Vidmar, Maša; Niklas, Frank; Schneider, Wolfgang; Hasselhorn, Marcus

On-entry assessment of school competencies and academic achievement. A comparison between Slovenia and Germany

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
European journal of psychology of education 32 (2017) 2, S. 311-331

Bitte verwenden Sie beim Zitieren folgende URN /
Please use the following URN for citation:
urn:nbn:de:0111-pedocs-174345 - *http://nbn-resolving.org/urn:nbn:de:0111-pedocs-174345*
DOI: 10.1007/s10212-016-0294-9 - *http://dx.doi.org/10.1007/s10212-016-0294-9*

On-entry assessment of school competencies and academic achievement: A comparison between Slovenia and Germany

Abstract

The foundation of school success is laid early in children's lives. Consequently, assessments of academic precursors may help to identify children in need of additional support. Such early assessments could also be interesting from an international perspective when educational systems are compared. This analysis is used to inform on the comparability of Slovenian and German versions of the English on-school-entry assessment tool "Performance Indicators in Primary School" (PIPS, Tymms and Albone 2002). PIPS was also used to predict later academic achievement in the two national samples. The German sample consisted of 468 children with a mean age of about 6;6 years at school entry (48.7% girls). In Slovenia, 328 children (49% girls) were assessed (mean age of about 6;3 years at school entry). Multi-group confirmatory factor analyses for PIPS did not support weak measurement invariance. However, results indicated that the number of factors as well as the pattern of loadings seem to be comparable. Further research is needed to examine in which respects PIPS might work as a tool for international comparisons. Structural equation modelling indicated that PIPS can be used as a predictor of academic achievement and that overall academic achievement could be predicted best by early numeracy. PIPS measures of literacy and numeracy skills were specific and significant predictors of children's later language and math achievement in Grade 1.

Key words: Performance Indicators in Primary School (PIPS), academic achievement, primary school children, early numeracy, early literacy, international comparison, on-entry assessment

To perform well in academic settings is a major developmental task for children and youth in modern Western societies (Masten et al. 1995). The foundation of success in school is laid fairly early in the school career of children, and early achievement in primary school is among the best predictors for later academic achievement (Schneider 2009; Stern 2009). But even before school enrolment an enormous variability exists in young children's cognitive performance and also in other school-related competencies (Dowker 2005; Schneider et al. 2014). These competencies consist of basic skills and abilities as well as knowledge and attitudes that help a child to function successfully in a school environment, both academically and socially (Hair et al. 2006; Wilkinson et al. 1998).
Consequently, early assessments of school-relevant competencies (i.e., at the start of formal education) might help to identify children who are at risk of developing problems later in school, to prevent these problems by informing tailored interventions, and to support teachers in their lesson planning. Such assessments might also be interesting from an international perspective on which we elaborate in more detail below.
In this study we focus on cognitive competencies as they have been proven to be of upmost importance (Duncan et al. 2007). However, this does not diminish the importance of other school competencies such as social or emotional skills (cf. Niklas 2011).

Early numeracy and early literacy

To identify children at risk for later school failure early, it is important to assess precursors of academic achievement. Over the last decades researchers have identified specific cognitive competencies that predict later academic achievement. For instance, early numeracy refers to counting, geometric reasoning, understanding part–whole relation, digit identification, whereas early literacy refers to phonological and linguistic awareness, letter recognition, vocabulary and familiarity with the function of print (McWayne et al. 2004; Niklas 2011). These competencies have been found to predict academic achievement in the early years of primary school (e.g., Aunio and Niemivirta 2010; Betts et al. 2009; Byrnes and Wasik 2009). In a Slovene study, specific on-entry competencies were consistently the best predictors of academic achievement in Grade 1, even after including children's extraversion, conscientiousness, intelligence, school attitudes, self-regulation, parenting practices, and parental involvement in the analytic model (Vidmar 2010).

An important predictor for later linguistic competencies in school is early letter knowledge (e.g., Näslund and Schneider 1996; Torppa et al. 2006, Preßler et al. 2014). For instance, Torppa and colleagues (2006) showed that children's early letter knowledge acted as a mediator between children's actual reading competence and precursors such as phonological awareness and vocabulary.

Specific precursors of later mathematical competencies have also been identified. In particular the early understanding of quantities, numbers, and the relation between quantities and numbers seem to be important (Geary 2011; Krajewski and Schneider 2009). The ability to understand these concepts is a specific predictor of mathematical achievement above and beyond the contribution of more general abilities such as intelligence. Intelligence on the other hand is a fairly unspecific but still important predictor of general academic achievement as well as mathematical and literacy competencies (Krajewski and Schneider 2009; Schneider and Näslund 1999; Schneider et al. 2014).

Assessment of early cognitive competencies from a cross-national perspective

Instruments for on-entry assessments of children's cognitive competencies not only need to be objective, reliable and valid (including the prediction of later academic achievement), but they also must be suitable for younger children as well as economical, so that all children starting school can readily be tested. Given their importance for later academic achievement, such instruments should focus on early mathematical and literacy abilities. Another aspect to be considered is the comparability of test results across different samples from different regions or countries. Such comparisons can provide baseline information about the students entering compulsory education and thus may be an indicator for the effectiveness of early childhood education and care systems or for identifying an age-range for a successful school-entry (Tymms et al. 2004); information that will be of interest for policy makers and educators. Fair and relevant international comparisons of students' attainments are helpful tools for countries to learn from each other. For instance, the Governing Board of OECD PISA (Programme for International Students Assessment) has shown interest in iPIPS, an international study on early learning outcomes. The aim of iPIPS is the linkage between PISA results and early years' assessments (Schleicher 2013). However, so far only very few test instruments for younger children exist, that fulfil all these aspects, despite the fact that such an instrument would help to contextualise and interpret assessments later in school (Tymms et al. 2014).

An economical instrument for the assessment of on-entry competencies that considers specific precursors such as those described above and also allows for the international comparison is the "Performance Indicators for Primary Schools" (PIPS, Tymms and Albone 2002). This test instrument assesses both initial linguistic and mathematical competencies at school entry and the further development of these competencies during Grade 1. It can be used in a computerized or a booklet version. PIPS has been used in many different countries for almost two decades (see Tymms and Albone 2002). More than 4500 schools worldwide apply this test currently, and more than one million children have been assessed across several countries (Bäuerlein et al. 2010; Bäuerlein et al. 2014; Tymms et al. 2014). Both retest-reliability and validity are warranted, and the test score predicts later academic achievement very well (Tymms and Wylde 2004). During the test duration of about 25 minutes "early reading/literacy", "early mathematics/numeracy" and "phonological awareness" are assessed. Although PIPS is a well-established test instrument in English-speaking countries (e.g., Wildy and Styles 2008) new versions in German and Slovenian have been developed only recently (Bäuerlein et al. 2010, 2014; Vidmar and Zupančič 2006, 2007a).

The present study and research questions

Whereas assessments and international comparisons of older children are fairly common and suitable instruments exist, there are almost no such tools for children starting formal schooling. Although the PIPS instrument has been used successfully in several countries to assess young children, international comparisons have been rarely conducted and done mostly between English-speaking countries (e.g., Tymms et al. 2004; Tymms et al. 2014). This study aims to fill in this gap by examining the PIPS instrument in two non-English speaking countries in Europe, and at the same time informing future endeavours in the development of international early year assessments.

In this study PIPS was applied on both a Slovene and a German sample, and equivalence of the two national adaptations was tested. In both countries children enter school at about the same age. However, in Germany slightly more children aged 3 and 4 attend formal early childhood education and care settings (e.g., OECD 2014).

The second aim was to analyse how well the assessment is able to predict general academic achievement as well as specific academic achievement in math and language one and two years after school entry. With this approach we tried to answer the following research questions:

1.) Is there sufficient overlap between the different versions of PIPS used in Slovenia and Germany to do cross-country comparisons and are the two versions measuring the same latent constructs? Do children from both countries differ in their abilities?

2.) Is the PIPS instrument a good predictor of later academic achievement in school in both countries?

We expected that a cross-country comparison between Slovenia and Germany would be possible in general. Here, comparisons for the mathematical tasks should be readily possible, whereas comparisons for the language based tasks should be more difficult and probably not possible for phonological awareness and vocabulary (Tymms et al. 2014). In addition, we expected children to perform at about the same performance level. Further, we expected PIPS to be a good predictor of academic achievement in school in general, and the mathematical and literacy subtests to be specific predictors of mathematical and language achievement.

Method

Both studies were carried out in the context of larger longitudinal studies. Here, only relevant data are analysed and presented. In both cases assessments took place within seven weeks after school enrolment (T1), and at the ends of Grade 1 (T2) and Grade 2 (T3).

Participants

German Sample

Data assessment in Germany was carried out in the context of the project “School-prepared child” (Hasselhorn et al. 2012). In total, 465 children (48.5% girls) could be tested at T1. At the time of school enrolment the children were between 67 to 96 months old (M = 77.7 months, SD = 4.1) and 46.8% had a migration background, indicating that at least one of their parents were born outside of Germany. The socioeconomic status (SES) was assessed with the “Wegener-scale” (Wegener 1988) an occupation prestige score that ranges from 20 (unskilled labour) up to 186.8 (physician). In the German sample the average highest prestige score in a household was M = 74.2 (SD = 32.7).

Slovene sample

In the Slovene study, 326 children (48.2% girls), their mothers and teachers participated. At T1 children were between 69 and 89 months old (M = 74.6; SD = 3.6 months). At T1 mothers provided data on their education (M = 12.5 years; SD = 3.1 years; used as an indicator of SES) and home language (91% of mothers reported Slovenian as home language). Children’s teachers (N = 87) rated the children at T2 and T3.

Instrumentation

On-entry assessment

In both samples the PIPS instrument (CEM Centre 2005, a, b; Tymms and Albone 2002) was used to assess mathematical and linguistic competencies at the beginning of Grade 1. For instance, the early reading score consists of vocabulary, letter identification, ideas about reading, recognition of words, and reading of sentences or short stories. Children had to repeat syllables and words and identify rhymes for the assessment of phonological awareness. Mathematical competencies were obtained using addition and subtraction tasks, naming of numbers, calculations with coins, and continuing of rows of numbers in a meaningful way. The psychometric properties of the English version have been well-established (test-retest reliability: r_{tt} =.90 to .98; see CEM Centre 2001; Tymms et al. 2004; prediction of later academic achievement see Tymms 1999; Tymms 2001; Tymms et al. 2007).

The German version of PIPS („Fähigkeitsindikatoren Primarschule“, FIPS) is a computer-administered test instrument with good psychometric properties, which has been piloted and normed in the last few years (Bäuerlein et al. 2010, 2014).

The Slovene version of PIPS assessment (Vidmar and Zupančič 2006, 2007a) consists of a booklet that includes instructions for administrators and pictures to be used in the assessment.

German and Slovene PIPS versions are very similar, with 14 and 12 different aspects of literacy and numeracy measured in both versions. Some differences can be found due to the different assessment format (computerized versus booklet) and national specifics (see below and Table 1 for comparison).

PIPS comparability for the two national versions

Some subtests appeared only in one of the two versions (e.g., Counting in Slovenia) and were dropped from the comparison. Eleven subtests were similar across the two versions; however, two “Phonological awareness” subtests and “Vocabulary” were dropped based on recommendations by Tymms et al. (2014).

An additional caveat in comparing the two versions was that the Slovene version had been modified based on the results of a pilot study (for details see Vidmar and Tymms 2009). Items that were *not* modified were kept in further analyses. The eight subtests that were compared are shown in Table 1. In three subtests some items were not exactly the same, but these subtests were still kept in the analyses: Letter identification (the Slovene version has upper case letters only and the German version has upper and lower case letters), Reading (the sentences and stories were mere translations from English resulting in different number of words across the two versions), and Number identification (different two- and three-digit numbers were used).

Academic achievement

In the German sample, at T2 and T3 reading abilities of the children were assessed with a silent reading task (“Würzburger Leise Lese Probe - Revision”, WLLP, Schneider et al. 2011, r_{tt} = .93 for Grade 1 to Grade 4). Here, children had to read words and then mark the one matching picture out of four given pictures. In addition, children’s spelling was assessed by two standardized dictations at T2 and T3 (“Deutscher Rechtschreibtest für das erste und zweite Schuljahr”, DERET, Stock and Schneider 2008, Cronbach’s α = .92). At T3 children’s reading comprehension was assessed with a standardized test, assessing word, sentence, and text comprehension (“Ein Leseverständnistest für Erst- bis Sechstklässler”, ELFE, Lenhard and Schneider 2006, Cronbach’s α = .92-.97).

Children’s mathematical abilities at T2 and T3 were obtained with the curriculum-based and standardized tests DEMAT 1 and DEMAT 2+ (“Deutscher Mathematiktest für erste/zweite

Klassen”; Krajewski et al. 2002, Cronbach’s α = .89; Krajewski et al. 2004, Cronbach’s α = .93). The psychometric characteristics of all these tests are excellent.

In the Slovene sample academic achievement at T2 and T3 was assessed with teacher-rated Attainment of Performance Standards: Slovene, Math and Nature & Society (Vidmar and Zupančič 2007b; Zupančič 2006). These standards were curriculum-based and were rated on a 6-point scale (from 0 = does not achieve to 5 = overachieves greatly). The child’s score for each subject is calculated as a mean of teacher’s ratings on the respective items (depending on the subject and grade between 25 and 62 items were rated). Cronbach’s alpha ranged from .98 to .99. Previous studies supported the validity of the instrument (e.g. Zupančič and Kavčič 2007b, 2008).

Intelligence

In the German sample, the Columbia Mental Maturity Scale (CMM, Burgemeister et al. 1972) was used to assess nonverbal intelligence 18 months before school entry. The children had to identify the odd picture in an array of four or five pictures (e.g. one spoon and three forks). Reliability and prognostic validity of the CMM are warranted (Esser 2002).

In the Slovene sample, the Colored Progressive Matrices (CPM, Raven et al. 1999) were administered in Grade 1 to assess general nonverbal cognitive ability. Children had to identify the correct section that completes the pattern in the picture. The reliability and validity of the instrument have been well established (Raven et al. 1999; Zupančič and Kavčič 2007a).

Questionnaires

In both samples information on education or occupation of the parents (socioeconomic status) as well as information on migrant background was obtained by parent surveys.

Procedure

In Germany, intelligence was measured in an individual assessment and parents answered a questionnaire 18 months prior to school entry. At T1 individual assessments, and at T2 and T3 group assessments were carried out.

In Slovenia, at T1 trained research assistants administered the on-entry assessment. Later in Grade 1, trained research assistants also administered intelligence tests individually and parents completed a questionnaire. At T2 and T3 children's teachers received instruments to report academic achievement via mail.

Overview of statistical analyses

SPSS was used to calculate descriptive statistics and correlations. We conducted confirmatory factor analyses (CFAs) for PIPS using Mplus, which was also used for structural equation modelling (SEM) to analyse the prediction of academic achievement from on-entry assessment. Two types of models were tested: (1) general models in which PIPS early literacy and numeracy predicted overall academic achievement in Grades 1 and 2, and (2) specific models in which the prediction from early literacy to language-related academic achievement was analysed as well as the prediction from early numeracy to math-related academic achievement. In addition, in both samples children's gender, intelligence and SES were used in the models as covariates. Mplus uses full information maximum likelihood (FIML) to assess parameters in the model. This algorithm also deals with the problem of missing values (Schafer and Graham 2002). When using FIML, algorithm parameter estimates and associated standard errors are based on all the information available (Peugh and Enders 2004). This means that even participants with missing data were kept in the analyses.

The fit of the models was evaluated using multiple indices: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). There are different guidelines to determine an acceptable (good or adequate/mediocre) fit of a model: e.g., CFI > .90 or .95, RMSEA < .05 or .06 or .08, SRMR < .08 (see Browne and Cudeck 1993; Hu and Bentler 1999; Hair et al. 1998; Marcoulides and Hershberger 1997). Some researchers (e.g. Kenny 2014) suggested RMSEA = .10 as the cut-off for not well fitting models. However, recent studies by Chen and colleagues (2008) and Fan and Sivo (2007) illustrated problems with universal cut-off points for fit indices. In our analyses, we used the following fairly lenient cut-offs: CFI > .90, RMSEA < .10 and SRMR < .08.

Results

Descriptive results

Means and standard deviations for the study variables are presented in Tables 1 and 2.

-please insert Tables 1 and 2 about here-

Table 1 indicates significant mean differences of the PIPS subtests between the Slovene and German samples. For each country, all indicators of academic achievement were combined in a latent construct for overall academic achievement. In regard to the language model, teacher-reported attainment of curriculum standards for Slovenian language and DERET, WLLP, and ELFE in the German sample were used. For the math model, teacher-reported attainment of curriculum standards for Math were used in the Slovene sample, and the tests DEMAT1 and 2 were used in the German sample.

Following the recommendation of Curran et al. (1996), no variables needed to be transformed due to excessive skew or kurtosis (the only exception with excessive skew and kurtosis was Reading in the German sample).

Correlations

Correlations among measures of literacy, numeracy and academic achievement were calculated. All but one PIPS early literacy indicators were positively interrelated. Correlations varied from r = .21 to .67 (all ps < .001) in the Slovene sample, and from r = .10

to .58 in the German sample (all *ps* < .05); only the correlation between the subtests "Ideas about Reading" and "Writing" was non-significant in the latter sample. Positive correlations were also found among PIPS early numeracy indicators (r = .52 to .56 in Slovenia and r = .28 to .39 in Germany; all p < .001). As expected, academic achievement indicators showed high stability across time (r = .78 to .84 in Slovenia and $|r|$ = .33 to .77 in Germany; all p < .001) and were highly associated with each other within time (r = .37 to .67 in Slovenia and $|r|$ = .35 to .68 in Germany; all p < .001; negative correlations in the German sample were observed as a child's results in the standardized dictation was measured by the number of misspellings). Correlations between PIPS indicators and academic achievement indicators are shown in Table 3.

-please insert Table 3 about here-

As can be seen in Table 3, all PIPS indicators were positively correlated with academic achievement indicators in the Slovene sample (r = .19 to .52; all p < .001). In the German sample, correlation coefficients were somewhat smaller but still mostly significant (r = .10 to .47; all p < .05). Non-significant correlations were found for the subtests "Writing", "Ideas about Reading", and "Calculations with pictures" with academic achievement indicators in Grade 2.

In both samples, all indicators correlated significantly in the expected direction with intelligence (r = .23 to .51 in Slovenia and $|r|$ = .15 to .39 in Germany; all *p*s < .01). Whereas correlations with SES were also all significant in the Slovene sample, they were lower and more often non-significant in the German sample. Correlations with gender were significant for about half of the indicators in both samples. In general, girls had higher scores on language-related indicators, whereas boys had higher scores on math-related indicators.

Confirmatory Factor Analyses (CFAs) and Structural Equation Models (SEMs)

CFAs for PIPS indicators were conducted for both samples in one step[1]. The model with two latent constructs – PIPS early literacy and PIPS early numeracy (indicators are presented in Table 1) demonstrated the following fit: CFI = .92, SRMR = .06, RMSEA = .11, χ^2 (15) = 155.67, *p* < .001. The fit indices did not indicate a perfect fit. All model-estimated loadings were significant (*p* < .001), and the correlation between the latent constructs was .71 (*p* < .001). Measurement errors of some indicators were correlated as indicated by the modification indices. Fit indices before modification indices were as follows: CFI = .84, SRMR = .08, RMSEA = .14, χ^2 (19) = 303.01, *p* < .001; by adding these correlations we followed common practice to avoid developing the incorrect models (MacCallum 1986); only one correlation was added at a time, and only in cases where there was sufficient justification. Within early literacy, 'Ideas about Reading' were correlated with 'Letter identification', 'Word recognition' and 'Reading'; within early numeracy, 'Calculations with pictures' was correlated with 'Math'. These correlations of measurement errors were included because the two subtests of early numeracy measured arithmetic abilities in children, whereas the third subtest 'Number identification' measured just basic numeracy knowledge. 'Ideas about Reading' was closely related to the other literacy subtests as some of the items measured similar knowledge, such as "Can you show me a word/sentence?". Despite a relatively high correlation between the literacy and numeracy constructs we decided to keep them as separate constructs, given that we were interested in whether early numeracy and literacy differed in how well they predicted later academic achievement.

---

[1] Because children were nested within schools (39 and 28 schools in the Slovene and German sample, respectively), multilevel models were employed in preliminary analyses to examine design effects (computed based on intra-class correlation: 1 + (mean cluster size – 1) · ICC*;* Snijders and Bosker 1999) for PIPS indicators. The great majority of variables demonstrated design effects smaller than 2 indicating a low effect of school clustering (Muthen 1999). Consequently, multi-level models were not employed.

In a next step, multi-group CFAs were conducted to test factorial invariance across the two groups (i.e., Slovene and German sample). This procedure tests construct comparability and assesses whether test scores measure the same construct of interest on the same metric across both countries (Wu et al. 2007). First, configural (nonmetric) invariance was tested to explore whether the same pattern of fixed and free loadings holds for each group (Widaman and Reise 1997). Fit indices demonstrated an acceptable fit of this model: CFI = .94, SRMR = .04, RMSEA = 0.10, $\chi 2$ (30) = 143.66, $p < .001$. This result could be considered as support of configural invariance (Wu et al. 2007), indicating that the number of factors as well as the pattern of loadings seems to be comparable in the Slovene and German models. In the following step, weak metric invariance was tested. Loadings of indicators of the latent variable were constrained to be equal across the two national samples. The $\chi 2$-difference test ($\Delta\chi 2(df) = \chi 2(df)weak - \chi 2(df)configural$; $\Delta\chi 2(6) = 50.05$, $p < .001$) was significant. Indicating that the assumption of weak metric invariance was not confirmed. Other recommended cut-off points for testing loading invariance were also indicative of non-invariance: ΔCFI = -.024, ΔRMSEA = .007, ΔSRMR = .038 (see Chen 2007). Therefore weak invariance was rejected.

Next, SEM was used to analyse the prediction of academic achievement from on-entry numeracy and literacy (PIPS). The models were run separately for the German and Slovene samples. Two types of models were tested: (1) general models in which early literacy and numeracy predicted overall academic achievement (see Figure 1) and specific models; i.e., a literacy-related model and a math-related model (see Figures 2 and 3).
Numeracy and literacy indicators were the same in both samples (see Table 1). The correlated measurement errors from CFA were kept in these analyses; some were also added based on modification indices. In a final step covariates were included in the models.

-Please insert Figure 1 about here-

The SEM model for overall academic achievement in both samples demonstrated acceptable fit (CFI = .94, SRMR = .05, RMSEA = .09, χ^2 (67) = 241.83, $p < .001$ for Slovenia; CFI = .91, SRMR = .07, RMSEA = .08, χ^2 (80) = 323.43, $p < .001$ for Germany). Explained variance for academic achievement in Grade 1 was $R^2 = .28$ and .48 and in Grade 2 $R^2 = .46$ and .95 for the Slovene and the German sample, respectively ($p < .001$). Results indicated that early numeracy was predictive of academic achievement in Grade 1 in both samples. In the Slovene sample, early numeracy also predicted academic achievement in Grade 2. In the German sample, the only predictor of academic achievement in Grade 2 was academic achievement at Grade 1. Early literacy did not significantly predict academic achievement in the Slovene sample, whereas in the German sample early literacy was a significant predictor of academic achievement in Grade 1. Academic achievement showed moderate (Slovenia) and very high (German) stability.

In the models depicted in Figure 1 gender, intelligence and SES were added as covariates (paths from covariate to all latent constructs were added in the models; non-significant paths were dropped). Results were similar to the ones presented in Figure 1. Fit indices were somewhat lower but still adequate; the size and significance of estimated coefficients changed for one path in each sample (as indicated in Figure 1 with #).

In a next step, the specific models (literacy-related and math-related) were analysed for each national sample. The same indicators that were used in the general models were also used in the specific models. The exception was math academic achievement in the German sample; here, scores on subtests of the DEMAT were used in the specific model instead of the total DEMAT score. The correlated measurement errors from CFA remained the same as above.

-please insert Figure 2 about here-

The literacy models in both samples demonstrated acceptable fit (CFI = .97, SRMR = .03, RMSEA = .08, χ^2 (12) = 35.31, $p < .001$ for Slovenia; CFI = .98, RMSEA = .05, SRMR = .04, χ^2 (32) = 73.54, $p < .001$ for Germany). The explained variance for academic achievement in Grade 1 was R^2 = .96 and .32 and in Grade 2 R^2 = .96 and .82 for the Slovene and German samples, respectively. In both samples early literacy predicted language achievement in Grade 1 and Grade 1 achievement predicted Grade 2 achievement.

In the models depicted in Figure 2 gender, intelligence and SES were added as covariates using the same procedure as described for general models. The model-estimated coefficients and fit indices in both samples remained very similar to the ones shown in Figure 2.

-Please insert Figure 3 about here-

The math models demonstrated an almost acceptable fit for the Slovene sample (CFI = .96, SRMR = .04, RMSEA = .11, χ^2 (5) = 23.29, $p < .001$) and an acceptable fit for the German sample (CFI = .91, SRMR = .06, RMSEA = .05, χ^2 (205) = 465.26, $p < .001$). Explained variance for math academic achievement in Grade 1 was R^2 = .93 and .59, and in Grade 2 R^2 = .95 and .57 for the Slovene and German sample, respectively. Our results indicate that early numeracy may be a good predictor of math achievement in Grade 1, and Grade 1 math achievement predicted Grade 2 math achievement.

Again gender, intelligence and SES were added as covariates using the same procedure as described for the general models. The model-estimated coefficients and fit indices in both samples were very similar to the ones shown in Figure 3.

Discussion

The test instrument Performance Indicators for Primary Schools (PIPS) has been applied successfully in thousands of schools worldwide for a long time (Bäuerlein et al. 2010; Tymms and Albone 2002). PIPS has been proven to be a reliable and valid instrument at school-entry that predicts children’s later academic achievement very well (Tymms and Wylde 2004) and that can be used for cross-country comparisons (Tymms et al. 2014). However, little was known about the characteristics of the recently developed adaptations of PIPS for Germany and Slovenia (see Bäuerlein et al. 2010, 2014; Vidmar and Zupančič

2006, 2007a). The present study aimed at filling this gap by testing the comparability of the two national versions (using multi-group CFA) and examining the prediction of academic achievement in primary school (using SEM).

Comparability of the two language versions

The descriptive comparison of the Slovene and German PIPS versions revealed that eight subtests assessing children's literacy and numeracy competencies were very similar across the two adaptations. Although comparisons of these mean subtests across the two countries showed some differences in child performance, this is not indicative of measurement invariance as differences in subtests can reflect true differences in latent constructs. These differences need to be interpreted cautiously given the results of the multi-group CFA. Whereas, support for configural invariance was found, indicating that the number of factors as well as the pattern of loadings seems to be comparable in the Slovene and German models, weak invariance was not found (the size of the loadings could not be constrained equally across the two samples). This indicates that one unit change in the item score is *not* scaled to an equal unit change in the factor score across groups (Wu et al. 2007). In other words, factor loadings contribute differently to the means, thus preventing valid and comparable score estimates, thereby also preventing the comparison of literacy and numeracy means.

Further comparisons using different samples are needed to test whether measurement equivalence of the German and Slovene PIPS is indeed only configural and in which regard PIPS may be used for direct comparisons of educational systems.

Prediction of academic achievement

The second aim of the study was to investigate the prediction of academic achievement based on PIPS numeracy and literacy scores. Reliably assessing children's literacy and numeracy competencies early is an important step in order to prevent negative educational tracks for children (Cuna and Heckman 2010). Providing teachers with information on children’s early abilities helps them to adjust their teaching accordingly and thus support their pupils in an adequate and positive way. Such feedback is much appreciated by teachers (Bäuerlein et al. 2014).

Our results indicate that general academic achievement could be best predicted by early numeracy. This replicates Tymms’ findings on the sample of English pupils (Tymms et al. 2007). Despite different approaches in the measurement of academic achievement (i.e., teacher reports versus standardized tests) similar findings were obtained supporting the robustness of the finding across both countries. The findings align with results reported for the English version, for which high correlations and regression coefficients between PIPS scores and reading and math achievement measured three years (Tymms 1999, 2001) and seven years later (Tymms et al. 2007) were found.

A comparison of the results for the German and Slovene sample showed that in the German sample, early literacy explained about 9% of the variance of Grade 1 academic achievement when covariates were not included in the model. However, in the Slovene sample, early literacy was not a significant predictor when early numeracy also was included in the model. This might be due to the fact that in Slovenia general academic achievement was assessed with only one of the three indicators related to language, whereas in the German sample two out of the three indicators were language-related.

In addition, in the Slovene sample early numeracy predicted Grade 2 academic achievement when covariates were not included in the model. This is probably due to lower stability of the academic achievement construct compared to the German model.

In regard to the specific models that showed better model fits, PIPS numeracy was a strong predictor of math academic achievement in Grade 1 in both samples, even after controlling for gender, SES and intelligence. Similarly, PIPS literacy was predictive of language academic achievement in Grade 1 in both samples. Tymms and colleagues (2007) found that early numeracy was most predictive of math achievement seven years later; however, early literacy was not a significant predictor of reading achievement seven years later when early numeracy and some other predictors were included in the regression model.

The results may be viewed in support of predictive validity of the PIPS assessment for academic achievement in Grade 1 in Germany and Slovenia; prediction of academic achievement in Grade 2 (after controlling for the stability of the achievement construct) was not significant. The results indicate that early numeracy abilities could be the most important predictor of further academic success; a finding robust across samples and achievement measures (Duncan et al. 2007).

Further investigation of assessment equivalence is needed as well as the prediction of later academic achievement. Testing measurement invariance across any two national versions of PIPS is a necessary prerequisite for comparisons across countries. Since in our study data up to Grade 2 was assessed, future studies should aim to analyse whether children's academic achievement in higher Grades can also be predicted by PIPS. Preliminary analyses using the German PIPS version indicate that on-entry assessments with PIPS are good predictors of spelling abilities until the end of primary school (Ehm and Hasselhorn 2014).

Limitations

Several limitations mark this study. Firstly, the Slovene and the German PIPS version differed in regard to several subtests and in some cases also in regard to individual items. Consequently, not all relevant early child competencies could be compared (e.g., Torppa et al. 2006) and the comparison of results in both countries has to be interpreted with caution. Secondly, in the Slovene sample no standardized child assessments of academic achievement were conducted, instead teacher ratings were used. However, these ratings have been shown to be reliable and valid (Vidmar 2011; Zupančič and Kavčič 2008). Thirdly, PIPS assesses precursors of academic achievement at school entry, that is, after children develop important early competencies and after first interventions may already have been occurred (e.g. Chittleborough et al. 2014; Niklas et al. in press; Niklas and Schneider 2015). For instance, Chittleborough and colleagues (2014) showed that early interventions before school entry can improve skills necessary for educational success and reduce socioeconomic inequality. However, PIPS offers a good and economical alternative to be used universally at school entry, and thus can still be used earlier than many other available standardized assessment tools.

Strengths

Despite these limitations, this study has several strengths. To our knowledge, this is the first comparative study on PIPS for samples from countries where English is not the first language (see Tymms et al. 2014; Wildy and Styles 2008). The analyses indicate that despite restricted comparability, PIPS data from different samples and from different countries can be described with the same number of factors as well as the pattern of loadings (although the loadings are on a different scale). PIPS can also predict general and specific academic achievement.

All analyses were conducted with larger sample sizes, using advanced statistical tools (FIML, SEM). Reliable and valid test instruments were used to assess children's academic achievement. In addition, we controlled for child and family characteristics. The findings also indicated that PIPS is a reliable and valid test instrument, not only in the English but also in the Slovene and German context.

Conclusion

In this study, recently developed Slovene and German versions of the English test instrument PIPS were analysed and compared. For some of our models, we did not find a perfect fit. However, our results indicate that the number of factors as well as the pattern of loadings seems to be comparable in the Slovene and German models (but the literacy and numeracy means could not be compared directly). PIPS still proved to be a reliable and valid school entry assessment tool in both contexts. This indicates that instruments can be useful within different cultures even when the measurement equivalence is not met as is the case with PIPS (Van de Vijver and Leung 1997). Further research needs to test in which regards PIPS might work as a tool for international comparisons.

Overall, our results also indicate that academic achievement seems to be closely associated with PIPS numeracy test scores in both samples (after controlling for intelligence and SES, see also Duncan et al. 2007). Early assessments of children‘s literacy and numeracy abilities with PIPS were specific and significant predictors of children’s later language and math achievement in Grade 1. Consequently, this assessment tool seems suited to support teacher’s work with young school children to identify strengths and weaknesses of these children and to provide them with the early support they need.

References

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in Grade one by early numeracy. *Learning and Individual Differences, 20*, 427-435.

Bäuerlein, K., Faust, G., Franz, U., Jost, M., & Schneider, W. (2010). Fähigkeitsindikatoren Primarschule (FIPS) – Erprobung und Weiterentwicklung einer internationalen computerbasierten Schulanfangsdiagnose [Performance Indicators for Primary Schools (PIPS) – testing and development of an international computerized test instrument for the beginning of school]. In M. Hasselhorn & W. Schneider (Eds.), *Frühprognose schulischer Kompetenzen, Tests und Trends, N.F., Vol. 9* (pp. 85-108). Göttingen: Hogrefe.

Bäuerlein, K., Niklas, F. & Schneider, W. (2014). Fähigkeitsindikatoren Primarschule (FIPS) - Überprüfung des Lernerfolgs in der ersten Klasse [Performance Indicators for Primary Schools (PIPS) – Evaluation of learning success in Grade 1]. In M. Hasselhorn, W. Schneider & Trautwein, U. (Eds.), *Lernverlaufsdiagnostik, Tests und Trends, N.F., Vol. 12* (pp. 127-144). Göttingen: Hogrefe.

Betts, J., Pickart, M., & Heistad, D. (2009). Construct and predictive validity evidence for curriculum-based measures of early literacy and numeracy skills in kindergarten. *Journal of Psychoeducational Assessment, 27*(2), 83-95.

Browne M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.). *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Burgemeister, B., Blum, L., & Lorge, J. (1972). *Columbia Mental Maturity Scale*. New York: Harcourt Brace Jovanovich.

Byrnes, J. P., & Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third Grades: An opportunity–propensity analysis. *Contemporary Educational Psychology*, *34*, 167-183.

CEM Centre (2001). *Performance Indicators in Primary Schools* (Technical report 2001). Durham, UK: University of Durham.

CEM Centre (2005a). *Performance Indicators in Primary Schools: On-entry baseline assessment [Booklet format].* Durham, UK: University of Durham.

CEM Centre (2005b). *Performance Indicators in Primary Schools: On-entry baseline assessment: Follow-up [Booklet format].* Durham, UK: University of Durham.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cut-off points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494.

Chittleborough, C. R., Mittinty, M. N., Lawlor, D. A., & Lynch, J. W. (2014). Effects of simulated interventions to improve school entry academic skills on socioeconomic inequalities in educational achievement. *Child Development, 85*(6), 2247-2262.

Cuna, F., & Heckman, J. (2010). *Investing in our young people*. Cambridge, MA: National Bureau of Economic Research.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16-29.

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*(4), 324-332.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446.

Ehm, J. H., & Hasselhorn, M. (2014, November). Prediction of spelling abilities in grade 4: Results from the German longitudinal study „Schulreifes Kind". Presentation held at the Scottish Educational Research Association (SERA) Annual Conference 2014, Edinburgh, UK.

Esser, G. (2002). *BUEVA – Basisdiagnostik für umschriebene Entwicklungsstörungen im Vorschulalter* [BUEVA – basis diagnostics for developmental disorders at pre-school age]. Göttingen: Beltz.

Fan, X., & Sivo, S. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509-529.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47* (6), 1539-1552.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first Grade. *Early Childhood Research Quarterly, 21,* 431–454.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hasselhorn, M., Schöler, H., Schneider, W., Ehm, J.-H., Johnson, M., Keppler, I., Krebs, K., Niklas, F., Randhawa, E., Schmiedeler, S., Segerer, R., & Wagner, H. (2012). Gezielte Zusatzförderung im Modellprojekt "Schulreifes Kind" – Auswirkungen auf

Schulbereitschaft und schulischen Lernerfolg [Focused additional support in the project „Schoolready child" – impact on school readiness and school success]. *Frühe Bildung, 1*(1), 3-10

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Kenny, D. A. (2014). Measuring Model Fit. Retrieved February 3, 2016 from URL: http://davidakenny.net/cm/fit.htm

Krajewski, K., Küspert, P., Schneider, W., & Visé, M. (2002). *Deutscher Mathematiktest für erste Klassen (DEMAT 1+)*. [German mathematical test for Grade 1]. Göttingen: Hogrefe.

Krajewski, K., Liehm, S., & Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+).* [German mathematical test for Grade 2]. Göttingen: Hogrefe.

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction,* 1*9*, 513-526.

Lenhard, W., & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler*. [A reading comprehension test for Grade 1 to 6]. Göttingen: Hogrefe.

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100*, 107-120.

Marcoulides, G. A., & Hershberger, S. L. (1997). *Multivariate statistical methods: A first course.* Mahwah, NJ: Erlbaum.

Masten, A. S., Coatsworth, J. D., Neemann, J., Gest, S. D., Tellegen, A., & Garmezy, N. (1995). The structure and coherence of competence from childhood through adolescence. *Child Development, 66,* 1635-1659.

McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology, 40*(4), 633-645.

Muthén, L. K. (1999, November 16). Multilevel data/complex sample: Intraclass correlations [4th post]. Retrieved February 3, 2016 from URL: http://www.statmodel.com/discussion/messages/12/18.html?1238555255

Näslund, J. C., & Schneider, W. (1996). Kindergarten letter knowledge, phonological skills, and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology*, *62*, 30-59.

Niklas, F. (2011). *Vorläuferfertigkeiten im Vorschulalter zur Vorhersage der Schulfähigkeit, späterer Rechenschwäche und Lese- und Rechtschreibschwäche. Diagnostik, Zusammenhänge und Entwicklung in Anbetracht der bevorstehenden Einschulung [*Precursors of school readiness and for the prediction of later dyscalculia and dyslexia. Diagnosis, interrelations and development considering imminent school enrolment*]*. Hamburg: Dr. Kovac.

Niklas, F., Cohrssen, C., & Tayler, C. (in press). Improving preschoolers' numerical abilities by enhancing the home numeracy environment. *Early Education and Development.*

Niklas, F., & Schneider, W. (2015). With a little help: improving kindergarten children's vocabulary by enhancing the home literacy environment. *Reading and Writing: An Interdisciplinary Journal, 28*(4), 491-508.

OECD (2014). *Education at a glance 2014: OECD Indicators*. OECD Publishing. doi: http://dx.doi.org/10.1787/eag-2014-en

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556.

Preßler, A.-L., Könen, T., Hasselhorn, M., & Krajewski, K. (2014). Cognitive preconditions of early reading and spelling: A latent-variable approach with longitudinal data. *Reading and Writing, 27,* 383-406.

Raven, J., Raven, J. C., & Court, J. H. (1999). *Priročnik za Ravenove progresivne matrice in besedne lestvice. 2. zvezek. Barvne progresivne matrice*. [Manual for Raven's progressive matrices and vocabulary scales. Section 2. Colored progressive matrices] Ljubljana: Center za psihodiagnostična sredstva.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Schleicher, A. (2013). *Beyond PISA 2015: A longer-term strategy of PISA. 36th meeting of the PISA Governing Board*. EDU/PISA/GB(2013)14. Retrieved October 3, 2015 from http://www.oecd.org/callsfortenders/ANNEX%20A.pdf

Schneider, W. (2009). The development of reading and spelling: Relevant precursors, developmental changes, and individual differences. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood: Findings from a 20 year longitudinal study* (pp. 199-220). New York: Psychology Press.

Schneider, W., Blanke, I., Faust, V., & Küspert, P. (2011). *Würzburger Leise Leseprobe-Revision (WLLP-R). Ein Gruppentest für die Grundschule* [Würzburger silent reading test – revision (WLLP-R). A group test for primary school]. Hogrefe: Göttingen.

Schneider, W., & Näslund, J. C. (1999). The impact of early phonological processing skills on reading and spelling in school: Evidence from the Munich Longitudinal Study. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 126–147). Cambridge, UK: Cambridge University Press.

Schneider, W., Niklas, F., & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learning and Individual Differences*, *32*, 156-162.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling.* London: Sage.

Stern, E. (2009). The development of mathematical competencies. Sources of individual differences and their developmental trajectories. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood: Findings from a 20 year longitudinal study* (pp. 221-237). New York: Psychology Press.

Stock, C., & Schneider, W. (2008). *Deutscher Rechtschreibtest für das erste und zweite Schuljahr (DERET 1-2+)* [German Spelling Test for Grade 1 and 2]. Göttingen: Hogrefe.

Torppa, M., Poikkeus, A.-M., Laakso, M.-L., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to Grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology*, *42*(6), 1128-1142.

Tymms, P. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading, 22*(1), 27-36.

Tymms, P. (2001). The development of a computer-adaptive assessment in the early years. *Educational and Child Psychology, 18*(3), 20-30.

Tymms, P., & Albone, S. (2002). Performance indicators in primary school. In A. J. Visscher & R. Coe (Eds.), *School improvement through performance feedback* (pp. 191-218). Lisse, Holland: Swets & Zeitlinger.

Tymms, P., Merrell, C., Hawker, D., & Nicholson, F. (2014). *Performance Indicators in Primary Schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions.* Durham, UK: University of Durham.

Tymms, P., Merrell, C., Henderson, B., Albone, S., & Jones, P. (2007). *Links between children's starting points and finishing points in primary school.* Paper presented at

12th Conference of European Association for Research on Learning and Instruction, Budapest, Hungary.

Tymms, P., Merrell, C., & Jones, P. (2004). Using baseline assessment data to make international comparisons. *British Educational Research Journal, 30*(5), 673-689.

Tymms, P., & Wylde, M. (2004). Basisprüfverfahren und Dauerbeobachtung in der Grundschule. [Basis test instrument and permanent monitoring in primary schools]. In G. Faust, M. Götz, H. Hacker, & H.-G. Roßbach (Eds.), *Anschlussfähige Bildungsprozesse im Elementar- und Primarbereich* (pp. 190-203). Bad Heilbrunn: Julius Klinkhardt.

Van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross cultural psychology* (Vol. 1, pp. 257-300). Boston: Allyn & Bacon.

Vidmar, M. (2010). *Dejavniki socialnega vedenja in učne uspešnosti v srednjem otroštvu* [The precursors of social behavior and academic achievement in middle childhood. Unpublished doctoral dissertation]. Neobjavljena doktorska disertacija, Univerza v Ljubljani, Filozofska fakulteta, Oddelek za psihologijo, Ljubljana.

Vidmar, M. (2011). Socialna kompetentnost in učna uspešnost v prvih razredih osnovne šole [Social competence and academic achievement in the first Grades]. *Šolsko polje* [School Field], *22*(1-2), 31-54.

Vidmar, M., & Tymms, P. (2009). *Performance indicators in primary schools: on-entry baseline assessment and follow-up material (PIPS-BA): introducing PIPS into Slovenian context - background, translation, adaptation, and application of the assessment*. Ljubljana: Pedagoški inštitut.

Vidmar, M. & Zupančič, M. (2006). *PIPS: Preizkus temeljnih kompetentnosti otrok ob vstopu v šolo – začetni preizkus in spremljanje* [PIPS: On-Entry Baseline Assessment and Follow-up. Unpublished]. Neobjavljeno gradivo, Pedagoški inštitut, Ljubljana.

Vidmar, M., & Zupančič, M. (2007a). *PIPS: Preizkus temeljnih kompetentnosti otrok ob vstopu v šolo – dodatek* [PIPS: Extension. Unpublished]. Neobjavljeno gradivo, Pedagoški inštitut, Ljubljana.

Vidmar, M., & Zupančič, M. (2007b). *Predloge za oceno doseženih standardov znanja v 2. razredu*. Neobjavljeno gradivo [Attainment of Performance Standards in Slovene, Math and Nature & Society: 2nd Grade. Unpublished]. Neobjavljeno gradivo, Pedagoški inštitut, Ljubljana.

Wegener, B. (1988). *Kritik des Prestiges* [Criticism of the prestige]. Opladen: Westdeutscher Verlag.

Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.

Wildy, H. & Styles, I. (2008). Measuring what students entering school know and can do: PIPS Australia 2006-2007. *Australian Journal of Early Childhood, 33*(4), 43-52

Wilkinson, J. E., Watt, J., Napuk, A., & Normand, B. (1998). Baseline assessment literature review and pre-school record keeping in Scotland. *Interchange, 55*, 4-19.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation, 12*(3), 1-26.

Zupančič, M., & Kavčič, T. (2007a) *Otroci od vrtca do šole: razvoj osebnosti in socialnega vedenja ter učna uspešnost prvošolcev* [Children from preschool to school: the development of personality and social behaviour, and academic achievement in the first Grade]. Ljubljana: Znanstvenoraziskovalni inštitut Filozofske fakultete.

Zupančič, M., & Kavčič, T. (2007b). Vzdolžna in sočasna napoved učnih dosežkov pri prvošolcih. [Longitudinal and concurrent prediction of academic achievement in the first Grade.] *Šolsko polje, 18*(5/6), 141-170.

Zupančič, M., & Kavčič, T. (2008). Dejavniki okolja, otrokovo socialno vedenje v vrtcu/šoli in učna uspešnost v prvem razredu. [Environmental factors, children's social behavior in (pre)school and academic achievement in the first Grade.] *Šolsko polje, 19*(1/2), 51-79.

Zupančič, M. (2006). *Predloge za oceno doseženih standardov znanja v 1. razredu.* [Neobjavljeno gradivo.] [Attainment of Performance Standards in Slovene, Math and Nature & Society: 1st Grade. Unpublished.] Ljubljana: Filozofska fakulteta, Oddelek za psihologijo.

Table 1. *Descriptives of the PIPS subtests and comparisons between the samples*

Area (latent variable) / **Subtest (indicator)**	**N of items**	**Slovene sample** *M*	*SD*	**German sample** *M*	*SD*	***t-test***	***Effect size***
Early literacy							
Writing	1	4.10	0.90	4.00	0.58	n.s.	.13
Ideas about reading	5	1.66	1.27	3.07	1.45	G>S	-1.04
Letter identification	26 (27)[a]	58.34	37.59	30.94	29.25	S>G	.80
Word recognition	6	0.98	1.31	1.28	1.13	G>S	-.25
Reading	69 (82)[a]	13.44	29.99	6.96	22.28	S>G	.24
Early numeracy							
Calculations with pictures	5	3.63	1.36	4.59	1.02	G>S	-.77
Number identification	14	6.84	3.84	8.86	2.84	G>S	-.58
Math	3	0.83	0.84	1.25	1.03	G>S	-.36

Notes. Number of items in parentheses denotes number of items in the German version (when the number of items differed for the two versions).

[a] *M* and *SD* are expressed as percentages of the maximum score (due to different maximum scores in the two samples).

Table 2. *Descriptives for academic achievement indicators*

Variable	Grade 1		Grade 2	
	M	*SD*	*M*	*SD*
Slovene sample				
Slovene language	2.91	0.68	2.97	0.70
Math	3.00	0.67	3.09	0.63
Nature & Society	3.11	0.63	3.10	0.60
German sample				
Spelling (DERET)[a]	10.27	6.91	17.11	9.92
Math (DEMAT)	24.10	7.42	19.89	8.70
Read (WLLP)	37.73	16.36	69.64	21.76
Read (ELFE)	/	/	42.07	16.56

[a] Reverse coding (low scores indicate fewer mistakes in the dictation)

Table 3. Correlations between PIPS and academic achievement indicators

			Early literacy					Early numeracy		
			Writing	IAR	Letters	Words	Reading	Calculations	Numbers	Math
SLO	G1	Slovene	.33**	.38**	.45**	.39**	.37**	.42**	.46**	.42**
		Math	.26**	.30**	.39**	.29**	.25**	.36**	.45**	.36**
		Nature & Society	.19**	.34**	.29**	.23**	.19**	.28**	.27**	.27**
	G2	Slovene	.37**	.37**	.52**	.49**	.38**	.45**	.50**	.39**
		Math	.30**	.32**	.36**	.34**	.23**	.43**	.47**	.32**
		Nature & Society	.29**	.36**	.38**	.35**	.19**	.38**	.39**	.32**
GER	G1	Spelling (DERET)	-.18**	-.12*	-.42**	-.35**	-.26**	-.20**	-.39**	-.24**
		Reading (WLLP)	.11*	.21**	.39**	.31**	.32**	.16**	.34**	.28**
		Math (DEMAT)	.20**	.13**	.31**	.28**	.23**	.32**	.40**	.39**
	G2	Spelling (DERET)	-.21**	-.06	-.38**	-.37**	-.23**	-.11*	-.33**	-.22**
		Reading (WLLP)	.09	.15**	.29**	.24**	.20**	.03	.26**	.25**
		Math (DEMAT)	.17**	.16**	.27**	.24**	.22**	.33**	.47**	.46**
		Reading (ELFE)	.12*	.16**	.41**	.30**	.25**	.10*	.36**	.26**

Notes. SLO = Slovene sample, GER = German sample, G = Grade, IAR = ideas about reading. *N* (Germany) = 416-426. *N* (Slovenia) = 315-321.

* $p < .05$; ** $p < .01$.

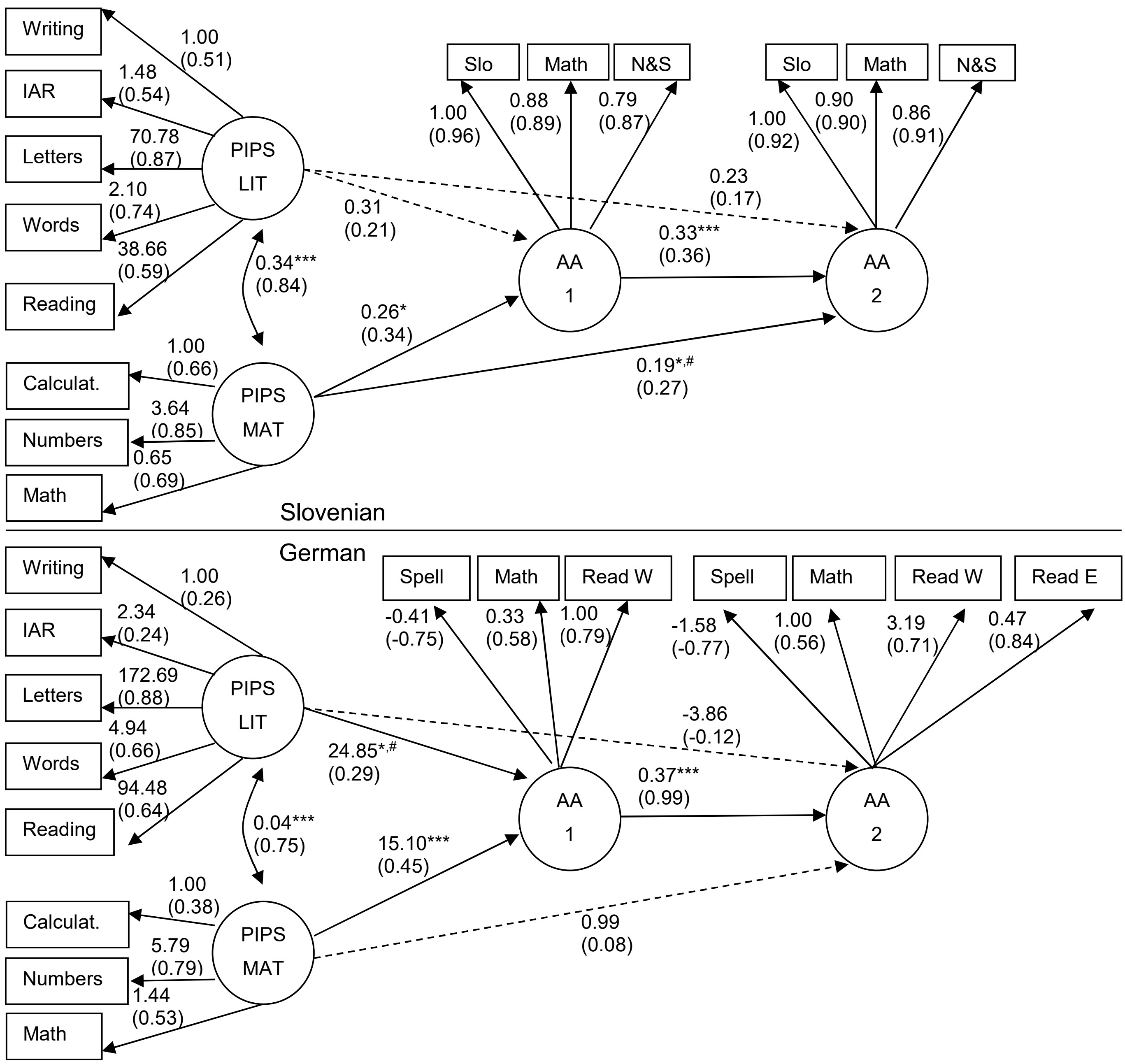

Figure 1. Structural equation model for the prediction of overall academic achievement. The numbers are unstandardized coefficient estimates (standardized coefficient estimates in parentheses). All indicator loadings were significant ($p < .001$). Solid lines represent significant paths or correlations; dashed lines indicate non-significant paths. # indicates paths that became non-significant after covariates were introduced into the model. IAR = ideas about reading, Slo = Slovenian language, N & S = Nature & Society, AA = academic achievement, LIT = literacy, NUM = numeracy, Read W = WLLP silent reading task, Read E = ELFE reading comprehension task.

*$p < .05$; ***$p < .001$

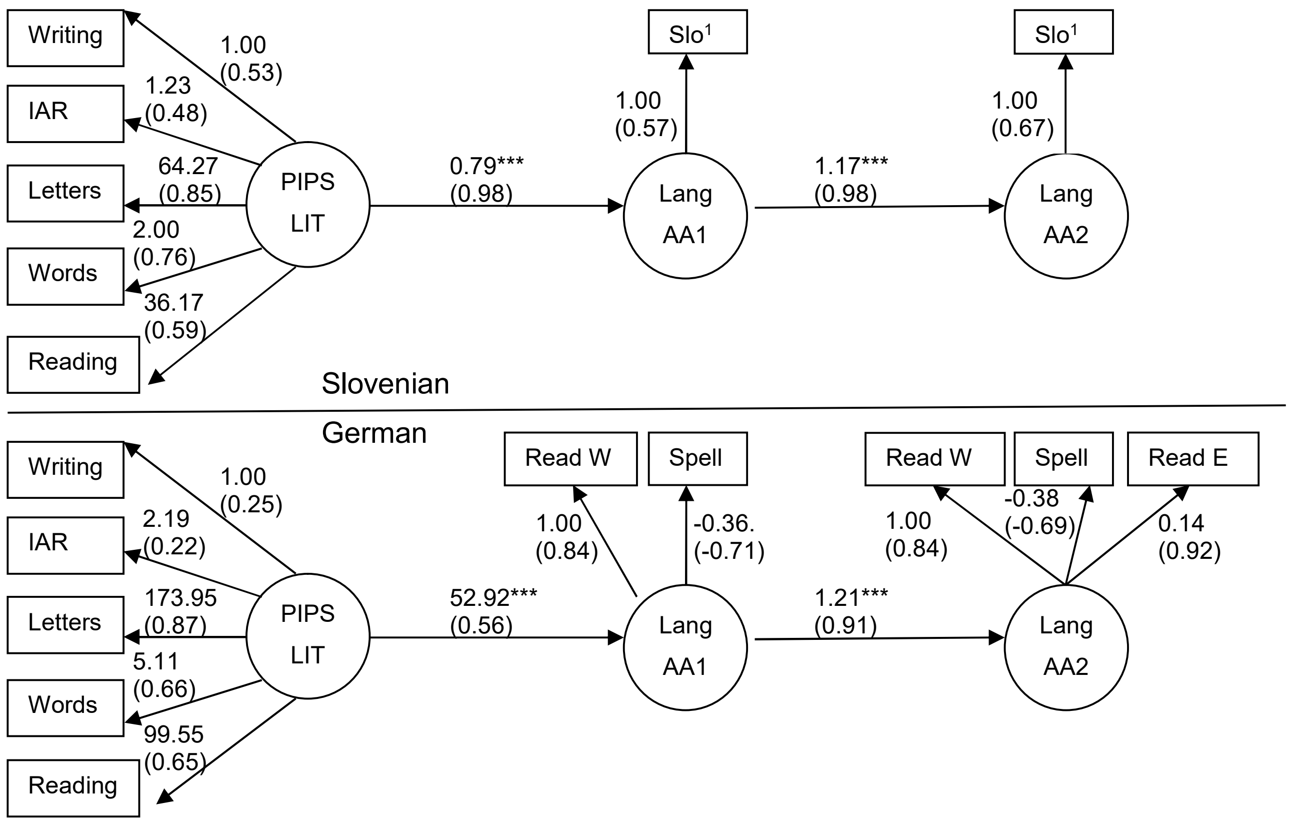

Figure 2. Structural equation model for the prediction of language academic achievement. The numbers are unstandardized coefficient estimates (standardized coefficient estimates in parentheses). All indicator loadings were significant (*p* < .001). Solid lines represent significant paths or correlations; dashed lines indicate non-significant paths. Lang = language (other abbreviations see Figure 1).

[1] The indicator is calculated as the mean of teacher's rating on several items (see Method).

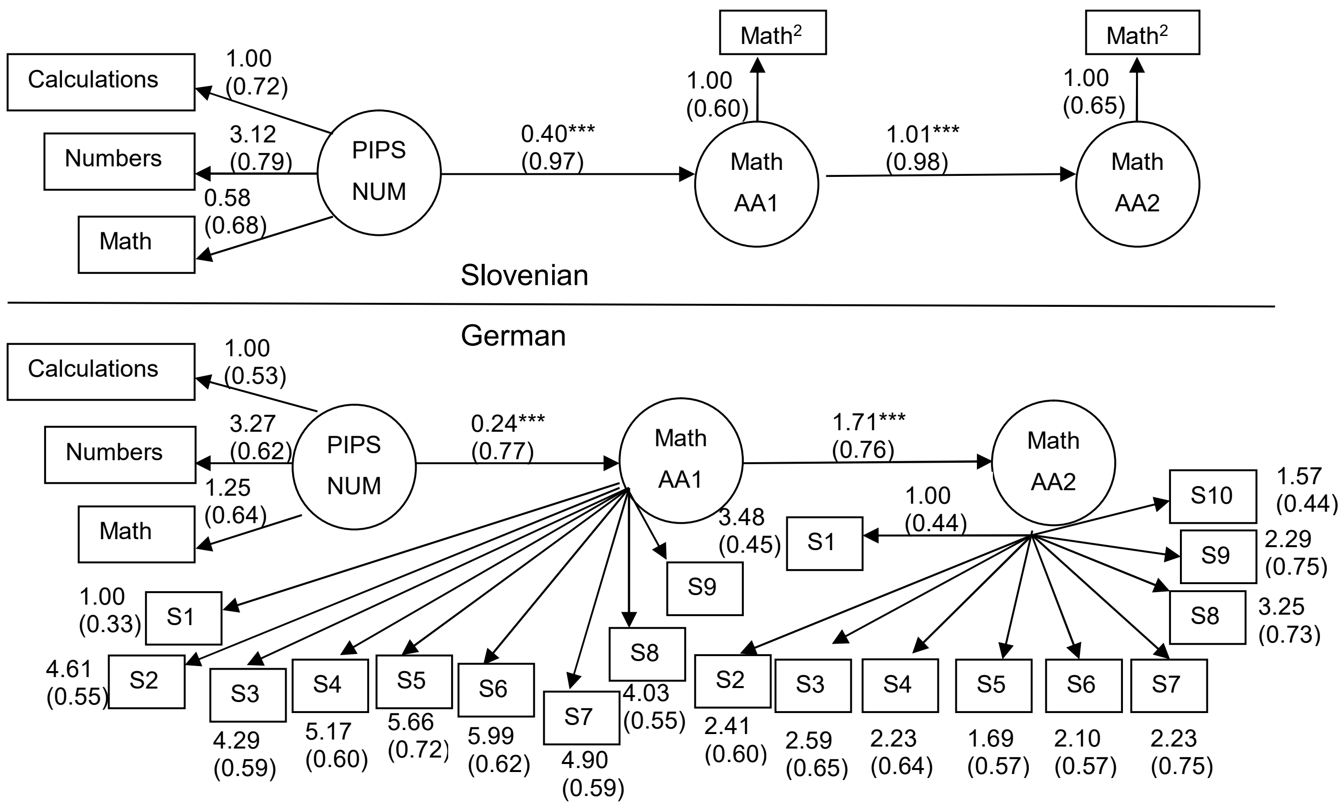

Figure 3. Structural equation model for the prediction of math academic achievement. The numbers are unstandardized coefficient estimates (standardized coefficient estimates in parentheses). All indicator loadings were significant (*p* < .001). Solid lines represent significant paths or correlations; dashed lines indicate non-significant paths. AA = academic achievement, NUM = numeracy, S = subtest.

[2] The indicator is calculated as the mean of teacher's rating on several items (see Method).