Can a Short Online Test Diagnose Student Thinking?

Vicki Steinle
University of Melbourne
v.steinle@unimelb.edu.au

Kaye Stacey
University of Melbourne
k.stacey@unimelb.edu.au

Beth Price
University of Melbourne
priceb@unimelb.edu.au

This paper provides evidence that a short, fully online, well-constructed diagnostic test based on research literature can give teachers information about their students' thinking and strategies that is sufficiently accurate to use for formative assessment purposes. The example is a test for students beginning to learn to solve equations. The main goal is to inform their teacher of the solving strategies each can use. Accuracy data from 3010 students is analysed to judge how well the overall performance of each group matches predictions. Alongside a previous analysis of misconceptions and common errors, we show this test gives teachers a good picture of most students to plan their lessons.

The purpose of this paper is to provide evidence that a short, fully online diagnostic test can give teachers information about their students' thinking and strategies that is sufficiently accurate to use for formative assessment purposes, i.e., to identify individual student's understandings and to plan future teaching that builds on their current knowledge. Of course, we do not claim that a computerised test is the best way to learn about an individual's mathematical thinking and knowledge, but our focus is on developing quick and easy tools to give teachers a 'good enough' guide for teaching an up-coming topic to all their students. Over many years, we have created *Specific Mathematics Assessments that Reveal Thinking* (SMART::tests) on 66 topics for middle years students.

In this paper, using a test of solving equations as an example, we illustrate one part of the evaluation of these tests from the large-scale data, here using the accuracy of student responses to assess its adequacy as a diagnostic tool and to identify refinements. To evaluate a SMART::test, we also conduct an in-depth analysis of students' common errors and misconceptions, as demonstrated for the present test by Steinle, Stacey, et al. (2022). Small-scale mixed methods studies have also been undertaken to validate SMART::tests (e.g., Klingbeil et al., 2024)

SMART::tests (www.smartvic.com) have been described elsewhere (e.g., Stacey et al., 2018). For this paper, the key points are that volunteer teachers assign the tests, which students can do at school or at home when needed. All aspects of the test are fully automated. The diagnosis is not just based on the correctness of items but also considers actual responses. The teacher receives, for each student, a 'stage of learning' and possibly some flags which indicate that there is evidence of a misconception, a missing conception or a common (procedural) error. Each test has a highly specific focus on a central aspect of a topic that we believe influences students' underlying understandings. The focus is on what mathematics education researchers have identified lies behind students' work and affects their success, not what is easy to see from a standard test. Professional learning is an important goal, by helping to deepen teachers' knowledge of hidden cognitive obstacles as they see how their own students think and see teaching suggestions for students at each stage.

In the next section, we describe the Australian curriculum expectations that are the focus of this test; briefly describe research which identifies an important learning challenge; give a mathematical analysis of the three strategies that students should learn, and explain the diagnosis made by the test. The method section briefly describes the test and its construction,

(2024). In J. Višňovská, E. Ross, & S. Getenet (Eds.), *Surfing the waves of mathematics education. Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia* (pp. 495–502). Gold Coast: MERGA.

and the data collection. We then look in turn at the performance of students in each of the diagnosed stages, examining whether their performance supports the diagnosis supplied.

Space limitations restrict this paper to reporting only on analyses of accuracy statistics—the numbers of students who are correct, incorrect or omit each item. This paper is supplemented by data from a detailed study of errors within the item responses (Steinle, Stacey et al., 2022). In this paper, it is valuable to see how much accuracy statistics can reveal.

Solving Linear Equations—Mathematical Analysis and Research

Curriculum Expectations

This paper focusses on the SMART::test *Solving Linear Equations*. Its target group is students beginning to learn (symbolic) algebra. The Australian Curriculum: Mathematics (Version 9) (Australian Curriculum, Assessment Reporting Authority [ACARA], 2022) has one main content descriptor for learning to solve linear equations, placed at Year 7 “AC9M7A03 “Solve one-variable linear equations with natural number solutions; verify the solution by substitution” (https://acara.edu.au/curriculum).

Supporting this, the (non-compulsory) elaborations suggest that students might use substitution to determine whether a given number is a solution, or solve using *backtracking* (called *unwinding* in this paper), or balance equations perhaps illustrated with a concrete model. The only given example of an equation is $3x + 7 = 19$ which has one occurrence of the pronumeral, and a natural number solution.

These elaborations for AC9M7A03 point to three different strategies for solving equations, all of which significantly contribute to mathematical progress. In order of cognitive complexity, these are substituting, unwinding then balancing. Both substituting and unwinding are foreshadowed in Year 6 by AC9M6A02 which is about finding unknown numbers in numerical equations (e.g., an equation where a square placeholder is written instead of the unknown number). At Year 8, AC9M8A01 highlights algebraic manipulation including rearranging, simplifying linear expressions, and using inverse properties (presumably to solve equations). These capabilities open up the range of equations that can be solved, especially by the balancing method.

Literature Review

The fundamental research on the teaching of equation solving was conducted several decades ago, although related research continues (e.g., Linsell 2010; Knuth et al., 2006). At that time, researchers were interested in characterising the essential elements of ‘algebraic thinking’ with a goal to move teachers’ and researchers’ conceptualisation of the school topic of algebra away from emphasis on the manipulation of symbols towards understanding the fundamental concepts that students must develop. This movement led to the formalisation of an algebra strand in primary school curriculum descriptions. Within the topic of solving linear equations, Filloy and Rojano (1989) observed there was a ‘didactic cut’ between the demands of solving ‘arithmetic equations’ that can be solved by successively unwinding operations on numbers (e.g., $Ax \pm B = C$, $x/A = B$) and other ‘non-arithmetical equations’ like $Ax \pm B = Cx \pm D$ (for A, B, C, D positive numbers). As other researchers have, they showed that this ‘didactic cut’ was hard for students to cross, not just because of new procedures to be followed but because the second required a different way of thinking about equations. Kieran (1992) described this distinction as being (p. 392) between (i) procedural operations with “ostensibly” algebraic items, i.e., “arithmetic operations carried out on numbers to yield numbers” and (ii) structural operations “that are carried out, not on numbers, but on algebraic expressions” which can have the result being an algebraic expression. Herscovics and Linchevski (1994) referred to a similar ‘cognitive gap’ between arithmetic and algebra, (p. 59) which is “characterised as the students’ inability to spontaneously operate with or on the unknown”. They observed that Year 7 students' strategies and success changed depending on the location of the unknown in the equation, especially for subtraction and division. For example, the most frequent strategy for solving $n - 13 = 24$ was to add 13 to 24 (using the inverse operation), but $37 - n = 18$ was solved by a variety of strategies such as "complementary subtrahend" $(37 - 18)$ using knowledge of subtraction. They concluded that the difference in success solving these two similar equations resulted from students "working around the unknown at a purely numerical level" (p. 70). This literature provides a justification for the importance of this test, as well as the fundamental concepts that illuminate the three strategies and enable the design of the *Solving Linear Equations* test. Note that there is no indication of the 'cognitive gap' in the curriculum statements, so helping teachers understand it is a major goal of this test.

Analysis of Strategies

Table 1 analyses the three strategies, their roles, and the characteristics of items that the test uses to identify the strategies that students are able to use. In practice, experts use all these strategies, often interchangeably in the case of unwinding and balancing, but for beginners they are separate skills drawing on separate concepts. The various statements about students in the early stages of learning algebra in Table 1 draw on evidence from the literature and our pilot studies (Steinle, Price, & Stacey, 2022).

Table 1

Characteristics of Three Equation Solving Strategies

Feature	Substituting	Unwinding$^{\alpha}$	Balancing$^{\beta}$
General statement	A guess and check strategy, sometimes with an 'improve' step.	Read notation as specifying a sequence of operations applied to an unknown number.	A series of transformed equations including simplifying and reorganising steps.
Example: Solve $2n + 5 = 20$	Guess $n = 3$ Substitute: $2 \times 3 + 5 = 11$. Compare result: 11 is too small, try larger *n*, etc.	Unknown was multiplied by 2, then 5 was added to give 20. To undo, first subtract 5 (gives 15) and divide by 2 (gives 7.5)	$2n + 5 = 20$ $2n + 5 - 5 = 20 - 5$ $2n = 15$ $2n \div 2 = 15 \div 2$ $n = 7.5$
Operations used	Given operations	Inverse operations	Inverse operations
Operations on and with	Numbers	Numbers	Numbers and pronumerals
Applicability	Can be used to solve any equation, but beginners only successful with small natural number solution.	Beginners only successful with one occurrence of the variable easily seen to be at the start of the 'story'.	Powerful for a wide variety of equations.
Main teaching purposes	Reading algebraic notation; understanding what a solution is; introducing tables of values.	Introducing inverse operations; connecting to 'function machine' diagrams.	Widely applicable principle for equation solving.

$^{\alpha}$Used to describe a process which is applied without any preliminary algebraic manipulation.

$^{\beta}$Often referred to as do the same to both sides and process often abbreviated to change side-change sign.

Table 1 includes a simple example and then specifies various salient characteristics from the literature. The example does not show operating on or with pronumerals and expressions, which is required when there is more than one occurrence of the pronumeral (e.g., to solve $5x + 9 = 2x$). Not noted is the fact that the logic of balancing additionally requires a conception of operating structurally on a whole equation to gradually transform it to a form where the solution is obvious. To stress the importance of all the strategies to progress in algebra, the table also briefly outlines their curriculum functions. The applicability row is the key to the diagnosis made by the test. We exploit the limitations of the methods, in particular for beginners, to identify what strategies students can use.

Method

The present test is an outcome of considerable preliminary investigation, pilot testing, and data analysis as recorded by Steinle, Price, and Stacey (2022). Each item requires the student to type the solution to the equation. Table 2 gives the 14 equations, noting features of equation structure and the solution, as well as a label which indicates which of the five groups each equation belongs to. Students are allocated to one of five stages according to accuracy of responses to items in Groups A to D, as described in Table 3. Items in Group E provide another part of the report to teachers. In specific terms, the research question for this paper is to determine the extent to which the performance of the students diagnosed in each stage matches the predictions based on the strategies.

Space does not allow a thorough discussion of the choice of items, but we note a few general points. The pilot work and literature highlighted how much beginning students are affected by apparently small changes of algebraic structure (e.g., changing $14 - 2x$ to $2x - 14$). Therefore, to be able to attribute changed success to just the identified features of structure (and hence strategy), the items in the diagnostic Groups A, B, C and D are as close to the canonical linear form as possible. Because teachers want to know about how students deal with various algebraic notations, the Group E items were included but not used for the main diagnosis. The number of equations were reduced to a minimum, to keep the average time of completion under 15 minutes. This test is Version 1; there is also a parallel test (Version 2) which can be used as a post-test, which has been similarly analysed. Results are similar to those reported here.

The data was collected from the 3010 anonymous students in the classes of volunteer teachers who assigned this test during the years 2016 to 2018. Most students were in Year 8 or 9. In general, the teachers seemed to have allocated the test to the intended target group, since the median test score was 7/14. We have no information about the conditions for taking the test, which may have varied significantly. Some students had trouble typing fraction and decimal answers (e.g., typing 1\5 instead of 1/5 or 1.1.625 instead of 1.625). Teachers can access any completed SMART::test to see why a student has received an unexpected result.

Results

In this section, we examine the extent to which the performance of the students diagnosed in each stage matches the predictions based on the strategies. Table 2 shows which items are expected to be within the capability of students who can use each strategy. The validity of the tests requires evidence that students in each diagnosed stage can use the strategy specified in Table 3. Of course, beginning algebra students make many slips, so that giving an incorrect response is not always indicative of a lack of knowledge of a strategy. Hence we have used the word ‘feasible’ in Table 2 for an appropriate strategy rather than ‘correct’. However, we expect that the predicted successes of Table 2 should give a good guide to the observed success on all items, including Group E items not used for diagnosis.

Table 2

Items Showing Equations, Details of Structure and Applicability of Strategies (Test Version 1)

Label	Equation	Equation Structure: Pronumeral Location: Left, Right	Equation Structure: Explicit Oper-ations	Solution	Strategy Applicability: Substitute	Strategy Applicability: Unwind	Strategy Applicability: Balance
A1	$3x + 8 = 23$	1, 0	+	5	F	F	F
A2	$4x + 9 = 37$	1, 0	+	7	F	F	F
B1	$5x + 7 = 15$	1, 0	+	1.6	A	F	F
B2	$8x + 3 = 16$	1, 0	+	1.625	A	F	F
C1	$8x + 5 = 3x + 14$	1, 1	+ +	1.8	A	N	F
C2	$12x + 2 = 8x + 15$	1, 1	+ +	3.25	A	N	F
D1	$7x - 11 = 2x - 4$	1, 1	− −	1.4	A	N	F
D2	$12 - 11x = 5 - x$	1, 1	− −	0.7	A	N	F
E1	$7x - 2 = 16$	1, 0	−	$^{18}/_{7}$	A	F	F
E2	$14 - 2x = 8$	1, 0	−	3	F	N	F
E3	$3x + 6 + 2x = 7$	2, 0	+ +	0.2	A	N	F
E4	$\frac{x+2}{5} = 3$	1, 0	+ ÷α	13	F	F	F
E5	$\frac{x}{3} + 1 = 5$	1, 0	÷α +	12	F	F	F
E6	$4(x - 3) = 21$	1, 0	− ×α	8.25	A	A	F

F: strategy is feasible, A: strategy is awkward, N: strategy not possible without initial algebra.

αFraction or brackets may or may not be seen as ‘explicit’ operations.

Notes. E1 is only item with solution longer than three decimal places. Pronumeral *a* was used in Version 1.

Table 3

Rubric for Allocating Stages to Students Based on Scores on Groups of Items

Stage	Description of stages	Score on group of items: A	B	C	D	E
0	Not yet at Stage 1	0	−	−	−	−
1	Students can solve simple linear equations that are easy to solve by repeat substituting;	1,2	0	−	−	−
2	…and can solve linear equations with more difficult solutions requiring systematic strategy (such as unwinding);	1,2	1,2	0	−	−
3	…and can solve linear equations involving addition only, with the pronumeral on both sides (balancing strategy needed);	1,2	1,2	1,2	0	−
4	…and can further solve linear equations involving subtraction with the pronumeral on both sides and non-integer solutions.	1,2	1,2	1,2	1,2	−

Note. “−” indicates that group is not used in the stage diagnosis.

Table 4 shows the accuracy for each item by stage, reporting two measures: Facility 1 is the percentage of all students who are correct on an item and Facility 3 is the percentage of the students who respond who are correct. If there are any omissions, Facility 3 is larger. When reading the tables, it is useful to note that algebraic manipulation shows the omission rate for an item is (Facility 3 − Facility 1) / Facility 3. If Facility 3 − Facility 1, there are no omissions; if Facility 3 = 1.5 × Facility 1, one third of students omitted; and if Facility 3 = 2 × Facility 1, the same number of students omitted the item as responded. Omissions generally increase later in the test although students also look through the test selecting items that look easy. (See Steinle, Stacey, et al. (2022) if you are curious about Facility 2.)

Table 4

Facility Range (Facility 1~Facility 3) of Test Items by Diagnosed Stage

Label	**Equation**	**Solution**	**Diagnosed stage**				
			Stage 0 (*n*=338)	**Stage 1 (*n*=575)**	**Stage 2 (*n*=762)**	**Stage 3 (*n*=287)**	**Stage 4 (*n*=1048)**
A1	$3x + 8 = 23$	5	rubric 0%	94%~94%	96%~96%	98%~98%	97%~97%
A2	$4x + 9 = 37$	7	rubric 0%	87%~89%	95%~96%	95%~95%	97%~97%
B1	$5x + 7 = 15$	1.6	3%~4%	rubric 0%	95%~95%	97%~97%	97%~97%
B2	$8x + 3 = 16$	1.625	3%~3%	rubric 0%	83%~88%	87%~90%	95%~95%
C1	$8x + 5 = 3x + 14$	1.8	6%~9%	12%~22%	rubric 0%	90%~91%	96%~96%
C2	$12x + 2 = 8x + 15$	3.25	5%~8%	7%~15%	rubric 0%	63%~76%	94%~94%
D1	$7x - 11 = 2x - 4$	1.4	4%~7%	6%~13%	3%~9%	rubric 0%	92%~93%
D2	$12 - 11x = 5 - x$	0.7	3%~5%	5%~10%	3%~9%	rubric 0%	69%~72%
E1	$7x - 2 = 16$	$^{18}/_{7}$	4%~6%	11%~20%	40%~61%	53%~70%	81%~83%
E2	$14 - 2x = 8$	3	7%~10%	27%~46%	33%~54%	39%~52%	77%~81%
E3	$3x + 6 + 2x = 7$	0.2	2%~4%	7%~15%	19%~44%	39%~61%	81%~87%
E4	$\frac{x+2}{5} = 3$	13	9%~16%	23%~46%	35%~68%	52%~78%	79%~87%
E5	$\frac{x}{3} + 1 = 5$	12	10%~19%	22%~46%	34%~66%	46%~70%	76%~84%
E6	$4(x - 3) = 21$	8.25	3%~6%	5%~11%	18%~39%	37%~59%	72%~80%

The shading (from Table 2) indicates the applicability (F, A, N) of the hypothesised strategy for each stage. 'Rubric 0%' indicates the item must be incorrect for student to be allocated to that Stage (as in Table 3).

Stages 3 and 4: Do Responses Match the Balancing Strategy?

Stage 3 and 4 students were predicted (Table 2) to do well on all items (except Stage 3 on Group D). They have very high performance on Group A, B and C and higher performance than other stages on all other items. After A1, Stage 4 students had a somewhat higher facility range than Stage 3 on every item. We see that Stage 4 students are more secure than Stage 3 in their balancing strategy (e.g., comparing Group C facilities) and the Group E facilities show they are better able to manage algebraic variations. Importantly, the facility ranges for Group D reveal the importance difference between subtraction of a number (D1) and subtraction of a pronumeral (D2) even for these relatively accomplished students (as noted by Herscovics & Linchevski, 1994). This difference is also apparent in the decrease in facility between E1 and E2 for students in Stages 2, 3 and 4. Only Stage 0 and Stage 1 students had higher facility on E2 than on E1, supporting the hypothesis that these students are substituting numbers rather than working with the algebra and hence found the item with a natural number solution easier.

Stage 3 students' responses generally fit the predicted pattern, with high success on items in Groups A, B, and C, and moderate success on Group E. We have no good explanation for the large drop in facility for C2 compared to C1. Overall, the Stage 3 pattern is consistent with students still developing the balancing strategy and the algebraic skills that are necessary to manage variations in algebraic structure (see, for example E3 and E6).

Stage 2: Do Responses Match the Unwinding Strategy?

The *a priori* analysis in Table 2 predicted the unwinding strategy would give success with eight items but, assuming no prior algebraic manipulation to put the equation in an amenable form, it would fail for six items (C1, C2, D1, D2, E2, E3) and that E6 would be difficult. The facility ranges for the Group A, B, C and D items match the prediction. Items E1, E4, E5 were predicted to be feasible with the unwinding strategy (Table 2), and around two thirds of those Stage 2 students who attempted these were successful. Item E6 can also be solved by unwinding, but success relies on reading the brackets notation appropriately; first 3 is subtracted from the unknown number then the answer is multiplied by 4. Perhaps this is why the facility was much lower. Items E2 and E3 cannot be solved by unwinding without preliminary algebraic manipulation. After collecting like terms, E3 belongs to Group B where this group of students have already shown very high facility, but only 44% of those who responded to E3 were successful and the omission rate was high. This is consistent with students learning unwinding strategies with clear stories of what happened to the unknown number, written in algebra (e.g., E4, E5, A, and B items), and not yet doing preliminary algebraic manipulation. Stage 2 students did much better on item E2 than predicted with nearly half being successful; we assume they turned to substitution (natural number answer), which may also have boosted facilities of E4 and E5. The omission rate was high for all E items, supporting the picture of these students just beginning to learn equation solving, still having little exposure to variations of algebraic form.

Stage 1: Do Responses Match the Substituting Strategy?

The *a priori* analysis in Table 2 predicted the substituting strategy would be successful when solutions were small natural numbers, but in other cases would be awkward for beginners. Locating a non-integer solution takes a relatively long time and is prone to error because more complex arithmetic is required. This is well supported by the data. All the items with Facility 3 over 40% have small natural number solutions. All other facilities are under 25%. The omission rate is high at around one third after Group A items, which is consistent with students taking a long time to complete the questions. Further support for use of substituting was presented by the error analysis (Steinle, Stacey et al., 2022). Stage 1 students often gave an answer of 1 more or less than the correct answer (so a numerical slip) or answered N.5 when the correct answer was between N and N + 1 (e.g., to give an answer 3.5 instead of 3.25). We hypothesise that this response was used to indicate that they know the answer is somewhere between 3 and 4.

Stage 0: What do These Students Know?

Students were put into Stage 0 if they answered neither A1 nor A2 correctly. It is to be expected that a small number of capable students made slips (mathematical or response entry) on both A1 and A2 and therefore are misdiagnosed by the strictly hierarchical rubric we have used (see Table 3). A small rate of misdiagnosis is supported by the observation that every item in Groups B, C and D was answered correctly by at least 3% of the Stage 0 students. The items with the highest facility for Stage 0 students are E2, E4, and E5, all of which have small natural number solutions, indicating some students were substituting. For E4 and E5, around 10% of students were correct. These are the only items that do not include implicit multiplication. We conclude that at least 10% of the students diagnosed as Stage 0 are using substitution (and we estimate about a half), but they do not reliably interpret implicit multiplication. In support of this, the error analysis (Steinle, Stacey et al., 2022) found that incorrect responses consistent with interpreting mx as $m + x$ were given by over one-third of Stage 0 students.

Conclusion

The purpose of this paper was to demonstrate that a short online test can diagnose students' thinking sufficiently well to help a teacher plan teaching. The constraints were that the test must be short, student responses must be machine-readable, and the diagnosis given to the teacher must be easily understood and important. The literature review and curriculum analysis demonstrated the importance of the strategies for solving linear equations. This paper illustrated the contribution of analysis of overall accuracy to validating the resulting diagnoses; error analysis reported elsewhere supplements this. The match between the theoretical predictions of item facility and the actual facilities provided evidence that most students in Stage 1 use substituting only, most students in Stage 2 can (also) use unwinding and that students in Stages 3 and 4 can use balancing, at least with basic algebraic structures. The distinction between Stage 3 and 4 students was not as clear-cut as expected, so could be refined. The test results also highlighted the importance of students frequently working with linear equations that are not written in the canonical ax + b form.

Acknowledgements

We thank Dr Eugene Gvozdenko for his technical work in delivering this test and results and the teachers of the anonymous students who contributed the data. Ethics approval 1239210.2 was granted by the University of Melbourne and teachers provided their consent for the data to be used. The SMART::tests have been used in commercialised products, in which the authors have no ongoing financial stake.

References

Filloy, E. & Rojano, T. (1989). Solving equations, the transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25.

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics*, *27*(1), 59–78.

Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 390–419). New York: MacMillan Publishing.

Klingbeil, K., Rösken, F., Barzel, B., Schacht, F., Stacey, K., Steinle, V. & Thurm, D. (2024). Validity of multiple-choice digital formative assessment for assessing students' (mis)conceptions: Evidence from a mixed-methods study in algebra. *ZDM Mathematics Education*. https://doi.org/10.1007/s11858-024-01556-0

Knuth, E. J., Stephen, A. S., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equals sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*, 297–312.

Linsell, C. (2010). Secondary numeracy project students' development of algebraic knowledge and strategies. *Findings from the New Zealand numeracy development projects 2009* (pp. 100–116). Wellington: Ministry of Education.

Stacey, K., Steinle, V., Price, B., & Gvozdenko, E. (2018). Specific mathematics assessments that reveal thinking: An online tool to build teachers' diagnostic competence and support teaching. In T. Leuders, J. Leuders, & K. Philipp (Eds.), *Diagnostic competence of mathematics teachers. Unpacking a complex construct in teacher education and teacher practice* (pp. 241–262). New York: Springer.

Steinle, V., Price, E., & Stacey, K. (2022). *SMART::test solving linear equations: development and student results* [Technical report]. University of Melbourne. https://doi.org/10.26188/20278920

Steinle, V., Stacey, K., & Price, B. (2022). Beyond accuracy: A process for analysis of constructed responses in large datasets and insights into students' equation solving. *Journal of Educational Research in Mathematics*, *32*(3), 1–28. https://doi.org/10.29275/jerm.2022.32.3.1