Gender Differences in the Development of Numerical Skills

in Four European Countries

Helga Krinzinger[1], Liane Kaufmann[2], Jacques Gregoire[3], Annemie Desoete[4], Hans-Christoph Nuerk[5] and Klaus Willmes[6]

[1] University Hospital of the RWTH Aachen University,Aachen, Germany, [2] Institute of Psychology, University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria, [3] Faculty of Psychology, Université Catholique de Louvain,Louvain-La-Neuve, Belgium, [4] Ghent University and University College Arteveldehogeschool, Ghent, Belgium, [5] Institute of Psychology, Eberhard Karls University, Tuebingen,Germany, [6] Section Neuropsychology, Neurological Clinic, University Hospital of the RWTH Aachen, Germany

ABSTRACT
Gender differences have been reported frequently for mathematical performance of boys and girls but usually not before adolescence. Therefore, developmental origins of a possible male advantage in numerical skills are understudied. In a large sample of second graders from four European countries, we found significant gender differences in favour of boys in writing multi-digit Arabic numerals from dictation, and in subtraction and multiplication. We discuss the latter findings with reference to a possible male advantage in visual-spatial cognition ability.

KEYWORDS
mathematics; cultural differences; elementary school

The Open University

Gender Differences in the Development of Numerical Skills

in Four European Countries

INTRODUCTION

Gender differences have been reported frequently in mathematical performance (e.g., Mills, Abland, & Stumpf, 1993; Robinson et al., 1996, Robinson et al., 1997; IEA, 1996; IEA, 2007) especially for mathematical problem-solving (Kimura, 2000). In modern society, mathematical ability has a stronger effect on employment rates and wages than does literacy (Dowker, 2005). Very little is known, however, about the developmental origins of gender differences in the domain of mathematics.

Cultural Mediation of Gender Differences in Mathematics

One of the first and most influential meta-analyses studying this topic found that gender differences in mathematical problem-solving emerge only at the high-school level (Hyde, Fennema, & Lamon, 1990). For example, in a study of regular elementary and secondary schools in Belgium, a great deal of overlap was found in the different mathematical performances of girls and boys (Desoete, 2007; 2008). Gender differences had a limited impact on mathematical problem-solving at the elementary school age ($N = 2255$). Boys were found to be better at several mathematical tasks in grades 2 and 3, whereas girls were at their best in grade 4. Boys, however, performed better in almost all mathematical tasks at the end of secondary school in Flanders ($N = 796$). Furthermore, when compared to older studies (Hyde, Fennema, & Lamon, 1990), the respective gender differences were larger in selected samples of highly precocious individuals and smaller in studies conducted after 1974. This finding suggests that socio-cultural factors underlie a male advantage in mathematical reasoning. Several studies have focused on possible mediators of culturally determined gender differences in mathematics and found that (1) lower parental expectations of girls' mathematics achievement influenced their attitudes towards mathematics (Eccles & Jacobs, 1986), (2) the coupling of self-belief about mathematical abilities, actual aptitude, and attitudes towards mathematics (Denissen, Zarrett, & Eccles, 2007) increases over time, (3) mathematics anxiety was already higher in girls than in boys in elementary school (Krinzinger et al., 2007), and (4) stereotype threats were indeed a significant predictor of individuals' mathematics performance scores (Osborne, 2001).

Evidence for cultural mediation of the gender gap in math comes from a recent study (Guiso et al., 2008) that compared the results of over 246,000 15-year-old boys and girls from over 40 countries who took part in the 2003 PISA (Program for International Student Assessment) study (OECD, 2003) and correlated these results with measures of gender equality, such as the World Economic Forum's Gender Gap Index (GGI) (Hausmann, Tyson, & Zahidi, 2006). The most important result of this study was that "the gender gap in math scores disappears in countries with a more gender-equal culture" (Guiso et al., 2008, p. 1164). A similar study by Else-Quest, Hyde, and Linn (2010) on data from both the 2003 PISA study and TIMSS 2003 came to the same conclusion. These findings denote a clear statement within the

nature-nurture debate about the male advantage in mathematical performance being due to socio-cultural factors.

BIOLOGICAL MEDIATION OF GENDER DIFFERENCES IN MATHEMATICS
Another line of research focuses more on possible biological reasons for the male advantage in mathematics performance. Most prominently, Geary (1996) argues that sexual selection during human phylogeny, and the associated mechanisms like sex hormones, results in a male advantage in spatial processing. A meta-analysis by Linn and Petersen (1985) found pronounced performance differences between males and females in spatial perception ($d = .44$; 81 studies) and mental rotation ($d = .94$; 18 studies). This male advantage may cause boys to be more interested in spatial objects and numbers than girls and more likely to use spatial solution strategies for mathematical problem-solving (Geary, 1996). This spatial cognition hypothesis was supported by several studies conducted by Casey and her co-workers. In a large sample of young adults (Casey, Nuttal, & Benbow, 1995), they observed that gender differences in the Scholastic Aptitude Test-Math (SAT-M) were mediated by mental rotation ability. In a sample of female college students, they also observed that mental rotation ability was accounted for by an interaction of familial handedness patterns as an indicator of biologically based individual differences and the choice of college major as indicator of environmental factors (Casey, 1996). Finally, in a large sample of 8th graders, they found that girls' lower mathematics scores were explained fully by their relatively poor spatial-mechanical skills (74% of total indirect effects) and by their lower mathematics self-confidence (26% of total indirect effects; Casey, Nuttall, & Pezaris, 2001). Rosselli et al. (2008) also found that, in children from age 7 onwards, spatial abilities were mediating gender effects for mental mathematical operations and solving arithmetic problems. Like Geary (1996), Casey et al. (2001) interpreted their own findings through different problem-solving styles of male and female subjects, with the latter being less likely to use spatial strategies that are more effective than algorithmic strategies for mathematical problem-solving (van Garderen, 2006). This hypothesis was confirmed by investigating boys' and girls' mathematical problem-solving strategies directly in a sample of high-school students (Gallagher et al., 2000) and by a longitudinal study in primary school children from 1st to 3rd grade (Fennema et al., 1998). Furthermore, differential relationships between spatial numerical representations and calculation skills in male and female third graders have been reported (Lonnemann et al., 2008).

To conclude, the spatial cognition hypothesis can be seen as a good example of the psychobiosocial model proposed by Halpern (1997): biological predispositions, individual interests, and social environmental factors interact during a cognitive learning process and result in gender differences in mathematics.

ALTERNATIVE EXPLANATIONS FOR GENDER DIFFERENCES IN MATHEMATICS
Hedges and Nowell (1995) provided an alternative explanation for the overrepresentation of male subjects in the higher performance ranges of mathematical achievement. In an analysis of six studies from the USA about different cognitive tasks using national probability samples, they observed that

boys' test scores have generally larger variance and that boys typically outnumber girls among the high-scoring groups. An analysis of the PISA 2003 mathematics assessment in the United States (OECD, 2003) by Liu, Wilson, and Paek (2008) yielded a different result. In a sample of over 5,000 15-year-old students, gender differences were generally small, but performance distributions of boys and girls differed according to the four mathematical domains tested. The domain "Space and Shape" (e.g., mental rotation of objects) displayed the largest gender gap (effect size $d = .14$), with more girls in the lower performance ranges and more boys in the higher performance ranges. In the domain "Change and Relationship" (e.g., translation between representation formats, understanding of fundamental relationships and types of change), girls were overrepresented in the middle performance range and underrepresented in the higher performance range ($d = .10$). No significant gender differences were found for either the mean performance ($d = .04$) or the performance distribution in the domain "Quantity" (mostly computational items). The domain "Uncertainty" (e.g., calculation of probabilities) showed the second-highest overall gender difference ($d = .12$), resulting from more girls belonging to relatively low performing groups and more boys belonging to the middle performing groups.

Therefore, gender differences do not arise solely from higher performance variability among boys and are not found in all numerical domains. This finding of differential gender differences in various mathematical tasks indicates that using several tasks may be more promising than global math performance measures in revealing possible causes for respective gender differences. For example, it may be possible that culturally or biologically mediated factors influence performance on two mathematical tasks differentially for boys and girls and that these effects will be intermingled or even neutralized for global performance measures.

Royer et al. (1999) suggested a completely different explanation for the better scores of male college students on standardized math achievement tests. They found a male advantage in math fact retrieval and concluded that higher computational fluency should leave more time and cognitive resources for solving more complex tasks in math tests, especially under time pressure. By testing both the spatial cognition hypothesis and the math fact retrieval hypothesis directly in a large sample of college students, Geary et al. (2000) found that both spatial cognition and computational fluency were mediating the male advantage in arithmetical reasoning.

OBJECTIVES OF THIS STUDY
Despite these ongoing debates about gender differences in mathematical abilities and the scientific efforts to unveil their possible causes, surprisingly little is known about their developmental origins. Only very rarely have gender differences in math performance been reported for primary school children (e.g., Fennema et al., 1998; Rosselli et al., 2008).

In our study, we wanted to uncover (1) whether gender differences in numerical tasks can be found even at young ages, and if so, (2) whether they are comparable for different countries.

METHODS
Participants
Overall, we tested 220 children in the middle of 2nd grade from Germany, Austria, Belgium, and France on four different numerical tasks. The sample was part of the standardization samples of the dyscalculia test TEDI-MATH (Van Nieuwenhoven, Grégoire, & Noël, 2001). The TEDI-MATH is a multi-componential dyscalculia test based on cognitive neuropsychological models of number processing and calculation and has been tested for conceptual accuracy and clinical relevance in previous studies (e.g., Stock, Desoete, & Roeyers, 2009).

In general, recruitment of the standardization samples took place in typical primary schools. Austrian children were, apart from a small subsample from Vienna, recruited from the rural area around Innsbruck and came from mainly middle-class socio-economic backgrounds. In Germany and Flanders (Flemish-speaking part of Belgium), the participating schools were spread across the whole country, and children came from mixed socio-economic backgrounds. In those samples, all children were asked to participate to represent the whole typical performance range. Children from Wallonia (French-speaking region of Belgium) and France were randomly selected from all over their respective country, but children who were one or more years behind the regular grade corresponding to their age, exhibited intellectual or sensorial disabilities, or who displayed behavioral troubles were excluded. The mean age of the 220 children was 91 months (standard deviation $SD = 5$), ranging from 83 to 113 months. Sample characteristics of the five different groups can be obtained from Table 1.

Table 1:
Sample characteristics: Number of children and age, in months, for each country

	Germany	Austria	Belgium	France	Mean age (*SD*)
Boys	30	16	29	24	92 (6)
Girls	36	22	38	25	91 (5)
Total	66	38	67	49	91 (5)
Mean age (*SD)*	94 (4)	95 (5)	89 (5)	89 (2)	

SD: standard deviation

French and Belgian children did not differ in their mean age (all $p > .05$, Bonferroni corrected) but were significantly younger than the German and Austrian children (all $p < .05$, Bonferroni corrected). Linear regression analyses for each of the four experimental tasks revealed that chronological age was not a significant predictor of performance on any task (all adjusted $R^2 < .01$, all $p > .12$).

GENERAL PROCEDURE AND TASKS
In a quiet, separate room in each of their schools, all children were assessed individually with the TEDI-MATH (Van Nieuwenhoven et al., 2001) by a trained tester. Written and informed consent was obtained from all parents or caregivers.

The multi-componential dyscalculia test TEDI-MATH demonstrated good psychometric properties (Kaufmann et al., 2009). For this study, we concentrated only on subtests with at least 10 items in the original French version (most subtests have more items in the Flemish and the German versions). It is important to note that the TEDI-MATH is not a curriculum-based test, but builds on neuro-cognitive models of number processing and calculation (most importantly, the Triple-Code model by Dehaene, 1992). Therefore, items of the subtests may be ahead of the respective national mathematics curricula. Of the six subtests that fulfilled this criterion for the respective age group, including Base-10 representation using sticks, Understanding units and tens, Writing Arabic numerals from dictation, Reading Arabic numerals, Subtraction, and Multiplication, we did not analyze Reading Arabic numerals because of severe ceiling effects. We also did not analyze Base-10 representation with sticks because this task requires strong verbal skills, as children have to explain their decisions for each item. In the Writing Arabic numerals from dictation-task, children were asked to write down 20 different Arabic numerals (3 single-digit, 9 double-digit, 8 three-digit numbers) that were presented verbally one by one by the tester. For the task Understanding units and tens, children were presented with Arabic multi-digit numbers on a sheet of paper and asked to point at the unit-digit or the decade-digit, respectively (3 double-digit and 2 three-digit numbers each). Subtraction and Multiplication were both administered by showing the children one calculation problem after another on a sheet of paper, asking them to solve them mentally as fast and accurately as possible and to give the correct answer verbally.

All items were scored with 1 for a correct and 0 for an incorrect answer. All tasks were terminated after 5 consecutive errors. In this case, all other subsequent items were scored as incorrect. The specific items of all four tasks can be obtained from Table 2.

ANALYSIS
We conducted several 2x4 analyses of variance (ANOVA) (Kirk, 1995) with the factors of gender (male and female) and country (Germany, Austria, Belgium and France) on the number of correct items for each of the four tasks. The assumptions underlying an ANOVA are independence of cases (guaranteed by individual testing in our study), normally distributed data, and homogeneity of variances. Homogeneity of variances could be assumed – using Levene's test - for the tasks Writing Arabic numerals from dictation and Subtraction (both p > .20) , but not for the tasks Understanding units and tens and Multiplication (both p < .001), and data was generally not normally distributed (all p < .016). However, as reported by Glass, Peckham, and Sanders (1972), results of ANOVAs are very robust against violations of normal distribution and variance homogeneity if sample sizes are large enough (n > 10), which is the case in our study.

The (in our case: two-factorial) ANOVA simultaneously compares mean performance rates for different groups (e.g., male/female and different countries) and tests for a possible interaction between the factors. In our case, this would mean the following: If the factor of gender was significant (indicated by a high *F*-value and a significance level *p* < .05), then the mean performance rates between

boys and girls should not be assumed to be equal. If the factor of country was significant, then the mean performance rates between specific countries should not be assumed to be the same. If the interaction between the two factors was significant, then gender differences should not be assumed to be equal for specific countries.

Item Nr.	Writing Arabic numerals from dictation	Understanding units and tens	Subtraction	Multiplication
1	4	28 [a]	4-2	1x7
2	7	13 [a]	9-5	6x1
3	1	10 [a]	5-3	2x4
4	11	520 [a]	6-6	3x3
5	40	709 [a]	4-0	8x0
6	16	20 [b]	16-4	3x5
7	30	15 [b]	27-6	4x4
8	73	37 [b]	40-20	10x2
9	13	650 [b]	36-10	0x3
10	68	405 [b]	44-26	3x10
11	80			
12	25			
13	200			
14	109			
15	150			
16	101			
17	700			
18	643			
19	190			
20	951			

[a] Recognition of unit-digit required; [b] Recognition of decade-digit required.

Table 2: Items of the four tasks Writing Arabic numerals from dictation, Understanding units and tens, Subtraction, and Multiplication

RESULTS
Writing Arabic numerals from dictation
The mean number of correct items achieved by each group (Germany, Austria, Belgium, and France), including the separation of girls and boys, in the task Writing Arabic numerals from dictation can be found in Figure 1 (all boxplots were created using SPSS Statistics 20).
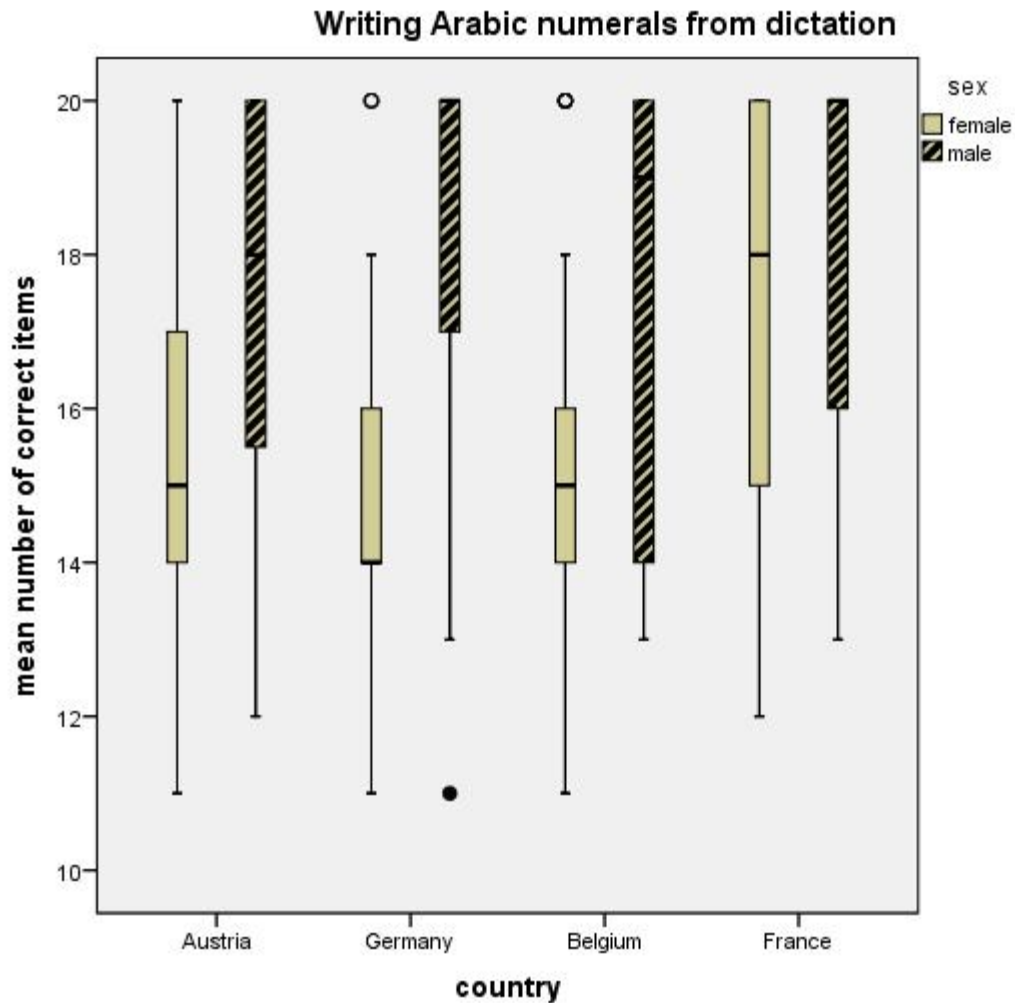
Figure 1: Mean number of correct items in Writing Arabic numerals from dictation by country (Austria, Germany, Belgium, and France), separated by boys and girls depicted by boxplots.

A 2 (gender) x 4 (country) ANOVA revealed both a significant main effect of gender in favor of boys [$F(1, 219)$[1] $= 34.97$, $p < .001$] and a significant effect of country [$F(3, 219) = 3.69$, $p = .013$], but no significant interaction [$F(3, 219) = 1.74$, $p = .16$] for Writing Arabic numerals from dictation.

The mean number of correct items achieved by each group (Germany, Austria, Belgium, and France), including the separation of girls and boys, in the task Understanding units and tens can be found in Figure 2.

---

[1] The numbers in the brackets given after the F-value indicate degrees of freedom for this analysis.
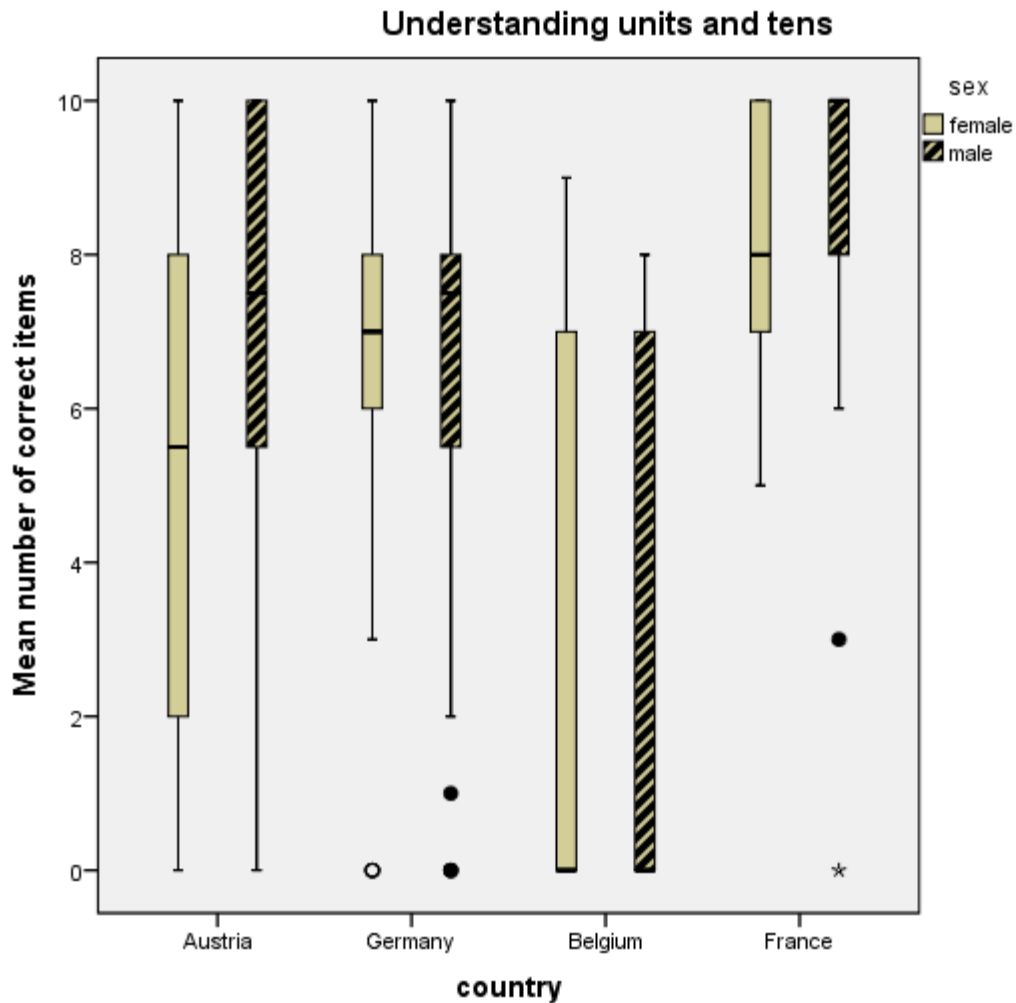
Figure 2: Mean number of correct items in units and tens by country (Austria, Germany, Belgium, and France), separated by boys and girls depicted by boxplots.

A 2 (gender) x 4 (country) ANOVA revealed no significant main effect of gender [$F(1, 219) = 2.17$, $p = .140$], a significant effect of country [$F(3, 219) = 30.57$, $p < .001$], and no significant interaction [$F(3, 219) = 0.48$, $p = .70$] for Understanding units and tens.

The mean number of correct items achieved by each group (Germany, Austria, Belgium, and France), including the separation of girls and boys, in the task Subtraction can be found in Figure 3.
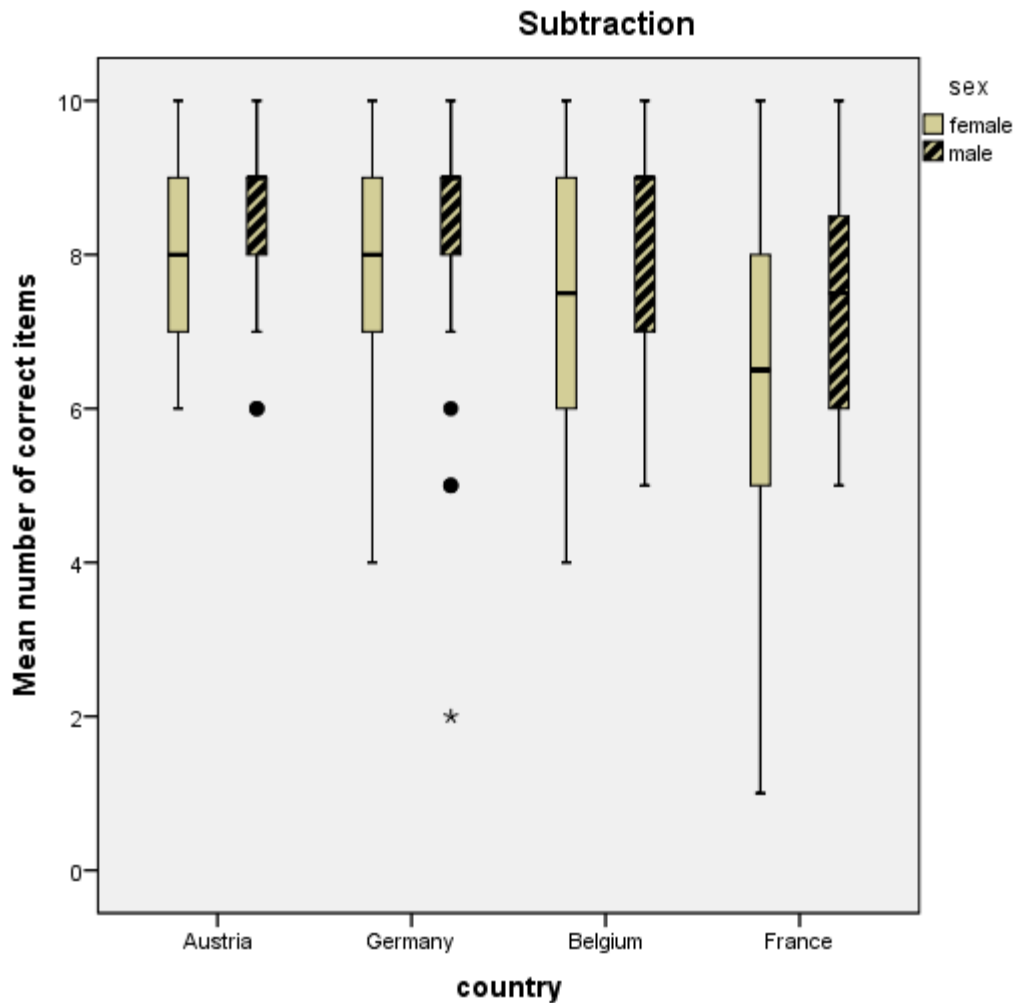
Figure 3: Mean number of correct items in Subtraction by country (Austria, Germany, Belgium, and France), separated by boys and girls depicted by boxplots.

A 2 (gender) x 4 (country) ANOVA revealed a significant main effect of gender in favor of boys [$F(1, 219) = 10.55$, $p = .001$], and a significant effect of country [$F(3, 219) = 7.36$, $p < .001$], but no significant interaction [$F(3, 219) = 0.54$, $p = .659$] for Subtraction.

The mean number of correct items achieved by each group (Germany, Austria, Belgium, and France), including the separation of girls and boys, in the task Multiplication can be found in Figure 4.
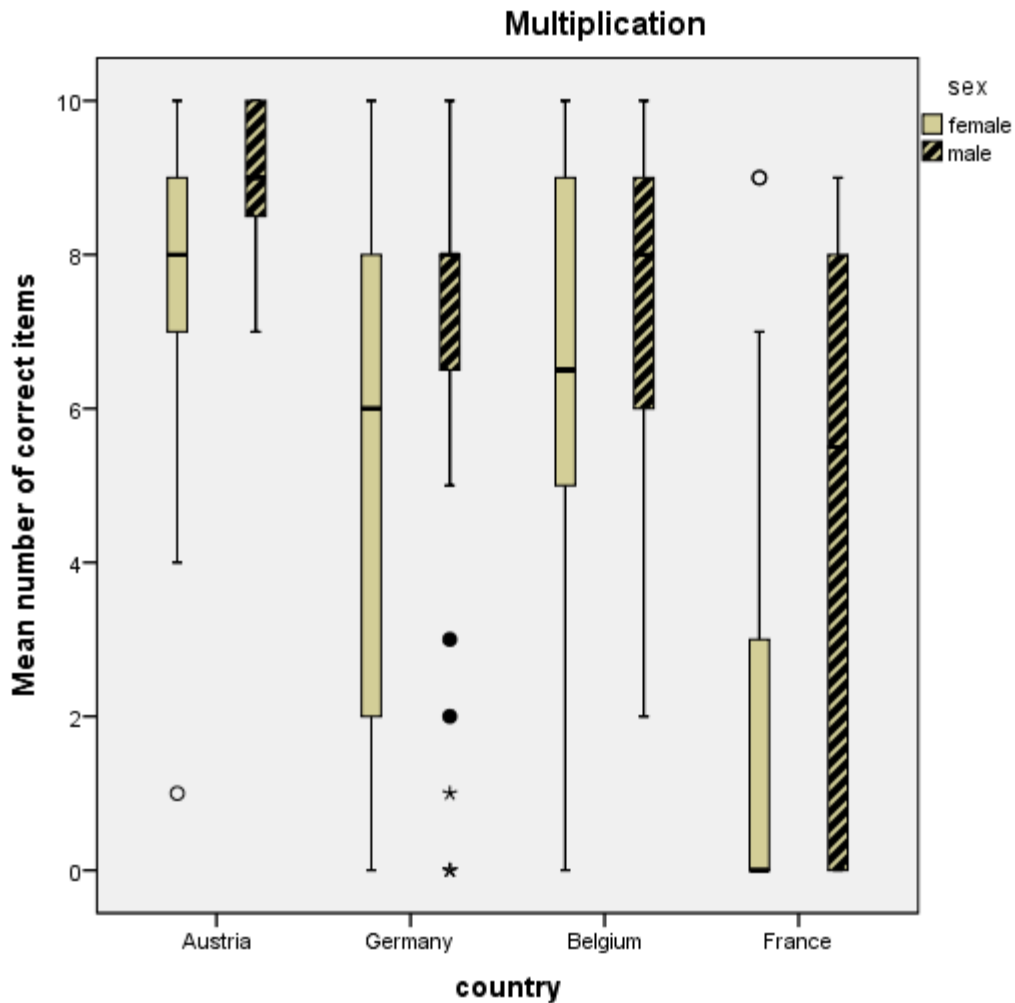
Figure 4: Mean number of correct items in Multiplication by country (Austria, Germany, Belgium, and France), separated by boys and girls depicted by boxplots.

A 2 (gender) x 4 (country) ANOVA revealed a significant main effect of gender in favor of boys [$F(1, 219) = 17.99$, $p < .001$] and a significant effect of country [$F(3, 219) = 27.38$, $p < .001$] but no significant interaction [$F(3, 219) = 0.68$, $p = .567$] for Multiplication.

DISCUSSION
The two main goals of this study were (1) to find out whether gender differences in favor of boys would be found in various mathematical tasks in a large sample of second graders, and if so, (2) to examine whether these gender differences would be equal across different countries.

Returning to our first research question, it was not clear whether we could find gender differences in mathematics in such a young sample, as most studies report

that a male advantage in mathematical tasks emerges only in, or after, adolescence, and only for specific tasks, such as mathematical problem-solving or geometry (for a meta-analysis see Hyde et al., 1990; but see also Fennema et al., 1998; Rosselli et al., 2008).

In our sample of young, primary school children, the observation of a significant effect of gender in favor of boys on three out of four tasks (Writing Arabic numerals from dictation, Subtraction, and Multiplication) was not necessarily anticipated and thus constitutes an interesting finding.

In this context, one must note that, in all three tasks with significant gender differences, the most difficult items were ahead of the respective national mathematics curricula and that children were not expected to solve them correctly. In other words, the performance differences between boys and girls were attributed to skills they had not been explicitly taught. This is in agreement with the frequent finding of a male advantage in standardized mathematical problem-solving tasks, but not for school achievement tests in mathematics (Kimura, 2000). This observation, however, does not solve the question of why boys are better in some mathematical tasks even at such a young age.

With respect to our second research question, we could not find any differences for the respective male advantages in task performance between our samples from four different European countries. This finding suggests that gender differences in numerical tasks in primary school seem to be a robust cultural finding.

With our study design, we could not infer whether the origin of the male advantage in numerical tasks in our sample was socio-cultural. Other (more cognitive) perspectives, sometimes with an emphasis on biological causes (for an overview, see Geary, 1996), provide alternative explanations of why male subjects outperform females on at least some mathematical skills. An influential line of research claims that men and boys have better visual-spatial abilities or rely more strongly on spatial problem-solving strategies compared to women and girls which provides them with an advantage for mathematical problem-solving (Casey et al., 1995, 2001; Fennema et al., 1998; Rosselli et al., 2008) or numerical representations (Lonnemann et al., 2008).

In general, it is not easy to explain how better visual-spatial abilities could help children in learning how to write multi-digit Arabic numerals. Children have to acquire a basic understanding of the regularity of the Arabic notational system of numbers for successful multi-digit number processing. As this regularity has inherent spatial features (as the term "place-value" already indicates), it is plausible that a preference for spatial thinking strategies or better visual-spatial abilities should provide individuals with an advantage for learning the place-value system of Arabic numerals.

This interpretation is supported by the finding that first graders' visual-spatial working memory capacity influenced significantly the number of one type of syntactic transcoding errors they made, specifically additive decomposition errors

(e.g., writing '360' as '30060'), whereas phonological working memory capacity was neither related to overall error rates nor to any specific error type (Zuber et al., 2009).

Concerning the observed gender differences in Multiplication, spatial problem-solving strategies might play a supporting role when children have very little or no formal experience with this new type of arithmetic operation and have not yet memorized the multiplication tables, which is the case for our sample. We may speculate that this initial male advantage in solving multiplication problems could even be the reason why better fact retrieval abilities have been reported for adult male subjects (Royer et al., 1999). Finally, because even though the multiplication task is influenced by place-value information (consistency effect: see Domahs, Delazer, & Nuerk, 2006; Domahs et al., 2007; Verguts & Fias, 2005), a better place-value understanding may also aid performance in the multiplication tasks.

To summarize, the interesting gender differences in our sample could potentially be explained by boys' stronger place-value understanding, which may, as the literature suggests, be based on their stronger visual-spatial abilities. Therefore, we propose that it is a worthwhile endeavor to examine which specific spatial abilities influence which specific numerical representations (e.g., place-value understanding).

CONCLUSION

In a large sample of second graders from four European countries, we found significant gender differences in favor of boys in writing multi-digit Arabic numerals from dictation, and in subtraction and multiplication. The advantages that boys have for some numerical tasks did not differ among the four European countries.

The gender differences might be a result of a male advantage in visual-spatial abilities or the boys' preference for using spatial thinking strategies. Future studies should investigate the role of spatial cognition on mathematics learned at an early age because a better understanding of any factor influencing the development of numerical cognition may help improve mathematics education in primary schools.

REFERENCES

Casey, M. B. (1996) 'Understanding individual differences in spatial ability within females: A nature/nurture interactionist framework', *Developmental Review,* vol. 16, pp. 241-260.

Casey, M. B., Nuttall, R., & Benbow, C. P. (1995) 'The influence of spatial ability on gender differences in mathematics college entrance test-scores across diverse samples', *Developmental Psychology,* vol. 31, pp. 697-705.

Casey, M.B., Nuttall, R.N., & Pezaris, E. (2001) 'Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items', *Journal for Research in Mathematics Education,* vol. 32, pp. 28-57.

Dehaene, S. (1992). 'Varieties of numerical abilities'. *Cognition,* vol. 44, pp. 1-42.

Denissen, J.J.A., Zarrett, N.R., & Eccles, J.S. (2007) 'I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest', *Child Development,* vol. 78, pp. 430-447.

Desoete, A. (2007) 'Students with mathematical disabilities in Belgium: from definition, classification and assessment to STICORDI devices' in Scruggs, T.E. & Mastropieri, M.A. (eds) *Advances in Learning and Behavioral Disabilities*, Amsterdam & Oxford, Elsevier Press.

Desoete, A. (2008) 'Do birth order, family size and gender affect arithmetic achievement in elementary school?', *Electronic Journal of Research in Educational Psychology,* vol. 6, no. 1, pp. 135-156.

Domahs, F., Delazer, M., & Nuerk, H.-C., (2006) 'What makes multiplication facts difficult – Problem size or neighbourhood consistency?', *Experimental Psychology,* vol. 53, pp. 275-282.

Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K. & Nuerk, H.-C. (2007) 'Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence', *Behavioral and Brain Functions,* vol. 3, no. 66, doi:10.1186/1744-9081-3-66.

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

Eccles, J.S. & Jacobs, J.E. (1986) 'Social forces shape math attitudes and performance', *Journal of Women in Culture and Society,* vol. 11, pp. 367-380.

Else-Quest, N.M., Hyde, J.S., & Linn, M.C. (2010) 'Cross-national patterns of gender differences in mathematics: a meta-analysis', *Psychological Bulletin,* vol. 136, pp. 103-127.

Fennema, E., Carpenter, T.P., Jakobs, V.R., Franke, M.L., & Lewi, L.W. (1998) 'A longitudinal study of gender differences in young children's mathematical thinking', *Educational Researcher*, vol. 27, pp. 16-31.

Gallagher, A.M., De Lisi, R., Holst, P.C., Gillicuddy-De Lisi, A.V., Morely, M., & Cahalan, C. (2000) 'Gender differences in advanced mathematical problem solving', *Journal of Experimental Child Psychology*, vol. 75, pp. 165-190.

Geary, D.C. (1996) 'Sexual selection and sex differences in mathematical abilities', *Behavioral and Brain Sciences*, vol. 19, pp. 229-284.

Geary, D.C., Saults, S.J., Liu, F., & Hoard, M.K. (2000). 'Sex differences in spatial cognition, computational fluency, and arithmetical reasoning'. *Journal of Experimental Child Psychology*, 77, 337-353.

Geary, D.C., Bow Thomas, C.C., Liu, F., & Siegler, R.S. (1996) 'Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling', *Child Development*, vol. 67, pp. 2022-2044.

Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). 'Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance.' *Review of Educational Research,* vol. 42, pp. 237-288.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008) 'Culture, gender, and math', *Science*, vol. 320, pp. 1164-1165.

Halpern, D.F. (1997) 'Sex differences in intelligence. Implications for education', *American Psychologist*, vol. 52, pp. 1091-1102.

Hausmann, R., Tyson, L.D., & Zahidi, S. (2006) *The Global Gender Gap Report 2006*, World Economic Forum.

Hedges, L.V. & Nowell A. (1995) 'Sex differences in mental test scores, variability, and numbers of high-scoring individuals', *Science*, vol. 269, pp. 41-45.

Hyde, J.S., Fennema, E., & Lamon, S.J. (1990) 'Gender differences in mathematics performance - a metaanalysis', *Psychological Bulletin*, vol. 107, pp. 139-155.

IEA (1996) *Third International Mathematics and Science Study (TIMSS III),* Chestnut Hill, TIMSS International Study Center Boston College.

IEA (2007) *Third International Mathematics and Science Study (TIMSS 2007),* Chestnut Hill, TIMSS International Study Center Boston College.

Kaufmann, L., Nuerk, H.-C., Graf, M., Krinzinger, H., Delazer, M., & Willmes, K. (2009) *TEDI-MATH - Test zur Erfassung numerisch-rechnerischer Fertigkeiten vom Kindergarten bis zur 3. Klasse,* Bern, Verlag Hans Huber, Hogrefe AG.

Kimura, D. (2000) *Sex and Cognition*. Cambridge, MA, MIT Press.

Kirk, R.E. (1995). *Experimental Design: Procedures For The Behavioral Sciences* (3 ed.). Pacific Grove, CA, USA: Brooks/Cole.

Krinzinger, H., Kaufmann, L., Dowker, A., Thomas, G., Graf, M., Nuerk, H.-C., & Willmes, K. (2007) 'Deutschsprachige Version des Fragebogens für Rechenangst (FRA) für 6- bis 9-jährige Kinder', *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, vol. 35, pp. 341–351.

Linn, M.C. & Petersen, A.C. (1985) 'Emergence and characterization of sex differences in spatial ability: A meta- analysis', *Child Development*, vol. 56, pp. 1479-98.

Liu, O.L., Wilson, M., & Paek, I. (2008) 'A multidimensional Rasch analysis of gender differences in PISA mathematics', *Journal of Applied Measurement*, vol. 1, pp. 18-35.

Lonnemann, J., Krinzinger, H., Knops, A., & Willmes, K. (2008) 'Spatial representations of numbers in children and their connection with calculation abilities', *Cortex*, vol. 44, pp. 420-428.

Mills, C.J., Abland, K.E., & Stumpf, H. (1993) 'Gender differences in academically talented young students' mathematical reasoning: Patterns across age and subskills', *Journal of Educational Psychology*, vol. 85, pp. 340-346.

OECD (2003) *Programme for International Student Assessment (PISA), 2nd Assessment,* Paris, France, OECD.

Osborne, J.W. (2001) 'Testing stereotype threat: Does anxiety explain race and sex differences in achievement?', *Contemporary Educational Psychology*, vol. 26, *pp.* 291-310.

Robinson, N.M., Abbott, R.D., Berninger, V.W., & Busse, J. (1996) 'The structure of abilities in math-precocious young children: Gender similarities and differences', *Journal of Educational Psychology*, vol. 88, pp. 341-352.

Robinson, N.M., Abbott, R.D., Berninger, V.W., Busse, J., & Mukhopadhyay, S. (1997) 'Developmental changes in mathematically precocious young children: Longitudinal and gender effects', *Gifted Child Quarterly*, vol. 41, pp. 145-158.

Rosselli, M., Ardila, A., Matute, E., & Inozemtseva, O. (2008) 'Gender differences and cognitive correlates of mathematical skills in school-aged children' *Child Neuropsychology*, vol. 15, pp. 216-231.

Royer, J.M., Tronsky, L.N., Chan, Y., Jackson, S.J., & Marchant, H. (1999) 'Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance', *Contemporary Educational Psychology*, vol. 24, pp. 181-266.

Stock, P., Desoete, A., & Roeyers, H. (2009) 'Predicting arithmetic abilities: The role of preparatory arithmetic markers and intelligence', *Journal of Psychoeducational Assessment*, vol. 27, pp. 237-251.

van Garderen, D. (2006) 'Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities', *Journal of Learning Disabilities*, vol. 39, pp. 496-506.

Van Nieuwenhoven, C., Grégoire, J., & Noël, M.P. (2001) *TEDI-MATH. Test Diagnostique des Compétences de Base en Mathématiques*, Paris, ECPA.

Verguts, T. & Fias, W. (2005) 'Neighbourhood effects in mental arithmetic', *Psychology Science,* vol. 47, pp. 132-140.

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009) 'On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity', *Journal of Experimental Child Psychology*, vol. 102, pp. 60-77.