# A readability metric for computer-generated mathematics

*Scott MacHaffie, Robin McLeod, Bill Roberts, Philip Todd, Leigh Anderson*

## Table of contents

- Abstract
- Introduction
- Methodology
- Results
    - Factored versus expanded
    - Intermediate expressions
- Metric
- Conclusion
- Acknowledgements
- References

## Abstract

We performed an empirical study to determine some of the factors which make an algebraic expression, and also a system of algebraic expressions, difficult to understand. We created a metric of the complexity of an individual expression, and of a system of expressions, which has a high positive correlation with our data. We used this metric in a simplification function.

## Introduction

We developed a geometry system which automatically derives algebraic expressions for geometric quantities. These results were too cumbersome to use directly, so we tried simplifying these expressions using symbolic algebra systems. This was unsuccessful because either the systems were unable to simplify the results due to their size or the results were still too complex to be understood. The problem was in the software and not inherent in the examples under study, for several of the examples could be done quite easily with pencil and paper. Two issues came up in this work: the need for better simplification techniques and the need for a judicious use of intermediate variables in systems of expressions.

Our software maintained systems of algebraic expressions as a sequence of intermediate variables. The software was able to automatically perform a variety of different simplification steps. To drive the automatic simplification strategy, we wanted to find a metric for the human readability of a sequence of expressions. Our simplification algorithm could then perform a manipulation on the expressions, test the complexity of the outcome, and back out of the simplification if the complexity increased.

Mathematical expressions can be written in different forms and although the forms are mathematically equivalent, they are not necessarily equally understandable. For example, the equations:

$$scr_1 = x(x(x+3)+3)+1$$
$$scr_2 = x^3 + 3x^2 + 3x + 1$$
$$scr_3 = (x+1)^3$$

all represent the same mathematical expression but they have differing cognitive values. To some extent, the cognitive values are context dependent. The purpose of this study, however, is to identify context independent factors.

There is no literature on what makes an algebraic expression easy to understand. There is work on the readability of computer programs, such as Pennington [2] and Tenny [3], and there is work on simplification of math equations, such as Landau [1].

# Methodology

For our study we used college mathematics, science, and engineering students. We created a questionnaire consisting of three sections. Section 1 was designed to obtain a comparison of three different simplification methods. Section 2 was designed to investigate single expressions, comparing factored versus non-factored expressions and the amount of nesting present in an expression. Section 3 was designed to investigate using more but simpler intermediate expressions versus using fewer complex expressions.

We recruited 189 students from local colleges and universities to participate in our study. Of these students, 120 were engineering majors taking a variety of 2nd through 4th year courses. The other 69 were community college students taking an advanced algebra course.

# Results

The results of section 1 were inconclusive due to insufficient difference between the stimuli. The nesting results from section 2 were also inconclusive. The other results from sections 2 and 3 were more useful.

## Factored versus expanded

The results from section 2 showed a clear preference for a factored expression over an expanded one. The preference was so strong that few of our examples produced a preference for the expanded form.

$$(x+y)(x-y) \qquad x^2 - y^2$$
$$(x+y)(x+2y) \qquad (x^2 + 3xy + 2y^2)$$

The subjects preferred the expanded form (right column) for the first pair of equations, but even for something as simple as the second pair of equations, they preferred the factored form (left column).

### Intermediate expressions

A strong preference was discovered for expressing mathematics as a longer sequence of simpler intermediate variables and a simpler final expression, over a shorter sequence of more complex expressions. Only when intermediate expressions were overly trivial, e.g. $d = \frac{b}{c}$ was there a preference for a smaller number of more complex expressions.

# Metric

For a single expression, we evaluated 7 metrics against our human preference data. The metric with the strongest correlation was counting the number of terms but excluding exponents. Terms are defined as primitive components of the expression (numbers or variables) e.g. *12x* + *b* has 3 terms: *12*, *x*, and *b*.

For systems of expressions we evaluated 4 metrics for correlation with our human preference data. The one with the highest correlation was obtained by using the best metric for single expressions, applying this metric to each expression in the system in turn, taking the largest value and adding the number of intermediate expressions. For example,

$$a = x + 2y - 3 - z$$
$$b = y\sqrt{a^2 - x^2}$$
$$c = x^2 + y^2 + z^2$$
$$f(x,y,z) = \frac{a^2}{b} - \frac{c}{b}$$

The lines have metrics of 5, 3, 3, and 4, respectively. There are 3 intermediate expressions, so this system has a metric of 8.

$$f(x,y,z) = \frac{(x+2y-3-z)^2 - (x^2+y^2+z^2)}{y\sqrt{(2y-3-z)(2y+2x-z-3)}}$$

This expression is mathematically equivalent to the previous system of expressions. The metric for this expression is 19. There are no intermediate expressions, so the metric for the system is also 19. The first system with the metric of 8 was preferred by 71% of our subjects. 7% of the subjects thought the systems of expressions were about equally preferable. Only 23% of the subjects preferred the second expression, with no intermediate variables.

This metric, the worst individual expression metric plus the number of intermediate expressions, had a Pearson's correlation of 0.96 with our human data in terms of which expressions were preferred.

# Conclusion

In this empirical study we have lain the groundwork for the study of the important problem of the readability of computer-generated algebra. The initial problem, an interactive geometry and algebra system, forced us to develop a variety of new simplification techniques. We used the metric we developed as the heuristic to guide which simplifications to apply. The fact that our prototype system can simplify complex sequences of expressions in many cases indicates that the metrics we have used are valuable, even though our understanding of the human readability problem is far from complete. It is our hope that this initial study will prompt further research.

## Acknowledgments

This research was sponsored by the National Science Foundation, project #DMI-9460654.

## References

1. Landau, S. Simplification of Nested Radicals, *SIAM Journal of Computation*, 21, 1 (Feb. 1992), pp. 85-100.

2. Pennington, N. Comprehension Strategies in Programming. *Empirical Studies of Programmers second workshop*, G. Olson, S. Sheppard, E. Soloway, Eds., Ablex, 1987, pp. 114-131.

3. Tenny, T. Program Readability: Procedures Versus Comment, *IEEE Transactions on Software Engineering*, 14, 9 (Sept. 1988), pp. 1271-1279.