"Quaderni di Ricerca in Didattica (Mathematics)", n°20 suppl 1, 2010
G.R.I.M. (Department of Mathematics, University of Palermo, Italy)
A.S.I. 5 Proceedings 5-7- November 2010

A re-analysis of TIMSS data using Statistical Implicative Analysis

Reinhard Oldenburg
Goethe University Frankfurt, Germany
Postal address: Senckenberganlage 9, D-60325 Frankfurt, Germany

E-mail: oldenbur@math.uni-frankfurt.de

Abstract. An often noted difference between teachers' experience and empirical research concerns the dimensionality of mathematical proficiency. Several large scale studies (PISA, TIMMS) as well as smaller studies found that the one-dimensional Rasch model is appropriate to model students' performance. On the other hand, teachers' experience is that several mathematical sub-domains are only weakly coupled. The dimensionality of a construct may be assessed by various models. This paper concentrates on Statistical Implicative Analysis and finds a structure within mathematical proficiency that is largely compatible with an intuitive classification.

Résumé. Une différence souvent soulignée entre l'expérience des enseignants et la recherche empirique concerne la dimensionnalité des aptitudes en mathématiques. Des études à grande échelle (PISA, TIMMS,) tout autant que de plus petites études, montrent que le modèle à une dimension de Rasch est approprié au modèle des performances des élèves. D'un autre côté, l'expérience des enseignants énonce que les multiples sous-domaines mathématiques sont plutôt faiblement liés. La dimensionnalité d'une construction serait imposée par différents modèles. Ce papier se concentre sur l'Analyse Statistique Implicative et découvre une structure au sein des aptitudes en mathématiques qui est grandement compatible avec une classification intuitive.

1 Introduction

It is common among teachers to believe that students fall into types that perform different in different mathematical disciplines, e.g. some students are said to do well in geometry but are weak in arithmetic. Large empirical studies on the other hand employ the one-dimensional Rasch model and usually find that it is sufficient to assume one latent capability variable for each student. Even more support for the hypothesis of one-dimensionality of mathematical proficiency came from the work of (Neubrand & Neubrand 2004). They analyzed German PISA data by grouping items into two categories. Items were classified as either technical items (requiring calculations or procedures to be applied) or modeling items. They found that the score variables derived from this classification are highly correlated with r=0.98.

In a study with undergraduate students (Engelbrecht, J. et al. 2005) a classification of mathematical tasks into conceptual and procedural tasks has been conducted and led to a medium correlation of r=0.59. Belter (2009) as well finds a high correlation, although no explicit value is given.

On the other hand, in (Oldenburg 2009) we observed that within the sub-domain of algebra, the performance on syntactical items (i.e. items that require operations that apply to the surface structure of an algebraic object) and semantical items (i.e. items that require an understanding of the meaning of expressions and equations) are only related very weakly.

There are other classifications of algebraic competencies. McCullum (2007) considered test items based on the classification of the National Research Council's report 'Adding it up'. Procedural fluency according to this classification is mainly tested by what we call syntactical items. Our semantical items test a combination of conceptual understanding and strategic competence.

The concept of syntactical algebraic item has been used in this study, because a lot of tasks in standard school books fall in this category and there is an ongoing discussion about the relevance of these tasks. In a sense, syntactical item is one that can easily be done by a computer algebra system, while semantical items need a link between algebraic objects and real other mental objects and concepts.

The present study re-analyzes results from TIMSS 2003 to further investigate the question of the inner structure of mathematical proficiency. The use of the TIMSS database gives a solid foundation as it uses approved predefined items that can't have a bias regarding our research question. The method of Statistical Implicative Analysis (SIA) is used to evaluate the data because it is able to detect structure in contrast to merely confirming or rejecting existing hypothesis. Especially we try to identify key concepts that are involved in several implicative relations so that they can be assumed to be fundamental either as prerequisites or as indicators.

2 Data preparation

TIMSS measured mathematical and scientific knowledge of representative samples from 40+ countries. We used the study from 2003 where the population consists of 8th graders. The TIMSS 2003 data set is available with full documentation (http://timss.bc.edu/timss2003.html). In order to reduce general variance in the data we selected a single country (England). The TIMSS test uses a multi-matrix design, thus items are compiled into books and each student is assigned one such book. For the SIA analysis we selected two books (books 2 (with data from 193 students) and 5 (with data from 212 students)) at random. From these books we selected the subset of items on mathematical questions for which the item text has been released. A binary 0/1 encoding has been applied to reduce the information given in the data. The raw data include information on the distractor selected selected by the student. However, for most items it seemed difficult to attribute partial credit to these wrong distractors. For free-form answers the database records a very rough classification (e.g. missing accuracy) of the type of error. Again it seems tricky to use a partial credit approach. Thus the binary encoding seems to be adequate.

The SIA analysis has been performed by the CHIC (Gras 2010) software.
For further analysis the items have been classified into one or more of the following categories:

- Arith: Items with rely mostly on calculations with numbers
- Algebra: Items that need algebraic thinking and symbolism
- Text: Items that need substantial understanding and interpretation of text
- Syntactical: Items that can be performed by evoking a schema triggered by the pattern of an expression
- Semantical: Items that require associating meaning with mathematical objects, i.e. modeling or interpretation items.
- Geo: Geometrical items
- Fun: Items that require working with functions and/or their graphs.
- Probability: Items that involve stochastic thinking

Table 1. Items from book2.

Item	Shortcut	Description	Classification
M022148	Time1	Time calculation	Arith
M022156	NumW1	Numerical word problem	Text; Arith
M022202	Geo1	Angle argumentation	Geo
M022253	AlgSyn1	Solution of linear equation	Algebra; Syntactical
M012013	Geo2	Counting blocks (spatial)	Geo
M012015	Geo3	Argument with properties of a trapezoid	Geo
M022135	Fun1	Read cooling curve	Fun
M022191	AlgSem1	Not formalized reasoning with proportion and inequalities	Algebra; Semantical
M022196	AlgSem2	Substitution of numerical values and evaluate truth of statements	Algebra; (Semantical, Syntactical)
M012016	Arith1	Ordering of rational numbers	Arith
M012017	AlgArith	Generalized arithmetic on patters	Arith
M012027	Time2	Time calculation	Arith
M012030	AlgSem3	Relation between length	Algebra, Geometry, Semantical
M022142	Geo4	Determine equal angles	Geo
M022188	Time3	Compare units of time	Arith
M012028	Arith2	Decimal system	Arith
M022139	Arith3	Percentage	Arith
M022154	Geo5	Rotation	Geo
M022189	Data1	Read bar diagram	Representation
M022251	AlgSem4	Model number by expression	Semantical; Expression
M012014	Data2	Interpret circle diagram	Representations
M012025	Data3	Interpret distance-time-diagram	Fun
M012026	Geo6	Determine angle in triangle from sum of angles	Geo
M022144	Arith4	Rounding	Arith
M022194	NumW2	Numerical word problem	Arith
M022198	Arith5	Sorting	Arith
M012029	AlgSem5	Verbalize rule	Algebra; Semantical
M022146	Prob1	Expected number	Probability
M022185	AlgSyn2	Simplification of difference	Algebra; Syntactical
M022199	Arith6	Ratios	Arith

Some comments on this classification: It is not always easy to decide whether syntactical or semantical aspects are dominant in an item. E.g. the item AlgSem2 has been classified both to be syntactical and semantical. It correlates with r=0.3 (Spearman correlation) both with the other syntactical as well as the other semantical items. Thus it has been excluded from both scales.

Another problem is that the number of categories to be used had to be answered pragmatically. In principle it would be desirable to have a much finer classification, e.g. the items on arithmetic include calculations with time and length and with rational and decimal numbers. However, the limited number of items requires a coarser classification.

For further analysis the number of correct items in each category has been calculated.

Table 2. Items from book5.

Item	Shortcut	Description	Classification
M032079	AT1	Direct/indirect arithmetic	Arith, Text
M032652	AT2	Word problem, arithmetic	Arith, Text
M032228	A1	Percentage inverse	Arith
M032044	Alm1	Algebraic Modeling	Algebra, Semantical
M032046	Als1	Solution of equation, Algebraic	Algebra, Syntactical
M032545	Alm2	Algebraic Modeling	Algebra, Semantical
M032649A	Ts1	Velocity calculation	Text; Syntactical
M032649B	T1	Velocity calculation in structured situation, possibly algebraic	Text
M032533	PS1	Proportional reasoning	Semantical
M032678	G1	Geometry; area calculation	Geo
M032403	G2	Divide Triangle	Geo
M032261	G3	Decide similarity of triangles	Geo
M032489	G4	Unfolding Polyhedron	Geo
M032588	G5	Find point in coordinate system	Geo

3 Results

This section contains the results obtained mainly by Statistical Inference Analysis (SIA) but also with classical correlation measures. In reporting SIA results we adopt a sloppy language by saying e.g. item A implies item B which should be understood as: There is a quasi implication that says that every students who solves A successfully solves B as well.

3.1 SIA analysis of book 5

The similarity analysis carried out with Chic did not recover the categories and item classification, e.g. not all geometry items are found to form a cluster in the sense of Chic. This is the expected and desired behaviour: As noted above, categories include a wide range of items that require different cognitive steps in their solution. It is thus especially interesting to interpret the empirical similarities discovered by data mining.

SIA revealed (see Figure 1) that geometry items G4, G5 are rather distinct from the other geometry items. This is easily interpreted as G4 requires spatial thinking and G5 tests essentially understanding of the Cartesian coordinate system. It may surprise that an item from geometry (G1) and an algebraic item (Alm2) are classified close together. Similar observations can be made in many other situations suggesting that certain types of geometric and algebraic reasoning processes are similar. In fact, the geometry item G1 requires calculation of an unknown length from a given length. Although the item does not ask the students to formalise this knowledge in algebraic language, it requires the students to think about relations of known and unknown quantities and this is essential for semantical algebraic items too. In fact, the algebraic item Alm2 asks students to express relations between known total prices and unknown individual process of oranges and lemons by a system of equations. The items text reads: "*At a market, 7 oranges and 4 lemons cost 43 zeds, and 11 oranges and 12 lemons cost 79zeds. Using x to represent the cost of an orange and y to represent the cost of a lemon, write two equations that could be used to find the values of x and y.*" The structure is more complex than the geometric problem, but the ability to operate on the unknown and to give it a certain meaning is required by both items.

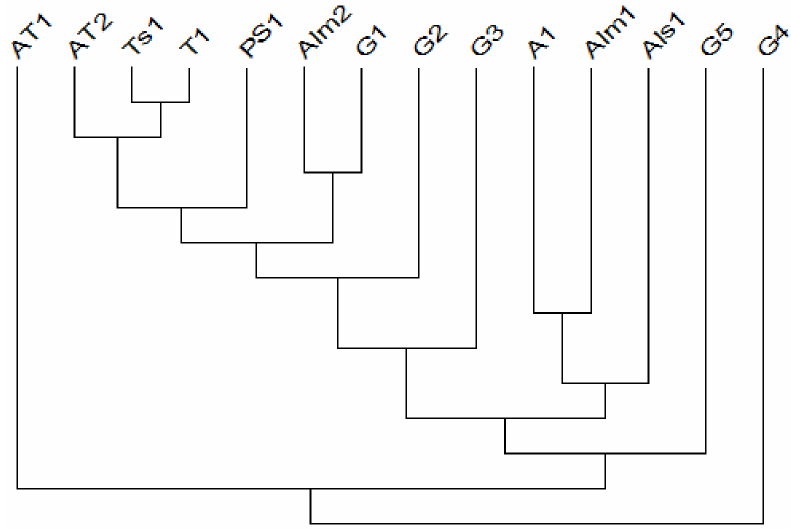

Figure 1. Similarity Graph for book 5

This completes our discussion of "surprising" results of the clustering. The remaining grouping is mainly "as expected". This should give confidence in the method. Next we shall interpret the implicative graph for book 5 (Figure 2). Almost all variables are involved in strong implicative relations (implicative intensities ≥0.99). The remaining ones (AT1,Als1,G4), i.e. items that are relatively weakly coupled, consists of a syntactical algebra item, one on spatial geometry and a modest arithmetic problem. It is reasonable that they don't correlate with the remaining items that much.

There are several items that are highly linked to others. Item Alm2 is one of these and its relation to geometrical questions has already been mentioned above. From the implicative analysis its importance is even greater as it implies other algebraic and geometric items as well. Another item that implies others in an interesting way is PS1. This is a rather simple item on numerical proportional reasoning: “A machine uses 2.4liters of gasoline for very 30hours of operation. How many litres will the machine use in 100 hours?” This observation, that items on proportionality are good predictors can be made in other places (e.g. other books not reported here) as well. Such items are more representative for mathematical proficiency in large than many others.

Several other conclusions from this graph confirm common knowledge about mathematical items. E.g. it is clear that T1→Ts1holds because Ts1requires a subset of the competencies of T1. Generally speaking, usually task difficulty decreases from top to bottom. An exception to this rule is Alm1 which is not that easy but requires prerequisites that can be detected by the other items in the test.

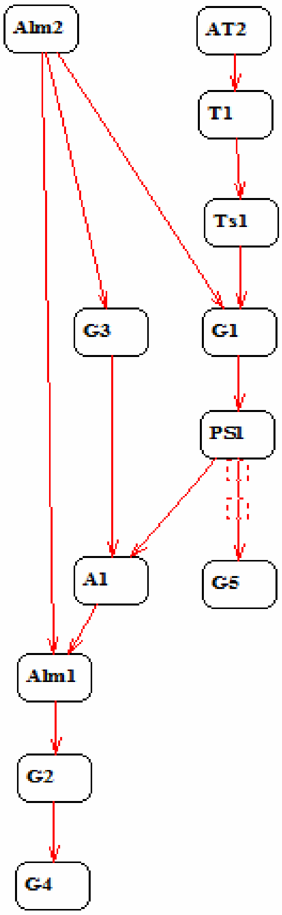

Figure 2. Implicative Graph for book 5

For the sake of completeness figure 3gives the implicative tree for this data set. However, it seems that in the present case this does reveal very much extra information with one notable exception: In this setup the items split up into two groups. The group on the right can be characterized as those items that need abstract reasoning with concepts while the left group lacks this is level of abstract reasoning.

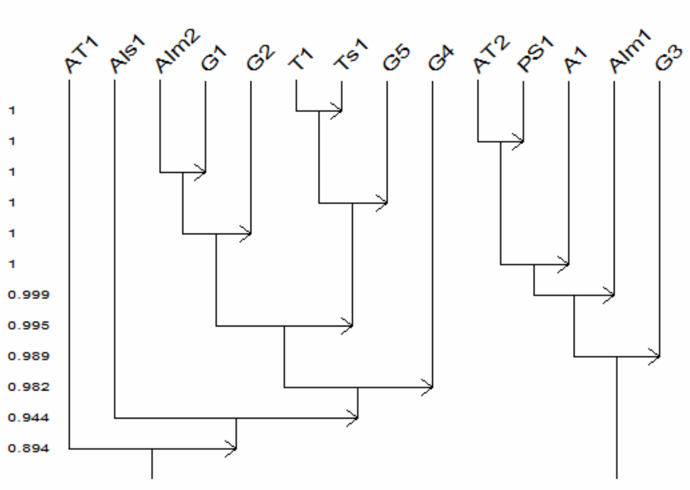

Figure 3. Implicative Tree for book 5

3.2 SIA analysis of book 2

We now report the same analysis for book 2. The similarity graph is given in figure 4.

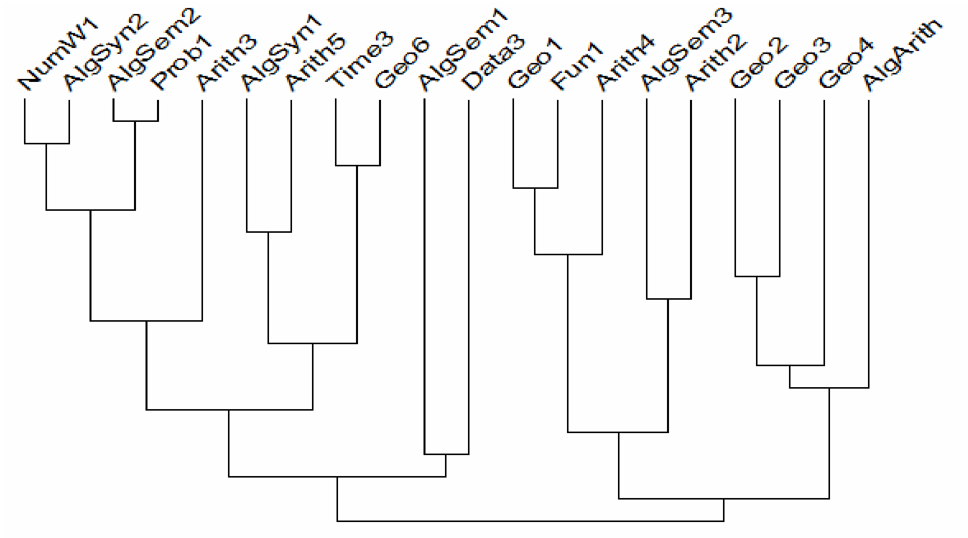

Figure 4a. Similarity Graph for book 2, part a

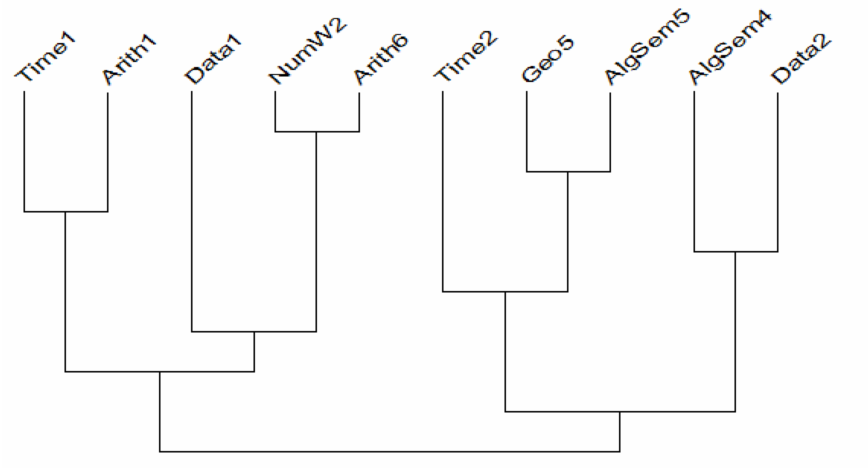

Figure 4b. Similarity Graph for book 2, part b

Again, much of this structure is easy to understand: Items Time1 and Arith1 are close because they require calculations with numbers. The same holds true for NumW2 and Arith6. However, it is interesting that these two two-item groups are not very close to each other. A closer look shows that NumW2 and Arith6 require explicit operation with wither large or abstract numbers, while Time1 and Arith are more concrete and require understanding and interpretation rather than longer calculation.

Time3 and Geo6 are close together, maybe because they both are concerned with measuring. Fun1 and Geo 1 are completely different in content area but the processes required are similar as in both cases one has to decide from the text which information to read off from a diagram and how to combine it. Geo1 also has the property that known and unknown quantities have to be put in relation. In this situation we argued above that the item should be similar to certain semantic algebraic items. This does not come out very strongly in the above diagram, but at least there is item AlgSem3 which again shows the need to combine known and unknown quantities.

Somewhat surprising is the similarity between the mental rotation task Geo5 and item AlgSem5 which asks students to verbalize a rule. Moreover it is interesting to note there are two pairs of arithmetical and syntactical algebraic items (Arith5-AlgSynth1, NumW1-AlgSynth2).Taking into account the discussion of book5 above one may result at the general conclusion that the similarity reflects less the subject and more kind of mental operation required. This makes it very interesting for research in didactics. The implicative graph (figure 5) gives rise to various interesting observations. First, note that of the 30 mathematical items only

13 are involved in strong implicative relationships. Item NumW1implies three other items. It is a very simple task: “A scoup holds 1/5kg of lour. How many scoups of flour are needed to fill a bag with 6kg of flour?” It can be solved using either proportional reasoning or conceptual knowledge about division. This confirms the observation made in the discussion of book 5 on the importance of such items. On the other hand, AlgSyn1 (it asks to solve $4\cdot(x+5)=80$ for x) is an item that is implied by three others, but all of them don’t have obvious properties that explain this fact. It is easy to understand that AlgSyn1 and AlgSem4 imply Arith2, an item on argumentation with decimal places. Also note that in the implicative graph the fact mentioned above that reasoning with geometric known and unknown quantities is expressed very clearly by the fact that Geo1 implies both AlgSem3 and AlgSem4.

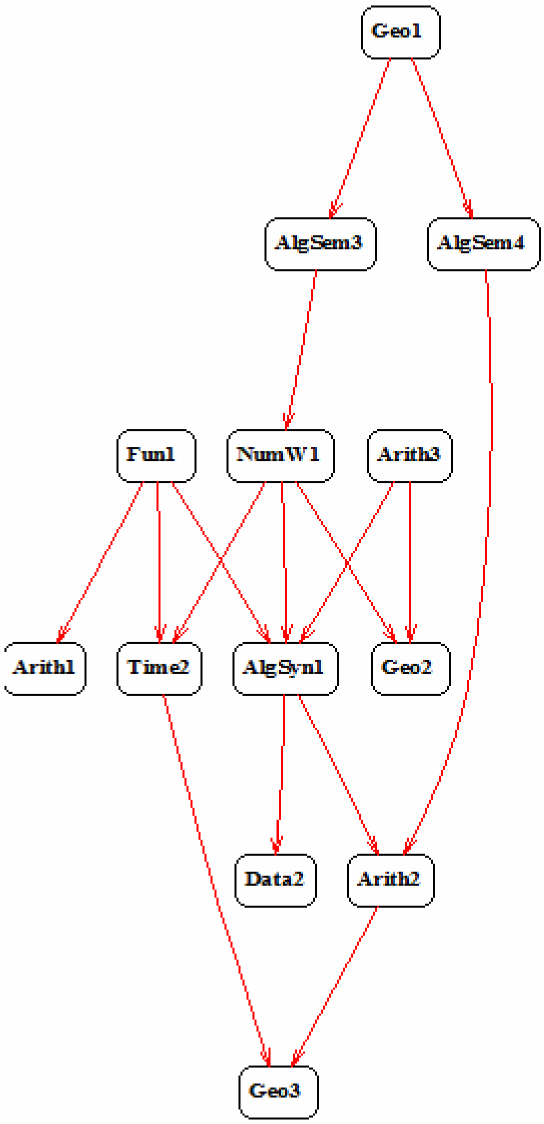

Figure 5. Implicative Graph for book 2

3.3 Correlation analysis of book 2

The variables formed according to the classification given in the table 1 above lead to the following correlation table (Spearman correlation) among sub-disciplines:

Table 3. Correlations between scales.

	Arithmetic	Geometry	Functions	Algebra
Arithmetic	1	0.51	0.36	0.57
Geometry		1	0.35	0.52
Functions			1	0.32

These correlations are high but definitely different from something close to 1.This shows that the distinct areas cannot be considered to be the same. Even more insight provides the following table that splits up syntactical and semantical parts of algebra. It shows that semantical algebra and syntactical algebra correlate less among each other than with other domains. This supports the result from a self-conducted algebra test as reported in (Oldenburg 2009). It also supports the observation made above that arithmetic and syntactical algebra items are rather close together.

Table 4. Correlations between scales with algebra split up.

	Sem alg	Syn. Alg	Arithmetic	Geometry	Functions
Semantic algebra	1	0.33	0.44	0.41	0.27
Syntactic algebra		1	0.46	0.38	0.27
Arithmetic			1	0.51	0.36
Geometry				1	0.35

However, performing SIA on this data gives a remarkable difference to our previous results (Oldenburg 2009) that syntactical aspects of algebra are rather unrelated to the rest of algebra. Here an implicative graph over these variables (figure 6) at index level ≥0.9 gives Geometric items a central role but couples syntactic algebra directly and strongly to it, as the only arrow of index ≥0.95 is that of syntactical algebra to geometry. This observation deservers further investigation.

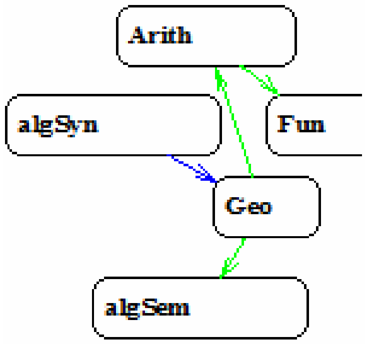

Figure 6. Implicative Graph

We didn't carry out similar studies for the item of book5 because of the lesser total number of math items.

4 Discussion

The results of this study show that SIA is an interesting method for data mining. It produces many results that are easy to understand and thus the method deserves confidence. Regarding the original research question whether mathematical proficiency can be modeled as being one dimensional the results of this study clearly show that there is an interesting and non-trivial inner structure. A large number of conclusions have been presented in section 3. They should be taken as suggestions coming from this analysis but not as proven facts. Further studies are required to investigate which conclusions are stable when considering other countries (which may be explained by different classroom culture) and other tests. The study also suggests extending the CHIC software to handle multi-matrix designs.

A very important result which we didn't expect is that obviously for the grouping of items by similarity as found empirically by Chic the domain of the item is less important than the type of argumentation. It this result can be confirmed by further study it opens up many new directions for research with possible implications for the organization of teaching mathematics. Especially it should be worth to come up with a complete model of mathematical tasks and their components. If such a model existed it could perhaps be used to predict the difficulty of items and to suggest learning paths.

References

Belter, A. (2009): The Impact of Teaching Algebra with a Focus on Procedural Understanding. *Journal of Undergraduate Research XII* .

Engelbrecht, J. et al. (2005): Undergraduate students' performance and confidence in procedural and conceptual mathematics. *Int. J. Math. Educ. Sci. Techno*l. 36, No. 7, 701-712.

Gras R., Suzuki, E., Guillet, F., Spagnolo, F. (Eds.) (2008). *Statistical Implicative Analysis. Theory and applications.* Berlin: Springer-Verlag.

Gras, R. (2010): CHIC 5.0 (Software).

McCallum, W. (2007): Assessing the Strands of Student Proficiency in Elementary Algebra. In: Schoenfeld (Ed..),, *A.: Assessing Mathematical Proficiency*. Cambridge: Cambridge.

Neubrand, J. & Neubrand, M. (2004). Innere Strukturen Mathematischer Leistung im PISA-2000-Test. In M. Neubrand et al. (Eds.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland.* (pp.87-108).Wiesbaden: VS Verlag.

Oldenburg, R. (2009): Structure of algebraic competencies. Proceedings of the Conference of European Researches in *Mathematics Education* (CERME) Lyon.

TIMSS Webseite. http://timss.bc.edu/timss2003.html. Accessed 25/04/2010