STATISTICAL IMPLICATIVE ANALYSIS OF STUDENTS'ALGEBRA PERFORMANCE

Reinhard Oldenburg, Augsburg University, Germany

Kaye Stacey, University of Melbourne, Australia

Abstract

The data mining method of Statistical Implicative Analysis is used to reveal details of the inner structure of algebraic competence as seen in data collected from beginning algebra students as part of the development of the SMART online diagnostic tests. The implicative analysis of the data shows the logical relation between algebraic sub-competencies. This paper reports item level analysis, comparing different cut-off levels for implication intensity, and a follow-up analysis of scales constructed from the clusters revealed. Finally in a novel application of statistical implicative analysis, the paper identifies pairs of scales that together imply an outcome, but do not do so separately. This allows the identification of abilities that are necessary foundations for other parts of algebraic proficiency.

https://doi.org/10.32388/FL8ENG

INTRODUCTION

The learning of symbolic algebra is a complex endeavor, and students tend to struggle with many concepts. Research reports have highlighted many difficulties, including problems with the equal sign (Kieran 1981), understanding the conventions of use of algebraic letters (Küchemann 1979), writing equations to solve problems (Stacey and MacGregor 1999), algebraic expressions (Graham and Thomas 2000) and developing structure sense (Arcavi 1994). Less is known about the structural relations between components of algebraic thinking (Oldenburg 2009, 2010). What concepts and skills support the understanding of other concepts and skills? Such knowledge could prove to be useful for the design of teaching sequences. The present paper uses data gathered from items administered online and a statistical data mining method to draw some conclusions. The data mining method used is that of statistical implicative analysis (SIA). This method searches for relations of implication in the data set. Such a statistical implication A→B says that high values on A are an indicator that B will have high values, but not necessarily that high values of B indicate high values of A. A common instance is when knowledge tested for B (e.g., counting to 10) is pre-requisite for knowledge tested for A (e.g., counting from 50 to 60). This contrasts with correlation and all analysis and hypothesis testing methods that build on it because correlation is a symmetric measure and thus detects only similarity but no directional information. Moreover, it is able to deal with complex structures that cannot be modeled by the linear hierarchies used in Rasch models (Stacey & Steinle 2006).

METHOD: STATISTICAL IMPLICATIVE ANALYSIS

Statistical methods based on the correlation matrix can reveal important insights into linear structures between observed or even latent variables. However, they are not adequate when one is interested in detecting the direction of an influence because the correlation matrix is itself symmetric. Therefore, several methods from data mining that are capable of giving information about the direction of a relationship may interest researchers in educational fields. Such methods do not look for correlations or similarities within data but for implications.

Two of these methods that have attracted some interest are Inductive Item Tree Analysis and Statistical Implicative Analysis. The former method is based on knowledge space theory (Doignon & Falmagne 1985), and algorithms to carry it out have been developed and implemented by Schrepp (2003) and Ünlü & Sargin (2010). These algorithms try to find a transitively consistent graph of implications (i.e., a→b and b→ c guarantee that a → c) between items. Unfortunately, these approaches are only suited for data sets of very modest size, as the computational complexity is very high. Moreover, studies with simulated data have shown that the power in detecting implications is only slightly better than for the method of Statistical Implicative Analysis that will be described now.

Statistical Implicative Analysis (SIA)

The SIA method (Gras et al. 2008) was invented by Regis Gras and has already been widely applied in many fields, including educational research. The book by Gras et al. (2008) contains several chapters on mathematics education, including the chapter on algebra knowledge by Croset, Trgalova and Nicaud (2008). They used SIA to discover what properties of algebraic expressions predispose students to evoke particular rules (including false rules). Malisani and Spagnolo (2009) investigated how the kind of approach students take to open-ended text problems implies what kind of result they produce.

There is a reference implementation of the basic algorithm and various extensions in the software CHIC by Gras et al. (2008). Unfortunately, this software cannot be applied directly to the present data because it cannot handle missing data, i.e., items omitted by students. Therefore, the method has been re-implemented by the first author inside the statistical programming system R. The difficulty of missing data was able to be circumvented, because in our analysis we only considered implications between two variables and therefore, in analyzing the implication between two specific items we can restrict the data set to those students that gave answers to both items. This is certainly adequate if the omissions are due to the fact that the students did not consider an item at all, but if the students considered the item and left it blank because they didn't know what to answer, it would be better to count this as wrong answer. However, this was not possible in the present study, so we only use pairwise non-missing data.

There are two versions of SIA, the classical and the entropic version. We now describe the classic version which is simpler and then add some remarks on the entropic version. Let us assume we have n observed cases and determined for each case the values of two binary variables A and B (so the data we have consists of two n -dimensional vectors with entries 0 and 1) and we want to decide whether it is reasonable to conclude that A implies B, in the following sense: If in a case the value of A is 1, then the value of B is likely to be 1 as well. As a logical implication $A \rightarrow B$ is always true unless A is true (written $A = 1$) but B is false (written $B = 0$); a single case with $A = 1 \wedge B = 0$ (using the logical "and" sign $\wedge$) shows that the implication does not hold in the strict sense of logic. However, in a statistical approach, some counter examples may result from random noise and thus we compare the number of observed counterexamples $n_{A \wedge \bar{B}}$ with the number of counterexamples that one would expect if the variables were independent. As counterexamples to the implication $A \rightarrow B$ are only given by cases with $A = 1 \wedge B = 0$, this expected number is $\frac{n_A}{n} \cdot \frac{n_{\bar{B}}}{n} \cdot n = \frac{n_A \cdot n_{\bar{B}}}{n}$ where n_A is the number of cases with $A = 1$ and $n_{\bar{B}}$ is the number of cases with $B = 0$ (note the bar over B means logical not). In SIA one uses the following test quantity to make the comparison:

$$q(A, B) = \frac{n_{A \wedge \bar{B}} - \frac{n_A \cdot n_{\bar{B}}}{n}}{\sqrt{\frac{n_A \cdot n_{\bar{B}}}{n}}}$$

The implication intensity is then defined by the probability $\varphi(A, B)$ of getting $q(A, B)$ as large as the observed one (i.e., getting a smaller number of counter examples) if the variables A and B were independent:

$$\varphi(A, B) = \frac{1}{\sqrt{2\pi}} \cdot \int_{q(A,B)}^{\infty} \mathrm{e}^{-\frac{t^2}{2}} \cdot dt$$

$\varphi(A, B)$ is a probability and thus has values between 0 and 1 and values close to 1 indicate empirical support for the implication. The graphs that will be shown below are defined by having a directed edge from A to B (in other words, showing that A implies B) if $\varphi(A, B)$ is larger than $1 - \alpha$, where α is the cut-off value (e.g., 0.05).

Some comments on this measure of implication intensity are in order. As all calculations are done with pairwise data (first A and B, then B and C, then A and C), it may be the case that both $A \rightarrow B$ and $B \rightarrow C$ have φ above a certain cut-off value but $A \rightarrow C$ does not. This is a drawback of the method only if you suppose that there should be strict transitive implicative relations behind the scene, and counterexamples are assumed to be due only to measurement error. In this case, one would like to reveal the consistent logical theory, and this should have transitivity of implication. On the other hand, not all counter-examples may be due to measurement errors, i.e., the hypothesis A may capture enough to (statistically) guarantee B, but B may include some other features that are needed to ensure C, although A alone is not enough. Thus, a failure of transitivity indicates that A alone is too "small" to be a good indicator for C. Another way to put this: The counterexamples to $A \rightarrow B$ and $B \rightarrow C$ each may be few, but they may combine to weaken $A \rightarrow C$ considerably.

One may wonder why SIA works with the rather complicated expression q instead of just the rate of counter examples. The reason is that if A is very difficult (i.e. n_A small) and B is very easy (i.e., $n_{\bar{B}}$ small) it is almost impossible to have any counterexamples. In fact, from a purely logical point of view, a trivial item (i.e., that can be solved by everyone) is implied by all items and an impossible item that is not solved by anyone implies all other items. The denominator in the definition of q is designed in such a way that it compensates for this effect. In fact, these two special cases will result in zero implicative intensity, simply substituting the extreme special case $n_A = 0$ makes q undefined as it results in 0/0, but in this case, one can transform the expression to see that in the limit $n_A \rightarrow 0$ we have $q \rightarrow 0$.

A simple extension of the method is required to deal with data with interval scale variables as may result from partial credit ratings that are expressed as numbers in the interval [0,1], so that we have $A_i, B_i \in [0,1]$ for each of case i. In this case the expression for q has to be calculated by a slightly more complex formula:

$$q(A, B) = \frac{\sum_i A_i(1 - B_i) - \frac{n_A \cdot n_{\bar{B}}}{n}}{\sqrt{\frac{(n^2 {s_A}^2 + {n_A}^2) \cdot (n^2 {s_{\bar{B}}}^2 + {n_{\bar{B}}}^2)}{n^3}}}$$

Here, $n_A = \sum_i A_i$, $n_{\bar{B}} = \sum_i (1 - B_i)$, and s_A is the standard deviation of A while $s_{\bar{B}} = s_B$ is the standard deviation of B.

Especially for large data sets and for data from partial credits it has been shown that φ may not be the best measure, as it may overestimate some implications. Gras et al. (2008) present an improved version that measures entropy to judge the implication intensity. The calculations are somewhat more difficult, but the interpretation is similar. The fact that the actual values of φ can be markedly different when applying the two methods hints at a general problem of the method:

The actual values depend on factors such as the method and the sample size, and hence sensible comparisons should be made only between implications measured with the same method in the same data set.

Application of SIA to algebra items

In this paper, SIA is used to analyze logical implication between test items that are dichotomously coded as correct (score 1) or incorrect (score 0). SIA detects pairs of items A and B where having item A correct implies a high probability of having item B correct as well. This is written as $A \rightarrow B$. A useful interpretation appeals to the logically equivalent form $\bar{B} \rightarrow \bar{A}$ which indicates that having item B correct is (statistically) necessary for having item A correct. Educationally, such an implication may be an intrinsic necessity (e.g., a pre-requisite where the knowledge tested by item B is clearly part of the knowledge tested by item A) or the result of a more subtle influence of learning B on learning A (these seem the most interesting outcomes for educational research) or it may be simply an empirical observation perhaps explained by a third factor (e.g., all students who can do advanced calculus items can also do basic spelling, due to learning age). Generally, we do not expect both $A \rightarrow B$ and $B \rightarrow A$ unless the items are testing knowledge that is similar (parallel items) for the student group. For example, if item A is 'solve $2x - 7 = 9$' and item B is 'solve $3x + 4 = 19$' and item C is 'solve $2x - 7 = 3x + 4$', we expect $A \rightarrow B$ and $B \rightarrow A$ (at least to some extent, although empirical data shows that this is far from a perfect implication), and $C \rightarrow A$ and $C \rightarrow B$. For a population of students in the early stages of learning algebra we do not expect that $A \rightarrow C$ and $B \rightarrow C$ because it is known that solving equations with variables occurring more than once is conceptually different and needs to be learned separately (see the discussion of arithmetical vs. non-arithmetical equations in Filloy (2008, chapter 4)).

METHOD: TEST AND DATA

The 'SMART::tests' (abbreviation of "specific mathematics assessments that reveal thinking") have been developed by Kaye Stacey, Vicki Steinle, Beth Price and Eugene Gvozdenko at the University of Melbourne, They are online tests (www.smartvic.com) on a broad range of school mathematics topics and are intended to be used by teachers who assign their students to selected tests as formative assessment when appropriate in their teaching programs (Stacey et al. 2009). The tests address a wide range of mathematical conceptions, are delivered and assessed by computer, and report to teachers on individual student's stage of learning and common errors and misconceptions.

The data used for this study consists of the responses of 1634 students in Years 7, 8 and 9, from a varied group of secondary schools in Victoria (Australia) in the years 2008 and 2009. For the items reported in this paper, the number of responses per item varied from 305 to 693 (mean 475).

For the present study we focus on the 37 algebra-related items from a large set of items used as a first step in the development of the final tests. As part of the initial item calibration and evaluation process, students were presented with a large number of items (more than could be done in the time available) in different orders, so all items were encountered early in the test by some students. Some students chose which items to answer themselves, but some teachers selected the sections to be answered by their classes. All the items are in multiple choice format. Table 1 lists some of the items; the Appendix lists all items in the form presented to students along with item success rates (percent of responses correct). Due to the multiple-choice format, students are not expected to have any technical problems with entering their answers. However, they may leave items blank, and there will be a variety of reasons for this: accidental omission, deliberate omissions due to students' perception of item difficulty, running out of time or being instructed by their teacher to only do certain parts. The cluster description in Table 1 is derived from theoretical content analysis.

Table 1. Sample items.

Item	**Item task**	**Cluster description**
I506_1	Is $\frac{12\times 17}{6\times 17} = 2$? (true or false)	Arithmetic
I506_6	Is $3 \times (2 \times 5) = 6 \times 15$? (true or false)	Arithmetic
I506_10	Is $\frac{17000-17}{17} = 1000 - 17$? (true or false)	Arithmetic
I479_1	Is $\frac{33x}{11x} = 3$? (true or false)	Transformational algebra
I479_10	Is $8d - 3g = 3g - 8d$? (true or false)	Transformational algebra
I490	Sue weighs 1kg less than Chris. Chris weighs *y* kg. We write Sue's weight in algebra as: a) *y*-1; b) *x*; c) 1-*y*; d) 24; e) 1*y*; f) 90 (choose one answer)	Set up expressions representing quantities
I495	Sam is *s* cm shorter than Eva. Eva is 95cm tall. We can write Sam's height in algebra as: a) 95-*s*; b) 95s; c) *s*-95; d) *s*; e) 94; f) –*s*95 (choose one answer)	Set up expressions representing quantities
I497	David is 10cm taller than Con. Con is *h* cm tall. We write David's height in algebra as: a) 10+*h*; b) 10*h*; c) *r*; d) *C*=*D*+10; e) *h*10; f) 18 (choose one answer)	Set up expressions representing quantities
I494	At a bike shop, there are *b* bikes (2 wheels each) and *t* trikes (3 wheels each). Choose	Set up relational equations

	the equation that says that there are a total of 100 wheels. Choices: a) $b+t=100$; b) $2b+3t=100$; c) $35b+10t=100$	
I498_1	Some students had to find some values of x to make this equation true. $x+x+x=12$. Mark the work of each student as correct or incorrect: Mary wrote $x=2$, $x=5$ and $x=5$ (Mary is correct/ incorrect)	Variable conventions, obeying variable rules
I498_2	Like 498_1: Millie wrote $x=9$, $x=2$ and $x=1$ (Millie is correct/ incorrect)	Variable conventions, obeying variable rules
I498_3	Like 498_1: Mandy wrote $x=4$ (Mandy is correct/ incorrect)	Variable conventions, obeying variable rules
I575	p stands for an unknown number. Write in mathematical symbols: 'Multiply p by 6, then add 10 to the result.' a) $p\times 6=x+10$; b) $(p+6)^{10}$ c) $6p10$; d) $6p+10$; e) $p\times(6+10)$; f) $16p$; g) $p16$ (choose one answer)	Set up expressions with numbers (involving order of operations and implicit multiplication).
I581	Shortened formulation: Bicycle rent with fixed fee of $25 and a charge of $8 per hour. Which expression for the cost would the teacher prefer?	Set up functional equations

The 37 items for this study are all relevant to learning algebra. Items with the same initial item number (e.g., I506_1 and I506_2) were presented together (see Appendix) and tend to represent related knowledge. The items were based on research literature to test for stages of learning and known misconceptions. After initial experimentation using qualitative and quantitative methods, all items had been turned into multiple choice format. Diagnosis of misconceptions by the full SMART test system depends on actual responses not only the number of correct answers (Price et al. 2013).

The data preparation consisted of re-coding the entries of the students-items matrix to three values: missing, 0 (incorrect) and 1 (correct). No further pre-processing or hypothesis formation is necessary to apply the method of SIA.

RESULTS

Item level analysis

The first analysis that has been carried out is a plain calculation of implication intensities and representing them as a graph. Interestingly, the items related by implications turn out to be of similar types that can be easily interpreted. For a cut-off of α=0.005, i.e., displaying only implication arrows if $\phi \geq 1 - \alpha = 0.995$, the graph is shown in figure 1. There are 20 items shown in the graph. The remaining 37 – 20 = 17 items are not linked to others at this cut-off level. This means that implications between these variables may exist, but their implication intensity is less than 99.5% and hence in this presentation they are deemed to be unlinked.

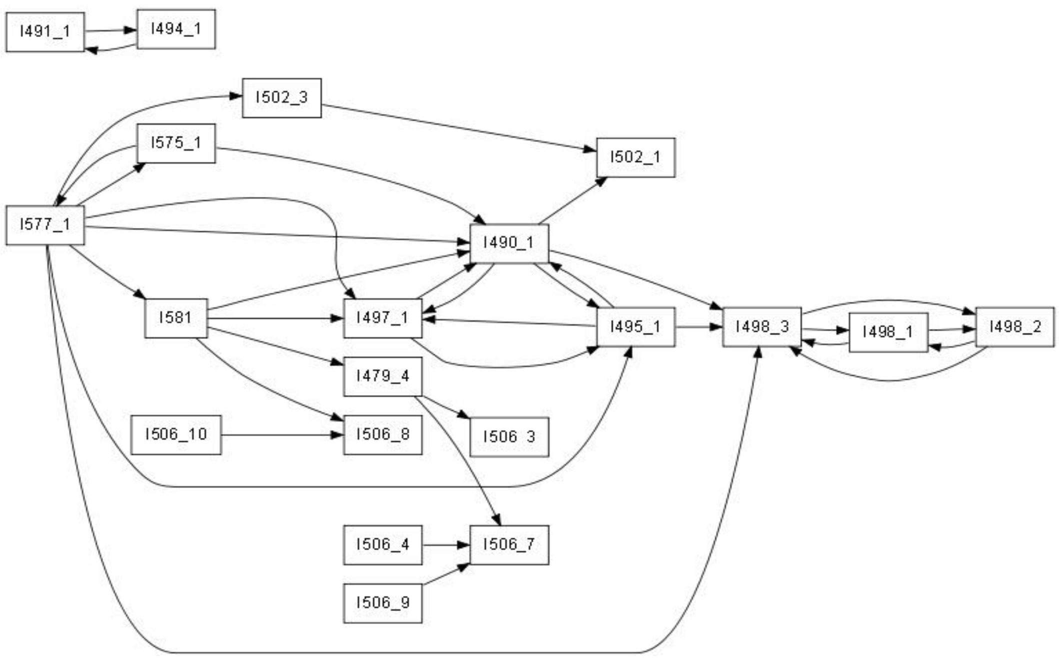


Figure 1: The implicative graph for cut-off of α=0.005

The interesting fact is that this structure, created from the data without any pre-formulated hypothesis, can be easily interpreted: Items that are joined by mutual arrows are – in the statistical sense of SIA – equivalent. For example, I490_1, I495_1 and I497_1 form such a cluster of almost equivalent items. In fact, these are all about setting up expressions representing quantities ($y - 1,\ 95 - s,\ h + 10$) and thus their equivalence is in line with expectations. There are differences in cognitive demand between these items (e.g., the success rates are 61%, 54%, 57% respectively) but students correct on one are very highly likely to be correct on the others. Next, the three parts of I498 are found by SIA to be (almost) equivalent and their content allows grouping them under the cluster of items testing variable conventions. Between these clusters there are two arrows going from the first to the second cluster, i.e., indicating that students who can successfully set up expressions will not violate variable conventions. Or, put the other way round: Obeying variable conventions seems to be a necessary condition for being successful in setting up expressions.

Another interesting feature is that there are arrows leading from I577 and I575 to nearly all other items. We speculate that this is because the algebraic notation required in these items is more complex as it involves two operations where the order is important and implicit multiplication needs to be understood. Hence, these items are difficult and mastering them is a good indicator for broader algebraic proficiency. The full interpretation of the graph by grouping most items into seven clusters based on links in the graphs is given in Figure 2.

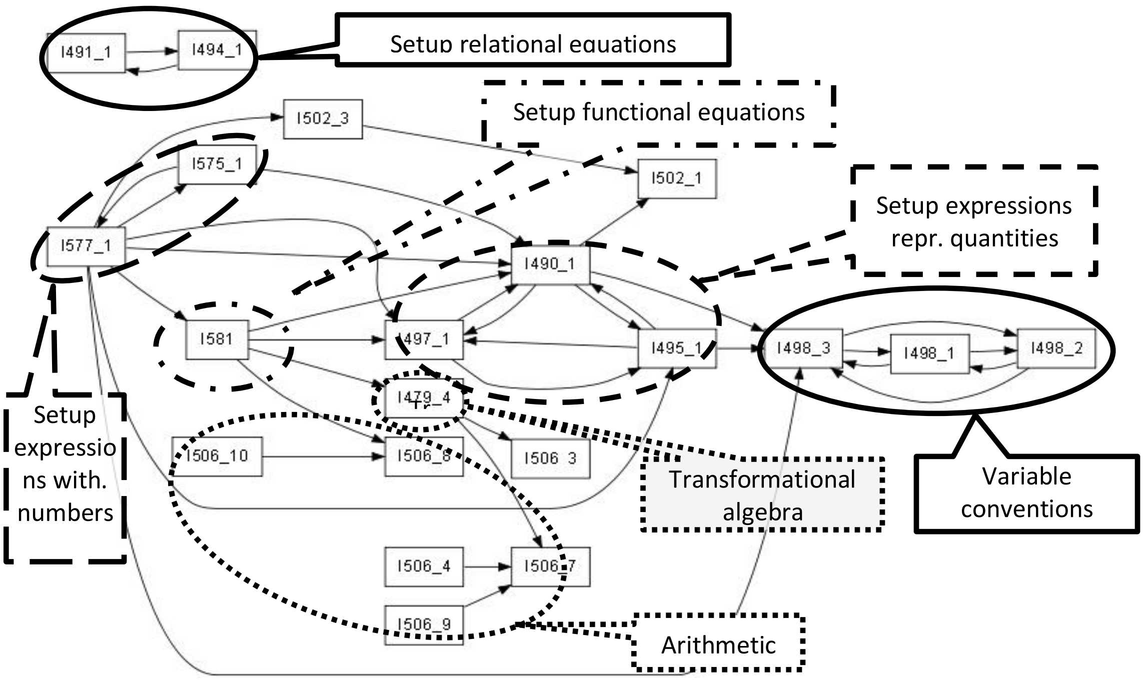


Figure 2: The graph from figure 1 with cluster interpretations added.

With more tolerance (i.e., a higher cut-off α), more arrows will be displayed, thus integrating items that are linked to others on weaker evidence. For α=0.01, new items and arrows show up, but the overall structure and interpretation remains the same (see Figure 3). For example, the cluster "arithmetic" grows but its relations to other clusters stays rather weak. In Fig. 3 "variable conventions I" are those related to the unique variable-value principle (in one calculation each symbol represents just one number), while variable conventions II relate to values for variables not being specified by naming convention. These two aspects of variable conventions appear to be rather independent of each other. But again, variable convention II is implied by others, but not vice versa, indicating that it is to some extent a necessary condition for success in algebra. Moreover, Fig. 3 reveals that some aspects of transformational algebra together form an island like setting up relational equations.

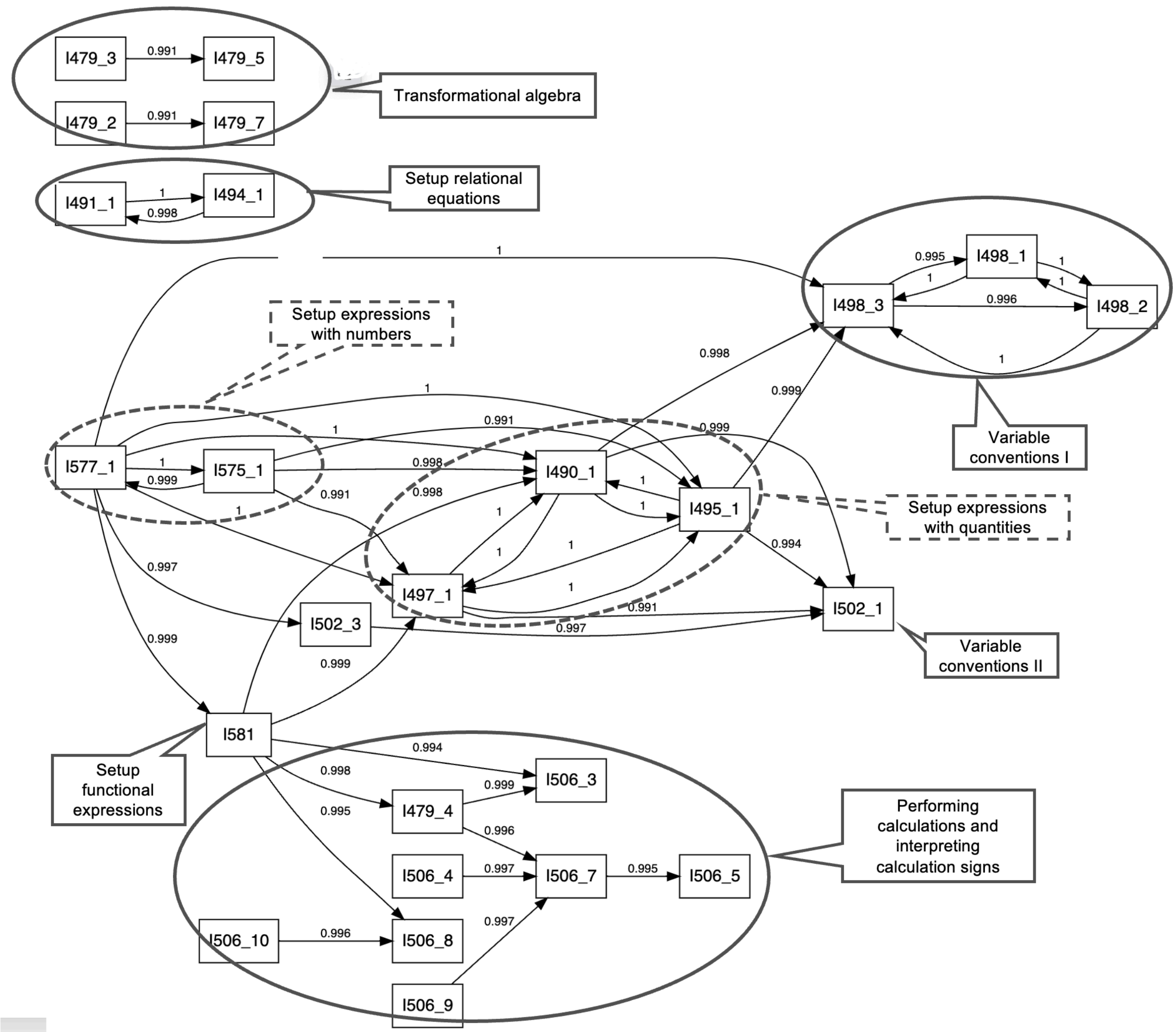


Figure 3: The implicative graph for cut-off of α=0.01 and showing implication intensities.

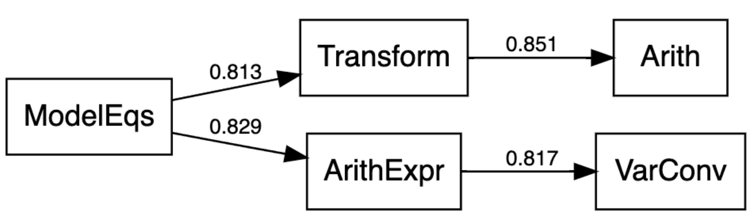
Analysis by scales

To explore this structure further we defined scales by averaging over non-missing responses to items that assess related content as confirmed by inspection and revealed by the former analysis (last section). The grouping of items into scales is informed by the results of the SIA analysis given above, but it differs slightly to reflect theoretical considerations on the items. The resulting scales are shown in Table 2. Note that Cronbach alpha values are rather low for several of the scales, but this is expected as items are not constructed according to the idea of a one-dimensional construct, but rather to explore students' thinking broadly. Moreover, subsequent analysis will not use classical multivariate analysis and will not draw conclusions from classical hypothesis tests using scale values. Hence, the low Cronbach alpha are not an issue of real concern here. There is one item, I493 (see Appendix), which did not fit well into any cluster or scale because it is the only item that presents an equation and then asks for the interpretation of the variables in it.

Table 2: Scales formed from the items

Scale name	**Description**	**Items**	**Cronbach alpha**
Arith	Arithmetical calculations	I506_1,…, , I506_10	0.61
Transform	Algebraic transformations	I479, 10 sub-items	0.36 (0.46 with _08 removed)
VarConv	Conventions of variables naming (e.g., unique value)	I502, 3 sub-items and I498, 3-sub-items	-0.08
ModelExprs	Use expressions to model real world situations	I497_1, I490_1, I495_1, I581	0.62
ArithExpr	Use expressions to express mathematical calculations	I575_1, I577_1	0.57
ModelEqs	Model situation with an equation in one variable	I579, I580, I491, I494	0.18
InterpretEq	Interpret the meaning of an equation	I493	-

Normalized sum scales for the scales in Table 2 have been calculated by averaging over the non-missing entries in the data set. Because of this procedure, the number of usable cases was larger than in the item-based approach above. Hence, the entropic version of SIA was used to calculate implication weights. The entropic version and the fact that partial credits have less variance than the dichotomous 0-1 coded items result in lower implication intensities and hence a cut-off α=0.2 was used to produce the graph at the top of Figure 4. This graph connects 5 of the 7 clusters. With a cut-off of α=0.25, the graph shown at the bottom of Figure 4 is produced. This shows implications between all 7 clusters.



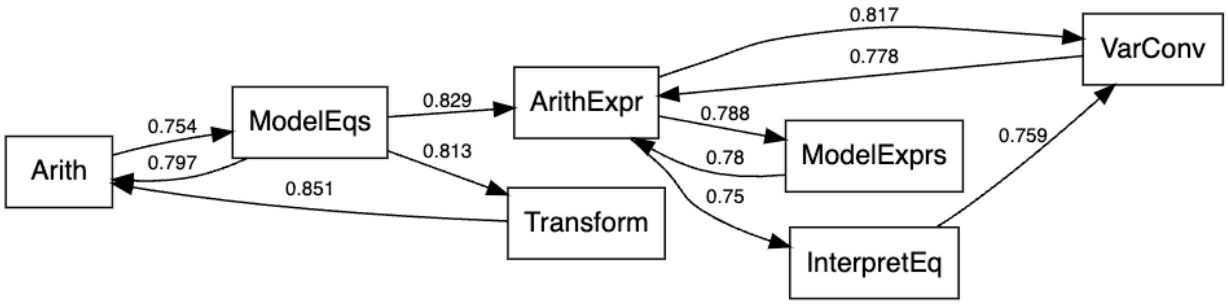


Figure 4: The entropic implicative graphs for scales with cut-off α=0.2 (above) and with α=0.25 (below)

The results largely confirm the conclusions drawn above. First, note that obeying variable conventions (VarConv) is implied only by ArithExpr at cutoff α=0.2, or turned around, without knowing the conventions of variables, fewer other things can be achieved (lowering the cut-off to 0.25 one sees also InterpretEq implies VarConv). The same holds true for Arith which is a necessary condition, especially for Transform (transformational algebra) which is in turn necessary for ModelEqs (modelling with equations). This is a part of algebra that requires extensive knowledge and skills, so success in it implies (in a statistical sense) that one will be successful in almost all other areas of algebra as well. ModelExprs (modeling with expressions) is required via ArithExpr (setting up arithmetical expressions) for ModelEqs (modeling with equations), which seems quite sensible.

Analysis by pairs of scales

The last kind of implicative analysis applied to the data was the search for pairs of scales that together imply good performance on a third scale. This kind of application of SIA has, to the best of our knowledge, never been used before. We are looking for implications of the form $A \wedge B \rightarrow C$, i.e., the presence of two indicators that together imply some consequence. As scales are encoded as values from the unit interval [0,1] the logical conjunction can simply be realized by multiplication: Given an interval scaled $A_i, B_i \in [0,1]$ one sets $(A \wedge B)_i := A_i \cdot B_i$. because in the dichotomic case the product is only 1 if both values are 1, and otherwise it is only close to 1 if both numbers are close to 1.

Note that, if $A \rightarrow C$ or $B \rightarrow C$, then in ordinary logic $A \wedge B \rightarrow C$ holds, so that it is not a surprise when the SIA also shows this relationship; it does not decrease entropy or increase knowledge. Thus, of greatest interest are the cases where both $A \rightarrow C$ and $B \rightarrow C$ have an implication intensity below the threshold, but $A \wedge B \rightarrow C$ is above. The following table lists the implications with intensity >0.85 in decreasing order, and marks those with a double asterisk **, where $A \rightarrow C$ and $B \rightarrow C$ both have implication weight below 0.8 and marks with a single asterisk those where both are below 0.85.

Table 3. List of *A∧B→C implications with intensity >0.85*

Implication $A \wedge B \rightarrow C$	Intensity	Implication $A \rightarrow C$	Intensity	Implication $B \rightarrow C$	Intensity	
ArithExpr ∧InterpretEq → ModelExprs	0.915	ArithExpr → ModelExprs	0.788	InterpretEq → ModelExprs	<0.75	**
Transform ∧ ModelExprs → Arith 0.882	0.882	Transform → Arith	0.851	ModelExprs → Arith	<0.75	
Transform ∧ ModelEqs → Arith	0.877	Transform → Arith	0.851	ModelEqs → Arith	0.797	
ArithExpr ∧ InterpretEq → VarConv	0.868	ArithExpr → VarConv	0.817	InterpretEq → VarConv	0.759	*
ArithExpr ∧ ModelExprs → VarConv	0.861	ArithExpr → VarConv	0.817	ModelExprs → VarConv	<0.75	*

The first row (with the double asterisk) shows that neither setting up arithmetical expressions nor the ability to interpret equations alone is sufficient to be good at modelling with expressions, but the combination is. This seems sensible as it shows that modelling with expressions involves their construction (ArithExpr) and interpretation (InterpretEq). The second and third row give information that is not new, because being good in transformational algebra already implies what is found here. The final two rows in Table 3 also make sense, as they add an element of algebra knowledge to the knowledge on arithmetical expressions.

DISCUSSION AND CONCLUSION

The results from this study reveal a rich structure of implications among students' ability to deal with algebraic tasks. The un-biased data mining method SIA carried out by the first author reveals clusters of items that are easily understood to be connected by the established understanding of the learning of algebra which guided the construction of the items by the second author and colleagues working on the SMART tests. Moreover, besides implications that are to be expected (e.g., that obeying variable conventions is a necessity for further insightful algebraic work), it shows some implications that are interesting – and also some missing implications that are worthy of note. Both present and absent implications may improve understanding of the learning of algebra. One such observation is that the modeling of situations with equations that define general relations (i.e., not just functional relations) between variables are an island that is not closely connected to the rest of the items (when analyzed at the item level). This use of algebra is, however, important in many modeling situations, and hence it should be given special attention in teaching. Moreover, setting up expressions that represent numbers (or encode number operations) implies many other dimensions and can thus be seen as an important ability to develop. Interesting missing implications are between canceling in number fractions (such as $\frac{12\times 17}{6\times 17}$) and in rational expressions (e.g., $\frac{33x}{11x}$). This may be because algebraic letters in calculations distract or dismay beginners so that they give up immediately. Alternatively, they may assume that results in algebra should contain a variable.

The results in this paper are to be treated with some caution. One reason is that it is not yet settled which implicative intensities should be considered to be sufficiently high to have importance. Another reason is that there are variations of the method, and it is not clear what version produces ’best‘ results. We have dealt with the issue of not having a broadly accepted ‘significance level’ by giving the graphs for different cutoffs, which gives an optical impression of the stronger and the weaker implications. Another reason is we have interpreted the observed implications by linking them back to known phenomena of learning algebra, and these could be questioned. Thus, one should investigate if similar relations occur in other test data on algebraic proficiency and investigate further how results fit with the theoretical understanding of the learning of algebra.

Acknowledgement

The data was collected by Kaye Stacey, Vicki Steinle, Beth Price and Eugene Gvozdenko of The University of Melbourne during the development of SMART::tests (www.smartvic.com), as part of the 2008 Grant Australian Research Council Linkage Grant *Supporting personalised learning in secondary schools through the use of specific mathematics assessments that reveal thinking* We thank the anonymous teachers and students from various schools in Melbourne for their participation.

References

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

Croset, M.-C., Trgalova, J., &Nicaud, J.-F. (2008). Student's Algebraic Knowledge Modelling: Algebraic Context as Cause of Student's Actions. In: Gras, R., Suzuki, E., Guillet, F., Spagnolo, F. (eds) *Statistical Implicative Analysis*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78983-3_4

Doignon, J.-P. &Falmagne, J.-Cl. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23:175–196.

Filloy, E., Puig, L., & Rojano, T. (2008). *Educational Algebra.* New York. Springer.

Graham, A.T. & M.O.J. Thomas. (2000). Building a versatile understanding of algebraic variables with a graphic calculator. *Educational Studies in Mathematics*, 41, 265-82.

Gras, R., Suzuki, E., Guillet, F., & Spagnolo (Eds.) (2008). *Statistical Implicative Analysis: Theory and Applications*. New York. Springer.

Küchemann, D. (1979). Children’s Understanding of Numerical Variables, *Mathematics in School, 7*, 23-26.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, 12(3), 317-326.

Malisani, E., & Spagnolo, F. (2009). From arithmetical thought to algebraic thought: The role of the "variable." *Educational Studies in Mathematics*, 71, 19–41.

Oldenburg, R. (2009). Structure of algebraic competencies. In V. Durand-Guerrier, S. SouryLavergne & F. Arzarello (Eds.). *Proceedings of CERME 6* (pp. 579-588). Lyon: Institut National de Recherche Pedagogique.

Oldenburg, R. (2010). A re-analysis of TIMSS data using Statistical implicative analysis. *Quaderni di Ricerca in Didattica (Mathematics),* n°20 suppl 1, 2010, 411-424.

Price, B., Stacey, K., Steinle, V., & Gvozdenko, E. (2013). SMART online assessments for teaching mathematics. *Mathematics Teaching*, 235(4), 10–15.

Schrepp, M. (2003). A Method for the Analysis of Hierarchical Dependencies Between Items of a Questionnaire. *Methods of Psychological Research Online*, 19, 43–79.

Stacey, K. & MacGregor. M. (1999). Learning the algebraic method of solving problems. *The Journal of Mathematical Behavior*, 18(2), 149-67.

Stacey, K. & Steinle, V. (2006). A case of the inapplicability of the Rasch Model: Mapping conceptual learning. *Mathematics Education Research Journal*, 18(2), 77 – 92.

Stacey, K., Price, B., Steinle, V., Chick, H., & Gvozdenko, E. (2009). SMART Assessment for Learning. *Proceedings of Annual Conference of the International Society for Design and Development in Education.*http://www.isdde.org/isdde/cairns/pdf/papers/isdde09_stacey.pdf

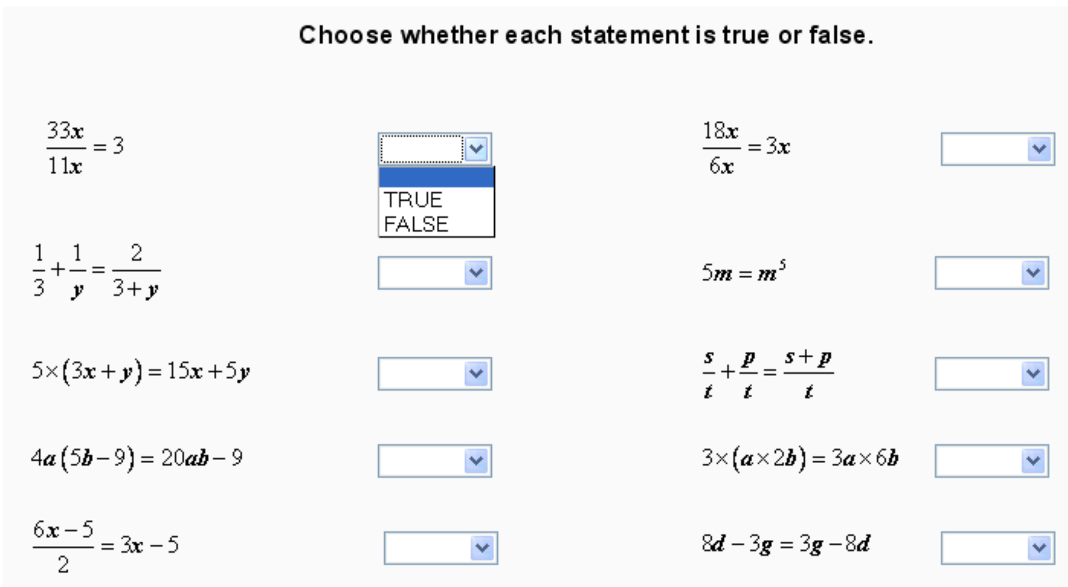
Ünlü, A., & Sargin, A. (2010). DAKS: An R package for data analysis methods in knowledge space theory. *Journal of Statistical Software*. 37(2), 1-31.

APPENDIX: SCREEN DUMPS AND SUCCESS RATES OF ITEMS

This Appendix lists the 37 items as presented online to the students (in then-current Moodle) together with their success rates. The order of items here is by item number, whereas students were presented with items in random sequences. The success rate is defined as the number of correct responses as a percent of the number of all responses (i.e., ignoring omissions).

Item 479: Success rates in the same tabular structure are:

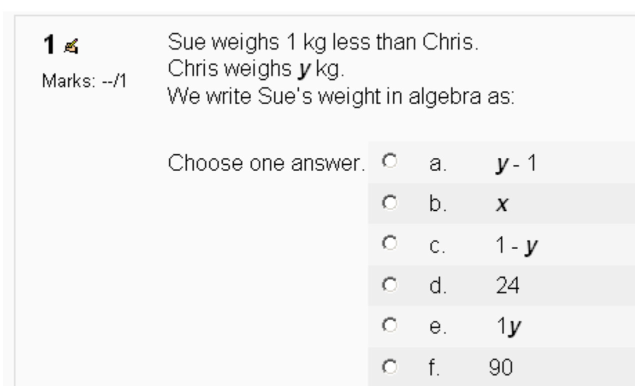
69%	42%
47%	60%
63%	69%
63%	37%
51%	73%



Choose whether each statement is true or false.

$\frac{33x}{11x}=3$	[ ] TRUE / FALSE	$\frac{18x}{6x}=3x$	[ ]
$\frac{1}{3}+\frac{1}{y}=\frac{2}{3+y}$	[ ]	$5m=m^5$	[ ]
$5\times(3x+y)=15x+5y$	[ ]	$\frac{s}{t}+\frac{p}{t}=\frac{s+p}{t}$	[ ]
$4a(5b-9)=20ab-9$	[ ]	$3\times(a\times 2b)=3a\times 6b$	[ ]
$\frac{6x-5}{2}=3x-5$	[ ]	$8d-3g=3g-8d$	[ ]

Item 490: Success rate 61%



1
Marks: --/1

Sue weighs 1 kg less than Chris.
Chris weighs ***y*** kg.
We write Sue's weight in algebra as:

Choose one answer.
- a. ***y*** - 1
- b. ***x***
- c. 1 - ***y***
- d. 24
- e. 1***y***
- f. 90

Item 491: Success rate 23%

1

Marks: --/1

For my garden, I bought $\boldsymbol{r}$ roses and $\boldsymbol{g}$ geraniums.
The roses cost $4 each.
The geraniums cost $5 each.
Choose the equation that says that the total cost was $70.

Choose one answer.

- a. $\boldsymbol{r} + \boldsymbol{g} = 70$
- b. $4\boldsymbol{r} + 5\boldsymbol{g} = 70$
- c. $10\boldsymbol{r} + 6\boldsymbol{g} = 70$

Item 493: Success rate 18%

1

Marks: --/1

Lucy bought 6 doughnuts for 12 dollars.
She wanted to work out how much each doughnut cost.
She wrote the equation $6\boldsymbol{d} = 12$.
In Lucy's equation, $\boldsymbol{d}$ stands for:

Choose one answer.

- a. the number of doughnuts
- b. doughnuts
- c. the cost of one doughnut
- d. one doughnut
- e. dollars

Item 494: Success rate 31%

1

Marks: --/1

At a bike shop there are $\boldsymbol{b}$ bikes (2 wheels each) and $\boldsymbol{t}$ trikes (3 wheels each).
Choose the equation that says that there are a total of 100 wheels.

Choose one answer.

- a. $\boldsymbol{b} + \boldsymbol{t} = 100$
- b. $2\boldsymbol{b} + 3\boldsymbol{t} = 100$
- c. $35\boldsymbol{b} + 10\boldsymbol{t} = 100$

Item 495: Success rate 54%

1
Marks: --/1

Sam is ***s*** cm shorter than Eva.
Eva is 95 cm tall.
We can write Sam's height in algebra as:

Choose one answer.

- a. 95 - ***s***
- b. 95***s***
- c. ***s*** - 95
- d. ***s***
- e. 94
- f. -***s***95

Item 497: Success rate 57%

1
Marks: --/1

David is 10 cm taller than Con.
Con is *h* cm tall.
We write David's height in algebra as:

Choose one answer.

- 10+***h***
- 10***h***
- ***r***
- ***C*** = ***D*** + 10
- ***h***10
- 18

Item 498: Success rates 47%, 51%, 70%

1
Marks: --/3

Some students had to find some values of x to make this equation true:
x + x + x = 12
Mark the work of each student.

Mary wrote x = 2, x = 5 and x = 5 [▾]

Millie wrote x = 9, x = 2 and x = 1 [▾]

Mandy wrote x = 4 [▾]

Item 502: Success rates 82%, 61%, 85%

1 Marks: --/3

Some students had to give some values of x and y to make this equation true.

$x + y = 16$

Mark the work of each student.

John gave $x = 6$, $y = 10$

Jack gave $x = 8$ and $y = 8$

James gave $x = 9$ and $y = 7$

Item 506: Success rates in the same tabular structure are:

65%	61%
76%	61%
80%	73%
74%	76%
68%	69%

Choose whether each statement is true or false.

$\frac{12\times17}{6\times17}=2$ (TRUE / FALSE)

$\frac{15\times100}{5\times100}=300$

$5\times917=5^{917}$

$\frac{1}{30}+\frac{1}{40}=\frac{2}{70}$

$\frac{6}{73}+\frac{5}{73}=\frac{11}{73}$

$3\times(2\times5)=6\times15$

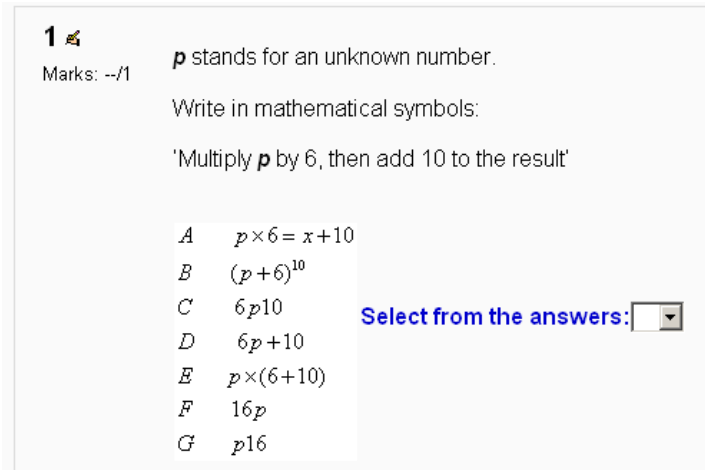
$92-43=43-92$

$2\times(300+7)=600+14$

$\frac{300+6}{3}=100+2$

$\frac{17000-17}{17}=1000-17$

Item 575: Success rate 37%


1

Marks: --/1

p stands for an unknown number.

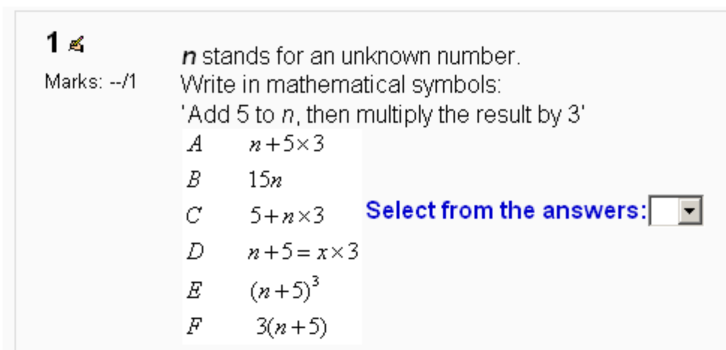
Write in mathematical symbols:

'Multiply ***p*** by 6, then add 10 to the result'

A	$p\times 6 = x + 10$
B	$(p+6)^{10}$
C	$6p10$
D	$6p+10$
E	$p\times(6+10)$
F	$16p$
G	$p16$

Select from the answers:

Item 577: Success rate 17%


1

Marks: --/1

n stands for an unknown number.
Write in mathematical symbols:
'Add 5 to *n*, then multiply the result by 3'

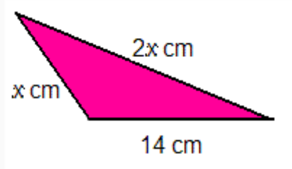
A	$n+5\times 3$
B	$15n$
C	$5+n\times 3$
D	$n+5 = x\times 3$
E	$(n+5)^3$
F	$3(n+5)$

Select from the answers:

Item 579: Success rate 77%

1 Marks: --/1

The perimeter of this triangle is 44 cm.
You are going to work out x using algebra.
Which of these equations would you write **first**?



Choose one answer.

- a. $x + y + z = 44$
- b. $x + 2x + 14 = 44$
- c. $44x - 14x = 30$
- d. $x = \frac{44-14}{3}$

Item 580: Success rate 36%

A bus took people on a 3 day tour.

The distance travelled on Day 2 was 85 km further than on Day 1.
The distance travelled on Day 3 was 125 km further than on Day 1.
The total distance was 1410 km.

You are going to use algebra to work out how far the bus went on the first day.
Which of these equations would you write first?

Choose one answer.

- a. $x + 85 + x + 125 = 1410$
- b. $x + (x + 85) + (x + 125) = 1410$
- c. $x + 85 + 125 = 1410$
- d. $x = \frac{1410-85-125}{3}$

Item 581: Success rate 36%

The cost of hiring a bicycle is made up of a fixed fee of $25 and a usage charge of $8 per hour.

Which way would your teacher prefer to see the rule for this cost written?

Choose one answer.

- a. $25C + 8t$ where the C is the fixed fee and t is the time used in hours
- b. $8t + 25$ where t is the number of hours used
- c. $C = 25 + 8t$ where C is the total hiring cost in dollars and t is the number of hours used
- d. $C = 25 + h = 8t$ where C is the total hiring cost in dollars, h is the hiring charge and t is the number of hours used
- e. $C + 8t$ where the C dollars is the fixed fee to hire a bike in dollars and t is the time used in hours