FrontlineLearningResearch
An official journal of EARLI

Basic Arithmetical Skills of Students with Learning Disabilities in the Secondary Special Schools: An Exploratory Study covering Fifth to Ninth Grade

Markus Gebhardt[a], Fabian Zehner[a], Marco G. P. Hessels[b]

[a]TU München, Germany
[b]University of Geneva, Switzerland

Abstract

The mission of German special schools is to enhance the education of students with Special Educational Needs in the area of Learning (SEN-L). However, recent studies indicate that students with SEN-L from special schools show difficulties in basic arithmetical operations, and the development of basic mathematical skills during secondary special school is not warranted. This study presents a newly developed test of basic arithmetical skills, based on already established tests. The test examines the arithmetical skills of students with SEN-L from fifth to ninth grade. The sample consisted of 110 students from three special schools in Munich. Testing took place in January and June 2013. The test shows to be an effective tool that reliably and precisely assesses students' performance across different grades. The test items can be used without creating floor and ceiling effects among fifth to ninth grade students with SEN-L. The items' conformity to the dichotomous Rasch model is demonstrated. The students' skills turn out to be very heterogeneous, both overall and within grades. Many of the students do not even master basic arithmetical skills that are taught in primary school, although achievement improves in higher grades.

Keywords: Arithmetical Skills; Curriculum Based Measurement; Special Needs

Corresponding author: Markus Gebhardt, School of Education, TU München, Arcisstraße 21, 80333 München, Germany, markus.gebhardt@tum.de

1.   Schooling of Students with SEN

The schooling of children with *Special Educational Needs* (SEN) is a controversial issue in school policies (European Agency for Development in Special Needs Education, 2007). It has been shown that students in integrative educational settings show superior school performance (particularly in mathematics) and, in the long run, show greater social skills than students in special schools (Baker, Wang, & Walberg, 1995; Carlberg & Kavele, 1980; Eckhart, Haeberlin, Sahli Lozano, & Blanc, P., 2011; Haeberlin, Blanc, Eckhart, & Sahli-Lozano, 2012; Haeberlin, Bless, Moser, & Klaghofer, 1991; Merk, 1982; Wang & Baker, 1986). Longitudinal research among students with SEN in German-speaking regions showed a delay in school achievement of at least two years compared to children of a corresponding grade in a regular school (Haeberlin et al., 1991). The *Hamburg School trials* showed that the performance gap appeared in second grade and increased up to fourth grade, even in classes with particularly good inclusive care (Hinz, Katzenbach, Rauer, Schuck, Wocken, & Wudtke, 1998). Cross-sectional studies confirm these findings (Tent, Witt, Bürger, & Zschoche-Lieberum, 1991; Wocken, 2000, 2005; Wocken & Gröhlich, 2007). Seventh grade students with SEN-L in special schools did not accomplish the requirements of fifth grade students in a general-education secondary school (Hauptschule; Wocken, 2000). In Germany in 2010, however, only 22% of the students with SEN and 23% of the students with *SEN in the area of Learning* (SEN-L) were in integrative settings (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2010). Nevertheless, the integration rate is rising slowly.

In the USA, the statistics about the school performance of students with SEN draw a similar picture. In the *Special Education Elementary Longitudinal Study* (SEELS; Schiller, Sandford, & Blackorby, 2008), children with SEN between the ages of 10 and 17 (N=5400) were observed over a period of six years. Results showed that 60% of students with Learning Disabilities (LD) in segregated settings and 32% of students with LD in integrative classes achieved the lowest performance level in mathematics (lower than the 20th percentile; Schiller et al., 2008). In secondary school, the performance gap between the students with and without SEN continues to widen. In ninth grade, the delay ranges from 3 to 4.9 years on average for students with LD, 1 to 3 years for students with *emotional disturbance* and more than five years for students with *intellectual disabilities* (Blackorby, Chorost, Garza, & Guzman, 2003). The individual growth over three school years varies widely, but in general, there are no significant differences in the magnitude of growth between the students with different types of SEN (Blackorby et al., 2003). This kind of longitudinal study is missing in the German speaking countries.

2.   Identification of Students with SEN-L in Germany

In almost all school systems, children with SEN are identified to give them a legal right to additional resources and support in school, but the concepts of LD vary widely from country to country. As a consequence, the size of the population of children with diagnosed LD is different in any given country (Sideridis, 2007). In the USA, for example, 5% of the student population is classified as having LD (Hallahan, Lloyd, Kauffman, Weiss, & Martinez, 2005). In Germany, 3% of all students are identified as students with SEN-L (KMK Statistics, 2010). These students have basic difficulties in various learning areas. Traditionally, in German-speaking countries, next to pervasive difficulties in school learning, an IQ below 85 (but above 70, thus excluding intellectual disability) was considered as the most effective diagnostic criterion of SEN-L, since this allowed a general "objective" assessment of a child's cognitive performance without using school indicators (Grünke, 2004). The categorization of students with SEN-L in Germany is similar to the international definition of LD by Lloyd, Keller, and Hung (2007). This definition refers to significant academic difficulties in school, for which neither other disabilities (e.g., sensory impairment, intellectual disability or emotional and behavioral disorders) nor lack of schooling can be found as cause (Lloyd et al., 2007). Students with a diagnosed *dyslexia* or *dyscalculia* are not identified as students with SEN in Germany (Büttner & Hasselhorn, 2011). Identification of students with SEN-L and, therefore, the

allocation of special educational resources to the school only applies to children with severe learning difficulties (Klauer & Lauth, 1997; Schröder, 2008).

Since the diagnosis of SEN-L appears not caused by somatic-medical reasons, but rather by the specific criteria of a given school system, the diagnosis of SEN-L is under constant legitimacy pressure. IQ testing has been criticized since the 1970s (Bundschuh, 2010), both by psychologists and, especially, by teachers and educational practitioners, and consequently, IQ is no longer used as the sole indicator of SEN-L in present governmental recommendations in Germany. Nevertheless, many researchers still regard low intellectual abilities as the most important aspect of diagnosing SEN-L (Kretschmann, 2006) and recommend the administration of a language-free IQ test in addition to standardized academic achievement tests as part of the diagnostic process (Kany & Schöler, 2009; Kottmann, 2006). We hope that the instrument under construction that is presented in this article will provide an additional means for improved objective diagnosis of SEN-L in the future.

3.   Basic Mathematical Skills

One third of the students with SEN-L, who have graduated from special schools, cannot handle numbers adequately and also have great trouble solving simple division tasks (Lehmann & Hoffmann, 2009). Students show problems with the understanding of word problems, division, the decimal system, and the doubling or halving of numbers (Moser Opitz, 2007). The lack of elementary arithmetic skills is mainly responsible for mathematical difficulties in secondary school. Basic mathematical skills require knowledge of quantity and numbers as well as operation rules (Ehlert, Fritz, Arndt, & Leutner, 2013; Ennemoser, Krajewski, & Schmidt, 2011). A cross-sectional study by Krajewski and Ennemoser (2010) showed that basic skills are not only acquired in elementary school, but also trained in secondary school classes. However, the level of mastery of these basic skills of students in different school tracks is very diverse. High school fifth graders in Gymnasium (grammar school) show better mastered basic skills than students in the eighth grade of Hauptschule (lower track of secondary school; Ennemoser et al., 2011). Only one study exists in integrative classes which includes students with SEN. An Austrian study carried out in urban integrative classes showed that the level of basic skills was also very heterogeneous (Gebhardt, Schwab, Schaupp, Rossmann, & Gasteiger-Klicpera, 2012). Even pupils without SEN-L had great difficulties in basic arithmetic. As a matter of fact, more than 30% of the regular students (without SEN) in fifth grade scored more than one standard deviation below the mean on a standardized school test (lower than the 16th percentile). Students with SEN-L were able to solve tasks regarding additions and subtractions, but had significant problems with tasks concerning multiplications and divisions in the number range up to 10,000 (Gebhardt et al., 2012). In German-speaking regions, research on the academic performance of students with SEN-L is mostly performed in intervention studies (Hecht, Sinner, Kuhl, & Ennemoser, 2011; Moog, 1993, 1995; Moog & Schulz, 1997, 2005; Sinner & Kuhl, 2010). These studies generally observed significant effects immediately following the interventions, but follow-up results again showed large differences between students with SEN-L and regular students with learning difficulties. When the training in basic mathematical skills ended, the students with SEN-L regressed to the same low level they showed before the intervention (Hecht et al., 2011; Sinner & Kuhl, 2010). All intervention studies used grade based standardized school-tests, which were constructed with classical test theory. However, when overlooking these various studies, which show the specific difficulties of students with SEN-L, it would be very useful to have one diagnostic tool that addresses the various arithmetic sub-skills and that is specifically tailored to this special population.

4.   Research Question

Special needs students show a one- to three-year delay in their development of basic arithmetic skills. The problem with standardized school tests is that they were developed and standardized for average students in the regular curriculum and, as a consequence, have difficulty displaying the academic growth of

students with SEN-L. Adapting such tests raises challenges with respect to the measurement's discriminatory power (e.g., ceiling and floor effects).

Another possibility is to use *Curriculum-based measurements* (CBM) to examine academic growth of students with SEN (Deno, 2003). Tests that are actually available were constructed with *classical test theory*. However, to measure academic progress, item response theory would be the better option (Klauer, 2011; Wilbert & Linnemann, 2011) since these models avoid certain methodological flaws that are associated with tests constructed with classical test theory (such as unreliability of the change scores and incomparability of the scale units of the subsequent measures). Our goal is to longitudinally assess the students' arithmetic skills and to evaluate the achievements of students of different ages, both criterion-based and norm-based. This can be achieved by using instruments that show conformity to specific models from item response theory.

Assessing basic arithmetical skills, the instrument developed in the longitudinal study on student development in integrative classes SILKE (Schulische Integration im Längsschnitt – KompetenzEntwicklung bei SchülerInnen mit und ohne SPF in der Sekundarstufe I; Academic integration in a longitudinal study – development of competences of students with and without SEN in secondary schools; Gebhardt, 2013; Gebhardt, Schwab, Krammer, & Gasteiger-Klicpera, 2012; Gebhardt, Schwab, Schaupp et al, 2012; Schwab, 2013), is used in this study to assess SEN-L students in separated special schools. In contrast to students without SEN, students in these special secondary schools are still explicitly taught in elementary arithmetical skills and these need to be addressed in the test.

The aims of this pilot study, hence, are the following:

− Apply the instrument *assessing basic arithmetical skills* to assess the arithmetical skills of a sample of SEN-L students and evaluate the scale's conformity to the *dichotomous Rasch model*.
− Explore the instrument's characteristics regarding discriminatory power, as well as classical psychometric criteria.
− Explore the basic arithmetical achievement of students with SEN-L in special schools, especially in respect to its development across the secondary school grades (cross-sectional), across one school year (longitudinal), as well as the interaction between these two factors.

5. Method

5.1 Design and Sample

The study was carried out in three special schools in Munich in January and June 2012, which constitute the middle (t1) and the end (t2) of the school term, respectively. At both times of measurement, 62 male and 48 female students ($N = 110$) with SEN-L from fifth to ninth grade were tested with the same instruments. At t1, students were 13.9 years old on average ($SD = 1.6$). Students took tests in groups in sessions of about 15 to 20 minutes, but they could take as much time as needed. If a student did not answer an item, the test administrator reminded the student to do his very best to do so. As all items comprise free response formats guessing behavior can be neglected. Table 1 shows the distribution of the sample across grades.

Table 1

Distribution of participants across school grades

Grade	n	Female	Male	Age
5	20 (18%)	35%	65%	11.9 (0.6)
6	23 (21%)	48%	52%	13.1 (0.7)
7	14 (13%)	43%	57%	13.8 (0.6)
8	33 (30%)	48%	52%	15.0 (0.6)
9	16 (15%)	38%	62%	16.0 (0.6)
Total	110 (100%)	44%	56%	13.9(1.6)

5.2   Instruments

On the basis of the arithmetic tests *Eggenberger Rechentest 3+ (ERT 3+*; Holzer, Schaupp, & Lenart, 2010) and *ERT 4+* (Schaupp, Lenart, & Holzer, 2010), an instrument was devised that consists of the ERT-scales, as well as additional, newly constructed items to handle the large heterogeneity in the target population and to avoid floor and ceiling effects. The ERT was originally designed to assess arithmetical skills at the end of the third (3+) and the fourth grade (4+) of elementary school. Ennemoser et al. (2011) differentiate arithmetic skills into knowledge of quantity as well as numbers and operation rules. In the currently devised instrument this differentiation is reflected in its subtests: *Knowledge of quantity* is represented by the subtests *writing numbers from dictation* and *number series*; *numbers and operation rules* is represented by the subtests *Basic arithmetical skills* and *word problems*. For the adapted instrument, the 12 items of the ERT 4+ subtest *number series* were used, which measures knowledge about the place-value system. Furthermore, the subtest *Basic numeracy* (comprising 13 items) was used, dealing with addition, subtraction, multiplication and division. The *placeholder task* is another subtest taken from ERT 4+, consisting of 6 items in which 2 numbers are given and the student has to find the third (e.g., ___ + 8 = 21). The subtest *word problems* comprise 9 items and was taken from ERT 3+ to match the students' levels and to avoid floor effects. Table 2 presents the final instrument with its four subtests.

Table 2

Subtests of the final instrument before item-selection procedure

	Subtest	Origin	*n* Items
	Basic arithmetical skills	ERT 4+: Basic numeracy	13
		ERT 4+: Placeholder	6
		Constructed by authors	15
	Word problems	ERT 3+: Word problems	9
Precursors	*Number series*	ERT 4+ Number series	12
		Constructed by authors	2
	Writing numbers from dictation	Constructed by authors	14

5.3   Analyses

To test the subtests' unidimensionality, the data were checked for conformity to the dichotomous Rasch model. This means, all items pertaining to the same subtest were scaled in one model. Then, to check the models' conformity with regard to *specific objectivity*, the independence of item parameters across subsamples was evaluated. These subsamples were chosen using two split criteria: raw score median (thus creating two achievement groups) and gender (Kubinger, 2005). Andersen's *Likelihood Ratio Test* (LRT;

Andersen, 1973), which is based on Conditional Maximum Likelihood estimates, was used to indicate items' conformity or non-conformity. For testing the items' fit to the model, the so-called *Waldtest* was used, which indicates the item parameter's deviance from the model while taking the estimates' standard error into account (Fischer & Scheiblechner, 1970). All analyses reported in this article were conducted with the Software R (R Core Team, 2013) and more specifically the package *eRm* (Mair, Hatzinger, & Maier, 2012) which was used for estimating item parameters and calculation of goodness of fit tests, as well as the package *PP* (Reif, 2012) for estimating person parameters.

To analyze students' ability and development in arithmetical skills, the person (ability) parameters were estimated using the item parameters from *t1*. These allowed to estimate person parameters for *t1* as well as for *t2* and, consequently, to map these abilities on one scale. In this case, *Warm Maximum Likelihood* estimates were used, as these allow for the estimation of extreme abilities, especially regarding possible 0 scores in the SEN-L group.

6.   Results

6.1   Scaling and Item-Selection Procedure

The scaling process was based on the data of t1 and afterwards crosschecked with the data of t2, taking into account its interdependency. After removing two items from the subtest *word problems* and one item from each of the other subtest, all items showed conformity to the dichotomous Rasch model. The subsequent quasi-cross-validation using t2 data was also successful. Only for Word Problems and Writing numbers from dictation the Gender effect reached significance, but all other tests were not significant. Table 3 presents the statistical values of the final Rasch models for the four subtests. The Andersen LRTs showed to be not significant for the final selection of items (1% level of significance was chosen to avoid accumulation of type-I-errors; cf. Kubinger, 2005), which indicates conformity to the dichotomous Rasch model, both with respect to t1 and t2 data. As the Andersen LRT uses CML-estimates, item parameters could not be estimated for items that were solved by all or never solved in the subsamples (the number of items is labeled with NA in Table 3).

Table 3

Statistical values of the final Rasch models for the four subtests

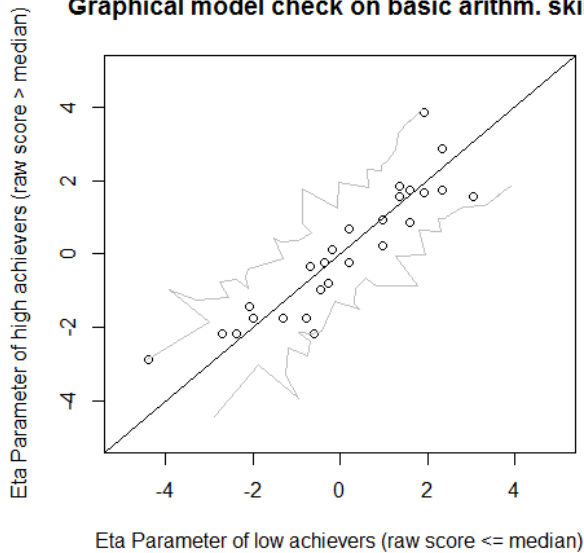|		Split criterion	LRT χ^2	df	$\chi^2_{\alpha=.01}$	p	Items	
							Removed	NA
Basic arithmetical skills	t1	Raw Score Median	42.6	28	48.3	.04	1	3
		Gender	45.8	31	52.2	.04		0
	t2	Raw Score Median	31.1	28	48.3	.32		3
		Gender	23.5	32	53.5	.86		0
Number series	t1	Raw Score Median	9.4	10	23.2	.50	1	2
		Gender	16.2	11	24.7	.14		1
	t2	Raw Score Median	10.8	8	20.1	.22		4
		Gender	22.6	11	24.7	.02		1
Word problems	t1	Raw Score Median	6.6	3	11.3	.08	2	3
		Gender	12.1	5	15.1	.03		1
	t2	Raw Score Median	5.0	4	13.3	.28		2
		Gender	14.6	5	15.1	.01		1

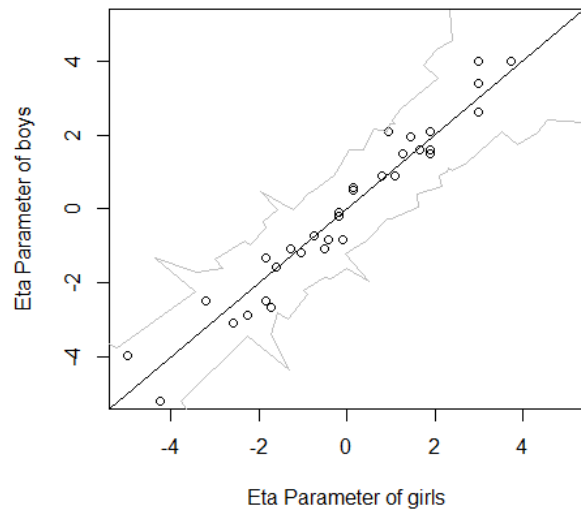|								
Writing numbers from dictation	t1	Raw Score Median	8.8	6	16.8	.18		6
		Gender	12.5	11	24.7	.33	1	1
	t2	Raw Score Median	16.8	7	18.5	.02		5
		Gender	25.7	12	26.2	.01		0

Note. All tests show to be not significant at 1% level, indicating conformity to the Rasch model. The NA-column indicates the number of items that could not be evaluated due to 0% or 100% correct in the subsample.

To illustrate the results of the item-selection procedure, a graphical representation of the model check of the subtest *basic arithmetical skills* at t2 is shown in Figure 1. Nearly all items are situated in the region of acceptable deviance, which is indicated by the gray control line. Acceptable deviance is defined in regard to the standard error of estimations in the respective area on the logit scale (cf. Wright & Stone, 1999). Furthermore, the standard errors of the estimations appear to be in an acceptable range (min = 0.2, mean = 0.3, max = 0.8, across all subtests and both times of measurement).

Graphical model check on basic arithm. skills

Eta Parameter of high achievers (raw score > median)

Eta Parameter of low achievers (raw score <= median)

Graphical model check on basic arithm. skills

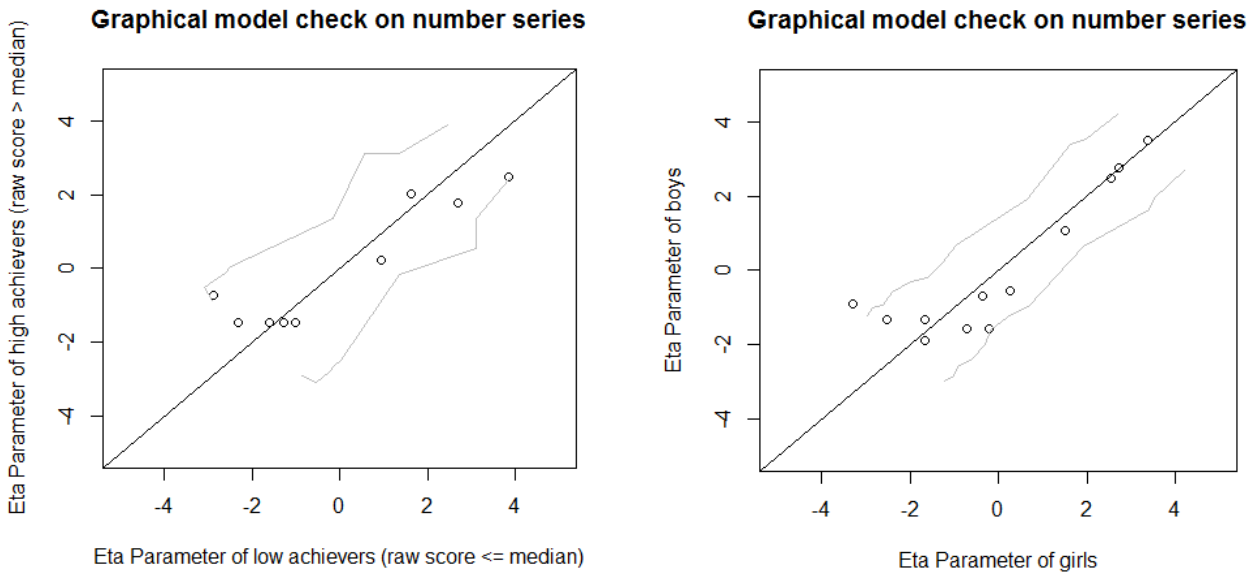Eta Parameter of boys

Eta Parameter of girls

Figure 1. Graphical model checks of the subtest basic arithmetical skills (top) and number series (bottom) by raw score-median (left) and gender (right). The gray line indicates the limit of acceptable deviance for single items (cf. text).

Finally, the total instrument with the four subtests comprising 33, 8, 12 and 13 items, respectively, also showed conformity to the dichotomous Rasch model. Table 4 shows that the items present a wide range of difficulty levels, both overall and across grades in nearly every subtest, leading to a reliable assessment across a broad range of ability. Only the subtest *writing numbers from dictation* shows a more narrow range of item difficulty for 9th graders, which might lead to a small ceiling effect for these students.

Table 4

Proportion correct within subtests across grades, including all selected items.

Grade	Basic arithmetic			Number series			Word problems			Writing numbers		
	Lo	Hi	M	Lo	Hi	M	Lo	Hi	M	Lo	Hi	M
5	.00	.85	.27	.00	1.00	.40	.00	.65	.24	.11	.95	.50
6	.00	.91	.39	.04	1.00	.57	.00	.91	.30	.30	1.00	.69
7	.00	.93	.46	.14	1.00	.70	.00	1.00	.41	.36	1.00	.79
8	.00	.97	.54	.21	.97	.67	.03	1.00	.44	.42	1.00	.81
9	.00	1.00	.68	.44	1.00	.85	.19	1.00	.62	.81	1.00	.97

Note. Lo = Lowest value, Hi = Highest value, M = Mean

The subtest reliabilities (Cronbach α) are presented on the diagonal of Table 5. The reliabilities vary from .72 to .92, which is above the conventional cut-off-value of .80, except for the subtest *word problems*, of which the reliability is still very acceptable. It should be mentioned that items that function conform the Rasch model are, as such, internally consistent because unidimensionality is included in the theoretical formulation of the model. Table 5 further reports high inter-correlations between the subtests, ranging from .64 between *number series* and *writing numbers from dictation* to .74 between *number series* and *word problems*.

Table 5

Reliabilities and inter-correlations between the subtests at t1

	(1)	(2)	(3)	(4)
Basic arithmetic skills (1)	.92	.75	.74	.72
Number series (2)		.86	.67	.64
Word problems (3)			.72	.66
Writing numbers (4)				.85

Note. The subtests' Cronbach α is presented on the diagonal.

6.2   Basic Arithmetical Achievement of Students with SEN-L

Students' achievement, in the form of their person (ability) parameter, was very heterogeneous in every subtest and in every grade. Person parameters referring to the subtest *basic arithmetical skills* showed standard deviations from 1.7 (on the logit scale) in grade six to 2.4 in grade nine. The dispersion in achievement did not show a trend across grades in terms of reduced or increased standard deviations. Linear regression shows that achievement in every subtest at t1 is predicted by grade (α = .05), with effects ranging from β = 0.47 in *word problems* to β = 0.58 in *basic arithmetical skills*. These relations were also significant at t2, but decreased in effect size, which were now ranging from β = 0.37 in *word problems* to β = 0.41 in *writing numbers from dictation*. The moderate relationships between grade and ability confirm the instrument's developmental validity. However, it must be noted that students from grades 7 and 8 showed very similar levels of achievement in every subtest and at both measurement points, except for *basic arithmetical skills*, in which 8th grader scored 0.7 logits higher than 7th graders at t1, but this difference vanished at t2.

When shifting from cross-sectional analysis to a longitudinal analysis of the development of achievement from t1 to t2, further differences between the subtests become evident. Two subtests appeared to group together with regard to development of mean achievement: In the *basic arithmetical skills* and the *writing numbers for dictation* subtests, students from lower grades somewhat improved over time, while those from higher grades regressed (see Figure 2). In the other two subtests, *number series* and *word problems*, students from every grade improved over time. However, these are descriptive tendencies and in terms of significance only *number series* showed a longitudinal main effect (d = 0.22). An ANOVA for repeated measurements shows that the factor time plays a significant role, $F(1, 101)$ = 8.6, p = .00, η^2 = .08. Although the interaction term did not reach significance, especially students from grade five increased in their achievements (+1.3 logits). A significant interaction effect between development and grade was found in *basic arithmetical skills*: $F(1, 101)$ = 3.9, p = .01, η^2 = .14. This indicates that students in lower grades improve their basic arithmetical skills over time while those in higher grades do not, or even drop in performance (d = 0.34 for 5th graders, d = 0.20 for 6th graders, d = -0.12 for 7th graders, d = -0.53 for 8th graders, d = -0.41 for 9th graders).
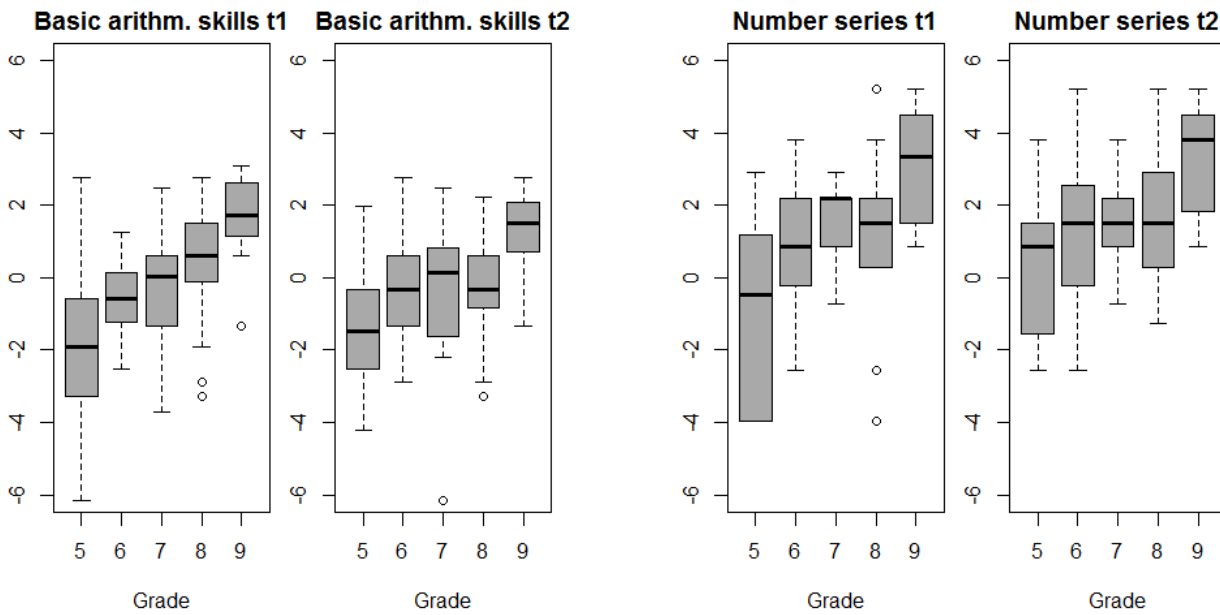
Figure 2. Development of person (ability) parameter distributions from t1 to t2 of the two subtests *basic arithmetical skills* (left) and *number series* (right), for each grade separately.

Research in the field of special education is particularly interested in the students' performance variations. Table 6 shows the students' mean ability parameters on all 4 subtests and for all grades separately, in the context of temporal development. The fifth and sixth graders show improvement on all subtests. However, the mean scores of students in 7th, 8th and 9th grade decreased in *basic arithmetical skills* and *writing numbers for dictation*, but remained stable or improved on *number series* and *word problems*. Overall, a regression-to-the-mean-effect was found. I.e., students with low scores tended to improve their scores whereas students with high scores tended to show a decrease at t2. This was confirmed by weak to moderate negative correlation between learning gains (t2 - t1) and achievement at t1 in *basic arithmetic skills* ($r = -.55$), *number series* ($r = -.38$), *word problems* ($r = -.38$) and *writing numbers* ($r = -.35$).

Table 6

Mean (M) values and standard deviations (SD) of person parameters per grade at t1 and t2

Grade	Basic arithmetic skills				Number series			
	M t1	M t2	SD t1	SD t2	M t1	M t2	SD t1	SD t2
5	-2.0	-1.4	2.2	1.7	-1.0	0.4	2.4	1.9
6	-0.6	-0.4	1.0	1.4	0.7	1.2	1.8	2.2
7	-0.2	-0.4	1.6	2.2	1.6	1.7	1.1	1.1
8	0.5	-0.2	1.5	1.3	1.4	1.9	1.8	1.8
9	1.7	1.2	1.1	1.2	3.2	3.3	1.5	1.5

Grade	Word problems				Writing numbers f. dictation			
	M t1	M t2	SD t1	SD t2	M t1	M t2	SD t1	SD t2
5	-2.4	-1.9	2.2	2.5	0.3	0.6	2.2	2.6
6	-1.6	-1.3	1.7	1.7	2.0	2.3	2.0	2.0
7	-0.7	-0.3	1.9	2.4	3.0	2.6	1.5	1.8
8	-0.4	-0.4	1.7	2.3	3.0	2.4	1.8	2.0
9	1.0	1.3	2.4	2.1	4.8	4.3	1.0	1.3

7.   Discussion

The instrument described in this article showed conformity to the dichotomous Rasch model. It also did not show remarkable ceiling or floor effects and, thus, allowed to measure basic arithmetical performance of students with SEN-L in special schools. Only the newly constructed subtest *writing numbers from dictation* showed a somewhat narrow range of item difficulties for 9th graders. This is not unexpected, since these students should already have acquired the basic competence of knowledge of quantity (Krajewski & Ennemoser, 2010). It would further be questionable if additional, more difficult items would measure the same construct. Two items of the subtest *word problems*, which was taken from the ERT 3+, had to be rejected and this scale should be further improved. Nevertheless, the instrument showed similar results as those found in the SILKE study in integrative classes (Gebhardt, 2013; Gebhardt, Schwab, Schaupp et al., 2012; Schwab, in press) and allowed a first exploration of the basic performance of students with SEN-L in special schools. Generally, students with SEN-L lag several years behind their peers without SEN. They are still learning what the other students learn in primary school and especially the basics of multiplication and division are taught to them in secondary school (see also Moser Opitz, 2007).

The inter-correlations of the subtest showed that the performance levels were similar across the subtests and, empirically, it would be sufficient to describe a student with only one scale score, indicating arithmetical ability. However, since the subtest scores are indicative of the development of different arithmetical skills, these should provide support for fitting an appropriate arithmetic curriculum of students with SEN-L. Thus, the results should help improve the construction of real curriculum based measurement of arithmetic for students with learning disabilities.

The instrument discriminated between the grades. Although the grade level showed medium effects on all subtests at t1 and t2, the heterogeneity of student performance within the grades was very large. This means that it is necessary to have different mathematical problems with varying levels of complexity available to be able foster the mathematical abilities of all students (Moser Opitz, 2007). Similar findings were described previously in several intervention studies (Hecht et al., 2011; Moog & Schulz, 1997, 2005; Sinner & Kuhl, 2010), but until now, the arithmetical performance of students with SEN had not been measured with a Rasch scaled standardized test.

One important finding of the longitudinal results was that students from every grade improved on the subtests *number series* and *word problems*, while only the 5th and 6th graders improved on the subtests *writing numbers from dictation* and *basic arithmetical skills*. This might be explained by the fact that the curriculum in 5th and 6th grade includes teaching basic arithmetical skills, whereas the curriculum of grades 7 to 9 prepares the students for vocational training. In these grades, basic skills are no longer explicitly trained, but instead, new operations such as fractions are introduced. As the old skills are not explicitly consolidated, basic arithmetic skills (including writing numbers form dictation) ~~and~~ from 3rd grade in primary school may again become a challenge for students in the 9th grade of special schools (see, e.g., Steiner, 2009). Another factor influencing the results, might be that the special school students who are performing well in 5th and/or 6th grade can attain integrative classes in 7th grade. Since such students "disappear" to other classes or schools, the cross-sectional data presented here cannot be interpreted in the same way as real longitudinal data. The present data must be viewed as giving explorative information, also when considering the relatively small sample that was included in this study. A much larger sample must be tested to draw stronger conclusions.

Finally, the development of basic arithmetical skills in this study was relatively limited. This underlines the challenge of teaching basic arithmetical skills in special schools and the, currently, rather limited success. Instruments such as the one presented in this article, that allow the continuous measurement of a series of arithmetical skills in secondary special education, may help to further develop evidence based interventions that are tailored to the needs of the students. When measurement and intervention are adapted to the needs of the students, they can jointly help in improving the students' arithmetic abilities.

Keypoints

- Students with special educational needs from special schools show difficulties in basic arithmetical operations.
- A newly developed Rasch scaled instrument allows the reliable measurement of basic arithmetical skills of students with SEN-L in secondary education.
- Students' skills turn out to be very heterogeneous, both overall and within grades.
- Many students do not even master arithmetical skills that are taught in primary school, although achievement improves in higher grades.

References

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, *38*(1), 123–140. Doi: 10.1007/BF02291180

Baker, E. T., Wang, M. C. & Walberg, H. J. (1995). The Effect of Inclusion on Learning. *Educational Leadership, 52*(4), 33–35.

Blackorby, J., Chorost, M., Garza, N., & Guzman, A. M. (2003). The Academic Performance of Secondary School Students with Disabilities. In U.S. Department of Education (Eds.), *The Achievement of Youth with Disabilities during Secondary School. A Report from the National Longitudinal Transition Study 2.* Menlo Park, CA: SRI International. Retrieved from http://www.seels.net/designdocs/SEELS_W1W3_FINAL.pdf

Bundschuh, K. (2010). *Einführung in die sonderpädagogische Diagnostik (7th ed.)*. München: E. Reinhardt.

Büttner, G. & Hasselhorn, M. (2011). Learning Disabilities: Debates on definitions, causes, subtypes and responses. *International Journal of Disability, Development and Education, 58*(1), 75–87. Doi: 10.1080/1034912X.2011.548476

Carlberg, C. & Kavale, K. (1980). The Efficacy of Special versus Regular Class Placement for Exceptional Children: A Meta-Analysis. *The Journal of Special Education, 14*(3), 295–309. Doi: 10.1177/002246698001400304

Deno (2003). Curriculum-based measurment. *Journal of Special Education*, *37*, 184–192. Doi:10.1177/00224669030370030801

Eckhart, M., Haeberlin, U., Sahli Lozano, C., & Blanc, P. (2011). *Langzeitwirkungen der schulischen Integration.* [Long-term effects of school integration]. Bern, Switzerland: Haupt Verlag.

Ehlert, A., Fritz, A., Arndt, D. & Leutner, D. (2013). Arithmetische Basiskompetenzen von Schülerinnen und Schülern in den Klassen 5 bis 7 der Sekundarstufe. *Journal für Mathematik-Didaktik, 34*(2), 237–263. Doi:10.1007/s13138-013-0055-0

Ennemoser, M., Krajewski, K., & Schmidt, S. (2011). Entwicklung und Bedeutung von Menge-Zahlen-Kompetenzen und eines basalen Konventions- und Regelwissens in der Klasse 5 bis 9. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *34*(4), 228–242. Doi: 10.1026/0049-8637/a000055

European Agency for Development in Special Needs Education (2007). *Assessment in Inclusive Settings. Key Issues for Policy and Practice.* Odense, Denmark: European Agency for Development in Special Needs Education.

Fischer, G. H., & Scheiblechner, H. H. (1970). Algorithmen und Programme für das probabilistische Testmodel von Rasch. [Algorithms and programs for Rasch's probabilistic test model.]. *Psychologische Beiträge*, *12*, 23–51.

Gebhardt, M. (2013). *Integration und schulische Leistungen in Grazer Sekundarstufenklassen. Eine empirische explorative Pilotstudie*. Wien: Lit Verlag.

Gebhardt, M., Schwab, S., Krammer, M., & Gasteiger-Klicpera, B. (2012). Achievement and integration of students with special needs (SEN) in the fifth grade. *Journal of Special Education and Rehabilitation, 13*, 7–19. Doi: 10.2478/v10215-011-0022-6

Gebhardt, M., Schwab, S., Schaupp, H., Rossmann, P., & Gasteiger-Klicpera, B. (2012). Heterogene Gruppen in mathematischen Grundfertigkeiten: Eine explorative Erkundung der Fähigkeiten im Grundrechnen in Integrationsklassen der 5. Schulstufe. *Zeitschrift für Inklusion (online)*, (1-2). Available from www.inklusion-online.net/index.php/inklusion/article/view/155/147

Grünke, M. (2004). Lernbehinderung. In G. W. Lauth, & M. Grünke (Eds.), *Interventionen bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (pp. 65–77). Göttingen: Hogrefe, Verl. für Psychologie.

Haeberlin, U., Blanc, P., Eckhart, M., & Sahli-Lozano, C. (2012, May). *Intégration scolaire d'enfants en difficultés d'apprentissage: effets à long terme. Information sur la recherche éducationnelle, CSRE, n° 12:021*. [School integration of children with learning difficulties: long-term effects. Information about educational research, CSRE, n° 12:021]. Retrieved May 31, 2013 from www.skbf-csre.ch/de/bildungsforschung/datenbank/.

Haeberlin, U., Bless, G., Moser, U., & Klaghofer, R. (1991). *Die Integration von Lernbehinderten: Versuche, Theorien, Forschungen, Enttäuschungen, Hoffnungen*. Bern: Haupt.

Hallahan, D. P., Lloyd, J. W., Kauffman, J. M., Weiss, M. P., & Martinez. (2005). *Learning disabilities: Foundations, characteristics, and effective teaching*. Needham Heights: Allyn & Bacon.

Hecht, T., Sinner, D., Kuhl, J., & Ennemoser, M. (2011). Differenzielle Effekte eines Trainings der mathematischen Basiskompetenzen bei kognitiv schwachen Grundschülern und Schülern der Förderschule mit dem Schwerpunkt Lernen – Reanalysen zweier Studien. *Empirische Sonderpädagogik, (4)*, 308–323.

Hinz, A., Katzenbach, D., Rauer, W., Schuck, K. D., Wocken, H. & Wudtke, H. (1998). *Die Integrative Grundschule im sozialen Brennpunkt: Ergebnisse eines Hamburger Schulversuchs*. Hamburg: Hamburger Buchwerkstatt.

Holzer, N., Schaupp, H., & Lenart, F. (2010). *Eggenberger Rechentest (ERT 3+): Diagnostikum für Dyskalkulie für das Ende der 3. Schulstufe bis Mitte der 4. Schulstufe*. Bern: Huber.

Kany, W., & Schöler, H. (2009). *Diagnostik schulischer Lern- und Leistungsschwierigkeiten: Ein Leitfaden mit einer Anleitung zur Gutachtenerstellung*. Stuttgart: Kohlhammer.

Klauer (2011). Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik*, (3), 207–224

Klauer, K. J., & Lauth, G. W. (1997). Lernbehinderungen und Leistungsschwierigkeiten bei Schülern. In F. E. Weinert (Eds.), *Enzyklopädie der Psychologie, Psychologie des Unterrichts und der Schule* (pp. 701–738). Göttingen: Hogrefe.

Kubinger, K. D. (2005). Psychological test calibration using the Rasch model – Some critical suggestions on traditional approaches. *International Journal of Testing*, 5(4), 377–394.

Lauth, G. W., & Grünke, M. (Eds.). (2004). *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis*. Göttingen: Hogrefe, Verl. für Psychologie.

Kottmann, B. (2006). *Selektion in die Sonderschule: Das Verfahren zur Feststellung von sonderpädagogischem Förderbedarf als Gegenstand empirischer Forschung*. Bad Heilbrunn: Klinkhardt.

Krajewski, K., & Ennemoser, M. (2010). Entwicklung mathematischer Basiskompetenzen in der Sekundarstufe. *Empirische Pädagogik, 24*(4), 353–370.

Kretschmann, R. (2006). Diagnostik bei Lernbehinderungen. In U. Petermann, & F. Petermann (Eds.), *Diagnostik sonderpädagogischen Förderbedarfs* (pp. 139–162). Göttingen: Hogrefe.

Lehmann, R., & Hoffmann, E. (2009). *Berliner Erhebung arbeitsrelevanter Basiskompetenzen von Schülerinnen und Schüler und Schüler mit dem Förderbedarf „Lernen"*. Münster: Waxmann.

Lloyd, J. W., Keller, C., & Hung, L. (2007). International Understanding of Learning Disabilities. *Learning Disabilities Research & Practice*, 22(3), 159–160. Doi: 10.1111/j.1540-5826.2007.00240.

Mair, P., Hatzinger, R., & Maier, M. J. (2012). *eRm: Extended Rasch Modeling*. R package version 0.15–1.

Merk (1982). Lernschwierigkeiten – Zur Effizienz von Fördermaßnahmen an Grund- und Lernbehindertenschulen. *Heilpädagogische Forschung,* 1, S. 53–69.

Moog, W. (1993). Schwachstellen beim Addieren – eine Erhebung bei lernbehinderten Sonderschülern. *Zeitschrift für Heilpädagogik, 44,* 534–554.

Moog, W. (1995). Flexibilisierung von Zahlbegriffen und Zählhandlungen – Ein Übungsprogramm. *Heilpädagogische Forschung, 21*(3), 113–121.

Moog, W., & Schulz, A. (1997). Das Dortmunder Zahlbegriffstraining – Lernwirksamkeit bei rechenschwachen Grundschülern. *Sonderpädagogik, 27*(2), 60–68.

Moog, W., & Schulz, A. (2005). *Zahlen begreifen: Diagnose und Förderung bei Kindern mit Rechenschwäche* (2. überarb. Aufl.). Weinheim: Beltz.

Moser Opitz, E. (2007). *Rechenschwäche – Dyskalkulie: Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen und Schüler*. Bern: Haupt.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Reif, M. (2012). *PP: Person Parameter estimation*. R package version 0.2.

Schaupp, H., Lenart, F., & Holzer, N. (2010). *Eggenberger Rechentest (ERT 4+): Diagnostikum für Dyskalkulie für das Ende der 4. Schulstufe bis der Mitte der 5. Schulstufe*. Bern: Huber.

Schiller, E., Sanford, C. & Blackorby, J. (2008). *The Achievments of Youth with Disabilities during Secondary School: A Report from the National Longitudinal Transition Study 2* (U.S. Department of Education, Hrsg.). Retrieved from http://www.seels.net/info_reports/SEELS_LearnDisability_%20SPEC_TOPIC_REPORT.12.19.08ww _FINAL.pdf.

Schröder, U. (2008. *Lernbehindertenpädagogik: Grundlagen und Perspektiven sonderpädagogischer Lernhilfe* (2nd ed.). Stuttgart: Kohlhammer.

Schwab, S. (2013). *Schulische Integration, soziale Partizipation und emotionales Wohlbefinden in der Schule - Ergebnisse einer empirischen Längsschnittstudie*. Berlin: Lit-Verlag.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2010). *Sonderpädagogische Förderung in Schulen 1999 bis 2008: Dokumentation Nr. 189 – März 2010*. Retrieved from http://www.kmk.org/fileadmin/pdf/Statistik/Dok_189_SoPaeFoe_2008.pdf.

Sideridis, G. D. (2007). International Approaches to Learning Disabilities: More Alike or More Different? *Learning Disabilities Research & Practice*, *22*(3), 210–215. Doi: 10.1111/j.1540-5826.2007.00249.x

Sinner, D., & Kuhl, J. (2010). Förderung mathematischer Basiskompetenzen in der Grundstufe der Schule für Lernhilfe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42*(4), 241–251.

Steiner, G. (2009). Forgetting while learning: A plea for specific consolidation. *Journal of Cognitive Education and Psychology, 8*, 117–127.

Tent, L., Witt, M., Bürger, W., & Zschoche-Lieberum, C. (1991). Ist die Schule für Lernbehinderte überholt? *Heilpädagogische Forschung*, (1), 289-320.

Wang, M. C. & Baker, E. T. (1985-86). Mainstreaming Programs: Design Features and Effects. *The Journal of Special Education, 19*(4), 503–521.

Wilbert, J., & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik, 3*, 225–242.

Wocken, H. (2005). *Andere Länder, andere Schüler?: Vergleichende Untersuchungen von Förderschülern in den Bundesländern Brandenburg, Hamburg und Niedersachen*. Potsdam. Retrieved from http://bidok.uibk.ac.at/library/wocken-forschungsbericht.html

Wocken, H. (2007). Fördert Förderschule? Eine empirische Rundreise durch Schulen für „optimale Förderung". In I. Demmer-Dieckmann, & A. Textor (Eds.), *Integrationsforschung und Bildungspolitik im Dialog* (pp. 35–60). Bad Heilbrunn: Klinkhardt.

Wocken, H. & Gröhlich, C. (2009). Kompetenzen von Schülerinnen und Schülern an Hamburger Förderschulen. In: W. Bos, M. Bonsen & C. Gröhlich (Hrsg.), *KESS 7 – Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7* (pp. 133–142). Münster: Waxmann.

Wright, B.D., & Stone, M.H. (1999). *Measurement essentials*. Wide Range Inc.: Wilmington. Retrieved from: http://www.rasch.org/measess/me-all.pdf