“p stands for packs” – diagnosing the letter as object misconception with SMART online tests

Katrin Klingbeil, Fabian Rösken, Bärbel Barzel

"*p* stands for packs" – diagnosing the *letter as object* misconception with SMART online tests

Katrin Klingbeil, Fabian Rösken and Bärbel Barzel

University of Duisburg-Essen, Germany; katrin.klingbeil@uni-due.de

Research on the challenges of learning algebra is extensive. Various typical errors and misconceptions, like the 'Letter as Object' misconception, have been identified. Making use of such findings for formative assessment is the core idea of the SMART online tests ('Specific Mathematics Assessments that Reveal Thinking') that have been developed at the University of Melbourne. Adapting the test 'Meaning of Letters' for use in Germany, we are investigating if these multiple-choice items indeed reveal the misconceptions of German students adequately. Based on student interviews, we added two open-ended items to the test and analysed the written explanations of an initial sample of 600 German grade 7 and 8 students. The analysis shows a satisfactory concordance between the automatic SMART diagnosis and the student explanations, and that the 'Letter as Object' misconception was highly prevalent among those students.

Keywords: Algebra, variables, letter as object misconception, online formative assessment.

Introduction

Variables are a fundamental concept in algebra which appears to be challenging for students to grasp. Many, for example, struggle with overcoming the *Letter as Object* misconception. An impactful possibility to support students in their learning is through formative assessment. For this, digital diagnostic tools like the SMART tests can offer a convenient solution if they are able to provide an adequate diagnosis of students' understanding. However, even if test items are carefully designed based on research, multiple-choice responses might be chosen for other than the intended reasons which could lead to inadequate diagnoses. Therefore, it is important to examine if multiple-choice items, especially after being translated and adapted for use in a different country, indeed reveal the misconceptions of assessed students adequately.

Theoretical background

Understanding algebraic letters

Drawing on a vast body of literature, Arcavi, Drijvers and Stacey "distinguish five facets of the concept of variable: a placeholder for a number, an unknown number, a varying quantity, a generalized number, and a parameter" (2017, p. 12). Across these facets, variables stand for or refer to one or more numerical values. However, algebra learners often struggle with this numerical interpretation and various typical errors and misconceptions have been identified. Steinle et al. (2009) distinguish between numerical and non-numerical misconceptions: incorrect numerical ways of thinking include, for example, the belief that the value of a letter is somehow connected with the alphabetical position of that letter, or the notion that different letters need to stand for different numbers. Examples for non-numerical misconception are *Letter not used* and *Letter as Object* (see Küchemann, 1981). The latter misconception has been repeatedly reported for decades (e.g. Akhtar & Steinle 2017; Küchemann, 1981; Rosnick, 1981). While various terms are used for (aspects of) this phenomenon,

such as *Letter as abbreviation* or *Letter as unit* (Akhtar & Steinle, 2017), for this paper, we follow the terminology of Küchemann who has described the *Letter as Object (LO)* misconception as the letter being "regarded as a shorthand for an object or as an object in its own right" (1981, p. 104). A well-known example for a problem eliciting *LO* was presented to 150 entering engineering students by Rosnick (1981):

> Write an equation, using the variables *S* and *P* to represent the following statement: "At this university there are six times as many students as professors." Use *S* for the number of students and *P* for the number of professors. (ibid., pp. 418f.)

37% of the students were not able to give a correct answer, the most common error being the reversed equation $6S = P$. Interviews supported the assumption that this error often roots in the incorrect interpretation of variables as standing for students respectively professors rather than for the number of students respectively the numbers of professors. Thus, the incorrect equation could be read as "there are six students for every one professor" (Rosnick, 1981, p. 419). Arcavi, Drijvers and Stacey (2017) emphasise that misleading teaching (e.g. the so-called "fruit salad algebra" where a stands for apples and b for bananas) can encourage this kind of interpretation and lead to serious problems for example when setting up equations.

As part of the foundational CSMS study on the mathematical understanding of secondary school students in the United Kingdom, Küchemann (1981) used an item that, in contrast to Rosnick's task, included addition (in the following referred to as "pencil" item):

> Blue pencils cost 5 pence each and red pencils cost 6 pence each. I buy some blue and some red pencils and altogether it costs me 90 pence. If b is the number of blue pencils bought and if r is the number of red pencils bought, what can you write down about b and r? (ibid, p. 107)

While only 10% of tested 14-year-old students provided the correct equation $5b + 6r = 90$, 17% gave $b + r = 90$ as an answer which might have been read as "blue pencils and red pencils together cost 90 pence" (*LO*). Interpreting b as "the number of blue pencils" is a possibility here, too; however, this would still imply a wrong understanding of equations with a number of pencils on one side of the equation, and the price of the pencils on the other. Interestingly, 6% of the students came up with another kind of equation: $6b + 10r = 90$ or $12b + 5r = 90$. These students had figured out a possible solution to the problem first and then used these values as coefficients in their equation (Küchemann, 1981). Since the letters are used as abbreviations for the involved objects ("12 blue pencils and 5 red pencils together cost 90 pence"), this is regarded as a special form of *LO*, which we will refer to as the *Solution as Coefficient* (*SAC*) misconception in the following.

Decades after the CSMS study, studies using Küchemann's items have resulted in similar findings on students' understanding of algebraic letters (e.g. Hodgen et al., 2010; Oldenburg, 2010).

Formative assessment with SMART online tests

SMART online tests have been developed at the University of Melbourne, Australia, since 2008 based on research. SMART enables easy provision and processing of diagnostic tasks. Since students also interact with the SMART system through inputs, in line with the framework suggested by Aldon et al. (in press), in terms of communication SMART would be categorised as assessment *through* and

with technology. Much more important, however, is the analysis: SMART offers an *extended analysis* that automatically analyses patterns between individual diagnostic items (Steinle et al., 2009) and thus allows insights into students' understanding by flagging levels of understanding and misconceptions of individual students. In addition, explanations, tasks, and teaching suggestions for the individual misconceptions and levels of understanding are provided to support teachers in planning targeted interventions. Since it is the teacher who makes these decisions, according to Aldon et al. (in press), SMART offers a *passive adaptation*.

In order to keep tests short, to make it easy to enter answers (e.g. without formula editors), and to enable a quick automatic analysis of student answers, SMART tests predominantly use multiple-choice (MC) items. Despite several known challenges of MC items, such as corrective feedback, working backwards, random guessing (Bridgeman, 1992), or the concern that the reasoning behind made choices cannot be assessed directly, SMART developers claim that MC items can be designed in a way to reveal students' thinking by analysing response patterns since "in contrast to careless errors, misconceptions [...] lead to predictable errors in student work" (Akhtar & Steinle, 2013, p. 36).

Following this premise, the SMART test *Meaning of Letters* has been developed to assess whether students know that letters, when used in algebra, stand for numbers. Three stages and *SAC*, a more sophisticated way in which students can exhibit the *LO* misconception, are reported by the automatic SMART analysis (see Table 1 for a shortened description). To be diagnosed at Stage 2, a student needs to answer all items correctly; for Stage 1, at least two correct answers are required. The misconception *SAC* is flagged additionally if one or more *SAC*-response options are chosen.

Table 1: Test *Meaning of Letters* – Stages of understanding and misconception

Stage	Stage 0	*LO* misconception in most items, rarely interpreting algebraic letters as standing for numbers
	Stage 1	Sometimes algebraic letters correctly interpreted as standing for numbers and sometimes *LO*
	Stage 2	Algebraic letters consistently interpreted correctly as standing for numbers, rather than as objects.
Miscon.	SAC	Solution as coefficient – These students believe that the coefficients in an equation indicate the solution to the problem and that the variables are shorthand for the objects involved; the equation is read as a kind of solution sentence.

The test contains six items of which two are adaptations of Küchemann's "pencil" item (Akhtar & Steinle, 2017), two are similar to the "Students and Professor" problem and two items ask students to explicitly decide on the meaning of an algebraic letter in a given equation describing a situation. In this paper, we focus on the item in Figure 1.

Biros are sold in packs of 3. Sam bought p packs and got b biros altogether. Choose the correct equation:  ☐ $b + p = 4$ ☐ $p = 3b$ ☐ $p = 3$ ☐ $3p = b$ ☐ $30b = 10p$	*Possible interpretation:* → One biro plus a pack of biros is 4 altogether. (*LO*) → A pack contains 3 biros. (*LO*) → A pack has 3. (*LO*) → Correct → Sam bought 10 packs and has 30 biros now. (*SAC*)

Figure 1: Item with possible interpretations of response options

In this item, similar to the "Students and Professor" problem, the pronumerals are the first letter of both involved objects which might make it more difficult to avoid an *LO* interpretation. Additionally,

the item stem uses the wording "Sam bought p packs" instead of, for example, "p is the number of packs Sam bought". Both potential hurdles are intentional as the test aims to inform teachers about indications of any misconceptions that might be present and not to prevent students from making this mistake. The last response option ($30b = 10p$), involving coefficients which are a possible solution to the problem, is derived from responses to Küchemann's "pencil" item. Note that the second option ($p = 3b$), despite omitting 1 as a coefficient, could also be interpreted as giving a solution ("Sam bought 1 pack with 3 biros"). In both cases, the letters are nevertheless interpreted as abbreviations for objects (biros, packs).

Methods

The presented research is part of the larger research project SMART[alpha] that is currently investigating how and the extent to which teachers' competencies, as well as students' understandings, develop through the use of SMART tests, exemplarily on the topic of understanding variables (Klingbeil et al., 2022a). Since automatic diagnosis is at the core of the SMART system, the underlying test items and the evaluation logic are examined more closely. In this paper, we focus on the research question whether the multiple-choice items of the SMART test *Meaning of letters* adequately capture the *Letter as Object* misconception of German 7th and 8th grade students. This question encompasses two angles on the term "adequately":

(1) If a certain misconception is detectable in the student's explanation, will the automatic diagnosis report this?
(2) If a certain misconception is reported by the diagnosis, how likely is it that this misconception will also be detectable in the student's explanation?

In a pilot, one class of German 8th grade students had been asked to fill in two SMART tests online, followed by interviews with selected students (n=5 for the *Meaning of Letters* test). The comparison of the automatic diagnosis and the qualitatively analysed interviews had shown a high concordance with regards to the diagnosed misconceptions; however, a few deviations, indications of guessing in MC items, and additional student difficulties (e.g. the invisible multiplication sign between coefficient and algebraic letter) had become apparent (Klingbeil et al., 2022b). Based on these results, we have decided to add two open-ended items to the original *Meaning of Letters* test, asking students to give reasons for their choice in the previous multiple-choice item. Thus, we aim to gain more insight into students' thinking and check the accuracy of the automatic diagnosis on a large scale. As part of our on-going main study, the test is being administered to 7th and 8th grade students across Germany. For this paper, we include initial data from 600 students who have completed the SMART test (version A) in December 2022 and January 2023 at the beginning of a teaching unit about variables, algebraic expressions and/or equations.

In order to compare the automatic diagnosis and students' written explanations, a first open coding procedure has been applied by two raters to code the explanations with regards to underlying (mis)conceptions. Afterwards, code definitions have been refined and deviations discussed until agreement between the two raters was reached (see Table 2 for the developed codes).

Table 2: Codes used for explanations

Code	Short explanation	Example
LO	clearly Letter as Object (*LO*)	"Because p stands for packs and b for all the biros." (p=3b)
LO-am	most probably LO, but ambiguous; sometimes partly correct	"Because there are three biros in one pack." (p=3b)
SAC	Coefficients in equation are interpreted as one/the solution (*SAC*), without explicit *LO*	"If you calculate 30:3 you get 10. Since it is sold in packs of three, I divided by 3." (30b=10p)
SAC+LO	SAC and explicit *LO*	"In 1 pack there are 3 biros. Everything times ten that makes 30 biros in 10 packs." (30b=10p)
CR	correct reasoning	"The number of packs times three gives the number of biros." (3p=b)
CR-am	correct in principle, but ambiguous	"The three stands for one pack and the b for the total number of biros." (3p=b)
NC	not clear/ambiguous (e.g. contradicting or partly erroneous)	"Since the 3 can be replaced by any number and thus it is shown 3/4/5/...p are equal to 9/12/15/... b" (3p=b)
OTH	other explanation (not LO, SAC or CR)	"Because the result must be b and this was the only option with b as the result." (3p=b)
NME	no meaningful explanation (no reason, guessed, nonsense or no answer at all)	"Because that's how it is"

First results

For a first overview, we consider the automatic diagnosis provided by the SMART system:

Table 3: Automatic SMART diagnosis by grade

Stage	All students (n=600)	Grade 7 (n=228)	Grade 8 (n=372)
0	61.2%	67.1%	57.5%
1	37.0%	32.0%	40.1%
2	1.8%	0.9%	2.4%

Additional misconception	All students (n=600)	Grade 7 (n=228)	Grade 8 (n=372)
SAC	68.7%	70.2%	67.8%
SAC at Stage 0	50.5%	55.3%	47.6%
SAC at Stage 1	18.2%	14.9%	20.2%

Remarkably, only 1.8% of all students have been diagnosed at Stage 2, i.e. all the other students have chosen at least one response option related to the *LO* misconception (see also Table 1). While this is not surprising for grade 7 students who are regularly encountering formal variables for the very first time at this point, grade 8 students already should have had opportunities to develop a viable understanding of algebraic letters. Hence, the difference between the two age groups regarding the diagnosed stages was to be expected (significant at $p < .05$ with $\chi^2(2, N = 600) = 6.42,\ p = .04$).

Comparing the automatic diagnosis with the coded explanations (see Table 4), the two different perspectives mentioned before are considered:

(1) If a certain misconception is detectable in the student's explanation, will the automatic diagnosis report this?
(2) If a certain misconception is reported by the diagnosis, how likely is it that this misconception will also be detectable in the student's explanation?

Table 4: Coded explanations by SMART diagnosis

	LO (n=255)	LO-am (n=107)	SAC+LO (n=45)	SAC (n=13)	CR (n=13)	CR-am (n=11)	NC (n=19)	OTH (n=9)	NME (n=128)
Stage 0 (n=367)	178	63	35	10	0	3	5	4	69
Stage 1 (n=222)	76	43	10	3	8	6	12	5	59
Stage 2 (n=11)	1	1	0	0	5	2	2	0	0
SAC (n=412)	184	80	45	12	4	5	6	6	70

Regarding perspective (1), it was found that students with *LO* and *LO-am* coded explanations (n=362) were all diagnosed at Stage 0 or 1 (which implies *LO* by definition) with the exception of one student each being at Stage 2. For explanations being coded as *SAC+LO* or *SAC* (n=58) the outcome was similar (with no student on Stage 2). Additionally, 98.3% of these students were flagged as having *SAC* by the SMART system. Of students with clearly correct explanations (*CR*, n=13) 38.5% were diagnosed at Stage 2, the rest at Stage 1.

In terms of perspective (2), Stage 2 students (n=11) wrote as many correct (*CR*) as ambiguous explanations (*CR-am*/*NC*/*LO-am*); furthermore, one student answer has been coded *LO*. While at Stage 2 no students have been found giving non-meaningful explanations, at Stage 0 18.8% and at Stage 1 26.6% of student answers fall into that category. Apart from that, Stage 0 students (n=367) gave predominantly *LO*-related (*LO*/*LO-am*/*SAC+LO*) reasons (75.2%) and no correct explanations. Of Stage 1 students' explanations (n=222) 58.1% were *LO*-related and 6.3% (rather) correct (*CR*/*CR-am*). While the diagnosed stages and the students' explanations tend to agree, the picture is different when looking at the *SAC* diagnosis: Only 13.8% of explanations from *SAC* diagnosed students have been coded as *SAC* or *SAC+LO*.

Considering the response options in more detail, it becomes apparent that students who chose $p = 3b$ as an answer (n=341) were very likely to also give an *LO*-related explanation (97.0% of 305 meaningful explanations, *LO*/*LO-am*/*SAC+LO*). Whereas this is similar for $p = 3$ (n=31; 95.2% of 21 meaningful explanations *LO*-related), of students who chose $b + p = 4$ (n=38) only 5 explanations have been coded as *LO*, while 27 explanations have been coded as non-meaningful. Students who chose the intended *SAC*-option $30b = 10p$ (n=78) showed *SAC* or *SAC+LO* in 91.7% of 60 meaningful explanations. The correct option $3p = b$ was chosen by 105 of 600 students; however, their explanations revealed a variety of reasons: 32.0% have been coded as (rather) correct (*CR*/*CR-am*), 34.7% as *LO*, 28.0% as ambiguous (*LO-am*/*NC*) and 5.3% as other (of 75 meaningful explanations).

Discussion & Outlook

Regarding the potential of SMART tests to provide teachers with information on their students' thinking by reporting existing misconceptions (perspective (1)), the results of our first analysis show a very high concordance between the SMART diagnosis and students' explanations for students exhibiting the *LO* or *SAC* misconception in their explanation: those students were almost exclusively diagnosed at Stage 0 or 1, and, if applicable, *SAC* additionally was flagged. Students with correct explanations were appropriately diagnosed at Stage 1 or 2.

Taking perspective (2), it could be observed that the majority of students being diagnosed at Stage 0 and 1 indeed gave *LO*-related explanations. Stage 2 students, however, partly still showed misconceptions or gave unclear reasons, so that teachers unfortunately could not rely on these students having a completely correct understanding. It is important to remember, though, that in this study only 11 students were diagnosed at Stage 2, so that no generalisations should be derived. Concerning *SAC* flagged students a discrepancy to their explanation becomes visible: most of them did not give any *SAC*-related reasons for their choice of response option. While this makes clear that it is not possible to draw conclusions from the automatic diagnosis to a certain explanation for one single item, it does not necessarily mean that the diagnosis is inadequate. Since choosing an *SAC* option in just one item

is enough to get an *SAC* diagnosis, it is possible that these students simply did not choose the *SAC* option in this investigated item but would show this misconception in one of the other items. In fact, we found that only 78 of 412 *SAC* diagnosed students did choose the *SAC* response option in this item (most of whom also gave *SAC*-related reasons). Thus, this particular item might not be as predictive for the *SAC* misconception as other items of the test. Consequently, the automatic diagnosis should not be tied to respectively compared to only one single item.

Principally, response options mainly seem to be chosen for the reasons assumed when designing the responses. This does not apply for the correct option and option $b + p = 4$ though. While the correct option was also chosen for *LO* reasons (interpreted as "One 3-pack contains biros"), only a few of the explanations for $b + p = 4$ were coded as the expected *LO* misconception. Remembering that the diagnosis is not resulting from only one item, it could be argued that this deviation is not problematic (especially when SMART is used formatively). A student choosing the correct option for the wrong reasons in this item would probably still show their *LO* misconception in items explicitly asking for the meaning of the letter. For option $b + p = 4$, however, it should be discussed if this response might not be included for the *LO*-diagnosis.

All in all, we observed a satisfactory concordance between the automatic SMART analysis and the misconceptions apparent in students' explanations, especially regarding perspective (1), since students with *LO* or *SAC* related explanations got flagged accordingly. Results for perspective (2), however, reinforce that SMART tests should not be used summatively because a diagnosed stage is no definite guarantee for a misconception being present respectively absent, but rather an indication for a possibly existing misconception that should be verified by further diagnosis. Thus, for a formative purpose, the automatic diagnosis is sufficient and can help teachers adapt their teaching to their students' needs. That this is necessary can be seen in the very high prevalence of the *LO* misconception among 7th and 8th grade students (according to the automatic diagnosis) that exceeded our expectations of 50–80% which were based on results from a similar SMART test with 1,400 Australian secondary school students (Akhtar & Steinle, 2017).

As a limitation, it needs to be considered that the students' free responses were mostly very short and often not unequivocally clear, making it difficult to distinguish between linguistic inaccuracies and manifest misconceptions. Therefore, also "ambiguous" codes were used. However, interpreting also ambiguous, *LO-am* coded explanations as indicating the *LO* misconception is corroborated by results from interviews conducted with five students in the pilot. For this first analysis of our initial data, we only included students' responses to one of the items, the automatic diagnosis, and their grade. In a next step, also the other five items, including another explanation, and further control variables, as well as parallel items in the test's B version will be taken into account.

References

Akhtar, Z., & Steinle, V. (2013). Probing students' numerical misconceptions in school algebra. In V. Steinle, L. Ball, & C. Bardini (Eds.), *Mathematics education: Yesterday, today and tomorrow – Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia* (pp. 36–43). MERGA.

Akhtar, Z., & Steinle, V. (2017). The prevalence of the 'letter as object' misconception in junior

secondary students. In A. Downton, S. Livy, & J. Hall (Eds.), *40 years on: We are still learning! – Proceedings of the 40th annual conference of the Mathematics Education Research Group of Australasia* (pp. 77–84). MERGA.

Aldon, G., Barzel, B., Cusi, A., & Olsher, S. (in press). Rethinking teachers' formative assessment practices within technology-enhanced classrooms. In B. Pepin, G. Gueudet, & J. Choppin (Eds.), *Handbook of digital resources in mathematics education*. Springer.

Arcavi, A., Drijvers, P., & Stacey, K. (2017). *The learning and teaching of algebra: Ideas, insights, and activities*. Routledge. https://doi.org/10.4324/9781315545189

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement*, *29*(3), 253–271. https://doi.org/10.1111/j.1745-3984.1992.tb00377.x

Hodgen, J., Küchemann, D., Brown, M., & Coe, R. (2010). Children's understandings of Algebra 30 Years on: What has Changed? In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education (CERME 6)* (pp. 539–548). Institut National de Recherche Pédagogique and ERME.

Klingbeil, K., Rösken, F., Barzel, B., Thurm, D., & Schacht, F. (2022a). SMART – online formative assessment: Professionalising teachers & enhancing students' understanding. In J. Hodgen, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)* (pp. 2781–2782). ERME / Free University of Bozen-Bolzano.

Klingbeil, K., Rösken, F., Barzel, B., & Schacht, F. (2022b). Potential and challenges of SMART as an online diagnostic tool in comparison to diagnostic interviews using the example of understanding variables. In H.-G. Weigand, A. Donevska-Todorova, E. Faggiano, P. Iannone, J. Medová, M. Tabach, & M. Turgut (Eds.), *MEDA3 Mathematics Education in Digital Age 3. Proceedings of the 13th ERME Topic Conference (ETC13) held on 7–9 September 2022 in Nitra, Slovakia*. (pp. 208–215).

Küchemann, D. (1981). Algebra. In K. M. Hart, M. L. Brown, D. E. Küchemann, D. Kerslake, G. Ruddock, & M. McCartney (Eds.), *Children's Understanding of Mathematics: 11–16* (pp. 102–119). John Murray.

Oldenburg, R. (2010). Structure of algebraic competencies. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education (CERME 6)* (pp. 579–588). Institut National de Recherche Pédagogique and ERME.

Rosnick, P. (1981). Some misconceptions concerning the concept of variable. *The Mathematics Teacher*, *74*(6), 418–420. https://doi.org/10.5951/MT.74.6.0418

Steinle, V., Gvozdenko, E., Price, B., Stacey, K., & Pierce, R. (2009). Investigating students' numerical misconceptions in algebra. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia (Vol. 2)* (pp. 491–498). MERGA.