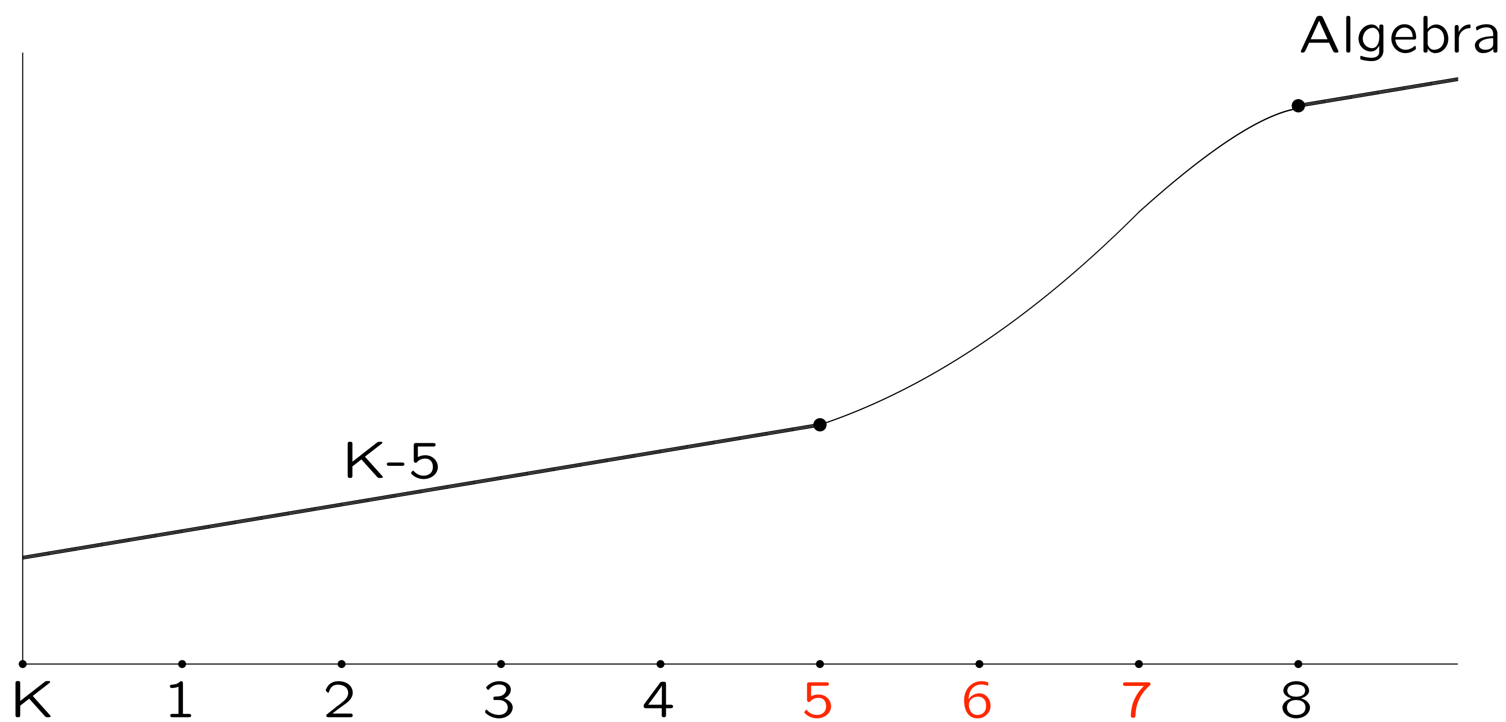
Notice the increase in abstraction as we go from left to right. Fractions and negative numbers (taught mainly in grades 5–7) are inherently abstract concepts, so that building on the abstraction of fractions and negative numbers would be an ideal way to acclimatize students to the abstraction of algebra.

However, the curriculum we have as of 2009 seems intent on bypassing the abstract nature of fractions and negative numbers. Instead it relies almost exclusively on the use of hands-on activities, analogies, and metaphors to teach these topics.

Graphically, we can present the situation this way:



To go from grade 5 to grade 8, one **might** gradually increase the use of symbols and elevate the level of abstraction to give students a smooth transition:

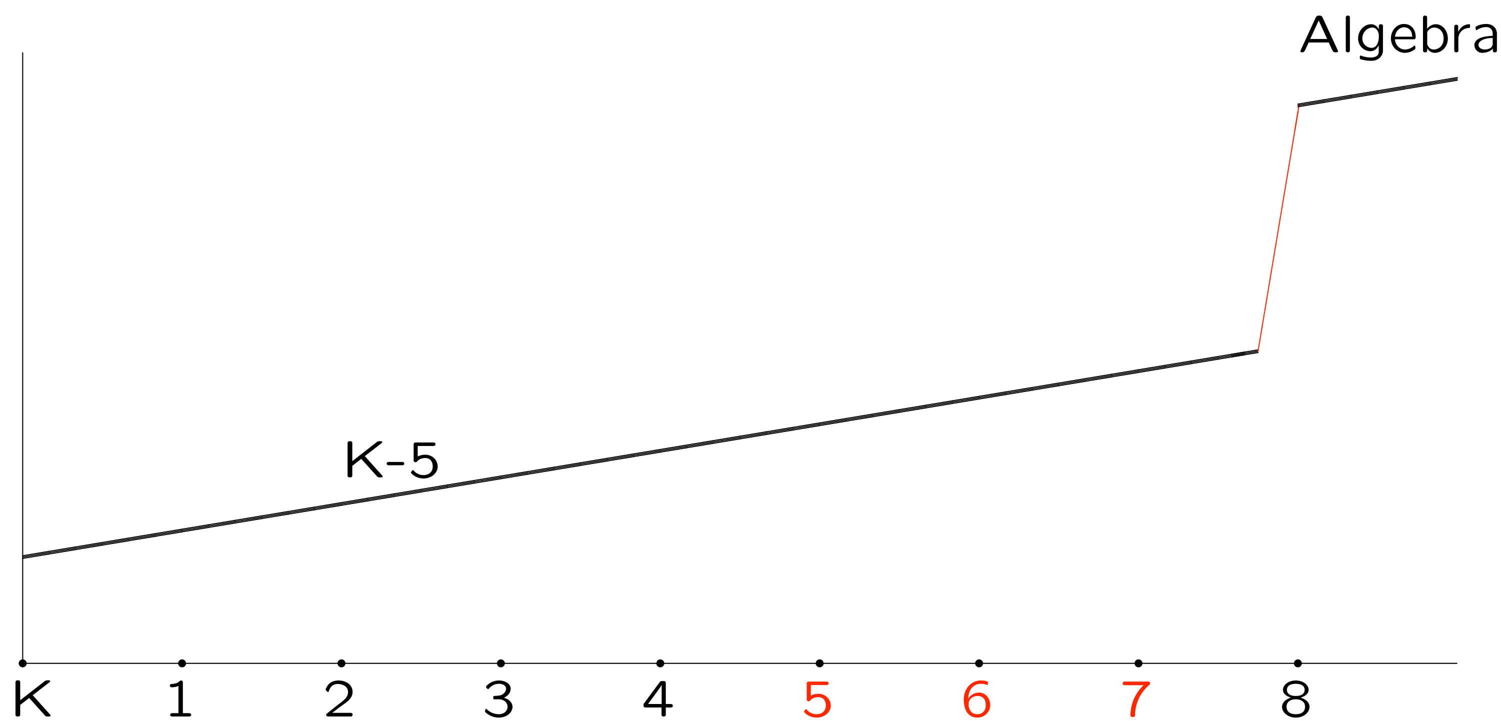


The reality is different.

Instead of teaching students that a fraction is a number, we teach fraction *exclusively* as a piece of pizza or part of the unit square. Instead of teaching why negative $\times$ negative $=$ positive by a careful application of the distributive law, we invent convoluted stories about how to make money by not paying a debt and then rely *exclusively* on those stories to teach this basic fact.

Instead of gradually increasing the use of symbols to state and explain basic facts about the four operations on fractions, we reduce most explanations only to those amenable to picture-drawings or hands-on activities. Thus only single-digit numbers are used most of the time for numerators and denominators.

The end result is a curriculum that flattens out in the critical grades of 5–7 and leaves students to negotiate the steep climb to algebra on their own:



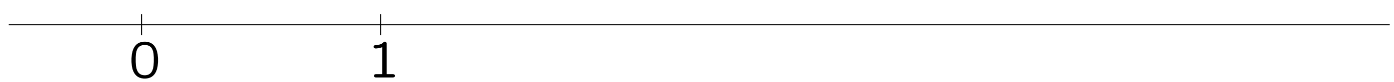
3. Two examples of what can be done

3a. *Addition of fractions*

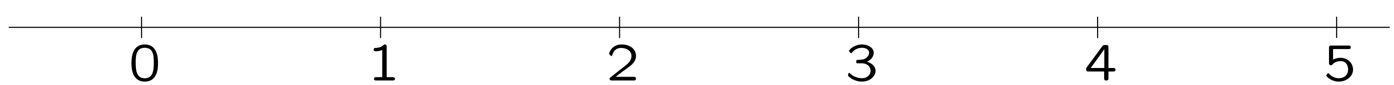
This topic is a bit more profound than is usually realized. We start from the beginning by giving a definition of a fraction. Let us first define all **fractions with denominator equal to 3.**

But why do we need a definition of fractions?
(i) If students have to add, subtract, multiply, and divide fractions, then they have to know what a fraction is. One cannot work with something when one doesn't even know what it is.
(ii) Reasoning in algebra and higher mathematics depends on precise definitions. Learning how to work with a precise definition of a fraction is an excellent introduction to algebra.

On a horizontal line, let two points be singled out. Identify the point to the left with 0 and the one to the right with 1. This segment, denoted by $[\mathbf{0, 1}]$, is called the **unit segment**.

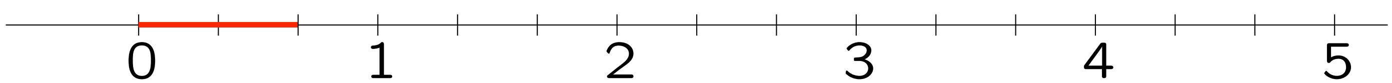


Now mark off equidistant points to the right of 1 as in a ruler, as shown, and identify the successive points with 2, 3, 4, ...

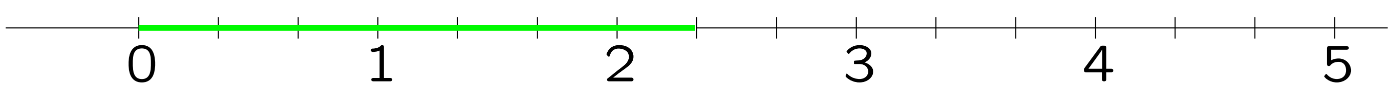


The line, with a sequence of equidistant points on the right identified with the whole numbers, is called the **number line**.

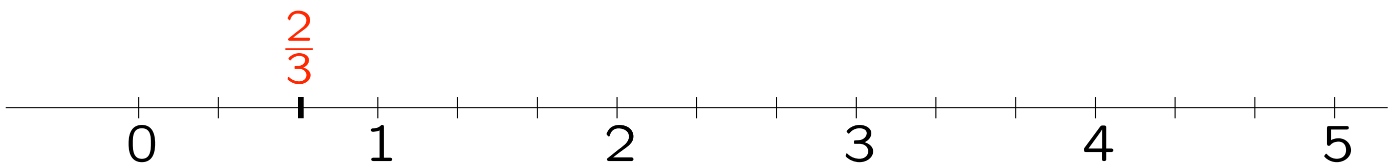
The unit segment $[0, 1]$ is taken to be **the whole**. Naturally, all other segments $[1, 2]$ (the segment between 1 and 2), $[2, 3]$, etc., can also be taken to be the whole. Divide each such segment into **thirds** (three segments of equal length); then we can count the number of thirds (the "parts") by going from left to right starting from 0. Thus, the red segment comprises *two thirds*:



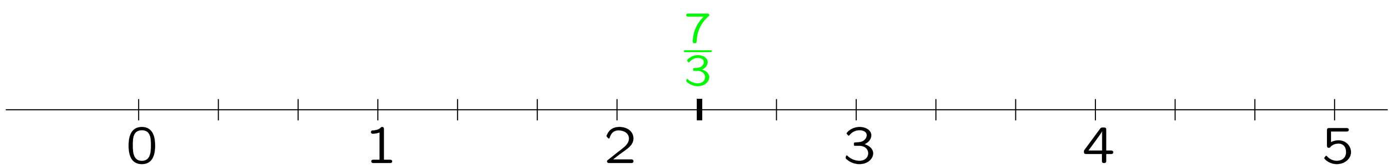
The following green segment comprises *seven thirds*:



Clearly, the red segment can be replaced by its *right* endpoint, which we naturally denote by $\frac{2}{3}$:



Likewise, the green segment will be replaced by its *right* endpoint, which is denoted by $\frac{7}{3}$:



Thus each "parts of a whole" (in the present context of *thirds*) is now replaced by a point on the number line. The point that is the *7th* point to the right of 0 is denoted by $\frac{7}{3}$. The point that is the *nth* point to the right of 0 is denoted by $\frac{n}{3}$.

The fractions with denominator equal to 5 are similarly placed on the number line: $\frac{8}{5}$ is the 8th point to the right of 0 in the sequence of *fifths*. And so on.

In general, if n is a positive integer, the fraction $\frac{3}{n}$ is the third point to the right of 0 among the n*ths* on the number line, and if m is a whole number, then the fraction $\frac{m}{n}$ is the mth point to the right of 0 among the n*ths* on the number line.

We also agree to identify $\frac{0}{n}$ *with* 0 *for any positive integer* n. In this way, all fractions are unambiguously placed on the number line.

Two things are noteworthy:

(i) The fractions with denominator 3 are *qualitatively* no different from the whole numbers: both are a sequence of equidistant points on the number line, and if we replace $\frac{1}{3}$ by 1, then the former sequence becomes the whole numbers.

(ii) **The number line is to fractions what one's fingers are to whole numbers:** It anchors students' intuition about fractions.

For the sake of conceptual clarity as well as ease of mathematical reasoning, we will henceforth **define** a fraction to be a point on the number line as described above.

You will see, presently, why this definition of a fraction is an advantage.

We now approach *the* fundamental fact about fractions.

Theorem on equivalent fractions *Given any two fractions* $\frac{m}{n}$ *and* $\frac{k}{\ell}$. *If there is a positive integer* c *so that*

$$m = c\,k, \qquad \text{and} \qquad n = c\,\ell$$

then the fractions are equal, i.e., $\dfrac{m}{n} \;=\; \dfrac{k}{\ell}$.

In your classroom, you wouldn't teach like this. You'd begin by saying, $\frac{3}{6}$ and $\frac{1}{2}$ are equal because

$$3 = 3 \times 1 \ \text{ and } \ 6 = 3 \times 2$$

$\frac{10}{12}$ and $\frac{5}{6}$ are equal because

$$10 = 2 \times 5 \ \text{ and } \ 12 = 2 \times 6$$

$\frac{14}{6}$ and $\frac{7}{3}$ are equal because

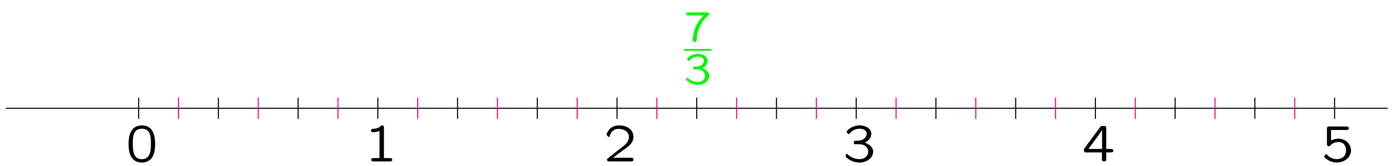
$$14 = 2 \times 7 \ \text{ and } \ 6 = 2 \times 3$$

You give many examples before stating the general fact above.

Let us prove $\frac{14}{6} = \frac{7}{3}$.

We must show that the 7th point to the right of 0 in the sequence of *thirds* is also the 14th point to the right of 0 in the sequence of *sixths*. (*Observe that with a clear-cut definition of a fraction, there is no ambiguity about what we must prove.*)

We divide each of the thirds into 2 equal parts:



The number line now has a sequence of *sixths*, and the 14th point to the right of 0 is therefore exactly the 7th point to the right of 0 in the sequence of *thirds*, as claimed.

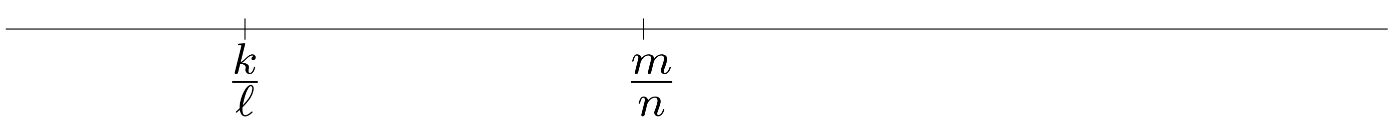
The reasoning in general for $\frac{ck}{c\ell} = \frac{k}{\ell}$ is exactly the same.

This theorem lies behind every statement about the operations of fractions. We now give some illustrations.

Here is a standard problem:

Which of $\frac{19}{54}$ *and* $\frac{6}{17}$ *is bigger?*

Before we try to answer this question, we ask what is meant by "bigger"? The number line allows us to give an unambiguous definition: a fraction $\frac{m}{n}$ is said to be **bigger than** another fraction $\frac{k}{\ell}$ if on the number line, $\frac{m}{n}$ is to the right of $\frac{k}{\ell}$.



Now we have to decide which of $\frac{19}{54}$ and $\frac{6}{17}$ lies to the right of the other on the number line. This is the point about definitions: if we agreed on "lying further to the right" as the meaning of "bigger", then we are obligated to follow through by showing one of $\frac{19}{54}$ and $\frac{6}{17}$ lies further to the right and not by appealing to another metaphor.

It must be recognized that the difficulty lies in having to compare the 19th point (to the right of 0) in the sequence of *54ths* with the 6th point (to the right of 0) in the sequence of *17ths*. How to compare a **54th** with a **17th**, i.e., $\frac{1}{54}$ with $\frac{1}{17}$?

Consider, for example, the analogous problem:

Which is longer, 19 feet or 6 meters?

You try to find *a common unit*. In this case, cm is good:

$$19 \text{ ft.} = 19 \times 12 \text{ in.} = 228 \text{ in.} = 228 \times 2.54 \text{ cm} = 579.12 \text{ cm}$$

Since 6 meters is 600 cm, we see that 6 meters is longer.

Thus, faced with comparing 19 **54ths** and 6 **17ths**, we try to find a common unit for $\frac{1}{54}$ and $\frac{1}{17}$. The Theorem says

$$\frac{1}{54} = \frac{17}{54 \times 17} \quad \text{and} \quad \frac{1}{17} = \frac{54}{54 \times 17}$$

So $\frac{1}{54\times17}$ will serve as a common unit for $\frac{1}{54}$ and $\frac{1}{17}$.

Now we apply the Theorem twice to get:

$$\frac{19}{54} = \frac{19 \times 17}{54} = \frac{323}{54 \times 17}$$

$$\frac{6}{17} = \frac{54 \times 6}{54 \times 17} = \frac{324}{54 \times 17}$$

In terms of $\frac{1}{54 \times 17}$, the 324th point to the right of 0 is clearly on the right side of the 323rd point. We therefore conclude that $\frac{6}{17}$ is bigger than $\frac{19}{54}$.

The basic idea of the above can be abstracted: given two fractions $\frac{m}{n}$ and $\frac{k}{\ell}$, the Theorem says we can rewrite them as two fractions with equal denominators,

$$\frac{kn}{\ell n} \quad \text{and} \quad \frac{\ell m}{\ell n}$$

Thus any two fractions may be regarded as two fractions with the same denominator. We call this **FFFP**, the **Fundamental Fact on Fraction Pairs**. FFFP has far reaching consequences. For examples, it immediately implies:

Cross-Multiplication Algorithm (CMA) : *Given any two fractions* $\frac{k}{\ell}$ *and* $\frac{m}{n}$,

$$\frac{k}{\ell} = \frac{m}{n} \text{ if and only if } kn = \ell m$$

$$\frac{k}{\ell} < \frac{m}{n} \text{ if and only if } kn < \ell m$$

We apply FFFP to the comparison of decimals. By definition, a **decimal** is a fraction whose denominator is a power of 10 written in the special notation introduced by the German Jesuit astronomer C. Clavius (1538-1612):

$$\text{write} \quad \frac{235}{10^2} \quad \text{as} \quad 2.35; \quad \text{write} \quad \frac{57}{10^4} \quad \text{as} \quad 0.0057$$

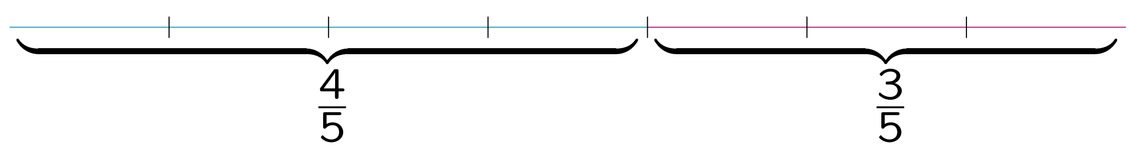
Therefore comparing decimals is a special case of comparing fractions (**tell your students that!**), with one advantage. If we want to compare, let us say, 0.12 with 0.098, there is no doubt as to what common denominator to use:

$$\text{rewrite} \quad \frac{12}{100} \quad \text{and} \quad \frac{98}{1000} \quad \text{as} \quad \frac{120}{1000} \quad \text{and} \quad \frac{98}{1000},$$

respectively. Clearly 0.12 is bigger.

We can finally consider the addition of fractions. First, how do we add whole numbers? $4+3$ is the length of the **concatenation** of one segment of length 4 followed by a second segment of length 3:

Now the meaning of $\frac{\mathbf{4}}{\mathbf{5}} + \frac{\mathbf{3}}{\mathbf{5}}$ should be the same: we define it to be the length of the **concatenation** of one segment of length $\frac{4}{5}$ followed by a second segment of length $\frac{3}{5}$:



The continuity from the definition of the addition of whole numbers to the definition of the addition of fractions is of critical importance for learning fractions.

At the moment, students in elementary school learn the addition of whole numbers as "combining things", but when they come to the addition of fractions, suddenly "addition" becomes an unfathomable maneuver about getting the least common multiple of the denominators and rewriting the numerator. They become disoriented and mathphobia would be a natural consequence.

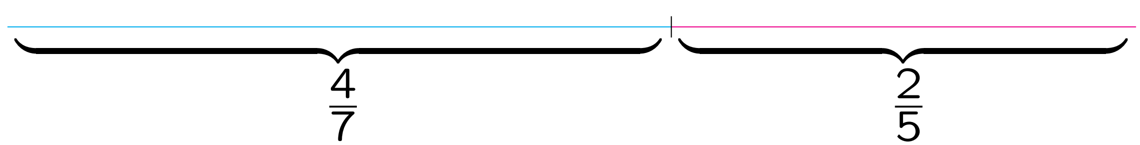
The above definition of fraction addition shows clearly why adding fractions still means "combining things".

By the definition,

$$\frac{4}{5}+\frac{3}{5} \;=\; \frac{4+3}{5}$$

which is the same addition as $4+3$ above, except that the number 1 is now replaced by $\frac{1}{5}$.

Next, we consider something more complicated: We define $\mathbf{\frac{4}{7}} + \mathbf{\frac{2}{5}}$ in exactly the same way: it is the length of the **concatenation** of one segment of length $\frac{4}{7}$ followed by a second segment of length $\frac{2}{5}$:



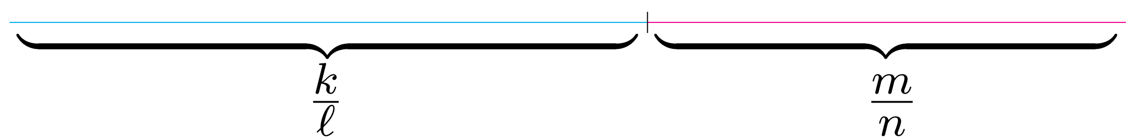
By definition, $\dfrac{4}{7} + \dfrac{2}{5}$ is the total length of 4 of the $\frac{1}{7}$'s and 2 of the $\frac{1}{5}$'s.

This is like adding 4 feet and 2 meters; we cannot find its exact value until we can find *a common unit for foot and meter*.

The same with fractions. FFFP tells us what to do: use $\frac{1}{7\times 5}$ as the common unit.

$$\frac{4}{7}+\frac{2}{5} \;=\; \frac{4\times 5}{7\times 5}+\frac{7\times 2}{7\times 5} \;=\; \frac{34}{35}$$

In general, we define the addition of $\frac{k}{\ell}$ and $\frac{m}{n}$ in exactly the same way: $\boldsymbol{\frac{k}{\ell}} + \boldsymbol{\frac{m}{n}}$ is the length of the **concatenation** of one segment of length $\frac{k}{\ell}$ followed by a second segment of length $\frac{m}{n}$:



By FFFP,

$$\frac{k}{\ell} + \frac{m}{n} = \frac{kn}{\ell n} + \frac{\ell m}{\ell n} = \frac{kn + \ell m}{\ell n}$$

Note that we have added fractions **without** once considering the Least Common Denominator. The LCD contributes nothing to the understanding of fraction addition.

A 1978 NAEP question in eighth grade:

Estimate $\dfrac{12}{13} + \dfrac{7}{8}$.

(1) 1

(2) 19

(3) 21

(4) I don't know

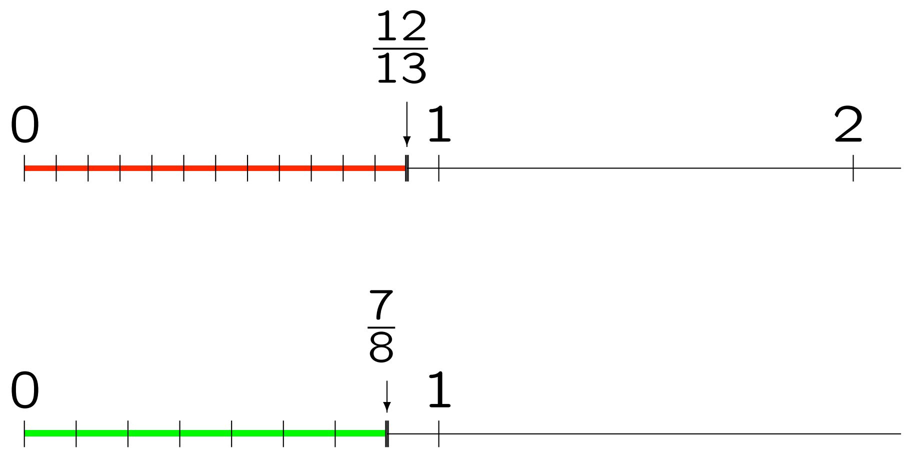
(5) 2

The statistics:

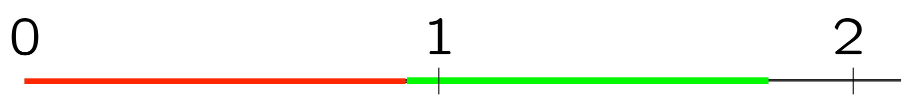
- 7% chose “1”.
- 28% chose “19”.
- 27% chose “21”.
- 14% chose “I don’t know”.
- 24% chose “2” (the correct answer).

Could this be exclusively the fault of our students?

Students need a mental image of a fraction to replace the mental image of a whole number given by the fingers on their hands.



Direct concatenation gives:



We now take up the more subtle aspects of the addition of fractions. Recall the above formula:

$$\frac{k}{\ell} + \frac{m}{n} \;=\; \frac{kn}{\ell n} + \frac{\ell m}{\ell n} \;=\; \frac{kn + \ell m}{\ell n}$$

Thus far, this is valid *only when* k, ℓ, m, n *are whole numbers.* But suppose k, ℓ, m, n are fractions?

First, we have to make sense of, for example, $\frac{k}{\ell}$ when k and ℓ are fractions. In this case, $\frac{k}{\ell}$ is a division of fractions and is called a **complex fraction**. Every complex fraction can be expressed as an ordinary fraction by use of invert-and-multiply. We have no time to discuss the **division of fractions** here, but there is a fairly long discussion of this in the September 2009 issue of the AFT house journal **American Educator**.

The fact that the formula for adding (ordinary) fractions is still valid even when the fractions are complex fractions is not difficult to prove. You can do it either by brute force (just invert and multiply all the way through; very tedious), or by abstract reasoning (short). The proof notwithstanding, this fact about complex fractions has to be taught explicitly, but at the moment it is not.

Here is one small reason why this should be taught: would you like to do the following computation as is,

$$\frac{1.5}{0.028} + \frac{42}{1.03} = \frac{(1.5 \times 1.03) + (42 \times 0.028)}{0.028 \times 1.03},$$

or would you prefer to change all the complex fractions to ordinary fractions before adding?

The validity of the formula for adding fractions,

$$\frac{k}{\ell} + \frac{m}{n} \;=\; \frac{kn}{\ell n} + \frac{\ell m}{\ell n} \;=\; \frac{kn + \ell m}{\ell n},$$

when k, ℓ, m, n are themselves *fractions* is important for a different reason. Wouldn't you like to do the following as is?

$$\frac{\sqrt{3}}{4} + \frac{1.2}{\sqrt{5}} \;=\; \frac{(\sqrt{3}\sqrt{5}) + (4 \times 1.2)}{4 \times \sqrt{5}}$$

What is at stake here is the validity of the formula, not only for all fractions, but also for (positive) irrational numbers such as $\sqrt{3}$ or $\sqrt{5}$. It turns out that the latter depends on the former. This highlights the importance of the formula for complex fractions.

What needs to be explicitly addressed in the school curriculum is a clear discussion of the following **Fundamental Assumption of School Mathematics (FASM)**:

If an identity between numbers holds for all fractions, then it holds for all (real) numbers ≥ 0.

FASM points out

why fractions are important in school mathematics,

why the arithmetic operations for complex fractions should be taught, and

the intrinsic limitation of school mathematics (we only teach fractions but not irrational numbers).

To summarize: When fractions are taught correctly, students

learn to use the symbolic notation naturally *(Theorem on equivalent fractions; CMA; formula for addition)*

learn abstract reasoning *(concept of fraction as point on number line; Theorem on equivalent fractions; CMA; formula for addition)*

learn the concept of generality *(Theorem on equivalent fractions; CMA; formula for addition)*

learn the importance of precision *(definition of fraction; addition of complex fractions; FASM)*

They get a good introduction to algebra.

3b. *Multiplication of rational numbers*

We want to explain why, if x, y are fractions, $(-x)(-y) = x\,y$.

We first prove the special case where x and y are whole numbers. The critical fact in this context is the following:

Theorem 1 $(-1)(-1) \;=\; 1$

Unlike other equalities in arithmetic, this equality will not be obtained by performing a straightforward computation. A little thinking is necessary, and it is the required *thinking* that throws students off.

We state our basic premise: we assume that for all rational numbers x,

(M1) Addition and multiplication of rational numbers satisfy the commutative, associative, and distributive laws.

(M2) $0 \cdot x \; = \; 0$.

(M3) $1 \cdot x \; = \; x$.

Consider the general question: How to show that a number b is equal to 1? One way is to show that $b + (-1) = 0$. Indeed, if this is true, i.e., if we know $b + (-1) = 0$, then we can add 1 to both sides to get $b + (-1) + 1 \; = \; 0 + 1$, from which $b = 1$ follows.

We now **prove Theorem 1.** Let $b = (-1)(-1)$ and we have to show $b = 1$. As remarked above, it suffices to prove that $b + (-1) = 0$. To this end, we compute:

$$\begin{aligned}
b + (-1) &= (-1)(-1) + 1 \cdot (-1) \qquad \text{(use (M3))} \\
&= (\,(-1) + 1\,)\,(-1) \qquad \text{(use dist. law)} \\
&= 0 \cdot (-1) \\
&= 0 \qquad \text{(use (M2))}
\end{aligned}$$

We are done.

It will be seen that the distributive law is the crucial ingredient that accounts for $(-x)(-y) = x\,y$.

Let us now prove that $(-3)(-4) = 3 \cdot 4 = 12$. We first prove:

$$(-1)(-4) \;=\; 4$$

This is because we can use the distributive law to expand, as follows:

$$\begin{aligned}
(-1)(-4) &= (-1)\left((-1)+(-1)+(-1)+(-1)\right) \\
&= (-1)(-1)+(-1)(-1)+(-1)(-1)+(-1)(-1) \\
&= 1+1+1+1 \qquad \text{(Theorem 1)} \\
&= 4
\end{aligned}$$

Therefore,

$$\begin{aligned}
(-3)(-4) &= \left((-1)+(-1)+(-1)\right)(-4) \\
&= (-1)(-4)+(-1)(-4)+(-1)(-4) \qquad \text{(dist. law)} \\
&= 4+4+4 \;=\; 3\cdot 4 \;=\; 12
\end{aligned}$$

We now prove in general that if m and n are whole numbers, then

$$(-m)(-n) = mn$$

As before, we first prove

$$(-1)(-n) = n$$

This is because

$$\begin{aligned}
(-1)(-n) &= (-1)(\underbrace{(-1)+\cdots+(-1)}_{n}) \\
&= \underbrace{(-1)(-1)+\cdots+(-1)(-1)}_{n} \qquad \text{(dist. law)} \\
&= \underbrace{1+\cdots+1}_{n} \qquad \text{(Theorem 1)} \\
&= n
\end{aligned}$$

Therefore,

$$
\begin{aligned}
(-m)(-n) &= (\underbrace{(-1)+\cdots+(-1)}_{m})\,(-n) \\
&= \underbrace{(-1)(-n)+\cdots+(-1)(-n)}_{m} \qquad \text{(dist. law)} \\
&= \underbrace{n+\cdots+n}_{m} \\
&= mn
\end{aligned}
$$

We now explain in general why:

$$\text{If } x,\ y \text{ are fractions, } \quad (-x)(-y) = x\,y\,.$$

We will need the analog of Theorem 1:

Theorem 2 *If y is a fraction, then* $(-1)(-y) = y$.

For the **proof**, observe that one way to prove a number b is equal to y is to prove that $b + (-y) = 0$. Because if this is true, then adding y to both sides gives $b + (-y) + y = 0 + y$, which then gives $b = y$. Now let $b = (-1)(-y)$. Then once again we use the distributive law:

$$b + (-y) = (-1)(-y) + 1 \cdot (-y) = ((-1) + 1)(-y) = 0 \cdot (-y) = 0$$

This completes the proof of Theorem 2.

So now, let x, y be any fractions. We will show $(-x)(-y) = x\,y$. By Theorem 2,

$$(-x)(-y) = (-1)x(-1)y = (-1)(-1)xy$$

By Theorem 1, $(-1)(-1) = 1$. Therefore

$$(-x)(-y) = (-1)(-1)xy = xy$$

We are done.

What we learned: The extensive use of the associative and commutative laws, and especially the distributive law, is a good introduction to algebra. The reasoning is abstract and general, as the identity $(-x)(-y) = xy$ is about *all* fractions.

References

College and Career Readiness Standards for Mathematics (released September 17, 2009) http://www.corestandards.org/Standards/index.htm

Foundations for Success: Final Report, The Mathematics Advisory Panel, U.S. Department of Education, Washington DC, 2008.
http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Report of the Task Group on Conceptual Knowledge and Skills, Chapter 3 in *Foundations for Success: Reports of the Task Groups and Sub-Committees*, The Mathematics Advisory Panel, U.S. Department of Education, Washington DC, 2008.
http://www.ed.gov/about/bdscomm/list/mathpanel/reports.html

Report of the Task Group on Learning Processes, Chapter 4 in *Foundations for Success: Reports of the Task Groups and Sub-Committees*, The Mathematics Advisory Panel, U.S. Department of Education, Washington DC, 2008.
http://www.ed.gov/about/bdscomm/list/mathpanel/reports.html

W. Schmid and H. Wu, The major topics of school algebra, March 31, 2008. http://math.berkeley.edu/~wu/NMPalgebra7.pdf

D. Willingham, What is developmentally appropriate practice? American Educator, Summer 2008, No. 2, pp. 34-39. http://www.aft.org/pubs-reports/american_educator/issues/summer08/willingham.pdf

H. Wu, How to prepare students for algebra, *American Educator*, Summer 2001, Vol. 25, No. 2, pp. 10-17. http://www.aft.org/pubs-reports/american_educator/summer2001/index.html

H. Wu, What's sophisticated about elementary mathematics? American Educator, Fall 2009, Vol. 33, No. 3, pp. 4-14.
http://www.aft.org/pubs-reports/american_educator/issues/fall2009/wu.pdf