2015

A latent profile analysis of math achievement, numerosity, and math anxiety in twins

Sara Hart, Jessica A. R. Logan, Lee A. Thompson, Yulia Kovas, Grainne McLoughlin, and Stephen Petrill

A latent profile analysis of math achievement, numerosity, and math anxiety in twins

S.A. Hart*

The Florida State University and Florida Center for Reading Research

J.A.R. Logan

The Ohio State University and Crane Center for Early Childhood Research and Policy

Lee Thompson

Case Western Reserve University

Yulia Kovas

Tomsk State University

Social, Genetic, and Developmental Psychiatry Research Centre, Institute of Psychiatry

Goldsmiths, University of London

Gráinne McLoughlin

Institute for Neural Computation, University of California San Diego

Social, Genetic, and Developmental Psychiatry Research Centre, Institute of Psychiatry

S.A. Petrill

The Ohio State University

*Corresponding Author: hart@psy.fsu.edu

Abstract

Underperformance in math is a problem with increasing prevalence, complex etiology, and severe repercussions. This study examined the etiological heterogeneity of math performance in a sample of 264 pairs of 12-year-old twins assessed on measures of math achievement, numerosity and math anxiety. Latent profile analysis indicated five groupings of individuals representing different patterns of math achievement, numerosity and math anxiety, coupled with differing degrees of familial transmission. These results suggest that there may be distinct profiles of math achievement, numerosity and anxiety; particularly for students who struggle in math.

A significant number of school-age children and adolescents underperform in math, which in turn limits available educational opportunities and ultimately occupational success (Geary, Hoard, Nugent & Bailey, 2012). In response, the literature suggests two categories of causal factors. The first includes information-processing structures and/or processes associated with math performance (Feigenson, Libertus & Halberda, 2013; Geary et al., 2012; Jordan, Glutting & Ramineni, 2010; Mazzocco, Feigenson & Halberda, 2011; Sasanguie, De Smedt, Defever & Reynvoet, 2012). From extant work in the area of information-processing, numerosity has emerged as a particularly interesting influence on math achievement. The second category involves the affective components of math performance (Fennema, 1989; Maloney & Beilock, 2012; McLoed, 1992). In particular, the construct of mathematical anxiety has a long history (e.g., Dreger & Aiken, 1957), and continues to draw considerable attention as an important influence on math performance (e.g., Maloney & Beilock, 2012). Bringing these two categories of casual factors together in the same study may advance our understanding of how and why children differ in their achievement outcomes, allowing for a better understanding of success and failure.

Halberda, Mazzocco and Feignenson (2008) were among the first to link numerosity and math achievement. In brief, numerosity is defined as an innate set of skills representing, but not limited to, the non-symbolic number approximation system which estimates large magnitudes (e.g., Halberda et al., 2008), and also the symbolic number approximation system which maps numerical symbols onto magnitudes (Geary et al., 2012; see also Butterworth, 2010; Sasanguie et al., 2012). It has been proposed that the non-symbolic number approximation system is an ancient domain specific skill that can be seen across human and non-human primates and requires no cultural input to develop (Verguts & Fias, 2004). A review of this work suggests that non-symbolic number approximation is often related to math achievement (De Smedt, Noël, Gilmore & Ansari, 2013), and experimental studies have indicated that training non-symbolic number approximation leads to better performance on math achievement measures (Hyde, Khanum & Spelke, 2014; Park & Brannon, 2013). However, the association between non-symbolic number approximation and math achievement is inconsistent across studies (De Smedt et al., 2013). In contrast, research involving symbolic number approximation, such as the skill of estimating where a number falls on a number line, is more conclusive. Children who are able to accurately map estimated number with the true linear nature of a number line tend to score higher on mathematics achievement measures (Booth & Siegler, 2006). The symbolic number approximation system has been traditionally seen as an incorporation of the non-symbolic number approximation system and a language-based processing system (Verguts & Fias, 2004). However, recent work suggests that non-symbolic number approximation may be more than just a building block and may differ from the symbolic number approximation system, which may contribute to individual differences in the impact of numerosity on math achievement (e.g., De Smedt et al., 2013).

In terms of affective influences on math performance, math anxiety is commonly defined as a negative emotional reaction to situations involving math performance or the thought of math performance (Ashcraft & Krause, 2007; Hembree, 1990; Maloney & Beilock, 2012; Richardson & Suinn, 1972). There are several important characteristics of math anxiety. First, math anxiety is relatively distinct from generalized anxiety and other forms of anxiety. A correlation of around .35 is typically found between math anxiety and generalized anxiety (Ashcraft & Ridley, 2005). Second, math anxiety is not associated with general cognitive ability or other non-mathematical cognitive abilities (Ashcraft & Ridley, 2005), suggesting that low general intelligence is not the primary cause of math anxiety (Ashcraft & Krause, 2007). Finally, highly math anxious individuals do poorly in math classes as well as on standardized tests of math, and in general have poor math performance outcomes (Betz, 1978; Sepie & Keeling, 1978).

How does math anxiety negatively impact math achievement? The attentional control theory (Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007) posits that the effects of anxiety cause attentional resources to be directed away from a goal (e.g., a math test) toward understanding the nature of the threat and the determination of how to minimize the threat. Therefore, attention is oriented toward the threat-related stimuli, likely in the case of math anxiety to be worrisome thoughts and other internal stimuli. These ruminations occupy the working memory system, and because math achievement is highly dependent on working memory, math achievement suffers (Ashcraft, 2002; Eysenck et al., 2007). Therefore, the prevailing expectation is that high math anxiety should impact complex cognitive skills like math achievement and not basic skills like numerosity (e.g., Ashcraft, 2002). That said, recent work exploring the relationship between math anxiety and numerosity suggests that adults with high math anxiety have difficulties with basic counting (Maloney et al., 2010) and have less precise representations of numerical magnitude (Maloney, Ansari, & Fugelsang, 2011). In general, given the mixed support for the association between non-symbolic number approximation and math achievement (e.g., De Smedt et al., 2013), it is likely the case that there would be a minimal association of non-symbolic number approximation with math anxiety.

Given the importance of math performance for both educational and occupational attainment, the main goal of this study is to go beyond simply looking at the correlations among math achievement, numerosity and math anxiety. Instead, this study used a statistical technique called Latent Profile Analyses (LPA; described more fully in Methods) to identify groups of children categorized based on similar profiles of math achievement, numerosity and math anxiety. Although we could not find published studies that examined the relations among the measures used here, LPA has been used to examine the relations among math achievement, numerosity, and general cognitive processing factors to determine the cognitive profiles of children in different achievement groups (Geary et al., 2009). The present study extends the work of Geary and his colleagues (2009). More specifically, we will determine if there are one or more profiles across numerosity and math anxiety which seem to be particularly associated with poor math achievement. Work in the cognitive and affective literatures suggests that a profile of low numerosity and high math anxiety might by negatively associated with math achievement, but this has not been empirically tested. The literature on numerosity is unclear concerning the association between non-symbolic number approximation and math achievement. The inclusion of both non-symbolic and symbolic number approximation in the latent profile analysis will allow a direct test of whether these two systems are differentially related to math achievement across groups of children. While the majority of the math anxiety literature indicates that math anxiety is associated with low math achievement no matter the underlying numerosity skill, a recent study indicates that there may be a link between math anxiety and numerosity (Maloney, Ansari, & Fugelsang, 2011). LPA can potentially identify subsets of children who have low numerosity and high math anxiety, and explore the specific role of symbolic versus non-symbolic number approximation.

The identification of distinct profiles of numerosity, math anxiety, and math achievement is an important first step in describing how these three constructs manifest in school-age children. Understanding the etiology of these profiles may contribute to the design of more effective customized interventions for poor achievement. Socio-familial influences impact both math anxiety and math achievement (e.g., Vukovic, Roberts & Wright, 2013). For instance, parents provide home learning environments, expectations on performance, and genes, which relate to math anxiety and achievement (Vukovic, Roberts & Wright, 2013). Behavioral genetic studies of twins also indicate that math achievement is familial (Alarcon et al., 1997; Hart et al., 2009; Oliver et al., 2004; Thompson, Detterman, & Plomin, 1991), and that math anxiety is influenced by both familial and child-specific environment effects (Wang et al., 2014). If latent profile analysis is applied in a twin study, the extent to which both members of a twin pair occur in the same profile indicates familial resemblance, characterized by both genetic and environmental transmissions (Eaves et al., 1993). Importantly, math performance may also be etiologically heterogeneous, in that some profiles may reflect patterns of traits and abilities which are passed on between members of a family, and some may reflect specific individual patterns (Eaves et al., 1993). This type of analysis has been used in the ADHD literature, where phenotypically heterogeneous classes of individuals across ADHD symptoms were created and familial transmission analyzed (e.g., Neuman et al., 2001; Rasmussen et al., 2002). Although there is no previous work specific to the present research question, the phenotypic and behavioral genetics literature on math performance more broadly suggest that familial transmission is likely.

To summarize, this study simultaneously examines the relations among math achievement, numerosity, and math anxiety in children. The goal is to determine if there are profiles representing differential relations of cognitive and/or affective components related to math performance, and additionally, if there are differential etiologies for the profiles. The first aim of this study is to determine if there are distinct profiles, or classes, of individuals who show similar patterns of performance across math achievement, numerosity and math anxiety. Previous published work has used correlation-based approaches and the results are mixed in regard to the relations among math achievement, numerosity, and math anxiety. It may be the case that there are differing subpopulations of children who show differential relations, explaining the inconsistent findings in the literature. The second aim of this study is to examine whether familial influences are differentially associated with different profiles of math achievement, numerosity, and math anxiety. In total, this work provides a better understanding of the complexity of math performance, focusing on both the multivariate nature and etiological heterogeneity underlying math.

Methods

Participants

Participants were drawn from the Western Reserve Reading and Math Project (Hart et al., 2009; Petrill et al., 2007), an ongoing longitudinal twin project in the state of Ohio. Twins were considered eligible for the initial project if they were at least enrolled in kindergarten but had not yet completed first grade. Recruiting was conducted through school nominations, Ohio State Birth Records, media advertisements and personal interactions. Schools were asked to send a packet of information to parents in their school system with twins, and 293 schools participated throughout the state of Ohio. Media advertisements in the Greater Cleveland Metropolitan Area were also used. And finally a social worker with longstanding ties to the community was also hired to assist in the recruitment of under-represented groups via face-to-face meetings with churches, community centers, and other service organizations. Originally, 379 families showed interest in the project by enrolling, but 65 families (17%) were never tested through a home visit.

After seven years of the project, additional twins were recruited into the project to increase statistical power and to account for (low) attrition. This additional recruitment was targeted in the Columbus, Ohio area, and was done via media advertisements, school nomination of twins, and word of mouth. After this recruitment, the project as a whole grew to 436 families who had at any point enrolled and therefore been given a participant identification number. Twin zygosity was determined by genotyping via buccal swab or saliva sample. Fourteen percent of the families did not consent to genotyping, and therefore zygosity information was collected by a questionnaire of twin physical similarity (Goldsmith, 1991).

Thus far, twins have been assessed across eight waves of annual home visits focusing on reading and math performance. The first three of the annual home visits concentrated on early reading development. A fourth visit focused on math skills at age 8.5 years and was conducted at the six month period before or after the third reading visit, depending on the age of the child. Two additional visits occurred which focused on the development of both reading and math achievement. Finally, the "wave 7" and "wave 8" home visits, the focus of this report, occurred within 45 days of each other, when the twins were approximately 12 years old (M = 12.25, SD = 1.20). They focused on reading achievement and language skills (wave 7) and numerosity and math achievement (wave 8). Additionally, a child questionnaire was given at wave 7 and included the math anxiety questionnaire; thus, minimizing the likelihood that the math achievement and numerosity data were biased by the administration of the math anxiety questionnaire, as the questionnaire was given approximately a month prior to math performance testing.

The present report uses data from 108 monozygotic (MZ; n = 59 female-pairs), 150 same-sex dizygotic (DZ; n = 85 female-pairs) and n = 6 undetermined (n = 5 female-pairs) twin pairs from the wave 7 (questionnaire) and 8 (numerosity and math achievement) testing cycles. This 264 pairs of twins represented all families who were still participating in the longitudinal study, with the other 172 families who had been originally assigned a participant number not active in this part of the study. These 172 families include 65 families who never participated in any testing. In total, these 264 pairs of twins represent 61% of the sample size of the larger project (*N* = 436 pairs), as calculated by any twin pair that ever enrolled in the project over the years. The families of the participants who were tested for the present sample (i.e., *n* = 264) were more likely to have a White mother ($\chi^2(4) = 14.93$, $p = .005$) and more likely to have a mother with at least a 4-year degree ($\chi^2(7) = 47.56$, $p < .000$) than the participants who were not tested (i.e., *n* = 172). Ninety-one percent of the present sample was White, 5% African American, and 2% Asian. Parent education varied widely but was slightly higher than the US average: 10% had a high school education or less, 16% had attended some college, 42% had a bachelor degree, 20% had some postgraduate education, and 5% did not specify.

Procedure

A large battery of math and reading measures and a child questionnaire were administered by two testers in the twins' home over two home visits one month apart. Each home visit began with parental consent and child assent. Wave 7, the reading/language oriented home visit also included the child questionnaire. The wave 8 math oriented home visit involved a battery of math achievement and numerosity measures. Each visit lasted approximately three hours. The present paper is based on data from the math achievement, numerosity and math anxiety measures collected over these two home visits. Although the larger project is longitudinal, this was the only time point for which these measures are available.

Math achievement. Two standardized math achievement measures were employed from the Woodcock-Johnson III Tests of Achievement (Woodcock, McGraw, & Mather, 2007). Calculation measures computation ability including addition, subtraction, multiplication, division and a combination of these. Applied Problems is a measure of problem solving. A child must read the problem which may contain extraneous information, decide which mathematical operation to use, and solve the problem. Internal consistency in this sample was high, Calculation Cronbach's $\alpha = .87$, Applied Problems Cronbach's $\alpha = .83$.

Numerosity. The Number-line Task is a numerosity measure of symbolic number approximation, in that it requires translation between numerical and spatial quantity without requiring specific knowledge of measurement units (Siegler & Opfer, 2003). The "Number-to-Position" task was used, in that for each of the 22 items a number was displayed above a 0-1000 number-line. The child estimated with a pencil mark where on the line that number belongs. A difference score was calculated by taking the absolute value of the difference between the child's answer and the number requested, and a mean difference score across items was calculated. After descriptive statistics were calculated, the mean difference was reverse scored (by multiplying by negative one; larger numbers conveyed a more accurate score) for all subsequent analyses. Split-half reliability, using a Spearman-Brown correction, in this sample was high, $r = .99$. Previous work has reported that the number-line task is moderately to highly related to other measures of estimation ($r = .38$-$.66$), as well as moderately associated with math achievement scores ($r = .44$-$.54$; Booth & Siegler, 2006).

The Dots Task is a method for measuring the ability of an individual to understand and manipulate numerical quantities non-symbolically, sometimes referred to as the approximate number system (Halberda & Feigenson, 2008; Libertus & Brannon, 2009). The Dots Task was administered on a laptop twice, once at the beginning, and once at the end, of the home visit. The participant was shown a series of dot arrays with intermixed blue and yellow dots of varying sizes and quantities. For each trial, pressing the space bar would result in a blank screen delay of 250ms followed by the stimulus dots array for 200ms. The child then had an unlimited amount of time to decide which color had more dots. The color and number of dots in each array was randomized across trials but always within 5 to 16 total dots with four possible ratio bins, 1:2, 3:4, 5:6 and 7:8. The first 10 trials were practice trials, followed by 40 randomly ordered test trials representing 10 trials per possible ratio bin (Halberda & Feigenson, 2008). The Weber fraction (*w*) score represents the degree of imprecision around the response to a given numerosity comparison (e.g., three yellow dots versus four blue dots), and lower scores represent higher approximate number system capabilities. For the final *w* score here, the averaged performance across both administrations and for all test trials was used. After descriptive statistics were calculated, the final *w* score was reverse scored (by multiplying by negative one) for all subsequent analyses. Although convergent validity evidence for *w* scores from the Dots Task has been shown to be moderate and statistically significant with other measures of non-symbolic number approximation (Price, Palmer, Battista & Ansari, 2012), the construct validity evidence of the measure is still inconclusive (e.g., Price et al., 2012).

Math Anxiety. During the reading/language home visit, twins are asked to fill out the Revised Math Anxiety Rating Scale for Elementary students (MARS-E; Suinn, Taylor & Edwards, 1988). The MARS-E consists of 26 items that are rated on a 5-point Likert scale (1 = not at all nervous; 5 = very nervous), measuring how tense or anxious children feel when they are engaged in math-related activities (e.g, “If you had to add up a cash register receipt after you bought several things, how nervous would you feel”). Internal consistency in this sample was high (Cronbach’s $\alpha = .89$). Criterion validity evidence has been reported to be sufficient, with statistically significant associations measured with math achievement tests ($r = -.31$; Suinn, Taylor & Edwards, 1988).

External variables. Two measures were used to measure the association of the latent class analysis membership to external variables. The first was the math Fluency subtest of the Woodcock-Johnson III Achievement test (Woodcock, McGraw, & Mather, 2001), which measures a participant's ability to answer addition, subtraction, and multiplication problems in a 3-minute time limit. Internal consistency in this sample was moderate, Cronbach's $\alpha = .66$. The second measure was a three-item Interest in Math scale given to the twins in the child questionnaire during the wave 7 home visit. The scale was drawn from the Organization for Economic Cooperation and Development's (OCED) Program for International Student Assessment (PISA; as described in Chiu & Zihua, 2008). The items included "When I do math, I sometimes get totally absorbed", "Math is important to me personally", and "Because doing math is fun, I wouldn't want to give it up", all scored on a 4-point Likert scale (1 = strongly disagree to 4 = strongly agree). A mean score was used, and internal consistency in this sample was moderate, Cronbach's $\alpha = .78$.

Analysis

Latent Profile Analysis (LPA), a type of Latent Class Analysis, is an empirically driven technique used to classify individuals into groups based on responses over multiple continuous indicators. In LPA, model fitting begins by the user setting the number of potential classes (i.e., profiles) to be estimated. Based on the user defined number of classes, the model determines the best possible group membership for each individual based on similarities across individuals and across indicators. For this study, the number of classes which could be expected was unknown. Thus, we used an exploratory method of determining the optimal number of classes (Muthén & Muthén, 1998-2007). We fitted several models to the data, each increasing the number of classes defined from 2 through 11 classes. We did this two ways, one correcting for family-level clustering via the cluster option in Mplus, and the other without regard to the individual's co-twin. As suggested by Muthén, Asparouhov and Rebello (2006), the only potential issue due to family-level clustering in an LPA is typically seen in a different number of classes being indicated between models with and without the clustering correction. The results from the clustering correction indicated that the same number of classes should be accepted as the model without clustering (see Supplemental materials for cluster correction results). To keep the simplest model, the model without correction for clustering was used in the final analyses (see also Vendlinski et al., 2014).

Using the Nylund, Asparouhov and Muthén (2007) guidelines for latent profile analysis, multiple steps were taken when determining the best fitting model. As a first step, the model with the lowest Bayesian information criterion (BIC) coupled with a statistically significant bootstrap likelihood ratio test (BLRT) was considered as potentially the best fitting model. Importantly, a statistically significant BLRT test is not sufficient to determine the overall best fitting model, but it was considered necessary in conjunction with BIC values. Using just the BLRT, the best fitting model is the (k – 1)-class model after the first statistically nonsignificant k-class model is determined. As a second step beyond model fit statistics, it is recommended by Jung and Wickrama (2008) that consideration is given towards successful model convergence, a high entropy value (greater than .8, or closest to 1.0; Ramaswamy et al., 1993), no less than 1% of the total participant count in a given class, and high posterior probabilities (close to 1.0). High posterior probabilities indicate that there is high confidence that an individual assigned to a given class actually belongs to that class. After the final model was selected, sample statistics (i.e., standard deviations and correlations) for each cluster were computed by weighting all observations by the posterior probabilities associated with the cluster (Pastor, Barron, Miller & Davis, 2007).

Raw data for every child available was first residualized with age, age squared, gender, total months-of-schooling (i.e., start of kindergarten to the month of testing for the wave 8 testing battery) and total months-of-schooling squared regressed out (to remain consistent across papers from this project; see Hart et al., 2009[1]), and residualized data were subsequently *z*-scored. Modeling was done in Mplus 7.0 (Muthén & Muthén, 1998-2007). Missing data was handled using the MLR (maximum likelihood estimation with robust standard errors) estimator, which yields unbiased estimates when the pattern of missing data is completely at random (Yuan & Bentler, 2000). After z-scoring, 6 cases for Dots Task and 14 cases for Number-line had scores less than three standard deviations below the mean (-3*SD*). These scores were recoded to missing, given the potential for extreme scores to skew the means of the latent classes, as well as possibly distorting the *p*-value of the BLRT (Nylund et al., 2007)[2]. Of the n = 264 families, 20 children were missing data on all measures used in the LPA and were therefore not included in the analyses. Number-line had the greatest proportion of missing data of the LPA measures at 10.7% (see Table 1). Little's MCAR test, including all measures of the LPA, was statistically nonsignificant, indicating that these data could be interpreted as missing completely at random

[1] Hart et al. (2009) indicated that in the broader twin project there is not a complete relationship between age and months-of-schooling completed. Given months-of-schooling was shown to be a statistically significant predictor of math outcomes beyond age, and months-of-schooling would serve to inflate familial resemblance estimates as it is known variance shared between twins, it was decided to control for not only age but months-of-schooling.

[2] There is no set method for how to deal with outliers such as these. Sensitivity analyses were conducted in that the present modeling was compared to the results of the same modeling where outliers had been recoded to -3*SD* (Cohen, Cohen, West, Aiken, 2001), rather than set to missing. The same five-class model is accepted, with all classes but class 1 remaining the same. Class 1 shifts to represent these additional low numerosity data points, with mean scores for both numerosity variables reflecting lower mean scores than is represented presently. Additionally, the achievement variables are closer to the mean in this alternative model.

($\chi^2(30) = 35.86$, $p = .213$; Little, 1988). All measures were coded so that for math achievement and numerosity, higher scores represented "better" performance and higher scores indicted higher math anxiety.

Post-hoc testing. After the best LPA model was selected, each individual in the data set was assigned to a class given his or her scores across the indicators. This class outcome variable was exported from Mplus and brought into SAS 9.4, and used in a series of five ANOVAs, one per math achievement, numerosity, and math anxiety measure. Of key interest, Tukey's post-hoc tests were analyzed to determine if there were statistically significant differences between the classes on each of the measures included in the LPA.

Class membership relations to external variables of math Fluency and Interest in Math. In an effort to test the validity of the classifications derived from the LPA class membership scores were used in two ANOVAs. The first ANOVA examined differences across the class memberships in an external measure of math achievement, specifically math Fluency, and the second in Interest in Math.

Familial resemblance. Because all children were used in the LPA and these data are from a twin study, we were able to determine if familial resemblance was related to class membership. The affected-status agreement statistic was calculated for each class to determine the extent to which twin pairs were concordant (or discordant) for that class (Waesche, Schatschneider, Maner, Ahmed & Wagner, 2011). The affected-status agreement is the proportion of twins where both members of the twin pair were assigned as being in a certain class compared to the number of individual twins who were assigned as being in that class but their twin was not. This calculation allows for a percent agreement (i.e., assignment to the class within a family) to be determined, with confidence intervals. The reason this statistic was used, rather than odds ratios as commonly reported for this type of analysis, is that it allows for a measurement of agreement of twin classification within a certain class which is not inflated by agreement of twin classification to all classes outside that certain class (which is the most likely case of agreement here). This was because the primary research question here was the extent to which twin pairs were concordant for a class if at least one twin member was assigned that class.

Using the affected-status agreement statistic, if the percent agreement was greater than 50% (represented by confidence intervals not overlapping .50), it was more likely that there was twin concordance for that given class, indicating that being in the same family was associated with a higher chance of being assigned to the same class (i.e., familial resemblance). If the percent agreement was less than 50% (represented by confidence intervals not overlapping .50), it was more likely that there was twin discordance for that given class, indicating being in the same family was associated with a lower chance of being assigned to the same class (i.e., individual specific effects). If the percent agreement was equal to 50% (represented by confidence intervals overlapping .50), then being in the same family did not affect the chance of being assigned to the same class[3].

Results

Descriptive statistics for all measures are presented in Table 1. In general, children scored at population average on Calculation but higher than average on Applied Problems and lower than average for Fluency (based on the norming samples of the original tests). The mean *w* score for the sample was .37, which represents an ability to differentiate bin ratios of

---

[3] Note that we considered a full decomposition of variance to estimate additive genetic, shared environmental and nonshared environmental influences on class membership by differentiating monozygotic versus dizygotic twins (i.e., rather than just familial or not). However, the final LPA model selected resulted in class sizes that were too small, and therefore underpowered, to do a full biometric analyses.

approximately three dots of one color to four dots of another color accurately (Halberda & Feigenson, 2008). This performance is lower than the typical adult ratio, which is to be expected in a sample of children (Halberda & Feigenson, 2008). The mean difference between the requested number and the drawn number for the Number-line Task was about 74 units, or 74 numbers off from the requested number on a number-line of 0 to 1000. Pearson correlations among the measures are presented in Table 2. In general, all associations among the measures used in the LPA were low to moderate and statistically significant, except the Dots Task which was not statistically significantly correlated with Math Anxiety, and the two numerosity measures were not statistically significantly correlated with the external validity check measure of Interest in Math. In general, math achievement measures appeared to be more strongly associated with math anxiety than the numerosity measures. This supports the previous literature suggesting that math performance outcomes are more susceptible to the negative effects of math anxiety (e.g., Ashcraft, 2002), although the current results indicate that numerosity, when defined as symbolic number approximation, may also be associated with math anxiety to a lesser degree (e.g., Maloney & Beilock, 2012).

Latent Profile Analyses

For this study, ten models were generated and compared, each testing a different number of possible classes (2 through 11; see Table 3). The model fitting results varied depending on what model fit statistic was considered. The 7 class model was the best fitting model according to the lowest BIC. BLRT tests were all statistically significant until the 11 class model, indicating that the best fitting model using this fit index was the 10 class model. Turning toward entropy, the 5 class model was the best fitting model, with the highest value of .85. Although the 4, 7 and 11 class model had entropy of .84 (therefore close to the 5 class model), the 4 class model did not fit as well across the other metrics compared to the 5 class model, the 7 class model had a class with approximately only 1% of the sample represented, and the 11 class model which had a statistically nonsignificant BLRT test (see Table 4). Finally, the posterior probabilities for each model were high (see Table 4). Balancing these findings, it was decided to accept the 5 class model, as it had a low BIC with a statistically significant BLRT (Nylund et al., 2007), had the highest entropy, reasonable representation of participants across classes, and good posterior probabilities. Results from the 5 class model are in Figure 1[4]. Data in Figure 1 are *z*-scored, allowing each profile's performance on the measures to be interpreted in *SD* units to the sample mean.

Class 1: Very Low Achievers with mixed Numerosity. The first class extracted was small ($n = 18$) and characterized by math achievement that was approximately 1.5 *SD* below average, with equally low Dots Task performance but only slightly below average Number-line performance. Additionally, this group had slightly high math anxiety. Results from this class indicated statistically nonsignificant correlations between all the measures (see Table 5).

Class 2: Low Symbolic Number Approximation. This class extracted the smallest proportion of children ($n = 17$), a group presenting very poor (2.20 *SD* below the mean) performance in the Number-line Task. The children in this class also performed below average on both math achievement measures (.35-.78 *SD* below the mean), although their performance on the Dots Task was slightly above average. These children had slightly greater than average math anxiety (about .28 *SD* above average). Sample statistics from this class indicated that there was a

[4] The 4 class model was a reasonable second choice for the final model. The classes from this model mirrored the 5 class model patterns, but did not include class 1. The remaining other classes in the 4 class model resulted in the same conclusions drawn here.

high and statistically significant positive correlation between the achievement measures (see Table 5). All other correlations were statistically nonsignificant.

Class 3: Low Non-symbolic Number Approximation. This class was small (*n* = 29), and represented children who scored poorly on the Dots Task (1.33 *SD* below the mean) despite about average performance on the achievement measures, slightly above average for the Number-line Task and lower than average math anxiety. The sample statistics indicated a moderate and statistically significant positive correlation between the achievement measures only (see Table 5).

Class 4: High Performers. This extracted group was the largest of the sample, representing about half of the participants (*n* = 295). This class performed above average on all of the math measures, and also had the lowest scores for Math Anxiety. There was a low and statistically significant positive correlation between the achievement measures (see Table 5).

Class 5: Anxious Low Achievers. The second largest extracted class (*n* = 149) indicated average performance for numerosity, yet low performance for math achievement. They were also notable for having the highest math anxiety levels (0.57 *SD* above average). There was a low statistically significant positive correlation between the Number-line Task and Applied Problems (see Table 5). All other correlations were statistically nonsignificant.

Post-hoc testing. Results from the ANOVAs indicated that there was a statistically significant relation between class membership and Calculation ($F(4, 484) = 102.44$, $p < .0001$), Applied Problems ($F(4, 500) = 155.78$, $p < .0001$), Dots Task ($F(4, 477) = 209.21$, $p < .0001$), Number-line Task ($F(4, 468) = 146.79$, $p < .0001$) and Math Anxiety ($F(4, 473) = 26.76$, $p < .0001$). Tukey's post-hoc testing was explored to test the seemingly small differences between the mean scores on some of the measures between the classes (see Figure 1, means in different ellipses are statistically significantly different from each other). Results across the math achievement measures were similar, with all classes being statistically significantly different from each other across the two measures except for class 2 and class 5, and for Calculation alone, class 2 and class 3 were not statistically significantly different. This indicates that class 1 performed statistically significantly worse on math achievement than the other four classes, and class 4 scored statistically significantly better than the other classes. For the Dots Task, the measure of non-symbolic number approximation, the two lowest and two highest performing classes were not statistically significantly different from each other, and although the scores were similar, class 5 was only statistically significantly different from class 4, but not class 2 (likely due to differences in sample size across the two classes). For the Number-line Task, the measure of symbolic number approximation, all classes were statistically significantly different from each other except class 3 and class 4, and class 1 and class 5. Finally, post-hoc results indicated that for math anxiety, there was a group of classes who had higher than average anxiety (classes 1, 2 and 5), and another mixed group (all but class 5).

Class membership relations to external variables of math Fluency and Interest in Math

Results from the two ANOVAs indicated that there was a statistically significant relation between class membership and math Fluency levels ($F(4, 493) = 37.55$, $p < .0001$), as well as between class membership and Interest in Math ($F(4, 376) = 5.46$, $p = .0003$). Z-score averages were plotted for math Fluency and Interest in Math by class for descriptive purposes in Figure 2. The ellipses in the figure represent the results of the Tukey's post-hoc pairwise comparison tests, in that means in different ellipses are statistically significantly different from each other. In total, the results offered some support to the validity of the classes assigned in the latent profile analysis.

Familial Resemblance

To test whether the extracted classes tended to run in families, the affected-status agreement statistic of individuals in a given class being concordant in that class with their co-twin was calculated (see Table 6). Descriptively, more twins were concordant for class membership than discordant for class 4 only (n = 224 versus n = 71). For the remaining classes, there were more discordant than concordant twins. The affected-status agreement statistic was calculated by the number of concordant twins divided by the total number of individuals assigned to that class. For example, for class 1, the affected-status agreement statistic was calculated by 6 divided by the sum of 12 + 6, or .33 (Waesche et al., 2011).

When the results, with confidence intervals, were examined, differential etiology of the classes was indicated. For individuals in class 2 and class 3, results indicated that for all individuals in the class, only 0% and 14% (respectively) were concordant with their twin for classification within that class. It was more likely that twins were discordant for these two classes, or, individual specific effects on class membership were indicated. For individuals in class 4, the affective status agreement statistic indicated that class membership was associated with a higher chance of both twins being assigned to the same class, in that of all individuals in the class, 76% were concordant by twin pair. For this class, class membership showed familial resemblance. Individuals in class 1 and class 5 had an affected-status agreement statistic in the direction of individual specific influences (33% and 46% of individuals within twin pairs were concordant respectively by class), although for both the confidence intervals overlapped with .50, indicating that being in the same family did not statistically affect class membership.

Discussion

In general, the goal of this study was to examine whether there are profiles of children representing differential relations of cognitive and/or affective components related to math performance, and additionally, whether there are differential familial etiology of those profiles. Previous work has been mostly limited to correlational designs, which assume a single underlying association among math achievement, numerosity and math anxiety. By using latent profile analysis, this study was able to identify multiple profiles representing different associations of math achievement, numerosity and math anxiety. The results indicated that five groupings of individuals with similar patterns of performance existed in these data.

Three classes, all small, appeared to represent groups of children who possessed various numerosity deficits (classes 1, 2 and 3). What is most interesting about these classes is that the relation of poor numerosity, measured as both non-symbolic and symbolic number approximation, to low achievement is not consistent. Some children showed low achievement with low non-symbolic number approximation only, others showed low achievement with low symbolic number approximation only, and finally some children showed average achievement with low non-symbolic number approximation. On one hand, the absence of a relation between numerosity and math achievement is not surprising, as many other factors have been indicated to be important in math achievement (e.g., Fuchs et al., 2005; Geary et al., 2007). However, it was expected that at least symbolic number approximation would be associated with math achievement, a finding commonly supported in the literature, compared to the more controversial association (or lack thereof) between non-symbolic number approximation and math achievement (De Smedt et al., 2013). This highlights the overall lack of understanding of the link between non-symbolic number approximation and symbolic number approximation to eventual math achievement (Siegler & Lortie-Forgues, 2014).

These classes also suggest that there appears to be a weak link at best between low numerosity and high math anxiety. When children had higher than average math anxiety, they consistently struggled with math achievement, and not necessarily with numerosity. This finding is not entirely surprising, as the initial conceptualization of math anxiety was that it relates solely to math achievement and not to basic numerosity skills (e.g., Ashcraft, 2002). However, in total, caution should be made in using the present findings to advocate either way in the debate of the role of numerosity in math achievement and/or anxiety, as these groups of children represented small proportions of the total sample and could represent unique, or possibly spurious, findings. This is especially important to remember when considering the results of class 1, as there was some indication that this class is not robust (i.e., it is dropped in the four-class solution model, and captures a slightly different class of children when outliers are treated differently).

One other group further differentiated low performers, representing the second largest grouping of children. Class 5 demonstrated that despite having typical numerosity, the children had achievement difficulties which were also associated with high math anxiety. Unfortunately, given our study design, the causal nature of the relation between low math achievement and high math anxiety is unknown. However, it is clear that a large group of children in this sample had high math anxiety and poor math achievement, seemingly despite typical underlying basic numerosity, further supporting an association between math achievement and math anxiety, which was also seen in the smaller classes (i.e., classes 1 and 2; Ashcraft, 2002). This finding supports the attentional control theory, which speaks to the role of math anxiety in diminishing the attentional resources of goal-directed attentional systems, such as math achievement (Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007).

The final group represented the largest proportion of the sample and the highest performers, class 4. This group performed above average on math achievement and numerosity and had the lowest math anxiety levels. Looking at this class in comparison with the others, it appears that individuals who performed above average on one measure also did well on the others, with some exceptions for the numerosity measures. This is in contrast to the low performance groups (i.e., classes 1, 2, 3, and 5), which were differentiated into different classes depending on individual measures. This suggests that there are many ways to perform poorly in math. The opposite did not appear to be as true. Coupled with this, high performers had low math anxiety. This indicates that in this sample there was no group of children who performed considerably above average on the math achievement and/or numerosity measures and also had high math anxiety.

The complexity of math performance was further explored by examining how familial influences affected classification. Interestingly, only the highest performers indicated familial influences, in that genetic and environmental influences shared between siblings support high performance in math achievement and numerosity, and low math anxiety. This supports previous work indicating both aspects of familial influences, additive genetic and shared environmental effects, on high math achievement (Petrill et al., 2009). For classes involving students who showed the average to lowest math achievement, child-specific environmental influences were implicated. Typically, the lower achievers also had math anxiety. This finding mirrors a recent finding that indicates that child-specific environmental influences are important sources of variance on math anxiety (Wang et al., 2014). Perhaps the math anxiety and achievement link, as described above based on the LPA findings across the classes, is through child-specific poor math-related experiences, such as individual interactions with math anxious teachers (Beilock et al., 2010), or through personal perceptions and expectancies (Meece, Wigfield & Eccles, 1990). In total, these results indicate etiological heterogeneity, in that there is no one etiology for good and/or poor math performance.

A limitation of this study is that the total sample size did not allow for more specific inferences concerning the genetic and environmental influences on the class membership (e.g., Neuman et al., 2001). Some classes were too small to further divide into monozygotic versus dizygotic twin pairings, which is what is needed to separate "familial resemblance" into the specific pieces of additive genetic and shared environmental effects. An additional possible limitation is the use of a sample of twins alone, which may be restricted in their representativeness of the general population of children (e.g., Rutter, Thorpe, Greenwood, Northstone & Golding, 2003). Despite this, it has been suggested that any major differences in cognitive performance between singletons and twins become statistically nonsignificant by the age of 5, and would therefore be negligible by the age of the present sample (Evans & Martin, 2000). Finally, by nature these analyses are sample specific, and these classes may not be found in other samples. We assert that other studies should attempt to replicate the present findings, to determine if these results are specific to this sample or not. The point of this work was to highlight the differential relations among the included measures beyond simple correlations, specifically to gain a better understanding of the complex nature of math performance. For future work, we propose to extend this model longitudinally to determine if the classifications found here hold up over time, or if there are developmental changes that occur across the classes.

The results from this work suggest that math performance is complex when considered in the context of math achievement, numerosity and math anxiety, and support the idea that poor achievement is associated with math anxiety. Additionally, there is not a clear association of numerosity, measured as non-symbolic or symbolic number approximation, to math achievement, other than if children are high achievers they also tend to have high numerosity. But low numerosity does not necessarily mean low achievement. Further, it was indicated that for the students in classes who struggled in some way, whether it be low achievement, low numerosity and/or high anxiety, there was at least a trend towards child-specific environmental influences on the given class. This finding indicates that individual children have different paths that lead to struggles with math, and these experiences may be child-specific. In total, these results have important implications for understanding why some children underperform in math, and may suggest that a one-dimensional, one-for-all intervention approach focusing on only cognitive or affective aspects may not work for all children who struggle. This present work suggests interventions that target multiple areas of math performance, for instance training on numerosity while also working on math achievement skills (e.g., calculation) with a component meant to reduce math anxiety, may be the most likely interventions to work with children who are showing math difficulties. We hypothesize that future studies which include multi-component interventions for math difficulties will be more successful than single component interventions, such as intervening on non-symbolic number approximation only. In conclusion, success in math seems to be uni-dimensional and runs in families, yet struggling in math seems to be complex and individual-specific.

Acknowledgements

This work was supported by NICHD grants HD038075, HD059215 and HD075460. The authors wish to thank the twins and their families for making this research possible.

References

Alarcon, M., DeFries, J.C., Light, J.G., & Pennington, B.F. (1997). A twin study of mathematics disability. *Journal of Learning Disabilities, 30*, 617-623. doi: 10.1177/002221949703000605

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185. doi: 10.1111/1467-8721.00196

Ashcraft, M. H., & Krause, J.A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243-248. doi:10.3758/bf03194059

Ashcraft, M. H., & Ridley, K. S. (2005). Math Anxiety and Its Cognitive Consequences. In J.I.D. Campbell (Ed.), *Handbook of Mathematical Cognition* (315-327). New York: Taylor & Francis.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860-1863. doi: 10.1073/pnas.0910967107

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology, 25*, 441-448. doi: 10.1037/0022-0167.25.5.441

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201. doi:10.1037/0012-1649.41.6.189

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*, 534–541. doi:10.1016/j.tics.2010.09.007

Chiu, M.M., & Xihua, Zeng. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction, 18*, 321-336. doi:10.1016/j.learninstruc.2007.06.003

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed). Mahwah, NJ: Erlbaum.

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48-55. doi:10.1016/j.tine.2013.06.001

Dreger, R. M., & Aiken, L. R. (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology, 48*(6), 344-351. doi: 10.1037/h0045894

Eaves, L. J., Silberg, J. L., Hewitt, J. K., Rutter, M., Meyer, J. M., Neale, M. C., & Pickles, A. (1993). Analyzing twin resemblance in multisymptom data: genetic applications of a latent class model for symptoms of conduct disorder in juvenile boys. *Behavior Genetics*, *23*(1), 5-19. doi: 10.1007/bf01067550

Evans, D.M., Martin, N.G. (2000). The validity of twin studies. *GeneScreen,* 1(2), 77-79. doi: 10.1046/j.1466-9218.2000.00027.x

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion, 6*, 409-434. doi: 10.1080/02699939208409696

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*, 336-353. doi:10.1037/1528-3542.7.2.336

Feigenson, L., Libertus, M.E., Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives, 7*(2), 74-79. doi: 10.1111/cdep.12019

Fennema, E. (1989). The study of affect and mathematics: A proposed generic model for research. In D.B. McLeod & V.M. Adams (Eds.), *Affect and Mathematical Problem Solving* (pp. 205-219). New York: Springer. doi: 10.1007/978-1-4612-3614-6_14

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 9*7, 493-513. doi: 10.1037/0022-0663.97.3.493

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive development, 24*, 411-429. doi:10.1016/j.cogdev.2009.10.001

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343-1359. doi: 10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology,104*, 206-223. doi:10.1037/a0025398

Goldsmith, H. Hill. (1991). A zygosity questionnaire for young twins: A research note. *Behavior Genetics, 21*, 257-269. doi: 10.1007/bf01065819

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*, 1457-1465. doi:10.1037/a0012682

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature.* 455, 665-668. doi: 10.1038/nature07246

Hart, S. A., Petrill, S.A., Thompson, L.A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology, 101*, 388-402. doi:10.1037/a0015115

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *1*, 33-46. http://www.jstor.org/stable/749455

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*, 92-107. doi:10.1016/j.cognition.2013.12.007

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 2*0, 82-88. doi:10.1016/j.lindif.2009.07.004

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*, 302-317. doi: 10.1111/j.1751-9004.2007.00054.x

Kaplan, D. (1989). Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance. *Educational and Psychological Measurement, 49*, 579-586. doi: 10.1177/001316448904900308:

Libertus, M.E., & Brannon, E.M. (2009). Behavioral and neural basis of number sense in infancy. *Current Directions in Psychological Science, 18*, 346-351. doi: 10.1111/j.1467-8721.2009.01665.x

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198-1202. doi: 10.1080/01621459.1988.10478722

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767-778. doi: 10.1093/biomet/88.3.767

Maloney, E. A., Ansari, D., & Fugelsang, J.A. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *The Quarterly Journal of Experimental Psychology, 64*, 10-16. doi:10.1080/17470218.2010.533278

Maloney, E. A., & Beilock, S.L. (2012). Math anxiety: who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences, 16*, 404-406. doi:10.1016/j.tics.2012.06.008

Maloney, E.A., Risko, E.F., Preston, F., Ansari, D., Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica, 123*, 154-161. doi:10.1016/j.actpsy.2010.01.006

Maloney, E. A., Risko, E.F., Ansari, D., & Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition, 114*, 293-297. doi: 10.1016/j.cognition.2009.09.013

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*, 1224-1237. doi: 10.1111/j.1467-8624.2011.01608.x

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 575-596). New York: MacMillan Publishing Company.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*, 60-70. doi:10.1037//0022-0663.82.1.60

Muthén, B., Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics*, *9*, 313-324. doi: 10.1375/twin.9.3.313

Muthén, L.K., & Muthén, B. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Neuman, R. J., Heath, A., Reich, W., Bucholz, K. K., Madden, P. A., Sun, L., ... & Hudziak, J. J. (2001). Latent class analysis of ADHD and comorbid symptoms in a population sample of adolescent female twins. *Journal of Child Psychology and Psychiatry, 42*, 933-942. doi: 10.1111/1469-7610.00789

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535-569. doi: 10.1080/10705510701575396

Oliver, B., Harlaar, N., Hayiou Thomas, M.E., Kovas, Y, Walker, S.O., Petrill, S.A., . . . Plomin, R. (2004). A Twin Study of Teacher-Reported Mathematics Performance and Low Performance in 7-Year-Olds. *Journal of Educational Psychology, 96*, 504-517. doi:10.1037/0022-0663.96.3.504

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science, 24*, 2013-2019. doi: 10.1177/0956797613482944

Pastor, D.A., Barron, K.E., Miller, B.J., Davis, S.L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*, 8-47. doi:10.1016/j.cedpsych.2006.10.003

Petrill, S.A., Deater-Deckard, K., Thompson, L.A., Schatschneider, C., DeThorne, L.S., & Vandenbergh, D.J. (2007). Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. *Reading and Writing, 20*, 127-146. doi: 10.1007/s11145-006-9021-2

Petrill, S. A., Kovas, Y., Hart, S. A., Thompson, L. A., & Plomin, R. (2009). The genetic and environmental etiology of high math performance in 10-year-old twins. *Behavior genetics, 39*, 371-379. doi: 10.1007/s10519-009-9258-z

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta psychologica, 140*, 50-57. doi: 10.1016/j.actpsy.2012.02.008

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*, 103-124. doi: 10.1287/mksc.12.1.103

Rasmussen, E. R., Neuman, R. J., Heath, A. C., Levy, F., Hay, D. A., & Todd, R. D. (2002). Replication of the latent class structure of Attention-Deficit/Hyperactivity Disorder (ADHD) subtypes in a sample of Australian twins. *Journal of Child Psychology and Psychiatry, 43*, 1018-1028. doi: 10.1111/1469-7610.00229

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology, 19*, 551-554. doi: 10.1037/h0033456

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K., & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Twin–singleton differences in language, and obstetric risks. *Journal of Children Psychology and Psychiatry, 44,* 326-341. doi: 10.1111/1469-7610.00125

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*, 344-357. doi: 10.1111/j.2044-835X.2011.02048.x

Sepie, A. C., & Keeling, B. (1978). The Relationship between Types of Anxiety and Under-Achievement in Mathematics. *Journal of Educational Research, 72*, 15-19. doi: 10.1080/00220671.1978.10885111

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives*, *8*, 144-150. doi: 10.1111/cdep.12077

Siegler, R.S., & Opfer, J.E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237-243. doi: 10.1111/1467-9280.02438

Suinn, R.M., Taylor, S., Edwards, R.W. (1988). Suinn mathematics anxeity rating scale for elementary school students (MARS-E): Psychometric and normative data. *Educational and Psychological Measurement, 48*, 979-986. doi: 10.1177/0013164488484013

Thompson, L.A., Detterman, D.K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science, 2*, 158-165. doi: 10.1111/j.1467-9280.1991.tb00124.x

Vendlinski, M. K., Javaras, K. N., Van Hulle, C. A., Lemery-Chalfant, K., Maier, R., Davidson, R. J., & Goldsmith, H. H. (2014). Relative Influence of Genetics and Shared Environment on Child Mental Health Symptoms Depends on Comorbidity. *PloS One*, *9(7),* e103080. doi: 10.1371/journal.pone.0103080

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: a neural model. *Journal of Cognitive Neuroscience, 16*, 1493-1504. doi:10.1162/0898929042568497

Vukovic, R. K., Roberts, S. O., & Green Wright, L. (2013). From Parental Involvement to Children's Mathematical Performance: The Role of Mathematics Anxiety. *Early Education & Development*, *24*, 446-467. doi: 10.1080/10409289.2012.693430

Waesche, J. S. B., Schatschneider, C., Maner, J., Ahmed, Y., & Wagner, R. (2011). Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities*, *44*, 296-307. doi: 10.1177/0022219410392048

Wang, Z., Hart, S.A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L.A., Plomin, R., McLoughlin, G., Bartlett, C.W., Lyons, I.M., & Petrill, S.A. (2014). Who is afriad of math? Two sources of genetic variance for mathematical anxiety. *The Journal of Child Psychology and Psychiatry, 55,* 1056-1064. doi: 10.1111/jcpp.12224

Woodcock, R.W., McGraw, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing. doi: 10.1177/003435520104400407

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology, 30*, 165-200.

doi: 10.1111/0081-1750.00078

Table 1. *Descriptive statistics for all measures*

Variable	Mean	*SD*	Min.	Max.	Skew	*n*
Calculation	100.80	15.52	52.00	146.00	-0.23	493
Applied Problems	106.37	11.80	60.00	143.00	-0.40	509
Dots Task	0.37	0.17	0.09	1.14	1.85	484
Number-Line Task	73.63	55.18	16.23	339.16	2.28	475
Math Anxiety	48.77	16.44	14.00	126.00	0.92	506
Fluency	83.01	25.77	18.00	160.00	0.40	502
Interest in Math	2.59	0.75	1.00	4.00	0.00	389

Note. Standard scores are displayed for Calculation, Applied Problems and Fluency for sample comparison purposes. *n* represents the sample size available on each measure, which does not equal total available sample size across measures.

Table 2. *Pearson correlations among the measures.*

Variable	1	2	3	4	5	6	7
1. Calculation	--						
2. Applied Problems	.66*	--					
3. Dots Task	.31*	.29*	--				
4. Number-line Task	.20*	.34*	.10*	--			
5. Math Anxiety	-.32*	-.35*	-.07	-.16*	--		
6. Fluency	.60*	.49*	.27*	.16*	-.24*	--	
7. Interest in Math	.25*	.25*	.07	.06	-.36*	.24*	--

Note. Correlations marked with an asterisk (*) were statistically significant at $p < .05$.

Table 3. *Model fit indices for the ten tested models*

Model	Log Likelihood (LL)	Free Parameters	Bayesian Information Criterion (BIC)	Bootstrap likelihood ratio test (BLRT) *p*-value	Entropy
2 classes	-2851.85	16	5803.39	.00	.76
3 classes	-2775.99	22	5689.04	.00	.82
4 classes	-2702.02	28	5578.49	.00	.84
5 classes	**-2677.64**	**34**	**5567.12**	**.00**	**.85**
6 classes	-2654.97	40	5559.17	.00	.81
7 classes	-2627.95	46	5542.50	.00	.84
8 classes	-2610.20	52	5544.39	.00	.82
9 classes	-2592.52	58	5546.41	.00	.83
10 classes	-2580.60	64	5559.95	.03	.82
11 classes	-2567.75	70	5571.50	.05	.84

Note. Bold indicates best fitting model chosen.

Table 4. *Highest average posterior probabilities associated with class membership for each of the ten models*

	2 class model		3 class model		4 class model		5 class model		6 class model		7 class model		8 class model		9 class model		10 class model		11 class model	
Class	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.	*n*	p.p.
1	332	.94	18	.97	41	.89	18	.88	14	.94	13	.96	34	.86	7	.98	31	.91	6	.99
2	176	.92	171	.89	304	.93	17	.99	35	.87	14	.95	12	.99	96	.84	14	.93	12	.94
3			319	.94	146	.86	29	.85	17	.99	35	.88	15	.92	5	1.00	237	.85	14	.93
4					17	.99	295	.92	270	.85	254	.87	98	.85	251	.85	12	.99	101	.85
5							149	.88	117	.89	66	.81	5	1.00	12	.99	72	.79	12	.99
6									55	.81	120	.89	253	.85	14	.92	11	.84	238	.86
7											6	.91	30	.77	59	.80	7	.98	6	.93
8													61	.81	31	.92	57	.75	1	1.00
9															33	.76	5	1.00	28	.77
10																	62	.81	21	.88
11																			69	.81

Note. *n* = sample size of each class. p.p. = posterior probability.

Table 5. *Sample statistics for each class.*

Class 1	Calculation	Applied Problems	Dots Task	Number-line Task	Math Anxiety
Calculation	0.80				
Applied Problems	0.13 [-.36, .56]	0.89			
Dots Task	-0.15 [-.58, .34]	-0.21 [-.62, .28]	0.58		
Number-line Task	-0.31 [-.68, .18]	-0.17 [-.59, .32]	-0.33 [-.69, .16]	0.60	
Math Anxiety	0.13 [-.36, .56]	-0.26 [-.65, .24]	0.33 [-.16, .69]	-0.24 [-.63, .26]	0.99
Class 2	**Calculation**	**Applied Problems**	**Dots Task**	**Number-line Task**	**Math Anxiety**
Calculation	0.96				
Applied Problems	0.67* [.28, .87]	1.02			
Dots Task	-0.04 [-.51, .45]	0.09 [-.41, .55]	0.30		
Number-line Task	-0.16 [-.59, .35]	-0.05 [-.52, .44]	-0.21 [-.62, .30]	0.54	
Math Anxiety	-0.36 [-.72, .15]	-0.13 [-.57, .37]	0.43 [-.06, .75]	-0.27 [-.66, .24]	0.79
Class 3	**Calculation**	**Applied Problems**	**Dots Task**	**Number-line Task**	**Math Anxiety**
Calculation	0.81				
Applied Problems	0.47* [.13, .71]	0.55			
Dots Task	-0.24 [-.56, .14]	-0.02 [-.38, .35]	0.55		
Number-line Task	-0.14 [-.48, .24]	0.08 [-.29, .13]	-0.17 [-.50, .21]	0.45	
Math Anxiety	-0.19 [-.52, .19]	-0.47 [-.71, .13]	-0.09 [-.44, .29]	0.16 [-.22, .50]	0.94
Class 4	**Calculation**	**Applied Problems**	**Dots Task**	**Number-line Task**	**Math Anxiety**
Calculation	0.72				
Applied Problems	0.39* [.29, .48]	0.68			
Dots Task	0.06 [-.06, .17]	0.03 [-.09, .14]	0.35		
Number-line Task	-0.05 [-.16, .07]	0.10 [-.02, .21]	0.08 [-.04, .19]	0.43	

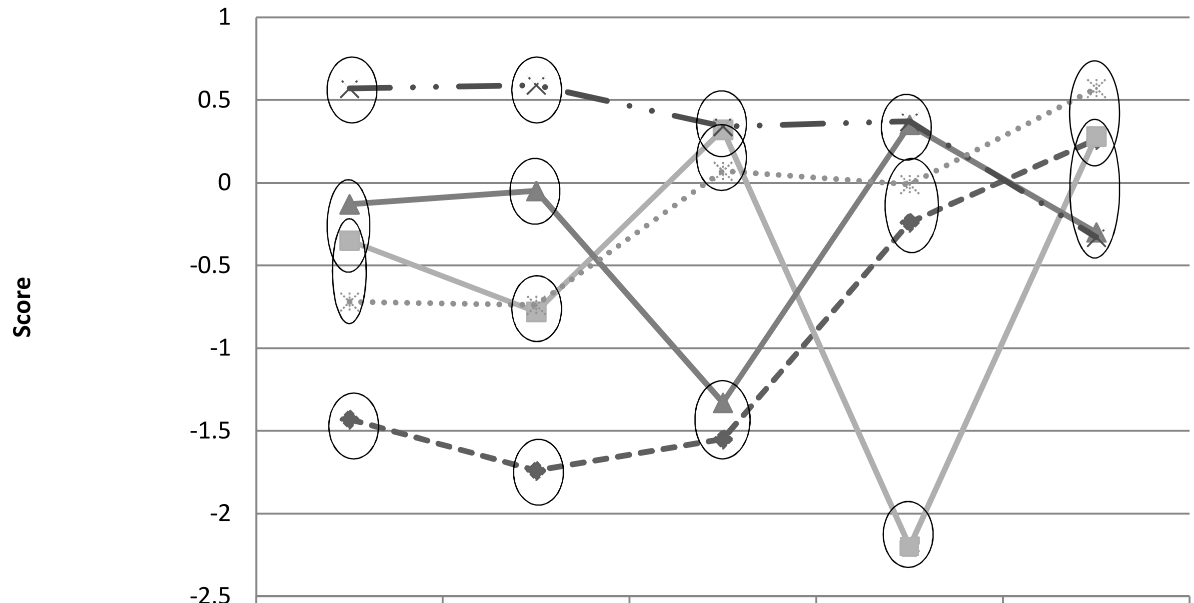
Math Anxiety	-0.02 [-.13, .10]	-0.03 [-.14, .09]	-0.03 [-.14, .09]	-0.06 [-.17, .06]	0.79
Class 5	Calculation	Applied Problems	Dots Task	Number-line Task	Math Anxiety
Calculation	0.72				
Applied Problems	0.16 [.00, .31]	0.69			
Dots Task	0.07 [-.09, .23]	-0.06 [-.22, .10]	0.39		
Number-line Task	-0.05 [-.21, .11]	0.20* [.04, .35]	0.02 [-.14, .18]	0.49	
Math Anxiety	-0.08 [-.24, .08]	-0.05 [-.21, .11]	-0.01 [-.17, .15]	0.21 [-.05, .36]	1.08

Note. *S*tandard deviations are displayed on the diagonal and correlations [95% confidence intervals], below the diagonal. Confidence intervals were calculated using Fisher's *z*-transformation. Correlations marked with an asterisk (*) were statistically significant at $p < .05$, as determined by 95% confidence intervals not bounding zero.

Table 6. *Affected-status agreement of individuals in a given class being concordant in that class with their co-twin*

Class	*n* For Class Membership[a]	Affected-status agreement	Confidence Interval
1	6/12	.33	.12, .55
2	0/17	.00	.00, .00
3	4/25	.14	.01, .26
4	224/71	.76	.71, .81
5	68/81	.46	.38, .54

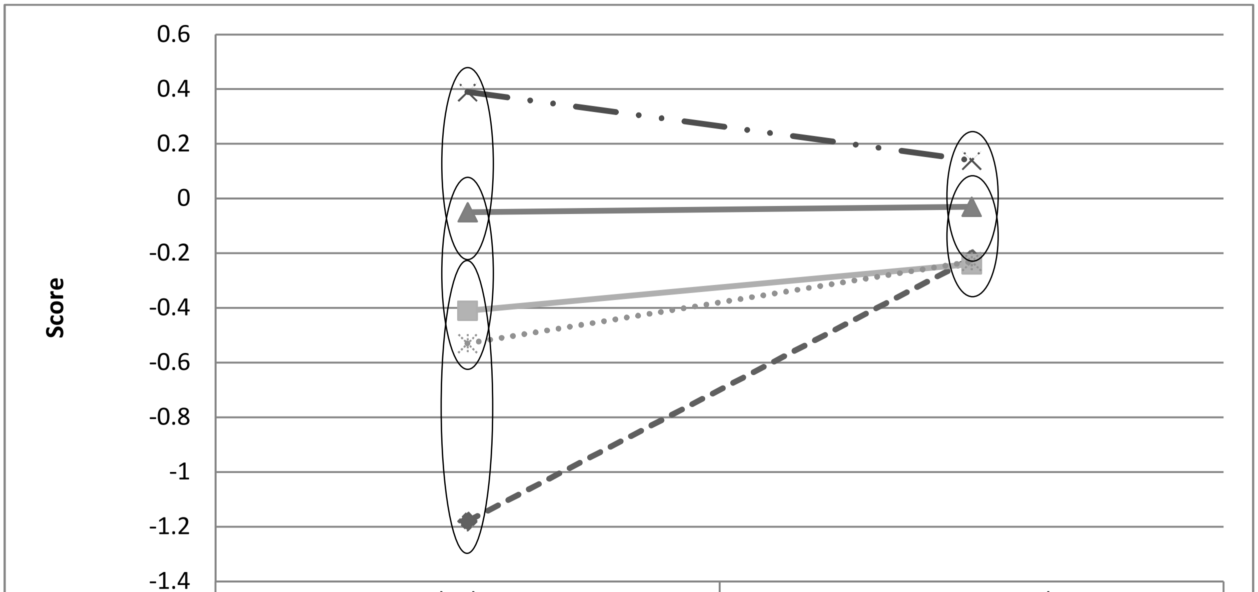
Note. [a] (*n* for twins concordant for class membership)/(*n* for twins discordant for class membership).



	Calculation	Applied Problems	Dots Task	Number-line Task	Math Anxiety
Class 1 n=18	-1.43	-1.74	-1.55	-0.24	0.26
Class 2 n=17	-0.35	-0.78	0.32	-2.20	0.28
Class 3 n=29	-0.13	-0.05	-1.33	0.35	-0.30
Class 4 n=295	0.57	0.59	0.34	0.37	-0.33
Class 5 n=149	-0.72	-0.74	0.07	-0.01	0.57

Figure 1. The final Latent Profile Analysis results, including the sample size of each class and mean *z*-scores. Higher values for math anxiety indicate higher levels of math anxiety.

Note. Means in different ellipses are statistically significantly different from one another.



	Math Fluency	Interest in Math
Class 1	-1.18	-0.22
Class 2	-0.41	-0.24
Class 3	-0.05	-0.03
Class 4	0.39	0.14
Class 5	-0.53	-0.23

Figure 2. Class membership relations to external measures of math Fluency and Interest in Math (*z*-scored).

Note. Means in different ellipses are statistically significantly different from one another.

Supplementary Table 1. *Model fit indices for the ten tested models, controlling for the family-level clustering of the data.*

	LL	Free Parameters	BIC	VLMR *p*-value	Entropy
2 classes	-2656.16	16	5344.31	.34	.76
3 classes	-2583.19	22	5210.69	.55	.82
4 classes	-2514.53	28	5201.07	.34	.84
5 classes	**-2493.07**	**34**	**5195.27**	**.33**	**.85**
6 classes	-2470.02	40	5186.06	.48	.80
7 classes	-2442.50	46	5167.93	.66	.83
8 classes	-2425.08	52	5169.99	.50	.84
9 classes	-2411.59	58	5179.91	.47	.82
10 classes	-2396.84	64	5187.32	.51	.82
11 classes	-2385.39	70	5201.32	.49	.83

Note. Bold indicates best fitting model chosen. As Tech14 (the BLRT) is not available using the clustering option in Mplus, here the Vuong-Lo-Mendall-Rubin likelihood difference test (VLMR) is reported (Lo, Mendell & Rubin, 2001).