Do fuzzy-logic non-linear models provide a benefit for the modelling of algebraic competency?

Reinhard Oldenburg

Do fuzzy-logic non-linear models provide a benefit for the modelling of algebraic competency?

Reinhard Oldenburg[1]

[1]Reinhard Oldenburg (Math education, Math department, Augsburg University, Germany)

Abstract

Statistical models used in mathematics education are often linear and latent variables are often assumed to be normally distributed. The present paper argues that by relaxing these constraints one may use models that fit the data better than linear ones and provide more insight into the domain. It combines research on statistical methodology with research on the competence structure within algebra. The methodological innovation is that models with latent variables from the unit interval are considered which allows to model relations by means of fuzzy logic. Estimation techniques for such models are discussed to the extend necessary for the present study. To assess the benefit of this modelling technique data from an algebra test is re-analyzed. It is shown that non-linear models have greater explanatory power and give interesting didactical insights. Moreover, model comparison allows to differentiate between different theoretical constructs related to algebraic understanding. Finally, a research program is outlined that aims at the development of a universal algebra competence model that can be applied to test data from various algebra tests.

Keywords: Competence model; Non-linear Model; Algebraic competences; Estimation technique.

Introduction

It is potentially beneficial for teaching to have an evidence-based understanding of the different component of algebraic competency. There are many theoretical constructs that aim at structuring school algebra: Usiskin (1988) defined four conceptions of algebra, Kieran (2004) has distinguished generational activities, transformational activities, and global/meta-level activities. Even more specific are many classifications of what variables are and how they are used. Küchemann (1979) has describe six ways to treat letters in algebraic formulas, Arcavi (1994) has related the understanding of variables and expressions, Epp (2011) has investigated the usefulness of different understanding of variables and (Oldenburg, 2019) is a classification approach for variables. The same holds true for other algebraic concepts, e.g. Bardini et al. (2013) distinguish different uses, and thus different understandings, of the equal sign, Mason and Sutherland (2002) introduce key aspects of school algebra. This list could easily be prolonged much more but this should suffice to give support for the following observations: There are many conceptual approaches to structure algebraic competencies, but statistical models are usually restricted to special cases, i.e. tests are constructed to explore the constructs of a special conception and it is usually not possible to apply one model to data from other tests. For example, an item that was designed to measure a latent construct from one theory may load on two latent constructs or even their combination from another theory. The present paper is the first in a planned series that aims at improving the situation by enhancing the flexibility of modelling. The full program will be outlined at the end of this paper.

Many statistical models used in algebra education research come from the class of structural equation models (SEM, see Hoyle 2012). Traditionally, these models haven been linear and non-linear versions gain only slowly support and are often restricted to very special model classes (Dijkstra & Schermelleh-Engel, 2014; Kelava et al., 2011; Umbach et al., 2017). The key innovation of the present paper is to use a more flexible estimation technique that allows to fit models even to data sets that provide no perfect measurement models for the constructs. The paper first describes this method, then the context from algebra education and finally applies the method to a test data set. The purpose is twofold: First, the algebraic test data are a benchmark to judge the usefulness of the estimation method on real world data. Second, the outcomes from fitting the model allow some conclusions that are of interest for algebra education. To this end, data from an algebra test is re-evaluated.

Hence, the research questions are the following:

- RQ1: Does a non-linear model fit the data from the algebra test better than a linear model?

- RQ2: Does the non-linear model give interesting and plausible insights for the understanding of algebraic thinking?
- RQ3: Can non-linear cross-loading rich models be used to analyze tests that don't fit the model well?

Materials and Methods

This section describes the new method that will be applied to algebra test data in later sections. The motivation is to have the greatest possible flexibility in modelling. Especially, linear structural models or their non-linear extensions are not capable of describing the models that will be applied here.

This section first describes the general model class (further details are given in Oldenburg (2020), and in Oldenburg (2021)) and comments a bit on model estimation and fit evaluation.

While most latent variable models use techniques to eliminate the values of latent variables for individual cases, the approach taken here is a weighted least square method that estimates values of parameters such as path loadings together with values of latent variables for each case. This approach is computationally demanding but feasible even for relatively large data sets.

The model consists of a set of equations $g_l = 0, l = 1 \dots m$ that are fitted to a data matrix $A_{i,j} \in \mathbb{R}, i = 1..n, j = 1..k$ of values for case i and item j. The columns of A are thus seen as n realizations of manifest random variables x_j. The model assumes that latent variables $\eta_q, q = 1..Q$ (which take on values for each case) and model parameters $p_s, s = 1..S$, exist such that the equations $g_l(\{x_j\}, \{\eta_q\}, \{p_s\}) = 0$ hold. Due to measurement error the equations for the actual data $g_l(\{A\}, \{Z\}, \{p_s\}) =: \epsilon_l \approx 0$ hold only approximately. Here Z is the matrix of the values of the latent variables in case i for η_q, i.e. $Z_{i,q} \in \mathbb{R}$.

A weighted least square approach is then used to minimize the objective function involving weights $w_l > 0, \sum_l w_l = 1$:

$$F_w(\{Z_{i,q}\}, \{p_s\}) := \sum_{l=1}^m w_l \cdot \sum_{i=1}^n \epsilon_{i,l}^2$$

This minimization problem may be restricted by constraints on parameters and/or latent variables. The minimizer of this function gives estimates both for the latent variables as well as the model parameters.

There are several strategies to choose the weights in the above objective function. If one knew the error variances $\sigma(\epsilon_l)^2$, it would be sensible to set $w_l := 1/\sigma(\epsilon_l)^2$, because this allows together with some more assumptions to prove that the estimate is the maximum likelihood estimation (see Oldenburg, 2021).

The following strategies have been investigated elsewhere (Oldenburg, 2021):

W_1: This is unweighted least squares, i.e. $w_l = \frac{1}{m}$ are chosen to be the same for all equations.

W_2: This is a two-step-strategy: After a first estimation round with strategy W_1 one has values of $\epsilon_{i,l}$ and they give estimates of the error variances that can be used in a second round.

W_o: This is a self-consistency strategy: The weights w with constraints $w_l > 0, \sum_{l=1}^m w_l = 1$ and a proportionality factor $K > 0$ are included into the parameters to be estimated from minimizing

$$F^o(\{Z, p, w, K\}) := \frac{F_w(\{Z_{i,q}\}, \{p_s\})}{\left(\prod_{l=1}^m w_l \cdot (1-w_l)\right)^{1/m}} + P \cdot \sum_{l=1}^m \left(w_l - \frac{K}{\epsilon_l^2}\right)^2$$

Here P is some large penalty number.

W_a: Again, this is a self-consistent strategy but with a different algorithm. The weights are not part of the optimization process but are viewed as parameters. Starting from the uniform weight vector $w^{(0)} = \left(\frac{1}{m}, \dots, \frac{1}{m}\right)$ the target function is minimized and a-posteriori error variance estimates $\sigma\left(\epsilon_l^{(0)}\right)^2$ are obtained. It the weights are optimal, then the sequence $w_l^{(0)} \cdot \sigma\left(\epsilon_l^{(0)}\right)^2, l = 1..m,$ will have minimal variance. Hence, for the next iterative step, $w^{(1)}$ is obtained from $w^{(0)}$ by changing the weights in the direction indicated by the deviation of $w_l^{(0)} \cdot$

$\sigma\left(\epsilon_i^{(0)}\right)^2$ from its mean value. This gives a new weight vector $w^{(1)}$ and in another minimization step new estimates are calculated. The process then iterates until no reduction of variance can be found.

For fit comparison, the mean error residual R per case and per equation $R := \frac{F_{min}}{n \cdot m}$ can be used. A second, similar measure is the data fit measure D which is calculated in the following way: After having calculated all estimates for parameters and latent variables these are plugged into the model equations and from this the data are prognosed as if they were unknown. Then, the Euclidean distance between the data vector and the prognosed data vector is calculated and averaged. The resulting number should be as close as possible to 0.

Now, this general estimation framework will be specialized to the situation needed in this paper. The latent variables in this model are certain algebraic competences $\eta_q, q = 1..Q$, e.g. η_1 be the ability to perform substitutions and η_2 the ability to transform algebraic expressions correctly. The values of all latent variables are constraint to the unit interval, i.e. $\eta_q \in [0,1]$ for each case. This allows to write down model equations e.g. of the following kind: $x_1 = b_1 + c_1 \cdot \eta_1 + c_2 \cdot \eta_1 \cdot \eta_2 + \epsilon_1$, $x_1 = b_2 + c_3 \cdot \eta_2 + c_4 \cdot \eta_2^2 + \epsilon_2$. Here $x_1 \in [0,1]$ is an observed variable. The parameters $b_i, c_i \in \mathbb{R}$ have the following interpretations: b_i are the offsets, i.e. the expected score on item x without having any competence described by the following variables and averaging over all latent variables not in the equation. Hence, $b_1 \neq b_2$ can occur, but experience shows that in most cases they differ only slightly. c_1 is the path coefficient that measures the influence of competence η_1 on x – this is the same interpretation as in linear SEM. Given that the latent variable values are from the unit interval their product $\eta_1 \cdot \eta_2$ can be interpreted in the sense of fuzzy logic (Zahdeh, 1965) as a measure for having both competences and c_2 is the path weight of the influence of this combination. Similarly, η_2^2 can be interpreted as a measure for having a particular strong competence η_2 and c_4 is its corresponding path weight. In the models estimated below all path coefficients are restricted to be non-negative, i.e. $c_i \geq 0, i \geq 1$. This reflects the idea that having a competence should not have negative effects. A further pragmatic reason is that these constraints eliminate certain special cases of under-determined systems. A similar argument could be made for offsets too, but on the other hand, allowing negative offsets b_i gives the freedom for situations in which a little bit of a competence η is actually not enough to predict any chance for solving the item but in which some higher value is needed to make the estimate larger than 0.

For better interpretation of path coefficients, one may normalize them by rescaling with the standard deviations of the involved variables. In the above example e.g., one has standardized coefficients $\hat{c}_3 := \frac{\sigma(\eta_1 \cdot \eta_2)}{\sigma(x_1)} \cdot c_3$. Standardized variables $\hat{x} = \frac{x}{\sigma(x)}$ will also be used, so that $x = c \cdot \eta$ is equivalent to $\hat{x} = \hat{c} \cdot \hat{\eta}$.

Empirical input: Test and data

A somewhat improved version of the test described in (Oldenburg, 2009) was used with 171 10th grade students from a German school. This is a full cross section of this school but maybe not representative for Germany as a whole, although the school is quite typical for Germany. The purpose of the data analysis is not to make claims or conjectures about the students but to test the method described above and to derive some insights into algebraic thinking. The test was administered by the teachers and written for diagnostic purposes. The anonymized data that are used for this analysis do not contain any personal data, so that the sample cannot be characterized by means of the data. However, one can assume equal sex distribution and average age of about 16-17 years. By the time of the test, education on elementary algebra is completed and in the next school year they'll start with (informal) calculus. The test was originally constructed with the algebra model of Oldenburg (2009) in mind. It contains 21 tasks, some of which have sub-items so that there is a total of 45 items. Most of them are encoded as incorrect/correct (0, 1) by the teachers, while some are encoded using partial credits which have been rescaled to the unit interval so that all observed variables are from [0,1]. There are several items shared with the test designed by Küchemann (1979). There are no multiple-choice items, all require a free-form answer but in most cases the correct answer (e.g., an expression or equation) is unambiguously defined.

The model dimensions

The model that will be investigated here has six latent dimensions of algebraic competencies. This model has also been fit successfully to other algebra test data as will be reported elsewhere. The dimensions as well as

example items that load directly on the latent variables (i.e. there es an equation of the type observed $= c_0 + c_1 \cdot$ latent $+ \epsilon$) are given in table 1.

Table 1: The latent variables of the model

Name & Shortcut	Description	Example item
Calculation Calc	Perform syntactic transformations	A14d: Simplify $(x - x^3) \cdot (x + x^3)$
Functions Fun	Work with function graphs and tables	A3: Match depicted graphs and functions $f_1(x) = 2x - 5$, $f_2(x) = 4 - x^2$, $f_3(x) = \frac{1}{x}$, $f_4(x) = 5 - x$, $f_5(x) = 4 - 4x^2$, $f_6(x) = \sqrt{x}$
Reference Ref	Use variables to refer to quantities	Determine the circumference of a polygon given in a picture with sides labeled by letters
Relations Rel	Relate quantities via one equation or several equations	If $u = x + y, x \cdot u = v$, what can be said about v? (from Küchemann 1979)
Quantification Quant	Make statements about domains for which a statement is valid	What is larger, $n + 2$ or $2n$? (from Küchemann 1979)
Substitution Subs	Perform Substitutions	If $f(x) = x^3 - 2$, then what is $f(x + 1)$?

For each of the items in the test I judged which of these six latent variables may explain individually or in combination success on the item. From this a model equation was generated. The set of these model equations is the non-linear model. By omitting all non-linear expressions from the model equations, a linear version of it was derived. Both models were fit using all strategies described above.

Note that the original test was not constructed for this model, so many items are not made to measure exactly one of these dimensions. Hence, a lot of possible cross-loadings will be included in the model equations.

Evaluating the estimation method

To assess model fit, the remaining error variance per case R as defined above was calculated. Moreover, the data fit measure D was determined. All results are given in table 2.

Table 2: Residual errors for linear and non-linear models

Method	W_1	W_2	W_o	W_a
R linear model	0.144	0.107	0.141	0.112
R non-linear model	0.133	0.074	0.126	0.104
R Improvement	7.8%	44.2%	12.5%	7.0%
D linear model	0.098	0.099	0.105	0.101
D non-linear model	0.093	0.096	0.102	0.094

Results indicate that by all measures and strategies the non-linear model fits the data better than the linear model. Note that for a completely uninformative model, $R = D = 0.25$. Hence, both models explain a large part of the data variance and the additional reduction of residual errors by more than 7% for the non-linear models

can be considered substantial. Also note that W_2 gives very good values of R for the non-linear model but this is likely to be a result of overfitting.

In principle one can try to fit the linear model with standard SEM software like lavan. However, with the many cross-loadings that result from the fact that the test was not designed exactly for these constructs, none of the usual SEM estimation methods converge. This shows that the new method is capable of delivering sensible results where traditional methods are not applicable.

Summarizing these observations, RQ1 can be answered affirmatively.

Insights for algebra education

For the detailed analysis I'll restrict to results from two strategies, W_1, W_a. The reason to include uniform estimation in this subset is due to the fact that it reflects in a sense best what the modeler wrote down. The other methods adjust weights by estimated error variances and this may not be necessary here, because manifest and latent variables are from [0,1] so that error variances can also be expected to be in a similar range. Another reason for including the uniform strategy W_1 is that is shows the best data fit. Among the weigh-adjusting methods I choose W_a because it also produces good data ft, and under some distributional assumptions can be expected to converge to the optimal value (although this conjecture is not proved but plausible from simulations).

From the estimates for the six latent variables correlations can be calculated. Results are given in table 3.

Table 3: Correlation matrices of the latent variable $Subs, Quant, Ref, Rel, Fun, Calc$

Strategy	Linear model	Non-linear model
W_1	$\begin{pmatrix} 1. & 0.58 & 0.48 & 0.53 & 0.49 & 0.36 \\ 0.58 & 1. & 0.59 & 0.59 & 0.37 & 0.41 \\ 0.48 & 0.59 & 1. & 0.66 & 0.39 & 0.4 \\ 0.53 & 0.59 & 0.66 & 1. & 0.41 & 0.88 \\ 0.49 & 0.37 & 0.39 & 0.41 & 1. & 0.31 \\ 0.36 & 0.41 & 0.4 & 0.88 & 0.31 & 1. \end{pmatrix}$,	$\begin{pmatrix} 1. & 0.21 & 0.2 & 0.34 & 0.44 & 0.15 \\ 0.21 & 1. & 0.26 & 0.44 & 0.24 & 0.27 \\ 0.2 & 0.26 & 1. & 0.37 & 0.32 & 0.23 \\ 0.34 & 0.44 & 0.37 & 1. & 0.42 & 0.58 \\ 0.44 & 0.24 & 0.32 & 0.42 & 1. & 0.26 \\ 0.15 & 0.27 & 0.23 & 0.58 & 0.26 & 1. \end{pmatrix}$
W_a	$\begin{pmatrix} 1. & 0.44 & 0.41 & 0.36 & 0.48 & 0.29 \\ 0.44 & 1. & 0.48 & 0.29 & 0.33 & 0.22 \\ 0.41 & 0.48 & 1. & 0.34 & 0.32 & 0.24 \\ 0.36 & 0.29 & 0.34 & 1. & 0.28 & 0.99 \\ 0.48 & 0.33 & 0.32 & 0.28 & 1. & 0.24 \\ 0.29 & 0.22 & 0.24 & 0.99 & 0.24 & 1. \end{pmatrix}$,	$\begin{pmatrix} 1. & 0.24 & 0.11 & 0.33 & 0.38 & 0.25 \\ 0.24 & 1. & 0.26 & 0.42 & 0.24 & 0.3 \\ 0.11 & 0.26 & 1. & 0.23 & 0.26 & 0.24 \\ 0.33 & 0.42 & 0.23 & 1. & 0.42 & 0.34 \\ 0.38 & 0.24 & 0.26 & 0.42 & 1. & 0.26 \\ 0.25 & 0.3 & 0.24 & 0.34 & 0.26 & 1. \end{pmatrix}$

The results in table 3 indicate that there is a substantial difference between linear and non-linear models but also that the estimation methods differ to some extent.

First, one can observe that with the linear model the latent constructs Rel and $Calc$ cannot be separated, however, for the non-linear model they are clearly distinct (although the estimate of the correlation coefficient differs between both methods).

Now let's look at some items in detail:

Item A13: It is known that $x = 6$ is a solution of $(x + 1)^3 + x = 349$. Use this to find a solution of $(5x + 1)^3 + 5x = 349$.

For this item the following equations were set up for the non-linear model:

$$A13 = b_{13a} + c_{13a} \cdot Subs + c_{13b} \cdot Rel \cdot Subs + \epsilon_{13a}$$

$$A13 = b_{13b} + c_{13c} \cdot Calc + c_{13d} \cdot Calc \cdot Subs + \epsilon_{13b}$$

The equations for the linear model are obtained by omitting the last summand in each equation. Some explanations are in order why the equations have been set up in this manner: First, from my own expert rating,

I found it plausible that the latent competencies $Subs, Calc, Rel$ are most likely to explain success in this item, especially that a logical combination of being good in substitution and in the other two competencies would explain success. Hence, one could think about writing down a single combined equation like

$$A13 = b_{13c} + c_{13e} \cdot Subs + c_{13f} \cdot Calc + c_{13g} \cdot Rel + c_{13h} \cdot Rel \cdot Subs + c_{13i} \cdot Calc \cdot Subs + \epsilon_{13c}$$

This is possible in the described modelling framework and gives interesting insights, too. However, this approach would give a single error estimate while having several equations allows to detect sources of large errors more precisely. Based on this argument one might be tempted to write down the equations

$$A13 = b_{13j} + c_{13j} \cdot Subs + \epsilon_{13j}, \ A13 = b_{13k} + c_{1kf} \cdot Calc + \epsilon_{13k}, A13 = b_{13l} + c_{13l} \cdot Rel + \epsilon_{13l}$$

$$A13 = b_{13m} + c_{13m} \cdot Rel \cdot Subs + \epsilon_{13m}, \ A13 = b_{13n} + c_{13n} \cdot Calc \cdot Subs + \epsilon_{13n}$$

Again, this is a sensible approach that is possible and interesting. However, it has the disadvantage that comparison with the linear model is not so easy. Of course, one can restrict to the subset of linear equations as well, but this will imply that the linear and the non-linear model differ in the number of equations, and this makes model comparison tricky. Hence, the two equations given above were used. An approach with a third equation with Rel entering linearly brought bad model fit for this additional equation and was hence discarded.

This completes the justification why the above given equations are used. The results of the estimation process are given in table 4.

Table 4: Estimates of model equations for item 13. Coefficients are rounded to two decimal places and terms are omitted if coefficient is rounded to 0. Error Variances are given in parentheses

	Linear model	Non-linear model
W_1, non-normalized	A13=0.02+0.4 Subs (0.12) A13=0.01+0.43 Calc (0.13)	A13=-0.17+0.26 Subs+0.72 Rel Subs (0.10) A13=-0.29+0.73 Calc Subs (0.12)
W_1, normalized	A13=0.05 +0.32 Subs A13=0.03 +0.19 Calc	A13=-0.47+0.12 Subs+0.39 Rel Subs A13=-0.79+0.34 Calc Subs
W_a, non-normalized	A13=0.02 +0.44 Subs (0.12) A13=0.03 +0.39 Calc (0.13)	A13=0.06 +0.71 Rel Subs (0.11) A13=-0.36+0.42 Calc+0.31 Calc Subs (0.12)
W_a, normalized	A13=0.04 +0.33 Subs A13=0.09 +0.11 Calc	A13=0.17 +0.43 Rel Subs A13=-0.98+0.05 Calc+0.25 Calc Subs

Now for the interpretation of these results: For the linear model the normalized equations indicate that both competencies, $Subs$ and $Calc$ have explanatory power. However, in the linear model $Subs$ seems to be more important, as its normalized path coefficients is larger than the one for $Calc$. Both estimation methods give somewhat different estimates, but this conclusion is independent of the method.

The results for the non-linear model emphasize the importance of $Subs$ as well, but in addition show that combinations of having two competencies give much stronger path coefficients than single competencies. Especially $Rel \cdot Subs$, i.e., being good both at substitutions and relational thinking explains success best. Again, this conclusion is independent of the estimation method. For W_a $Calc$ and for W_1 $Subs$ have some additional explanation power but the path weights are small, and in fact much smaller than that of $Calc \cdot Subs$. Hence, one may conclude that substitution in combination with other abilities is an important part of algebraic competency.

Now that the formation and interpretation of model equations are explained in detail for item 13, some of the other items are discussed more briefly.

Item 10: Assume a, b are positive real numbers. Assume that $\frac{1}{a} + \frac{1}{b} = 2$ always holds. What can be said about a if b decreases?

Proposed model equations are:

$$A_{10} = b_{10a} + c_{10a} \cdot Quant + c_{10b} \cdot Quant \cdot Rel + \epsilon_{10a}$$

$$A_{10} = b_{10b} + c_{10c} \cdot Subs + c_{10d} \cdot Quant \cdot Subs + \epsilon_{10b}$$

The findings from fitting the model are contained in table 5. The interpretation is that in the linear model, $Quant$ is the most explaining latent variable, but $Subs$ contributes as well. For the first equation, the possible explanation $Quant \cdot Rel$ is not useful, but for the second equation $Quant \cdot Subs$ turns out to be the strongest predictor. The large offsets together with the large error variances indicate, however, that there may be some additional explanatory variable needed.

Table 5: Estimates of model equations for item 10.

	Linear model	Non-linear model
W_1, non-normalized	A10=-0.66+2.52 Quant (0.18) A10=0.3 +0.58 Subs (0.21)	A10=0.09 +6.37 Quant (0.14) A10=0.16 +7.09 Quant Subs (0.15)
W_1, normalized	A10=-1.37+0.48 Quant A10=0.63 +0.34 Subs	A10=0.19 +0.63 Quant A10=0.32 +0.6 Quant Subs
W_a, non-normalized	A10=-0.13+1.46 Quant (0.22) A10=0.31 +0.63 Subs (0.21)	A10=0.43 +2.68 Quant (0.20) A10=0.41 +0.09 Subs+3.02 Quant Subs (0.21)
W_a, normalized	A10=-0.13+1.46 Quant A10=0.31 +0.63 Subs	A10=0.89 +0.37 Quant A10=0.84 +0.06 Subs+0.32 Quant Subs

Item 16: Note that $7 \cdot 9 = 8^2 - 1$ and $11 \cdot 13 = 12^2 - 1$. Formulate a rule and justify.

Proposed equations:

$$A_{16} = b_{16a} + c_{16a} \cdot Ref + c_{10b} \cdot Quant \cdot Rel + \epsilon_{10a}$$

$$A_{16} = b_{16b} + c_{16c} \cdot Rel + c_{10d} \cdot Quant \cdot Subs + \epsilon_{10b}$$

Results are given in table 6. Interpretation is straightforward: The linear model clearly shows that the competencies Ref and Rel are important but predict most strongly when in addition a strong competency in quantification is available.

Table 6: Estimates of model equations for item 16.

	Linear model	Non-linear model
W_1, non-normalized	A16=-0.06+0.74 Ref (0.12) A16=-9.81+19.75 Rel (0.13)	A16=0.08 +6.02 Quant Ref (0.10) A16=0.14 +7.46 Quant Rel (0.10)
W_1, normalized	A16=-0.14+0.43 Ref A16=-25.13+0.4 Rel	A16=0.2 +0.56 Quant Ref A16=0.37 +0.56 Quant Rel
W_a, non-normalized	A16=0.04 +0.47 Ref (0.14) A16=-0.87+2.47 Rel (0.15)	A16=0.22 +3.64 Quant Ref (0.12) A16=0.22 +0.16 Rel+3.83 Quant Rel (0.12)
W_a, normalized	A16=0.1 +0.32 Ref A16=-2.23+0.21 Rel	A16=0.57 +0.46 Quant Ref A16=0.57 +0.1 Rel+0.38 Quant Rel

As a last example, let's look at an item where a quadratic term in one competency gives a good estimation:

Item 17c: Given the function $f(x) = x^3 - 2$ calculate $f(x + 1)$.

The main difficulty in this item is to do the substitution correctly. Hence, the proposed model is:

$$A_{17c} = b_{17c} + c_{17ca} \cdot Subs + c_{17cb} \cdot Subs^2 + \epsilon_{17c}$$

The linear model estimated by W_a gives the normalized estimate A17c=-0.08+0.53 Subs   (error variance 0.17) while the non-linear estimate is A17c=-0.43+0.92 Subs2 (error variance 0.04). This shows that getting this substitution done right requires a higher level of $Subs$ than most other tasks: Only if you are really good in substitution you can tackle this item (which confused many students because of the double role of x). Also note the substantial drop of error variance and that the linear term got almost zero loading in the nonlinear model.

Of course, much more insights can be drawn from the fitted model but for the purpose of this publication four items should be sufficient to underpin the positive answer to research questions 2 and 3.

Conclusions and Outlook

The results given above allow to answer all three research questions affirmatively. The new estimation method allowed to estimate linear models with many cross-loadings that cannot be estimated by traditional SEM techniques. Moreover, the nonlinear model fits the data much better than the linear one and supports the hypothesis, that for many items the best prediction of success is given by the fuzzy logic conjunction of having two competencies or by having one competency to a very high degree (to a higher power). Especially, it is interesting to note that substitution is an important algebraic competency and that it is especially important when combined with abilities in other areas of algebra such as quantification, calculation and relational thinking.

However, it must be stated clearly that the results given above must be read with caution. First, the method applied is new and this is the first application of it to real world data. It may be that the good performance seen in simulation studies may not carry over to such data. However, the fact that results are quite reasonable indicate that this worry may be unnecessary. Thus, several further methodological research goals are steered up:

- Deeper theoretical underpinning of the method

- Develop a Bayesian approach to estimating unit interval constrained latent variable models as a second method to increase trust in results

- Application of the method to more data sets

  Regarding the results about algebraic competency, one additionally should consider that the data set was only of moderate size and that the test was not developed for the model tested. Thus, the following research goals arise:

- Check other models. Especially, find if some items are best explained by products of three or even more latent variables

- Check if the found importance of substation can also be supported by other data sets

- Test other models proposed in the literature that can be translated into such a nonlinear structure

Perspectives: A research program

If the critical questions raised in the preceding section can be answered satisfactorily the following research possibilities opens up. As cross-loadings are less problematic compared to traditional SEM and logical conjunction can be modelled it is relatives easy to adapt the measurement model of the above six-dimensional model to the items from other existing algebra tests. The central idea of the research program is therefore to find a model of algebraic competency (maybe a modification of the above model) that can be fitted to the data from as many algebra tests as possible. The idea behind this is, of course, that the structure of algebraic competency should be a construct that is the same for many people educated in a similar algebraic culture. So

while we currently often have "one study: one model : one test" the idea is to have one general model, i.e. "several studies and tests: one model". Of course, not all data sets will be useful in the end. For example, the PISA data set is too sparse to be estimated by my method (although full data is not required but stable numerical results require a certain density). By following this research program it is hoped that a general model can be constructed that is supported by a great variety of data, and, vice versa, explains a large variety of data by some simple principles.

Data Availability

Data, test items (in German) and Mathematica programs to perform the analysis are available at https://myweb.rz.uni-augsburg.de/~oldenbre/AlgebraTestData.zip .

References

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

2. Bardini. C., Oldenburg, R., Pierce, R., & Stacey, K. (2013) Technology prompts new Understandings: The Case of Equality. *Proceedings of MERGA 2013*.

3. Dijkstra, T. K. and Schermelleh-Engel, K. (2014). Consistent partial least squares for non-linear structural equation models. *Psychometrika*, 79(4), 585–604. DOI: 10.1007/s11336-013-9370-0

4. Epp, S. (2011): Variables in mathematics education. In P. Blackburn (Ed.). *Tools for teaching logic*. TICTTL 2011. Berlin: Springer LNCS 6680, 54–61. DOI: 10.1007/978-3-642-21350-2_7

5. Hoyle, R. H. (Ed.) (2012). *Handbook of Structural Equation Modeling*. Guilford Press. https://doi.org/10.1080/10705511.2013.769397

6. Kelava, A., Werner, C. S., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., Cham, H., Aiken, L. S., and West, S. G. (2011). Advanced non-linear latent variable modeling: Distribution analytic lms and qml estimators of interaction and quadratic effects. *Structural Equation Modeling*, 18, 465–491. https://doi.org/10.1080/10705511.2011.582408

7. Kieran C. (2004) The Core of Algebra: Reflections on its Main Activities. In: Stacey K., Chick H., Kendal M. (eds) *The Future of the Teaching and Learning of Algebra*. The 12thICMI Study. New ICMI Study Series, vol 8. Springer, Dordrecht. https://doi.org/10.1007/1-4020-8131-6_2

8. Küchemann, D. (1979). Children's Understanding of Numerical Variables, *Mathematics in School*, 7, pp. 23–26.

9. Mason, J., & Sutherland, R. (2002). *Key Aspects of Teaching Algebra in Schools*. Qualifications and Curriculum Authority..

10. Oldenburg, R. (2009). Structure of Algebraic Competencies. In V. Durand-Guerrier, S. Soury-Lavergne & F. Arzarello (Eds.), CERME 6: Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education.

11. Oldenburg, R. (2019). A classification scheme for Variables. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.). *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education*. Utrecht: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

12. Oldenburg, R. (2020). Structural Equation Modelling – Comparing Two Approaches. *The Mathematica Journal*.

13. Oldenburg, R. (2021). Case based error variance corrected estimation of structural models. Submitted to *Statistical Modelling*.

14. Umbach, N., Naumann, K., Brandt, H., and Kelava, A. (2017). Fitting non-linear structural equation models in r with package nlsem. *Journal of Statistical Software*, 77, 7.

15. Usiskin, Z. (1988). Conceptions of school algebra and uses of variables. In A. F. Coxford (Ed.), *The Ideas of algebra, K-12* (pp. 8–19). Reston, VA: NCTM.

16. Zadeh, L. A. (1965). Fuzzy Sets, *Information and Control*, Vol. 8, pp. 338-353.

Conflicts of Interest

I declare that there no conflict of interests.