

The Development of Students' Algebraic Proficiency

Irene van Stiphout

Eindhoven School of Education

Paul Drijvers

Utrecht University

Koeno Gravemeijer

Eindhoven School of Education

Students' algebraic proficiency is debated worldwide. To investigate the development of algebraic proficiency in Dutch secondary education, we set up a study, in which 1020 students in grades 8 – 12 took four algebra tests over a period of one year. Rasch analysis of the results shows that the students do make progress throughout the assessment, but that this progress is small. A qualitative analysis of test items that invite structure sense reveals that students' lack of structure sense may explain the results: the majority of the students were not able to deal flexibly with the mathematical structure of expressions and equations. More attention to structure sense in algebra education is recommended.

Keywords: algebra; algebraic skills; Rasch scale; secondary education; structure sense; symbol sense

Student achievement in algebra is a worldwide concern. International comparative studies such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) induced review studies on how to improve students' algebraic proficiency (e.g., National Mathematics Advisory Panel, 2008). It is widely accepted that this proficiency includes not only procedural skills but also involves algebraic insights.

In the Netherlands, the discussion on algebraic proficiency focused on the level of basic algebraic skills in the transition from secondary education to higher education. Complaints were heard that students are not proficient in basic algebraic algorithms and cannot apply them correctly. As a result, educators and politicians called for a stronger emphasis on procedural skills (Van Gastel et al., 2007). The following examples from McCallum (2010), however, illustrate that the demands in higher education exceed the level of superficial procedural fluency:

- recognizing that $P \cdot \left(1 + \frac{r}{12}\right)^{12n}$ is linear in P (finance);
- identifying $\frac{n(n+1)(2n+1)}{6}$ as being a cubic polynomial with leading coefficient $\frac{1}{3}$ (calculus);
- observing that $L_0 \sqrt{1 - \left(\frac{v}{c}\right)^2}$ vanishes when $v = c$ (physics);
- understanding that $\frac{\sigma}{\sqrt{n}}$ halves when n is multiplied by 4 (statistics).

To address this worldwide and national debate on algebra achievement, and on procedural fluency and conceptual understanding in particular, we decided to investigate the development of algebraic proficiency in Dutch secondary education.

What is Algebraic Proficiency?

As an introduction to our research, we will first discuss what we understand by proficiency. In doing so, we focus on two aspects, namely the relation between procedural fluency and conceptual understanding, and the notion of structure sense.

The Relation between Procedural Fluency and Conceptual Understanding

The distinction between procedural fluency and conceptual understanding is central in discussions on algebraic proficiency. Skemp (1976) distinguished knowing how to apply the rules and algorithms correctly (instrumental understanding) and knowing both what to do and why (relational understanding). Kilpatrick et al. (2001) see procedural fluency and conceptual understanding as two of five strands of mathematical proficiency, along with strategic competence, adaptive reasoning, and productive disposition. To Hiebert and Lefevre (1986), procedural knowledge comprises the formal language (including the symbols), and algorithms and other rules.

It is widely accepted that procedural fluency and conceptual understanding have to go hand in hand: algebraic expertise encompasses a continuum which ranges from basic skills such as procedural work for which a local focus and algebraic calculations suffice, to strategic work which requires a global focus and algebraic reasoning and conceptual understanding. The latter aspects are probably the hardest to learn and to teach, but at the same time, the above examples of McCallum (2010) show the importance of flexible skills such as the ability to read through symbolic expressions and to cleverly select and use symbolic representations. In line with this, Sfard and Linchevski (1994) argue that flexible manipulation skills can be seen as a function of the versatility of available interpretations, and the adaptability of the perspective. In their view, these abilities are part of a structural mode of thinking.

Symbol Sense and Structure Sense

To capture the flexible skills that are involved in algebraic proficiency, Arcavi (1994, 2005) introduced the notion of symbol sense. He defines symbol sense as a complex feel for symbols that includes a positive attitude towards symbols and a global view (or Gestalt view) of expressions. Part of this global view is the ability to read through symbols. As an example of this, Arcavi (1994) discusses the equation $(2x+3)/(4x+6) = 2$. Reading through the symbols reveals that the left-hand side of the equation equals $\frac{1}{2}$ for all $x \neq -\frac{1}{2}$, because the numerator equals half the denominator. Inspecting the equation before starting to solve it with the purpose of gaining a feeling for the meaning of the problem is seen as an instance of symbol sense.

As a second example, Arcavi (1994) discusses Wenger's equation, $v\sqrt{u} = 1 + 2v\sqrt{1+u}$, which is to be solved for v (Wenger, 1987). The difficulty here is to recognize this equation

as linear in v and to overcome the visual salience of the square roots in the equation (Kirscher & Awtry, 2004). This requires identifying parts of the expression as units, an ability that is referred to as the Generalized Substitution Principle (Wenger, 1987).

To solve Wenger's equation, the ability to recognize its – in this case linear – structure is crucial. This ability is labeled as structure sense. The term structure sense is introduced by Linchevski and Livneh (1999, p. 191) to describe the ability “to use equivalent structures of an expression flexibly and creatively.” In high school algebra, structure sense encompasses a collection of abilities, such as: recognize a structure, see a part of an expression as a unit; divide an expression into meaningful sub-expressions; recognize which manipulation is possible and useful to perform; and choose appropriate manipulations that make the best use of the structure (Hoch & Dreyfus, 2004, 2006). Novotná and Hoch (2008) define structure sense as students' ability to (1) recognize a familiar structure in its simplest form, (2) deal with a compound term as a single entity and through an appropriate substitution recognize a familiar structure in a more complex form, and (3) choose appropriate manipulations to make best use of a structure.

We believe that structure sense, as defined by Novotná and Hoch (2008), is such an important aspect of algebraic proficiency that it is worth studying in more detail. In contrast to Novotná and Hoch, we consider structure sense to be part of symbol sense rather than being separated ability, namely the part of symbol sense that involves seeing structures and patterns in algebraic expressions and equations, which is needed while carrying out algebraic manipulations such as simplifying expressions and solving equations.

Research Questions

In light of the above, we have formulated the following research questions.

1. How does students' algebraic proficiency develop from a cross-sectional perspective?
2. How does students' algebraic proficiency develop from an individual perspective?
3. How does students' algebraic proficiency develop in terms of structure sense?

How to Investigate Research Questions?

To address these questions, a set of test items was designed. Four tests consisting of subsets of these items were administered over a calendar year in a partly cross-sectional and partly longitudinal design. Within the cross-sectional perspective, we first determined the distributions of the scores of groups of students at a given times, and then compared those distributions. Within the longitudinal perspective, we looked at how the scores of individual students developed.

Test Design

Because the complaints made by students as well as educators primarily concern algebraic skills taught at the lower secondary level, the algebra test items focus on algebraic skills taught in grades 7 – 9 (Van Gastel et al., 2007, 2010). These items are based on the attainment targets formulated by the Dutch Ministry of Education, Culture and Science, combined with the theoretical considerations discussed above, and are inspired by Dutch textbook series and literature. In Dutch upper secondary education, students choose one of four different streams, each with a different mathematics curriculum. Because of these different programs, the items cover only the common algebra topics, which include expanding brackets, simplifying expressions, and solving equations. The algebra items included range from basic skills to symbol sense in general, and structure sense in particular. The complete list of algebra tasks can be found in the Appendix. In addition, numerical tasks are included that relate to the transition from arithmetic to algebra. These items are not addressed in this article; they are, however, included in the Rasch scales discussed below.

From the set of items, four tests were composed. Each of these tests consisted of 12 to 16 items and was designed to be completed in half an hour, in order not to overburden the students and teachers. The four tests consisted of open questions, to be worked out with paper and pencil. In this way, we avoided students' guessing answers. During the tests, calculators or notes were not allowed.

Test Administration

We assessed students in March 2008, May 2008, October 2008, and February 2009. Students of grades 8, 9, 10 and 11 (ages 13-16) participated in the first and second assessment. After the summer vacation, these students were in grades 9 up to 12 in October 2008 and February 2009. To be able to follow individual students, we have used similar items in different assessments to enable an anchor design.

Table 1

Number of students taking the tests

Grade	March 2008	May 2008	October 2008	February 2009	Part. 4×	Part. ≥ 1
8/9	164	227	173	171	94	266
9/10	163	160	129	114	56	217
10/11	243	185	163	144	90	268
11/12	244	204	188	72	37	269
Total	814	776	653	501	277	1020

Table 1 provides an overview of the numbers of students who took the tests. Four schools participated, all making use of one of the two mostly used Dutch text book series and in that

sense representative. From each school, two classes of each grade were involved. In total, 1020 students participated at least once and 277 students took all four tests.

During data collection a curriculum reform took place. The new curriculum for the grade 10/11 cohort pays more attention to algebraic skills than the old one for cohort grade 11/12, most notably: solving equations, simplifying and calculating with fractions and square roots (Ctwo, 2009). In the interpretation of the findings, this curriculum change will need to be taken into account.

Rasch Scales for Algebraic Proficiency

After data collection, the students' written answers were coded 1 for correct and 0 for incorrect. Doubtful cases were discussed with colleagues. To analyse the test results, we used the Rasch model, a one parameter item response model (Rasch, 1980; Bond & Fox, 2007; Linacre, 2009). With a Rasch analysis, one linear scale is created on which both persons are situated according to their ability and items according to their difficulty. On this scale, not only the order but also the distances between the items and the students have meaning. Rasch theory supposes that the probability of a person giving a correct answer on an item is a logistic function of the difference between that person's ability and the difficulty of the item. The probability P_{ni} of person n with ability B_n to correctly answer item i with difficulty D_i

$$\text{is given by } P_{ni} = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}}.$$

Both the ability of the persons and the difficulty of the items are measured in so-called units of log odds ratios, or logits. The local origin of the Rasch scale is usually situated in the center of the range of item difficulties. If the ability equals the item difficulty, that is, if

$$B_n = D_i, \text{ then } P_{ni} = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}} = \frac{e^0}{1 + e^0} = \frac{1}{2}.$$

For each assessment, including the arithmetic assignments, we created its corresponding Rasch scale. Next, we connected the four Rasch scales by using anchor items, i.e., similar items in the different assessments. As a result, items of all four assessments are placed on one scale. Also, students of different assessments have a Rasch measure on one scale.

To determine which items students master, we need to decide what we view as mastery. We consider a probability of 80% of answering an item correctly as an expression of mastering that item. From the Rasch model it follows that a probability of 0.8 of person n answering item i correctly corresponds to an ability B_n which is 1.39 logit higher than the

difficulty D_i of item i because if $B_n - D_i = 1.39$, then $P_{ni} = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}} = \frac{e^{1.39}}{1 + e^{1.39}} \approx 0.8$. As a

consequence, we consider that students with a measure at least 1.39 higher than the measure of the item master that particular item. The fit of the Rasch model to the data was checked. With respect to the reliability, we found values of .70, .70, .66 and .68 for assessments 1, 2, 3 and 4, respectively. These reliability scores can be compared to Cronbach's alpha.

Differences between Student Cohorts

To address the first question on the development of the different cohorts of students, we performed a cross-sectional comparison of grades 8 through 12. Figure 2 shows the percentiles of Rasch measures in logits of the four cohorts of students. As the number of participating students varied over the assessments, the bars between the dashed lines only partly represent the same students.

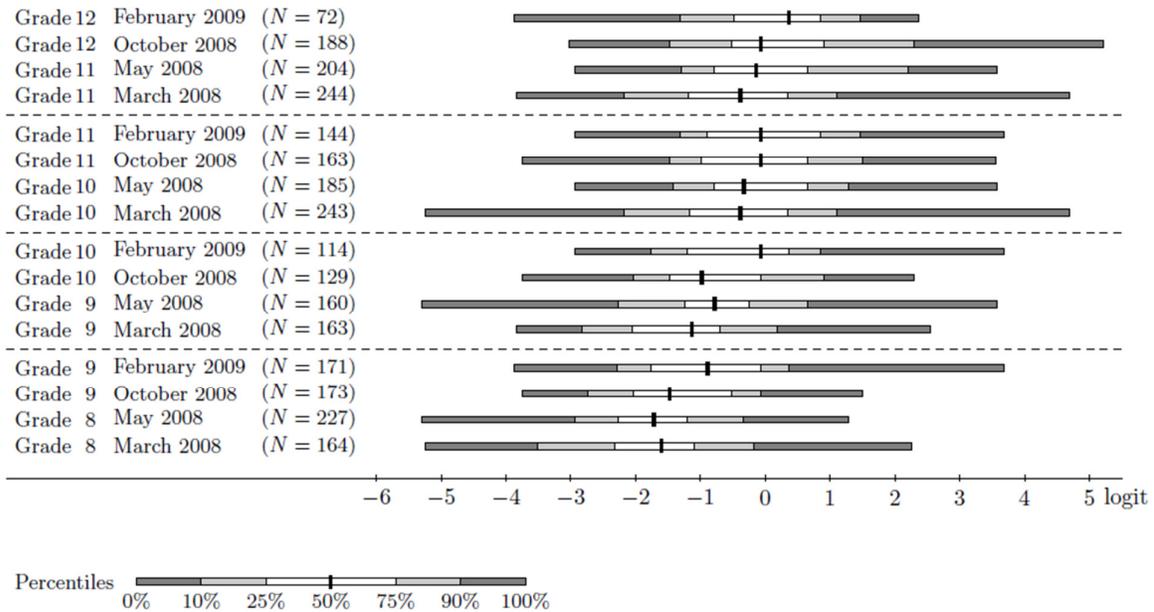


Figure 1. Cross-sectional percentiles of all grades in all assessments.

Figure 1 shows that the averages of the different assessments generally increase with the grades. For the central 50% of the students (the white parts of the bars), the difference between the lowest average (grade 8, May 2008) and the highest average (grade 12, February 2009) is approximately two logits. Thus, the dispersion within the twelve or sixteen different assignments is rather small with regard to the difference between the worst and the best scoring student (the length of the whole bar). If we focus on 80% of the students (i.e., leaving out the best and worst 10%), we see that they are within a range of at most 3.8 logits of each other. Grade 9/10 performs better than grade 8/9 and grade 10/11 performs better than grade 9/10. The difference between grade 10/11 and grade 11/12 is less obvious. Here we note that the curriculum of students in grade 11/12 differs from the curriculum of students of the other grades on algebraic skills, which makes it hard to interpret this lack of difference between grades 10/11 and grades 11/12.

To sum up, the cross-sectional analysis showed that there is progress and there is only little dispersion among the central 50% of the students. There seems to be a growth in ability between the cohort of grade 10/11 and that of grade 11/12, which might be a positive effect of the curriculum changes in the Netherlands.

Individual Proficiency Development

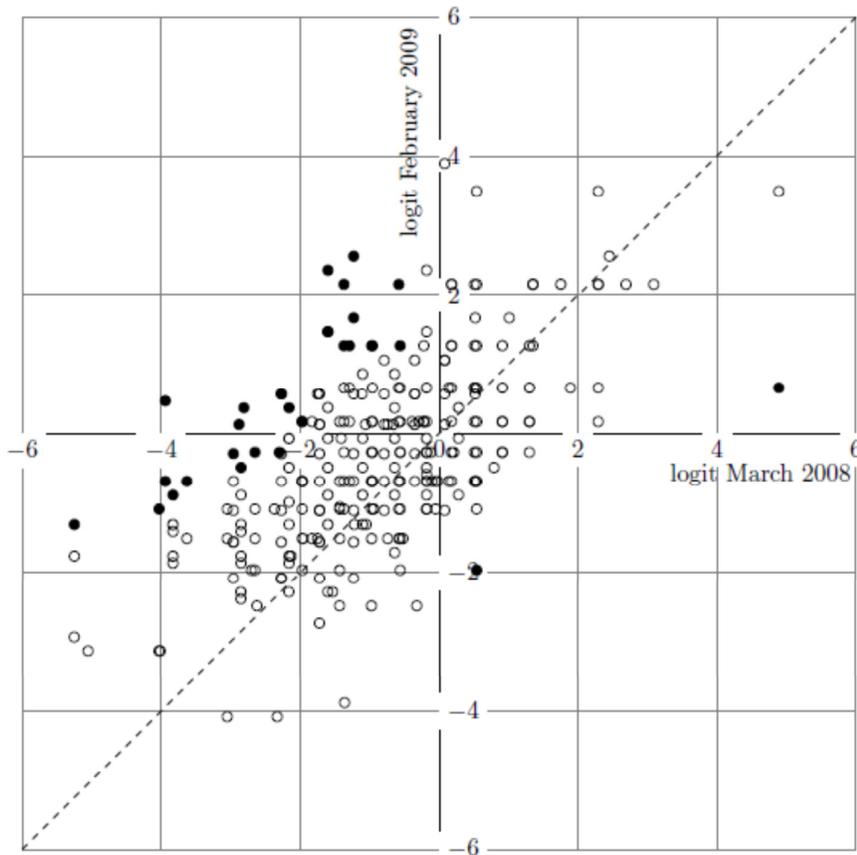


Figure 2. Rasch measures of students in the first and fourth assessment ($N=390$).

To address the second question on the development of individual students' algebraic proficiency over time, Figure 2 shows the Rasch measurements in logits of individual students in the first assessment (horizontally) against the similar results of the fourth assessment (vertically). Each dot represents one student who participated in both the first and the fourth assessment. Students above the dashed line improved their performance, whereas students below the dashed line show a decreasing ability. The open dots represent students who did not make significant difference; the filled dots represent students who made significant progress or retrogress. Based on the 95% confidence intervals provided by the Rasch analysis, the results show that the majority of the students (358 out of 390) did not make significant progress. Some students (30 out of 390) did make progress, whereas 2 students retrogressed.

Summarizing, this analysis showed that individual students make progress during a calendar year, but that only few students made significant progress.

A Closer Look on Structure Sense Items

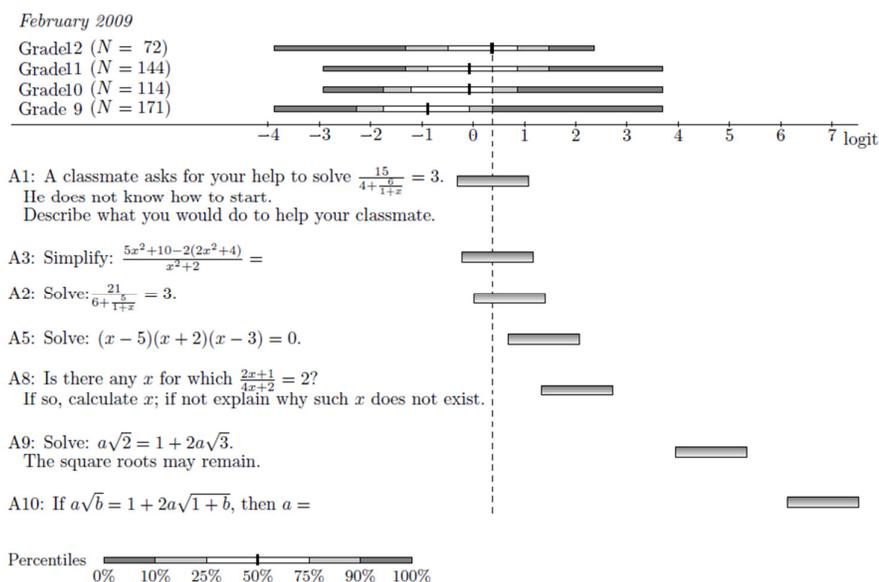


Figure 3. February 2009 student percentiles and structure sense tasks on the Rasch scale.

To investigate students' proficiency development in a more qualitative way, we analyze in greater detail the tasks that are tailored to structure sense. We selected ten tasks from the four assessments for which recognizing the algebraic structure of an expression or seeing a part of an expression as a unit really pays off. Figure 3 provides an overview of student measures of the fourth assessment in February 2009, and the task difficulties. For clarity's sake, in case of similar tasks, we included only one task in the Figure, resulting in seven tasks: one involving simplification an expression, the other six on solving equations.

Central in Figure 3 is the horizontal axis with logits as units. The gray scaled bars above the axis represent student ability in percentiles in the fourth assessment in February 2009. The bars below the horizontal axis represent the difficulty of the tasks on the Rasch scale. The left-hand side of a bar corresponds to a probability of 0.50 of answering the corresponding task correctly, which is usually referred to as the Rasch measure of the task; the right-hand side of the bar corresponds to a probability of 0.80 of answering that task correctly. Tasks for which the corresponding bars lie on the left-hand side of the dashed line are mastered by at least 50% of the grade 12 students. Figure 3 shows that all seven tasks are mastered by less than 25% of the grade 12 students.

Below, we discuss the students' performance on each of the seven tasks.

Items A1 and A2: Fraction in Equation

In the first assessment (March 2008), students were asked to report the first clue they would provide a classmate to help solve the equation $\frac{15}{4+\frac{6}{1+x}} = 3$. The Rasch measure of

this task, A1, is -0.31 logit (probability of success 0.50). The way the task was formulated

left room for answers such as “I would call the teacher for help.” This kind of answer revealed problems in the dichotomous coding of student answers. To avoid this kind of answer, a formal version, A2, was included in the third assessment. The Rasch measure of the latter version is 0.01 logit, so the tasks did not differ much.

To solve these equations, different strategies can be used. For example, students may use the cover-up method and cover the denominator. In this way, the equation becomes $\frac{15}{W} = 3$, which is easily solved, providing $W \neq 0$. Following this line of thought, the next step would be $4 + \Delta = 5$ which implies $\frac{6}{1+x} = 1$. This equation in turn could be solved by covering the denominator, thus yielding $\frac{6}{\diamond} = 1$, which implies $1+x=6$. An example of a more formal strategy is to multiply the numerator and the denominator both with $1+x$, or multiply both sides of the equation with the denominator $4 + \frac{6}{1+x}$. These strategies have in common that students have to identify a part of the equation (the denominator, or a part of the denominator) as an object, which is seen as an expression of structure sense.

Item A3: Simplification

In the second assessment (May 2008), we asked students to simplify the expression

$$\frac{5x^2 + 10 - 2(2x^2 + 4)}{x^2 + 2}$$

The Rasch measure of this task is -0.22 logit (probability of success 0.50). A similar task, A4, with Rasch measure -0.50 , was included in the fourth assessment (February 2009).

This formula has the structure $\frac{5W-4W}{W}$ which leads to $\frac{x^2+2}{x^2+2}$. But simply expanding the brackets and taking similar terms together yields the same. In both cases the next step is to recognize that this fraction consists of two similar expressions that are divided, so the fraction yields one.

Vereenvoudig zo ver mogelijk:

$$\frac{5x^2 + 10 - 2(2x^2 + 4)}{x^2 + 2} =$$

$$\frac{5x^2 + 8(2x^2 + 4)}{x^2 + 2}$$

$$\frac{5x^2 + 16x^2 + 32}{x^2 + 2}$$

$$\frac{21x^2 + 32}{x^2 + 2}$$

Vereenvoudig zo ver mogelijk:

$$\frac{5x^2 + 10 - 2(2x^2 + 4)}{x^2 + 2} =$$

$$\frac{5x^2 + 10 - 4x^2 - 8}{x^2 + 2} =$$

$$\frac{x^2 + 2}{x^2 + 2} = 1$$

Translation: Vereenvoudig zo ver mogelijk: Simplify

Figure 4. Work of a grade 8 (left) and a grade 10 (right) student.

An example of the difficulties students experienced is presented in Figure 4. The grade 8 student rewrote the numerator $5x^2 + 10 - 2(2x^2 + 4)$ to $5x^2 + 8(2x^2 + 4)$ by erroneously taking 10 and -2 together. To this student, the algebraic structure of the numerator was not clear. Another type of error is shown in the right screen of Figure 4. This grade 10 student correctly simplified the expression to $\frac{x^2 + 2}{x^2 + 2}$. Then he concluded that $\frac{x^2 + 2}{x^2 + 2}$ equals zero instead of one. This error might stem from an inability to see the algebraic structure of the expression, but other explanations such as a deep misunderstanding of fractions or the student being used to the form “expression = 0” seem also reasonable.

Item A5: Factorization

In the second assessment, students were asked to solve the equation

$$(x-5)(x+2)(x-3) = 0.$$

This requires students to identify the underlying algebraic structure, which involves three factors on the left side of the equation, and the product of these three factors equals zero. The structure is $A \cdot B \cdot C = 0$, which implies $A = 0$ or $B = 0$ or $C = 0$. The Rasch measure of this task is 0.68 logit (probability of success 0.50). A similar task, A6, was included in the fourth assessment. This task was perceived as less difficult, with a Rasch measure of -0.60 logit, probably due to a test-retest bias. The difficulty of these tasks can be explained by the students' tendency to expand the brackets, after which they could not find the factorization. Some of the students came to see the error and recovered by giving the correct answer.

8. Los op:
 $(x-1)(x+3)(x-4) = 0.$
 $(x^2 - 2x - 3)(x-4) = 0$
 $x^3 - 4x^2 - 3x^2 - 12x - x^2 + 4x - 3x + 12 = 0$
 $x^3 - 2x^2 + x + 12 = 0$
 ~~$x^3 - 2x^2 + x + 12 = 0$~~
 $x = 1 \quad \vee \quad x = -3 \quad \vee \quad x = 4$

Translation: Los op: Solve

Figure 5. Work of grade 12 student.

Figure 5 shows such a work of a grade 12 student who started to solve the equation by expanding the brackets. In the second line of this work, he correctly wrote that

$$x^3 - 4x^2 + 3x^2 - 12x - x^2 + 4x - 3x + 12 = 0.$$

Moving to the next line, he forgot the factor $-12x$, so erroneously concluded that $x^3 - 2x^2 + x + 12 = 0$. In the next line, which is scratched, he started to factor out x . Apparently, then he realized that the solutions can be found more easily and wrote the correct solutions down.

The students' tendency to expand the brackets indicates the visual salience of the brackets. The underlying structure of $A \cdot B \cdot C = 0$ is overlooked.

Item A8: Arcavi's Equation

In the section on structure sense, we addressed the equation $(2x+3)/(4x+6) = 2$. Arcavi (1994) argues that resisting the impulse to immediately solve the equation, and instead to try to read meaning into the symbols, is an expression of symbol sense, and in our opinion of structure sense in particular. In this example, it requires that students recognize the form $\frac{A}{2A}$ in the expression $(2x+3)/(4x+6)$. To do so, it is necessary to see $2x+3$ as a unit. However, realizing that $(2x+3)/(4x+6)$ equals $\frac{1}{2}$ is not necessary for the solving process. Just solving the equation and finding the solution $x = -1\frac{1}{2}$ and realizing that the denominator equals zero, is a correct procedure for solving this equation. So, reading through the equation and seeing the structure is practical, but not necessary.

The ability to see $\frac{A}{2A}$ manifests more structure sense than finding the solution for which the denominator equals zero and then concluding that the equation does not have a solution. Since more structure sense is supposed to be an expression of a higher level of algebraic proficiency, we expect students with more symbol sense to have a higher Rasch measure, and thus we expect students who see $\frac{A}{2A}$ to have a higher Rasch measure than students who found the solution for which the denominator equals zero. To investigate this relation between these strategies and students' proficiency, we performed a qualitative analysis of the written answers of the students. As we were interested in the strategies used to solve this equation, we restricted the analysis to the correct answers. The analysis yielded three categories of students. First, students that argued that the quotient equals $\frac{1}{2}$. These students recognized the structure $\frac{A}{2A}$ in the equation. The second category contains students who found the solution for which the denominator equals zero. These students correctly argued that this solution is no solution. The third category contains students who gave another argument. For example, these students only wrote "no." Another example is a student who argued that $(2x+1)/(4x+2)$ equals a fraction between 0 and 1. From the 501 participating students in the fourth assessment, 105 gave a correct solution. Half of these students used the strategy of the first category. The other students were nearly evenly distributed between the other categories.

Figure 6 shows the relation between strategy and ability. Again, the central line represents the Rasch scale of proficiency. The gray scaled bars below the axis represent percentiles of students' Rasch measures in the fourth assessment in February 2009. The dots above the axis represent students in a particular grade with a particular strategy.

From the figure we see that, although more students solve the task correctly using the strategy "quotient equals $\frac{1}{2}$ " than using the strategy "denominator zero," the Rasch measure

of the students using the former strategy is not higher than those using the latter. In other words, the strategy that is supposed to be a manifestation of more symbol sense does not imply a higher Rasch measure.

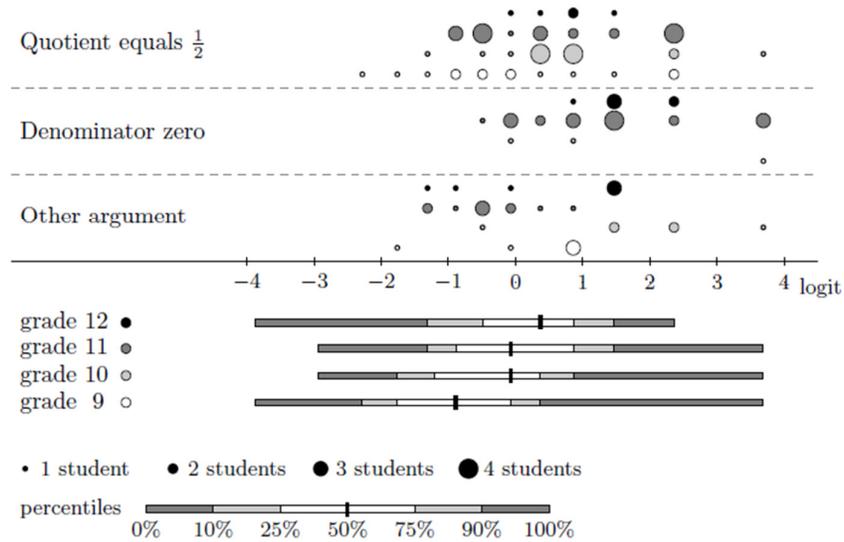


Figure 6. Students' ability combined with the strategy used.

In our view, this means that either the relation between strategy and structure sense is not as strict as the literature suggests, or the relation between structure sense and the underlying latent variable of the Rasch scale is weak. In the former case, the strategy students use might depend on the strategy they think they are expected to use based on the recipes Dutch textbooks provide. Students might think that they have to use these recipes as part of an (implicit) didactical contract (Brousseau, 1990). From this point of view, the more broadly applicable strategy "denominator zero" might be viewed as more valuable, because this strategy could also be used in case the equation had been $(2x+1)/(3x+2) = \frac{1}{2}$. However, this single case cannot justify radical conclusions; further research on the relation between strategies, structure sense and algebraic proficiency seems appropriate.

Items A9 and A10: Wenger's Equation

The two items with the highest difficulty, A9 and A10, are adapted versions of Wenger's equation. Figure 3 shows that these two tasks were far beyond the ability of all students. In the first assessment, we included Wenger's equation in which we replaced the letters u and v for readability with the letters a and b . We did not ask students to solve $a\sqrt{b} = 1 + 2a\sqrt{1+b}$ for a , because Dutch students are not familiar with this kind of question. Instead, we asked students to rewrite $a\sqrt{b} = 1 + 2a\sqrt{1+b}$ as an expression of the form $a = \dots$. The analysis revealed that this way of asking is ambiguous to students. For example, students divided both sides of the equation by \sqrt{b} , which yields to $a = \frac{1 + 2a\sqrt{1+b}}{\sqrt{b}}$. This is not the kind of answer we intended to see, but somehow meets the purport of the question. Exactly two students out

of the 650 of grades 9, 10 and 11 were able to solve this equation. These two students recognized the linear form of the equation and gave the correct answer, $a = \frac{1}{\sqrt{b} - 2\sqrt{1+b}}$.

In the third assessment, an adapted version of the equation was included. In this version, we substituted $b = 2$, so the equation to be solved became $a\sqrt{2} = 1 + 2a\sqrt{3}$. Only 19 out of the 653 participating students were able to solve this task.

The difficulty of Wenger's equations is explained by students' inability to recognize the linear form of the equation. This requires students to sense the symbols as arranged in a special pattern, which is an expression of structure sense. Wenger (1987) found that students are able to perform manipulations correctly, but these manipulations do not lead to a solution. Rather, students go round in circles, create more complex expressions and then reduce these terms. The square roots that serve as coefficients in this equation may be interpreted by the students as a signal to square both sides of the equation. The unfamiliarity of students with square roots as coefficients can enhance the visual salience of square roots (Kirscher & Awtry, 2004).

We agree that recognizing the linear form is crucial for solving the equation. However, in our view, there are other hurdles. First, the unconventional sequence of symbols in $a\sqrt{2} = 1 + 2a\sqrt{3}$. In this equation, the variable is followed by the numerical coefficient in the left-hand side of the equation. In the right-hand side, the variable is in the middle of the numerical coefficient. This is an unusual sequence of symbols for Dutch students, because in Dutch textbook series, the numerical coefficient usually precedes the variable. In this way, the equation would have been $\sqrt{2}a = 1 + 2\sqrt{3}a$. We realize that this is only a slight difference for an expert. But we believe that to students, this change of order perhaps makes the difference between being able or not being able to solve the equation. The role of the order of symbols in exercises would be an interesting topic of further studies.

The second hurdle concerns the different roles of the variables a and b in the equation. The a in the equation serves as unknown, whereas the b serves as a variable. The versatility of the use of variables is a well known difficulty in mathematics and has been studied by many researchers (e.g., Matz, 1982; Janvier, 1996; Rosnick, 1981; Wagner, 1983; Trigueros & Ursini, 1999; Drijvers, 2003; Ursini & Trigueros, 2001; Schoenfeld & Arcavi, 1988). The skill of flexibly dealing with the different roles of variables can be seen as part of a broad view on equations. The results for these tasks suggest that students do not have such a broad view.

Summarizing, our findings confirm that Wenger's equation presents students with several difficulties. These difficulties all concern the ability to recognize the linear structure of the equation which can be seen as a part of structure sense. The students' performance on the two linear equations does not allow us to conclude which of these difficulties is paramount, but in our view, solving these equations requires structure sense –which the students apparently lack.

Conclusion and Discussion

In this article we set out to answer the following research questions.

1. How does students' algebraic proficiency develop from a cross-sectional perspective?
2. How does students' algebraic proficiency develop from an individual perspective?
3. How does students' algebraic proficiency develop in terms of structure sense?

In answer to the first question, we found that the Dutch student cohorts involved in the study made some progress. In general we concluded that students mastered simple tasks, but tasks become too complicated rather quickly. The difference between grade 10/11 and grade 11/12 is small. This might be explained by the aforementioned curriculum change: students in grade 8/9, grade 9/10 and grade 10/11 followed a curriculum that includes more algebra than the program of students of grade 11/12. This curriculum change might have had a positive effect on students' algebraic proficiency of the students from grade 8/9 through grade 10/11. Furthermore, there is little dispersion among the middle 50% of the students.

Related to the second question, the analysis of the results yields that the majority of the individual participants did make progress from the first to the fourth assessment, so in a period of one year. However, this progress was not significant for the majority of the students. The answer to the third question is that there are only few tasks that were mastered by the majority of the students. Although students showed progress both cross-sectionally and longitudinally, this progress did not encompass the majority of the items. In other words, the majority of the items were too difficult for students of grade 8, and were still too difficult for students of grade 12. Furthermore, the range of items that the majority of the students mastered did not include tasks that involve conceptual aspects of algebraic proficiency, and for structure sense in particular. In answering the third research question, we found that the majority of the students was not able to deal flexibly with the mathematical structure of expressions. Also, the notion of structure sense proved valuable to interpret student results and to explain student difficulties. For example, the inability to solve the equation $(x-5)(x+2)(x-3)=0$ can be understood as the inability to recognize the mathematical structure $A \cdot B \cdot C = 0$ that implies $A=0$ or $B=0$ or $C=0$, which is a lack of structure sense.

To put those results in perspective, we have to take into account some limitations of the study. First, because we did not want to place too heavy a load on the teachers and the students, we chose to keep the number of test items relatively low. As a consequence, differences would have had to have been quite large to be significant. Second, a curriculum change took place during the data collection. The new curriculum pays more attention to algebra. We took this change into account by concluding that the curriculum change might have a positive effect if the growth continues.

If we reflect on the study's results, we consider them as disappointing in that the students hardly develop structure sense. The results suggest that Dutch students hardly develop structure sense as is evident from the students' inability to

- see that $15/(4 + (6/(1 + x))) = 3$ can be read as $15/W = 3$, which means that $W=15:3$ (providing $W \neq 0$);
- observe that the expressions $5x^2 + 10$ and $2x^2 + 4$ are multiples of $x^2 + 2$;
- recognize the expression $(x - 5)(x + 2)(x - 3)$ as a multiplication;

- recognize the equation of Wenger as linear.

These examples reflect the demands of higher education—as expressed by the examples of McCallum (2010) mentioned in the introduction—which exceed the level of procedural fluency. Students tend to choose the routine way for solving problems, and do not manage to step out of the procedure in order to reconnect to the underlying meaning when needed (Arcavi, 1994).

The students' tendency to focus on routine procedures might be a consequence of the didactical contract to which the textbooks contribute. These textbooks tend to focus on the procedures and not so much on symbol sense and structure sense (Van Stiphout, Drijvers & Gravemeijer, 2013).

A limited feel for the structure of expressions, and for (sub)expressions as objects (Sfard 1991) is an obstacle for reaching a higher level of conceptual understanding in which the structure and ambiguous nature of the algebraic objects are central. Reaching this higher level is inherently difficult and involves a shift of thinking. How to reach this higher level is a core concern of the mathematics education research community. We may conclude that the call from educators and politicians for more attention to routine and procedural skills will not solve the students' problems, because the problems with the more difficult items do not primarily stem from a lack of procedural skills, but more from a lack of conceptual understanding.

Acknowledgements

This article is based on Chapters 2 and 3 of the first author's PhD study: Van Stiphout, I. M. (2011). *The development of algebraic proficiency* (Doctoral dissertation). Eindhoven School of Education, Eindhoven, The Netherlands.

References

- Arcavi, A. (1994). Symbol sense: Informal sense-making in formal mathematics. *For the Learning of Mathematics*, 14(3), 24-35.
- Arcavi, A. (2005). Developing and using symbol sense in mathematics. *For the Learning of Mathematics*, 25(2), 42-47.
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences (Second Edition)*. Mahwah, NJ: Lawrence Erlbaum.
- Brousseau, G. (1990). Le contrat didactique: le milieu. [The didactical contract: The setting]. *Recherches en didactique des mathématiques*, 9(3), 309-336.
- Ctwo. (2009). *Rapport tussenevaluatie van de 2007-programma's wiskunde havo/vwo*. [Report on the intermediate evaluation of the 2007 programs mathematics havo/vwo]. Retrieved 15 December 2013 from <http://www.ctwo.nl/>.
- Drijvers, P. (2003). *Learning algebra in a computer environment: Design research on the understanding of the concept of parameter* (Doctoral dissertation). University of Utrecht, The Netherlands.

- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum.
- Hoch, M., & Dreyfus, T. (2004). Structure sense in high school algebra: The effect of brackets. *Paper presented at the 28th Conference of the International Group for the Psychology of Mathematics Education*. Bergen, Norway: PME.
- Hoch, M., & Dreyfus, T. (2006). Structure sense versus manipulation skills: An unexpected result. *Paper presented at the 30th Conference of the International Group for the Psychology of Mathematics Education*. Prague, Czech Republic: PME.
- Janvier, C. (1996). Modeling and the initiation into algebra. In N. Bednarz et al. (Ed.) *Approaches to algebra* (pp. 225-236). Dordrecht: Kluwer Academic Publishers.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up*. Washington, D.C.: National Academy Press.
- Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education*, 35(4), 224-257.
- Linacre, J. M. (2009). *A user's guide to Winsteps*. Retrieved August 31, 2012, from <http://www.winsteps.com/winsteps.htm>
- Linchevski, L., & Livneh, D. (1999). Structure sense: The relationship between algebraic and numerical contexts. *Educational Studies in Mathematics*, 40(2), 173-196.
- Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman, & J. Brown (Eds.), *Intelligent Tutoring Systems* (pp. 25-50). London: Academic Press.
- McCallum, W. (2010). Restoring and balancing. In Z. Usiskin, K. Andersen, & N. Zotto (Eds.), *Future curricular trends in school algebra and geometry: Proceedings of a conference* (pp. 277-286). Charlotte, NC: Information Age Publishing Inc.
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, D.C.: U.S. Department of Education.
- Novotná, J., & Hoch, M. (2008). How structure sense for algebraic expressions or equations is related to structure sense for abstract algebra. *Mathematics Education Research Journal*, 20(2), 93-104.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Rosnick, P. (1981). Some misconceptions concerning the concept of variable. *Mathematics Teacher*, 74(6), 418-420.
- Schoenfeld, A., & Arcavi, A. (1988). On the meaning of variable. *Mathematics Teacher*, 81(6), 420-427.
- Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22(1), 1-36.
- Sfard, A., & Linchevski, L. (1994). The gains and pitfalls of reification: The case of algebra. *Educational Studies in Mathematics*, 26(2-3), 191-228.

- Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 1-7.
- Trigueros, M., & Ursini, S. (1999). Does the understanding of variable evolve through schooling? *Paper presented at the 23rd Conference of the International Group for the Psychology of Mathematics Education*. Haifa, Israel: Technion.
- Ursini, S., & Trigueros, M. (2001). A model for the use of variable in elementary algebra. *Paper presented at the 25th Conference of the International Group for the Psychology of Mathematics Education*. Utrecht: Freudenthal Institute.
- Van Gastel, L., Cuypers, H., Jonker, V., Van de Vrie, E., & Van der Zanden, P. (2007). *Eindrapport nationale kennisbank basisvaardigheden wiskunde*. Amsterdam: Consortium NKBW.
- Van Gastel, L., Cuypers, H., Grift, Y., Kaper, W., Van der Kooij, H., Tempelaar, D. et al. (2010). *Aansluitmonitor Wiskunde VO-HO*. Amsterdam: Consortium NKBW.
- Van Stiphout, I. M. (2011). *The development of algebraic proficiency* (Doctoral dissertation). Eindhoven School of Education, Eindhoven, The Netherlands.
- Van Stiphout, I. M., Drijvers, P. & Gravemeijer, K. (2013). The implementation of contexts in Dutch textbook series: a double didactical track? In Lindmeier, A.M. & Heinze, A. (Eds.). *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 4, pp. 337-344. Kiel, Germany: PME.
- Wagner, S. (1983). What are these things called variables? *Mathematics Teacher*, 76, 474-479.
- Wenger, R. (1987). Cognitive science and algebra learning. In A. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 217-251). Hillsdale, NJ: Lawrence Erlbaum.

Authors

Irene van Stiphout, **PhD.**, Eindhoven School of Education, Cito Amsterdamseweg 13, P.O. Box 1034, 6801 MG Arnhem, The Netherlands; Irene.vanStiphout@cito.nl

Paul Drijvers, **Associate Professor**, Freudenthal Institute for Science and Mathematics Education, Faculty of Science, Utrecht University, Princetonplein 5, Office 367, PO Box 85170, 3508 AD Utrecht; p.drijvers@uu.nl

Koeno Gravemeijer, **Full Professor Emeritus**, Eindhoven School of Education, Technische Universiteit Eindhoven P.O. Box 513 5600 MB Eindhoven; k.gravemeijer@tue.nl

Appendix: Test Items

Tasks with an A in the measure are tasks that have served as anchor items in the Rasch analysis. Tasks are arranged by increasing measure. The Rasch measure of a task corresponds to a probability of 0.50 to answer that task correctly.

Test Item	Included in test version	Rasch Measure (logits)
Expand the brackets: $-4(3a + b) =$	March 2008	-3.46 A
Expand the brackets: $-5(2p + q) =$	May 2008	-3.46 A
Expand the brackets: $-3(4p + q) =$	October 2008	-3.46 A
Expand the brackets: $-4(5p + q) =$	February 2009	-3.46 A
Simplify: $-2(3x - y) + 3(-4y - 2) =$	March 2008	-1.47 A
Simplify: $-3(2a - b) + 4(-2b - 3) =$	May 2008	-1.47 A
Simplify: $-4(2x - y) + 2(-3y - 4) =$	October 2008	-1.47 A
Simplify: $-2(4x - y) + 3(-2y - 4) =$	February 2009	-1.47 A
You know the operations plus, minus, multiplication and division. We introduce an operation, diamond, and $a \diamond b$ is defined as follows. For two numbers a and b , we say $a \diamond b = a^2 - a \cdot b$. Does $a \diamond b = b \diamond a$ hold for all numbers a and b ?	October 2008	-1.24
Simplify $(3a^2 + 2a + 7)(a + 8)$. Show your work.	March 2008	-0.85 A
Simplify $(2x^2 + 3x + 4)(x + 6)$. Show your work.	May 2008	-0.85 A
Simplify $(2a^2 + 4a + 3)(a + 5)$. Show your work.	October 2008	-0.85 A
Simplify $(3a^2 + 4a + 5)(a + 7)$. Show your work.	February 2009	-0.85 A
Solve: $(x - 1)(x + 3)(x - 4) = 0$.	February 2009	-0.60
Simplify: $\frac{7x^2+7-3(2x^2+2)}{x^2+1} =$	February 2009	-0.50
Solve: $(x - 5)(x - 7) = 3$.	March 2008	-0.48
A classmate asks for your help in solving $\frac{15}{4+\frac{6}{1+x}} = 3$. He does not know how to start. Describe what you would do to help your classmate.	March 2008	-0.31

Test Item	Included in test version	Rasch Measure (logits)
Simplify: $\frac{5x^2+10-2(2x^2+4)}{x^2+2}$	May 2008	-0.22
Solve: $\frac{21}{6+\frac{5}{1+x}} = 3$.	October 2008	0.01
Martijn claims that $Q = \sqrt{P-2}$ implies $P = Q^2 + 2$. Explain why you do or do not agree with Martijn.	May 2008	0.07
Solve: $x^2 - 5x + 6\frac{1}{4} = 0$	March 2008	0.33
Solve: $(x-5)(x+2)(x-3) = 0$	May 2008	0.68
Substitute $a = -1$ and $b = -2$ in $-(ab^2)^3 - 2(a^2b)^2$.	October 2008	0.72
Substitute $a = -2$ and $b = -1$ in $-(a^2b)^3 - 2(ab^2)^2$.	March 2008	0.91
Martijn claims that $\sqrt{a^2} = a$ holds for all numbers, a . Explain why you do or do not agree with Martijn.	March 2008	1.00
Is there any x for which $\frac{2x+1}{4x+2} = 2$? If so, calculate x ; if not, explain why such an x does not exist.	February 2009	1.34
Is there any x for which $\frac{2x+3}{4x+6} = 2$? If so, calculate x ; if not, explain why such an x does not exist.	March 2008	1.37
Rewrite the formula $P = \frac{1}{Q} + 5$ as a formula of the form $Q = \dots$ something with $P \dots$	March 2008	1.43
Solve: $2(3x+2) = 3(2x-1) + 7$.	October 2008	2.01
Solve: $a\sqrt{2} = 1 + 2a\sqrt{3}$. The square roots may remain.	October 2008	3.94
If $a\sqrt{b} = 1 + 2a\sqrt{1+b}$, then $a =$	March 2008	6.13