Assessing Algebraic Expectation

Lynda Ball and Kaye Stacey
University of Melbourne
<l.ball@edfac.unimelb.edu.au>
<k.stacey@unimelb.edu.au>

Robyn Pierce
University of Ballarat
<r.pierce@ballarat.edu.au>

An important component of the algebra needed to do mathematics with a computer algebra system (CAS) is symbol sense, analogous to ‘number sense’ in arithmetic with a four function calculator. This paper presents an assessment of a component of symbol sense: algebraic expectation. We describe the results of 169 Year 11 students including interesting individual items and correlations with other measures. Initial analysis shows the test is successful in monitoring both algebraic skills and students’ certainty in these skills, both of which affect success with using CAS.

Introduction

The work reported in this paper is motivated by the need to monitor students’ acquisition of basic algebra skills when computer algebra systems (CAS) are available for doing, teaching and learning mathematics. Their use is approved for Mathematical Methods (CAS), a new Year 11/12 subject for the Victorian Certificate of Education. A cohort of students, now in Year 11 at three schools, is expected to undertake Year 12 examinations in the new subject in 2002. The CAS calculators are graphics calculators that additionally have a symbolic manipulation facility, which can handle algebraic manipulations and equation solving, exact calculations (e.g. with surds and logs), differentiation and integration, matrix operations etc. Table 1 gives an example of the new capabilities. The subject is being monitored by the CAS-CAT project (http://www.edfac.unimelb.edu.au/DSME/CAS-CAT). Stacey, McCrae, Chick, Asp & Leigh-Lancaster (2000) provides further details.

Table 1
Sample Differences in Equation Solving with Graphics and CAS Calculators.

Equation	Graphics Calculator	CAS Calculator
$3e^{0.5t} = 4$	t = 0.58 can be located as intercept	$t = 2\,ln\frac{4}{3}$ or t = 0.58
$3e^{at} = b$	unable to solve this	$t = \frac{ln\frac{b}{3}}{a}$ or $t = \frac{ln\,b - ln\,3}{a}$

One of the most contentious aspects of the introduction of CAS into school mathematics concerns its relationship to the development of basic algebra skills. Our work on Mathematical Methods (CAS) specifically raises four questions.

Question 1. What basic algebra skills will people (students, professional users of mathematics etc) need if they can be assured of a level of access to CAS that is similar to the level of access that people now have to scientific calculators? A sensible answer to this question is critical for decisions about content. In the literature, Herget, Heugl, Kutzler and Lehman (2000) have been prominent in proposing answers to this question.

Question 2. What basic algebra skills are needed to use a CAS effectively? Students will need to have adequate skills from Year 10 or develop them early in Year 11.

Question 3. Will the basic algebra skills of students who learn mathematics with CAS fall below an acceptable level (however defined)? Students now consolidate algebraic skills in Years 11 and 12, but CAS use might interfere with this.

Question 4. Can CAS enhance the basic algebra skills of students, in ways that parallel the pedagogical possibilities of scientific or graphics calculators for number and graphing? The possibility of enhancing teaching and learning is a major attraction for teachers.

This paper outlines part of our response to these questions. Firstly, following Pierce (2000) we propose that an important component of the algebra needed when CAS is used (Questions 1 and 2 above) will be 'symbol sense', rather analogous to the 'number sense' that is needed for doing arithmetic with a four function calculator. A key skill is to recognise equivalent expressions quickly. Secondly, we report on a trial of a 'Quick Quiz' created to test this skill, reporting the overall results, interesting individual items and correlations with other measures and evaluating its success as a test instrument. We intend to use this test to monitor students' algebra skills, as required by Questions 3 and 4 above. Throughout this paper, we use the word 'skills' in a general sense including knowledge of fundamental concepts, the ability to apply knowledge and algebraic manipulation.

Algebraic Insight for Using CAS

Arcarvi's (1994) concept of 'symbol sense' can be used to describe the understanding of algebra required for working in partnership with technology. For solving problems using mathematics, students need manipulative skills, but they also need the ability to formulate problems in mathematical terms, plan problem solutions, monitor progress towards a solution and then interpret solutions. Symbol sense refers to this knowledge, beyond technical skills, which is related to algebra. Pierce (2000) and Pierce and Stacey (2001) have defined 'algebraic insight' as that part of symbol sense that is directly required to work with CAS in solving a problem already formulated algebraically. (It is not concerned with translating to and from the real world problem).

Pierce divides algebraic insight into two sections: first *algebraic expectation* for working entirely within the symbolic representation and second, the abilities needed for linking numerical and graphical representations with the symbolic. This paper is concerned only with algebraic expectation.

Pierce has analysed algebraic expectation and observes that important components relate to knowing basic conventions and properties of operations, and being able to identify structure and key features of algebraic objects. For example, entering expressions into a CAS requires identification of their structure, especially to make use of brackets (or other grouping devices on some calculators). For example, the expression in Table 1 may need to be entered as *ln(b/3)/a* or *(ln(b)–ln(3))/a*. Students need algebraic expectation to monitor the succession of expressions appearing on a CAS screen, making on-going rough checks for mathematical sense. The degree of a polynomial, for example, indicates the number of linear factors. It is also essential that they can quickly recognise equivalent forms of expressions, especially since CAS does not always present results in a conventional manner. All of these are components of algebraic expectation.

Developing the Quick Quiz

For monitoring students' algebraic skills in the CAS-CAT project, three instruments are being developed. The 24-item Quick Quiz, which is the subject of this paper, assesses algebraic expectation and is supplemented by a 6-item Constructed Response Test. In addition, a separate test (not reported here) of both multiple choice and constructed response items assesses the ability to link representations.

We noted above that CAS users need to recognise simple equivalences quickly, so that they are not derailed by the trivial but are alerted to the significant unexpected. The Quick Quiz tries to capture the essence of making quick, real time decisions by presenting items with a restricted time to respond. To do this, the test was administered using PowerPoint with a fixed time of 10 seconds for each question.

One of the common instances when students have to recognise equivalence of algebraic expressions is in checking their work from the back of a textbook. Very often a student's answer is not in the same form as the answer in the book and so the student needs to decide whether the answer is correct or not. This situation was used to supply a realistic context for the Quick Quiz. Students were told that they would be presented with a series of slides, each showing two expressions and that they should think of them as being a mythical student's answer to a problem and the textbook answer. Their task was to decide if the student was definitely wrong, probably wrong, probably right or definitely right. If they could not decide, they were asked not to guess but to choose the option 'no idea'. It was also explained that the intention was not to look for special cases, but to consider general expressions. For example, in item 1, x/y is indeed equal to y/x when $x = \pm y$, but not in general. To reduce confusion of right/wrong and correct/incorrect in reporting results, we will refer to items where the mythical student's answer matched the textbook answer as true items and the other items as false. We will abbreviate students' responses as *df* (definitely wrong/false), *pf* (probably wrong), *ni* (no idea), *pt* (probably right/true) or *dt* (definitely right).

Algebraic insight is intended to refer to skills required for real-time monitoring of CAS procedures by a problem solver to identify errors and to evaluate answers. Thus the students were given the options of 'definitely' and 'probably' because we expected that a measure of confidence in equivalence would be important in assessing algebraic insight. We assume that students who are very sure that false items are true would be less likely than those who are less sure to identify related errors. Similarly, students who are very sure of a true equivalence will be more likely to check it quickly and accurately and move on to subsequent steps in a solution than those who are less sure.

Items cannot be unambiguously allocated to Pierce's framework for algebraic insight. However, most items involve knowledge of basic conventions and properties of operations, at least 7 items involve being able to identify structure and at least 5 items involve use of key features of algebraic objects. Items are given in Table 2.

First Major Trial of the Quick Quiz

Methodology

In the three project schools, the Quick Quiz was administered to 169 students in nine Year 11[1] classes undertaking Mathematical Methods (the continuing non-CAS subject) and

[1] Some students are accelerated Year 10 students.

Mathematical Methods (CAS). This was done by the first and third authors who began by explaining the purpose of the test to the students. They explained that students would have to concentrate very hard as each slide would only be shown for ten seconds and that there would not be enough time to do working out but that they were expected to make a quick judgment based on their mathematical experience. The students did not use any technology in completing the tests. Two sample items were demonstrated before the timed sequence began. The Quick Quiz was completed in approximately 4 minutes and, after a short break, students completed the Constructed Response Test. Because students have had only a small exposure to CAS at this stage, no distinction is made here between the two mathematics subjects. In the future, we will look for group differences in the post-testing.

Results by Item

Table 2 summarises the results for each item, ordered according to the percentage of students who answered the question correctly: the item facility. An item was judged to have been answered correctly if the response was *dt* or *pt* for a true item and *df* or *pf* for a false item. So, for example, Item 6 (the comparison of $5m$ with m^5) is false, and 92% of students answered correctly *df* or *pf*. Table 2 also shows two measures that incorporate students' certainty about their answers. First, scores were weighted by certainty when the responses were scored as *dt* (+2), *pt* (+1), *ni* (0), *pf* (-1) or *df* (-2) for true items and the negative of these for false items. Equally definite correct and incorrect responses cancel out on this measure (called the weighted score), but they accumulate with the certainty index. This is defined as the percentage of students who answer *dt* or *df* to an item. The number of characters in an item is a count of how many characters appeared for each item. For example, in item 24, $\sqrt{xy}$ is counted as three characters and $\sqrt{x+y}$ as four, giving a total of seven characters. The last column is [100 × no. of *dt* /(no. of *dt* or *pt*)] for a true item (similar for false items).

Table 2 shows that the items covered a wide range of difficulty. The correlation between the two measures of item difficulty (item facility and average weighted score by item) is 0.99, so item facility is used subsequently. True and false items were similarly difficult overall (average item facility 47% for false and 45% for true items). There was a moderate tendency for short items to be easier than long items (correlation of –0.41 between item facility and number of characters). Some of the short items were very well done (e.g. items 1, 6, 24) but other short items (e.g. 20 and 4) highlighted fundamental problems. Table 2 also shows that the items varied markedly on the certainty index. For item 1, for example, 80% of students answered *dt* or *df*, whereas for item 19, less than 30% did.

Figure 1a shows that certainty and item facility measure different properties of the items (correlation 0.56). The scattergram shows that items with the highest facility have high certainty index (items 6, 22, 24). Amongst the items with lower item facility, the success seems related to the (low) certainty index for true items, but there is a group of false items with very low success and high certainty (especially items 4 and 11). We conclude that students are unlikely to identify an error of this nature in their work.

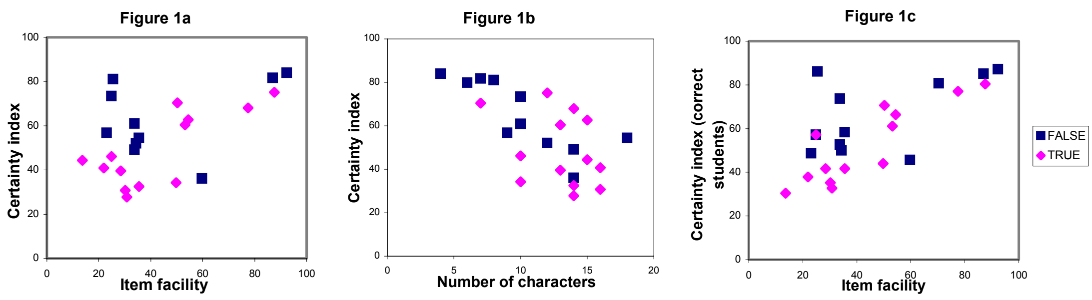
Table 2

Summary by Item of Responses to the Quick Quiz.

Item no.	Item	True or False	Item facility	Av. wtd item score	Cert. index	No. of chars.	% of correct who are certain
6	$5m \quad m^5$	F	92	1.7	84	4	87.2
22	$-2y+6 \quad -2(y-3)$	T	88	1.4	75	12	80.4
24	$\sqrt{xy} \quad \sqrt{x+y}$	F	87	1.4	82	7	85.0
13	$2f-g+3f-g \quad 5f-2g$	T	78	1.2	68	14	77.1
1	$x \div y \quad y \div x$	F	70	0.8	80	6	80.7
14	$sin(2x) \quad (sin 2)x$	F	60	0.5	36	14	45.5
23	$(b+a)^2 \quad a^2+b^2+2ab$	T	54	0.3	63	15	66.3
2	$n^2(n^3+1) \quad n^5+n^2$	T	53	0.2	60	13	61.1
5	$\frac{12x}{6x} \quad 2$	T	50	0.1	70	7	70.6
8	$h=x^2 \quad x=\pm\sqrt{h}$	T	50	0.3	34	10	44.0
17	$6+(4a+2b)\div 2 \quad 6+2a+b$	T	36	–0.1	33	17	41.7
15	$2x^2-y^2 \quad (2x-y)(2x+y)$	F	36	–0.3	54	18	58.3
10	$\frac{1}{3}+\frac{1}{y} \quad \frac{2}{3+y}$	F	34	–0.4	52	12	50.0
3	$\frac{2+x}{y+2} \quad \frac{x}{y}$	F	34	–0.4	61	10	73.7
9	$\frac{a-b}{d-c} \quad \frac{a}{d}-\frac{b}{c}$	F	34	–0.4	49	14	52.6
19	$2x+3=y \quad x=\frac{3-y}{-2}$	T	31	–0.2	28	14	32.7
18	$(x+y)\div 2\times x \quad \frac{x^2+xy}{2}$	T	30	–0.3	31	16	35.3
7	$\sqrt{16x-4y} \quad 2\sqrt{4x-y}$	T	28	–0.4	40	13	41.7
4	$\frac{12x}{6x} \quad 2x$	F	25	–0.8	81	8	86.0
16	$\frac{4+b}{4} \quad 1+\frac{b}{4}$	T	25	–0.6	46	10	57.1
11	$a+p\div q \quad \frac{a+p}{q}$	F	25	–0.9	73	10	57.1
12	$\frac{s}{t}+\frac{p}{t} \quad \frac{-(s+p)}{-t}$	T	22	–0.7	41	16	37.8
20	$\sqrt{2}+\sqrt{y} \quad \sqrt{2+y}$	F	23	–0.8	57	9	48.7
21	$\frac{2a^3+5}{a^2} \quad 2a+\frac{5}{a^2}$	T	14	–1	44	15	30.4

A correlation of –0.66 shows a strong relationship between certainty index and number of characters in Figure 1b, stronger than the relationship of item facility to number of characters (correlation –0.41). The graph shows that students tend to be very certain about their response (whether correct or incorrect) for short items. After about ten characters, the number of characters predicts the certainty index less well. It would appear that when an item contains a small number of characters the students decide immediately and confidently about whether the item is true or false. This immediacy of response may be the response that is necessary to give a true indication of student algebraic insight. Short items can expose both student algebraic strengths and weaknesses.

Figure 1c indicates the certainty of those students who responded correctly to an item. To interpret this graph, consider item facility as an indicator of item difficulty. In general the certainty index (correct students) follows the general certainty index, although it is a little higher. Only for the hardest items are the students who are correct less certain than all students (see Table 2), perhaps indicating well-placed caution. Like the whole sample, these students are more certain of the difficult false items than the difficult true items.



r = 0.56 **r = –0.66** **r = 0.45**

Figure 1. Scattergrams of certainty index, item facility, number of characters and certainty index (correct students) showing true and false items (N = 24 items).

Relationship to Results of the Constructed Response Test

The Quick Quiz is essentially a test of recognition, which is important when using CAS, but it is not sufficient for expert use. Just as a student using a four-function calculator needs to be able to do simple arithmetic unaided, a student using CAS must be able to carry out simple algebraic operations. We therefore also designed a simple Constructed Response Test, to supplement the recognition test of the Quick Quiz. The items involved index laws, substitution and solving a linear equation (all done well with average scores over 4 out of 5) simple factorisation and quadratic equation solving (average scores over 2 out of 5) and transposition of formulae (very poorly done). The Quick Quiz and the Constructed Response Test had a correlation of 0.49, indicating that the Quick Quiz cannot be used as the sole measure of algebraic skills. Figure 2a shows that the best students did very well on both tests, but for others there was little relationship: moderate and poor results on one test corresponded to both moderate and poor results on the other. The next step in our investigation will be to try to locate evidence of poor algebraic insight in the students' constructed responses and other work.

The Correctness of Relatively Certain and Uncertain Students

For the results by student the correlation between the total certainty of their responses (scored from 0 to 48) and the % correct for the Quick Quiz was low (r = 0.34). Figure 2b shows that for students who were most certain about their responses, the percentage correct ranged from 25% to 85%, showing that amongst the most certain, there were both very good and very poor students. Uncertain students had a smaller range for their percentage of correct responses ranging from approximately 25% correct to 60% correct. In summary, relatively uncertain students scored moderately or poorly, whereas there was high variability of scores amongst the relatively certain students. Teachers need to help some students increase their confidence in their work, whilst others need to become more cautious. In subsequent testing we will find out who learns the most.

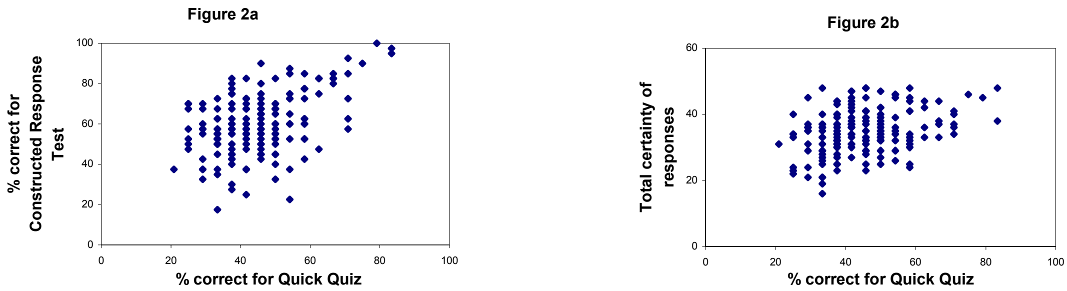


Figure 2. Scattergrams of % correct for Constructed Response Test, % correct for Quick Quiz and total certainty of responses (N = 169 students).

Discussion of Individual Items

In this section we compare results for a few related items. Items 4 and 5 both asked about $12x/6x$, but the 'textbook answers' were different. Very few students answered Item 4 correctly, although the certainty index was very high. The error (that $12x/6x = 2x$) has been reported in the literature and probably results from students wanting to preserve the x as a record of the 'units' of the question. In a previous Melbourne research study, one student who had found the sides of a triangle by solving the equation $2x + x + 14 = 44$ (so $x = 10$) reported the side length as $10x$ cm because, as she said, 'It's an exxy type of centimetres'. (Many of the other items also tap into misconceptions well established in the literature. Space precludes full reporting.) There was an audible response from the students when item 5 appeared and the consequent 'second thoughts' that the repetition stimulated may have been responsible for the lower certainty index for item 5. The correctness of response still only reached 50%. One wonders how students who are unsure of basic equivalences such as this can operate in any algebraic environment, with or without CAS. This type of item really seems to test algebraic insight as no real working is required and students instantly respond. The high certainty would seem to confirm a quick response by students. In refining an algebraic insight test it may be preferable to include items of this nature where it would be anticipated that students (nearly) instantly respond to the item rather than having to work out an answer as in item 19, which we would not use again.

Most of the items used only material expected to be covered in the junior secondary school. In this way, a test of algebraic insight has to take into account the level of the students: a test for other students may have quite different items. Notational items 6 and 24 were well done as was linear factoring (22) and collecting like terms (13). Simple reorganisations such as the slightly unusual presentation of $(a+b)^2$ in item 23 were unexpectedly difficult. All fraction items involving addition and subtraction (10, 3, 9, 16, 11, 12, 21) had very low facility (13% to 35%, average 28%), but the relatively high certainty indices (average 47) show that many students were unaware of their errors. One of the key factors in being able to give the correct response for these items is the ability to move easily between different equivalent expressions. When CAS is available, this is going to be an essential skill for students, as the CAS calculator will often give fraction answers in a 'non-standard' form or in a different form to that which would be obtained using a 'by-hand' method.

Conclusion and Future Work

This paper has reported the results of a large sample of students on a test designed to assess algebraic expectation. The scenario for the test [comparing a mythical student's answer with a textbook answer] worked well, by providing a realistic context for the items. Asking students to respond with a measure of certainty in their answers was also useful, but not for measuring the facility of items. Instead, the certainty index demonstrated a different feature of the items and enabled us to identify algebraic manipulations that will trap many students (e.g. that $12x/6x = 2x$, the fraction equivalences). As we monitor algebraic skills over the two years of VCE, we hope to see improvement in student scores, better-placed certainty and an overall increase in certainty. Many students are uncertain about basic algebra. We hope that they will develop confidence and be given support through using CAS, rather than being tentative CAS users. The analysis of the results by the crude measurement of characters has lead us to propose that the Quick Quiz would work as well or better if we used items requiring only 'recognition' and no more than one manipulation step. The final criterion for the usefulness of our measure of algebraic expectation is a future investigation of whether we see evidence of students' good or poor algebraic expectation in their class work and other tests.

Acknowledgement

The authors acknowledge financial support of the Australian Research Council, the Victorian Curriculum and Assessment Authority, Hewlett-Packard Australia Ltd, Shriro Australia (Casio) and Texas Instruments Australia Ltd, the scholarly input of Lewis Berenson (Holon Academic Institute of Technology, Israel) and our CAS-CAT colleagues Gary Asp, Peter Flynn, David Leigh-Lancaster, Margaret Kendal and Barry McCrae and the teachers and students in the project schools.

References

Arcavi, A. (1994). Symbol sense: Informal sense-making in formal mathematics. *For the Learning of Mathematics, 14* (3), 24-35.

Herget, W., Heugl, H., Kutzler, B., & Lehmann, E. (2000). Indispensable manual calculation skills in the CAS environment. *Micromath,* Autumn 2000, 8-17.

Pierce, R. (2000, July). Algebraic insight and students' use of DERIVE. *Computer Algebra in Mathematics Education* (Proceedings of the 4th international DERIVE and TI89/92 conference). CD-ROM. Liverpool, UK.

Pierce, R. & Stacey, K. (2001). *A framework for algebraic insight*. Manuscript submitted for publication.

Stacey, K., McCrae, B., Chick, H., Asp, G. & Leigh-Lancaster, D. (2000). Research-led policy change for technologically-active senior mathematics assessment. In J. Bana and A. Chapman (Eds.), *Mathematics Education Beyond 2000* (Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia, pp. 572-579). Fremantle: MERGA.